On the Governance of Federated Platforms

Pedro Bustamante[1], Marcela Gomez, Prashant Krishnamurthy, Michael J. Madison, Ilia Murtazashvili, Balaji Palanisamy, Ali F. Palida, Martin Weiss[2]

[1]Carnegie Mellon University, Pittsburgh, PA 15213
[2]University of Pittsburgh, Pittsburgh, PA 15260

TRPC 51: The 51st Annual Telecommunications Policy Research Conference

Abstract We consider the governance of federated systems (interacting autonomous platforms) associated with social media. A salient feature is that the individual platforms (or "servers") may operate under different "acceptable use" principles (moderation, content types, and user match with these are parameters of interest). These principles may be articulated in advance for each server and are enforced through content moderation. We analyze these through the lens of Ostromian notions of polycentrism as well as with the GKC framework. Toward this analysis we use data from Mastodon and an agent-based simulation of platforms using NetLogo. The data analysis shows that the number of instances or servers has plateaued in Mastodon despite the backlash against Twitter. Our agent-based model provides us with the observation that there is little churn in users once they settle on a platform. We discuss several ways to extend our model of federated systems as well as lay a groundwork for comparisons of federated and centralized social media systems.

Keywords federated systems, governance, polycentricity, Governing Knowledge Commons

1 Introduction

Online social networks (OSNs) are impersonal relations over the Internet. These social networks are sometimes referred to as the new public or town square. They are also subject to substantial public debate. The biggest OSNs – Facebook, Twitter, Instagram – typically have a highly corporate business model. Partly because of these business models, concerns have been raised about data privacy and data monetization, as user data is used for advertising, as well as security and concerns about moderation. An alternative framework for social media governance are Decentralized Online Social Networks (DOSNs). The defining features of these decentralized social networks is that they rely on open-source software (anyone can set up a server) and communication protocols that enable interconnections among servers that embrace the same protocol. In DOSNs, users can set up their own server (instance) and interact with users of other servers (instances). Our focus is on the governance of these decentralized Internet communities, collectively referred to as the Fediverse.

The Fediverse is a shorthand for describing the federated technology systems that have existed in various forms over the years, including recent advances in federated machine learning for performance and privacy reasons. The recent attention to tech-platform dominance as well as the ownership change in Twitter has brought increased attention to federated microblogging services like Mastodon [1] [2]. We can imagine that businesses and organizations in the future Metaverse will also use a federated approach. By "federated," we mean a set of self-governing units that also interoperate. Over the past year, there has been substantial attention paid to the Fediverse, a term that suggests a conglomeration of self-governing interoperating platforms, at this time primarily for microblogging. A second example is that of interacting communities, with sensors and the data they generate with the Internet of Things, for sustainability, local good, and community benefits. A final example of a generalized federated platform would be several businesses in the same "realm"[1] creating a unit of their own (e.g., landscaping services) interoperating with suppliers of landscaping material, gig workers, homeowners, etc., each with their own platforms.

The Fediverse gained considerable attention following Elon Musk's acquisition of Twitter in October 2022 as a significant fraction of Twitter users sought alternatives to this centralized media platform. At the time, the platform that was most heavily regarded as an alternative was Mastodon [1], a federated system that had many similarities with Twitter, though also significant differences. Threads, owned by Facebook's parent company, Meta, is a centralized system that has recently emerged as a centralized competitor to Twitter.

Mastodon is comparable to Twitter and Reddit, though Mastodon communities are independent; Mastodon provides content warnings; and users can interact even if they belong to different servers because of a client-to-server and server-to-server communication protocol. A consequence is an extensive followership mechanism and three timelines: a home timeline (toots generated by followed users); a local timeline (toots yielded within the same instance); and a federated timeline (all public toots from all users known to the instance where the user is registered). Mastodon also allows users to apply rules and policies on generated content, with administrators declaring both main topics of their instances and prohibited content as well as place content warnings or spoilers that are displayed before it is viewed or not, as well as closing registration for instances (private instances), though that does not prevent user interactions as allowed above which are governed by a specific protocol.

There are also significant puzzles, which we describe and explore below using *governance* as our umbrella concept. Why do some OSNs and federated systems thrive, and others fail? Why do some succeed at small scales but not at larger ones? What other dilemmas and associated risks do these systems face, and how are they overcome? Mastodon has thousands of servers, yet only a few of the biggest ones attain scale. Studies of Free/Libre and Open-Source Software (FLOSS) suggested that while open-source software systems contribute to innovation, they often get swallowed up by larger platforms and like any enterprise, they sometimes fail despite initial promise. Given the substantial public discourse surrounding Twitter, or X as it is now known, consideration of the sustainability of federated platforms such as Mastodon is a question with significant social, economic, political, and legal implications.

Governance in federated systems refers to the overarching framework of formal and informal rules, norms, processes, and structures that guide the management, sharing, and utilization of data across the entities of the system. It includes establishing roles, responsibilities, and decision-making mechanisms for participants and members to ensure effective coordination and collaboration among the participating entities. In general, governance frameworks in federated information systems can encompass policies, standards, and protocols that address data ownership, access rights, security, privacy, conflict resolution, and, where applicable, compliance with regulatory requirements.

In this paper, we analyze governance primarily in terms of content moderation. A salient feature is that the individual platforms (or "servers") may operate under different "acceptable use" principles. These

[1] The term "realm" has been used in the federated version of the Kerberos security protocol.

principles may be articulated in advance for each server and are enforced through content moderation. These will be analyzed through the lens of Ostromian notions of polycentrism as well as with the GKC framework.

2 What are Federated Systems?

Federated networks consist of different communities on different servers that interoperate with each other, rather than existing on a single platform. Early attempts include FLOSS communities, as well as open-source and open-code communities. Mastodon combines these features with social media. The Fediverse consists of millions of accounts distributed over thousands of servers (instances). In the federation model, servers communicate through the same protocol so that participants can interact with other servers (hence a massive social network, called the Fediverse). Users can follow other accounts without needing accounts of their own. Dimensions include ethics, organizational structure, underlying technologies, features, source code access, or special interest communities they support. A shared feature is that they provide an alternative to centralized social media, as well as decentralization and interoperability.

Federated systems, also known as federated architectures or federated networks, are decentralized systems in which multiple independent entities or organizations collaborate to provide a unified service or platform. These entities, often referred to as nodes, servers, or instances, retain their autonomy and control over their respective parts of the system while interconnecting with other nodes to enable seamless communication and sharing of resources using a shared protocol. The primary feature is that there is no centralized entity that controls all the platforms or services, and there is autonomy (to an extent) on how each platform behaves.

There are several examples of federated systems in use today. We briefly describe two such systems here.

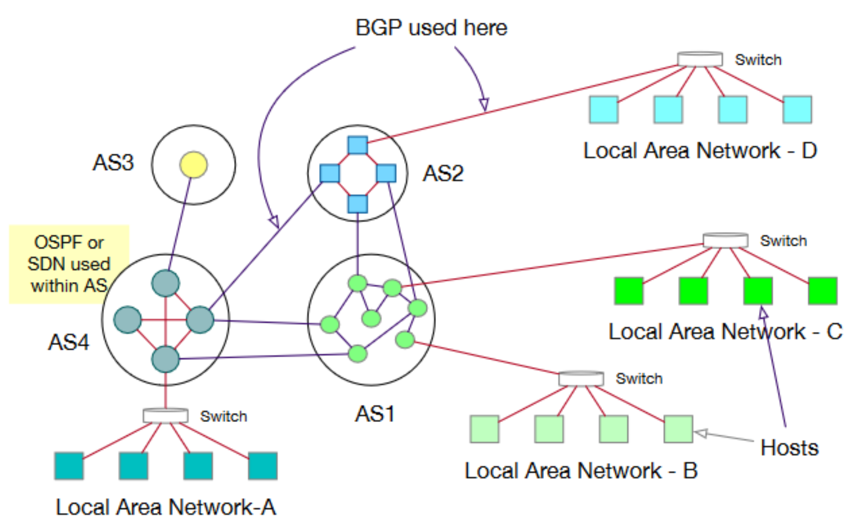

Figure 1: Simplified Schematic of the Internet

The Internet itself is a federated system comprising autonomous systems that work with each other to relay packets from a source to a destination (see Figure 1). Within an autonomous system (AS), controlled by one administrative entity, routing protocols such as Open Shortest Path First (OSPF) or Routing Internet Protocol (RIP) are used to move packets between routers (e.g., green circles in AS1) toward their destinations. Each autonomous system could use different routing protocols within itself, and in recent years, could also implement a logically centralized software-defined networking (SDN) approach to routing packets. The Border Gateway Protocol (BGP) is universally accepted by these autonomous systems for advertising access to different networks and finding paths on the global Internet. BGP is the "glue" that guarantees the interconnection of networks of the Internet. In other words, local administrative autonomy, and the ability to work with other autonomous groups is a hallmark of the Internet. One may even view this architecture as polycentric when nation-states exert control over the networks in their geographical domain.

A second example is the use of federated single sign-on where service providers (SPs) work with identity providers (IDPs) to authenticate users and provide services to them. The users are registered with IDPs such as Google, Meta, or Apple and authenticate with them when they need services from an SP. This way, users do not have to have separate logins and passwords with each SP. Depending on the sharing agreements between SPs and IDPs, the degree of privacy and tracking of users may vary. IDPs however always have information about user behavior. The standards, such as Security Assertion Markup Language (SAML) allow interoperability between SPs and IDPs.

We can infer certain characteristics from these two examples that we will elaborate on later, namely (i) local autonomy of systems (ii) independence for end users (humans, organizations, etc.) to join multiple systems – although there may be heavy preferences for one such system (iii) possibility of churn (users and organizations may change service providers or platforms) and (iv) standards for interoperability and acceptable behavior between locally autonomous systems.

3 Federated Social Media Platforms

The concept of federated systems presents an alternative approach to the traditional centralized model employed by popular social media platforms. In a federated social media network, individual instances, operated by different organizations or individuals, form a network by interconnecting with each other. Users to create profiles, share content, and interact with others within an instance; however, users from different instances can still communicate and connect with each other across the federated network. The two major purposes of such social media networks are (i) to gain influence and mindshare through followers and (ii) to disseminate content widely.

3.1 Architecture and Attributes

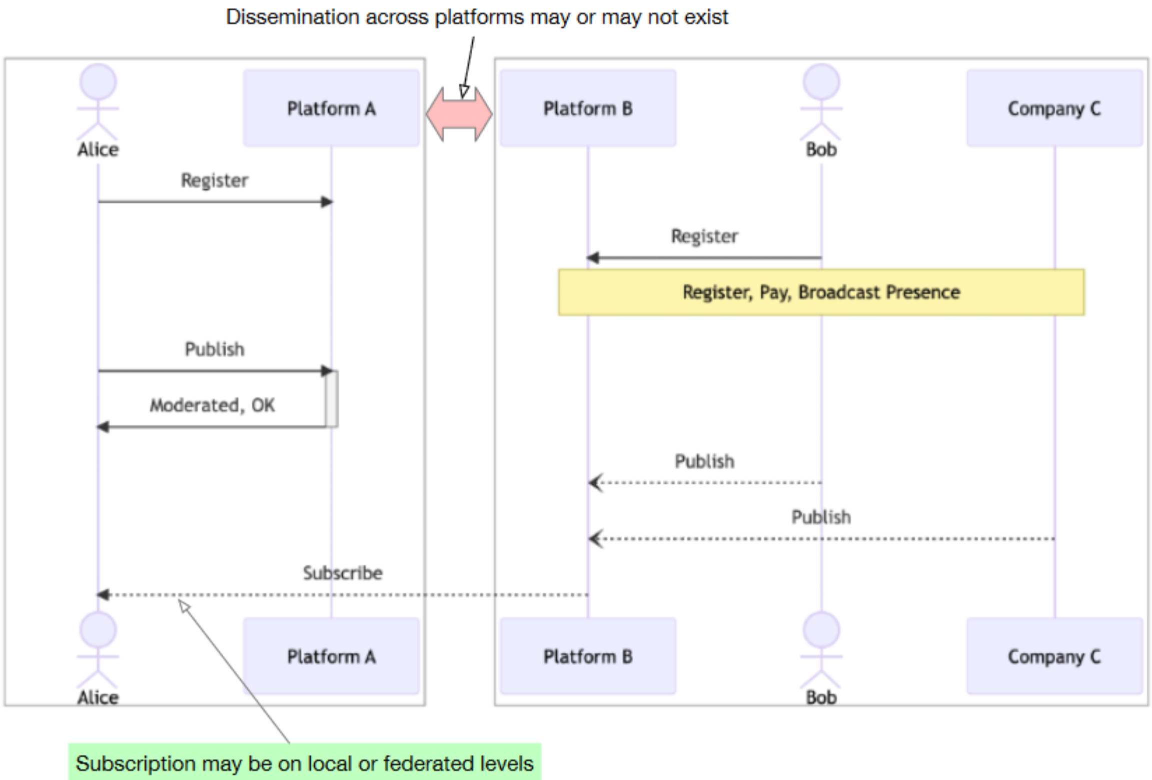

Figure 2: Simplified schematic of a Fediverse with two Platforms

The general architecture of a federated social media network is shown in Figure 2. There are platforms (also called nodes, servers, or instances – which could be in principle distributed on the Internet) that are "online" for the most part. Such platforms act as "brokers" in the publish-scribe models [3]. On the left-hand side of Figure 2, Alice, a user, registers with Platform A where she posts her content. Platform A may have some rules for moderating content. Such rules may moderate content before it is published or moderated later. There may be human intervention, keyword-based moderation, or the use of Artificial Intelligence (AI) to moderate content. Platform A may have mechanisms to allow users registered with other Platforms to subscribe to content and rules for such interactions.

Alice also subscribes to feeds of others, say Bob or Company C. Bob and Company C (on the right-hand side of Figure 1) are registered with Platform B. Bob is simply a user like Alice, but Company C may be a paid subscriber to Platform B. It has the advantages of verification (and perhaps a blue badge like that in Twitter) and perhaps better dissemination of its content to subscribers (priority, latency, advertisements, etc.).

Size: In general, the Fediverse has multiple platforms of multiple sizes. Size here could mean a variety of parameters – hosting and dissemination capacity (storage, servers, network access, availability, etc.), the number and types of participants (paid, unpaid, followers, content creators, etc.), and the vibrancy of the ecosystem (how frequently the platforms are being used and the engagement of the participants).

Content: Platforms may support publishing different types of data. Mastodon [2] typically supports microblogging (short textual content) while others such as PeerTube [4] support videos, news aggregation, images, file hosting, or longer articles.

Resilience: One aspect of federated systems is that platforms may appear or disappear without notice (user churn is a common characteristic of many social media platforms). With multiple platforms, it is less likely that content will not be available or the system will fail to serve the community. Users may be able to migrate to a different platform if their current platform does not serve their interests, and purposes, or fails due to economic reasons. However, depending on the how content may have been published, it may be available only on a single platform, which may still lead to failures.

Matching and Behavior: The reason for a user to join a platform may be based on interest or governance aspects (see below). It is possible that Alice picked platform A because it was the first one of which she was aware, because it matched her interests (e.g., the platform included discussion topics such as technology or radio spectrum), or because this was where her social network was present (e.g., many friends and coworkers are registered with the server). As one example, in [2], the authors investigate the behavior of users in Mastodon and discover that there are strong linguistic, cultural, and geographical affinity go platforms.

Governance: The governance (e.g., identity management and user verification, moderation, rules for federation, advantages to paid users, etc.) aspects are another characteristic of the Fediverse platforms. We have a broader discussion of the governance aspects in Section 4, but provide a few aspects of governance here. These are some of the aspects of the Fediverse that could further impact the association of users to specific platforms. The key feature of federated systems is that different platforms may have different rules or standards of behavior and appropriateness.

Identity management and user verification: Platforms may have varying rules for managing users. Users may or may not be verified as humans (for example, [2] reports that there is a 5% or so presence of bots in Mastodon). Even if verified as humans (through e-mails, captchas, etc.), they may adopt personas that do not exist in real life. Alternatively, like Meta's Facebook, platforms may require only humans to register with their real-world identities. Some platforms may require the authentication of users through public keys or a blockchain-like mechanism. That is, a human user is associated with a public-key private-key pair, and these have to be used to post content. In this way, there is some assurance that the entity posting content corresponds to the actual person who is supposed to be posting the content.

Moderation: Social media platforms have policies and rules on what content is allowed and what is not. These rules are used to "moderate" the content posted by subscribers. Such policies might influence the type of content that is allowed, the latency in publishing content, and the creation of communities with specific topics of interest. Moderation may be accomplished in a variety of ways, using AI, keywords, human approval, blacklists of users, etc. There are extremes in moderation possible. The social network Gab [5] has little moderation and has attracted many fringe users and allows hate speech, while Twitter, during the 2020 election period heavily moderated tweets that might have unverified information. Mozilla.social, a platform (or instance) on the Mastodon Fediverse has a strict set of moderation policies [6] that do not allow misinformation or hate speech in addition to content that has criminal activity in it. It also forbids impersonation or violations of trademarks and copyrights.

Federation rules: The previous paragraph briefly discussed moderation *within* a platform. Moderation across platforms is possible as part of federation rules. Federation rules are how platforms interact with each other. In some protocols like NOSTR (discussed in Section 3.2), the platforms (called relays) do not talk with each other. It is entirely up to the clients (or users) to publish and subscribe to multiple platforms based on their preferences. In others, such as Mastodon, individual platforms have dissociated with Gab as they do not agree with the content policies of Gab [7]. This isolation of Gab from other Mastodon servers is achieved through a process known as "silencing" or "blocking." When an instance decides to silence another server, it essentially severs the connection between its users and the other server. In this case, this means that content from Gab is no longer visible to users on the silenced instance, and interactions with Gab users are limited or disabled. One question that arises is whether there are automated ways of “matching” policies across platforms and rules to allow or block content based on these policies. On the Internet, for instance, firewall rules have been used to block specific domains, IP addresses, or protocols (say transport layer port numbers). These rules, in themselves, have become extremely complicated [8], and misconfigurations have led to security problems.

Business Models: In federated systems, the incentive for users to “pay” to join a platform reduces the presence of platforms that do not charge users to join, whether to reach an audience or access content. A strategy to gain users would (at least initially) make a platform as friction-free to join as possible. In fact, barriers such as payment or restrictions on the number or types of content that can be accessed can irk users and cause them to consider alternatives. The recent emergence of Meta’s Threads platform as a serious competitor to Twitter is attributed in part to the restrictions that were increasingly placed on the members of the Twitter community. Size, discussed briefly above, may influence business models. In Internet interconnection, backbone ISPs "peer" with networks of similar size and scope but require "transit" payments from smaller networks. In a similar manner, it may be possible for larger platforms to charge a fee to smaller platforms for access to content. It is possible that a platform that becomes exceeding popular (based on the rich get richer phenomena in network science) eventually monopolizes content and users. In such a situation, we can imagine a situation similar to current Twitter where the barriers of newly creating engagement may force users to bear the inconvenience of paying the large platform for reaching an audience or retrieving content. For example, Threads has seen a significant drop in engagement (by 70%) since the initial surge in users joining the platform [9].

3.2 Interoperability Protocols

For the various platform in a federated system to work together, it is necessary for them to be able to communicate with each other seamlessly. It then becomes necessary for the platforms to adopt a standard for exchanging metadata or other kinds of information related to content, users, policies, etc. There are several protocols that have emerged in recent years to support the operation of federated systems. We describe some of them below.

ActivityPub is a standard developed by the World Wide Web Consortium (W3C) [10]. It allows servers of different platforms to talk with each other, as well as user software (clients) to talk with the servers. Direct messages are possible between users (called Actors in ActivityPub), as well as public messages to all followers. Implementations can have partial functionality (client-only, server-only, or federated server). For example, a client-only implementation can converse with a server-only implementation in a closed system without federation, but extending the functionality is supposed to be straightforward. Messages are in JSON ActivityStreams format making it standards compliant with web technologies. Mastodon and Peertube both use ActivityPub – allowing a mix of microblogging and video in a federated system. ActivityPub has also been used as a protocol for file hosting services.

NOSTR calls itself a “decentralized social network with a chance of working” [11]. Unlike ActivityPub, NOSTR platforms (called relays) do not talk with each other (as shown in Figure 3). The discovery of content and publishers has to happen differently in NOSTR. Since relays do not talk with each other, they can locally enforce moderation and other policies agnostic of other relays.

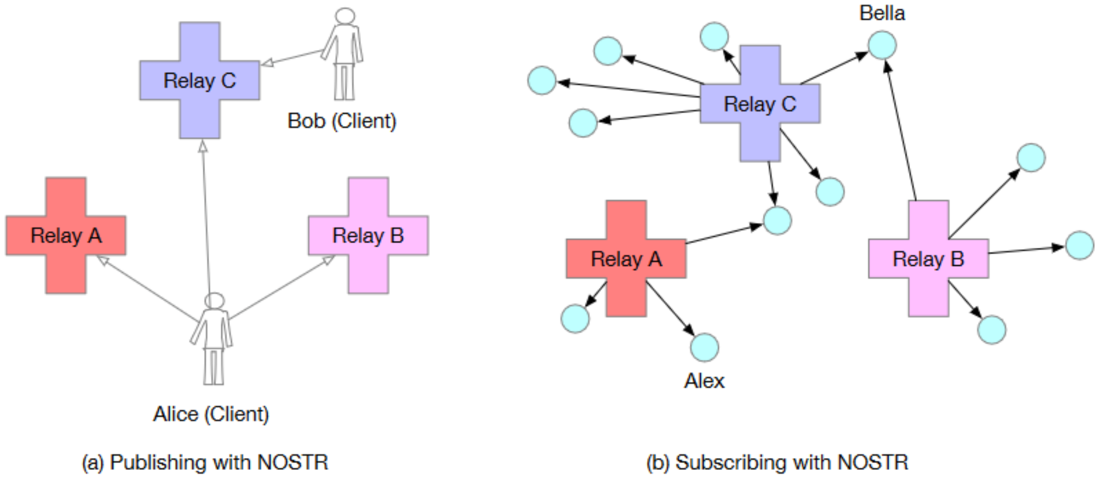

Figure 3: Schematic to illustrate the operation of NOSTR

NOSTR supports many variations in behavior. In Figure 3, relays A and B may be public relays that accept anyone as users. Relay C may be a commercial (paid) relay that is more selective in who can publish. Commercial relays may offer higher availability. To discover content or users, clients may publish a list of relays where they are publishing. Since clients can publish with multiple relays, there is redundancy if one or more relays censor a client or go away.

The **AT** or Authenticated Transfer protocol [12] is being developed by Bluesky, a federated alternative microblogging service. The protocol depends on public key encryption for managing identities and data. The system depends on HTTP-based APIs for interactions between clients and servers as well as interactions between servers of different platforms. Like ActivityPub and unlike NOSTR, servers can converse with other servers.

3.3 Related Work

There have been several recent studies on decentralized online social network platforms. La Cava et. al [13] studies user behavior in Mastodon by analyzing the dichotomy between information consumption and production in the network. In another work, the authors analyze the network of Mastodon instances to study the distinguishing traits of the underlying federative mechanism. Zignani et. al [2] analyzes the social data gathered from Mastodon and shows that the decentralization in the social network leads to a structure different from the one in a typical centralized social network. Guidi et. al [14] analyzes the Steemit online social network platform to show that users are incentivized to produce content even though the richest users are not necessarily the most active ones producing content. A recent by work Li *et al* [15] estimated the monetary value of the online volunteer work completed by all volunteer moderators on Reddit. Following the recent acquisition of Twitter, La Cava et. al [16] study the migration of Twitter users to Mastodon and observed that the temporal trace of their migrations is compatible with a phenomenon of social influence, as described by a compartmental epidemic model of information diffusion. Recent work has also focused on the architecture, infrastructure, and protocols of social media and their impact on transparency, accountability, and fairness. Ermoshina et. al [17] analyzes the role that informational architectures and infrastructures in federated social media platforms play in content moderation processes. The authors particularly analyze the Fediverse as an alternative model for content distribution and moderation. Singh et al [18] discuss how automated tools are being used by major technology companies to shape the content we see and engage with online, and how internet platforms, policymakers, and researchers can promote greater fairness, accountability, and transparency around these algorithmic decision-making practices. Rozenshtein [7] analyzes the specific issue of content moderation on the Fediverse, using Mastodon as a case study to draw out the advantages and disadvantages of the federated content-moderation approach as compared to the currently dominant closed-platform model. Ilik et. al [19] study three World Wide Web Consortium–recommended protocols namely the Linked Data Notifications protocol, the ActivityPub protocol, and the WebSub protocol to create an information-sharing pipeline to provide access to the most up-to-date information to all stakeholders. There has also been recent work on the ethical aspects of decentralized technologies. Allen et. al [20] discuss the ethical issues in decentralized social technologies and a framework for making ethical choices to leverage opportunities for transformation while avoiding the potentially problematic consequences.

Research on governance of decentralized platforms, including the Governing Knowledge Commons (GKC) [21] literature, is also related. These works initially considered open-source technology communities. More recently, they consider blockchains. A shared theme is that openness and decentralization do not necessarily solve governance challenges associated with centralized or closed systems. Governance is still required by virtue of social dilemmas created by openness and decentralization, among other things. Rather than see distributed networks as an "governance-less" alternative, they emphasize that there is governance, and that the diversity of rules explains their relative performance. In this regard, such perspectives are an extension of Ostrom's comparative institutional analysis.

4 A Governance Perspective on the Fediverse

One aspect of the governance of the Internet is the degree of centralization. The Fediverse represents a shift from centralized (and oftentimes gigantic) social networks to smaller, interconnected instances. It shares features with the political theory of federations in which networks of various political actors decide to cooperate collectively, with substantial autonomy among various units. Both federations and the Fediverse have a centralized structure: federations involve a central government, with local government autonomy, while the Fediverse has a central protocol. The shared feature is that the local units – local governments or servers – enjoy substantial autonomy.

Polycentric systems recognize a role for local autonomy that enables adjustments that facilitate experimentation, especially for collective goods subject to rivalry to an extent [22]. In the context of technology, this is often described from the perspective of the Governing Knowledge Commons (GKC) research framework, a perspective which recognizes that communities integrated through shared technology generate governance dilemmas. Those dilemmas are primarily associated with the fact that knowledge, information, and data resources are abundant and yet confront challenges with management that are akin to dissipation of the resource. Further, as a descriptive endeavor, many of the systems to manage knowledge, especially Internet communities, are decentralized. As well, regulation of these systems, including the Internet, tends to be polycentric, with emphasis on collaborative, multi-stakeholder governance that can avoid high costs of adversarial policymaking as well as foster synergy among stakeholders [23].

Another aspect of governance concerns relations in the Fediverse itself. Governance of emergent technologies is sometimes seen as mainly spontaneous and lacking in the conventional issues normally associated with governance dilemmas. Indeed, the anarchist view of technology as disruptive tends to see technology as enabling freedom from governance challenges. Over time, this initial anarchist view has given way to a belief that these decentralized technologies require governance, or self-governance. The GKC framework is especially useful in this regard. The GKC framework is a method of researching resources that are the produces of the human mind, including knowledge and information in scientific domains, domains related to arts and culture, and resource domains defined largely by their human-generated character. These "constructed cultural commons" require governance, which refers to groups of people or communities with shared access or use of resources who manage their behavior via an established set of formal and informal rules and norms. One of the defining features is the institutionalized sharing of resources among community members.

Some examples include FLOSS, as mentioned earlier, which is an open-source community that largely relies on self-governance. Blockchain networks are another example. Blockchains are distributed networks. Still, they depend on an involved governance [24]. They also have the feature of a knowledge commons in that their use depends on freely available uses. Given these shared resources, a critical question is how they are governed, both internally (among users within a network) and their external relations.

Federated platforms, from this perspective, have the critical feature of a knowledge commons. These social networks are domains of social interaction, including arts and culture. They are human-generated, or user-generated. Though centralized networks are also human-generated, the focus on users is especially a feature of knowledge commons. They are also defined by the shared use of a resource. That resource might be the protocol or network. It is shared, though sharing does not imply the absence of rules. The extent to which federated systems like Mastodon can define membership (private networks) is an example of such sharing of resources with the presence of boundaries.

Another feature of the GKC perspective is that analysis begins by consideration of the specific community under consideration. In this regard, research questions, while informed by some general ideas about governance of decentralized communities, tailors analysis to the specific community under consideration. In the case of federated systems, some of these questions suggested by a GKC perspective include the following:

1. How do governance aspects (e.g., moderation), influence the size and movement among instances?
2. How does an increase in the number of instances in a federated system influence its performance?
3. How is the discovery process of federated systems characterized?

Each of these questions raises significant questions about the governance of the Fediverse. The first question on governance aspects reflects that users not only choose federated versus centralized systems based on governance considerations but that within the Fediverse, governance varies substantially.

The second question recognizes that there may be an optimal size of federated systems, where "size" can refer to the number of instances people interact with. Unlike a political federation, which is typically stable and with relatively few interacting units, the Fediverse is open-ended. That open-endedness provides for choice but may also contribute to information overload that makes choice challenging. This has been analyzed from the Governing Knowledge Commons perspective [21], which was developed in part to analyze open-source networks such as FLOSS.

The third question involves the ways in which users find opportunities in the Fediverse. What is clear is that the Fediverse creates the possibility of many communities shared with a protocol. What is less clear is how users discover what is best.

Our model focuses on the first aspect, with an emphasis on moderation. It illustrates some of the dynamics of federated systems as well as serves as a basis for extensions, which we also consider. Our model can also be used to address the second and third questions. One approach is to use the GKC perspective [25] to analyze these systems.

5 Modeling Federated Systems

We built a preliminary agent-based model using NetLogo to simulate the behavior of participants and the impact that governance of platforms might have on the evolution of this behavior. Toward this, we first conducted an exploratory data analysis of Mastodon by obtaining data from the statistics API that is facilitated by Mastodon. In other words, all the data is reported by Mastodon itself. Next, we conduct a simulation using agent-based modeling.

5.1 Data Analysis

The data analysis of Mastodon is composed of two parts. We begin with a general analysis of the evolution of the number of registered users, active users, and servers in the platform. Then, we examine server characteristics. For this second part, we focused the analysis on a smaller subset of the current ~10K servers. The sample is 500 servers (i.e., instances in Mastodon) and we extrapolate the results to the 10k server population given that the sample consists of the largest servers.

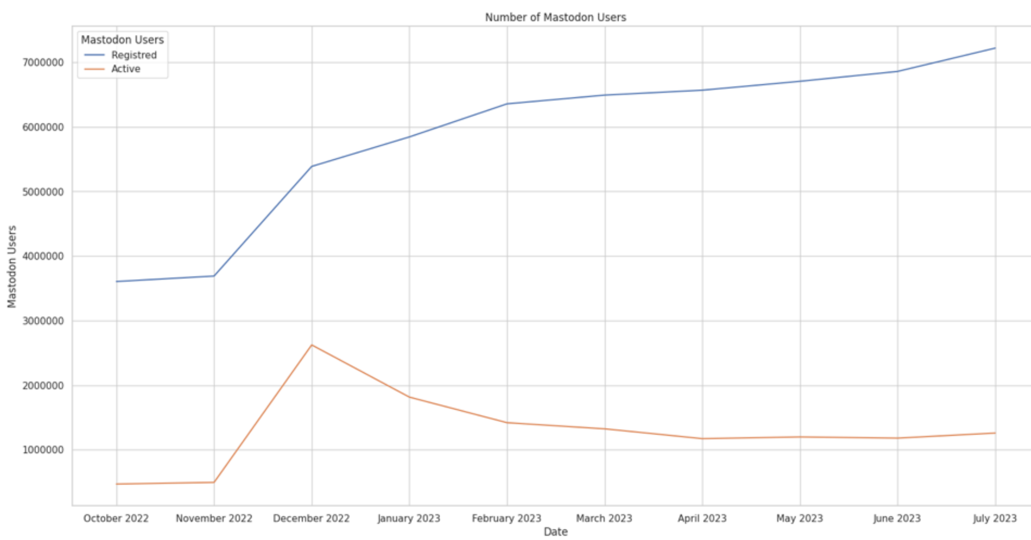

Figure 4: Number of Active and registered users in the Mastodon Fediverse

Figure 4 shows the evolution of the platform's registered and active users beginning in October 2022. We note that the number of registered users has continuously increased in the last year. This increase coincides with Elon Musk buying Twitter and a lot of users trying to find an alternative (Mastodon gained around 3 million new users in about three months). On the other hand, we can see a similar increase in the number of active users around Nov/Dec 2022, but it has not been sustained. The number of active users has fallen since the "twitter-migration". Although recently the number of active users has remained constant around a million users, it has not followed the same trend as the number of registered users. This is consistent with other studies regarding the initial popularity of the platform as a twitter-like alternative, but without the intervention of a centralized entity.

Another important component of any Fediverse, including Mastodon, is the number of servers. In Mastodon, a user can select one or more servers, known as instances, to create an account. And as we detailed in the previous sections, a user can connect to other servers using the available communication mechanisms such as ActivityPub. It is worth mentioning that each server (instances in Mastodon) can define its own rules and moderation policies. In the case of the servers, we see a similar behavior to the number of active users, where a lot of new servers were created when Twitter was experiencing the ownership change. Nevertheless, the number of active servers has been steadily decreasing since its peak around March 2023. The current number of active servers (as of July 2023) is around 9k (shown in Figure 5).

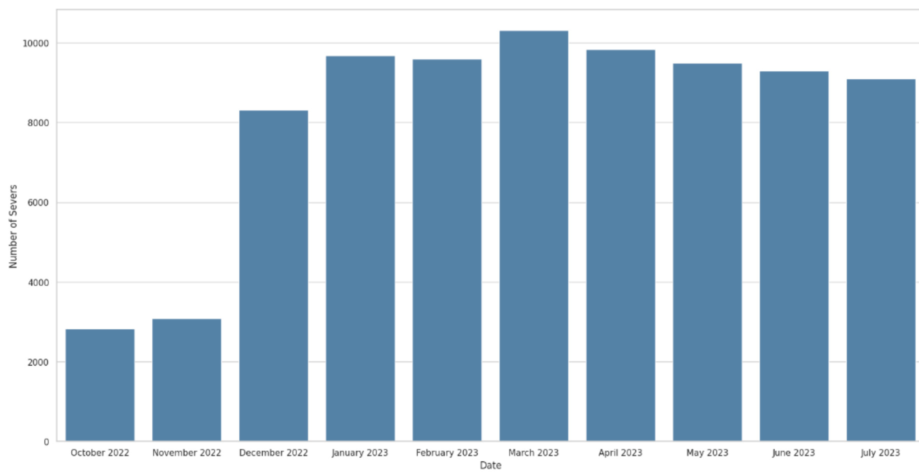

Figure 5: Number of servers (instances) in the Mastodon Fediverse

The second part of the Mastodon Data Analysis involves a sample of 500 servers (instances). This includes popular instances such as *mastodon.social* and *pawoo.net*[2]. We were interested in two important things regarding the individual servers in the Mastodon Fediverse, namely the user distribution (including active and registered users) and the rule distribution.

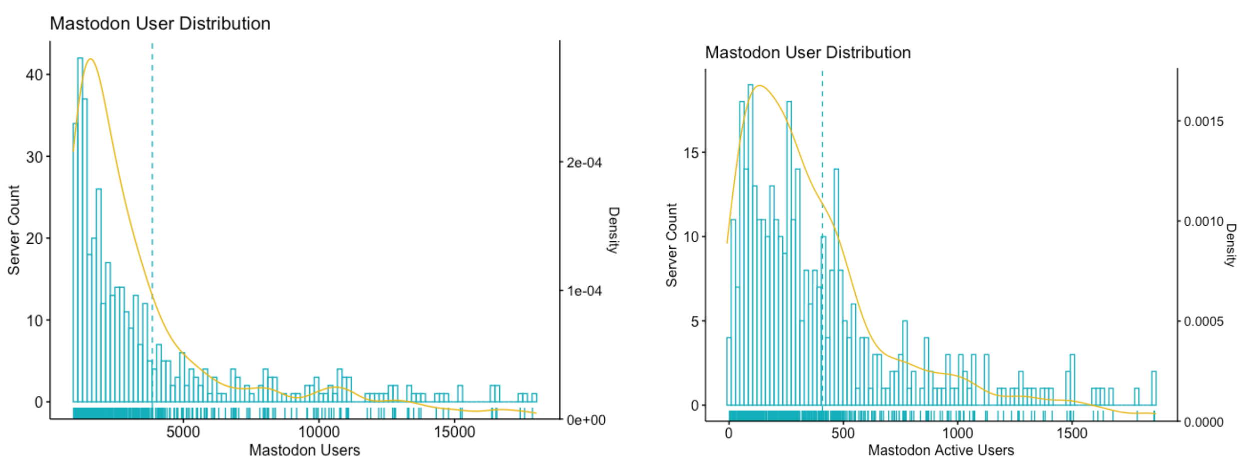

Figure 6: User Distribution in the Mastodon's Servers (instances)

The graph on the left of Figure 6 shows a histogram/density of the (registered) users in each of the instances (servers) at Mastodon. It is important to note that this graph excludes the outliers that tend to have more than 20,000 users (e.g., *mastodon.social*) as these produce a distorted graph that is difficult to interpret. As

[2] Each of these servers have more than 1 million users.

you can see in the graph, the average number of users (dashed blue line) in each instance is less than 5,000 registered users. In fact, most servers have less than 10,000 registered users. The graph on the right of Figure 6 also shows a histogram/density plot, but it shows the distribution of the active users in the different servers (instances). In the same light as the previous distribution, this graph excludes servers that are considered outliers with thousands of active users. This graph follows the same pattern as registered users, where most instances have less than 500 active users.

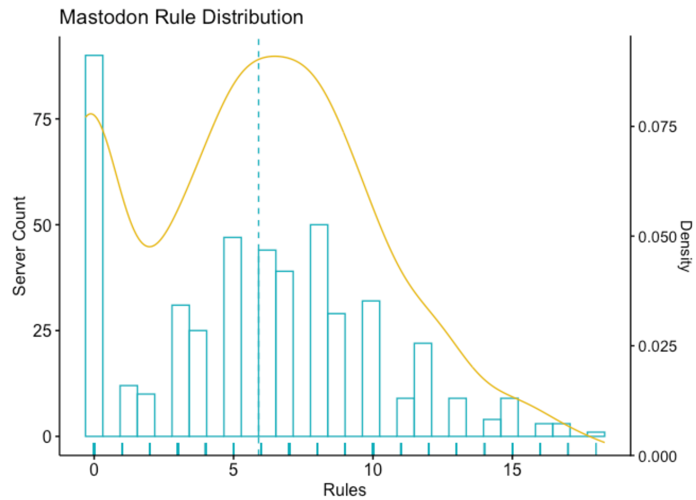

Figure 7: Rule Distribution in Mastodon's Servers (instances)

As we can observe, both the number of registered and active users in Mastodon's servers follow a power-law-like distribution. In other words, a small number of users is clustered at the top of a distribution. This implies that a small number of servers are in fact the "gatekeepers" for most users in the Mastodon Fediverse.

The final part of the analysis is the *Rule* analysis. This is a central feature of the Fediverse: users can choose servers that best match their preferences; we also captured this in the NetLogo model, where we assume that instances with more rules might be more prone to moderation, at least ex-ante when a user is deciding to join a server. Each server defines a set of rules that the user accepts before joining a particular server. Most of these rules are similar to those of other social network platforms. For instance *mastodon.social* (the most popular server in the Mastodon Fediverse) has seven (7) rules:

- *"Sexually explicit or violent media must be marked as sensitive when posting*
- *No racism, sexism, homophobia, transphobia, xenophobia, or casteism*
- *No incitement of violence or promotion of violent ideologies*
- *No harassment, dogpiling or doxing of other users*
- *Do not share intentionally false or misleading information"*

Figure 7 shows the histogram/density of the number of rules in each server. In the same way as before, this graph excludes the outliers. While counting the number of rules does not address the diversity of rules, it is suggestive of variations in rules of different servers. Most servers have around 6-7 rules, which tend to be the same rules as *mastodon.social*. However, there are several servers that have no rules. We believe this is very interesting and can lead to a lot of problems in the platform. On the other hand, there are a number of instances with a larger number of rules that try to avoid many of the issues regarding the lack of moderation.

5.2 Agent-Based Model of Fediverse Structures

To dive deeper into the workings of Fediverse systems, we develop an agent-based model (ABM) that captures essential elements of moderation in the system. In our model, we define two main types of agents, namely Fediverse users and Fediverse servers. Servers replicate the behavior of decentralized Fediverse platforms such as Mastodon. Each server defines a set of rules that will apply to each user joining its network. In addition, each server has a moderation policy to enforce the agreed-upon rules. It is worth noting that the number of rules tends to remain constant within each server, while the moderation activities and policies are influenced by multiple factors (e.g., number of users, moderation mechanisms, user activity, etc.). In the initial version of our model, both the number of rules and the level of moderation remain constant throughout the different agent simulations. In addition, we define a "spectrum" of servers, in which each server selects its own number of rules and moderation policies based on a min-max scale. The second type of agents in our model are Fediverse users. Each user is willing to join the Fediverse platform through one of the available servers based on their moderation preferences – each user has a moderation preference based on a min-max scale. As previously explained, there are multiple reasons for a user to join a particular server (i.e., the user's social network preferences) in a real-world Fediverse platform. In our model, we assume that all servers have similar social characteristics, hence, all the servers in the model discuss similar topics, have similar social network characteristics, have the same language, etc. From the perspective of each user, the main difference among servers is their number of rules and moderation policies related to these rules.

In the setup stage of our model, several servers and users are created in the ABM. Each user agent has a moderation preference, and each server agent has an initial set of rules and a moderation policy. Before joining a server, each user only knows the number of rules in each server as the moderation policies are difficult to infer without joining a server. In this way, we assume that users who prefer higher moderation, will prefer a higher number of rules. As we can observe in Section 5.1, popular platforms in the Fediverse fall into the class of scale-free networks, meaning that they have power-law (or scale-free) degree distributions. In this light, to model a user joining a server we used a modified version of the Barabási–Albert Preferential Attachment network model [26]. In our model, users join a server with a probability that is proportional to the number of rules that the existing servers have. This behavior allows us to match users with similar moderation preferences to servers with similar moderation policies.

Once the initial setup of the model is complete, users join the Fediverse through their preferred server according to moderation levels. Even though the number of rules can be a good indicator of the moderation policies of a server, moderation in any social system goes beyond the ex-ante set of rules. We also consider this in our model by including a moderation factor for each server defined as a min-max scale. The moderation factor reflects the moderation policies implemented by each server, which might be based on multiple factors such as the number of registered and active users. Nevertheless, a given user discovers these moderation policies only after they have joined a particular server.

After the initial setup, the agents are free to interact in the Fediverse. In our model, users might not be *happy* with the moderation policies of their server and might decide to join a different server in the network. We also capture this in our model using a similar mechanism as the one used in Schelling's model [27] of segregation. In our model, users seek another server when there is a significant difference between their moderation preference and the server's moderation policy. To avoid the possibility of infinite loops in our model and reflect real-world user behavior, after a number $\left(\frac{NumServers}{2} + 1\right)$ of server-changes, users leave the network since they seem unable to find a server that matches their moderation preferences.

5.3 Experiment Setup and Analysis

To capture all possible combinations of the factors in the system, we implement a Full Factorial Experimental Design. We select two main variables as the inputs in our experiments. We select the initial number of users and servers in the Fediverse. In addition, we define the limits for the different user and server variables in the model (see Table 1). From the variables and levels in our experimental setup, we define various scenarios. First, we simulate scenarios with a single server to emulate a centralized system (e.g., Twitter and Threads) and compare them to traditional Fediverse structures with multiple servers. We also simulate scenarios where the number of users is greater than the number of servers, and vice versa. To evaluate the results of our experiments, we have selected three important characteristics of the network: the number of users that left the Fediverse, the degree distribution of the servers, and the number of users that are not able to find a server within the Fediverse. We believe that these outcomes best summarize the interactions of the agents and macro-behavior we aim to study.

Experiment Setup	
Variables	**Levels**
Number of Fediverse Users	[10, 100, 1000]
Number of Fediverse Servers	[1, 10, 100]
Model Outcomes	
Number of Fediverse Users that left the platform	
Average server degree distribution	
Number of server changes (in each interaction)	
Number of Fediverse users without a server	

Table 1: Experimental Setup of the ABM simulations

In what follows, we present the results obtained for each of the network metrics presented above.

Users leaving the platform: In our model, after a user has switched servers a predefined number of times, – calculated as number of servers / 2 + 1 – if they have not found a suitable server, they would leave the platform. This allows us to evaluate how easy it is for users to find a server they will be interested in joining, and more importantly, remain connected to that server. We find that in scenarios with fewer servers, a larger number of users leave the platform after the period predefined in the model. Indeed, between 50% and 80% of users leave the platform when there is only one server (see Figure 8). As the number of servers increases, scenarios with a larger number of users show higher leaving rates than the scenario with only 10 users; however, in scenarios with 100 and 1000 users, the leaving rate is ~45% and ~25% for cases with 10 and 100 servers, respectively.

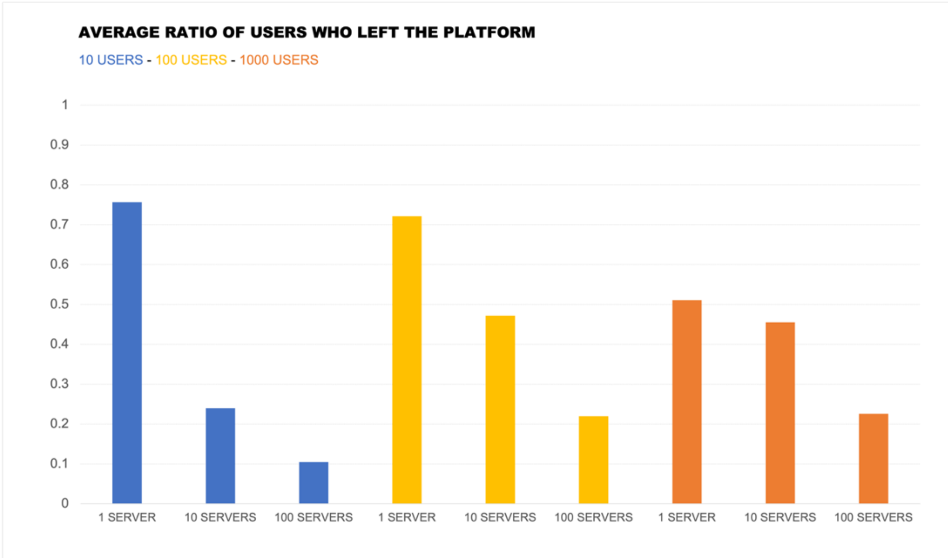

Figure 8: Average ratio of users leaving the platform after a predefined number of unsuccessful attempts to find a suitable servers.

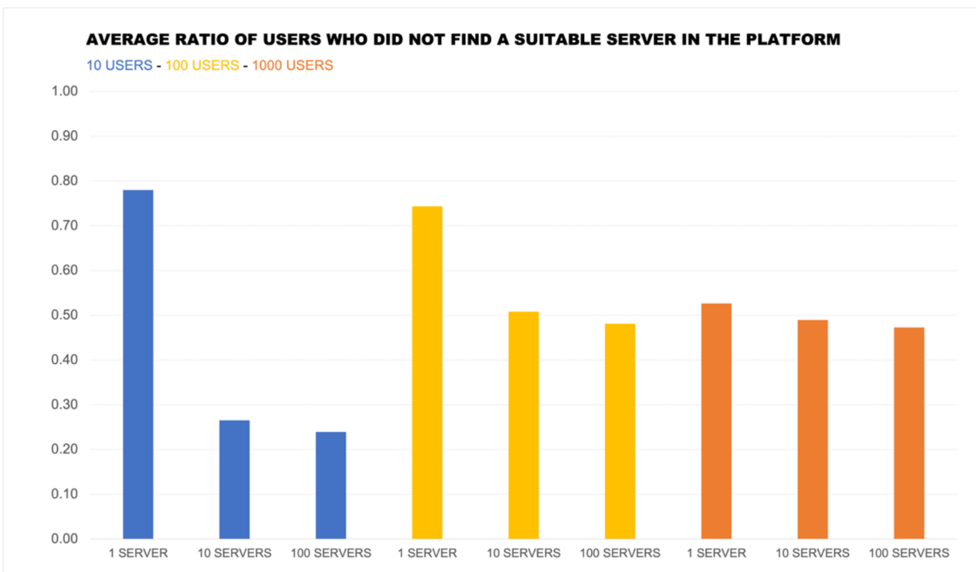

Figure 9: Model results for average ratio of users that did not find a suitable server in the Fediverse platform. In scenarios with higher numbers of users and servers, the ratio of matched users approaches 50%.

Users who did not leave the platform but have not found a server: In this case (see Figure 9), we evaluated the number of users who were not matched to any servers at every point in time (i.e., at each tick of the simulation). Hence, these users have not left or won't necessarily leave the platform, as in the previous case. We assume that if a user does not find a suitable server, it prefers not to join one. We will refer to these users as 'unmatched users.' As expected, scenarios with fewer servers show a higher number of unmatched users. Compared to the scenario with only 10 users, the number of unmatched users is lower than in scenarios with 100 and 1000 users. However, this trend stabilizes as the number of servers increases. This is particularly the case in scenarios with 1000 users, where there are ~50% of unmatched users in all three – 1, 10, and 1000 – server levels.

Average degree of each server in the platform: This metric (Figure 10) allows us to explore the number of users connected to each server throughout the simulation time. We find that, on average, as the number of servers increases, the connections of each individual server decrease; however, we also find that as the number of users and servers increases, the number of users covered with the available servers remains around 50%. In the 1000-user scenario, this percentage drops only to 47.3% when there is only one server. This allows us to infer that even in the presence of a larger set of server options, users do not necessarily switch or search for other alternatives once they found a suitable server. This may bring to our attention the complexity of exploring a large range of server options in the Fediverse.

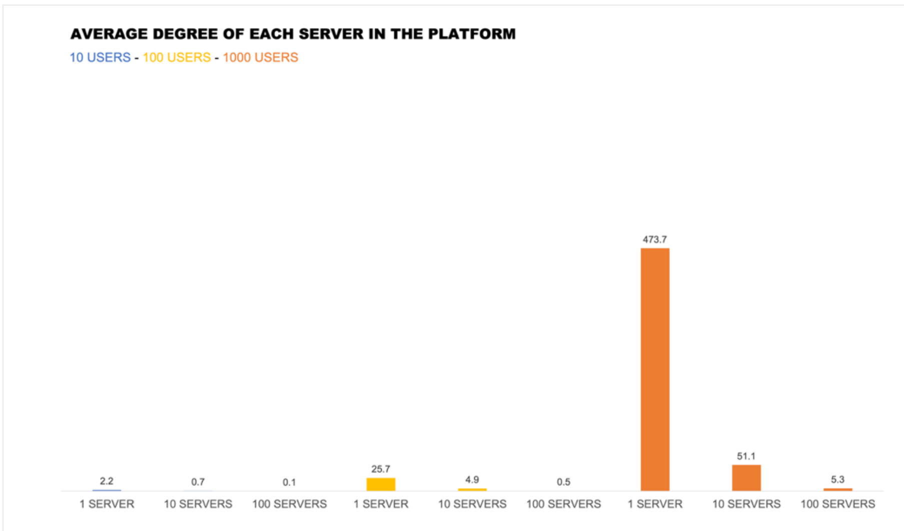

Figure 10: Average degree of each server in the platform allows us to investigate the number of users that are connected to a given server throughout the simulation period. As the number of users increase, we find that approximately half of the users are connected to a server. This is particularly the case of scenarios with 100 and 1000 users combined with 10 and 100 servers.

6 Discussion and Analysis

Federated systems are characterized by power sharing between autonomous local units and central-coordination agencies. Social networks are important for stimulating information flow within and between communities and are often organized in federations. As institutions are necessary to govern information flow, physical social interactions often abide by social norms determining what information can be communicated on which channels, who can communicate with whom, and which modes of communication are admissible. OSNs have an additional advantage in that communication protocols can be formalized in automated processes and enforced by network operators. Fediverse goes a step further by providing users with the tools and infrastructure needed to generate customized OSNs for themselves. It thus provides a technology through which productive communities can self-generate and coordinate on their own terms.

A conceptual framework for studying federated information systems, e.g., Fediverse, would require a clear set of analytical elements (governance structures) for comparative analysis as well as a language and vocabulary for facilitating such analysis. The literature on common-pool resource management, particularly extended to the knowledge-commons literature, provides a promising direction of consideration and our ABM analysis provides an initial preview of how such frameworks can be operationalized in predicting hypothetical governance alternatives in the management of information. Our model particularly highlights how variation in communication protocols across information platforms affects the stability of a federated online social network. Perhaps it is possible to argue that while moderation and user match are important, user inertia and network externalities (discovery and availability of content) may reduce the mobility of users across platform instances.

7 Conclusion

Emergent technologies are considered from the perspective of centralized versus decentralized governance. We argued that federated systems depend on governance as well, and that a range of governance outcomes with respect to decentralized systems emerges from combinations of the details of system design, the specification of formal rules by each system, and user choices as to which system or subsystem to join. We analyzed these dynamics using a simple model in which instances vary in moderation. This approach offers insight into the ways in which governance influences the evolution of distributed systems.

An area for future research is to integrate the question of governance internal to federated systems with governance of systems external to it. This might include comparison of centralized systems with decentralized systems where both vary in governance dimensions, including moderation, as well as governance aspects such as privacy and data monetization. Such comparative institutional analysis would offer an integrated perspective on governance of social networks.

8 Bibliography

[1] L. La Cava, S. Greco and A. Tagarelli, "Understanding the growth of the Fediverse through the lens of Mastodon," *Applied Network Science,* vol. 6, no. 64, 2021.

[2] M. Zignani, C. Quadri, S. Gaito, H. Cherifi and G. P. Rossi, "The Footprints of a “Mastodon”: How a Decentralized Architecture Influences Online Social Relationships," in *IEEE Infocom Workshops: CAOS 2019: Communications and Networking Aspects of Online Social Networks*, 2019.

[3] S. Tarkoma, Publish/subscribe systems: design and principles, John Wiley & Sons, 2012.

[4] L. Gerster, F. Arcostanzo, N. Prieto-Chavana, D. Hammer and C. Schwieter, "The Hydra on the Web: Challenges Associated with Extremist Use of the Fediverse – A Case Study of PeerTube," Institute for Strategic Dialogue, 2023.

[5] S. Zannettou, B. Bradlyn, E. D. Cristofaro, H. Kwak, M. Sirivianos, G. Stringhini and J. Blackburn, "What is Gab? A Bastion of Free Speech or an Alt-Right Echo Chamber?," in *The Third International Workshop on Cybersafety, Online Harassment, and Misinformation*, (Part of ACM WWW), Lyon, 2018.

[6] Mozilla, "Mozilla.Social | Content Moderation Policy," [Online]. Available: https://mozilla.social/content-moderation-policy. [Accessed 24 July 2023].

[7] A. Z. Rozenshtein, "Moderating the Fediverse: Content Moderation on Distributed Social Media," *Journal of Free Speech Law,* vol. To appear, 2023.

[8] A. Wool, "Firewall Configuration Errors Revisited," https://arxiv.org/pdf/0911.1240.pdf, 2009.

[9] L. Cho, "Threads Is Already Losing Its Allure for Users, Adding Urgency for New Features," *The Wall Street Journal,* 21 July 2023.

[10] W3C, "ActivityPub W3C Recommendation," 2018. [Online]. Available: https://www.w3.org/TR/activitypub/. [Accessed 27 06 2023].

[11] NOSTR, "A Decentralized Social Network with a Chance of Working," [Online]. Available: https://nostr.com. [Accessed 27 06 2023].

[12] A. Protocol, "The AT Protocol," [Online]. Available: https://atproto.com/. [Accessed 24 July 2023].

[13] L. La Cava, S. Greco and A. Tagarelli, "Network Analysis of the Information Consumption-Production Dichotomy in Mastodon User Behaviors," in *Proceedings of the Sixteenth International AAAI Conference on Web and Social Media (ICWSM 2022)*, 2022.

[14] B. Guidi, A. Michienzi and L. Ricci, "A Graph-Based Socioeconomic Analysis of Steemit," *IEEE Transactions on Computational Social Systems,* 2021.

[15] H. Li, B. Hecht and S. Chancellor, "Measuring the Monetary Value of Online Volunteer Work," in *Proceedings of the Sixteenth International AAAI Conference on Web and Social Media* , ICWSM, 2022.

[16] L. La Cava, L. M. Aiello and A. Tagarelli, "Get Out of the Nest! Drivers of Social Influence in the #TwitterMigration to Mastodon," in *https://doi.org/10.48550/arXiv.2305.19056* , Available on ArXiV, 2023.

[17] K. Ermoshina and F. Musiani, "Safer spaces by design? Federated architectures and alternative socio-technical models for content moderation," in *Annual Symposium of the Global Internet Governance Academic Network (GigaNet)*, 2022.

[18] S. Singh, "Everything in Moderation, An Analysis of How Internet Platforms Are Using Artificial Intelligence to Moderate User-Generated Content," newamerica.org, 2019.

[19] V. Ilik and L. Koster, "Information Sharing Pipeline," *The Serials Librarian,,* vol. 76, pp. 55-65, 2019.

[20] D. Allen, E. Frankel, W. Lim, D. Siddarth, J. Simons and E. G. Weyl, "Ethics of Decentralized Social Technologies: Lessons from the Web3 Wave," Edmond & Lily Safra Center for Ethics, Harvard University, 2023.

[21] B. M. Frischmann, M. J. Madison and K. J. & Strandburg, "Introduction," in *Governing knowledge commons*, New York: , Oxford University Press., 2014.

[22] E. Ostrom, "Beyond markets and states: Polycentric governance of complex economic systems," *American Economic Review,* vol. 100, no. 3, pp. 641-72, 2010.

[23] C. Feijóo, Y. Kwon, J. M. Bauer, E. Bohlin, B. Howell and R. Jain, "Harnessing artificial intelligence (AI) to increase wellbeing for all: The case for a new technology diplomacy," *Telecommunications Policy,* vol. 44, no. 6, 2020.

[24] I. Murtazashvili, J. B. Murtazashvili, M. B. Weiss and M. J. Madison, "Blockchain Networks as Knowledge Commons," *International Journal of the Commons,* vol. 16, no. 1, 2022.

[25] M. J. Madison, B. M. Frischmann and K. J. & Strandburg, "Constructing commons in the cultural environment," *Cornell Law Review ,* vol. 95, pp. 657-710, 2010.

[26] R. Albert and A.-L. Barabási, "Statistical mechanics of complex networks," *Reviews of modern physics,* vol. 74, no. 1, 2002.

[27] T. C. Schelling, "Models of segregation," *The American economic review,* pp. 488-493, 1969.