City, University of London

Quantifying Social Ties and Normative Behaviour in Online Fringe Communities

Author:
Amin Mekacher

First Supervisor: Prof. Andrea Baronchelli
Second Supervisor: Prof. Yang-Hui He

Examiners:
Prof. Mark Broom
and
Dr. Francesco Pierri

A doctoral dissertation submitted in fulfilment of the requirements for the degree of Doctor of Philosophy
in the
Department of Mathematics
School of Science and Technology

November 17, 2024

Contents

List of Figures

List of Tables

Abstract

Following the stellar rise of major social platforms in the online ecosystem, concerns have emerged regarding the content that is being shared to an international audience on such outlets. The public scrutiny led platform administrators to set up moderation policies to curtail the rise of hate speech, misinformation and online harassment. These policies triggered disenfranchised communities to settle on platforms that are not policing the content that is shared by their community, nowadays known as the *alt-tech ecosystem*.

Once primarily used as a meeting space for fringe online communities, these platforms have become major outlets for radical political movements. State presidents, such as Donald Trump and Jair Bolsonaro, have leveraged outlets belonging to the alt-tech ecosystem to smear their adversaries, raise funds and share controversial narratives.

In this thesis, we study the rise of this alternative ecosystem and measure its impact on user engagement, online discourse and the proliferation of extremist movements online.

First, we look at the emergence of Voat, a far-right platform, and how it was competing against Reddit, its mainstream counterpart. We characterise the impact of deplatforming policies from Reddit, the level of active engagement that emerges on Voat as well as the cross-community interactions that take place on the platform.

Second, we study cross-platform interactions by studying Gettr, a platform that has become a popular venue for right-wing state representatives in the US and Brazil. By looking at a user's activity on Twitter and Gettr, we assess how deplatforming leads to the emergence of a committed community on an alt-tech platform, and how different user cohorts use toxic discourse on both platforms.

Third, we assess the impact of this phenomenon beyond the West by studying Koo, an Indian-made social platform that gathered support from nationalist political representatives in India and Nigeria. We study the ability for an alt-tech platform to become an international venue and the interplay between language and culture in shaping online interactions.

We hope that this work will shed light on the social drivers that are catalysing the rise of this alternative online ecosystem, the societal risks that arise in an environment that does not enforce any content moderation, and the challenges that policy makers must tackle to ensure that these platforms do not become ideal venues for radicalisation.

Publications

This thesis is based on the following publications:

I. Amin Mekacher, Antonis Papasavva; *" I Can't Keep It Up." A Dataset from the Defunct Voat.co News Aggregator*; Proceedings of the International AAAI Conference on Web and Social Media 2022; 2022

II. Amin Mekacher, Max Falkenberg, Andrea Baronchelli; *The systemic impact of deplatforming on social media*; PNAS Nexus; 2023

III. Amin Mekacher, Max Falkenberg, Andrea Baronchelli; *The Koo Dataset: An Indian Microblogging Platform With Global Ambitions*; Proceedings of the International AAAI Conference on Web and Social Media 2024; 2024

IV. Amin Mekacher, Max Falkenberg, Andrea Baronchelli; *How Language, Culture, and Geography shape Online Dialogue: Insights from Koo*; arXiv; 2024

Other publications:

V. Amin Mekacher, Alberto Bracci, Matthieu Nadini, Mauro Martino, Laura Alessandretti, Luca Maria Aiello, Andrea Baronchelli; *Heterogeneous rarity patterns drive price dynamics in NFT collections*; Scientific Reports; 2022

VI. Max Falkenberg, Alessandro Galeazzi, Maddalena Torricelli, Niccolò Di Marco, Francesca Larosa, Madalina Sas, Amin Mekacher, Warren Pearce, Fabiana Zollo, Walter Quattrociocchi, Andrea Baronchelli; *Growing polarization around climate change on social media*; Nature Climate Change; 2022

Acknowledgements

This thesis is the result of work I could have never accomplished alone. I want to thank:

Andrea Baronchelli, who has guided me throughout the process and provided me with a lot of freedom to explore the research questions I am passionate about.

My colleagues at Ofcom, especially Holly, Amir, Jake, Yuval, Sinthu, and Almos, who provided me with exciting opportunities to get involved in the Online Safety Act and to engage in impactful work with respect to moderation policies for online services.

All the brilliant researchers I had the chance to collaborate with throughout my thesis and shared their expert knowledge. A special mention to Antonis, Luca and Laura.

Prof. Mark Broom and Dr. Francesco Pierri, for kindly agreeing to be the examiners for this thesis.

The Alan Turing Institute and especially all the great people I have met during my Enrichment Scheme placement. I especially want to thank Elliot, Tris, Hussein, Ellen, Zack, Youmna, and Sarada, for always being a friendly presence at the office.

All the friends I have met since moving in London, who made me feel welcome in the big city from day one and made my three years of PhD an unforgettable experience. I mostly want to thank Cristina, Torrun, Elora, Cinthia, Adam, Matteo, Ben, Alessio, Guillaume and Josh for turning London into my second home. A big shout also to my friends back home in Switzerland, who always stayed in touch remotely and cheered me up throughout my studies. Special mentions to Paloma, Vincent, Anne-Marie, François, Liv, Joëlle, Priscilla, Catherine, Nadim, Nicolas, Chloé and Anass.

The PhD and postdoc crowd in the mathematics department at City, for all the time we have spent commiserating about our lives as junior academics in the UK, while grabbing some delicious food from Exmouth Market.

Alberto, Max, Elohim, Maddalena, Laura and Matthieu, for working together, bashing our heads against the same challenges and drowning our issues in a pint.

Ma famille en Suisse, et en Tunisie, pour votre soutien constant, les coups de fil depuis Genève, Lucerne, Chebba, Tunis, tous les bons plats et les encouragements.

"Everything is relative, one man's absolute belief is another man's fairy tale."
– Salman Rushdie

Chapter 1

Introduction

One of the most recent shifts in our ultra-social society is the rise of a rich online ecosystem, with new services burgeoning and offering users the opportunity to share their content to a wide audience, beyond their social circle and their geographic location. This "many-to-many" communication pattern was seen as a way to catalyse the rise of a *Digital Democracy* [5], with minimal hurdles between content producers and consumers.

Parallel to the sharp increase in technology adoption across the globe, some ethical issues have started to arise regarding the content that is being shared on these online spaces [6]. The absence of gatekeeping systems implies that any content can be widely shared, with no moderator in the loop to check the validity or the safety of the content. As such, some online platforms were quickly swarmed with hateful and inappropriate content, with no safeguard in place to ensure that such content does not reach a vulnerable audience [7]. A totemic example of the dilemma can be found in Usenet's history. Born in 1980 as a platform promoting an "anything-goes" philosophy to promote unrestricted free speech [8], Usenet sustained several legal blows, with respect to their lack of moderation policies [9]. However, the jurisprudence draws a line between *primary publishers*, who write, edit and publish the material, and *secondary publishers*, who only distribute the material to a wider audience with no curation [10], with platforms such as Usenet being traditionally included within the latter category [11]. As a result, the US drafted Section 230, a legal act that provides online services with an immunity from being held liable in the US for curating the content shared on their platforms.

The debate around content moderation and safe practices in online ecosystems became all the more critical in recent years, following the rise of the *alt-tech ecosystem* [12]. With established social platforms, such as Twitter, Facebook and Reddit, becoming more stringent with respect to online harassment, misinformation and hate speech [13, 14], new platforms started emerging, with their aim being to offer an alternative outlet to communities that have been banned from mainstream outlets [15]. This process, known as *deplatforming*, caused significant fringes of the online user base to aggregate on new outlets, where content moderation is close to non-existent [16] and anti-social narratives are tolerated, if not incentivised by the collectivity [17].

Once a fringe phenomenon, alt-tech platforms have recently become a major player in political propaganda, by allowing radicalised users to weave a collective narrative and build a robust network, without being banned by the platform's administrators [18]. This has led to the rise of several prominent conspiratorial communities, such as the QAnon movement in the US [19], the white supremacist ideology on Stormfront [20] or the global incel (involuntary celibates) community on Reddit and specialised forums [21]. The recent rise in terrorist attacks and violent riots perpetrated by far-right sympathisers, from the 2011 Norway attacks to the *Unite the Right* rally in Charlottesville, Virginia, highlighted the security risks intertwined with an empowered and vocal online extremist movement [22, 23]. However, despite the urgency to understand better the content being shared on such platforms and the user interactions that arise, the scientific literature related to the social cohesion and the narratives being shared within these communities is still scarce.

This thesis aims to improve our understanding of this burgeoning online phenomenon, by mapping the alt-tech ecosystem and looking at the impact these new platforms have on the overall online environment. Our contributions can be summarised into four main angles. First, we provide an overview of several alt-tech platforms, and the events that prompted online communities to migrate to these ecosystems. By looking at Voat, Gettr and Koo, we provide a clear understanding of the unintended impacts of deplatforming policies, the velocity of a collective migration and the user engagement that is cultivated afterwards on the alt-tech platform. Second, we look at the user engagement that spans across these platforms and within mainstream platforms and provide a clear comparison of the content being shared on an established platform, versus the one disseminated on the alt-tech platform. Our analysis provides a nuanced overview of the consequences that deplatforming has in curbing the exposure to hateful and/or misleading narratives for a wider audience. Third, we also study how users engage on alt-tech platforms, by looking at the impact that social factors, such as language, social status and group belonging, play in shaping the interaction network. Finally, we provide extensive datasets of user-created content and user-to-user interactions taking place on several alt-tech platforms, by making them publicly accessible. Our hope is that it will provide the resources required by other researchers to look at other social phenomena emerging on alt-tech platforms. Our datasets contain information about user profiles, but also fine-grained metadata related to user-generated content and the interactions taking place within the community, thus providing a longitudinal overview of the platforms' user base.

Our research focuses on a question that is extremely time-dependent, as the major deplatforming events that took place over the last few years had an immediate ripple effect on the alt-tech ecosystem. In order to precisely understand their impact, our research uses unique datasets, comprising user profiles and user-to-user interactions taking place on alt-tech platforms. For instance, chapter 3 is based on user profiles, community profiles and user posts and comments from Voat, from the platform's launch in 2014 up to its shutdown in 2020 [24]. Therefore, we can identify peaks of registrations and activity over the lifetime of the platform and can compare it with events of interest taking place on Reddit. Likewise, chapter 4 uses a more granular dataset for Gettr, another US-centric alt-tech platform, comprising user profiles, posts, comments, but also likes and shares taking place on the platform. As such, this dataset allows us to map passive interactions on Gettr, and therefore to understand the role that lurkers play in alt-tech environments. A similar dataset has been collected for Koo, an Indian-based alt-tech platform whose community spans Nigeria and Brazil as well [25, 26]. Chapter 5 provides an overview of the dataset, whereas research questions related to the role of linguistic and cultural factors in Koo's interaction patterns are tackled with this dataset in chapter 6. Each study is based on novel unique datasets of unprecedented size and coverage of the studied systems. The data used throughout our research was collected by using open-access APIs provided by the platforms, and the data was subsequently made available to the academic community through data-sharing services.

The research presented in this thesis relies on several analytical tools to identify patterns of interest in user-to-user interactions and platform dynamics. Network analysis is used in chapters 3, 5 and 6 to map the interaction network on Voat and Koo, and highlight the homophilic behaviour that strongly dominates user-to-user interactions on both platforms, in terms of ideology and language respectively. Statistical ecology tools, such as the Kaplan-Meier curve and alpha and beta diversity metrics, are used in chapters 4 and 6 to better understand how engaged users are on the platform over time, and how their interactions predominantly take place within their social circle. Time-series analysis is also used in chapters 3, 4 and 6 to map the impact of deplatforming events in established platforms on the registration peaks that are observed by their alt-tech competitors, further highlighting the speed at which collective migrations take place after a community gets banned from a mainstream platform. Natural Language Processing algorithms are used in chapters 4 and 5 to identify the most salient topics discussed within communities from the US, Brazil, India and Nigeria that settled on alt-tech platforms. Our findings offer a unique understanding of the narratives that are cultivated on alt-tech platforms, and a comparison with the topics identified from Twitter activity in Gettr offers some insights on the discourse taking place on both mainstream and alt-tech platforms. We also employ techniques to analyse the news ecosystem that arise on alt-tech platforms in chapters 4 and 5, with a more detailed study of the reliability of popular news outlets and their monopoly on the news ecosystem that a user is exposed to on alt-tech platforms. Finally, we use toxicity analysis in chapter 4 to quantify the toxicity of the content posted by several user cohorts on Gettr over time, and how different the level of toxicity is compared to Twitter activity.

The thesis is structured around three main questions, detailed below, with each chapter tackling one of the questions. Chapter 2 provides a broader overview of the social media ecosystem and its recent evolution. The main literature around deplatforming, online radicalisation and social dynamics within social media is also presented. Chapters 4 and 6 are also accompanied by their additional material, which provides the reader with more in-depth analysis and robustness checks.

Chapters 3 and 4: How do alt-tech platforms react to massive deplatforming from established platforms?

One of the leading causes for the surge of popularity within the alt-tech ecosystem is the ban of influential figures and communities from established platforms, leading them to seek refuge with an unregulated competitor [27]. Some of these platforms have become very proficient at recruiting disenfranchised actors, who helped skyrocket their prominence in the overall ecosystem [28].

To approach this question, we will look at two alt-tech platforms, namely Voat and Gettr. While both were founded with similar intentions, their initial capital was drastically different: Voat was started as a student project, whereas Gettr had substantial backup from conservative politicians in the US and Brazil from day one [29]. Both, however, benefited from major bans taking place on Reddit and Twitter respectively. Voat became a hub for QAnon-related communities after their ban from Reddit in 2017 [30], whereas right-wing personalities, such as Joe Rogan and Marjorie Taylor Greene, cemented Gettr's place in the US far right ecosystem [31]. In chapters 3 and 4, we identify the popularity of both platforms in the wake of the deplatforming events that took place on their mainstream competitors. Afterwards, we assess the user engagement they nurture, by looking at the user interaction network and the most salient topics that are cultivated.

The content of these chapters is based on publications [I] and [II].

Chapter 5: How does the alt-tech ecosystem evolve outside of the West?

Despite being often portrayed as a Western phenomenon, alt-tech platforms have also become strong contenders in the online ecosystem for many other nations. This rise is primarily fueled by a desire to disentangle themselves from US-dominated platforms, where non-English speakers are often marginalised in terms of language support and content moderation [32]. Moreover, several countries have enacted policies to block US platforms from being accessed within their territory, such as China's Golden Shield Project [33] and Russia's striving for digital sovereignty [34].

Our analysis will focus on Koo, an Indian-based platform that quickly garnered support from the dominant national political party. With its initial aim being to embody a "language first" philosophy, where every user should be able to express themselves in their mother tongue, Koo ended up attracting international governments, following their disagreements with Twitter's content moderation. As such, Koo began to threaten the US-based platforms' hegemony with its international outlook that spanned India, Nigeria and Brazil.

The content of this chapter is based on publication [III].

Chapter 6: What kind of social ecosystem arises on an alt-tech platform?

After a community has been banned from a social media platform, there is an urgent need for its members to congregate on another platform. The process is often spearheaded by a few leader figures, resulting in an overall resilience against deplatforming policies [35]. However, due to their nature as a safe haven for banned users, alt-tech platforms are likely to attract communities with extremely diverse ideologies, and they might not be able to cohabit on a similar outlet. For example, after being banned from Reddit, members of the r/The_Donald subreddit, who were strong supporters of Donald Trump's political stances, migrated to Voat [36]. However, they were soon to be censored by other communities already present on the platform, leading them to move instead to their own message board [37].

As such, there is a need to understand how communities, which migrate to a similar alt-tech platform from various contexts, end up sharing a digital space. Our analysis looks at Koo, the Indian social platform that attracted communities from India, Nigeria and Brazil. By looking at language use from each active user, our aim is to map the presence of cross-linguistic interactions, which would indicate that there is an avenue for communities, emerging from very different cultural settings, to communicate beyond the linguistic barrier [38].

The content of this chapter is based on publication [IV].

Overall, our research improves our understanding of the online social ecosystem, and the impact of alt-tech platforms on online behaviours. This thesis is based on a very recent social phenomenon, and can therefore act as a stepping stone to further clarify the role that alt-tech platforms play in aggregating radicalised users on an unregulated online space. Interesting avenues for future research can explore the impact of government policies against alt-tech platforms, especially in a context where several of these platforms have refused to comply with direct orders to remove some of its content [39]. Some of these questions will be further discussed in chapter 7, where we review the main findings of this thesis, as well as potential studies that could further improve our results.

Chapter 2

Background

2.1 Social dynamics on social media

Since their advent in the early 2000s, online social platforms have changed the way we interact within our community and with a global audience. By offering the opportunity for anyone to produce new content and share it to an international public, social media has revolutionised the dynamics that lead to the rise of online communities [40, 41]. Many studies have highlighted how our online interactions follow a strongly assortative pattern, with communities being moulded around collective identities [42], shared hobbies [43] and similar political leanings [44]. The ability to instantly connect with like-minded individuals has made social platforms pivotal in shaping activist movements, related to social causes [45] or aiming to raise awareness about humanitarian disasters [46].

On the other side, this pattern of preferential attachment has also exacerbated the polarisation observed in online spaces. Previous studies have shown that online polarisation has led to a fragmentation of the online debate along partisan lines [47], given birth to online echo chambers [48], and has bolstered uncivil behaviour towards one's political opponents [49]. This trend has been shown to be gamed by outside actors, whose aim is to hijack democratic processes by further dividing users along their political opinions [50].

Previous research has also delved into the role that social media platforms play in disseminating and amplifying misinformation towards a wider audience. Public exposure to misleading information has been particularly critical during major political events, such as presidential elections [51] or national referendums [52]. This phenomenon is further exacerbated by the role played by state actors [53] and grassroots movements [54], which capitalise on their online audience to promulgate their agenda and gain political leverage [55].

As a result of the role played by social media in distributing misleading content, researchers have audited the algorithms used by major social platforms to recommend content to their users, and have found that recommender systems are systematically nudging users towards misleading or problematic content [56, 57, 58]. This behaviour is partially explained by the popularity bias that governs modern recommender systems, namely that overall popular content is being recommended more often than personalised content [59]. Recommender systems are also burdened by an exposure bias, with some users being over-represented in the content pool and therefore more often recommended than others [60].

As mentioned in the introduction, online platforms have started investing in content moderation systems [61], forcing them to become *de facto* curators of the content shared on their services [62]. As such, platform administrators have been constantly navigating the ethical line between censoring free speech, in favour of keeping their community safe from harmful or misleading content [63].

There is a growing debate regarding the best strategies that should be enforced by social media companies to adopt a sound platform governance policy [64], most notably with respect to the use of automated systems [65] instead of human moderators [66], with other systems combining both with a human-in-the-loop approach [67]. However, one of the most divisive questions relates to the punishments that should be applied to any user who infringes the platform's terms of service [68]. One of the most adopted strategies is to reduce the visibility of content that has been flagged as being controversial or harmful [69], or even ban repeated offenders from the service altogether [70].

2.2 Deplatforming

In the wake of extremist movements on social platforms and the increasing social and political pressure for site administrators to be held accountable for the content that is shared on their services, mainstream platforms have started clamping down on users and communities that infringe their terms of service. Bans have been used to remove users who were involved in cyberbullying [71], disinformation campaigns [72], and hate speech [73]. This strategy has been mostly aimed towards controversial public personalities, in the hope that the lack of a figurehead for the movement would reduce collective engagement with offensive speech [74].

Deplatforming has been shown to effectively curb the use of toxic language on mainstream platforms and to tackle online safety concerns [70]. However, its long-term consequences have been undermined by the rise of an alternative online ecosystem, with platforms catering to communities that have been banned from established platforms. From its infancy in the early years of the web2 environment, sparked by message boards, such as 4chan [75] and Stormfront [20], the alt-tech ecosystem now counts a plethora of platforms that have managed to attract a more significant user base [76]. In the US, platforms such as Gab [77, 78], Parler [79, 80], and Truth Social [81, 82] have grown to become US-centric powerhouses for conspiratorial movements such as QAnon [83] and the "Stop the Steal" [84] allegations, whereas the phenomenon has also become prevalent in Europe, with Odysee, Telegram and Bitchute [85, 86].

Whereas alt-tech platforms used to only host fringe communities, they rose to prominence in several political circles, by offering a safe space for members to interact without any scrutiny from content moderators [16]. As such, studies have highlighted their use as a recruitment ground for extremist movements [87], due to the laissez-faire behaviour of the platform owners [12]. The rise of populist movements on alt-tech platforms has also been correlated with the amplification of anti-democratic narratives [88] leading, in some cases, to political riots [89, 90].

Alternative platforms have also become a privileged venue for online extremist movements to raise funds from their followers. After being banned from crowdfunding platforms such as Patreon, GoFundMe, and Kickstarter [91], far right movements relied on their counterparts in the alternative ecosystem, such as Hatreon and Wesearchr [92]. This solution has allowed fringe actors to raise enough capital to fund controversial research projects [93], or to cover legal expenses for lawsuits [94]. The scale of the crowdsourcing potential of fringe outlets further highlights the potential for extremist movements to tap into the alt-tech ecosystem to mobilise their followers. It is essential to note that some misinformation superspreaders are also leveraging their online influence to earn an income via their content. These financial incentives to spread misleading information have notably been identified within wellness influencers on Instagram [95] and YouTube channels [96] involved in spreading anti-vaccine propaganda. Hoax stories related to the 2016 US presidential election have also been financially profitable for several online figures through ad revenues [97]. As such, we need to also take into consideration the financial drivers that can lead influencers to migrate to alt-tech platforms with their followers, after having been banned from established outlets.

2.3 Online radicalisation

By offering access to a trove of content and the possibility to connect with like-minded individuals, online services have also facilitated the interconnection between users who share similar fringe opinions related to racial issues, public health and anti-immigration policies [98]. With the opportunity to aggregate on a similar platform, such communities have become new recruiting grounds for extremist movements, now able to groom newcomers who are seeking new information [99].

The RECRO model [100] describes the phases undertaken by an online user, seeking alternative narratives related to personal triggers and vulnerabilities (*Reflection* and *Exploration* phases). Their journey might lead them to connect with a community that espouses a narrative that conflicts with the mainstream dogma (*Connection* phase), leading the user to adopt similar views and to seek new information to reinforce their new internal narrative (*Resolution* phase). Finally, they become themselves a messenger of the community, leading newcomers to follow the same radicalisation pipeline (*Operational* phase). This pipeline has been observed in practice within anti-vaccine communities on Reddit and Facebook [101].

Similarly, the 3N model highlights the three main psychological drivers that lead people to espouse extremist opinions [102]. The first component, *Needs*, represents all the social, psychological or emotional motives that the person seeks to satisfy by joining a radical movement. The second one, *Narratives*, groups every storytelling tool, such as conspiracy theories, used to frame an in-group / out-group dichotomy between the extremist community and mainstream social movements. Finally, the last one, *Networks*, highlights the social pressure and the sense of belonging that arises once a person becomes strongly intertwined with other members of the extremist community. Overall, the 3N model emphasises the role that radicalisation plays in bringing purpose to isolated individuals, by providing them with a belief system and a support network. The 3N model has notably been used to study online social networks, such as the impact of conspiracy theories on group dynamics on Twitter [103].

Moreover, this phenomenon has been exacerbated by a growing fear over social media censorship [104], leading users to seek information in less reputable sources to fight off the information vacuum in established social media platforms - a social phenomenon often called the Streisand effect [105]. The lack of available information regarding major events trigger the *exploration* phase of the RECRO model described previously. This process has been shown to be a strong magnet for conspiratorial communities after collective traumatic events [106], and to lead young men to adopt radicalised viewpoints to offset personal grievances [107]. More specifically, conspiracy theories are a powerful medium to lure new recruits and build a collective narrative, centred on a distrust of mainstream institutions. Previous research defines conspiracy theories as "radicalising multipliers", due to their propensity to demonise the out-group, label voices of moderation as being part of the conspiracy, and spur their members to violent action [108].

The last phase of the radicalisation process suggested in [100] is the *operational* phase - when a newly radicalised individual becomes themselves a messenger of the movement, by means of violent protest or by becoming part of the recruitment pipeline [109]. In the latter case, social media has been extensively used by extremist movements to amplify their narrative [110], reach out to socially vulnerable users [111] and to coordinate their actions [112]. This strategy was employed by several extremist movements across the globe, from ISIS's digital strategy [113] to Pegida in Germany [114]. Moreover, thanks to the ubiquity of social platforms in our day-to-day interactions and the communities we identify with, extremist movements have been able to tap more massively on other scenes to subvert their narratives in mainstream spaces, for example through the musical scene [115] or combat sports [116].

Online social platforms have also been shown to be intricately involved in the radicalisation pipeline, by providing their users with content that reinforces their biases. Several studies have identified the role played by YouTube's recommender system [117, 118] into recommending more extreme content to their users over time.

One of the leading questions that will be tackled in this work is the role that alt-tech platforms play in fueling more users into the radicalisation pipeline. By offering an online space where violent and controversial content is unmoderated, alt-tech platforms have been shown to be incubators of content that promotes violent extremism [119]. Moreover, by hosting communities that have been banned from mainstream competitors, alt-tech platforms are allowing extremist groups to retain an online presence and promote their narratives in a welcoming space [120]. Our analysis will provide further evidence of the mechanisms employed by alt-tech platforms to attract, retain and radicalise users, particularly for political gains.

2.4 Mainstreaming of fringe opinions on alt-tech platforms

The web 2.0 paradigm, which rose in the aftermath of the dotcom bubble, has led to the rise of a new online ecosystem, where the content is dominantly user-generated and publicly shared by the site provider. The rise of the web 2.0 has been strongly associated with the advent of an empowerment of social movements [121], which are now able to share their grievances with a wide audience, metastasise into online grassroots activism [122] and raise awareness about social and environmental challenges [123].

However, studies have also highlighted the pervasive role played by social media in normalising and amplifying hateful narratives, by laundering fringe opinions into the mainstream discourse [124, 125]. A recent example of such a process is the "It's Okay to be White" campaign, which was kickstarted on 4chan in 2017, before becoming an innocuous message used by white supremacists in political discourses [126]. Online platforms have become a major tool in providing an audience to extremist narratives, more so by providing radicalised communities with a safe space to weave their narrative [127]. One of the pervasive effects of similar social campaigns is the normalisation of extremist opinions in the public sphere, such as hate speech against minority groups [128] or the spread of conspiracy theories [129].

Despite being aggressively banned from mainstream social outlets, extremist communities are still able to capitalise on the alt-tech ecosystem to weave an international network of like-minded individuals and to further widen their audience [130, 131]. While established platforms are still being weaponised by extremist groups to optimise the exposure to their narratives, alternative outlets increasingly became pivotal to share radicalised discourse that would be "censored" on mainstream platforms [132], leading to a more politicised and radical discourse to become the norm on unmoderated platforms [78].

Alt-tech platforms have also started to become a conduit for right-wing narratives, meaning that extremist opinions are getting intertwined with moderate right-wing political stances [133]. For example, Rumble hosted the GOP debate in 2023 [134], highlighting an alliance between mainstream political movements and what was once seen as a fringe online movement. Similarly, politicians who are active on both mainstream and alt-tech platforms further blur the line between extremist and established conventions [135]. These strategies make it more socially acceptable for larger crowds to align with discriminatory - and often violent - political stances [136]. This mainstreaming of fringe opinions is operated in a similar fashion that David Duke, an influential member of the Ku Klux Klan, funnelled white supremacist claims in the Republican bloodstream while running to become Louisiana's State Representative in 1989 [137]. His own chapter of the movement, named the "Knights of the Ku Klux Klan", aimed at shifting the ideology of the movement into more acceptable focus points, by branding itself as a "white civil rights organisation" [138].

This work aims to further understand how alt-tech platforms are instrumental in shifting the Overton window - a concept adopted in political sciences to represent the range of policies that are publicly acceptable for a politician to endorse at a given time [139]. To do so, we look at the most prevalent topics within political communities and compare their prominence within alt-tech and mainstream platforms, in order to identify attempts to inject political stances into the public mind.

2.5 Social media beyond the West

While the previous studies were heavily focused on the rise of online populist movements in the Western hemisphere, similar trends have been observed internationally, with a similar position leveraged by the growing alt-tech ecosystem. Several political leaders from large nations, such as former Brazilian president Jair Bolsonaro or Indian Prime Minister Narendra Modi, have incorporated alt-tech platforms into their political strategies, in order to share misleading or hateful narratives about their political opponents [140, 141].

Combined with the fact that state-of-the-art content moderation systems deployed on social platforms are mostly optimised to work with English text [142], there is a growing feeling of disenfranchisement among non-Western communities, leading alt-tech platforms that offer support for vernacular languages to become valid competitors on the social media market [38].

This phenomenon is all the more crucial in India, where social media platforms have been extensively used by the dominant political party as a social mobilisation tool. Following a conflict with Twitter in 2021, after the platform refused to comply with takedown requests from the Indian government [143], the country's prime minister Narendra Modi has been shaping a stricter governance of online services in the country [144]. As a pioneer in social media usage for political campaigning [145], the Bharatiya Janata Party (BJP), with Modi as its leader, has been an early adopter of Indian-made social platforms, in a bid to increase its presence on non-English dominated social media outlets [146]. Previous studies have stressed out the role that harmful speech [147] and disinformation [148] play in the BJP's social media apparatus. As such, their presence on unregulated platforms, willing to take down any content that is flagged by the government [149], raises concerns related to the democratic process and the protection of ethnic and religious minorities in India.

The work presented in this thesis aims to further explore the opportunities that alt-tech platforms leverage by hosting non-Western communities. By looking at the rise of Koo, an Indian-based social platform, we provide a clearer picture of the growing adoption of the alt-tech ecosystem as a legitimate venue for major political movements around the world [38, 150].

2.6 Online lurking behaviour

Despite a significant fraction of the population worldwide being registered on at least one social media platform [151], this does not imply that they are actively contributing to the content shared online. Indeed, when considering the users visiting a social platform, most of them would be labelled as *lurkers*, meaning that they browse on the platform, read the content shared by their peers, but rarely actively take part in a conversation. In the U.S., research suggests that almost half of the adult users on Twitter are lurkers [152], although most estimates posit that this category encompasses 90% of the users [153].

Despite their lack of contribution to the production of novel content, lurkers still play an important role in the social media ecosystem. Rather than being active members of their community, they can be seen as active listeners, reflecting on the content they encounter online [154]. Moreover, a previous study has shown that lurkers can be mobilised by online political movements [155], whereas specific reward schemes can be efficient at "delurking" users [156]. To quantify their structural role in the social network, several centrality metrics have been defined to rank lurkers within a community [157].

Due to the nature of these communities, we can expect lurkers to abound on far-right platforms. Previous research suggested that newcomers on extremist forums often aspire to become full members of the community, and are therefore keen to listen and learn the best ways to earn their peers' respect [158]. Moreover, when comparing violent to non-violent online right-wing extremist groups, one notices that lurkers are more prevalent in the violent groups [159].

The collective engagement becomes all the more toxic when considering smear campaigns such as *Gamergate*, an anti-feminist campaign within the gaming community that started in 2014 [160]. Hateful narratives are more likely to lead to an engaged community [161], raising more concerns regarding the security risks that could arise within these communities once their members metastasise into an engaged collective.

Our analysis provides insights on the user engagement that takes place across alt-tech platforms, more specifically by looking at the unbalance in content production among users. Our results suggest that hyperactive users leverage their position within the community to hold a monopoly over the collective discourse, further deepening the divide between content producers and consumers.

2.7 Studied platforms

The alt-tech ecosystem has become an evolving phenomenon over the last few years, with new platforms weaponising the global health crisis related to Covid-19, and the related conspiracy theories, to grow their audience [162]. Although this thesis aims at providing novel insights related to the systemic impact alt-tech platforms leverage in the online ecosystem, our work will focus on a few case studies. One of our central objectives is to improve the access to extensive datasets within the academic community, to allow other research groups to carry out their own study on the alt-tech ecosystem. Therefore, we publicly released the data we collected from each platform studied in this thesis.

The following subsections provide the reader with historical and social context related to each of the platforms that will be discussed later on.

2.7.1 Voat: Countering Reddit's community bans

Launched in 2005, Reddit is a social news aggregation platform, where users can create and join communities of interest, called *subreddits*. These communities offer the opportunity for users to create posts, which can then be interacted with by their peers. Subreddits are also responsible for moderating any content that is being shared, with the task usually carried out by volunteers. In order to curb anti-social behaviour taking place on the platform, Reddit administrators and moderators are able to *quarantine* a subreddit, an intervention implying that the targeted subreddit will still be accessible and members will have the possibility to stay active, however the community will not appear as search results for any query on the platform. Combined with the decentralised moderation system operated on Reddit [163], the platform's interventions led to problematic communities becoming less civil against Reddit staff and other subreddits [164].

From 2015, Reddit became less tolerant towards subreddits that were repeatedly violating their anti-harassment policies. Communities such as r/fatpeoplehate, whose aim was to ridicule influencers from the body positivity movement, and r/The_Donald, a subreddit catered for Donald Trump supporters, were among the communities banned for their problematic behaviour [165, 166]. This led their thousands of members to be left without an outlet to network and keep their communities alive. However, they were soon welcomed on Voat, a Reddit clone founded in 2014 by Atif Colo as a Reddit free-speech alternative [167]. By capitalising on Reddit's moderation policies, Voat managed to attract a significant number of banned subreddits, leading to a community of more than 100,000 users settling on the platform.

Following the ban of QAnon-related subreddits, Voat was endorsed by influential leaders of the movement as their next destination, and the platform's community became mainly composed of QAnon sympathisers [168]. This led the platform to earn its reputation as a far-right hub, with a strong focus on the QAnon mythos and Trump-related paraphernalia [169]. More importantly, the Voat saga led to a reassessment of the pernicious impact that banning an online community can have in terms of the usage of toxic speech [170] and the health of the overall social media ecosystem [171].

Our work looks at the collective organisation that arises within a subreddit after getting banned and the velocity at which these communities migrated to Voat [130]. Afterwards, we provide some insights into the news ecosystem, showing the advent of an *epistemic bubble* on the platform - a place where sources are filtered by gatekeepers within the community, to only share resources that align with the dominant narrative of the community [172]. Moreover, we look at the structure of the interaction network on Voat and highlight the existence of siloed communities within the platform, suggesting that extreme communities on Voat scarcely interact with the overall community.

2.7.2 Gettr: The US far-right's political outlet

Following the wave of anti-institution sentiments that were exacerbated by the Covid-19 pandemic, members of the US Republican Party became infamous for inflaming their supporters around extremist narratives [173]. Some members of Congress, such as Georgia's congressional incumbent Marjorie Taylor Greene and Colorado's Lauren Boebert, rose to prominence among far-right circles for endorsing extremist opinions related to social issues and anti-immigration policies [174]. However, in the aftermath of the January 6 insurrection and Trump's ban from Twitter, politicians on the far-right end of the spectrum started questioning the agenda governing social platforms, which they accused of being biased in favour of the Democrats [175].

As such, the alt-tech ecosystem became an attractive online environment to share political messaging without the risk of censorship. Gettr, launched in July 2021, was created by Jason Miller, Donald Trump's former spokesman, with the intention to host a political dialogue that would be deemed as being misleading or harmful on established platforms [89]. Gettr was quickly joined by members of the US right-wing ecosystem, such as former White House's chief strategist Steve Bannon and InfoWars executive Alex Jones. However, the platform also capitalised on the strong ties between the Trump and Bolsonaro families and was endorsed by former Brazilian president Jair Bolsonaro, along with members of his cabinet [176].

In the following months, Gettr managed to extend its popularity in the overall social media ecosystem by attracting political figures and online influencers, such as podcaster Joe Rogan [177] and far-right French politician Florian Philippot [178]. With a collective discourse mostly revolving around Republican talking points [179], Gettr aimed to cultivate a community that is sympathetic with Trump's political stances, while also expanding beyond the US by sponsoring political rallies in Brazil [180]. Gettr's influence in Brazilian politics climaxed after Bolsonaro's loss at the November 2022 presidential election against his opponent, Lula. Following a call of arms that was broadcast on Gettr by prominent US conservative celebrities [181], such as Steve Bannon, Bolsonaro supporters stormed the Brazilian Congress on January 8, 2023, in a public riot strongly reminiscent of the events that took place at the US Capitol on January 6, 2021 [182].

After Elon Musk's takeover of X, formerly known as Twitter, and Donald Trump's launch of Truth Social in 2022, Gettr has suffered several financial blows and is allegedly close to shutting down, after massive layoffs took place in the company in January 2024 [183].

Our study looks at Gettr's ability to capitalise on deplatforming events taking place on Twitter, by measuring the influx of new users and their degree of engagement on Gettr [89]. Moreover, we also take a deeper look at users who are active on both Gettr and Twitter, to measure their propensity to target their political opponents on Twitter, and their usage of toxic language on both platforms.

2.7.3 Koo: A Twitter competitor outside the Western hemisphere

In the last few years, we also witnessed the growth of social media platforms outside of Europe and North America. These new communities are tapping into a demographic market that is often underfunded by established platforms, and are therefore threatening the US hegemony in the social media market. One of the most interesting developments is within the Indian market, where new platforms are providing support to vernacular languages, thus opening their doors to non-English speakers. Traditional platforms have been shown to underinvest in minority languages, leading underrepresented linguistic communities to be more exposed to harmful and misleading content [184, 185]. This fallacy is often caused by a lack of resources in said languages, challenging the possibility to train automated content moderation systems to flag problematic content [186].

Koo had become a flagship platform for the "Indian-made" ecosystem - aiming at offering an alternative to US-dominated social platforms to non-English speakers across India. After winning the Atmanirbhar App Innovation Challenge - a contest launched by the Ministry of Electronic & IT - Koo was labelled as one of the promises of the start-up community to create a "world class Made In India App" [187]. Moreover, the platform's investments in supporting Indian vernacular languages allowed Koo to capitalise on its national status [188]. One of the platform's advertisement campaign, named "#KooKiyaKya", showcases the opportunity for every Indian user to express themselves in their mother tongue on Koo [189].

Aside from India, Koo had also been able to market itself on the international market, notably in Nigeria and Brazil. Strong investments in the support for vernacular languages turned the platform into a cherished outlet for then Nigerian president Muhammadu Buhari, as well as Brazilian president Lula [25].

Due to its close ties with the BJP government, Koo had been accused of fostering the populist narratives that drive Narendra Modi's government policies [190]. For example, investigations have shown that Koo was used to share divise and hateful narratives in the run-up to the 2024 Indian elections, in June 2024 [191].

Despite the endorsements from the dominant political movement in India and its ambitions to rival Twitter in terms of online outreach, Koo struggled to keep users active in the long term. Its usage had plummeted in the last year, forcing the platform to lay off a large portion of its employees [192]. In June 2024, the platform underwent a funding crisis, with salaries being suspended for its entire workforce [193]. This was closely followed with Koo shutting down definitely on July 3, 2024, after running out of funds to keep their services afloat [194].

Our work provides a longitudinal analysis of Koo's popularity over the years, and the scale of the migrations taking place in India, Nigeria and Brazil, that fueled the platform's user base. By analysing cross-language interactions taking place on the platform, we show that minority language speakers were consigned to the periphery of the interaction network, indicating that they were heavily segregated from Hindi- and English- speakers. Moreover, conversations taking place within a linguistic community ddid not significantly overlap with other linguistic groups, indicating that languages on Koo were strongly siloed, despite the platform's ambition to become a meeting hall for vernacular language speakers.

Chapter 3

"I Can't Keep It Up." A Dataset from the Defunct Voat.co News Aggregator

Note: Throughout this chapter, I decided to censor the name of an online community, which corresponds to a racial slur.

Voat.co was a news aggregator website that shut down on December 25, 2020. The site had a troubled history and was known for hosting various banned subreddits. This chapter presents a dataset with over $2.3M$ submissions and $16.2M$ comments posted from $113K$ users in $7.1K$ *subverses* (the equivalent of a subreddit for Voat). Our dataset covers the whole lifetime of Voat, from its developing period starting on November 8, 2013, the day it was founded, April 2014, up until the day it shut down (December 25, 2020).

This work presents the largest and most complete publicly available Voat dataset, to the best of our knowledge. Along with the release of this dataset, we present a preliminary analysis covering posting activity and daily user and subverse registration on the platform so that researchers interested in our dataset can know what to expect.

Our data may prove helpful to false news dissemination studies as we analyse the links users share on the platform, finding that many communities rely on alternative news press, like Breitbart and GatewayPundit, for their daily discussions. In addition, we perform network analysis on user interactions finding that many users prefer not to interact with subverses outside their narrative interests, which could be helpful to researchers focusing on polarisation and echo chambers. Also, since Voat was one of the platforms banned Reddit communities migrated to, we are confident our dataset will motivate and assist researchers studying deplatforming. Finally, many hateful and conspiratorial communities were very popular on Voat, which makes our work valuable for researchers focusing on toxicity, conspiracy theories, cross-platform studies of social networks, and natural language processing.

3.1 Introduction

Social networks are a primary tool in today's society. They offer countless opportunities for people around the world to connect in various ways, find jobs, entertain themselves, catch up on world happenings, etc. At the same time, social networks sometimes offer a "safe-house" for people that want, among other things, to connect to like-minded individuals towards sharing hate and toxicity [164, 195], discussing controversial matters [196], and spreading misinformation and disinformation [197].

Mainstream social networks suffer from users and communities that organise these conversations on their platforms. A common "solution" the administrators resort to is to ban them–*deplatforming*. A social network that is known to have taken this action many times is Reddit, which banned more than $7K$ subreddits [198] from its platform; the first one being in 2014 [199].

Research on deplatforming shows that users that had their communities banned met on forums and even got more toxic than what they used to be [200]. Other than forums, banned users also move to social networks that allow controversial discussions. One of the platforms that many banned Reddit communities decided to migrate to was Voat.

Voat was a Reddit-esque social network founded in April 2014 and shut down in December 2020 [24]. Similar to Reddit, discussions on Voat are divided into various channels–*subverses*–the equivalent of a subreddit. Users can subscribe to as many subverses they wish but cannot moderate more than ten to prevent users gaining undue influence on the platform. User registration on Voat requires only a unique username and a password. Newcomers can upvote, downvote, and comment on existing submissions but cannot create new submissions under subverses until they achieve a certain amount of upvotes on all of their comments.

Since its foundation, Voat gradually gained popularity over its years of operation, especially after every Reddit cleansing [167, 201, 202, 203]. Overall, Voat is known for hosting banned extreme communities and users, providing a safe space for like-minded individuals to share their ideas "freely." Voat has attracted the interest of researchers before as it hosted communities like /v/fatpeoplehate, /v/CoonTown, and /v/N***** [204], /v/TheRedPill [205], /v/GreatAwakening [168], etc.

Data Release. In this work, we present, to the best of our knowledge, the largest and most complete dataset of Voat. We release a dataset [206] that consists of over $18.6M$ posts from $113K$ users in $7K$ subverses over the lifetime of Voat (November 2013 - December 2020). More specifically, our dataset is *four fold* as it contains the title, body, and metadata of submissions; content and metadata of comments; user profile data; and subverse profile data.

Relevance. Our dataset provides several opportunities to the research community. First, Voat was evidently the place many banned users and communities moved to after being banned from other platforms [204, 207]. To this end, our dataset can assist researchers that focus on deplatforming and user migration. Also, our dataset may aid researchers deepen our understanding on how and when these communities choose their new "home" after a ban. Second, our dataset covers numerous offline events like the 2016 and 2020 US Presidential Elections and debates, Brexit, Jeffrey Epstein's arrest, and various terrorist attacks and unrest around the world that can prove helpful in further analysis of these events. Third, since Voat was a supporter of online freedom of expression for extreme and hateful communities, it contains a variety of slang language and toxic content that can be useful towards understanding hateful communities.

Chapter organisation. The rest of the chapter is organised as follows. First, we briefly explain what Voat is and how it works in Section 3.2 before going through its history in Section 3.3. Then, we describe the process of parsing the data released by the Archive team at the Internet Archive Wayback Machine (IAWM) platform, along with the complimentary collection of additional user and subverse data in Section 3.4. We then describe the structure of our dataset in Section 3.5 and provide a statistical analysis of the dataset (Section 3.6), followed by reviewing related work (Section 3.7). The chapter concludes with Section 3.8.

3.2 What *was* Voat?

Voat was a Reddit-esque news aggregator launched in April 2014. The mascot of Voat resembles an angry goat, which was designed and freely offered to the website by a user of the site.

Subverses. Discussions on Voat occur in specific groups of interests called "subverses." Users could create subverses on-demand before June 2020, when the administrators disabled this functionality. When a user creates a subverse, they become its owner, hence having complete authority over the subverse: they can deactivate the subverse and appoint other co-owners and moderators. The moderators can delete submissions and comments posted by users and even ban users from posting on the subverse. The owners and moderators can also allow users to post anonymously in their subverse, which replaces the posters' username with a random multi-digit number; not unique to each user. To prevent users from gaining extreme influence on the platform, Voat limits the number of subverses one can own or moderate.

Users. Voat proclaimed itself as a free-speech platform that offered its users anonymity. When newcomers register a new account, Voat does not require any personal details to verify the account, like an email address or phone number. A user can insert a username and a password to register, but if they forget their password, there is no way to recover the account.

After registering a new account, users can subscribe to subverses of interest, comment, upvote, and downvote the comments and submissions but cannot post new submissions. To post a new submission, they first need to acquire ten Comment Contribution Points (CCP). To do so, newcomers post comments on existing submissions trying to collect a *net score* of ten upvotes on all of their comments (one downvote cancels one upvote). The privilege of posting submissions is not guaranteed as users may lose it if their CCP falls below ten. Although this functionality may discourage users from being toxic to each other, it might also prevent users from debating their opinions as others may disagree and downvote them. Voat users often refer to themselves as "goats" due to the platform's mascot.

Submissions and voting system. Voat was a news aggregator platform, hence users can create a new submission by posting a title and a description, accompanied by a link to a news source, optionally. If the poster provides a link, the submission's title becomes a hyperlink to the source website. The domain of the source website appears next to the submission's title, along with the poster's username, and the date and time the submission was posted.

Similar to Reddit, Voat offers a hierarchical, tree-like commenting system: other users can comment on the submission and the comments of other users. Users can upvote or downvote the submission or other users' comments. In contrast with Reddit, Voat displays the total number of upvotes and downvotes a submission or a comment received. Also, the downvote functionality on Voat is not the same as Reddit's: downvoted submissions and comments alert the moderators of spammy or illegal content so they can take action. This functionality enforces the establishment of echo chambers as users usually downvote content that does not align with their beliefs. This usually results in the downvoted user to either losing their submission posting privilege or even being banned from the subverse.

Content visibility. Voat attempted to provide its users with some ephemerality without deleting its content, but hiding it instead. Voat subverses filter submissions under three tabs, namely, hot, new, and top. Each subverse has 500 active submissions in 20 pages (0 to 19). Hot submissions are the ones that are currently active and discussed, new submissions are the ones that were posted most recently, and top submissions are the most popular submissions of the subverse (many comments). Many subverses disabled the functionality of these tabs, and the submissions shown across all three tabs are often the same, just in a different order. We note that our dataset does not contain the tab of submission's as tabs are merely filters and often change based on the status of the submission, e.g., from new to hot.

When a user creates a new submission on a subverse, it would typically appear first on the new tab on page 0. At the same time, the last submission of page 19 is archived but not deleted, meaning if one knows the link to that submission they can still reach it but cannot comment or vote it.

Voat API. Voat supported a JSON API service for some time, but its maintenance stopped in October 2020. To collect the submissions of a subverse, one had to request the API of a specific page number (0 to 19) of a subverse's tab. The response of the API would be the 25 submissions of that page without their comments. To collect the comments, one needs to request them using the submission ID number, in which the API responds with 25 comments at a time.

Thus, to collect all the submissions from a subverse, one needs to request all 20 pages for the three tabs separately from the API. As explained by [168], the API does not list the archived subverses and does not respond to requests where the page is above 19. However, if one knows the submission ID and the subverse it was posted in, they can request the API for that specific submission. Since submission IDs on Voat are incremental, one could theoretically collect all of Voat's submissions by requesting the API for each submission ID incrementally for more than 7.5K subverses; that is 7.5K requests, in the worst case, to collect a single submission. To the best of our knowledge, no study or work managed to collect the full Voat dataset.

SearchVoat. A website not associated with Voat, named searchvoat.co, used to collect the Voat submissions and comments for its users to browse[a]. The site does not support an API and does not allow web scraping. After Voat shut down, the site became a news aggregator, similar to Voat[b].

3.3 Voat's Troubled History

In this section, we present Voat's history as we believe it highlights the significance of our dataset. WhoaVerse was the original name of the website and it was founded in April 2014. The website was a hobby project of Atif Colo (Voat username @Atko). Justin Chastain later joined Colo as a co-founder (username @PuttItOut). The founders advertised the website as an alternative social network focusing on freedom of expression and speech, which satisfies its users' needs and wants. In December 2014, WhoaVerse changed its name to Voat and marked its mascot as an angry goat.

In June 2015, after Reddit banned various hateful subreddits [167], including /r/n***** and /r/fatpeoplehate, many Reddit users started registering accounts on Voat. The sudden influx of users overloaded the site, causing temporary down time [208].

On June 19, 2015, Voat's web hosting service, Host Europe[c], canceled Voat's contract claiming that the site is publicising abusive, insulting, youth-endangering content, along with illegal right-wing extremist content [209]. Some days later, PayPal froze Voat's payment processing services [210]. In response, Voat shut down four subverses, two of which hosted sexualised images of minors and the founders attributed the shutdown to political correctness [211]. The site moved to a different hosting provider and started accepting cryptocurrency donations.

In July 2015, Reddit banned a popular administrator that caused another influx of Reddit members registering with Voat, leading to more downtime. In an interview, Colo said that they "provide an alternative platform where users would not be censored and still say whatever they want" [212]. Voat was the target of DDoS attacks many times and experienced

[a] https://searchvoat.co/search.php
[b] https://searchvoat.co/forum/
[c] https://www.hosteurope.de/en/

numerous failures during its six years of operation. The most significant attack was in July 2015 [213]. Voat, Inc. became a registered corporation in the US in August 2015. Although Voat was based in Switzerland, the U.S. seemed like the best option as explained by Colo in a post: "US law with regards to free speech, by far beats every other candidate country we've researched."

In November 2016, more users relocated to Voat after Reddit banned the /r/pizzagate conspiracy theory subreddit [201]. In January 2017, Colo resigned as CEO of Voat due to time availability restrictions and was replaced by Chastain. Chastain ran a fundraiser campaign in May 2017 after announcing that Voat might have to shut down due to financial issues; Voat managed to stay online.

In November 2017, Reddit banned its incel community (/r/incel), and many of its followers reportedly moved to Voat [202]. About a year later, on September 12, 2018, Reddit banned numerous subreddits dedicated to the QAnon conspiracy theory, which again caused many QAnon adherents to migrate to Voat [168].

In April 2019, Voat's CEO Chastain asked Voat users to stop threatening people as he had been contacted by a "US agency" about the threats posted on the website[d]. In response, Voat users were not pleased to hear that Voat was working with agencies to remove Voat content and "limiting" the site's freedom of expression. Specifically, the first comment on the submission was an anti-Semitic slur, calling for the extermination of Jews [214].

Finally, on December 22, 2020, Voat announced again, now for the last time, that it would shut down due to lack of funding[e]. Chastain explained that he had been funding the site himself since March 2020 but had run out of money. On December 25, 2020, Voat shut down and its last submission was posted by Chastain, noting: "@Atko made the first post to Voat, so I am making the last"[f].

In Table 3.1, we list some aforementioned Reddit bans that probably affected Voat's activity. Some of these bans previously captured researchers' interest. We use these bans in our analysis in Section 3.6 to show whether Voat's activity was indeed affected.

3.4 Data Parsing and Data Collection

This section details the methodology and tools employed for our data collection infrastructure.

Submissions and Comments. Following Voat's shutdown on December 25, 2020, the Archive Team[g] released a set of Voat snapshot captures in Web ARChive (WARC) for-

[d] https://searchvoat.co/v/Voat/3178819

[e] https://searchvoat.co/v/announcements/4169936

[f] https://searchvoat.co/v/Voat/4174956

[g] For more details about the Archive team, see wiki.archiveteam.org

No	Date	Ban
1	May 9, 2014	/r/beatingwomen [215]
2	Sep 6, 2014	/r/TheFappening [215]
3	May 7, 2015	/r/n***** [204]
4	Jun 6, 2015	/r/fatpeoplehate [204]
5	Nov 23, 2016	/r/pizzagate [201]
6	Nov 7, 2017	/r/incel [202]
7	Mar 15, 2018	/r/CBTS_Stream [203]
8	Sep 18, 2018	/r/GreatAwakening [168]

Table 3.1: Reddit bans that reportedly affected Voat's activity.

	Count	# Users	# Subverses
Submissions	2,334,817	80,063	7,616
Comments	15,731,754	153,827	7,515
Subverses	7,094		
Users	108,451		

Table 3.2: Number of submissions, comments, user profiles, and subverse profiles in the IAWM dataset.

mat [216], hosted on the Internet Archive Wayback Machine (IAWM). These WARC captures include snapshots the IAWM captured over the lifetime of Voat. A WARC format file consists of single or multiple WARC records (snapshots), and it supports, among other things, the access and scraping of archived data. The files also hold revised and duplicated snapshots [217].

To parse these snapshots into structured data, we set up a Python script to parse the submissions and comments. In our case, every WARC file is a collection of various Voat snapshots the IAWM captured. To facilitate the smooth parsing of the WARC files, we used the *warcio* Python library[h]. This library offers a convenient and reliable way to read a WARC file by streaming every entry included in the file and automatically detecting the *payload*. The payload contains the capture itself, i.e., the HTML DOM tree code of the platform. Each WARC file includes the snapshot of the entire platform for a specific time and date, that is, thousands of submission pages for millions of submissions.

[h] https://pypi.org/project/warcio/

	Submissions	Comments	Users	Subverses
Total	2,380,262	16,263,309	113,431	7,095

Table 3.3: Released dataset.

Our parser captured the HTML DOM tree code of each page included in the WARC files serially. Then, it passed the HTML DOM tree to a function that uses the *beautifulsoup* Python library to read and store in JSON format the data and metadata of the submissions and comments, i.e., submission title and content, number of upvotes and downvotes, comments, etc[i]. We ensure that our parser only stores the latest submission version, as WARC files have duplicate data.

We note that although many languages appear in our dataset, the overwhelming majority of posts use the English language. In addition, our parser does not capture or store any visual media, like videos and pictures, since such files are not included in the snapshots. Hence, our dataset is not suitable for researchers focusing on visual media analysis.

User and subverse profiles. To complement our dataset, we also collected user and subverse profiles. A user profile includes data like username and registration date, whereas a subverse profile consists of data like subverse creation date, description, etc. To collect this data, we built a crawler using the IAWM API[j], beautifulsoup, and HTML requests[k].

Every user and subverse profile URL is unique, but they all start the same way: *voat.co/u* for the former and *voat.co/v* for the latter. First, we requested the IAWM API for all the snapshots whose URLs start like user or subverse URLs. We then collected the responses and parse them into JSON format, storing the latest snapshot the IAWM has in its database for every unique username and subverse profile URL.

The above process results in the dataset summarised in Table 3.2. We collected a dataset that consists of more than $2.3M$ submissions posted by $80K$ users in $7.6K$ subverses, and over $15.7M$ comments posted by almost $154K$ users. Note that IAWM does not have the profiles of about 500 subverses and hence we only manage to collect the profiles of $7.1K$ subverses (6.8% loss). In addition, we collected almost $108.5K$ unique user profiles.

Data collected via Voat API. In an attempt to complete our dataset, we merged it with the data collected for the [168] study. For that study, we collected $176K$ submissions and $1.45M$ comments posted from $28K$ users in 241 subverses. Our data collection infrastructure used Voat's API between May 2020 and October 2020, when Voat stopped the maintenance of its API. We found $45.5K$ submissions and $532K$ comments that were missing from the IAWM archive and incorporated them in the released dataset.

Some subverses on Voat offered anonymity to their users by replacing their username with a random eight-digit number (not a unique number for every user). The total number of users that commented or posted a submission (Table 3.2) does not include anonymous or deleted users. Hence, we assume that Voat's known user base is $155K$ users, at least, based on the data we collected from the IAWM. It is impossible to know the exact Voat user base since Voat never shared the complete list of user profiles, even when it supported

[i] https://pypi.org/project/beautifulsoup4/
[j] https://pypi.org/project/waybackpy/
[k] https://pypi.org/project/requests-html/

a data API service; to collect a user's profile, one needs to know the username. This means that we cannot acquire user profile data of "lurkers." Assuming the total known number of usernames is $155K$, we estimate that about 27.1% of the total users' profile data ($41.6K$) is either missing, or deleted profiles. However, [168] show that 13% of the $15K$ users being active in QAnon discussions deleted their profiles. Considering that many usernames were deleted every day on Voat, we estimate that this dataset offers the best representation of Voat's user base to date. Incorporating [168] user data with ours, we found $5K$ additional user and 1 subverse profiles. The final dataset presented and released with this work is detailed in Table 3.3.

Fair Principles. The data released and presented in this chapter aligns with the FAIR guiding principles for scientific data, as described below[l]:

- *Findable*: We assign a unique constant digital object identifier (DOI) to our dataset[206][m].
- *Accessible*: Our dataset is openly accessible.
- *Interoperable*: We use JSON format to store our dataset since it is widely used for storing data and can be used in various programming languages. We also provide a detailed description of our dataset's format in Section 3.5.
- *Reusable*: We provide all the available metadata along with our dataset and we extensively document them in this chapter, in Section 3.5.

Ethical Considerations. The data collected, presented, and released with this chapter are available on the Wayback Machine and also used to be accessible (without the need of a registered account) on Voat before it went down. The collection and release of this dataset does not violate Voat's or Wayback Machine's Terms of Service. Although some subverses on Voat allowed users to post anonymously, the overwhelming majority did not offer this functionality. Hence, we collect user profile data of $114K$ users. The only identification of these user profiles is the unique pseudo name, which is not personally identifiable information. Analysis of the activity generated on Voat to other services could potentially be used to de-anonymise users. We note that we followed standard ethical guidelines [218] and made no attempt to de-anonymise users.

3.5 Data Description

We now present the structure of our dataset, available at [206].

Our dataset consists of submission, comment, user profile, and subverse profile data. We release our data in various newline-delimited JSON files (`.ndjson`)[n]. Each line in a `.ndjson` file consists of a JSON object that holds various keys and values. Specifically, we release 7,616 `.ndjson` files, one for every subverse, that hold the submission data. Similarly, we

[l] `https://www.go-fair.org/fair-principles/`
[m] `10.5281/zenodo.5841668`
[n] `http://ndjson.org/`

release $7,515$ `.ndjson` files that have comment data. We inspect our dataset for the missing 101 subverses' comments and find that these subverses have no comment activity, only a small number of submissions. Also, a single `.ndjson` file is released for user profile data, and another for subverse profile data. In total, we release $15,133$ `.ndjson` files. Table 3.4 lists the keys, value data type, and description of our dataset files.

We choose to release the submission and comment data separately for every subverse as we believe it facilitates researchers that want to focus on specific communities.

We also use JSON to release our dataset as it is among the most optimal ways to store and share data as it has extensive documentation and is supported by all popular programming languages.

Key	**Value data type**	**Description**	**Key**	**Value data type**	**Description**
subverse_name_submissions.ndjson (7,616 files)			subverse_name_comments.ndjson (7,515 files)		
`title`	`string`	Submission's title	`body`	`string`	Comment content
`body`	`string`	Submission's content	`user`	`string`	Poster's username
`user`	`string`	Poster's username	`time`	`string`	Time of comment
`time`	`string`	Time of submission	`date`	`string`	Date of comment
`date`	`string`	Date of submission	`upvotes`	`integer`	Number of upvotes
`upvotes`	`integer`	Number of upvotes	`downvotes`	`integer`	Number of downvotes
`downvotes`	`integer`	Number of downvotes	`comment_id`	`integer`	Comment ID
`domain`	`string`	Linked domain	`depth`	`integer`	Depth level
`link`	`string`	Submission's URL	`subverse`	`string`	Subverse's name
`submission_id`	`integer`	Submission's ID	`root_submission`	`integer`	Parent submission ID
`subverse`	`string`	Subverse's name			
user_profiles.ndjson (1 file)			subverse_profiles.ndjson (1 file)		
`user`	`string`	Username	`subverse`	`string`	Subverse's name
`reg_date`	`string`	Registration date	`subscriber_count`	`integer`	Subscriber count
`moderates`	`list of strings`	Moderated subverses	`about`	`string`	Subverse's description
`owns`	`list of strings`	Created subverses	`date_created`	`string`	Subverse's creation date

Table 3.4: Description of the keys and data value types.

3.6 Data Analysis

In this Section we provide statistical analysis and visualisation of our dataset.

Posting Activity. First, we show the overall posting activity on Voat. Figure 3.1 shows the number of submissions and comments per day on the platform. The vertical red dotted lines represent the events listed in Table 3.1. Although the platform was officially launched in April 2014, the first-ever submission was posted by *@Atko* on November 8, 2013, on the /v/voatdev subverse, that focused on the development of Voat, and at the time, only seven users were posting.

The total number of submissions in 2013 is only 61. These submissions primarily include discussions between @Atko and @PuttItOut in the /v/voatdev subverse. When the platform was launched in 2014, the total number of submissions peaks to $5,268$, then $276K$ in 2015,

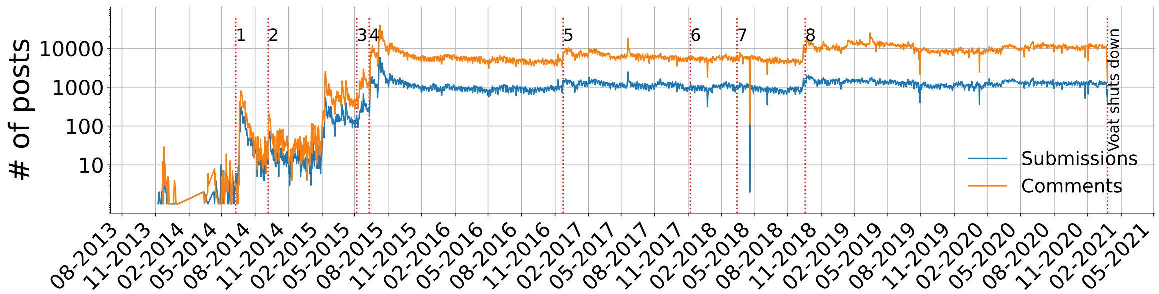


Figure 3.1: Number of all submissions and comments per day on Voat. Note log scale on y-axis. The red dashed lines represent some of the communities that got banned from Reddit, namely 1) /r/beatingwomen, 2) /r/TheFappening, 3) /r/n*****, 4) /r/fatpeoplehate, 5) /r/pizzagate, 6) /r/incel, 7) /r/CBTS_Stream, and 8) /r/GreatAwakening.

$324.8K$ in 2016, $397.2K$ in 2017, $382.9K$ in 2018, $439.2K$ in 2019, and for the last year, 2020, $421K$ submissions. Overall, there was no significant increase in activity on the platform after 2016.

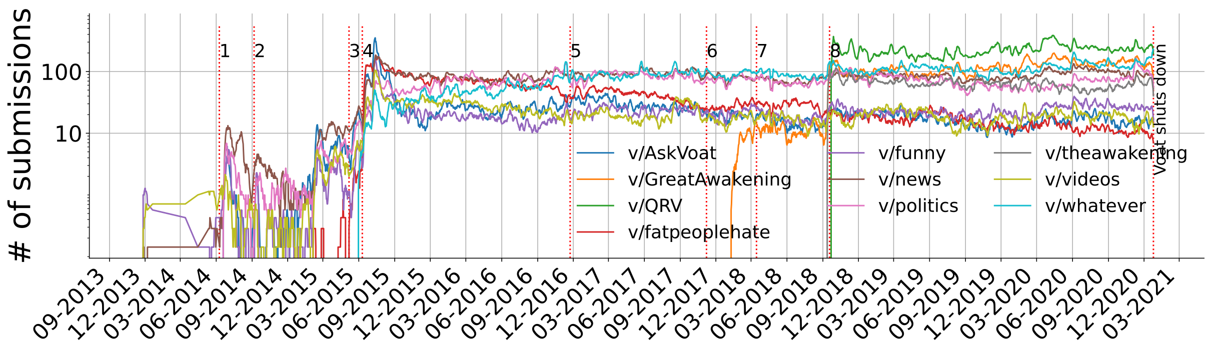
The most active day on the site is July 10, 2015, with $5.5K$ submissions. Manual inspection of our dataset indicates that discussions on that day focuses on Donald Trump, vaccine legislation, Reddit's CEO Ellen Pao resigning, and other world happenings. This date is very close to the date Reddit banned hateful communities like /r/fatpeoplehate and /r/n***** [167]. Shortly after Reddit banned these communities, Voat experienced heavy traffic and downtime [219].

Regarding comment activity, only 99 comments were posted in 2013, $13K$ in 2014, $1.6M$ in 2015, $1.8M$ in 2016, $2.1M$ in 2017, $2.4M$ in 2018, $3.8M$ in 2019, and $3.3M$ in 2020 Again, the date with the most comments on the platform is July 10, 2015, with $37.5K$ comments.

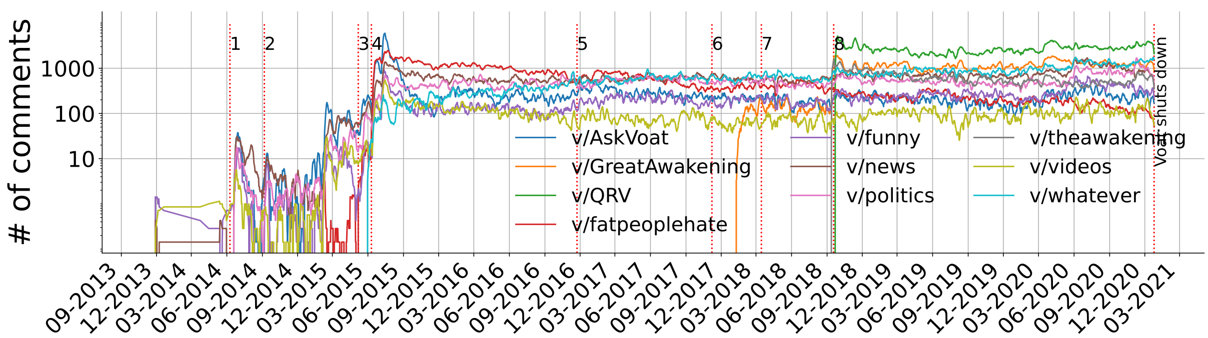
In addition, we show the overall activity on Voat in the top ten most subscribed subverses, namely, /v/AskVoat, /v/GreatAwakening, /v/QRV, /v/fatpeoplehate, /v/funny, /v/news, /v/politics, /v/theawakening, /v/videos, and /v/whatever, in Figure 3.2. We present this analysis to show how active the most popular subverses on Voat were, since we believe that researchers interested in our dataset might consider these findings useful. The vertical red dotted lines on the figure indicate the bans listed in Table 3.1. When the Reddit refugee crowd joined Voat (ban number 1, 3 and 4 from Table 3.1) many general discussion subverses like /v/AskVoat, /v/news, /v/politics, /v/videos, /v/funny, and /v/whatever became more active, indicating that this new influx of users bolstered the overall activity on the platform.

Interestingly, not all banned subreddits appeared on Voat shortly after a Reddit ban frenzy. The subverse /v/GreatAwakening was created on January 1, 2018, nine months before Reddit banned QAnon subreddits (ban no. 8). This subverse was the 10th most popular subverse when Voat shut down.

QAnon discussion on the platform boomed when /v/theawakening and /v/QRV first



(a) Submissions



(b) Comments

Figure 3.2: Seven day average number of a) submissions and b) comments per day on the top 10 most subscribed subverses on Voat. Note log scale on y-axis. The red dashed lines represent some of the communities that got banned from Reddit, namely 1) /r/beatingwomen, 2) /r/TheFappening, 3) /r/n*****, 4) /r/fatpeoplehate, 5) /r/pizzagate, 6) /r/incel, 7) /r/CBTS_Stream, and 8) /r/GreatAwakening.

appeared on Voat on September 12 and September 22, 2018, respectively, with approximately 200 submissions per day on /v/QRV alone. These three subverses turned out to be among the top 5 most active subverses on the platform, with /v/QRV being the most active in both daily submissions and comments on the whole Voat, within only ten days after being banned from Reddit [220].

The figures discussed in this subsection support the reports that Voat was among the main hubs for Reddit migrating communities. In addition, Figure 3.2 shows that other than general discussion subverses, the most subscribed subverses focused on hate speech (/v/fatpeoplehate) and conspiracy theories (/v/QRV, /v/theawakening, /v/TheGreatAwakening).

Submission Engagement. We set to discuss the engagement of the users on the platform. In Figure 3.3 we plot the Cumulative Distribution Functions (CDF) of the number of comments, upvotes, downvotes, and net votes (upvotes minus downvotes) per submission.

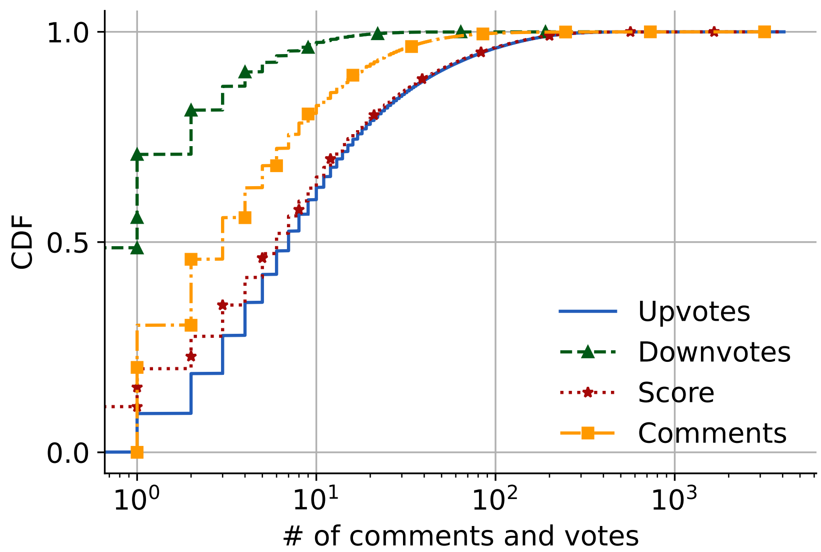


Figure 3.3: CDF of the number of comments, upvotes, downvotes, and net votes per submission.

Submissions on Voat get a median number of 3 comments, 7 upvotes, 1 downvotes, and a net score of 7. Comments receive a median 1 and 0 upvotes and downvotes respectively. The most upvoted submission reached over $4K$ upvotes, posted by Atko in /v/announcements in July 2015, explaining that Voat is experiencing heavy traffic due to Reddit bans. The most downvoted submission (392 downvotes) was posted in /v/politics with the title "Dear Media: Please Stop Normalizing The Alt-Right." The most liked comment noted that "someone isn't happy that Voat is succeeding" and reached $1.5K$ upvotes on a submission posted by Atko, discussing the DDoS attacks Voat was experiencing in July 2015. Last, the most disliked comment received 247 downvotes, posted by a user that was asking PuttItOut to reconsider the voting system of the site since they lost their submission posting privileges because of people downvoting them when posting their honest opinion. The user asks the CEOs:

> *[...]ask yourself: Are you fine with a website that caters to some of the most dangerous people currently walking the planet? Take a look at how depraved Trump supporters are, and ask yourself if free speech is worth the cost:[...]*

User registration and Subverse creation. In Figure 3.4 we plot the number of daily user and subverse registrations. The vertical dotted lines indicate bursty periods detected using Kleinberg's burst detection algorithm [221]. Kleinberg's algorithm models a time-series as an automaton, switching between two states: q_0, where events take place at a slow rate, and q_1, with events happening at a faster rate. By looking at the rate of events taking place at each time period, Kleinberg defines a cost function that the time-series needs to reach to move up to the state q_1.

We first define the baseline probability of an event taking place as $p_0 = \frac{R}{D}$, where R is

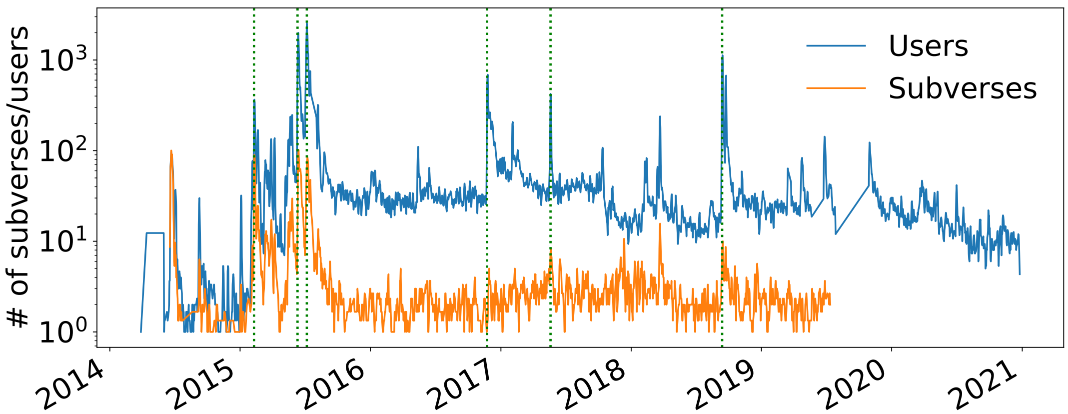


Figure 3.4: Number of users and subverses registered per day. Note log on y-axis. The green dashed lines indicate bursty periods within the user registration curve, computed using the Kleinberg's burst detection algorithm

the number of events taking place at a specific time, and D is the total number of events in the time-series. The probability of the bursty state is defined as $p_1 = s \cdot p_0$, where s is a coefficient indicating how high the rate of events needs to be to enter a bursty state.

The cost of transitioning from state q_0 and q_1 is defined as the sum of the goodness-of-fit between the observed proportion and the expected probability of each state, which we sum with the transition cost defined as

$$\tau = \gamma \cdot n \tag{3.1}$$

Where γ represents the difficulty of transitioning into higher states, n is the number of time points.

The bursty points are shown on Figure 3.4, when we set $s = 3$ and $\gamma = 2$. We notice that most of the bursty periods strongly match with the days when major subreddits were banned from Reddit. More precisely, there are three bursty points in 2015, on May the 9th, June the 11th and July the 7th. The first date matches with the ban date of /r/n*****, whereas the two others closely follow the time when /r/fatpeoplehate was banned from Reddit. Also, during the summer of 2015, Reddit changed their free speech and content policy [222] and the founder noted that "Reddit was not created to be a bastion of free speech."

Later on, the identified bursty peaks on November 23, 2016 and September 13, 2018, are strongly related with the dates when /r/pizzagate and /r/CBTS_Stream were banned, although we notice that users from /r/CBTS_Stream migrated to Voat shortly before they were banned from Reddit. This was prompted by the ban of other QAnon-related subreddits, most notably when r/GreatAwakening was banned on September 12, 2018.

While the peaks in registration overlap with significant bans that took place on Reddit, our analysis does not provide a glimpse of the velocity that one can observe when looking at singular deplatforming events. To fill this gap, we performed an Event Study Analysis (ESA), a statistical tool allowing us to compare the number of registrations that Voat experienced on the days prior and leading to a subreddit being deplatformed to an expected baseline. To do so, for four specific bans that took place on Reddit (namely /r/fatpeoplehate, /r/CBTS_Stream, /r/incel and /r/GreatAwakening), we compute the average number of registrations that took place on Voat in the two months prior to the ban as our historical baseline, and we compare this average to the number of daily registrations in the two weeks before and after the ban. The number of abnormal registrations therefore correspond to the difference between the observed number of registrations and the historical baseline.

As shown on Figure 3.5, the velocity of the platform migration is especially noticeable when looking at /r/fatpeoplehate and /r/incel, where the spike of abnormal registrations takes place in the four days following the ban, before plateauing at a lower level. This further suggests that, at a time when Voat was not yet known as a haven for banned Reddit communities, users would gradually join Voat after the ban took place. Interestingly, when looking at both QAnon related subreddits (/r/CBTS_Stream and /r/GreatAwakening), we notice that the number of abnormal registrations also increases in the days before the ban is enforced, indicating that the communities were anticipating to be deplatformed from Reddit. The influx of new users sharply decreases right after the ban, indicating a higher level of coordination in the collective migration than for /r/fatpeoplehate and /r/incel.

This analysis provides a glance at Voat's user base and subverse changes over the years. It is apparent that Reddit influenced Voat's activity and that the platform was among the preferred Reddit alternatives for banned users.

Links. Since Voat is a news aggregator platform, we analyse the domains the users posted on the site to show what kind of content the userbase of Voat consumed.

For each submission that redirects users to other domains, we retrieve the name of the subverse the submission is posted in, and the external link it redirects to. We count how many times a domain is shared in a community, keeping only the subverse and domain pairs that are the most recurrent in the dataset. Figure 3.6 displays an alluvial diagram showing, on the left, the considered subverses and, on the right, the top web domains shared on this subverse.

Most of the links that redirect users to Reddit were posted in /v/MeanwhileOnReddit. The subverse focusing on body-shaming, /v/fatpeoplehate, redirected users to Instagram, YouTube, and image sharing services (websites where users can upload images and share the link on other platforms). The /v/news subverse linked YouTube, Voat, online press outlets, and archiving services links. It is known that users in fringe communities avoid sharing the direct link to a website and prefer an archive link instead to avoid monetising

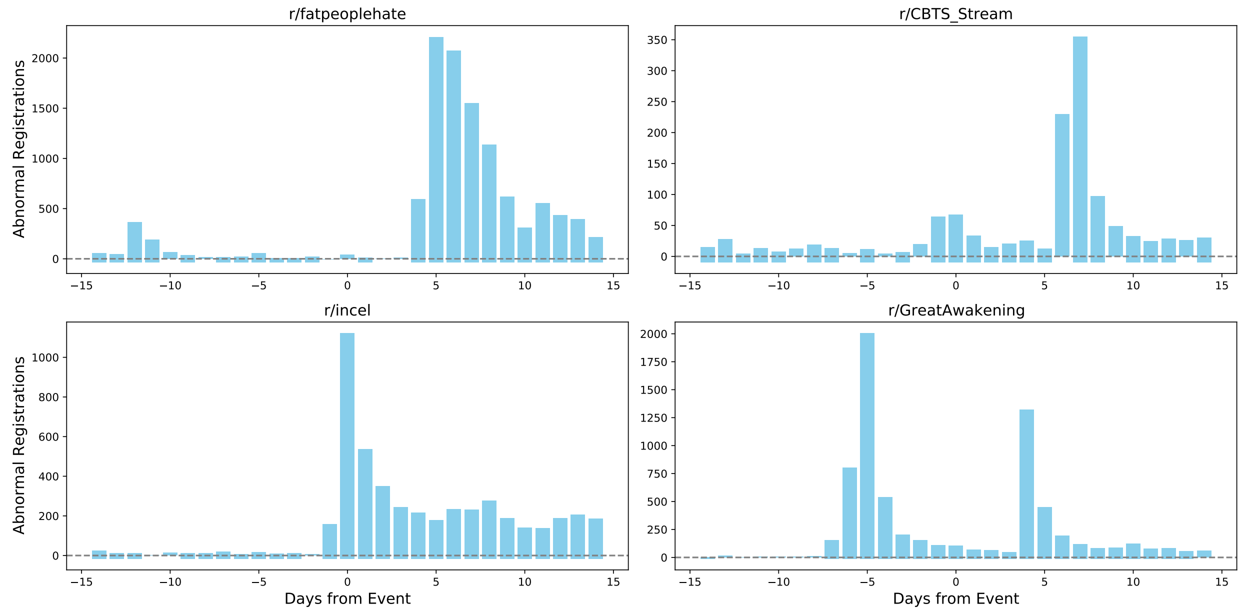


Figure 3.5: Number of abnormal registrations that took place on Voat in the lead-up to four major bans on Reddit (/r/pizzagate, /r/CBTS_Stream, /r/incel and /r/GreatAwakening). The number of abnormal registrations was computed using the Event Study Analysis framework, with a considered baseline of two months before the ban. The gray line represents the historical average of bans in the two-months period.

the website [223]. The majority of the alternative news links (Breitbart, GatewayPundit, and Zero Hedge) are posted on /v/news and /v/WorldToday. Most of the Twitter links on the website were posted in /v/QRV and /v/GreatAwakening. Most of the tweets include Donald Trump's tweets and other political discussions on Twitter (now X).

Overall, Voat users shared links to other social networks like Twitter and Instagram. News on the website was shared via legitimate online press outlets and other alternative news outlets, along with archiving services links. Most of the images on the platform were shared on /v/funny, /v/fatpeoplehate, and /v/whatever.

User Interactions. We now take a deep look into Voat's user ecosystem. We attempt to show how users form clusters based on the subverses they most often engaged with (posted a submission or a comment) to show whether the userbase of Voat is homogeneous or not. Further analysis on Voat's user base may shed light on what content users prefer to see on Voat and whether most Voat's subverses focused on hateful and politically incorrect content.

As shown in [30, 168], some users are responsible for a large amount of content being shared in some communities, leading to imbalances, influencing the content users consume on the platform. By analysing each user's interactions on Voat, we hope to obverse how all these various communities blended after a mass migration from Reddit, or if Voat was

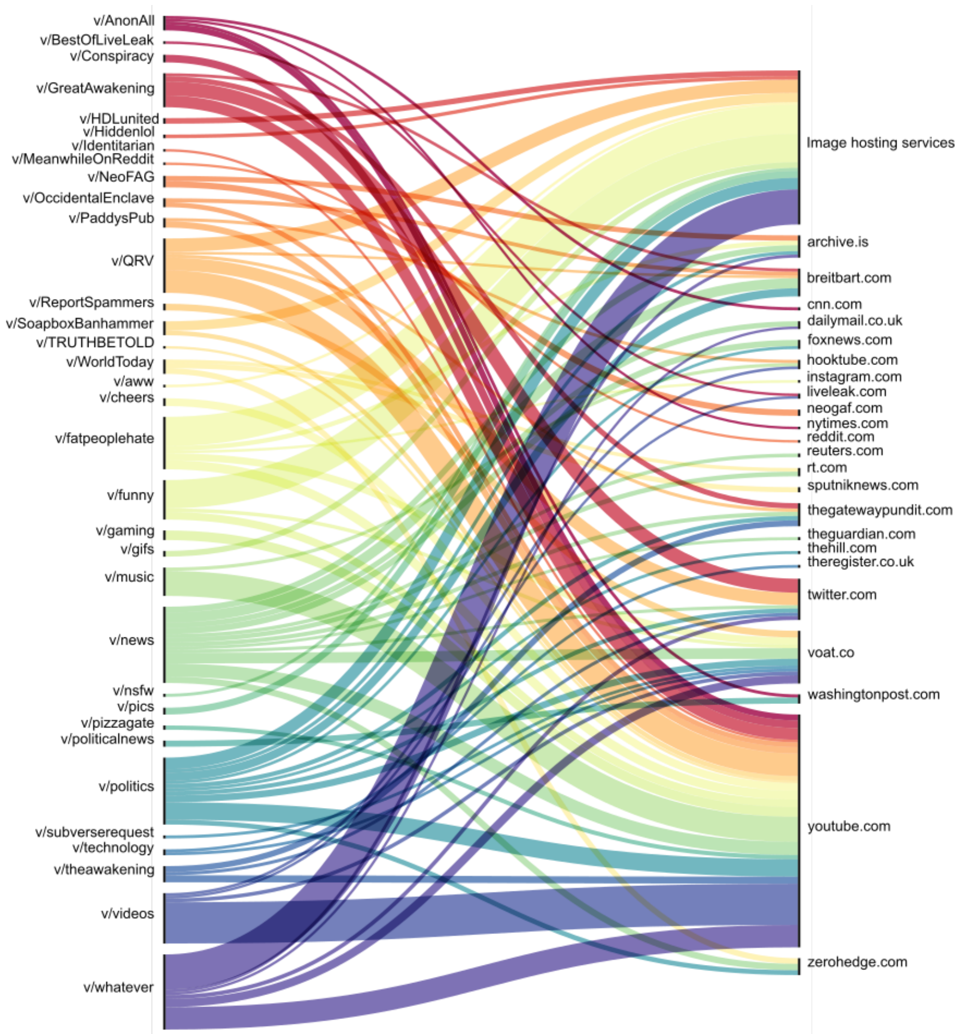


Figure 3.6: Alluvial diagram, displaying the top domain - subverse pairs on Voat. The width of the link indicates how many times the domain was shared on the subverse.

nothing more than an aggregate of small, selective echo chambers.

In Figure 3.7 we plot a graph network where nodes represent users, and the edges symbolise their interactions. For example, users are linked together if they participated in the same conversation, i.e., they both commented on the same submission, or one of them is the submitter while the other commented. The weight of the edge is given by the number of interactions shared by the two users, and the colour represents the subverse where the user participated the most.

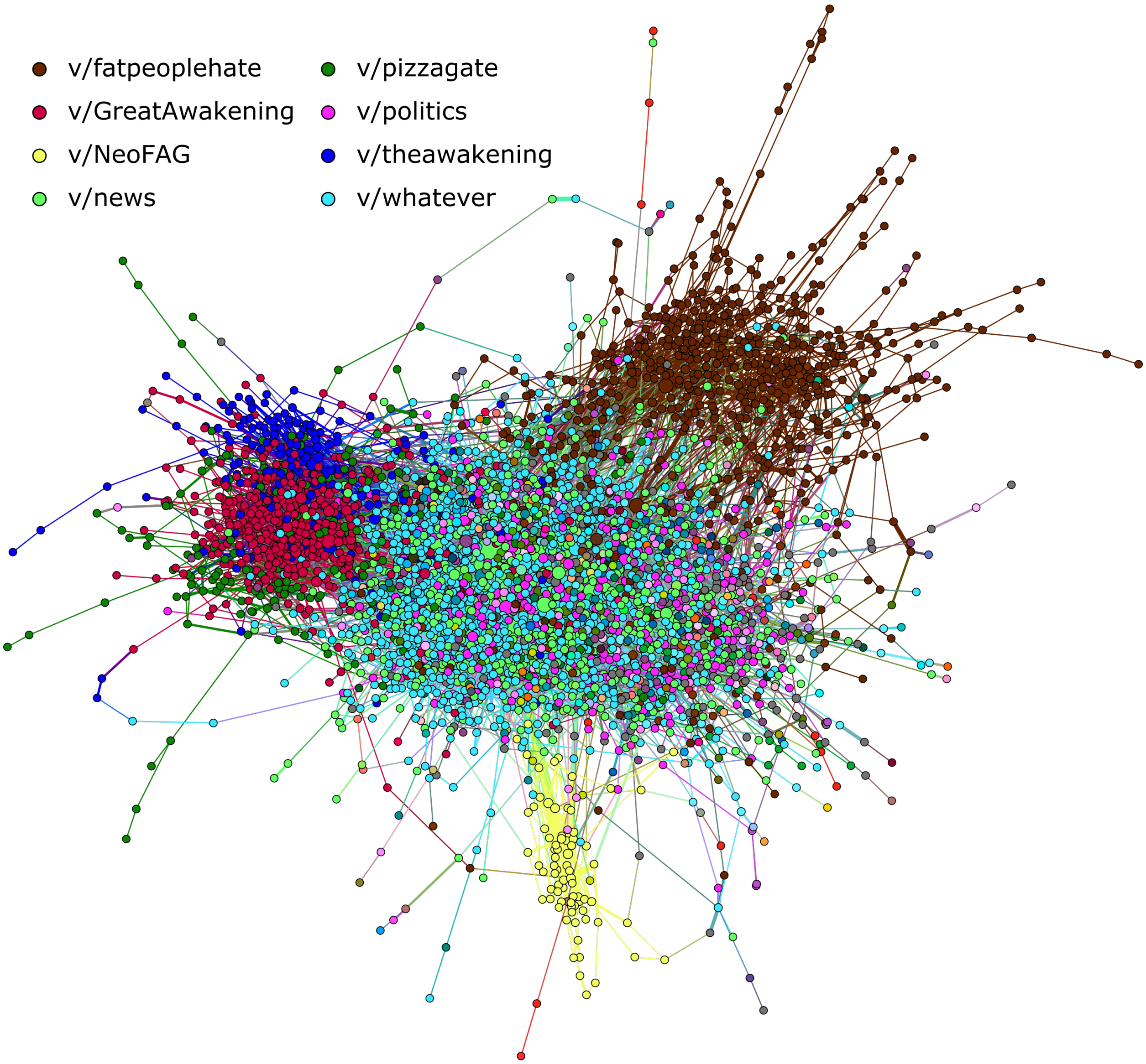


Figure 3.7: User and subverse interaction ecosystem.

The network is composed of a giant cluster, where most subverses are mixed together. This cluster includes /v/politics, /v/news, and /v/whatever, which is expected since these are general discussion subverses, and it is likely that many users meet there for general discussion. However, some subverses are strongly isolated in the network. For example, the /v/NeoFAG (yellow) community shows that most users tend to only engage within that subverse. Similarly, /v/GreatAwakening (red) and /v/theawakening (dark blue) seem to be clustered together and somewhat interacting with /v/pizzagate (dark green). Some users that engage with these three subverses also engage in the general discussion subverses, which is aligned with the findings of [168]. Last, /v/fatpeoplehate (brown) users also seem to form

Subverse	**EI-Homophily Index**
politics	0.50
news	0.40
whatever	0.23
theawakening	0.01
GreatAwakening	-0.25
pizzagate	-0.49
fatpeoplehate	-0.61
NeoFAG	-0.74

Table 3.5: Average homophily index between subverses and members.

their own cluster while infiltrating the general discussion subverses.

To measure the homophily of these communities, we used the EI homophily index, which is a metric that indicates how many members of a network favour in–group interactions rather than out–group ones. Given a specific node with E external edges, i.e., edges with nodes from the out–group, and I internal edges, i.e., edges with nodes from the in–group, the EI homophily index is given by the equation $EI = (E - I)/(E + I)$.

An index $EI = +1$ indicates that the node only interacts with members of the out–group, whereas $EI = -1$ applies to nodes that only interact within their in–group. Table 3.5 lists the average EI homophily index of the members of the subverses highlighted in the legend of Figure 3.7.

Users who are very active on subverses like /v/politics and /v/news have a high average EI homophily index, meaning they mostly interact with users from the out–group. The opposite can be said for the /v/theawakening, /v/GreatAwakening, /v/pizzagate, and especially, /v/NeoFAG and /v/fatpeoplehate. These communities do not converse much outside their social group. The EI index is almost zero for /v/theawakening, meaning its users interact as much with the out–group as with the in–group. By looking at the community, this can be explained by the fact that users from the communities gravitating around the QAnon narrative, i.e., /v/theawakening, and /v/GreatAwakening, are more connected than other communities. As a result, the external edges can be nothing more than crossovers between these two subverses. The userbases of /v/NeoFAG and /v/fatpeoplehate seem to be the ones that only prefer to interact with members of their community.

The isolation of a few communities on Voat can further be described by looking at the distribution of the EI-Homophily index within a subverse, as shown on Figure 3.8. As expected, we notice that the communities that are isolated in the interaction network also have a distribution skewed towards negative values. On the other side, most of the subverses peak at positive values of the EI-Homophily index, indicating that they belong to the giant component of the network.

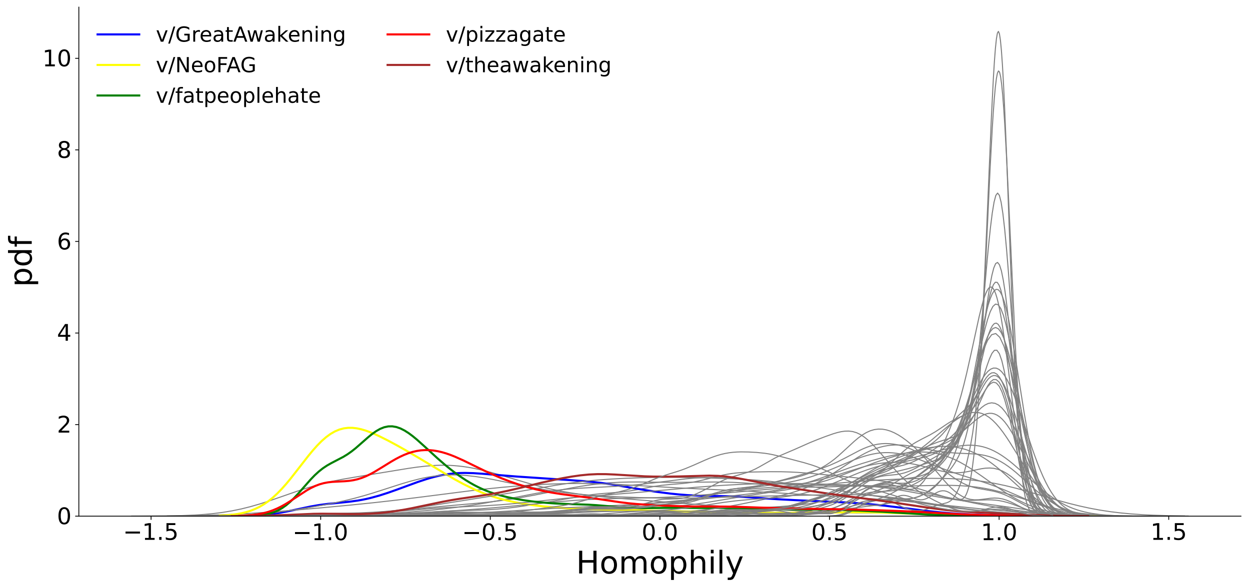


Figure 3.8: Distribution of the EI-Homophily index within subverses on Voat. Some of the subverses are highlighted in colour, whereas the others are plotted in gray.

We present this analysis to motivate researchers studying user interactions and echo chambers. Further research using our dataset may shed light on whether Voat was a bastion of echo chambers or not, along with what narratives users within these communities exchanged.

3.7 Related Work

In this section, we present existing work focusing on Voat, and other dataset papers similar to ours. Voat attracted the interest of researchers over the past years, especially after Reddit started banning communities in 2015. Although some studies mention that their dataset is available upon request, these datasets only include data from a couple of subverses that cover a short period of time. To the best of our knowledge, our Voat dataset is: 1) the only one to be openly and publicly available online; and 2) the most complete and largest one, covering the whole history of Voat, along with data of the users that ever posted a submission or a comment on the platform.

Voat research. [224] collect data from various platforms, including Voat and Reddit and perform computational analysis to identify the primary motivations that drive users to move to other platforms. [204] collect data from 4chan, Reddit, MetaFilter, and Voat and build a model to detect abusive content online. Subverses used in this work include /v/CoonTown,

/v/N*****, and /v/fatpeoplehate, all focused on hate towards individuals of specific body or race characteristics, created on Voat shortly after the 2015 Reddit bans [167]. Similarly, [205] collect data from Reddit, Voat, and three online forums to train a classifier that detects hateful speech. Their Voat dataset includes data from /v/CoonTown, /v/fatpeoplehate, and /v/TheRedPill. A study on deepfakes finds that pornographic deepfakes are mainly created for circulation within the community [225]. The study uses data from Voat's /v/DeepFake and the site mrdeepfakes.com, which both were created after Reddit banned the subreddit /r/DeepFakes in 2018 [226].

[227] compare the features of 872K comments from /v/politics, /v/television, and /v/travel, to Reddit and 4chan comments building a classifier that predicts the origin of the comments based on its style and content. [168] collect 0.5M posts from /v/GreatAwakening, /v/news, /v/politics, /v/funny, and /v/AskVoat to provide an empirical exploratory analysis of the QAnon community on Voat. They find, among other things, that /v/GreatAwakening is not as toxic as the general discussion subverses. Finally, [169] compare Voat's /v/GreatAwakening and /v/news posts to 4chan, 8kun, Reddit, and Q drops (posts posted by "Q," the mastermind behind the QAnon conspiracy theory) on a large scale study on QAnon. They find that Voat posts are as threatening as Q drops and that content creators on Reddit and Voat only consist of a small portion of the total community.

Other datasets. One of the largest Reddit datasets is the one of [228], which presents an archiving platform that collects Reddit data and makes them available to researchers since 2015. The same platform also published over 27.8K channels and 317M messages from 2.2M users from Telegram [229]. [77] release a dataset of 37M posts, 24.5M comments, and 819K user profiles collected from Gab. [230] published a dataset consisting of 183M posts and 13.25M user profiles from Parler, a Twitter (now X) alternative. Last, [207] present a dataset with over 3.3M threads and 134.5M posts from the Politically Incorrect board (/pol/) of the imageboard forum 4chan.

3.8 Conclusion

In this work, we present and release a Voat dataset comprising more than 2.38M submissions and 16.2M comments posted from 113K users in over 7K Voat subverses. We combine data collected from Voat API and IAWM released archives to complete the dataset to the best of our ability. Voat shut down on December 25, 2020, and its data are now otherwise inaccessible. In this work we also perform a preliminary analysis of the released dataset so researchers interested in it can know what to expect.

Overall, we hope this work further motivates and assists researchers focusing on deplatforming and how users organise migrations to other platfroms. In addition, our dataset could also help answer numerous questions about how "free–speech" sites operate, e.g., do moderators ban users that express opinions other than the ones aligned with the narratives of a subverse? How do users vote and how toxic are they towards such content? Do sites like these incentivise users to form echo chambers? What kind of content do users in these communities consume, etc.? Also, our dataset could assist multi–platform studies to understand similarities and differences of different communities. Last, since Voat was a bastion of free-speech, we are confident that access to our dataset could assist researchers towards training algorithms in natural language processing and detecting hate speech, fake news dissemination, conspiracy theories, etc. Finally, other than quantitative work, we hope that the data can also be used in qualitative studies of specific events, social theories, and communities.

Chapter 4

The Systemic Impact of Deplatforming on Social Media

Introduction

Voat's case study in the previous chapter highlights the velocity at which communities migrate to an alternative platform, after being banned from established social media outlets. This resilience against deplatforming policies raises a few questions related to the efficiency of such interventions from social media administrations. Not only does it indicate that communities can seamlessly regroup on free speech-friendly platforms, but alt-tech platforms are weakly regulated and poorly monitored, with evidence to suggest that they allow violent narratives to develop and thrive [231, 232, 233, 234, 235, 236, 237].

In this chapter, we present a unique dataset which further enhances our understanding of online collective displacements, focusing on a matched cohort of users who migrated from Twitter to Gettr, a Twitter-clone that has attracted many of Twitter's most high-profile suspended accounts including US congresswoman Marjorie Taylor-Greene, media executive Steve Bannon, and conspiracy theorist Alex Jones. Unlike Voat, Gettr leveraged a roster of influencers within the Republican sphere in the US and in Brazil, and has become an attractive online outlet.

Our dataset presents the near-complete evolution of Gettr from its founding in July 2021 to May 2022 including 15M posts from 785,360 active users who have posted at least once. Of these users, 6,152 are verified, 1,588 of which self-declare as active on Twitter (see Methods). For these 1,588 self-declared Twitter users with a verified Gettr account, we download their Twitter timeline from July 2021 to May 2022 totalling 12M tweets and retweets. These users represent the "*matched*" cohort, with analysis of their Gettr posts (Twitter tweets) referred to as "*matched Gettr*" ("*matched Twitter*") below. A manual check of these accounts confirms that 95% of matches across the two platforms are accurate, corresponding to the same individual or organisation (see Appendix). For the remaining verified Gettr users, we use the Twitter API to identify those accounts which have been suspended from the platform, assuming accounts share the same username on both platforms, totalling 454 accounts who constitute the "*banned*" cohort. Finally, all remaining users who are not verified on Gettr are part of the "*non-verified*" cohort.

In the remainder of the chapter, we will overview account activity and retention on Gettr, showing that the banned cohort are 5 times more active than the matched cohort. Despite this, our results will show that these two cohorts are structurally mixed on Gettr, sharing the same politically homogenous audience and posting similar content. Using matched cohort tweets, we will show that Gettr is generally representative of the US far-right, and that matched users are more toxic on Twitter than they are on Gettr. We find little evidence to support the view that users become more toxic as a result of their extended use of the fringe platform. Finally, we will highlight how Gettr had a global impact, outlining the structural changes in the Portuguese-language Gettr network that emerged in the run up to the January 2023 riots in Brazil.

Results

User acquisition and activity

We start by analysing how the three cohorts of "matched Gettr", "banned" and "non-verified" users joined Gettr. Figure 4.1A shows that user registrations were largely steady over time with two exceptions where registrations peaked: (i) July 2021 when the platform was founded, and (ii) January 2022 after the suspension of Marjorie Taylor-Greene and Robert Malone on Twitter [238], and following the announcement by Joe Rogan that he would be opening a Gettr account [239].

In Figure 4.1B, we show the fraction of accounts from each cohort who are active on any given day. For the matched cohort, we present their activity on both Gettr and Twitter. Focusing on the non-verified cohort, we see that a growing user base does not correlate with the growth of an engaged community, with, on average, 4% of the non-verified cohort active on any given day. On Gettr, 10% of the matched cohort are active on average, likely exceeding the value for the non-verified cohort because verified social media users are typically more active than other users [240]. However, on Twitter the matched cohort are significantly more active with 69% of accounts active any given day. The activity of the matched cohort on Twitter is stable, with no evidence of a reduction in activity following the January 2022 suspensions. For the banned cohort on Gettr, activity approaches the baseline of the matched cohort on Twitter, with 53% active daily, 5-times larger than for the matched cohort on Gettr, and 13-times larger than the non-verified cohort. These results are qualitatively robust if we consider exclusively English-language accounts or Portuguese-language accounts, the first and second largest Gettr demographics, respectively (see Appendix). A previous study has also shown a similar increase in activity for banned Twitter and Reddit users active on Gab

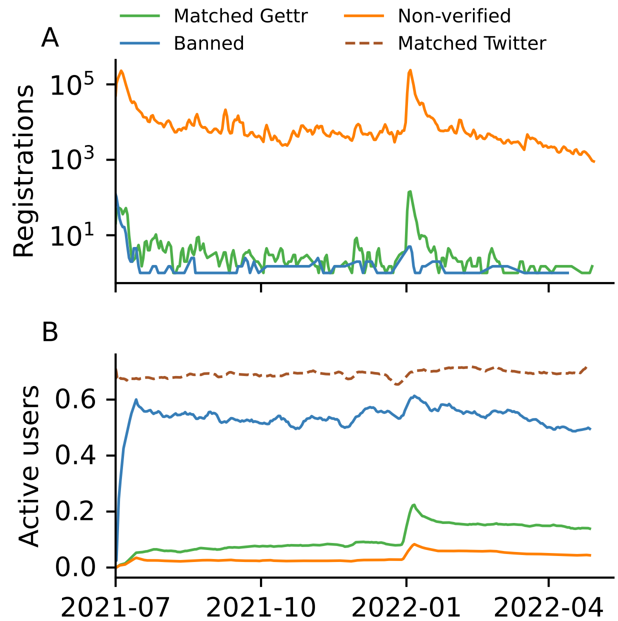


Figure 4.1: **User registrations and daily activity for each cohort.** (A) 3-day moving average of the daily number of users who registered on Gettr. The curve is displayed separately for the banned cohort (blue), the matched cohort (green) and other non-verified users on Gettr (orange). (B) 7-day moving average of the proportion of users from each cohort who were active on Gettr on a given day. The percentage of the matched cohort active on Twitter is also shown (dashed brown).

[241].

User retention on Gettr

We now focus on the retention of users on Gettr. In Fig. 4.2, panels A and B show the survival curves for the proportion of users who remain active a certain number of days after registration (see Methods) for key registration months (July 2021 and January 2022 where registrations peaked, see Fig. 4.1), while panel C shows the average retention of users in each cohort over time. Survival curves for other registration months are shown in the Appendix and follow the same pattern with higher banned retention than matched retention. The matched cohort are consistently active on Twitter with no evidence that users stop using the platform over time: 90% of the matched cohort are active in the first month covered by our dataset (07/21), and 98% of these users remain active in our dataset's final month (04/22). This highlights that the matched cohort are established Twitter users who are committed to the platform.

Figure 4.2 shows that the banned cohort have the highest retention on Gettr, independent of the month in which they joined the platform, whereas the non-verified cohort and the

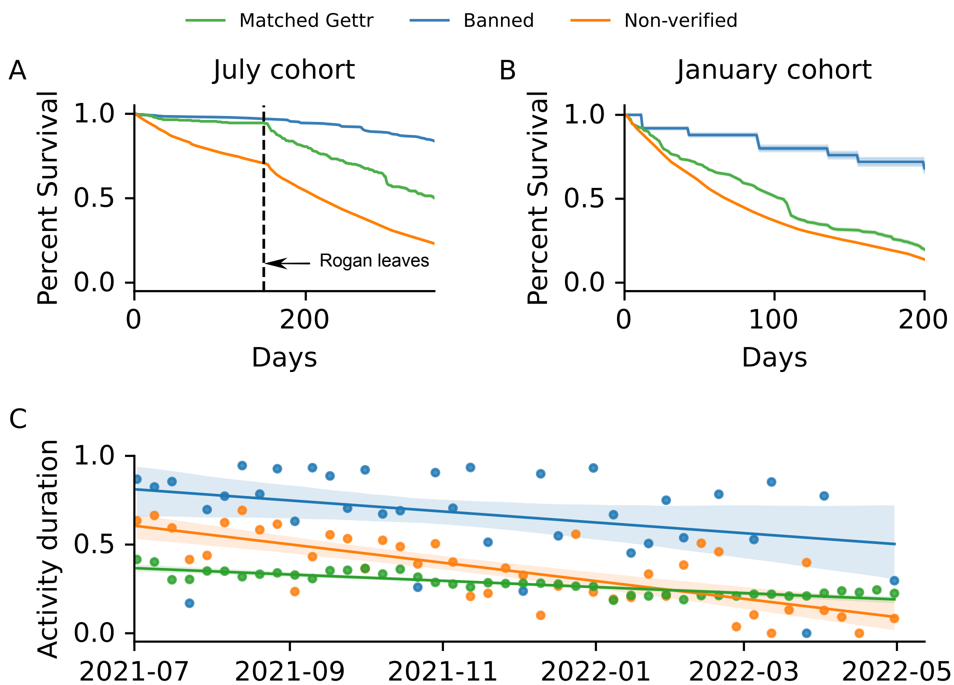


Figure 4.2: **User retention for key registration months and average retention by registration date over time.** (A) Kaplan-Meier survival curves for each user cohort showing the fraction of accounts who registered in July 2021 who remain active on Gettr a given number of days after registration for the banned cohort (blue), matched cohort (green) and the non-verified cohort (orange). The standard error of each curve is computed using Greenwood's formula [1] (see Methods). The dashed line corresponds to January 1, 2022, shortly before Joe Rogan joined Gettr. (B) Survival curves for January 2022. (C) Decay curves for user activity, showing the duration of their activity with respect to their registration date, normalised by the number of weeks to the end of our data collection period. Data for each cohort is fitted using linear regression ($y = ax + b$, $a = -0.007, [-0.014, 0]$, $b = 0.8, [0.65, 0.95]$ for banned users, $a = -0.011, [-0.015, -0.008]$, $b = 0.6, [0.52, 0.67]$ for matched users, and $a = -0.003, [-0.004, -0.002]$, $b = 0.36, [0.34, 0.37]$ for non-verified users; square brackets indicate 95% confidence interval, highlighted by shaded area.)

matched cohort become inactive at a faster rate. For the highlighted registration months, we note that the January curves fall off at a sharper rate than the July curves. For the July cohort, half of the newly registered users from the non-verified cohort become idle after 216 days, compared to only 68 days for the January cohort.

The event which clarifies these differences is the Marjorie Taylor Greene deplatforming on Twitter. This deplatforming was denounced by Joe Rogan who opened a Gettr account on January 2, 2022, resulting in a large migration of his supporters, and supporters of Marjorie Taylor Greene, to Gettr. However, after criticising the platform's policies [242], Rogan quit the platform on January 12. This ten-day period highlights how a single celebrity's endorsement resulted in a large migration to Gettr. However, the subsequent denouncement by Rogan not only resulted in many new users quitting the platform (those from the January 2022 cohort in panel B), but also resulted in many existing users quitting, see dashed line in panel A. Importantly, members of the banned cohort who registered in July 2021 did not leave Gettr at an enhanced rate after January 2022. This highlights that users who had the option to return to Twitter did so, but those who could not (due to suspension) continued to use Gettr.

Compared to previous Gettr studies which showed that users become idle shortly after registration [243], possibly due to the lack of engaging content [244], our results reveal the discrepancy between users banned from Twitter and users who remain active on Twitter, indicating that Gettr was most successful at retaining users who had lost their Twitter audience. Our results also show that deplatforming events of exceptional prominence can induce a significant influx of accounts into a fringe platform, but not necessarily a corresponding outflux from the dominant mainstream platform.

Gettr structure and content

In order to further clarify differences between banned and matched users, we now focus on the structure and content of the Gettr social network. We start by generating a topic model using Gettr posts [245] (see Methods). A table of topics and their description is provided in the Appendix. This shows that content on Gettr is dominated by issues of broad relevance to the US political right including (1) Covid-19 – one sixth of all Gettr content, approaching one third in some months – (2) deplatforming from Twitter and other social media platforms, (3) accusations of election fraud and the January 6th insurrection, and (4) broader issues regarding gender, abortion, gun-control, the US supreme court, and race.

Most topics discussed on Gettr are prominent in tweets authored by the matched cohort, however, three themes are disproportionately prominent on Gettr: (i) Accusations of election fraud surrounding the 2020 US election, (ii) resistance to Covid-19 vaccine mandates, particularly in relation to the "Freedom Convoy" protests in Canada, and (iii) the Russian invasion of Ukraine. These are topics which are known to have been targets of the Twitter content moderation team [246, 247, 248].

We now measure whether the banned and matched cohorts are structurally segregated (or polarised) to assess whether the cohorts share the same, or different, audiences on Gettr. This check is important since there is no a priori reason to assume that the banned and matched cohorts are drawn from the same ideological group. We measure structural segregation (or polarisation) using the latent ideology, a well established method which constructs a synthetic ideological spectrum from user interactions on the platform [249, 250, 251] (see Methods). This measure orders the network of interactions between a set of influencer accounts (the banned and matched cohorts combined) and a set of accounts who interact with them (the non-verified cohort). By merging the banned and matched cohort into a single group, we can measure differences in how the non-verified cohort interact with banned and matched users in an unbiased manner based on purely structural factors. Note that we exclude a small number of non-USA based accounts from the influencer set to avoid geographical conflation (see Methods).

The distribution of the latent ideology for the banned and matched cohort, and for the non-verified cohort, is shown in Fig. 4.3A. Both distributions are unimodal according to Hartigan's diptest [252]. We observe that the banned and matched distribution falls within the bounds of the broader non-verified distribution. The banned and matched distribution is, however, significantly narrower, a feature indicative of the network centrality of these users who play a central role in the general Gettr discussion. Non-verified Gettr users are found both at the core of the Gettr discussion and at the peripheries. The central role of banned and matched users is expected since verified social media accounts typically attain higher engagement than non-verified accounts [253, 254].

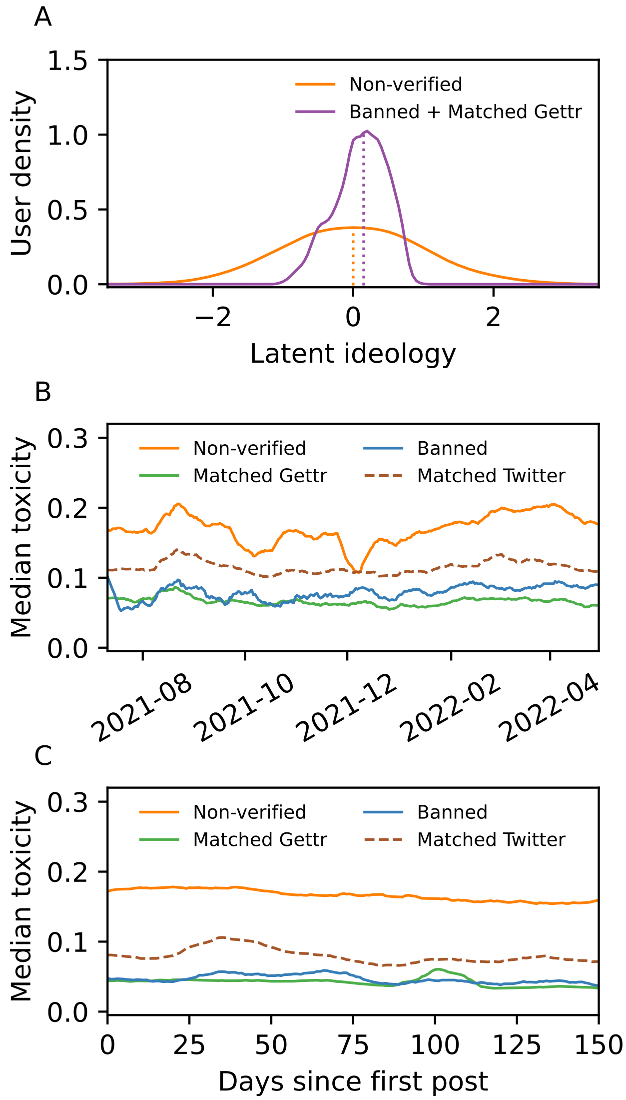


Figure 4.3: **The latent ideology of Gettr users, and the toxicity of Gettr posts and Twitter tweets.** (A) The latent ideology is calculated using the 500 most active banned and matched users on Gettr, merged into a single influencer cohort. Unit values on the x-axis correspond to the standard deviation of the ideology distribution for all users. Both distributions are unimodal when tested using Hartigan's diptest (multimodality not statistically significant for the non-verified cohort, $p = 0.99 > 0.01$, or banned and matched cohort, $p = 0.61 > 0.01$). (B) The median post toxicity each day for each user cohort (14-day moving average). Toxicity is calculated using the Google Perspective API [2] (see Methods). Median toxicity [lower and upper quartile] for the non-verified cohort, 0.17 $[0.06, 0.37]$, banned cohort, 0.05 $[0.02, 0.15]$, matched cohort on Gettr, 0.04 $[0.02, 0.11]$, and matched cohort on Twitter, 0.09 $[0.04, 0.22]$. (C) The median toxicity of posts authored a fixed number of days after a user first posted on Gettr (or Twitter; 14-day moving average). There is minimal evidence of a meaningful increase in user toxicity due to extended Gettr use (see Appendix).

The unimodal ideology, and the central position of the matched and banned cohorts, indicates that these users share a common audience on Gettr; segregated audiences would appear as a multi-modal ideology distribution (see examples in [250, 251]).

Content toxicity and twitter mentions

We now focus on the toxicity of posts from each user cohort, shown in Fig. 4.3B. Toxicity is calculated using the Google Perspective API [2] (see Methods). The panel shows the median daily toxicity of each cohort. Across the observation period as a whole, and by applying a bootstrapping procedure to ensure equal sample sizes (see Methods), we find that posts authored by the non-verified cohort are more toxic than posts by the matched cohort (KS-test $D = 0.36$, $p = 1.3 \times 10^{-57}$), and than posts by the banned cohort (KS-test $D = 0.32$, $p = 8.0 \times 10^{-47}$). We also find that the tweets authored by the matched cohort are more toxic than Gettr posts authored by the matched cohort (KS-test $D = 0.23$, $p = 8.7 \times 10^{-23}$). However, the difference between the toxicity of posts for the banned cohort and matched cohort on Gettr is not statistically significant (KS-test $D = 0.06$, $p = 0.09$).

In order to assess whether individual users are becoming more toxic over time due to their extended Gettr use, we plot the median toxicity of posts authored a fixed number of days after each user first posted on Gettr (or first posted on Twitter during our observation period), as shown in Fig. 4.3C. For the non-verified cohort, the gradient in the change of the toxiciy over time is not significantly different from zero (see Appendix). For the other cohorts, there is a statistically significant but very small non-zero gradient in the toxicity over time. However, this gradient is negligible when considered in the context of the much larger inter-quartile range of post toxicity values for each cohort (see Appendix).

Together, the results for the latent ideology, topic modelling, and toxicity show that, although there are significant differences in activity and retention between the banned and matched cohorts on Gettr (see Figs. 4.1 and 4.2), there is little that distinguishes their audience and content. This result confirms previous research which shows that fringe platforms are politically homogeneous; platforms with this property may be referred to as “echo-platforms” [48, 255]. In contrast, mainstream platforms are often politically diverse, but with opposed political groups confined to echo-chambers [48, 251, 254, 256, 257, 258, 259].

Considering the toxicity of posts for each topic, we find that topics with disproportionately high toxicity are related to race (e.g., Black Lives Matter; median post toxicity [lower and upper quartile] = 0.40 $[0.31, 0.52]$), focus on female US Democratic politicians (0.38 $[0.18, 0.58]$, and discuss gender issues (0.38 $[0.24, 0.51]$). All three topics are known to attract abusive content on social media [260, 261, 262].

We now explore possible reasons why the matched cohort are more toxic on Twitter than they are on Gettr. To do this, we analyse the Twitter accounts mentioned in tweets authored by the matched cohort. For each mentioned account, we compute the ratio between the number of users from the matched cohort who quote-tweet that account and the number of users from the matched cohort who quote-tweet or retweet that account. This ratio (referred to as the "quote-ratio" throughout) is instructive since there is evidence that retweets are often (but not exclusively, journalists being a known exception) used to endorse the message of the original author [251, 263], whereas quote tweets allow a user to comment on a message in either a positive, negative, or neutral manner. Negative "quoting" behaviour is a known method of communication with ideological opponents across polarised environments [264, 265]. Hence, a low quote-ratio (i.e., the account is disproportionately retweeted) indicates general endorsement by the matched cohort of users, whereas a high quote-ratio (i.e., the account is disproportionately quote-tweeted) indicates that the matched cohort are more likely to disagree with and hold a negative view of this account.

Figure 4.4 shows the toxicity of tweets authored by the matched cohort, binned according to their quote-ratio. We count each mentioner-mentionee pair only once for quote-tweets and once for retweets to avoid bias from highly active accounts, and only include accounts mentioned by at least five matched users. This reveals (i) that tweets authored by the matched cohort mentioning any Twitter account are more toxic than tweets which do not mention another account, and (ii) that tweets authored by the matched cohort are more toxic if they mention an account with a high quote-ratio than if they mention accounts with lower quote-ratios.

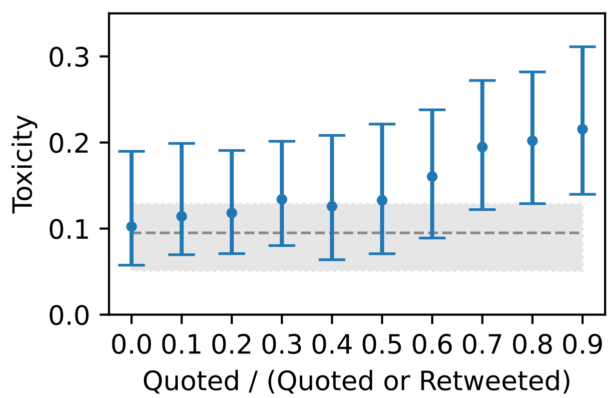


Figure 4.4: **Toxicity of tweets authored by the matched cohort mentioning other Twitter accounts, binned according to their quote-ratio.** The distribution of the quote-ratio is shown in Fig. 4.5. Each point indicates the median toxicity of tweets with a quote-ratio within the binned range $[x, x + 0.1)$. Error bars indicate the inter-quartile range. The dashed line indicates the median toxicity for all tweets (including those which do not mention another account) from the matched cohort, with the shaded region indicating the inter-quartile range; all data points lie above this line.

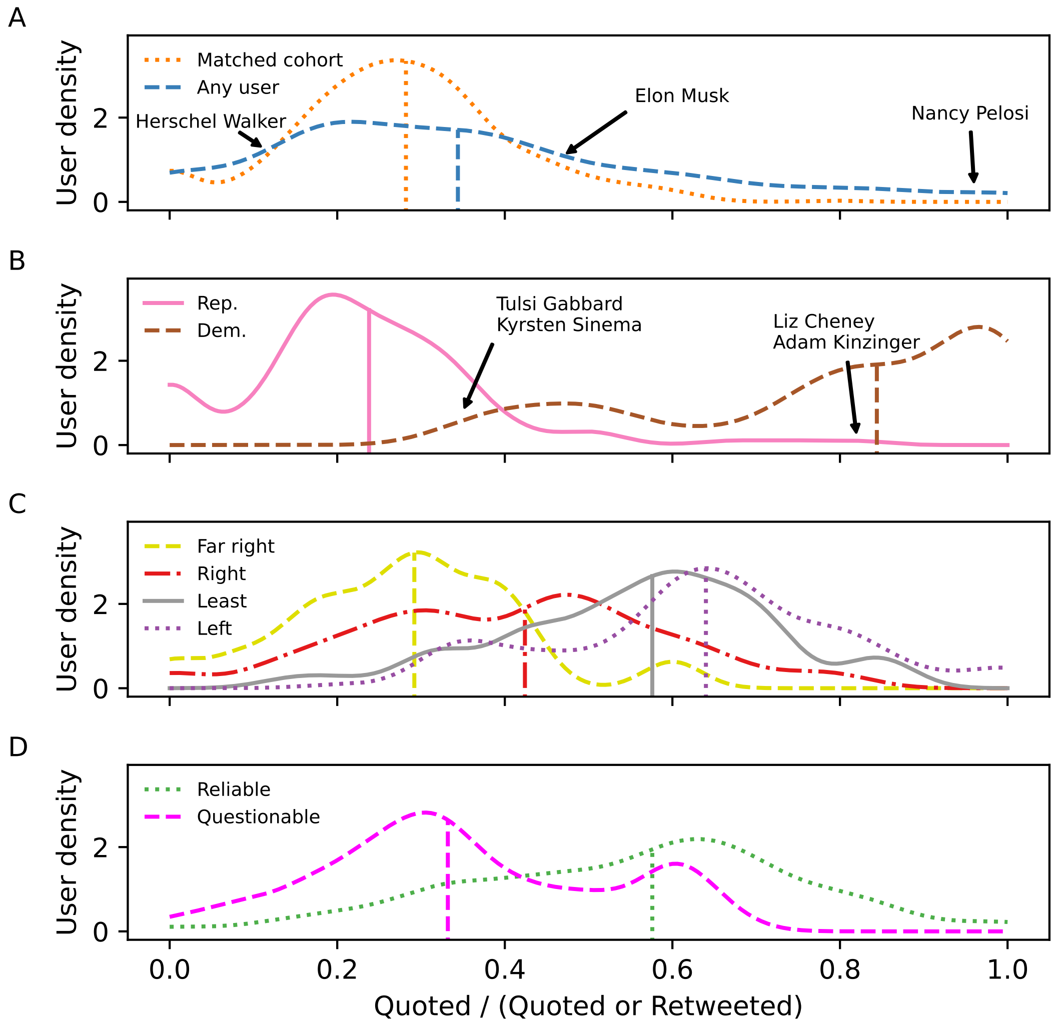


Figure 4.5: **The distribution of the quote-ratio of accounts mentioned on Twitter by the matched cohort.** (A) The quote-ratio distribution for all mentioned accounts (blue dashed), and for mentioned accounts who are part of the matched cohort of users (i.e., a matched user mentioning another matched user; orange dotted). (B) The quote-ratio distribution for Twitter accounts belonging to known elected US Republican (pink solid) and known elected US Democrat (brown dashed) politicians. (C) The quote-ratio distribution for Twitter accounts belonging to news media organisations who have been labelled with a political leaning by MBFC. Organisations are classified as left (purple dotted), least-biased (grey solid), right (red dot-dashed), or far-right (yellow dashed). (D) The same news media organisations, but broken down according to whether they are classified as a reliable or questionable by MBFC. Vertical lines mark the median of each distribution. Annotations indicate mentioned accounts of particular interest (see text).

To better understand this result, we plot the distribution of the quote-ratio broken down into four groups. Figure 4.5A shows the distribution of all users mentioned by the matched cohort, and the distribution for Twitter accounts who are also part of the matched cohort (i.e., a matched account mentioning another matched account). Three individuals are marked on the figure: (1) Republican 2022 Senate nominee Herschel Walker, the user with the lowest quote-ratio of prominent mentioned accounts (> 100 unique mentions), (2) Democratic speaker of the house Nancy Pelosi[a], the user with the largest quote-ratio (> 100 unique mentions), and (3) Elon Musk, the account with the most unique mentions.

Figure 4.5B shows elected US political accounts mentioned by the matched cohort, labelled using the dataset in [266], broken down by party affiliation. This shows that Republican politicians are disproportionately retweeted (i.e., endorsed) by the matched cohort, whereas Democrats are disproportionately quote-tweeted. The individuals marked on this panel are political outliers; Liz Cheney and Adam Kinzinger are the Republican politicians with the highest quote-ratios (> 10 unique mentions), whereas Tulsi Gabbard[b] and Kyrsten Sinema are the Democratic politicians with the lowest quote-ratios (> 10 unique mentions). This shows that these politicians do not align with the dominant position of their parties. Consequently, the matched cohort are more likely to endorse the Democratic outliers, and more likely to negatively quote-tweet the Republican outliers; Liz Cheney and Adam Kinzinger have been referred to as RINOs ("Republicans in name only") by their far-right opponents [267, 268].

Figure 4.5C shows the news media organisations mentioned by the matched cohort, grouping them according to their political leaning as classified by Media Bias / Fact Check (MBFC; see Methods). Previous research confirmed that MBFC classifications are similar to classifications from other reputable media rating organisations [269]. Finally, Fig. 4.5D repeats the analysis in panel C, but groups media outlets according to whether MBFC labels them as reliable or questionable.

Using the distribution of all mentions (the "any user" curve in Fig. 4.5A) as the baseline behaviour of the matched cohort, we find that, when tested using a two-sample Kolmogorov-Smirnov test, only the distributions of far-right media organisations in panel C (KS-test p-value $= 0.24 > 0.01$; Cohen's $d = 0.20$) and questionable media organisations in panel D (KS-test p-value $= 0.29 > 0.01$; Cohen's $d = 0.05$) are not significantly different from the baseline (see Appendix). This observation aligns with previous research showing that the US political right on Twitter are more likely to share questionable news sources, and are more likely to be suspended [270].

The Democrat politicians distribution has the largest statistical difference to the all-mention baseline (KS-test p-value $= 3 \times 10^{-16} < 0.01$; Cohen's $d = 2.34$). With the exception of Tulsi Gabbard, no Democratic politicians has a known Gettr account; 132 are mentioned on Twitter by the matched Gettr cohort. In contrast, 32 Republican politicians have been

[a]Speaker of the house during the timeframe of our dataset

[b]Democrat during our analysis timeframe; left the Democratic party in October 2022.

active on Gettr; 151 are mentioned on Twitter by the matched Gettr cohort.

Combining the evidence from the topic modelling and from the quote-ratio in Fig. 4.5 indicates that the matched cohort are aligned with the US far-right, often quote-tweeting, but not retweeting, their Democratic political opponents and moderate Republicans. In conjunction with the latent ideology in Fig. 4.3, this suggests that Gettr as a whole is generally representative of the US far-right. These results suggest that the ability to mention one's political opponents on Twitter is part of the reason that the matched cohort are more toxic on Twitter than they are on Gettr where direct interactions with political opponents are not possible [271, 272].

Discussion & Conclusion

In this chapter, we have analysed self-declared user-level matching to study the ban-induced platform migration from Twitter to Gettr. First, we showed that the banned cohort of users deplatformed from Twitter are more active on Gettr, and have higher platform retention, than the matched cohort who remain active on Twitter. Second, we revealed that Gettr content primarily discusses themes relevant to the US political right. Topics overepresented on Gettr are known to have resulted in account suspensions on Twitter. Third, we showed that the matched and banned cohorts share the same politically homogeneous Gettr audience. Finally, we found that matched users are more toxic on Twitter than they are on Gettr, and that these toxic tweets often directly mention political opponents. We find little evidence of a meaningful increase in user toxicity over time.

The fact that the banned and matched cohorts appear similar in every regard, apart from their activity and retention on Gettr, is evidence of the systemic impact of deplatforming. Fringe platforms offer a safe haven where deplatformed users are free to capitalise on their supporters following suspension from Twitter. However, in this politically homogeneous environment, users are essentially confined to an ideological "echo-platform" [48, 255] where they cannot engage and confront their political opponents. Our results hint that this ability to interact with opponents may be part of Twitter's appeal for far-right social media users, although more work is needed to fully clarify this observation.

When users are banned from mainstream platforms, they become wholly dependent on the fringe alternatives (despite Gettr also suspending some users, notably white supremacist Nicholas Fuentes [273]). This may pose a societal risk since fringe platforms are believed to facilitate the emergence of radical narratives and the spread of hate speech [79, 90]. A lack of monitoring can, therefore, mean that signs of collective upheaval are missed. The Brasília insurrection, which took place in January 2023 following Jair Bolsonaro's defeat in the presidential election, is an example of this, as the election fraud allegations were widely spread by Steven Bannon's podcast *The War Room*, streaming regularly on Gettr while being banned from Twitter [181, 274]; an analysis provided in the Appendix shows how Portuguese language Gettr activity rose in the weeks preceding the riots.

These results complement and add to existing work which considers the effect of mainstream deplatforming on users' behaviour on Gab primarily on the basis of content analysis [241]. However, while the Gab study finds an increase in user toxicity over time, we find little evidence of a similar increase on Gettr. This difference may indicate that changes in toxicity (or lack thereof) depend on the fringe platform used after suspension, rather than on deplatforming itself, but more work is needed to validate this hypothesis in the context of the wider social media ecosystem.

It is important to contextualise the scope of our findings, whose limitations are avenues for future work. First, the current study only considers the migration from Twitter to Gettr, since users of other platforms do not declare their Twitter use as standard. If data becomes available, future work should extend scrutiny to multiple other platforms, ideally in a unified study. Second, the Gettr matching feature only applies to verified users, a subset of the users who migrated from Twitter to Gettr. Analysing non-verified users migrating across platforms would clarify the differential impact of deplatforming on content creators as opposed to consumers. Finally, we cannot study tweets from the banned cohort since this data is not publicly available. Analysing this content would explain why some users are deplatformed, but others are not.

Overall, our study highlights how Gettr struggles to compete with Twitter when users have free choice to use either platform. However, the decision by Twitter to deplatform a user impacts how those users view Gettr as an alternative. We anticipate that future work will build on these observations and speculate that other fringe platforms will likely show a similar dependence on their mainstream competitors. This work is urgently needed given the potential risks posed to democracy by poorly regulated social media [275, 276].

Methods

Gettr data

The data used for this study has been collected using GoGettr [277], a public client developed by the Stanford Internet Observatory to give access to the Gettr API. This API allows to query user interactions, including the posts they like or share. User profiles were initially collected through a snowball sampling, by using highly popular accounts on the platform as seed users, and using the API to query their follower and following list, before repeating the same process for a random sample of the newly retrieved users. Repeating this process many times ensures that our dataset is near-complete for the studied time period.

To attract more users from Twitter, Gettr previously offered a feature that would automatically import a user's tweets upon creating an account. However, due to Twitter blocking this capability on July 10, 2021 [278], Gettr had to discontinue this feature. To ensure the accuracy of our results, any posts imported before July 10, 2021, and any Gettr post whose timestamp precedes the account's creation date were removed from our dataset.

To ensure our case study on the Brazilian right encompasses the Brasília insurrection, we expanded the data collection time frame for any user associated with the Brazilian community. The data collection was run in July 2022 for every user whose profile we have retrieved, and in January 2023 for the users in the Brazilian cohort.

Twitter data

For each verified Gettr account where the Gettr API references their Twitter followers in the account metadata, we check that the Twitter account with the same username is active using the Twitter API (see https://developer.twitter.com/en/products/twitter-api/academic-research). We identify accounts who were previously active on Twitter but are now banned from the "HTTP 403 Code 63" error message corresponding to suspended Twitter accounts. Other error messages are used for protected or not found accounts. Our study does not consider users banned from Twitter who did not join Gettr, or joined Gettr using a different username to their original Twitter username.

For each active account we download their Twitter timeline including all tweets and retweets in the period July 2021 to May 2022. This totals approximately 12 million tweets. Data was collected between September and October 2022, preceding Elon Musk's amnesty of suspended Twitter accounts.

User labelling

Throughout our analysis, we label verified Gettr users as being either "matched" or "banned", depending on whether their corresponding Twitter account is active or suspended. Any verified user who decided to link their Twitter account on their Gettr page has their Twitter follower count displayed on their profile, which can also be retrieved from the Gettr API. We stress that this self-declaration permits cross-platform matching since users can "reasonably expect" that their Gettr accounts will be associated with their Twitter accounts.

To match accounts across platforms, we assumed that users picked the same username on both Gettr and Twitter, and we used the Twitter API to retrieve their Twitter activity. A user is classified as "banned" if the Twitter user endpoint returned an error indicating that the account has been suspended (Error "HTTP 403 Code 63").

To validate the accuracy of this matching process, we manually check whether matched accounts correspond to the same organization or individual on both Gettr and Twitter (see Appendix). This reveals a 95% matching accuracy. Note that following Elon Musk's amnesty on banned accounts, approximately one third of accounts in the banned cohort have been reactivated. Of the top 100 most followed accounts on Gettr from the banned cohort, 33 have been unbanned on Twitter following the Musk amnesty, and all 33 are an exact match for the same individual or organisation across the two platforms. However, due to the discontinuation of the Twitter API for academic purposes, the detailed analysis of these unbanned accounts on Twitter is not possible.

For data privacy reasons, all analysis of users across platforms is aggregated at the cohort level; we do not present results for individual users.

Calculating account survival

The Kaplan-Meier estimate is a tool used to quantify the survival rate of a population (in our case, users active on a social platform) over time. For each time step t, we measure how many users become indefinitely inactive, and we quantify the survival rate as

$$\hat{S}(t) = \prod_{t_i \leq t} \left(1 - \frac{d_i}{n_i}\right), \tag{4.1}$$

where d_i represents the number of users who became inactive at time t_i, and n_i is the number of users who are still active up to time t_i. Greenwood's formula is used to estimate the confidence interval for the Kaplan-Meier estimate of the survivor function. For the study time t, the standard error is given by

$$\widehat{SE}_2(t) = \sqrt{\hat{S}^2(t) \sum_{t_i \leq t} \frac{d_i}{n_i(n_i - d_i)}}. \tag{4.2}$$

Topic Modelling

Gettr and Twitter content is analysed using the BERTopic topic modelling library [245]. This method extracts latent topics from an ensemble of documents (in our case Gettr posts and Twitter tweets). The base model uses pre-trained transformer-based language models to construct document embeddings which are then clustered.

These methods are known to struggle with very short documents which are common on micro-blogging sites. Hence, we train our topic model using exclusively Gettr posts which are longer than 100 characters. To avoid any single user dominating a specific topic, we limit the training set to no more than 50 posts from any given user.

Latent ideology

To calculate the latent ideology on Gettr, we use the method developed in [249, 250] and filtering procedures from [251]. The method uses a bipartite approach where it classifies Gettr accounts as influencers or regular users. It then generates an ordering of users based on the interaction patterns of regular users with the influencer set. In the current chapter, we select the matched and banned user sets as our influencers, and the remaining set of Gettr users as our cohort of regular users.

Two factors can conflate the latent ideology: (1) account geography, and (2) a lack of user-influencer interactions. Since we are interested in the segregation of the banned and matched cohorts based on political ideology, the former is problematic because country-specific communities on social media can appear structurally segregated from a related community in other countries, even if they are politically aligned. For this reason, we remove a small number of accounts associated with the UK and China from our set of Gettr influencers. In the latter case, a lack of user-influencer interactions can be problematic since influencers with few user interactions appear as erroneous outliers when computing the latent ideology, often because they do not take active part in the conversation. Hence, we restrict our influencer set to the 500 banned and matched accounts who receive the largest number of interactions from other users on Gettr. In the current study, we consider any interaction type including comments, shares and likes. The latent ideology is robust as long as the number of influencers used is larger than 200 accounts [251].

In order to assess the modality of the ideology distributions, we use Hartigan's diptest. This approach is used to identify polarised social media conversations and echo-chambers [250, 251]. Hartigan's diptest compares a test distribution against a matched unimodal distribution to assess distribution modality [252].

The test computes the distance, D, between the cumulative density of the test distribution and the cumulative density of the matched unimodal distribution. The D-statistic is accompanied by a p-value which quantifies whether the test distribution is significantly different to a matched unimodal distribution. A p-value of less than 0.01 indicates a multimodal distribution.

Toxicity analysis

The toxicity of Gettr and Twitter content is computed using the Google Perspective API [2], which has been used in several social media studies to assess platform toxicity [83, 279, 280].

Given a text input, the API returns a score between 0 and 1, indicating how likely a human moderator is to flag the text as being toxic. For our analysis, we used the flagship attribute "toxicity", which is defined as "[a] rude, disrespectful, or unreasonable comment that is likely to make people leave a discussion" [281].

When computing statistics for the toxicity of each user cohort, we apply a bootstrapping procedure to avoid erroneous results from variable cohort sizes. This is important since the distribution of post toxicity is fat tailed; there are far more posts with low toxicity than high toxicity. Therefore, a smaller set of posts from a user cohort may have a lower median toxicity, purely due to sampling effects. To avoid this conflation, bootstrapping is employed where equally sized samples are drawn from each cohort (usually 100 posts), and the median toxicity is computed for each sample. Then, the sampling procedure is repeated 100 times to compute the median and inter-quartile range for the sampled post toxicity.

News media classification using Media Bias / Fact Check

For the quote-ratio analysis in Fig. 4.5, we identify the Twitter handles of news media outlets and classify their political leaning using Media Bias / Fact Check (MBFC; see `https://mediabiasfactcheck.com/`). Ratings provided by MBFC are largely similar to other reputable media rating datasets [269].

MBFC classify news outlets under seven leaning categories: extreme left, left, center-left, least (media considered unbiased), center-right, right, extreme right. To ensure that we have enough news media outlets to enable the quote-ratio analysis, we group these classifications into four larger groups: Left – (extreme left, left, center-left), Least – (least), Right – (center-right, right), Far-right – (extreme right). Note that we have chosen to use the terminology "far-right" instead of "extreme-right" since the former is more common in the academic literature.

Chapter 5

The Koo Dataset: An Indian Microblogging Platform With Global Ambitions

Previous chapters have emphasised the increasing influence leveraged by alt-tech platforms in the overall social media ecosystem. However, our analysis so far mostly revolved around US-based communities. Therefore, there is a growing need to understand how alt-tech platforms thrive beyond Western nations.

Koo, a microblogging platform based in India, had emerged as a major new social network hosting high profile politicians from several countries (India, Brazil, Nigeria) and many internationally renowned celebrities. This chapter presents the largest publicly available Koo dataset, spanning from the platform's founding in early 2020 to September 2023, providing detailed metadata for 72M posts, 75M comments, 40M shares, 284M likes and 1.4M user profiles. Along with the release of the dataset, we provide an overview of the platform including a discussion of the news ecosystem on the platform, hashtag usage, and user engagement. Our results highlight the pivotal role that new platforms play in shaping online communities in emerging economies and the Global South, connecting local politicians and public figures with their followers. With Koo's former ambition to become the town hall for diverse non-English speaking communities, our dataset offers new opportunities for studying social media beyond a Western context.

Introduction

Launched in India in 2020, Koo had grown to become the second largest microblogging platform worldwide [282], and had attracted a number of political communities who have been critical of Twitter for censoring their discourse [283]. Most notably, Koo had become a leading social platform used by the Bharatiya Janata Party (BJP), the governing political party in India.

Given the dominance of the BJP on Koo, the platform appeared to be similar to many other alt-tech platforms with a focus on right-wing politics. However, beyond India, Koo had successfully attracted a politically diverse set of users, including both supporters and opponents of former Nigerian president Buhari, and of Brazilian president Lula. This success highlights how Koo had the potential to move beyond the politically homogeneous model of most alt-tech platforms to become a major politically heterogeneous venue for both political and apolitical discussions across multiple countries, challenging the US social media hegemony.

Data Release. In this work, we present the most extensive Koo dataset currently available, including 71.7M posts, 74.6M comments, 283.5M likes, 40.0M shares, and 1.4M user profiles. Our dataset spans from Koo's launch in early 2020 up until September 2023. Due to the ability to paginate through a user's complete timeline to access historical activity via the API, we are confident that our dataset provides a near-complete overview of the interactions carried out by users who were actively using the platform.

Relevance. Our dataset provides researchers with the opportunity to study an alt-tech platform based in India which had attracted an international community of high-profile politicians and celebrities, primarily from India, Brazil and Nigeria. With BJP politicians heavily endorsing Koo, our dataset enables the study of political rhetoric through the analysis of the political content shared online. Previous research has shown that the BJP promote islamophobic [284] and populist [285] rhetoric. It is therefore of interest to consider whether their claims on Koo aimed to polarise opinions and stoke hate in their audience. Moreover, with many news outlets shared on Koo, the dataset will enable the study of political discourses through an analysis of the alignment between news media outlets' editorial line and the national political parties' ideologies [286, 287]. Finally, the dataset can contribute to cross-platform comparisons of political rhetoric.

Chapter organization. In this chapter, we first explain Koo's structure before detailing its impact on the social media ecosystem. Second, we describe the method used to query posts, interactions and user profiles from Koo's public API before describing the structure of our dataset. Finally, we present results which provide an overview of the Koo platform, before discussing our results and the importance of our dataset in the conclusion.

What was Koo?

Koo was a multi-lingual microblogging platform launched in early 2020 by Bombinate Technologies, a company based in Bangalore, India. Social interactions on Koo worked in a very similar way as interactions on other microblogging platforms (e.g., X): Users could create an account and then follow, or be followed by, other users. User profiles could be personalised with a profile-picture and additional information including a user description, personal title, and links to other social platforms. User profiles also displayed the account username and creation date.

Once logged in, users could submit public posts, called *koos*. A user could like a post, comment on it, or share it (a *rekoo*). Koo also provided the ability to translate a koo to many languages. Since March 2023, the platform had integrated ChatGPT in a bid to increase the number of users actively creating content for the platform [288].

Figure 5.1 provides an example of a post on Koo. Posts had a 500 character limit and might include images and videos. The panel under each post indicated how many times the koo had been commented, liked and shared by other users, and allowed the viewer to share the koo on other platforms. The profile panel, located above the koo, displayed the user's username and their self-indicated title. The yellow tick at the right of the username, known as the yellow tick of eminence, was granted by Koo administrators to accounts considered "significant representative[s] of the Voices of the World" [289]. Other users could earn a green badge by self-verifying their account.

The demographics of Koo

Koo aimed at becoming a central outlet for non-English speaking communities online by targeting countries where the uptake of traditional social media has been low [290]. The platform had experienced three key surges in user registrations. First, in February 2021 Indian users joined Koo following an open conflict between Narendra Modi's cabinet and Twitter, after the government ordered Twitter to remove more than 1,000 accounts they alleged were responsible for spreading misinformation around the Indian farmers' protests [290]. Many government officials and celebrities endorsed the platform in the wake of the conflict, including the chief minister of Uttar Pradesh Yogi Adityanath and cricket player Virat Kohli. Indian Prime Minister Narendra Modi has expressed his support for Koo but did not sign up on the platform [291]. Previous research has highlighted that about 80% of the politicians who signed up on Koo are members of the BJP [292]. Consequently, Koo had been accused of supporting nationalist politics in India, despite its attempts to attract other political parties [149].

Second, in June 2021, Twitter temporarily suspended then Nigerian president Muhammadu Buhari from the platform after he threatened his political opponents with violence. This led Buhari's government to instate a nationwide Twitter ban on June 5, 2021 [293]. Afterwards, Koo experienced a significant uptake among Nigerian government officials and civilians, both supporters and opponents of the Buhari regime, leading the platform to roll out accessibility for the main vernacular languages spoken in Nigeria [25]. Koo quickly gained legitimacy within the country, becoming an official channel of communication for the government alongside Facebook and Instagram. Twitter was reinstated in Nigeria in January 2022, after they agreed to conditions set by the Nigerian government [294], leading to a reduction in Koo's usage from 289,000 monthly active Nigerian users to 40,000 by September 2022 [295].

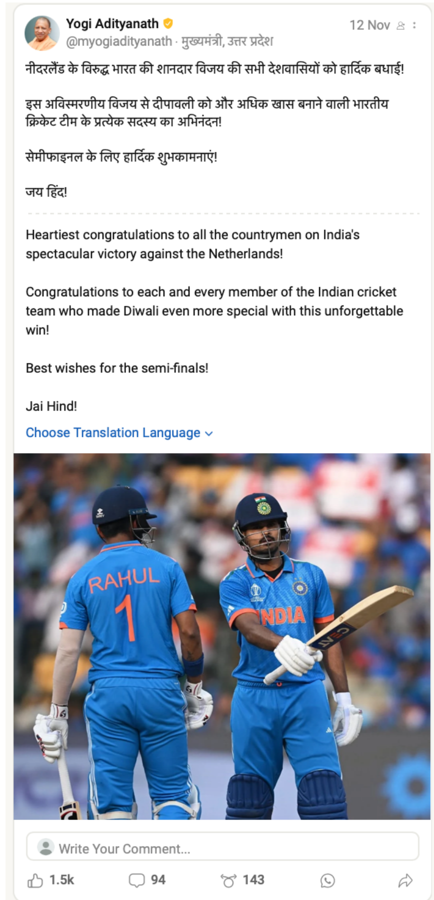


Figure 5.1: **An example of a koo.** The main panel includes the original post in Hindi, its translation in English and an image. The top panel provides information about the poster, including their user handle, profile picture, their self-declared title, the yellow tick of eminence (if applicable), and the post creation date. The bottom panel allows logged-in users to comment, share or like the post. Koo provides additional icons to share posts on other platforms.

Finally, in November 2022, shortly after Twitter was purchased by Elon Musk, Koo attracted a large Brazilian community following a Koo-related linguistic pun posted by Felipe Neto, a Brazilian YouTuber and online influencer [26]. Consequently, Brazilian president Luiz Inácio Lula da Silva and many of his supporters joined Koo; Lula gained over 50,000 followers in less than 4 hours [296] with Koo becoming the most downloaded app on the Apple App Store and the Google Play Store for a few days [297].

Table 5.1 shows the total number of posts, comments, likes and shares made in the top 10 languages on the platform, as well as the percentage of posts made in each language. Hindi makes up about half of the total messages, highlighting the dominance of the Indian community on Koo. English, Portuguese and Nigerian English are the only non-Indian languages which had significant use on the platform.

Language	Posts	Ratio (%)	Comments	Likes	Shares
Hindi	35,074,082	48.9	33,762,199	141,851,555	22,710,161
English	21,341,502	29.8	15,364,430	59,405,972	10,745,495
Portuguese	4,933,641	6.88	9,747,183	51,323,645	1,521,073
Telugu	2,127,279	2.97	2,218,213	1,975,845	104,001
Kannada	1,889,473	2.63	3,700,669	5,972,733	859,069
Marathi	1,286,933	1.79	913,997	5,237,631	128,886
Tamil	1,268,948	1.77	376,721	1,262,414	82,940
Bengali	1,194,315	1.67	1,031,905	2,912,268	203,597
Gujarati	1,014,200	1.41	504,975	2,991,499	103,317
Nigerian English	610,568	0.85	1,076,604	1,893,175	571,813

Table 5.1: The top 10 languages used on Koo. The ratio indicates the percentage of the total number of posts written in each language. Columns indicate, for each language, the number of comments, likes and shares associated with the language.

Focusing on high-profile Koo users, Table 5.2 shows the fraction of users with a green (verified) or yellow (account of eminence) badge for the 10 largest linguistic communities on Koo. Over a quarter of Portuguese-speaking users had received a green badge, the largest percentage for any linguistic community, indicating heavy use of Koo's self-verification feature. In contrast, only 0.33% of Nigerian English accounts were self-verified. Hindi-speaking users also displayed a high level of adoption for the self-verification process. For each linguistic community, only a small fraction of accounts were awarded the yellow tick of eminence, with a slight bias favouring Indian communities.

Language	Self-Verified (%)	Accounts of Eminence (%)	User Profiles
Hindi	10.8	0.79	552,941
English	7.33	1.37	212,244
Portuguese	27.96	0.5	178,471
Telugu	4.05	0.35	16,265
Kannada	2.95	0.93	25,547
Marathi	5.81	0.63	11,226
Tamil	5.3	1.15	5,021
Bengali	5.66	0.34	14,619
Gujarati	5.51	0.58	9,931
Nigerian English	0.33	0.31	17,184

Table 5.2: Ratio of self-verified accounts and those with a yellow tick of eminence for the top 10 linguistic communities on Koo. The table also indicates the number of user profiles provided for each linguistic community.

Studies have shown that two accounts who share a common set of social media interactions typically share similar political or ideological views [251, 298]. To map these connections on Koo, we construct the co-occurrence network of Koo users with the account of eminence badge following the methodology outlined in [298], see Fig. 5.2. Each node in the co-occurence network corresponds to a single eminent user, coloured according to their modal post language. Two eminent users are connected by an edge if their posts were liked by at least 50 common users. For visual clarity, we only show the giant connected component, and eminent users with edges to at least two other users.

The network shows that clusters are strongly influenced by linguistic factors: prominent users were more likely to be part of a similar social circle on Koo if they predominantly used the same language on the platform. Aside from the isolated Portuguese-speaking cluster and the Nigerian English speaking users, eminent users speaking Kannada, Telugu and Gujarati are also slightly disconnected from the English/Hindi dominant community. However, there is little evidence of structural polarization within linguistic communities on Koo, unlike on Twitter (now X) [298].

Data Collection

Koo's API. Koo provided a public API, which could be accessed by sending a GET request to **https://www.kooapp.com/api/**. All data included in this dataset was retrieved by querying the user profile endpoint: knowing a user ID or account handle, the API provided specific endpoints to retrieve their posts, comments, shares and likes.

User Profile Collection. In order to collect a user's activity, we first need to retrieve their profile to get their user ID. We started with a short list of manually collected high-profile users, including prominent politicians, who were active on Koo. Through their interactions on Koo (see below) we identify any user they had interacted with. We query the user profile of these accounts, and repeat this process iteratively, collecting more users through a snowball sampling method until we retrieve no new user profiles. A limitation of this data acquisition strategy is that we do not capture lurkers on Koo (users who used Koo passively). However, we do capture all the interactions for every user who interacted with any account in the giant connected component on Koo. The user profiles provided by the API included information on a user's username, display name, profile description and manually entered title.

User Activity. Given a user ID, Koo's API provided specialised endpoints to query the user's posts, comments, shares and likes. The endpoints were respectively **profile/createdKus**, **profile/commentsKus**, **profile/rekooKus** and **profile/likedKus**, with the user ID provided as a query parameter. We paginate through the results to ensure that all of a user's posts and interactions are collected.

Ethical Considerations. The data released with this chapter was collected using Koo's public API, in accordance with the platforms' Terms of Service. Koo did not offer the possibility for users to make their account private, therefore the data we are releasing contains publicly accessible information only. We do not release metadata related to geolocation. All

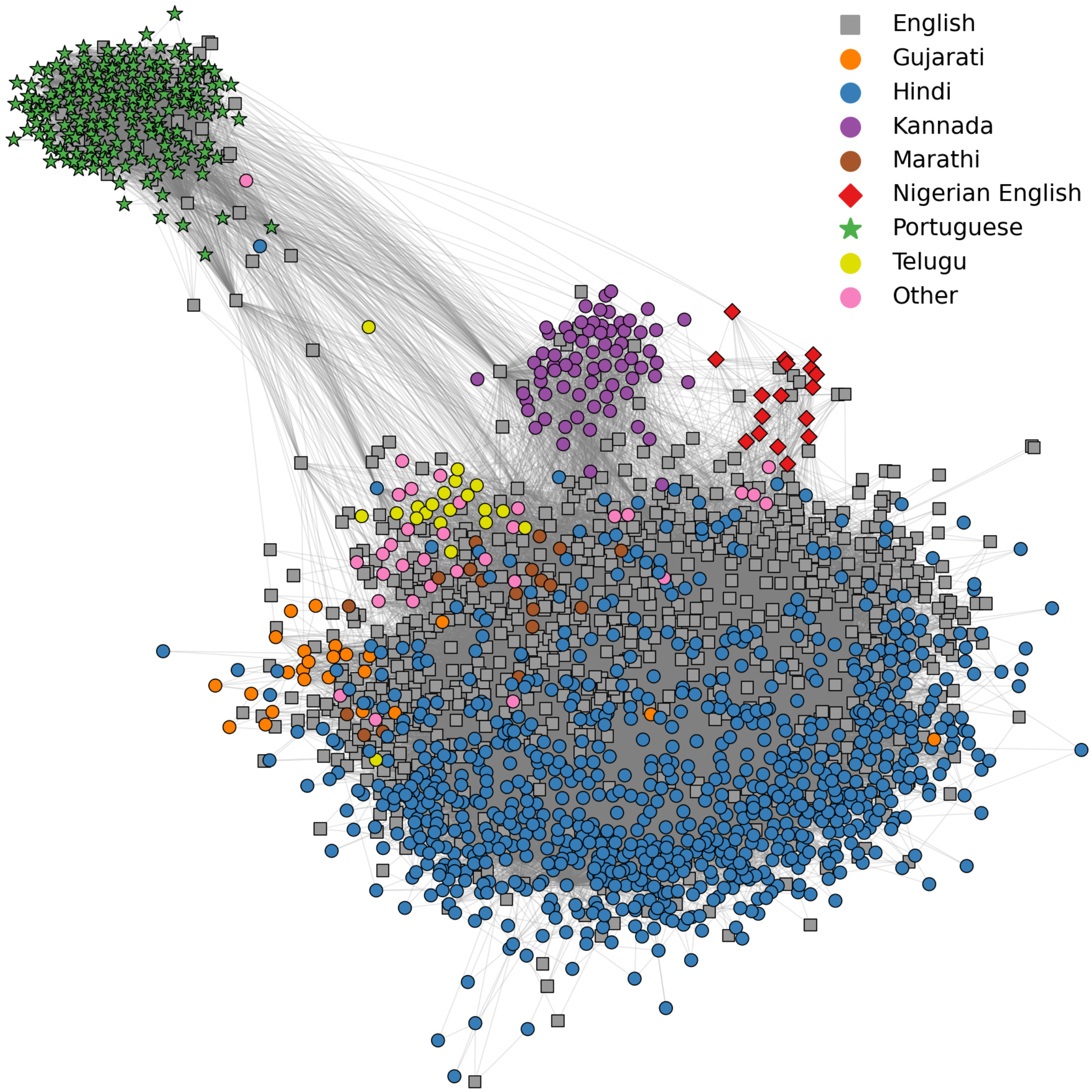


Figure 5.2: **Co-occurrence network of accounts of eminence.** Two eminent users are connected by an edge if at least 50 accounts on Koo interact with both of them. Nodes are coloured according to modal account language. Node shapes differentiate Indian and non-Indian languages.

data released are available publicly and results are aggregated across accounts. To ensure that our data collection pipeline and data sharing complies with the relevant regulations, we have completed a Data Protection Impact Assessment which has been approved by our institution.

Similar to the Voat dataset described in Chapter 3, we ensured that the released dataset aligns with the FAIR principles, to provide the scientific community with a comprehensible and usable dataset.

Data Description

We now present the structure of the dataset, available on a public repository [141]. We release the metadata related to each type of user activity, as well as the user profiles, in separate JSON files. Table 5.3 indicates, for each user activity, the fields included in the JSON schema, including the data type of each field and a short description. For user convenience, we provide language-specific files for each interaction type. In total, we release 34 files for comments, shares and likes, and 43 files for the posts. This difference is due to small linguistic communities with fewer than 5 posts and no other interaction types.

Results

In this section, we provide a quantitative overview of Koo including (1) an analysis of user engagement over time (2) an analysis of Koo's news media ecosystems, and (3) an analysis of user content. These results offer a more in-depth description of Koo's growth within the digital platform ecosystem and its potential to harbour a unique online community.

User retention. Previous studies have shown that, although alternative platforms are successful at attracting a large number of user registrations, many newcomers do not remain active on the platform and become idle within days of registering [89, 243]. Figure 5.3A shows the number of posts, comments, shares and likes, recorded on a daily basis on Koo. Figure 5.3B shows the daily number of active users over time.

After steady growth in 2020, the platform experienced a jump in activity in February 2021, attributed to BJP officials signing up on Koo with their followers. The daily activity count starts plateauing afterwards, but does not experience any substantial decrease. We see a second major increase in activity in November 2022 after Brazilian celebrities endorsed Koo, although activity levels quickly return to levels observed before the Brazilian migration. User engagement and activity does not fall substantially after any of the spikes in user registrations. This result may indicate low retention amongst users who joined Koo during the major collective migrations, a pattern that has been observed in other alt-tech outlets [89, 130].

News ecosystem. Next, we look at news media URLs shared on Koo. Previous studies have highlighted the emergence of online ecosystems where users are only exposed to a limited selection of sources, often due to information-filtering mechanisms [299, 300]. This phenomenon is commonly referred to as an *epistemic bubble* and has been implicated in online radicalisation processes [172] and populist discourses [301]. The question of news

Key	Type	Description
language_code_posts.json (43 files)		
id	string	ID of the post
creatorId	string	ID of the user who created the post
title	string	Content of the post
createdAt	int	UNIX timestamp of the post
handle	string	Handle of the user who created the post
language_code_likes.json (34 files)		
id	string	ID of the liked post
creatorId	string	ID of the user who created the post
createdAt	int	UNIX timestamp of the like
handle	string	Handle of the user who posted
liker_id	string	ID of the user who liked
language_code_comments.json (34 files)		
id	string	ID of the commented post
creatorId	string	ID of the user who created the post
title	string	Content of the comment
createdAt	int	UNIX timestamp of the comment
handle	string	Handle of the user who posted
commenter_id	string	ID of the user who commented
language_code_shares.json (34 files)		
id	string	ID of the shared post
creatorId	string	ID of the user who created the post
createdAt	int	UNIX timestamp of the share
handle	string	Handle of the user who posted
sharer_id	string	ID of the user who shared
koo_users.json (1 file)		
id	string	ID of the user
handle	string	Handle of the user
title	string	Self-given title
description	string	Profile description
createdAt	int	Timestamp of the account creation

Table 5.3: Description of the metadata in each data file.

diversity is of particular interest in the case of an Indian social platform that aimed to host a wide range of cultural backgrounds. Previous studies have underlined the ethical issues that arise when religious events are covered in a controversial way by mainstream outlets that do not promote media secularism [302, 303].

To measure the news ecosystem for each linguistic community, we separate posts into languages for the news media analysis. This approach allows us to map the news ecosystem shared among users who predominantly used the same language on the platform. Comparing the results across languages demonstrates an outlet's ability to seep through language barriers. Figure 5.4 shows the top 20 online domains shared on Koo and their respective prevalence within the major linguistic communities on the platform. We notice a strong tendency for Hindi news outlets to be part of the top domains on Koo, as Hindi speakers

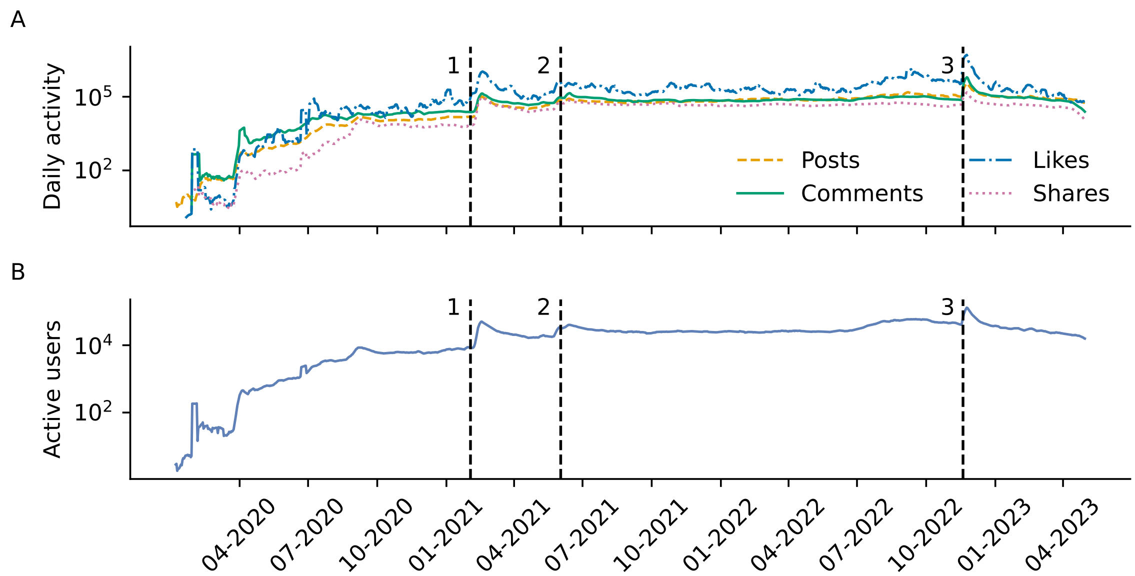


Figure 5.3: **Daily activity and number of active users.** A) 7-day moving window average of the amount of content (posts, comments, likes and shares) posted on Koo. B) 7-day moving average of the number of active users on a given day. A user is considered active if they created a new post or if they commented, shared or liked an existing post. The dashed lines indicate the events that led to the major collective migrations on Koo, namely 1) the Farmers' Protest in India; 2) Twitter getting banned in Nigeria and 3) Elon Musk's purchasing Twitter and the subsequent Brazilian migration.

represented almost half of Koo's user base. However, some websites were also broadly shared within other linguistic communities. This is the case of *ETV Bharat*, an Indian news channel that is available in 11 major Indian languages. Social media platforms, such as YouTube and Facebook, were also widely shared across linguistic communities.

Some news outlets are found to be very popular amongst smaller Indian linguistic communities. This is the case for *Eenadu*, the most prominent daily newspaper in Telugu. Along with *Sakshi*, it reaches more than 70% of the Telugu-speaking audience [304]. For the Tamil-speaking community, the lesser-known news media *OneIndia* was widely shared. *OneIndia* had been very prolific on Koo and had garnered a large following by systematically sharing links to its articles, thereby outranking some of the better known Tamil newspapers such as *Dinakaran*.

Interestingly, when looking at Nigerian English speakers, the most shared outlet is *Peoples Gazette*, an online newspaper launched in 2020 that has led several investigations about cases of corruption in Muhammadu Buhari's government [305]. The second most shared news outlet among Nigerian English speakers (exc. YouTube and Telegram) is *The Punch*, the most widely read newspaper in Nigeria, which is also highly critical of Buhari's politics [306].

These findings suggest that the Nigerian government was also followed by its dissenters when it migrated to Koo in June 2021 and had a less prominent influence on the relevant news ecosystem than their political opponents.

Portuguese speakers widely shared links to social media platforms, with Instagram, YouTube and Koo links being the most shared websites. These are followed by *G1 Globo*, one of the most popular news outlets in Brazil. Similar to the other major newspapers in Brazil, *Grupo Globo*, the conglomerate that owns *G1 Globo*, was highly critical of Jair Bolsonaro's political stances [307]. However, the network has also been accused of delegitimizing Lula and other leaders of the Workers' Party, such as former president Dilma Rousseff [308]. In a similar fashion as for Nigeria, these results suggest that the Brazilian community on Koo shared news media content which is primarily antagonistic to the current political regime. However, the Brazilian migration also included media actors that are supportive of Lula's regime. The second most shared news platfom within the Brazilian community is *Diario do Centro do Mundo*, a left-leaning digital news outlet criticised as "charlatans hired by the Workers' Party" [309].

The dominance of specific news outlets within linguistic communities, particularly in India, suggests that Koo may not have harboured, or been conducive to, a social environment that cultivated media pluralism. Media pluralism has often been defined as a hallmark of a healthy democracy [310], while also being an important tool for tackling misleading news [311]. To measure news diversity, we compute the Gini coefficient for the set of web domains shared by each linguistic community on Koo. A Gini coefficient of 1 indicates that only one news source was being shared within the community, whereas a coefficient of 0 would represent perfect equality across all news sources. Figure 5.5 shows the Gini coefficient for each linguistic community, plotted against its population size. The figure shows that larger linguist communities on Koo tend to have a Gini coefficient close to 1. This further highlights the dominance of individual news media outlets in each of these communities.

Hashtags analysis. Next, we look at the conversations that was hosted on Koo. Previous studies focused on the US alt-tech ecosystem have shown that populist themes are widely discussed in culturally homogeneous communities [312]. To identify the most salient topics users discuss, we perform an analysis of the hashtags shared by the major linguistic communities on Koo. Often used to bring visibility to a specific topic, hashtags have been shown to play an instrumental role on several social platforms, both in political campaigning and to rally citizens to activist movements. These social phenomena have been extensively studied in India [313, 314], Nigeria [315, 316] and Brazil [317, 318].

Table 5.4 shows the 20 most shared hashtags for the Hindi, English, Nigerian English and Portuguese speaking Koo communities. For Hindi and English speakers, we note that the most popular hashtags gravitate around Indian matters (#kooforindia, #india), suggesting that the English-speaking community were largely based in India. Most Hindi and English hashtags relate to national politics and economic matters (#narendramodi, #HindiNews, #Budget2022), which further highlights that Koo had succeeeded in hosting a sustainable

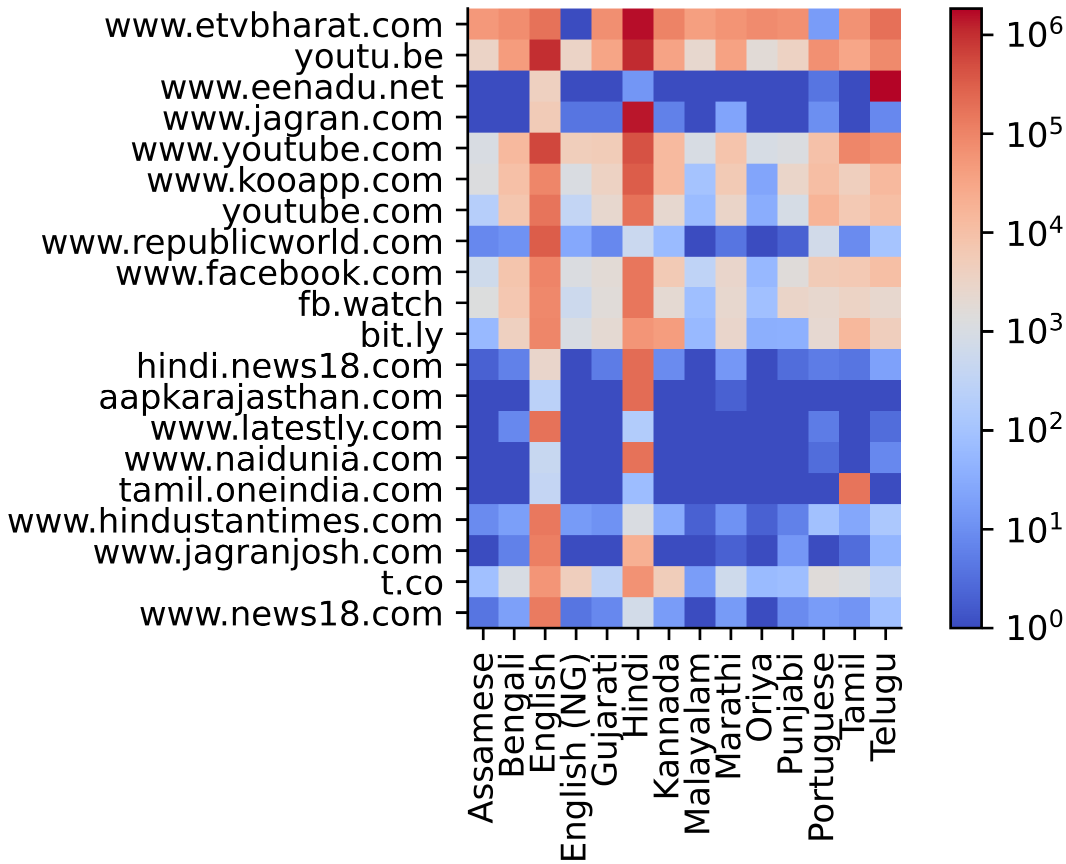


Figure 5.4: **Top-shared web domains and their prevalence in the dominant linguistic communities.** Number of links leading to a web domain shared by the top-10 linguistic communities on Koo. The top-20 shared domains are shown.

conversation within India. However, we also find apolitical hashtags, particularly related to Cricket in India (#Cricket, #iplauction2022). Finally, this analysis highlights interest in Rampal Singh Jatain (#Sat_Bhakti_Sandesh, #SantRampalJiMaharaj), an Indian religious cult leader who has been previously arrested for murder [319]. He signed up on Koo in August 2020 and had garnered more than 330k followers.

The Nigerian English-speaking community prominently used hashtags related to the Twitter ban and the migration to Koo in June 2021 (#letskoo, #bantwitter, #nigeriatwitterban). Furthermore, many hashtags also mention locations in the country, such as cities (#kano, #kaduna, #lagos) or regions (#arewa). These results show that the Twitter ban remained a major topic of debate, and that Nigerian users attempted to rebuild their social network on Koo by connecting with peers located in the same region of the country. National politics were also discussed (#Buhari, #buharitrain, #OsinbajoDay) and highlight the presence of both pro- and anti-Buhari cohorts on the platform. The #OsinbajoDay hashtag refers to

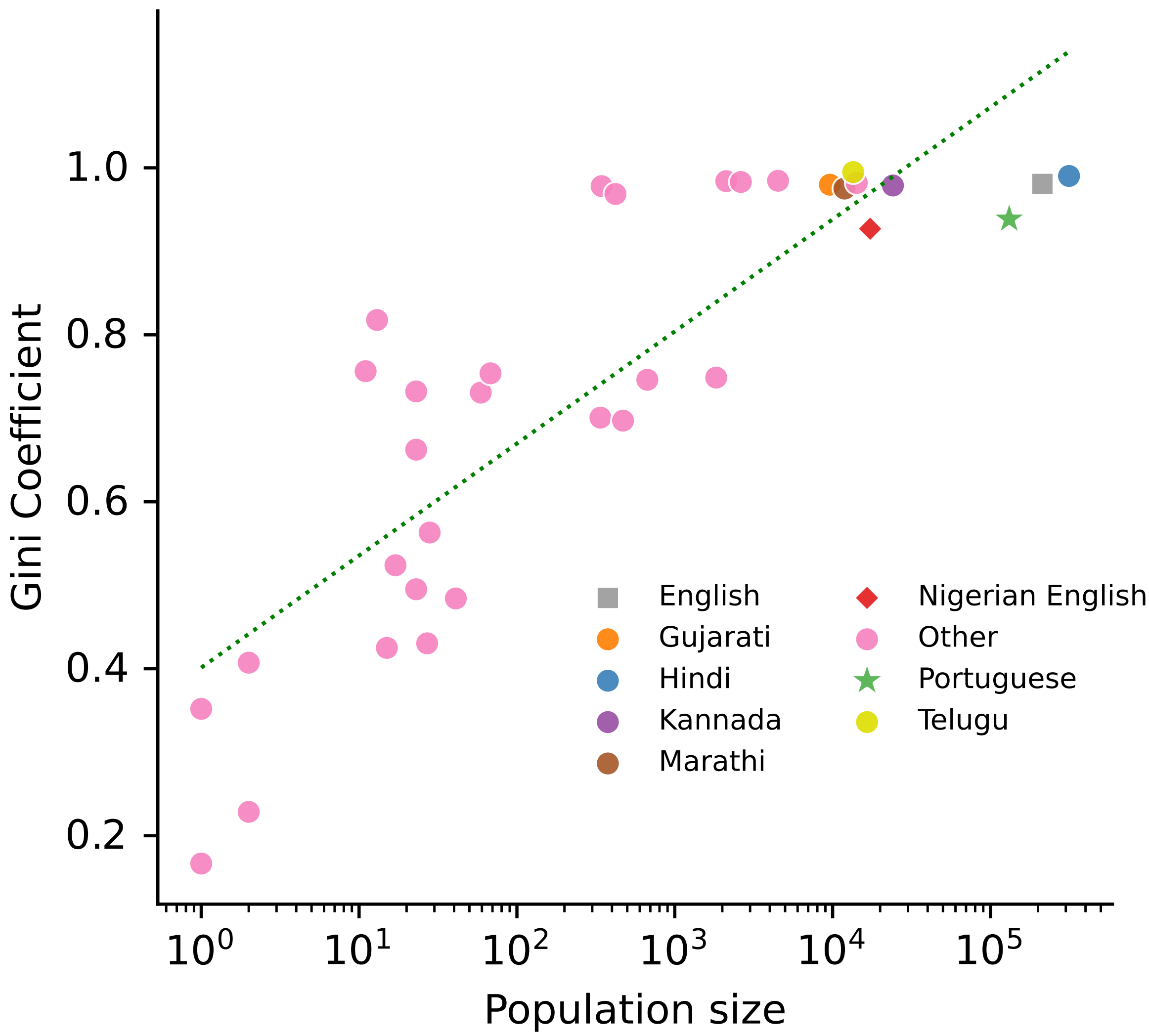


Figure 5.5: **Media plurality across linguistic communities on Koo.** The Gini coefficient for the news media web domains shared by each linguistic community, plotted against the population size of each linguistic community. A Gini coefficient close to 1 highlights a monopoly held by a single news source, whereas a Gini coefficient close to 0 indicates a more diverse link-sharing ecosystem. The dotted green line shows the linear fit for the correlation between the population size and the Gini coefficient. The linear regression has a slope $a = 0.134$, with a 95% confidence interval $[0.104, 0.164]$

then Vice President Yemi Osinbajo, hailed by grassroots campaigners as the right candidate to take over "the good works already started by Muhammadu Buhari" [320]. In contrast, the #buharitrain hashtag is used on social media to criticise president Buhari's railway project between the capital city of Abuja and its airport, with the investment considered a failure

[321].

The hashtags used by the Brazilian community highlight a desire from users to connect with their peers on Koo (#koobrasil and #sdv, Portuguese slang for “follow me back” on social platforms). This is further suggested by the prominence of hashtags related to Elon Musk’s purchase of Twitter (#elonmusktwitter, #layoffs, #riptwitter). The initial migration indeed occupied a significant space in the Brazilian conversation, with other prominent hashtags related to the linguistic pun that triggered the migration to Koo (#brasilnokoo, #tomandonokoo) [26]. Unlike the Indian and Nigerian communities, none of the most prevalent hashtags in the Brazilian community are related to contemporary political topics. Instead, mentions of the 2022 FIFA World Cup (#copadomundo2022, #fifaworldcup, #vaibrasil) and general conversations (#memes, #humor, #art) are more common. These results highlight that the Brazilian migration, unlike the Indian and the Nigerian ones, was not initially politically motivated.

Overall, the hashtags indicate that the main communities on Koo were all involved in conversations relating to the events which triggered their communities’ respective migrations to Koo. This highlights the impact that social media deplatforming can have on online discourses [89]. The stark difference in the prominence of political hashtags between the Brazilian community and the Nigerian / Indian communities underlines that the initial motivations leading to a platform migration can be both political and apolitical, and can heavily impact the dominant discourses that subsequently emerge on alt-tech platforms.

Hindi		**English**		**Nigerian English**		**Portuguese**	
Hashtag	**Ratio (%)**	**Hashtag**	**Ratio (%)**	**Hashtag**	**Ratio (%)**	**Hashtag**	**Ratio (%)**
koooftheday	0.77	koooftheday	0.62	nigeria	5.41	koobrasil	4.46
comedy	0.72	india	0.37	letskoo	3.38	brasilnokoo	3.78
memes	0.71	memes	0.32	Nigeria	2.34	koo	3.77
memeoftheday	0.71	India	0.31	bantwitter	1.42	copadomundo2022	3.61
dailymemes	0.42	koo	0.31	kano	1.32	brasil	2.94
Budgetmemes	0.37	comedy	0.30	kanoconnect	1.14	tomandonokoo	2.56
SaintRampalJi	0.31	kooforindia	0.29	koo	1.04	meme	1.57
HindiNews	0.29	memeoftheday	0.28	kaduna	1.01	koonobrasil	1.42
koo	0.25	narendramodi	0.27	lagos	0.95	riptwitter	1.30
SantRampalJiMaharaj	0.25	motivation	0.25	arewapeeps	0.87	memes	1.13
Sat_Bhakti_Sandesh*	0.24	dailymemes	0.25	nigeriatwitterban	0.82	elonmusktwitter	0.79
india	0.24	trending	0.23	kooyouropinion	0.81	humor	0.78
MpNews	0.20	Budget2022	0.20	june12	0.80	layoffs	0.77
entertainment	0.20	Budgetmemes	0.19	arewakooconnect	0.78	koobr	0.75
Navabharat	0.19	Cricket	0.18	ifollowback	0.67	fifaworldcup	0.75
Nan	0.19	news	0.18	OsinbajoDay	0.67	copadomundo	0.64
kooforindia	0.18	COVID19	0.18	covid19	0.66	vaibrasil	0.56
iplauction2022	0.18	SaintRampalJi	0.17	Buhari	0.64	Brasil	0.54
narendramodi	0.18	SantRampalJiMaharaj	0.17	Arewa	0.64	sdv	0.54
kookiyakya	0.18	life	0.15	buharitrain	0.63	art	0.53

Table 5.4: **Top hashtags used on Koo.** Prevalence of the top 20 hashtags used by the major linguistic communities on Koo. The percentage indicates the percentage of usage of the respective hashtag within the considered linguistic community. The hashtag indicated with the asterisk (*) is translated from Hindi.

Conclusion

In this work, we release and present a Koo dataset totalling over 72M posts, 399M user interactions, and 1.4M user profiles. The whole dataset was collected via Koo's undocumented public API.

Our analysis highlights the growth of a multi-lingual, cross-country online ecosystem, with collective migrations to the platform often spearheaded by influential politicians. Many narratives on the platform related to the events which triggered the initial migrations to Koo, resulting in prominent debates on the national political landscape of India and Nigeria, and to a lesser extent Brazil. For India, Koo content was heavily biased in support of the ruling BJP party. In contrast, both the Nigerian and Brazilian communities on Koo shared pro- and anti-Government content, although the Brazilian Koo discussion was less political than the others.

These collective migrations highlight how the success of an alt-tech fringe platform can be intimately linked to the decisions made by mainstream platforms and to platform- or regulator-driven deplatforming events (e.g., because Twitter banned users, or because regulators banned Twitter) [89]. Catering to these communities, Koo had become a prominent venue for political and social debates for several communities. However, despite a lack of news media and content diversity, and unlike most alt-tech fringe platforms, Koo had had success in attracting a politically heterogeneous user-base beyond India in both Nigeria and Brazil. This sets Koo apart from other emerging platforms based outside the US - especially those in China and Russia where politicians have emphasised the need for "Internet sovereignty" [322] - in that their ambition is to expand beyond their home jurisdiction. Consequently, a platform like Koo has the potential to challenge the current dominance of US-based social media platforms, a change which may have important consequences for social media regulation.

We anticipate that this dataset will support further research investigating digital ecosystems based outside the US, and how they impact political campaigning and news sharing. This is particularly important given the role Koo played in the 2024 Indian elections, which took place from April 19 to June 1 [323]. A study reports that the platform's moderation team did not react to openly hostile content, even after it was reported by the researchers [324]. With the hateful content reported targeting mostly female politicians and the Muslim community, this is further evidence that Koo allowed divisive and hateful content to proliferate on their service.

Finally, given Koo's linguistic diversity, we hope that our dataset will motivate researchers to further develop computational tools for studying social media narratives and news media sharing in underrepresented languages. This work will be crucial to enable a robust analysis of the online news media ecosystem, misinformation, and its potential impact on political campaigns across diverse markets.

Chapter 6

How Language, Culture, and Geography shape Online Dialogue: Insights from Koo

Introduction

The social media ecosystem, which has historically been Western-focused [325], has evolved significantly in recent years, with a rapidly growing number of the active users from non-Western countries and the Global South [326]. Despite this shift in demographics, major social media platforms such as Twitter (now X) and Facebook lack adequate support for many major vernacular languages [327] (e.g., in South-East Asia [328]), with platforms continuing to prioritise Western audiences. For instance, investments into English-language content moderation on X, Facebook and Instagram still heavily outstrip investments into moderation tools for other language's [329], indicating that platforms are less equipped to tackle, and are (arguably) less concerned about, harmful content posted in languages other than English. This is despite documented evidence that social media can play a key role in election campaigns, and the fact that India, the World's largest democracy which went through a major political election in June 2024, has become the largest market in terms of users for many leading social platforms including Instagram [330], Youtube [331] and Facebook [332].

This persistent failure to support many communities beyond the West was the *raison d'etre* of Koo, the Indian-based social platform discussed in the previous chapter. Koo aimed to champion a "language-first" approach, where each user is able to express themselves in their native language when connecting with their peers [333]. By upscaling their auto-translate tool, Koo aimed to offer an inclusive experience to speakers of less widely utilised languages, a feature not prioritised on the dominant US-based platforms. Indicative of this, Koo used to support 20 of India's 22 official languages, whereas X (previously Twitter) only supports 5 [334]. By appealing to political leaders across countries, mostly from India, Nigeria and Brazil, including some who have criticised US-based social media platforms, Koo had managed to attract a geographically diverse user-base, becoming the second largest microblogging platform globally after X [282, 335] before its shut down in July 2024 [194]. As such, it occupied an influential position in the social media ecosystem and offers a unique opportunity to study the role of language on social media.

Although often defined as an alt-tech platform, Koo had attracted a more international user base than US-based alt-tech platforms [130, 241, 336] leading to a more diverse community. Here, we extend this work by focusing explicitly on the role that language plays in shaping the structure of a non-Western social platform.

Previous social media studies have considered the role of language in shaping online interactions, but not in the context of India. Researchers have studied linguistic trends on Twitter (now X) and found that English-speaking posts were dominant on the platform when it mostly attracted users from Western nations [337, 338], whereas the usage of English became less prevalent when considering non-Western countries [339, 340]. Moreover, following the growth of communities outside English-speaking nations, different linguistic communities were shown to interact differently with a social platform's features, leading to distinct social structures [341]. However, despite the subsequent globalisation of social media, the formation of dyadic ties on Twitter was found to be strongly correlated with the linguistic background of a user, even between different English-speaking countries [298, 342]. Studies have also considered the influence of bilingual social media users on interaction networks, and whether users post in languages other than a platforms dominant language [343, 344]. Language diversity was further quantified by using geolocation data to map the language diversity in the Greater Manchester [345]. More recently, studies have compared literacy levels across regions on Facebook [346]. Our research contributes to this literature by studying an online platform striving to host vernacular languages, in the hope to harbour a sustainable multi-lingual community.

With 22 nationally-recognised languages, and over 100 languages with more than 10,000 native speakers, India is a unique case study to assess the impact of linguistic pluralism on user-to-user interactions online. India's linguistic history has been the focus of several studies, looking into the national linguistic landscape [347], the patterns of communication [348] and the socio-economic ramifications of a complex linguistic environment [349], but many of these methods have not been applied in the context of social media. Moreover, the rich linguistic composition of India also allows for an analysis of language use at a national scale, before looking at linguistic communities across several countries. Similar observations are valid for Nigeria, one of the other countries where Koo was adopted by government officials, where multilingualism plays a major role in social interactions in several areas [350], with about 500 vernacular languages spoken across the country [351].

In the remainder of this chapter we first assess the impact of the various collective migrations to Koo which shaped the platform into a multilingual venue. We then study how political incentives to migrate to Koo led to different degrees of user engagement over time. Afterwards, we look at the topology of the interaction network, while assessing the impact of the language barrier to foster cross-cultural communities. Finally, we look at user mobility across linguistic landscapes and how this relates to the richness of a community's online conversation, as well as the shared discourse between language pairs. Our findings suggest that linguistic and cultural factors were instrumental in bridging communities on Koo, with few interactions taking place across communities with different linguistic backgrounds.

Results

Platform migration and user retention

We begin by examining Koo's popularity over time in the online ecosystem. Figure 6.1 shows the daily number of registrations on Koo, from the launch of the platform in 2020 until early 2023. Major political and social events, which had an impact on Koo's outreach, are marked as dashed lines.

The first significant peak in registrations can be seen in February 2021. India was in the midst of the farmers' protest, a popular movement against a new set of laws adopted by the Parliament of India in September 2020. These events triggered a conflict between the Indian government and Twitter (now X) after BJP (the ruling political party) officials pressured Twitter to ban accounts linked to the popular movement [352]. Members of the government and BJP supporters in India subsequently signed up to Koo, as Twitter did not comply with their requests, and invited their community to follow suit [353]. As seen on the figure, the political movement managed to increase Koo's user base substantially [190]. The platform's willingness to comply with content take down orders issued by the government made Koo more attractive to BJP politicians, thereby cementing the dominance of BJP narratives on the platform [353].

The second burst in registrations dates from June 2021, when Nigerian then-President Muhammadu Buhari banned Twitter from the country and registered on Koo with members of his government [354], after Buhari's tweets were deleted for inciting violence against his political opponents [355]. Koo experienced a wide adoption from government officials in Nigeria, prompting the platform to hire vernacular speakers for content moderation purposes in Nigeria [25]. The government was also followed by a large number of Nigerian users who subsequently signed up to Koo, leading to an uptick in registrations. However, Koo had little success in attracting Nigerian celebrities or influencers, unlike in India where the platforms gained support from Bollywood actors and prominent cricket players [290]. Previous research has shown that celebrities' endorsement can catalyse a massive migration towards alt-tech platforms [89].

The last major peak in registrations took place in November 2022, shortly after Twitter was purchased by Elon Musk. Felipe Neto, a Brazilian influencer with over 16 million

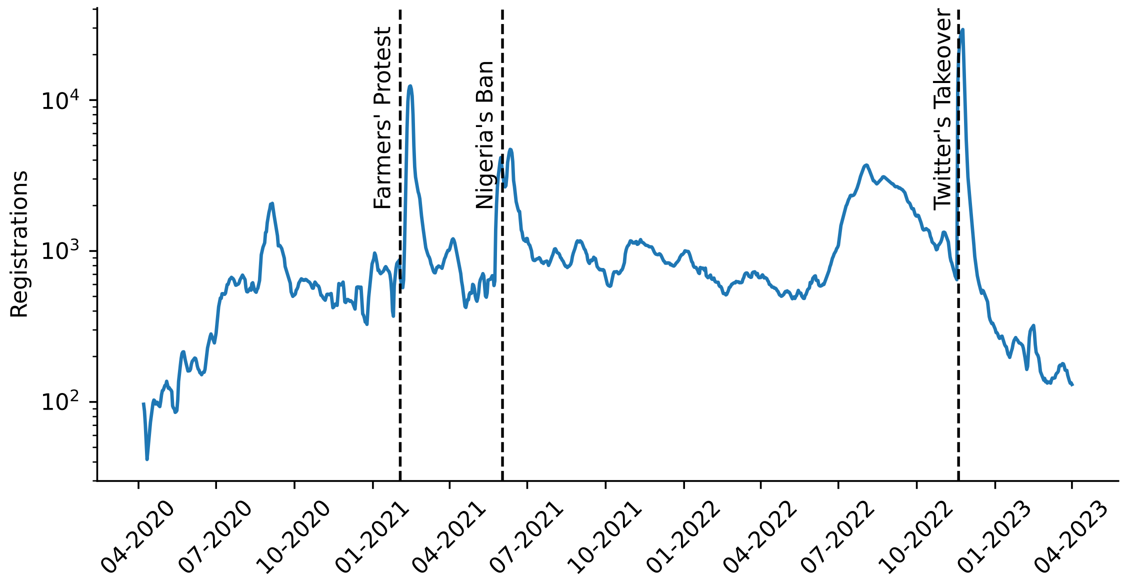


Figure 6.1: **Daily number of registrations on Koo, and the impact of collective migration.** 7-day moving average of the daily number of registrations on Koo, from the beginning of 2020 to early 2023. The dashed lines indicate, in order: the migration of BJP politicians and their supporters following the Indian Farmers' Protest in February 2021; the migration of the Nigerian government after Twitter was banned in the country in June 2021; the Brazilian community joining Koo in November 2022 after Elon Musk purchased X.

followers on Twitter, advertised his migration to Koo on Twitter (now X), which led to his followers signing up to the platform as well [26]. This collective movement was strengthened when Brazilian President Lula also registered on Koo [296]. In total, Koo's user base grew substantially in Brazil, with the Koo app downloaded over 1 million times in the space of 48 hours [356].

These three events have shaped the major linguistic communities on Koo. As shown in the previous chapter, each collective migration had temporarily boosted the daily activity on Koo, allowing the platform to become a serious competitor in the social media marketplace. A breakdown of the registration numbers broken down by language is also provided in the Appendix, showcasing the influx of Hindi, Nigerian English, and Portuguese users.

To assess the success of linguistic migrations, we measure user retention, i.e. how many users within a cohort are still active after a given number of days. Throughout our analysis, we match each user with the language they used the most when posting and commenting on the platform.

Figure 6.2 shows the Kaplan-Meier estimator, a tool used to visualise the retention curve of a population over time, computed for each linguistic community on Koo. Given a linguistic community, the Kaplan-Meier estimator indicates how many users were still active on Koo, a given number of days after they registered on the platform. The figure indicates that both Brazilian and Nigerian communities have a much lower retention than other linguistic communities on Koo, with 50% of the cohort becoming inactive within 16 days and 23 days of signing up to the platform, respectively. In contrast, it took 131 days to reach the same level of user retention when considering Hindi-speaking users, thus highlighting a strong difference in user engagement across linguistic communities and countries. The smaller linguistic clusters also display a higher survival rate than the Brazilian and Nigerian users, which suggest that the sustainability of a community does not only depend on its population size.

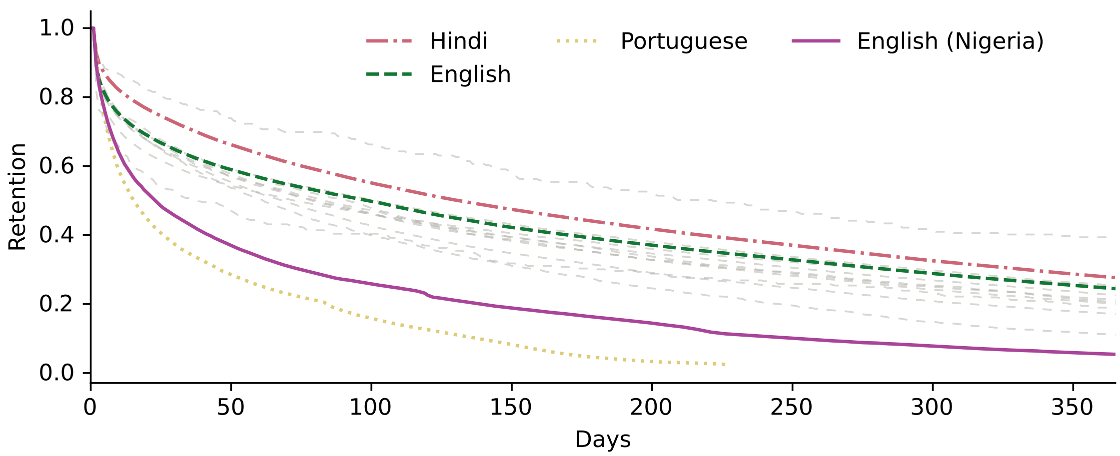


Figure 6.2: **Heterogeneous user retention for various linguistic communities.** Kaplan-Meier survival curves for the main linguistic communities on Koo, showing the fraction of users who remained active after a given number of days. For each user, we define "day zero" as being their registration date on Koo. Other linguistic communities are displayed in grey. The retention curve is displayed until the day that fewer than 1% of users from a linguistic community remain active.

The lack of long-term adoption in the Nigerian community can be explained by a popular resistance to the Twitter ban that was instated by Buhari's government in June 2021. Nigerian users managed to bypass the ban shortly after it was instated, with VPN usage becoming more common nation-wide [354]. Moreover, Koo did not receive sustained support from the Nigerian government. Muhammadu Buhari lifted the ban on Twitter in January 2022, after the platform and his government settled on an agreement [294]. Buhari stopped being active on Koo shortly afterwards, as did most of the government members who joined the platform. Koo's monthly active users in Nigeria fell by over 90%, suggesting that the platform failed to establish a foothold as sustainable as their popularity in India [295].

In the case of Brazil, the migration was not, primarily, triggered by political motivations in the same way as for India and Nigeria, but rather by a linguistic pun involving the word "koo" in Portuguese, although Brazilian President Lula did join the platform during its initial growth-phase. However, in general, the Brazilian community and Brazilian celebrities did not stay as engaged on Koo, when compared to celebrities from India [357]. Both Felipe Neto and Lula, who were the main drivers of the Brazilian migration on Koo, stayed active on Twitter and their followers can therefore still follow their feed without requiring access to Koo. Previous research has shown that users are less active on alt-tech platforms if they can still reach out to their followers via a mainstream outlet [89]. As of November 2023, Lula is still sporadically active on Koo, whereas Felipe Neto's last interaction dates to May 2023.

These results suggest that collective migrations to an alternative social platform can have mixed levels of success, depending on the motivations triggering the migration and the degree of approval it garners across the community. The Indian migration exemplifies the birth of a sustainable community on Koo, as it was led by government officials and garnered support from both national celebrities and BJP supporters. The Nigerian government followed a different pattern, where the low support for the X ban outside of Buhari's supporters led to a short-lived retention for most Nigerian users who signed up on Koo. In the same fashion, the Brazilian community shows a low level of user retention, which can be explained by a lack of social incentives to shift the political discourse to an alternative platform, and away from the dominant US-based platforms.

Language-use in the Koo interaction network

We now focus on user interactions on Koo and the landscape that emerges on a platform where many languages and cultures coexist, following individual migration decisions.

Figure 6.3A shows an interaction network, where two users are connected if they interacted (i.e. one of the users liked, shared or commented the other's post) on Koo, with the weight of an edge proportional to the number of interactions between a pair of users; for simplicity we treat the interaction network as undirected. The network layout, generated with a force-directed graph drawing method, highlights the strong segregation between several linguistic communities: Portuguese-speaking users (yellow) mostly interact with other Portuguese-speaking users, and likewise, Hindi-speaking users (blue) mostly interact with other Hindi-speaking users. On the other hand, the English-speaking cohort (green) acts as a bridge between Hindi, Portuguese and Nigerian English speakers (purple), as well as the smaller cohorts that can be seen in the periphery of the graph.

The network indicates a strong homophilic behaviour, meaning that users will mainly engage with members of their own linguistic community. Koo's auto-translate feature, which allows users to translate any post to the language of their choice, does not seem to mitigate the language assortativity that we observe on the platform. Homophilic patterns have been observed in many online social spaces, for example when considering information diffusion dynamics [358], community formation [359], and in political interaction networks [298].

We can highlight this homophilic behaviour by computing the k-cores of the network.

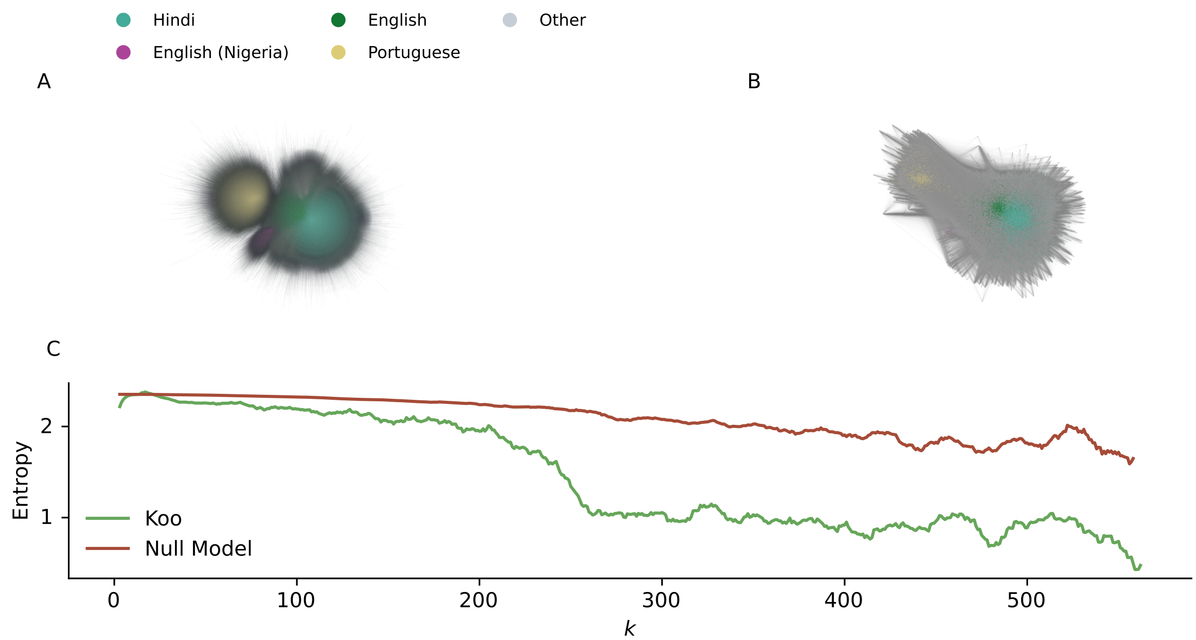


Figure 6.3: **The Koo interaction network and the impact of linguistic homophily on the network's structure.** Each node represents a user, and two nodes are connected if one of the users interacted with the other user's content. Users are coloured according to their modal language on the platform. The main linguistic communities are the Hindi-speaking users (blue), English-speaking users (green), Nigerian users (purple) and Portuguese-speaking users (yellow). The layout is generated by using a force-directed graph drawing method. A) The total interaction network. B) The k-core of the interaction network with $k = 150$. C) The Shannon entropy of the modal language of the nodes belonging to the k-core of the graph, with respect to the value of k. The entropy of the interaction network is compared to the value obtained in a null model, where we shuffle the modal language associated to each node in the network.

Given a network, its k-core for $k \in \mathbb{N}$ is defined as the sub-graph in which all vertices have a degree greater than or equal to k. k-core analysis offers a deeper view into the tightly connected components of a network, therefore identifying the nodes that are strongly interconnected or play an influential role in the overall network structure [360]. Figure 6.3B shows the interaction network from figure 6.3A but filtered so that only the $k \geq 150$ core is shown. This threshold allows us to filter out the periphery of the network, by only keeping the 5% most strongly connected nodes of the network.

This visualisation offers a better overview of the tightly interconnected components within the interaction network, with a strong core of English-speaking and Hindi-speaking users. Conversely, the Brazilian and Nigerian clusters are more isolated. The k-core also highlights that users belonging to smaller linguistic communities are scarcely included in the core of the network, with 95% of the core users belonging to the Hindi, English, Nigerian English and Portuguese-speaking clusters. This analysis further underlines that highly connected clusters mostly involve users who speak one of the dominant languages on the platform. Looking at the English-speaking community, we note that it is principally connected to Hindi speakers in the k-core, highlighting the instrumental role that the English language plays in Indian political communications [361]. On the other hand, both the Nigerian and the Brazilian communities dominantly communicate in their native language and are therefore isolated in the k-core.

To quantify the connection between linguistic communities, we generate the k-core for all values of k for which the k-core exists. We retrieve the modal language of each vertex included in the k-core and compute the Shannon entropy, to evaluate whether the k-core encompasses a diverse range of languages or is primarily dominated by a few major languages [362, 363]. Once we have computed the k-core, we retrieve the modal language of each user included in the core and calculate the entropy of the list of languages. Figure 6.3C displays the resulting entropy with respect to k. We notice that higher values of k lead to a lower entropy in the language composition of the k-core, suggesting that dense interactions on Koo take place mostly within homogeneous linguistic clusters.

To ensure that the lack of diversity in high-degree interactions is a characteristic of the interaction network on Koo, and not an erroneous finding, we define a null-model of the network, where the modal language is shuffled for the nodes in the network (preserving the language prevalence distribution), and the Shannon entropy is computed again for each value of k. Figure 6.3C displays the median Shannon entropy for each value of k, after running the null-model 1000 times. We notice that the entropy for the null-model does not sharply decrease for higher values of k, which reveals that this sharp decline in language diversity within the k-core is indeed a distinctive feature of the Koo interaction network. These findings further indicate that cross-linguistic interactions on Koo are rare with respect to same-language interactions.

To measure the prevalence of a linguistic community within the k-core for any value of k, vertices in the network can also be defined by their coreness, i.e., the maximum value of k for which they still belong to the k-core. In the Appendix, we show the distribution of the coreness of the users belonging to each linguistic community. Our analysis reveals that only Hindi- and English-speaking users are included in the highest cores. This result indicates that, despite the presence of a rich linguistic landscape on the platform, strong interaction ties on Koo are mainly driven by linguistic homophily, and that cross-linguistic interactions are rare.

Language homophily and multilingual activity

The structure of the interaction network indicates that the linguistic background of a user strongly influenced their interaction patterns on Koo. Previous studies have highlighted that many ties in social networks are strongly assortative, i.e., they connect individuals who share similar attributes in terms of cultural background or social status [364]. Language assortativity has been previously found to influence mating decisions [365], friendship ties among adolescents [366], and political communication between countries using a common language [298].

When considering a diverse population such as the Koo user base, a question to consider is the interaction patterns for members of minority linguistic groups. Previous studies have shown that minority groups in organisations rely on out-group interactions to be connected to the centre of the network [367], whereas belonging to an underrepresented social group leads to a stronger in-group identity in friendship networks [368]. As such, we will next look at linguistic behaviours on Koo, by measuring the propensity for users to interact within their linguistic community on the platform. We will also evaluate a user's likelihood of using their modal language when interacting with their peers on Koo.

To measure a user's adherence to their modal language relative to their propensity to use other languages, we define the commitment: given a user with N posts in their modal languages and M posts in other languages, the commitment is given by

$$C = \frac{N}{N+M}. \tag{6.1}$$

Commitment has previously been used in linguistic studies to assess the adoption of new linguistic norms, indicating that outdated norms are still persistently used by a minority of the population [369]. Similar findings have been highlighted in ethnographic studies looking at the adoption of new spelling rules in both Spanish and English-speaking nations [370, 371]. The literature therefore suggests the potential for smaller-scale communities to survive in a linguistic setup, despite the rise of a dominant linguistic framework.

The average commitment of a linguistic community, plotted against its population size, is displayed on Figure 6.4A. We notice a strong tendency for the average commitment of a user to their modal language to increase as the population size of their community increases. Among the highlighted communities, Nigerian English speakers display the highest average commitment ($C = 0.90$), and English-speaking users also display a high level of commitment ($C = 0.84$), despite their role as a bridging community between other languages. Portuguese and Hindi speakers also have a high commitment ($C = 0.87$ and $C = 0.88$, respectively), which indicates that their communication on Koo mostly relied on their modal language, emphasising the absence of cross-language interactions. However, smaller linguistic communities display a lower commitment to their modal language. For example, Indian communities, such as Odia speakers ($C = 0.73$), and non-Indian clusters such as Spanish speakers ($C = 0.74$), are less committed to their modal language. This finding highlights the need to communicate in other languages in order to be part of a community on the platform. The attractiveness of a language has previously been modelled by its number of speakers [372] suggesting that smaller linguistic communities are less likely to attract new speakers, thus fuelling their need to use other languages in order to be connected to core conversations on the platform.

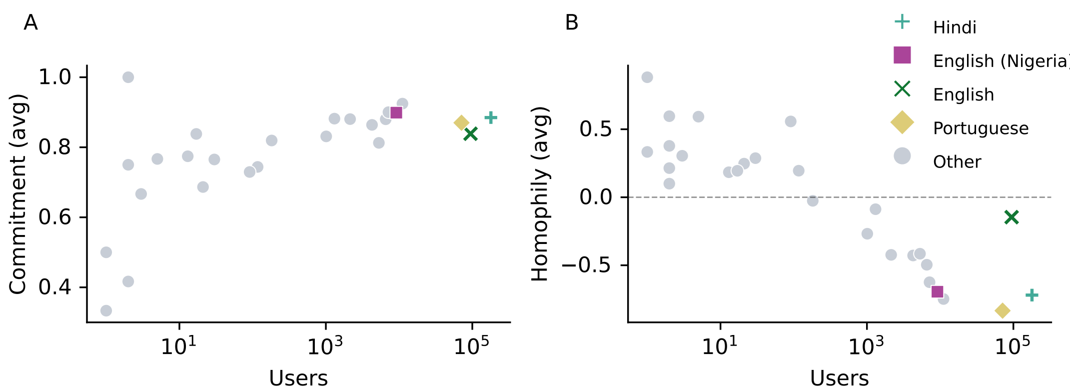


Figure 6.4: **EI-Homophily and language commitment and the impact of a community's size on its sustainability.** Number of users belonging to a linguistic community plotted against A) their commitment to their modal language, and B) their EI homophily index. Both metrics are averaged by the number of users for whom the language measured is their modal language. The coloured dots represent the Hindi-speaking community (blue), English (green), Portuguese (yellow) and Nigerian English (purple). The dashed line indicates an average homophily equal to 0.

The high level of commitment observed for the major linguistic communities, along with the topology of the interaction network, shows that there is a strong trend for interactions on the platform to involve two users with a similar linguistic background. To measure whether social interactions on Koo mostly took place within the confines of homogeneous linguistic clusters, we use the External-Internal (EI) homophily index [373]: given a node in the interaction network with E edges to their out-group (in this case, interactions with another linguistic community) and I edges with their in-group (members of the same linguistic community), their EI homophily index is given by

$$EI = \frac{E - I}{E + I}. \tag{6.2}$$

A node which only interacts within their in-group (same language) therefore has $EI = -1$, whereas a node which only interacts with their out-group (different languages) has $EI = 1$. The EI-homophily index has previously been used to measure people's tendency to interact with their politically-aligned peers on social media [374] and those sharing a similar vaccination status [375], with both studies showing an overall trend for people to cluster in homogeneous groups. A social network where the majority of interactions are intra-group links (i.e. with the EI-homophily index close to -1) is referred to as being *homophilic*, whereas it is referred to as *heterophilic* if the network displays several inter-group interactions (i.e. with an EI-homophily index close to 1).

The EI-homophily, averaged for each linguistic community by considering the modal language of each user, is shown in Figure 6.4B, and plotted against the size of each language's population. We notice that small communities have a positive homophily index, indicating that members of small linguistic clusters interact mostly with other linguistic communities. On the other hand, with the exception of the English-speaking community, larger linguistic communities are more involved in in-group interactions, leading to a negative EI-homophily index. For example, Portuguese-speaking users have an average EI-homophily index of -0.94, whereas it is equal to -0.66 for the Hindi-speaking users. This finding highlights the existence of siloed communities on Koo, where users' interactions were strongly influenced by language similarities. The strong assortativity with respect to language is further displayed in the layout of the interaction network in Figure 6.3A, where we see how disjointed the Portuguese-speaking and Hindi-speaking communities are with respect to other linguistic communities.

Looking at the average EI-homophily index for English-speaking communities in Figure 6.4B, we notice a stark difference between Nigerian English speakers ($EI = -0.64$) and other English speakers ($EI = -0.15$). This can also been observed in the interaction network in Figure 6.3A, where Nigerian English-speaking users are disconnected from the core of the network, whereas English speakers are strongly involved in cross-language interactions, leading to a higher average EI-homophily index than for Hindi and Portuguese speakers. Thus, the English language acted as a lingua franca on Koo, allowing users from diverse linguistic backgrounds to engage in inter-community interactions. However, the contrast in homophily between Nigerian English and English speakers also highlights that language was not the only factor playing a central role in shaping the interaction network. Cultural similarities are also influential in bridging users together on a social platform. Communities can be structured around specific salient topics of conversation related to the cultural and political landscape of a country, another feature of a platform hosting diverse demographics that we investigate below. Moreover, it is likely that most of the English speakers on Koo (excluding those who speak Nigerian English) had an Indian focus, due to the ubiquity of the English language in India's public affairs [376].

These findings can also be confirmed by looking at the distribution of the EI-Homophily index for each linguistic community, displayed in Figure 6.5. In line with our previous observations, we notice that a majority of users from the Nigerian, Brazilian and Hindi-speaking communities have an EI-Homophily index of -1, and were therefore only interacting with members of their linguistic community. Meanwhile, English-speaking users have a more balanced distributions, whereas smaller linguistic communities are more likely to aggregate at positive values for the index, therefore indicating their propensity to interact outside of their linguistic cluster.

The homophilic patterns observed in the interaction network suggest that communication across linguistic communities was rare on Koo. However, our analysis does not take into

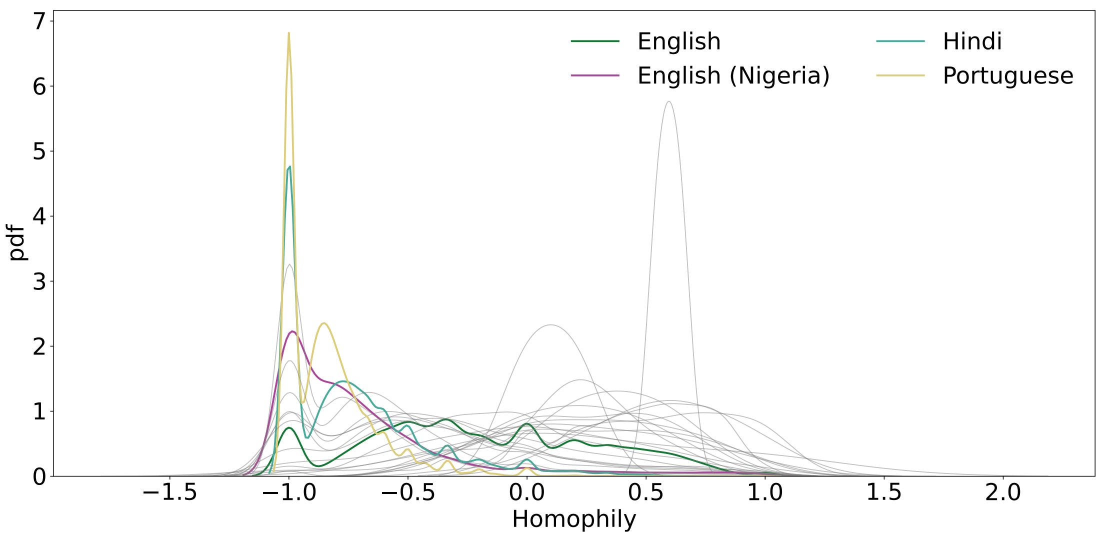


Figure 6.5: Distribution of the EI-Homophily index, for each linguistic community on Koo. The major linguistic communities are highlighted in colours, whereas all the other communities are plotted in gray.

account users who communicated on the platform in more than one language, and therefore belonged to more than one linguistic community on the platform. To measure the propensity for users to switch between languages, we map languages as the nodes of a network, with weighted edges representing the number of users who posted in both languages on Koo. This layout follows the principle of a global language network (GLN), which allows us to quantify indirect communications between pairs of languages by looking at the number of speakers they share [377]. By considering the number of modal speakers in two linguistic communities and the number of speakers they share, we use the phi coefficient to measure the association between the two languages. For two languages i and j, their phi coefficient Φ_{ij} [377] is given by:

$$\Phi_{ij} = \frac{M_{ij}N - M_i M_j}{\sqrt{(M_i M_j (N - M_i)(N - M_j))}}, \tag{6.3}$$

with N being the total number of users, M_i and M_j being the number of speakers of language i and j respectively, and M_{ij} being the number of bilingual speakers for languages i and j. A positive phi coefficient indicates that the number of bilingual speakers between languages i and j is higher than what could be expected based on their representation in our dataset, whereas a negative value indicates that the co-occurrence of both languages is underrepresented relative to the size of the communities. This metric therefore allows us to assess whether there is a stronger connection between two linguistic communities on Koo, than is expected due to chance alone.

To ensure that the link between two linguistic communities is significant, we use the t

statistic, defined as:

$$t_{ij} = \frac{\Phi_{ij}\sqrt{D-2}}{\sqrt{1-\Phi_{ij}^2}}, \tag{6.4}$$

where the degree of freedom D is defined as $D = \min(M_i, M_j)$. As all the linguistic communities we consider in our analysis have at least 20 modal speakers, we set $D = 20$. By setting $p = 0.05$, we can reject the null hypothesis, i.e., the number of links between two languages in the global language network is not statistically significant, if $t_{ij} \geq 1.72$ (one-tailed t-test). Any significant link in the network indicates that there are significantly more bilingual speakers between two languages than expected by chance.

Figure 6.6A displays the value of the phi coefficient between the main languages used on Koo. We notice that there is a positive association between the Indian languages used on Koo. Despite being strongly homophilic, the Hindi-speaking community has a positive correlation with several Indian languages, such as Gujarati and Marathi. Smaller linguistic communities in India also show a strong symbiosis in the network: languages such as Telugu, Kannada, Tamil and Assamese are strongly connected to one another, indicating that speakers of less prominent Indian languages on the platform were more likely to also communicate on Koo using another national language. Our findings are aligned with the results of the 2011 Indian language census, indicating a large number of bilingual and trilingual speakers among the smaller linguistic communities in the country [378]. On the other hand, both the Nigerian English and Portuguese-speaking communities have a negative correlation with respect to every other linguistic community, English excepted.

Figure 6.6B shows the results of the t-statistic, with significant links in the global language network highlighted in blue. We notice that both the Brazilian and Nigerian communities share non-significant links with the smaller linguistic communities on Koo, indicating that there is not a meaningful number of users who used both Portuguese or Nigerian English and another language on Koo - with the exception of English, which is significantly correlated to all languages. These results further suggest that the Brazilian and Nigerian communities were strongly isolated when it comes to language mobility, whereas a shared cultural background enabled Indian language speakers to navigate through different linguistic communities. While enhancing the access to posts written in a user's non-native language, the auto-translate feature on Koo did not appear to incentivise users to interact in languages outside their home country.

Discourse richness and similarity across languages

What about content? A natural question is whether stronger ties between two linguistic communities also implies that their respective discourses were similar. Moreover, some of the linguistic communities being larger than others, we hypothesise that a larger community should have a richer discourse. Studies have shown that the size of a community defines its propensity to sway the discussion topics in another community [379], and that nurturing a local discourse can allow a local community to claim their own governance [380].

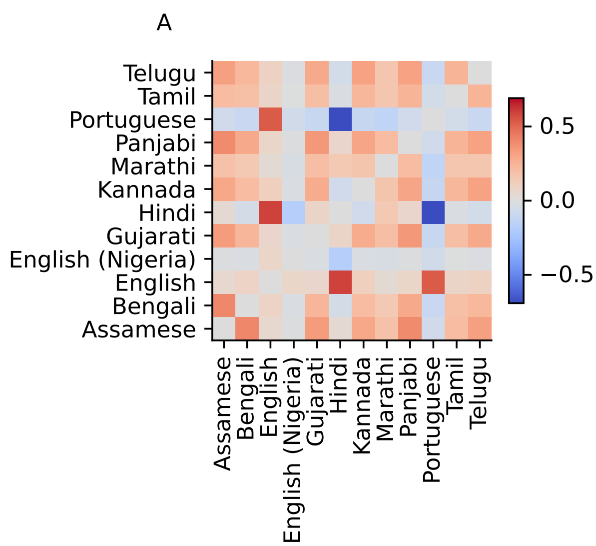


Figure 6.6: **Global language network and multilingual activity.** A) The correlation measured from the global language network. Two languages with a positive correlation share more connections than expected based on their respective number of speakers, and is negative otherwise. B) The t-statistic for each pair of languages in the global language network. Blue cells indicate that the link between the two languages is significant with respect to the t-statistic, whereas red cells highlight non-significant links. A link is considered significant if $p < 0.05$.

To answer these questions, we use diversity measures defined in ecology, which were defined to assess the richness of an environment by looking at the presence of various species and their respective prevalence [381], as well as how often these species can be found across different environments [382]. These methods were also previously used in linguistics research, for example to measure linguistic diversity between Canadian cities [383]. For our analysis, we consider hashtags used within a linguistic community as a proxy for the discourse. Hashtags have been shown to occupy a different linguistic function than words, sharing similarities across languages [384], thus allowing us to capture narratives shared by various linguistic communities on Koo. Hashtags are also a more reliable signal to measure the overlap of narratives across linguistic communities than plain text, as hashtags can be identified without the need to compare textual data from different languages. Our approach is further motivated by the presence of several low-resource languages - i.e. languages for which the overall scientific community lacks a large enough corpus to train NLP models - in our corpus, which are known to be under-represented in many large language models and can therefore lead to unreliable text classification [385].

To assess the richness of the discourse within a linguistic community, we use the alpha diversity, a measure used in ecology to quantify the diversity (or richness) of an environment, by looking at the species that are included in the observed ecosystem. In our case, we compute the alpha diversity of hashtags used by a linguistic community using the Chao1 estimator. We can also estimate the propensity for two linguistic communities to discuss similar topics by computing the beta diversity, another measure that assesses the number of species that are found in two different environments. Using the hashtags, we measure the beta diversity of the discourse between two linguistic communities with the Bray-Curtis dissimilarity index.

To measure the richness of the discourse within a linguistic community, we compute the alpha diversity with the Chao1 estimator $\hat{S}_{Chao1}$, which aims at providing a lower-bound estimation of the number of unseen species, in order to assess the total number of unique hashtags used by a linguistic community from our observations. The Chao1 estimator is defined as:

$$\hat{S}_{Chao1} = \begin{cases} S + \frac{N-1}{N}\frac{s_1^2}{2s_2} & s_2 > 0 \\ S + \frac{N-1}{N}\frac{s_1(s_1-1)}{2} & s_2 = 0 \end{cases}, \tag{6.5}$$

where S is the total number of hashtags, N is the number of unique hashtags and s_i is the number of hashtags that appear at least i times in the observations. This formula, however, only takes into account hashtags that appear once or twice in a linguistic community. To extend the amount of information used to estimate the richness of the conversation, we compute the improved Chao1 estimator $\hat{S}_{iChao1}$, which aims at correcting the first-order bias of $\hat{S}_{Chao1}$ by also including the number of triplets s_3 and quadruplets s_4. The improved Chao1 estimator is defined as [386]:

$$\hat{S}_{iChao1} = \hat{S}_{Chao1} + \frac{N-3}{4N}\frac{s_3}{s_4} \cdot max\big(s_1 - \frac{N-3}{N-1}\frac{s_2 s_3}{2s_4}, 0\big). \tag{6.6}$$

Figure 6.7A shows the improved Chao1 estimator for each linguistic community plotted against its population size. We measure a strong correlation ($R^2 = 0.92$) between the two variables, with bigger communities displaying a richer use of hashtags than smaller ones. There is, however, a non-negligible fluctuation in the measured richness for linguistic communities that share a similar population size, especially when looking at mid-sized communities.

To look more closely at this disparity, we display on figure 6.7B the improved Chao1 estimator per user, plotted against the population size. Points are coloured to highlight the Indian official languages. Interestingly, we notice that for a similar population size, the ratio is higher for Indian languages than for non-Indian languages (with English, a language widely used in Indian administration, being the exception).

These findings suggest that, with similar population sizes, Indian language speakers will give rise to a richer discourse on Koo than a non-Indian based linguistic community. These findings are especially insightful when considering that the Portuguese and Nigerian English speaking users were involved in far less rich conversations than their Indian peers, showcasing Koo's struggle to become a significant microblogging platform outside its native India [295].

The presence of a rich discourse among Indian communities does not necessarily indicate the existence of a national cohesion. The "divide and rule" policy, implemented by British imperials in India from 1857 to the country's independence in 1947, to sow divisions between ethnic groups in the country, led to long-standing cultural cleavages [387] and to the emergence of a communal, rather than national, sense of identity [388]. We can assess the influence of these divisions between communities on narrative content using the Bray-Curtis dissimilarity index [389], a measure of beta diversity, defined as

$$\beta_{AB} = \frac{\sum_i |p_i^A - p_i^B|}{\sum_i (p_i^A + p_i^B)}, \tag{6.7}$$

where p_i^A and p_i^B represent the relative frequency of hashtag i in the linguistic communities A and B, respectively. β_{AB} provides us with a measure of the dissimilarity between the two environments: if the two linguistic communities use no similar hashtags, their dissimilarity β_{AB} will be 1, whereas it will be equal to 0 if they use exactly the same hashtags.

Figure 6.7C shows the Bray-Curtis dissimilarity between the discourse in two linguistic communities. We notice that the Portuguese and Nigerian communities have a high dissimilarity when compared to the other languages, which indicates that there is a low overlap in the hashtags shared by Portuguese and Nigerian English speakers with respect to other linguistic communities. This finding further confirms their status as isolated clusters.

Looking at the Indian languages, we also notice a higher similarity shared between a few national languages, namely Gujarati, Bengali, Hindi and Marathi. These four languages are Indo-Aryan languages, thus sharing more linguistic similarities with each other than with Dravidian languages such as Telugu and Kannada. These results therefore outline the existence of a *Sprachbund*, a set of languages that share many similarities in their structure [390]. This concept seeps within the online discourse on Koo, where we saw a strong affinity between languages that share similar roots. Overall, Indian languages shared a more similar discourse with each other than with non-Indian languages on the platform, which might explain why the conversation was richer amongst these communities.

Discussion & Conclusion

In this chapter, we have analysed the emergence of a multi-lingual ecosystem on Koo, a microblogging platform based in India, which was shaped by successive migrations prompted by political and social events in India, Nigeria, and Brazil. We first looked at the impact of collective migrations to Koo in India, Nigeria and Brazil, and measured their respective success by comparing their impact on the daily registrations on Koo and their user retention over time. Second, we looked at the user interaction network and showed a strong linguistic segregation within isolated communities and the propensity for Hindi- and English-speaking users to be more central in the overall network. Third, we measured the average commitment and homophily for each linguistic community, and showed that linguistic communities with larger user-bases are more likely to be self-sufficient and dominantly display in-group interactions, with English being the exception due to its universal use across cultural communities. Fourth, we generated the global language network and highlighted the importance of linguistic crossovers between Indian languages. In contrast, Portuguese-speaking users only significantly overlap with English speaking users. Finally, we measured the richness and dissimilarity of linguistic clusters and noticed a strong interconnection across Indian languages, which also have a richer discourse than other communities with a comparable population size.

Our study offers a first insight into the emergence of a multi-lingual alt-tech platform, where the co-existence of diverse linguistic communities was driven by independent collective migrations. Despite Koo's ambitions to unite the non-English speaking world under a single banner, the language divide we measure both in terms of interactions and discourse similarity suggests that these communities were growing concurrently on the platform but without a strong tendency to overlap. While major languages could still thrive and be used by a sustainable community, less prevalent languages ended up being marginally used, leading their speakers to be less involved in the major conversations taking place on the platform. These findings might indicate that minority linguistic communities are disenfranchised from the central political discourse on social media. Previous inquiries have already raised concerns about the access to many public services in India for local language speakers [391]. Koo used to support 20 of India's 22 official languages, in a country with 122 languages listed in the national constitution.

Our analysis highlights the growth of the digital market in the Global South, which is often overlooked by the academic literature. Koo pledged to offer an online space catering to non-English speaking users and managed to attract key political figures from major emerging economies, leading the platform to become a serious competitor to X [392]. Moreover, our research adds more nuance to the literature on alternative platforms that has been mostly focused on right-wing narratives. Koo was used by both the Indian and the Brazilian governments, representing contrasting political orientations. However, the Indian and Brazilian communities on Koo were shown to hardly interact with one another, indicating scarce engagements across the political spectrum. Koo's struggles to attract members of political parties other than the BJP and their allies in India highlights the difficulties alt-tech platforms face in building an online space for politically diverse communities [149]. This is further exacerbated by the recent rise of policies related to digital sovereignty across a number of countries, which compromised Koo's ambition to become a unifying platform for communities outside the West [393].

Our findings are limited by several factors, which can be explored in future studies. First, our analysis is restricted to Koo, which was the most popular microblogging platform based in India, and the second most popular India-based social platform in general (after ShareChat). Social platforms based in other nations in the Global South may display different structural patterns and should also be studied. Future studies may also consider extending our analysis to other popular platforms based in India including ShareChat, 2go and Line. Second, our analysis relies on the language identified by Koo's automated system for each post and comment on Koo. This may be problematic given that automatic language detection has various levels of reliability depending on the source language [394]. Further work may consider the accuracy of Koo's language identification pipeline, and whether it results in specific biases in our analysis. Finally, our work focuses on a static overview of Koo's interaction network by looking at all the user activity that took place over the four years studied. This ensures that we are able to fairly compare diverse linguistic communities who joined Koo at different times. However, an analysis of the temporal evolution of communities on Koo would be equally valuable, in particular to identify how the network evolved following each collective migration.

Overall, our study improves our understanding of social interaction patterns within multilingual communities by looking beyond western social platforms. We anticipate that our work will inspire more research on the global social media ecosystem, to ensure that our findings with respect to alternative platforms are nuanced by cultural and linguistic factors. Finally, our analysis stresses the need to develop a wider range of tools to analyse social media content in minority languages, many of which are currently underrepresented on the internet, and understudied by academics.

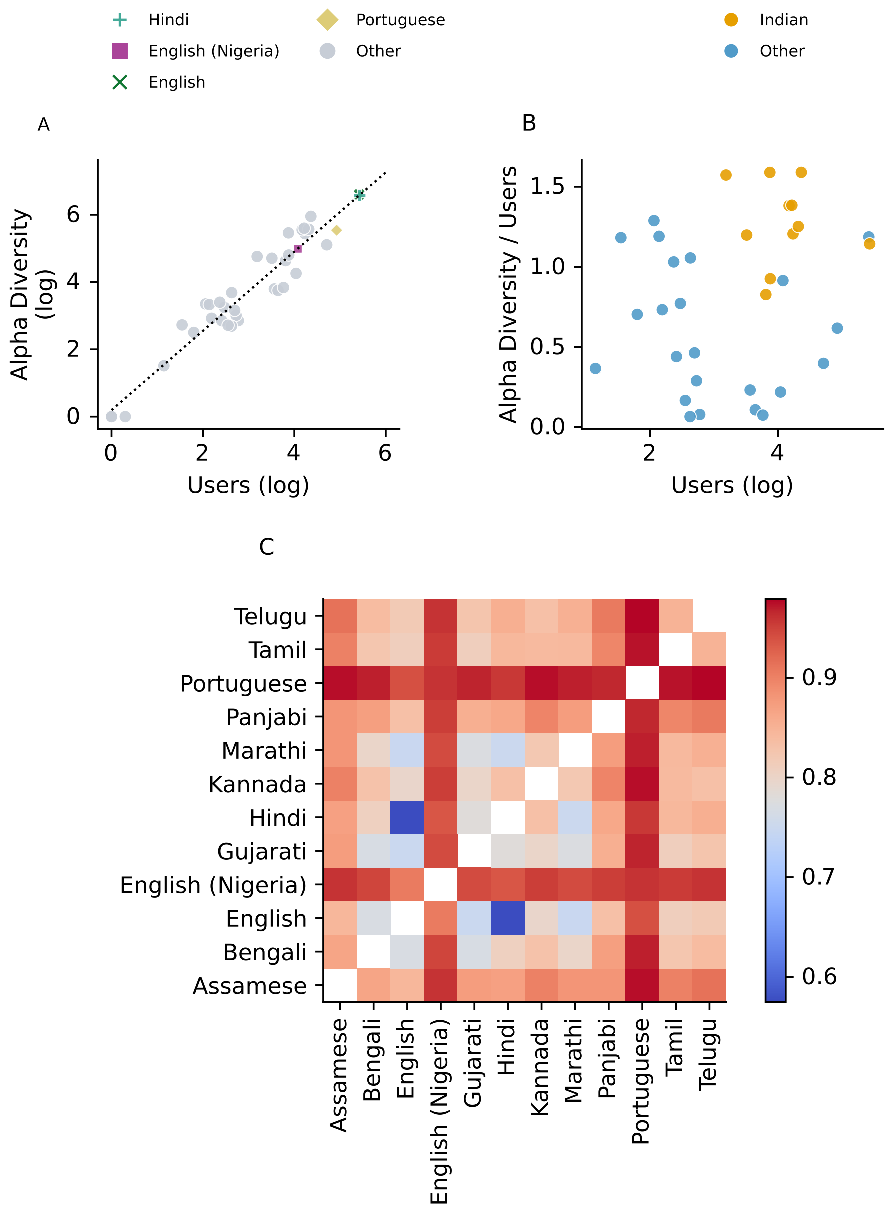


Figure 6.7: **Discourse richness and similarity across linguistic communities on Koo.** A) The alpha diversity of the discourse in a linguistic community, measured with the improved Chao1 estimator, plotted against the population size of the community, along with the linear fit (Spearman's $R^2 = 0.92$). Colours are used to indicate the main linguistic communities on Koo. B) The ratio between the improved Chao1 estimator and the population size plotted against the population size of the community. Colours indicate communities speaking an Indian language. C) The beta dissimilarity, measured with the Bray-Curtis index by considering the list of hashtags used by the largest linguistic communities on Koo and measuring their respective dissimilarity. Two communities with an index close to 0 use similar hashtags, whereas an index of 1 indicates that there is no overlap in the hashtags used by the communities.

Chapter 7

Conclusions

This thesis follows recent trends in content moderation for social platforms, and the new forms of online political engagement the mainstreaming of extremist ideologies harboured within the social media ecosystem. With the help of extensive datasets, providing detailed information on user-to-user interactions and user-generated content within alt-tech platforms, this work provides a timely comparison of the rise of several platforms, each with their own ambitions and catering to a specific fringe of the population. Our results illuminate several aspects of user engagement within alt-tech platforms that were previously understudied in the academic literature, and the main findings can be summarised along three main points.

First, we have improved our understanding of the global impact of deplatforming policies, by assessing how platforms such as Gettr, Voat and Koo capitalised on their mainstream competitors' bans on problematic communities. Each platform managed to attract a significant user base, by advertising their unregulated stance and their laissez-faire attitude towards content that would result in a ban on mainstream platforms. Moreover, the cases of Joe Rogan for Gettr, and Felipe Neto for Koo, highlight how online influencers can leverage their community by prompting a massive migration towards an alternative platform, even if it is not politically motivated. These findings raise a lot of debates related to the governance of online platforms, and the need to consider the health of the overall ecosystem when deciding to ban a high-profile user.

Second, we have provided novel insights regarding the nature of the communities that burgeon on alt-tech platforms, more so after a new user cohort migrates to the platform. For Gettr and Koo, we have highlighted how specific user cohorts stay more engaged over time on the alt-tech platform, either because they have been banned from established platforms, or because the alternative outlet is specifically catering to their needs. Very interestingly, for both Voat and Koo, we have also identified the birth of echo chambers within the platform, which can be triggered by ideological differences or by language barriers.

Finally, we have assessed that the rise of the alt-tech ecosystem in the Global South follows similar patterns of engagement than in Western communities. By looking at Koo's initial growth in India, and its subsequent expansions in Nigeria and Brazil, we stressed out the role that alt-tech platforms play in hosting nationalist political movements, after they ran into contentious situations with mainstream platforms' moderation teams.

Aside from highlighting the alt-tech platforms' opportunistic attitude towards deplatforming policies, our research also provided some insights on the role they play in radicalising their communities, in an online environment promoting free speech and amplifying fringe opinions. As such, Voat became a central hub in the dissemination and the study of the QAnon mythos [30], whereas Gettr was used by Bolsonaro supporters to sustain the threat of an election fraud, leading to the insurrection that targeted the Brazilian parliament in January 2023. We hope that our work will motivate more research to look at the online social media environment as a whole, rather than focusing on a single platform in the ecosystem.

Our main findings provide some explanations with respect to the rise of the alt-tech ecosystem over the last few years, as it morphed into an online venue for extremist political movements to gather support from their followers, either by crowdsourcing their actions or by organising demonstrations and riots. More fundamentally, our research can be summarised in three central findings.

Alt-tech platforms capitalise on massive deplatforming events from established platforms, by welcoming disenfranchised communities.

In chapters 3, 4 and 6, we established that alt-tech platforms' popularity primarily stem from their welcoming attitude towards communities that have been banned from mainstream platforms, or which feel censored by the platforms' moderation team. Voat's growth over the years is strongly punctuated by the bans that took place on Reddit, whereas both Gettr and Koo invested in high-profile politicians and influencers to attract new users.

Our findings raise some essential questions about the efficiency of deplatforming as a moderation policy. While it protects a vast majority of the active users of a platform from being exposed to harmful content, it provides unregulated alternative outlets with a significant flow of adopters, leading users to online echo chambers, where they are more likely to be radicalised and mobilised. Alt-tech platforms are also scarcely monitored by law enforcement and the academic community, which explains why telltale signs of a mobilisation for the US Capitol Riot did not raise any alarm within police forces or the U.S. Department of Homeland Security [395]. Moreover, as discussed in chapter 2, alt-tech platforms have become an ideal breeding ground to foster extremist ideologies, which are rarely challenged by a dissident voice within the platform.

The growth of the alt-tech ecosystem follows very similar trends in the Global South than within Western nations, by welcoming communities that feel censored by established platforms.

As mentioned in chapters 3 and 4, Voat and Gettr earned their status in the social media ecosystem by fostering far-right leaning communities, predominantly from the US, after they have been deplatformed from established platforms. As such, they swiftly profiteered off any major ban taking place on Reddit and Twitter respectively, and ended up hosting a diverse range of sub-communities.

On the other side, as covered in chapter 5 and 6, Koo strove to be a safe venue for diverse ethnic groups and political opinions. However, by hosting the BJP after the farmers' protest in India, the platform struggled to attract politicians from diverse political leanings [149], thus displaying a lack of political plurality.

Our results indicate that the alt-tech platform's role in the digital ecosystem mostly lies in its proclivity to act as a safe haven for online communities that have been censored or banned by mainstream platforms, while also catalysing the rise of conformist online societies.

Alt-tech platforms have evolved to become a patchwork of fragmented clusters, where siloed communities can coexist without any cross-interaction taking place.

In chapter 3, we looked at the interaction network between the users on Voat, highlighting the existence of a few siloed subverses, whose interactions with the overall user base are very sparse. These communities all moved from Reddit after being banned, and subsequently joined Voat. However, their proclivity towards online harassment, hate speech and conspiracy theories prompted them to act as outliers on the alt-tech platform, highlighting that users on such platforms do not adhere to unified boundaries with respect to toxic behaviour.

Meanwhile, in chapter 6, we showed that the interaction network on Koo is strongly shaped by linguistic preferences, where users are more likely to interact with their peers if they predominantly use the same language. This phenomenon is further exacerbated by cultural differences, as we notice that Nigerian and Brazilian users are strongly isolated from their Indian peers. Meanwhile, the English language acts as a lingua franca and plays a pivotal role in political communication on the platform, despite Koo's ambition to host non-English speaking communities across the world.

Overall, our results illuminate several aspects related to user engagement on alt-tech platforms, and the role it plays in shaping their group identity. We hope that these insights will help inform policies around deradicalisation strategies [396], content moderation on social media and the growing need to monitor extremist groups online.

The work presented in this thesis was initially motivated by the rise of bans taking place on established social media platforms and a lack of evidence regarding their efficiency. Deplatfroming has several positive consequences, such as the decrease in toxic language used on the platform where the ban took place [70]. Moreover, due to their smaller size, alt-tech platforms do not provide influential figures with an audience of a scale similar to the one they can interact with on mainstream outlets [241]. As such, deplatforming is a successful policy when considering the audience of the platform enforcing the bans, as its users see their exposure to anti-social behaviour being decreased. However, I argue that, in the light of the results shown in this thesis, deplatforming mostly implies moving the problematic users to new platforms, where toxic behaviour is condoned and oftentimes encouraged. As such, it makes it much harder for law enforcement to monitor these communities and their propensity to be involved in riots or lone wolf terrorism. More importantly, any user who becomes active on alt-tech platforms will no longer be exposed to nuanced and conflicting opinions, decreasing the odds of disrupting their radicalisation journey. Instead of capitalising all their efforts on identifying and banning problematic communities, I would urge platforms to consider implementing interventions that aim at deradicalising users, such as the nudges that X (then Twitter) tested in 2021 to incite users to avoid using toxic speech [397].

While our findings offer a major improvement to our understanding of the online extremist ecosystem, they also signpost several new avenues that need to be investigated in future research. We provide the reader with some potential research questions, that would further solidify our expertise in online radicalisation.

How do radicalised users stay engaged across platforms?

While deplatforming used to imply an initial ban from a mainstream platform, the recent growth of the alt-tech ecosystem spurred many political actors to be active on both mainstream and alt-tech platforms. However, due to the difficulties, both technical and ethical, in matching user accounts between platforms, there is little evidence to indicate whether their discourse follows similar patterns on both platforms. A leading hypothesis suggests that their discourse would be more toxic on alternative outlets, as they have to navigate established platforms' moderation policies to avoid being banned [171]. However, our analysis of Gettr highlights a higher propensity to use toxic speech on Twitter rather than Gettr, for some user cohorts. As such, more analysis is required to unravel whether there is also a strong difference in terms of the narratives being shared on both mainstream and alt-tech platforms. An alternative explanation, which we explored for Gettr and which was also studied with Parler [398], motivates the increase in hate speech on mainstream platform by the hostility that arises when extremist users interact with members of their out-group. Moreover, future work could build upon previous studies related to crowdsourcing within extremist movements [399], to assess whether content shared on alt-tech platform is more likely to include links to crowdfunding platforms.

Do alt-tech platforms share a common narrative, with communities coordinating across platforms?

One of the growing concerns related to the alt-tech ecosystem is that platforms seem to metastasise into an organic environment, running parallel to the mainstream social media ecosystem. For example, a previous study underlined how white supremacist servers on Discord were strongly interconnected in the run-up to the *Unite the Right* rally in Charlottesville, Virginia [131], with similar interaction patterns identified for the QAnon community [19].

With extensive datasets now publicly available for a significant number of alt-tech platforms, an essential question to investigate relates to assessing their similarity in terms of overall discourse, news ecosystem and patterns of user engagement. By doing so, we would be able to assess whether alt-tech platforms lead to a similar social hierarchy, with a strong divide between content producers and consumers. A comparative analysis would also provide some valuable insights on the factors that lead some alt-tech platforms to become successful online venues, whereas others struggle to keep an engaged community.

How does a user's behaviour change over time after they join an alt-tech platform?

As mentioned in chapter 2, users who join online extremist communities are likely to undergo an extensive radicalisation process, triggered by their exposure to alternative, oftentimes hateful narratives, until they become themselves messengers of the movement. Scholars aimed to understand whether this process leads to a shift of socialisation patterns outside of the extremist communities, with the evidence highlighting an adoption of far-right vocabulary and symbols within three months of exposure [400]. These results indicate that the process, starting from the initial exploration phase to the adoption of a fully radicalised ideology, is very swift and leads to a shift of a person's group identity.

These findings can be further explored by comparing how the adoption of a social identity varies across platforms. For example, *brigading* is a common phenomenon on Reddit, where one community starts invading another subreddit by spamming with posts and comments that further their agenda. This phenomenon has been extensively studied on Reddit [171, 401], but there has not been further research, aiming at investigating this phenomenon originating from other alt-tech platforms. Such a study could further broaden our understanding of the anti-social collective behaviours that are promoted on alt-tech platforms, the nature of their targets and the tools they use to game the mainstream platform's algorithm.

Overall, this thesis brought a more nuanced of the impact that deplatforming policies can have not only on a single platform, but on the overall social media ecosystem. As the alt-tech ecosystem is constantly growing and there is more appetite for influential politicians to become involved with these communities, there is an urgent need to understand how these platforms act as facilitators when users are following a radicalisation pipeline and whether deplatforming might lead disenfranchised communities to end up on non-moderated and non-monitored platforms. Aside from the risks that such a process might lead to an even more divided society, previous events have showcased how the online radicalisation process can also lead to offline unrest or violence.

Chapter A

Appendix to Chapter 4

A.0.1 Manual account labelling

Our primary analysis heavily relies on Gettr users voluntarily disclosing their corresponding Twitter accounts, enabling us to conduct a cross-platform comparison of their activities. To ensure the accuracy of account matching, we manually examined the 300 accounts with the highest follower count on Gettr, out of the 1588 matched users we have in the cohort. The main aim is to confirm that the accounts on both platforms belong to the same person. Manual matching is a subjective judgement based on whether an account uses the same, or similar, handle on both platforms, the same, or similar, profile picture, and uses the same, or similar, profile biography. Out of the 300 accounts checked manually, 285 were successfully matched, corresponding to a 95% match rate. This match rate is comparable to the match rate found in previous studies using a similar method [241].

Following Elon Musk's amnesty on suspended accounts on Twitter announced in November 2022 [402], we also manually checked the 100 banned accounts with the highest follower count on Gettr. 33 of these accounts have since seen their ban being lifted on Twitter, thus confirming that Elon Musk did indeed offer such an amnesty to banned users. Interestingly, we noticed that many of these accounts share very similar narratives, primarily centred to Donald Trump's political propaganda and conspiracy theories revolving around Covid-19. Due to changes in the policy related to the Twitter API, we cannot study the activity of these accounts on Twitter after their ban has been lifted.

A.0.2 User acquisition and activity

One of the main findings of our analysis is that banned users are, on average, more active than non-verified or matched users, a first indicator that the banned cohort is more likely to benefit from being able to register on fringe platforms with a less stringent moderation policy. In particular, we notice that matched users are seven times more active on Twitter than on Gettr, wheras banned users are five times more active than matched users on Gettr. To ensure that our results are not conflated with the presence of various linguistic communities in the platform, we run the same analysis separately for the two major demographics on Gettr, i.e. by considering only the English- and Portuguese-speaking users. The results for the English-speaking community are displayed on figure A.1, and show very similar trends than when considering the platform as a whole. Most notably, we still notice a peak of registrations in January 2022, related to Joe Rogan announcing his account on Gettr and which had a large impact on US-based users. Figure A.1B also shows a larger propensity for banned users to be active on Gettr, when compared to matched and non-verified users.

Figure A.2 shows the same results with the Portuguese-speaking cohort. An interesting observation on figure A.2A is the absence of a peak of registrations in January 2022. This suggests that Joe Rogan's social media outreach mostly consists of English-speaking users, and therefore his registration on Gettr did not lead to a surge of Portuguese-speaking users on Gettr. However, there is a noticeable peak of registrations in September 2021, which is the result of Gettr's involvement in the Brazil CPAC 2021, with its CEO Jason Miller traveling to Brazil to attend the event [180]. Figure A.2B highlights how banned users within the Brazilian community are also more active on Gettr than the other cohorts.

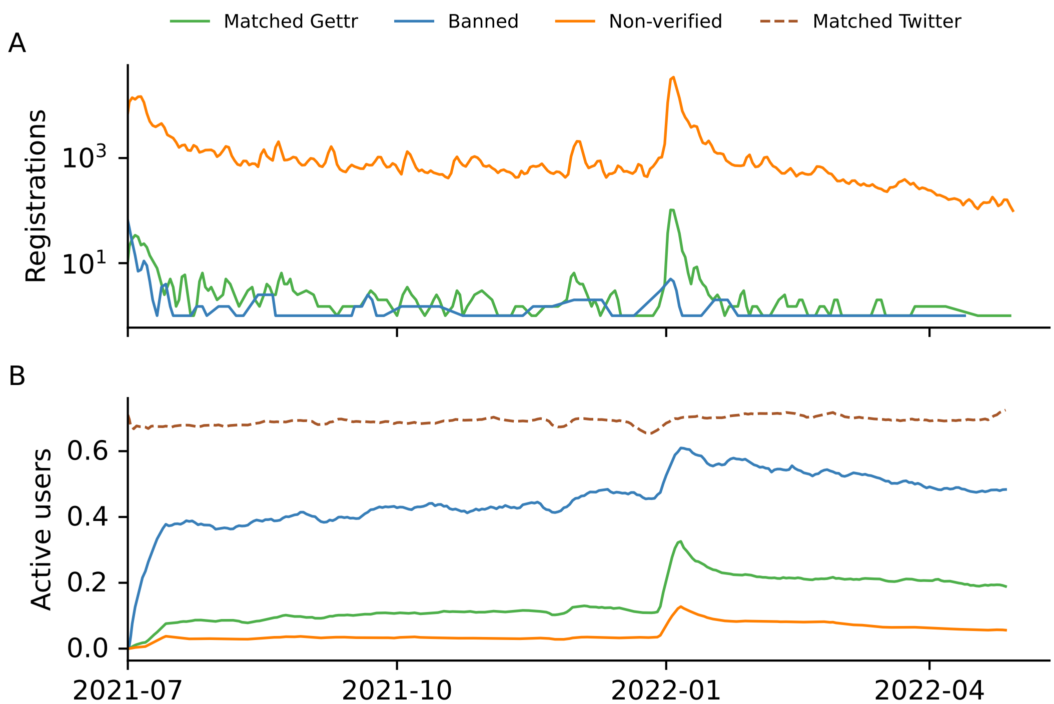


Figure A.1: **User registrations and daily activity for each cohort - English-speaking cohort.** (**a**) 3-day moving average of the daily number of users who registered on Gettr. The curve is displayed separately for the banned cohort (blue), the matched cohort (green) and other users who are not-verified on Gettr (orange). (**b**) 7-day moving average of the proportion of users from each cohort who were active on Gettr on a given day. The percentage of the matched cohort active on Twitter is also shown (dashed brown). Only English-speaking users are considered for this analysis.

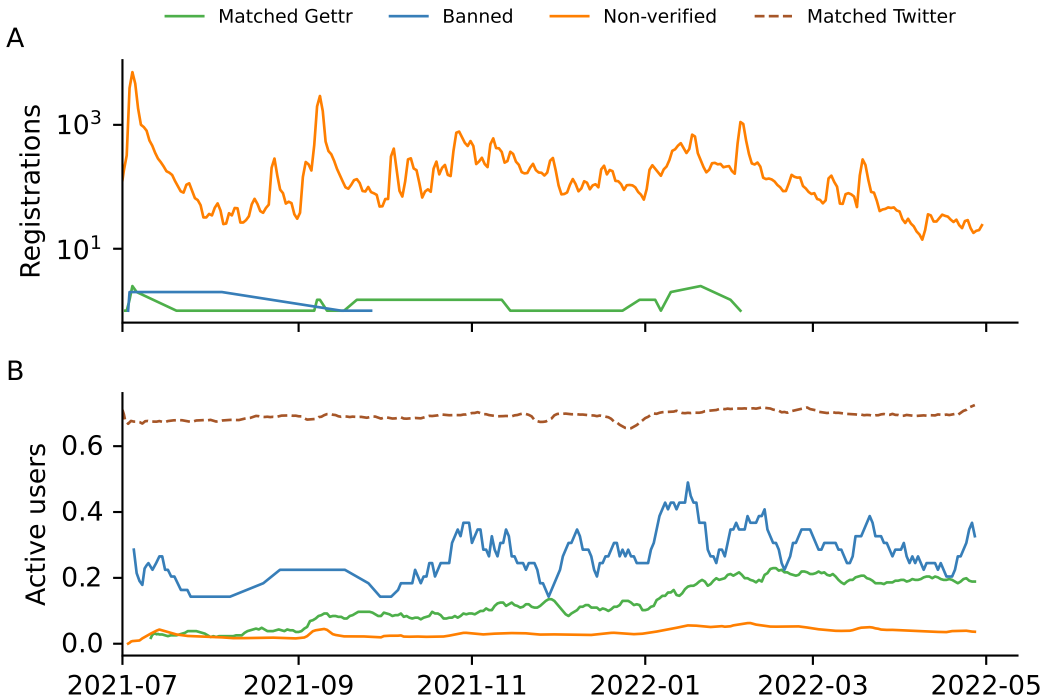


Figure A.2: **User registrations and daily activity for each cohort - Portuguese-speaking cohort.** (**a**) 3-day moving average of the daily number of users who registered on Gettr. The curve is displayed separately for the banned cohort (blue), the matched cohort (green) and other users who are not-verified on Gettr (orange). (**b**) 7-day moving average of the proportion of users from each cohort who were active on Gettr on a given day. The percentage of the matched cohort active on Twitter is also shown (dashed brown). Only Portuguese-speaking users are considered for this analysis.

A.0.3 User retention on Gettr

Our main analysis focuses on the two months with the highest numbers of new registrations: July 2021, when Gettr was first launched, and January 2022, when Joe Rogan joined the platform and led to almost a million new users to sign up as well. To ensure that our results also hold for any other month in our dataset, we compute the Kaplan-Meier estimate for every month up to April 2022, by separating the user cohorts mentioned in the main analysis (banned, matched and non-verified users). The results of the analysis are shown in figure A.3. We notice that banned users are more likely to be active on the platform for a longer period of time, regardless of the considered registration month, whereas users in the matched and non-verified cohorts become inactive shortly after joining Gettr.

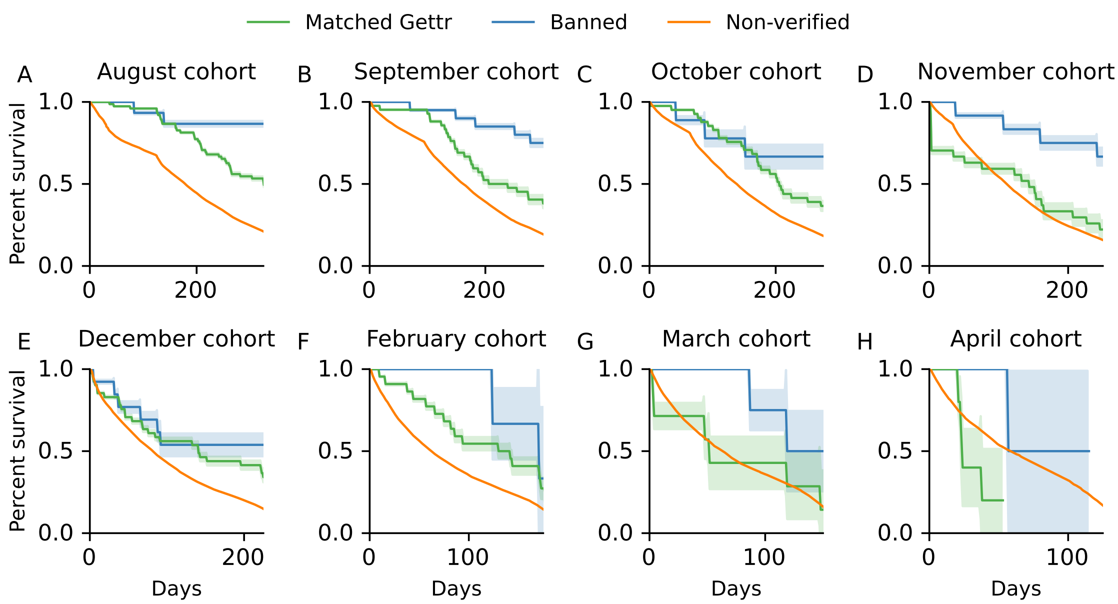


Figure A.3: **User retention for other registration months** (A) Kaplan-Meier survival curves for each user cohort showing the fraction of accounts who registered in August 2021 who remain active on Gettr a given number of days after registration for the banned cohort (blue), matched cohort (green) and the non-verified cohort (orange). The standard error of each curve is computed using Greenwood's formula [1] (see Methods). (B-H) Survival curves for other registration months, between September 2021 and April 2022 (excluding January 2022)

A.0.4 English-language Topic Modelling

To better understand Gettr content we employ a topic model, trained on English language Gettr posts using BERTopic [245] (see Methods), to characterise key discussion themes on the platform. The top 20 topics of classified Gettr content are shown in Table A.1 alongside a representative Gettr post for each topic extracted automatically by BERTopic. Using the same model, we classify matched Twitter content using the same model; the promimence of each topic on both Gettr and Twitter is shown in Table A.2, with the ratio column indicating whether a topic is over- or under-represented on Gettr relative to Twitter. We list the median toxicity of posts classified as part of that topic on Gettr and on Twitter in table A.3.

The topic modelling reveals that content on Gettr is dominated by topics of significant relevance to US politics, and in particular of relevance to the political right. Topic 1 (Covid-19, vaccines) corresponds to approximately one sixth of all classified content on Gettr, with this value approaching a third in some months. Topic 2 focuses explicitly on discussions around major social media platforms referencing the political right being deplatformed from Facebook and from Twitter.

It is worth considering those topics which are found disproportionately on Gettr, relative to the matched Twitter dataset. This includes topics related to the war in Ukraine (No. 3), the Canada Convoy protests which related to Covid-19 vaccine mandates (No. 4), and the 2020 US election and alegations of voter fraud, particularly in Arizona (No. 10). These are topics which are known to have been targets of the Twitter content moderation team; several accounts were suspended for (1) sharing pro-Russian content during the Ukraine war [246], (2) sharing Covid-19 misinformation [247], and (3) for supporting the January 2021 insurrection following the 2020 US elections [248]. Consequently, it is not surprising that these topics are more prominent on Gettr than they are on Twitter (note that we do not have access to deleted Twitter content). If we consider the toxicity of topics, we find that topics with disproportionately high toxicity are topic 20 on race and Black Lives Matter, topic 16 on female US Democratic politicians, and topic 14 on gender issues.

No.	Topic	Representative post
1	Covid-19, Vaccines, Virus, Fauci	"More proof the jab doesn't work as whoopiiiigasbag's overweight triple vaxxxed ass gets hijacked by the cofraud-19!"
2	Twitter, Gettr, Facebook, Musk	"Facebook is censoring conservatives like crazy. I literally said exactly the same thing are illegitimate president said and I was banned from facebook for bullying. It's so nuts."
3	Ukraine, Putin, Russia, War	"Klaus Schwab and George Soros have been quite loud about this but no one is listening because they are using Ukraine as a distraction. They are also using the Ukraine invasion as a catalyst for the NWO aka great reset believe me the NWO is pure evil."
4	Canada, Truckers, Trudeau, Convoy	"Canada is under attack from its federal and provincial governments who have been bought and sold by big pharma. "
5	God, Lord, Jesus, Christ	"Hey Christian! What type of vessel do you desire to be for the Lord? #ChooseChristAlways"
6	Biden, Joe, Administration, President	"I have never personally seen or heard a bigger moron than Biden. He simply isn't home. He needs to retire so the GOP can impeach hahadahoe Indian name."
7	China, CCP, Communist, Taiwan	"This is how ccp educates young kids to be a patriot in china. It reminds me that I was brainwashed and taught since the elementary school #TakedowntheCCP #CCPisnotchines"
8	Podcast, Watch, Rogan, Episode	"Check out our latest episode of the brothers tao podcast. We discuss the #Truckersforfreedom as well as the #whoopigoldberg situation. We also take a look at the joerogan saga and what it means for censorship."
9	Border, Illegal, Southern, Immigrants	"The absolute nerve of Biden saying we need to secure the border that he deliberately opened wide for criminal to invade our country is laughable. Actually its fkg disgusting."
10	Election, Fraud, Audit, 2020	"Question. Would the riots at the capitol be justified if the election was stolen? You know how I know it was stolen? All the same people who told us that trump stole the 2016 election says 2020 was fair…they lie."
11	Afghanistan, Taliban, Biden, Kabul	"Bin Laden is dead but Al Qaeda's #2 Ayman Al-zawahri came out to celebrate the 20th anniversary of 911. It wouldn't be so sad if the pentagon front row wasn't filled with turncoats."
12	Christmas, Art, Merry, Day	"My wife saw you on the warroom today and loves the colour of your lipstick. She would like to know the brand and what colour red."
13	Abortions, Baby, Roe, Wade	"If a woman leaves her state where abortion is illegal and kills her baby through abortion, can she be charged with murder when she comes back home?"
14	Gender, Woman, Trans, Sex	"I have 6 male dogs, 5 are neutered, none of my dogs ever identified as a female. None went to public school either though."
15	Freedom, Government, People, Country	"I highly suggest this article... Fight for freedom individual freedom for everybody like the constitution guarantees!"
16	Pelosi, Nancy, Woman, Hillary	"So she is saying fck the ppl and the voters. We are still passing of our agenda she's evil and should be in prison."
17	Supreme Court, Jackson, Scotus, Judge	"Can't believe they nominated Jackson porno sympathizer a complete libtard."
18	Guns, Gun Control, School, Shooting	"The government offered to buy my guns from me, but after a thorough background check of the buyer, I'm not comfortable with selling weapons to organized crime."
19	Insurrection, Jan 6th, Antifa, Prisoners	"Keep reading updates on the abuse of the Jan 6 political prisoners and am wondering when someone in power is going to put a stop to it. Is there not one judge/law enforcement official/politician who can challenge this lawless confinement!!!??!"
20	Black, White, Racism, BLM	"Presuming or insinuating a race is oppressed gives the impression that that race is sub-par. White democrats need to stop speaking on behalf of other races. Oppression is more of a social class issue than anything, it doesn't know race nor gender."

Table A.1: Topics extracted from Gettr posts using BertTopic. Each topic is accompanied by a representative post, extracted automatically by BertTopic. Posts have been minorly edited for clarity and to remove unnecessary text such as URLS.

No.	Topic	Gettr %	Twitter %	Ratio
1	Covid-19, Vaccines, Virus, Fauci	16.4	10.6	1.6
2	Twitter, Gettr, Facebook, Musk	4.6	3.9	1.2
3	**Ukraine, Putin, Russia, War**	3.9	1.6	**2.5**
4	**Canada, Truckers, Trudeau, Convoy**	3.4	1.3	**2.7**
5	God, Lord, Jesus, Christ	3.2	3.7	0.9
6	**Biden, Joe, Administration, President**	1.8	0.7	**2.5**
7	China, CCP, Communist, Taiwan	1.6	1.0	1.6
8	Podcast, Watch, Rogan, Episode	1.4	3.0	0.5
9	Border, Illegal, Southern, Immigrants	1.4	1.0	1.4
10	**Election, Fraud, Audit, 2020**	1.3	0.5	**2.9**
11	Afghanistan, Taliban, Biden, Kabul	1.0	1.0	0.9
12	Christmas, Art, Merry, Day	0.9	1.4	0.6
13	Abortions, Baby, Roe, Wade	0.8	1.2	0.7
14	Gender, Woman, Trans, Sex	0.8	0.9	0.9
15	**Freedom, Government, People, Country**	0.8	0.3	**3.0**
16	Pelosi, Nancy, Woman, Hillary	0.7	0.5	1.4
17	Supreme Court, Jackson, Scotus, Judge	0.7	0.4	1.6
18	Guns, Gun Control, School, Shooting	0.7	0.7	1.0
19	Insurrection, Jan 6th, Antifa, Prisoners	0.7	0.4	1.9
20	Black, White, Racism, BLM	0.7	0.6	1.2
Other	N/A	14.0	22.5	0.6
Outliers	N/A	39.2	42.9	0.9

Table A.2: **Topics on Gettr and their relative prominence in the matched Twitter dataset.** Topics are listed in order of size and are characterised by a small number of keywords identified using BERTopic, see Methods. The ratio column indicates the prominence of a topic on Gettr, divided by its prominence on the matched Twitter dataset. Topics highlighted in bold are more than twice as prominent on Gettr than on Twitter. Topics not in the top 20 are grouped in the "other" category. BERTopic classifies documents as "outliers" if a topic does not correspond to a defined category.

No.	Topic	Gettr Toxicity	Twitter Toxicity
1	Covid-19, Vaccines, Virus, Fauci	0.11	0.06
2	Twitter, Gettr, Facebook, Musk	0.16	0.10
3	Ukraine, Putin, Russia, War	0.23	0.11
4	Canada, Truckers, Trudeau, Convoy	0.15	0.05
5	God, Lord, Jesus, Christ	0.10	0.06
6	Biden, Joe, Administration, President	0.34	0.18
7	China, CCP, Communist, Taiwan	0.19	0.09
8	Podcast, Watch, Rogan, Episode	0.08	0.04
9	Border, Illegal, Southern, Immigrants	0.22	0.11
10	Election, Fraud, Audit, 2020	0.15	0.05
11	Afghanistan, Taliban, Biden, Kabul	0.25	0.16
12	Christmas, Art, Merry, Day	0.06	0.05
13	Abortions, Baby, Roe, Wade	0.30	0.13
14	Gender, Woman, Trans, Sex	0.38	0.25
15	Freedom, Government, People, Country	0.20	0.10
16	Pelosi, Nancy, Woman, Hillary	0.38	0.20
17	Supreme Court, Jackson, Scotus, Judge	0.29	0.10
18	Guns, Gun Control, School, Shooting	0.23	0.12
19	Insurrection, Jan 6th, Antifa, Prisoners	0.19	0.12
20	Black, White, Racism, BLM	0.40	0.36

Table A.3: **The median toxicity of posts classified as part of each topic on Gettr and Twitter.** The Gettr toxicity is computed using all Gettr posts. The Twitter toxicity is computed using only posts from the matched cohort. On both Gettr and Twitter, the two topic with the largest toxicity are topic 20 relating to race and topic 14 relating to gender. Topics regarding Democrat politicians are also disproportionately toxic.

A.0.5 Cohort toxicity over time

In Fig. 3C of the main manuscript, we show the median toxicity of posts authored a fixed number of days after a user authored their first post on Gettr (or after their first post in our observation time window on Twitter). To rigorously assess whether post toxicity is increasing or decreasing over time, we compute a linear fit ($y = ax + b$, where a and b are slope and intercept parameters) of the normalised daily toxicity using ordinary least squares. To compute an error on the fit, we compute a bootstrapped median of each daily median toxicity 100 times using 50% of the data with replacement, resulting in 100 individual fits on the daily normalised toxicity graph. This results in median fitting parameters (95% confidence interval in square brackets) for each cohort of:

- **Non-verified:** $a = 0.0007$, $[-0.0157, 0.0139]$ units toxicity per year. $b = 0.1687$, $[0.1643, 0.1732]$.
- **Banned:** $a = -0.0091$, $[-0.0152, -0.0022]$ units toxicity per year. $b = 0.0503$, $[0.0482, 0.0530]$.
- **Matched Gettr:** $a = -0.0120$, $[-0.0137, -0.0103]$ units toxicity per year. $b = 0.0450$, $[0.0443, 0.0458]$.
- **Matched Twitter:** $a = 0.0100$, $[0.0072, 0.0127]$ units toxicity per year. $b = 0.0787$, $[0.0775, 0.0801]$.

Relative to the inter-quartile range of the post toxicity across our observation period, the median yearly change in toxicity for each cohort corresponds to:

- **Non-verified:** $0.0007/0.31 \approx 0.2\%$.
- **Banned:** $0.0091/0.13 \approx 7\%$.
- **Matched Gettr:** $0.0120/0.09 \approx 13\%$.
- **Matched Twitter:** $0.0100/0.18 \approx 6\%$.

This clearly shows that linear fits for the annual change in cohort toxicity are significantly smaller than the daily variability in post toxicity within each cohort. Consequently, we can consider the annual change in toxicity for each cohort negligible in the context of the expected fluctuations in post toxicity.

A.0.6 Quote-ratio Statistics

Statistics for the difference between the all user quote-ratio distribution shown in Fig. 5A of the main chapter and each subdistribution is shown in table A.4. For each statistical comparison we use a 2-sample Kolmogorov-Smirnov test to assess the difference between the test distribution and the all user distribution. The table provides the Kolmogorov-Smirnov test and corresponding p-value. A p-value less that $p = 0.01$ indicates that the test distribution is significantly different to the all user baseline. To quantify the magnitude of the difference between the distributions, we use a non-parametric analogue of the Cohen's d effect size. A verbal descriptor for the effect size is provided according to the rules of thumb in [4]. Using this terminology quantifies the difference between the all-user baseline and questionable media sources as very small (and not statistically significant). In contrast, the difference between the all-user baseline and Democrat politicians is statistically significant and described as huge according to the Cohen's d rule of thumb.

Distribution	**KS Statistic**	**KS p-value**	**Cohen's d**	**Cohen's d descriptor**
Matched cohort	0.22	2.8e-35	0.28	Small
Republicans	0.32	1.8e-11	0.56	Medium
Democrats	0.63	3.1e-16	2.33	Huge
Far right	0.23	0.24	0.20	Small
Right	0.21	0.002	0.43	Small
Least	0.52	1.3e-13	1.16	Large
Left	0.55	7.1e-35	1.41	Very large
Reliable	0.44	1.3e-41	1.15	Large
Questionable	0.15	0.29	0.05	Very small

Table A.4: Statistics for the difference between the all user baseline distribution in Fig. 5 of the main chapter, and each subdistribution listed in the left-most column. For each comparison we provide the KS-test statistic, the corresponding p-value, the Cohen's d effect size, and a verbal descriptor for Cohen's d, according to best practice in [4].

A.0.7 Gettr's wider impact on right-wing politics - the case of Brazil

Journalistic reports have suggested that Gettr played a key role in facilitating the Brasília insurrection on January 8, 2023, following Jair Bolsonaro's defeat in the Brazilian Presidential elections [403, 404]. Here we investigate whether there is evidence for this role in the Gettr interaction network.

First, we study the power imbalance in the Portuguese language network by measuring the Gini coefficient of the degree distribution, shown in Fig. A.4A. The figure shows that the Gini coefficient peaked in the run-up to the Brasília riots, which is evidence that a handful of users were responsible for shaping the collective narrative of the Portuguese language Gettr community [405, 406].

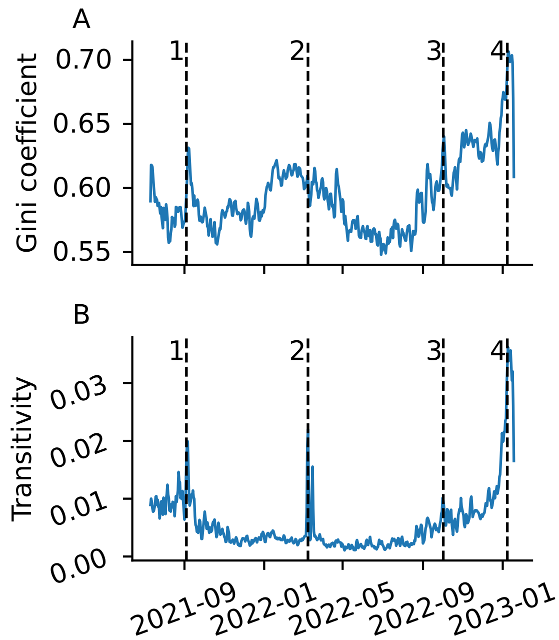


Figure A.4: **Evolution of the interaction network in the Brazilian community.** Analysis of the daily interaction network, generated by considering any interaction within a 1-day window. (A) Gini-coefficient of nodes in the giant connected component. (B) Transitivity of the giant component. Dashed lines correspond to key events related to Brazilian politics and Gettr's involvement: (1) 2021 CPAC Brazil Conference, (2) the "Ato pela terra" demonstration organised in Brasília against Bolsonaro's "Package of Destruction" laws [3], (3) the Brazilian presidential election, and (4) the Brazilian Congress attack in Brasília.

We now study grassroots engagement, measured by computing the transitivity of the Gettr interaction network, see Fig. A.4B. This measure increases when a community of users densely interact with one another [407, 408]. The figure shows that network transitivity peaked during CPAC 2021, where the Bolsonaro regime and Gettr shaped their close alliance [180], and in the days leading up to the Brasília riots. Applying a Portuguese-language topic model to the network reveals that users were discussing accusations of rigged elections and claims of a corrupt media in the lead up to the riots.

The peak in both the Gini coefficient and the transitivity shows that leading Bolsonaro allies successfully capitalised on accusations of election fraud to generate a grassroots movement on Gettr in the wake of Bolsonaro's defeat in the Brazilian elections. These results offer new quantitative insights which build on journalistic reports of Gettr's role in the riots. Critically, our results show that even when a platform appears largely inactive, a community of idle users can be mobilised within a short time period leading to real world harms. Note, however, that we cannot claim a direct link between deplatforming induced migrations to fringe social media and Gettr's potentially harmful role during the Brazilian insurrection.

A.0.8 Portuguese-language Topic Modelling

Our analysis on the Brazilian community on Gettr being focused on the impact online propaganda can have on offline upheavals, we provide the same analysis of the most salient topics mentioned by that community as we did for the English-speaking users. We use BERTopic to identify the key topics that shape the conversation within that linguistic cohort. The dominant ones are listed in table A.5. The topic modelling indicates that election frauds allegations are strongly discussed among the Brazilian community, which confirm the suspicions that Gettr was a key player in spreading that narrative. Notably, messages about the elections represent about 3.7% of the messages (No. 2), whereas Lula is strongly accused of stealing the elections (No. 4). Moreover, several actors who allegedly contributed in covering up the fraud, such as the mainstream media (No. 9) [409] or the Supreme Federal Court (No. 11) [410], are recurrently mentioned. Some of the key members of the Brazilian far-right circle, who are verified users on Gettr are also widely discussed within the community (No. 17). Interestingly, several of these topics are also disproportionately found within the English-speaking community (see table A.2). Some examples are the Covid-related conspiracy theories (No. 6), the Ukrainian war (No. 10), and the moral panic around abortion (No. 8). These findings further reinforce the overlap between the populist right in the United States and Brazil, and indicate that such narratives are platform-centric rather than bound to a specific demographic.

No.	**Topic**	**Gettr %**
1	Brazil, People, Country	14
2	Polls, Fraud, Vote, Elections	3.7
3	Lord, God, Jesus Christ, Truth	3.5
4	Thief, Bandit, Lula, Crime	3.4
5	War, Military, Patriots, Freedom	2.1
6	Health, Vaccines, Covid-19, Pfizer	1.7
7	Inflation, Salary, Economy, Budget	1.3
8	Children, Abortion, Gender, Pedophilia	1
9	Media, Press, Fake News, Journalists	1
10	Ukraine, Russia, Putin, NATO	0.9
11	Supreme Federal Court, Ministers, Constitution	0.7
12	Camp, Concentration Camp, Elderly, Children	0.6
13	Christians, Church, Pope, Satan	0.6
14	China, Communist Party, Covid-19	0.6
15	Patriots, Insiders, Terrorists, Left	0.6
16	Media, Press, Brazil, Truth	0.6
17	Jair Bolsonaro, Flavio Bolsonaro, Carla Zambelli, Carlos Jordy	0.5
Other	N/A	10.7
Outliers	N/A	54

Table A.5: **Topics which correspond to more than 0.5% of posts on Gettr in the Brazilian community.** Topics are listed in order of size and are characterised by a small number of keywords identified using BERTopic, see Methods. BERTopic classifies documents as "outliers" if a topic does not correspond to a defined category.

Chapter B

Appendix to Chapter 6

B.0.1 Linguistic communities' longitudinal evolution

We showed previously that Koo experienced several growth spurs, following collective migrations from various demographics. We show that these additional registrations also led to an increase of activity coming from these groups, by breaking down the daily activity by linguistic communities. Figure B.1 displays the moving average of daily registrations for each linguistic community. As we can see, the peaks that we noticed previously are indeed caused by the BJP migration, the Nigerian ban on Twitter and the Brazilian community moving to Koo. Interestingly, each peak also leads to a moderate growth of interest among other linguistic communities.

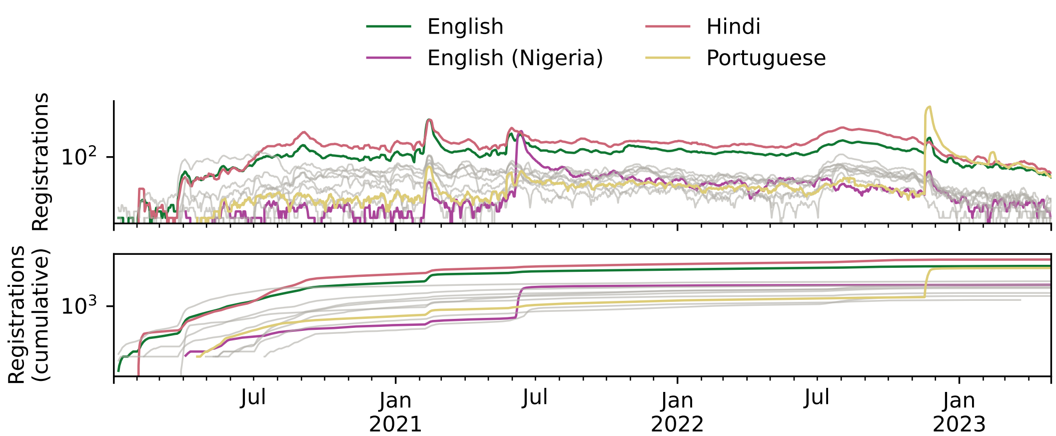


Figure B.1: **Registration activity from linguistic communities.** Time-series showing the 7-day moving average of the number of registrations made by each linguistic community. The major communities are highlighted in colour.

Figures B.2, B.3 and B.4 display the moving average for the number of comments, shares and likes made by each linguistic community on a daily basis, showcasing a burst of activity on days related to the collective migrations observed on Koo.

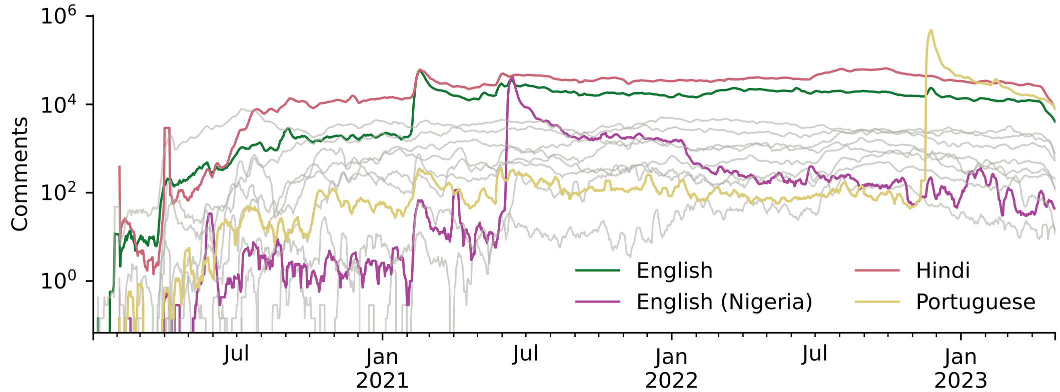


Figure B.2: **Commenting activity from linguistic communities.** Time-series showing the 7-day moving average of the number of comments made by each linguistic community. The major communities are highlighted in colour.

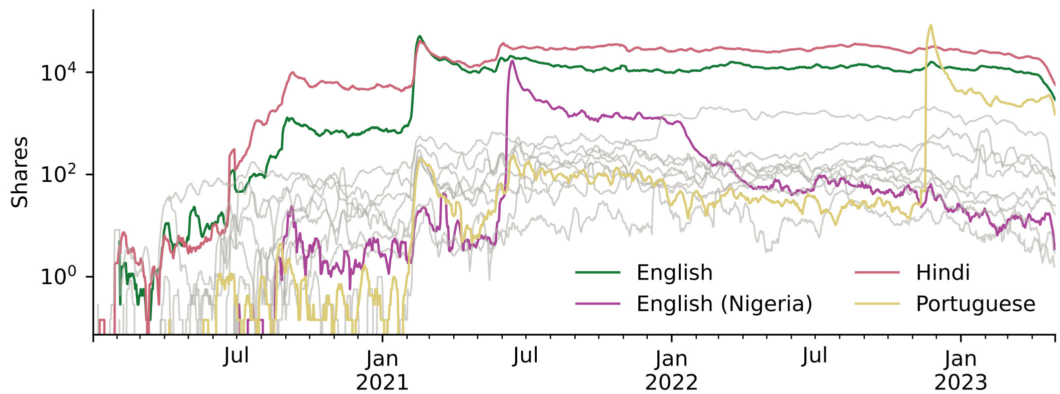


Figure B.3: **Sharing activity from linguistic communities.** Time-series showing the 7-day moving average of the number of shares made by each linguistic community. The major communities are highlighted in colour.

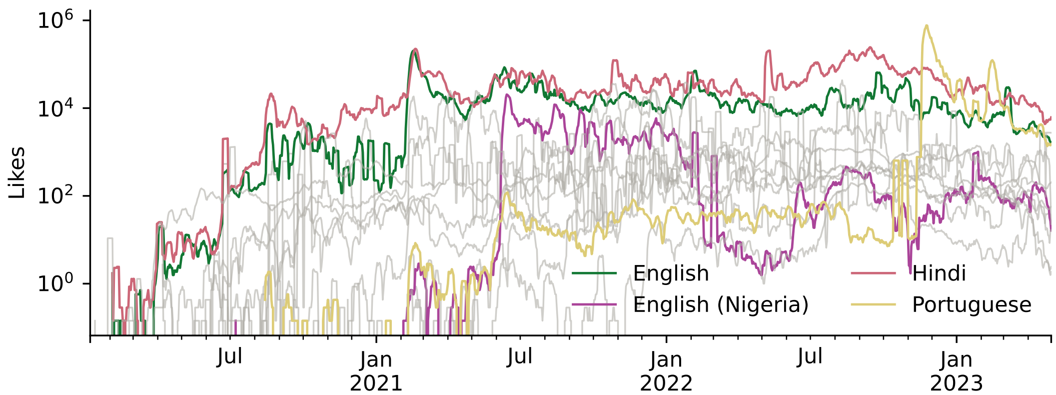


Figure B.4: **Liking activity from linguistic communities.** Time-series showing the 7-day moving average of the number of likes made by each linguistic community. The major communities are highlighted in colour.

B.0.2 Interaction network k-core analysis

Our analysis highlighted a strong dominance of English- and Hindi-speaking users on the k-core of the interaction network on Koo. Figure B.5 displays the distribution of users in the k-core of the network, for each linguistic community. The inset zooms on the highest values of k observed in the network, to further underline that only English and Hindi speakers are included in the highest cores.

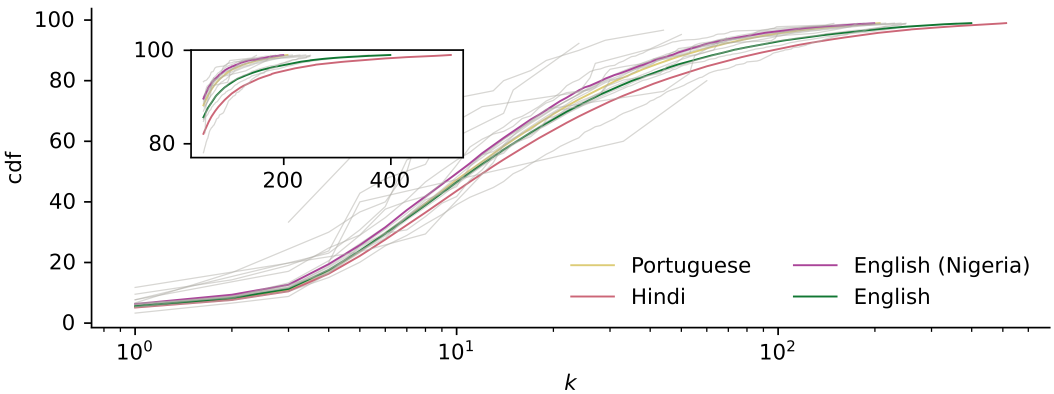


Figure B.5: **Linguistic composition of the k-core.** Cumulative distribution function of the percentage of users from a given linguistic community belonging to the k-core, for every value of k. Inset: Zoom on the distribution for $k \geq 50$.

Bibliography

[1] Alan B. Cantor. Projecting the standard error of the kaplan-meier estimator. *Statistics in Medicine*, 20(14):2091–2097, 2001.

[2] Google perspective api. https://perspectiveapi.com/ Accessed 3 February, 2023, 2023.

[3] Jan Rocha. Brazilians vs bolsonaro's 'package of destruction'. https://lab.org.uk/brazilians-vs-bolsonaros-package-of-destruction/ Accessed 25 July, 2023, 2022.

[4] Shlomo S Sawilowsky. New effect size rules of thumb. *Journal of modern applied statistical methods*, 8(2):26, 2009.

[5] Homero Gil de Zúñiga, Aaron Veenstra, Emily Vraga, and Dhavan Shah. Digital democracy: Reimagining pathways to political participation. *Journal of information technology & politics*, 7(1):36–51, 2010.

[6] Priyanjana Bengani, Mike Ananny, and Emily J Bell. Controlling the conversation: The ethics of social platforms and content moderation. *Columbia University Libraries*.

[7] Nina I Brown. Regulatory goldilocks: Finding the just and right fit for content moderation on social platforms. *Tex. A&M L. Rev.*, 8:451, 2020.

[8] Bryan Pfaffenberger. "if i want it, it's ok": Usenet and the (outer) limits of free speech. *The Information Society*, 12(4):365–386, 1996.

[9] Blake T. Bilstad. Obscenity and indecency on the usenet: The legal and political future of alt.sex.stories. *Journal of Computer-Mediated Communication*, 2(2):0–0, June 2006.

[10] Rasmus Kleis Nielsen. News media, search engines and social networking sites as varieties of online gatekeepers. In *Rethinking journalism again*, pages 93–108. Routledge, 2016.

[11] Jeffrey Taylor. Liability of usenet moderators for defamation published by others: Flinging the law of defamation into cyberspace. *Florida Law Review*, 1995.

[12] Joan Donovan, Becca Lewis, and Brian Friedberg. Parallel ports: Sociotechnical change from the alt-right to alt-tech. *transcript*, 2019.

[13] Emillie de Keulenaar, João C Magalhães, and Bharath Ganesh. Modulating moderation: A history of objectionability in twitter moderation practices. *Journal of Communication*, 73(3):273–287, 2023.

[14] Kyle Langvardt. Regulating online content moderation. *Geo. LJ*, 106:1353, 2017.

[15] Ryan Kor-Sins. The alt-right digital migration: A heterogeneous engineering approach to social media platform branding. *New Media & Society*, 25(9):2321–2338, 2023.

[16] Nicole Buckley and Joseph S. Schafer. "censorship-free" platforms: Evaluating content moderation policies and practices of alternative social media. *Vol 4: Special Issue 1: Media and the Far-Right*, (Vol 4, Issue 1), 2022.

[17] Heather J Williams, Alexandra T Evans, Jamie Ryan, Erik E Mueller, and Bryce Downing. The online extremist ecosystem. *RAND Corporation*, 2021.

[18] R.Y. Lazerson. The growing security threat from alternative platforms. https://csp.berkeley.edu/2023/05/16/the-growing-security-threat-from-alternative-platforms/ Accessed 1 May, 2024, 2023.

[19] Joseph Packer and Ethan Stoneman. Where we produce one, we produce all: The platform conspiracism of qanon. *Cultural Politics*, 17(3):255–278, 2021.

[20] Lorraine Bowman-Grieve. Exploring "stormfront": A virtual community of the radical right. *Studies in conflict & terrorism*, 32(11):989–1007, 2009.

[21] Manoel Horta Ribeiro, Jeremy Blackburn, Barry Bradlyn, Emiliano De Cristofaro, Gianluca Stringhini, Summer Long, Stephanie Greenberg, and Savvas Zannettou. The evolution of the manosphere across the web. *Proceedings of the International AAAI Conference on Web and Social Media*, 15(1):196–207, May 2021.

[22] Emily Turner-Graham. "breivik is my hero": The dystopian world of extreme right youth on the internet. *Australian Journal of Politics & History*, 60(3):416–430, 2014.

[23] Emily Blout and Patrick Burkart. White supremacist terrorism in charlottesville: Reconstructing 'unite the right'. *Studies in Conflict & Terrorism*, 46(9):1624–1652, 2023.

[24] Adi Robertson. 'free speech' reddit clone voat says it will shut down on christmas. https://www.theverge.com/2020/12/22/22195115/voat-free-speech-right-wing-reddit-clone-shutdown-investor Accessed 12 June, 2024, 2020.

[25] Idris Abubakar and Christopher Nilesh. Koo is selling itself as a twitter substitute in nigeria. https://restofworld.org/2021/koo-is-selling-itself-as-a-twitter-substitute-in-nigeria/ Accessed 15 August, 2023, 2021.

[26] Alex González Ormerod. Twitter drama has brazilians flocking to indian platform koo. https://restofworld.org/2022/twitter-brazil-koo/ Accessed 16 August, 2023, 2022.

[27] Kirill Bryanov, Dina Vasina, Yulia Pankova, and Victor Pakholkov. The other side of deplatforming: Right-wing telegram in the wake of trump's twitter ouster. In *Communications in Computer and Information Science*, pages 417–428. Springer International Publishing, 2022.

[28] Theo Wayt. Joe rogan helps twitter alternative gettr amass a million new users. https://nypost.com/2022/01/07/joe-rogan-helps-gettr-amass-a-million-new-users/ Accessed 30 April, 2024, 2022.

[29] Clare Malone. Can gettr become the online gathering place for trump's g.o.p.? https://www.newyorker.com/news/annals-of-communications/can-gettr-become-the-online-gathering-place-for-trumps-gop Accessed 15 October, 2022, 2022.

[30] Hatemail. Qoup d'état: What QAnon's Reddit Migration Tells Us About Misinformation. https://bit.ly/3eNa8Jq Accessed January 7, 2022, 2021.

[31] Alex Morris. Will marjorie taylor greene's twitter suspension matter? https://www.rollingstone.com/politics/politics-news/marjorie-taylor-greene-twitter-suspension-effective-1277806/ Accessed 30 April, 2024, 2022.

[32] Derek Lackaff and William J Moner. Local languages, global networks: Mobile design for minority language users. In *Proceedings of the 34th ACM International Conference on the Design of Communication*, pages 1–9, 2016.

[33] Zixue Tai. Casting the ubiquitous net of information control: Internet surveillance in china from golden shield to green dam. *International Journal of Advanced Pervasive and Ubiquitous Computing (IJAPUC)*, 2(1):53–70, 2010.

[34] Alexandra V Orlova. " digital sovereignty," anonymity and freedom of expression: Russia's fight to re-shape internet governance. *UC Davis J. Int'l L. & Pol'y*, 26:225, 2019.

[35] Theofilos Gkinopoulos, Stefano Pagliaro, Sofia Stathi, and Manuel Teresi. Leadership style, group orientation of conspiracy beliefs and uncertainty in the era of covid-19: Effects on moral leadership and leader identification and the moderating role of right-wing authoritarianism, Aug 2021.

[36] Craig Timberg and Elizabeth Dwoskin. Reddit closes long-running forum supporting president trump after years of policy violations. https://www.washingtonpost.com/technology/2020/06/29/reddit-closes-long-running-forum-supporting-president-trump-after-years-policy-violations/ Accessed 30 April, 2024, 2020.

[37] Robert Peck. The hate-fueled rise of r/the_donald—and its epic takedown. https:

//www.wired.com/story/the-hate-fueled-rise-of-rthe-donald-and-its-epic-takedown/ Accessed 30 April, 2024, 2020.

[38] Amin Mekacher, Max Falkenberg, and Andrea Baronchelli. How language, culture, and geography shape online dialogue: Insights from koo. *arXiv*, 2024.

[39] Gettr. Gettr committed to free speech fight in brazil after court rejects appeal. https://about.gettr.com/press/gettr-committed-to-free-speech-fight-in-brazil-after-court-rejects-appeal, Accessed 20 January, 2023, 2022.

[40] Anatoliy Gruzd, Jenna Jacobson, Barry Wellman, and Philip Mai. Understanding communities in an age of social media: the good, the bad, and the complicated. *Information, Communication & Society*, 19(9):1187–1193, 2016.

[41] Tim Finin, Anupam Joshi, Pranam Kolari, Akshay Java, Anubhav Kale, and Amit Karandikar. The information ecology of social media and online communities. *AI Magazine*, 29(3):77–77, 2008.

[42] Emad Khazraee and Alison N Novak. Digitally mediated protest: Social media affordances for collective identity construction. *Social Media+ Society*, 4(1):2056305118765740, 2018.

[43] Catherine M Ridings and David Gefen. Virtual community attraction: Why people hang out online. *Journal of Computer-mediated communication*, 10(1):JCMC10110, 2004.

[44] W Lance Bennett. The personalization of politics: Political identity, social media, and changing patterns of participation. *The annals of the American academy of political and social science*, 644(1):20–39, 2012.

[45] Manyu Li, Nadia Turki, Cassandra R Izaguirre, Chloe DeMahy, Brooklyn Labery Thibodeaux, and Taylor Gage. Twitter as a tool for social movement: An analysis of feminist activism on social media communities. *Journal of community psychology*, 49(3):854–868, 2021.

[46] Pam Shearing. Has social media revolutionized awareness of humanitarian abuses and the "responsibility to protect"? *St Antony's International Review*, 8(2):12–32, 2013.

[47] Luca Iandoli, Simonetta Primario, and Giuseppe Zollo. The impact of group polarization on the quality of online debate in social media: A systematic literature review. *Technological Forecasting and Social Change*, 170:120924, 2021.

[48] Matteo Cinelli, Gianmarco De Francisci Morales, Alessandro Galeazzi, Walter Quattrociocchi, and Michele Starnini. The echo chamber effect on social media. *Proceedings of the National Academy of Sciences*, 118(9):e2023301118, 2021.

[49] Lena Frischlich, Tim Schatto-Eckrodt, Svenja Boberg, and Florian Wintterlin. Roots of incivility: How personality, media use, and online experiences shape uncivil partic-

ipation. *Media and Communication*, 9(1):195–208, 2021.

[50] Almog Simchon, William J Brady, and Jay J Van Bavel. Troll and divide: the language of online polarization. *PNAS Nexus*, 1(1):pgac019, 03 2022.

[51] Nir Grinberg, Kenneth Joseph, Lisa Friedland, Briony Swire-Thompson, and David Lazer. Fake news on twitter during the 2016 u.s. presidential election. *Science*, 363(6425):374–378, 2019.

[52] Glenda Cooper. Populist rhetoric and media misinformation in the 2016 uk brexit referendum. In *The Routledge Companion to Media Disinformation and Populism*, pages 397–410. Routledge, 2021.

[53] Savvas Zannettou, Tristan Caulfield, Emiliano De Cristofaro, Michael Sirivianos, Gianluca Stringhini, and Jeremy Blackburn. Disinformation warfare: Understanding state-sponsored trolls on twitter and their influence on the web. In *Companion Proceedings of The 2019 World Wide Web Conference*, WWW '19, page 218–226, New York, NY, USA, 2019. Association for Computing Machinery.

[54] Aim Sinpeng, Ross Tapsell, et al. From grassroots activism to disinformation: Social media trends in southeast asia. *From Grassroots Activism to Disinformation*, pages 1–18, 2021.

[55] Chris J Vargo, Lei Guo, and Michelle A Amazeen. The agenda-setting power of fake news: A big data analysis of the online media landscape from 2014 to 2016. *New media & society*, 20(5):2028–2049, 2018.

[56] Muhsin Yesilada and Stephan Lewandowsky. A systematic review: The youtube recommender system and pathways to problematic content. *PsyArXiv*, 2021.

[57] Aylin Smellen. Investigating the effect of recommender system algorithms on the amplification of disinformation. *Radboud University*, 2022.

[58] Miriam Fernandez, Alejandro Bellogín, and Iván Cantador. Analysing the effect of recommendation algorithms on the spread of misinformation. In *Proceedings of the 16th ACM Web Science Conference*, pages 159–169, 2024.

[59] Mehdi Elahi, Danial Khosh Kholgh, Mohammad Sina Kiarostami, Sorush Saghari, Shiva Parsa Rad, and Marko Tkalčič. Investigating the impact of recommender systems on user-based and item-based popularity bias. *Information Processing & Management*, 58(5):102655, 2021.

[60] Masoud Mansoury, Himan Abdollahpouri, Mykola Pechenizkiy, Bamshad Mobasher, and Robin Burke. A graph-based approach for mitigating multi-sided exposure bias in recommender systems. *ACM Transactions on Information Systems (TOIS)*, 40(2):1–31, 2021.

[61] Tarleton Gillespie. *Custodians of the Internet: Platforms, content moderation, and the hidden decisions that shape social media*. Yale University Press, 2018.

[62] Enrique Armijo. Reasonableness as censorship: Section 230 reform, content moderation, and the first amendment. *Fla. L. Rev.*, 73:1199, 2021.

[63] Anastasia Kozyreva, Stefan M Herzog, Stephan Lewandowsky, Ralph Hertwig, Philipp Lorenz-Spreen, Mark Leiser, and Jason Reifler. Resolving content moderation dilemmas between free speech and harmful misinformation. *Proceedings of the National Academy of Sciences*, 120(7):e2210666120, 2023.

[64] Robert Gorwa, Reuben Binns, and Christian Katzenbach. Algorithmic content moderation: Technical and political challenges in the automation of platform governance. *Big Data & Society*, 7(1):2053951719897945, 2020.

[65] Tarleton Gillespie. Content moderation, ai, and the question of scale. *Big Data & Society*, 7(2):2053951720943234, 2020.

[66] Michael Randall Barnes. Online extremism, ai, and (human) content moderation. *Feminist Philosophy Quarterly*, 8(3/4), 2022.

[67] Daniel Link, Bernd Hellingrath, and Jie Ling. A human-is-the-loop approach for semi-automated content moderation. In *ISCRAM*, 2016.

[68] Greyson K Young. How much is too much: the difficulties of social media content moderation. *Information & Communications Technology Law*, 31(1):1–16, 2022.

[69] Sarah Myers West. Censored, suspended, shadowbanned: User interpretations of content moderation on social media platforms. *New Media & Society*, 20(11):4366–4383, 2018.

[70] Shagun Jhaver, Christian Boylston, Diyi Yang, and Amy Bruckman. Evaluating the effectiveness of deplatforming as a moderation strategy on twitter. *Proceedings of the ACM on Human-Computer Interaction*, 5(CSCW2):1–30, oct 2021.

[71] Michael Halpin, Norann Richard, Kayla Preston, Meghan Gosse, and Finlay Maguire. Men who hate women: The misogyny of involuntarily celibate men. *New Media & Society*, 0(0):14614448231176777, 0.

[72] H. Innes and M. Innes. De-platforming disinformation: conspiracy theories and their control. *Information, Communication & Society*, 26(6):1262–1280, October 2021.

[73] Daniel Robert Thomas and Laila A. Wahedi. Disrupting hate: The effect of deplatforming hate organizations on their online audience. *Proceedings of the National Academy of Sciences*, 120(24):e2214080120, 2023.

[74] Stefan D McCabe, Diogo Ferrari, Jon Green, David MJ Lazer, and Kevin M Esterling. Post-january 6th deplatforming reduced the reach of misinformation on twitter. *Nature*,

630(8015):132–140, 2024.

[75] Angela Nagle. *Kill all normies: Online culture wars from 4chan and Tumblr to Trump and the alt-right*. John Hunt Publishing, 2017.

[76] Gordon A Gow. Turning to alternative social media. *Social Media Research Methods*, page 568, 2021.

[77] Gabriel Fair and Ryan Wesslen. Shouting into the void: A database of the alternative social media platform gab. *Proceedings of the International AAAI Conference on Web and Social Media*, 13:608–610, July 2019.

[78] Ehsan Dehghan and Ashwin Nagappa. Politicization and radicalization of discourses in the alt-tech ecosystem: A case study on gab social. *Social Media + Society*, 8(3):20563051221113075, 2022.

[79] Luke Munn. More than a mob: Parler as preparatory media for the u.s. capitol storming. *First Monday*, February 2021.

[80] Max Aliapoulios, Emmi Bevensee, Jeremy Blackburn, Barry Bradlyn, Emiliano De Cristofaro, Gianluca Stringhini, and Savvas Zannettou. A large open dataset from the parler social network. *Proceedings of the International AAAI Conference on Web and Social Media*, 15:943–951, May 2021.

[81] Patrick Gerard, Nicholas Botzer, and Tim Weninger. Truth social dataset. *Proceedings of the International AAAI Conference on Web and Social Media*, 17:1034–1040, June 2023.

[82] Bobby Allyn. Google is now distributing truth social, trump's twitter alternative. https://www.npr.org/2022/10/14/1129024148/google-truth-social-trump Accessed 30 October, 2023, 2022.

[83] Andrea Sipka, Aniko Hannak, and Aleksandra Urman. Comparing the language of qanon-related content on parler, gab, and twitter. In *Proceedings of the 14th ACM Web Science Conference 2022*, WebSci '22, pages 411–421, New York, NY, USA, 2022. Association for Computing Machinery.

[84] Avinash Prabhu, Dipanwita Guhathakurta, Mallika Subramanian, Manvith Reddy, Shradha Sehgal, Tanvi Karandikar, Amogh Gulati, Udit Arora, Rajiv Ratn Shah, Ponnurangam Kumaraguru, et al. Capitol (pat) riots: A comparative study of twitter and parler. *arXiv preprint arXiv:2101.06914*, 2021.

[85] Milo Z. Trujillo, Maurício Gruppi, Cody Buntain, and Benjamin D. Horne. The MeLa BitChute dataset. *Proceedings of the International AAAI Conference on Web and Social Media*, 16:1342–1351, May 2022.

[86] Mark Sweney. Bitchute to continue hosting kremlin-backed rt channel. https://www.theguardian.com/business/2022/mar/09/bitchute-website-continue-hosting

-blocked-rt-channel Accessed 24 October, 2023, 2022.

[87] Mario Peucker and Thomas J Fisher. Mainstream media use for far-right mobilisation on the alt-tech online platform gab. *Media, Culture & Society*, 45(2):354–372, 2023.

[88] Melissa-Ellen Dowling. Far-right populism in alt-tech: A challenge for democracy? *new media & society*, 2023.

[89] Amin Mekacher, Max Falkenberg, and Andrea Baronchelli. The systemic impact of deplatforming on social media. *PNAS Nexus*, 2(11), 2023.

[90] Lynnette Hui Xian Ng, Iain Cruickshank, and Kathleen M Carley. Coordinating narratives and the capitol riots on parler. *arXiv preprint arXiv:2109.00945*, 2021.

[91] David Gilbert. Crowdfunding site patreon is purging far-right figures. https://www.vice.com/en/article/qvqeev/crowdfunding-site-patreon-is-purging-far-right-figures Accessed 25 June, 2024, 2018.

[92] Tom Keatinge, Florence Keen, and Kayla Izenman. Fundraising for right-wing extremist movements: How they raise funds and how to counter it. *The RUSI Journal*, 164(2):10–23, February 2019.

[93] Sophia Savva. Jordan peterson's federal funding denied, rebel media picks up the tab. https://thevarsity.ca/2017/05/01/jordan-petersons-federal-funding-denied-rebel-media-picks-up-the-tab/ Accessed 29 April, 2024, 2017.

[94] Matt Pearce. Neo-nazi website raises $150,000 to fight southern poverty law center lawsuit. https://www.latimes.com/nation/la-na-daily-stormer-20170602-story.html Accessed 30 April, 2024, 2017.

[95] Rachel E Moran, Anna L Swan, and Taylor Agajanian. Vaccine misinformation for profit: Conspiratorial wellness influencers and the monetization of alternative health. *International Journal of Communication*, 18:23, 2024.

[96] Dayane Fumiyo Tokojima Machado, Alexandre Fioravante de Siqueira, Natiely Rallo Shimizu, and Leda Gitahy. It-who-must-not-benamed: Covid-19 misinformation, tactics to profit from it and to evade content moderation on youtube. *Frontiers in Communication*, 7(1037432):1–14, 2021.

[97] Joshua A. Braun and Jessica L. Eklund. Fake news, real money: Ad tech platforms, profit-driven hoaxes, and the business of journalism. *Digital Journalism*, 7(1):1–21, January 2019.

[98] Alice Marwick, Benjamin Clancy, and Katherine Furl. Far-right online radicalization: A review of the literature. *The Bulletin of Technology & Public Life*, 2022.

[99] Emily Booth, Jooyoung Lee, Marian-Andrei Rizoiu, and Hany Farid. Conspiracy, misinformation, radicalisation: understanding the online pathway to indoctrination

and opportunities for intervention. *Journal of Sociology*, 2024.

[100] Loo Seng Neo. An internet-mediated pathway for online radicalisation. In *Combating Violent Extremism and Radicalization in the Digital Era*, pages 197–224. IGI Global, 2016.

[101] Nathalie Van Raemdonck. The echo chamber of anti-vaccination conspiracies: Mechanisms of radicalization on facebook and reddit. *Institute for Policy, Advocacy and Governance (IPAG) Knowledge Series, Forthcoming*, 2019.

[102] Arie W. Kruglanski, Michele J. Gelfand, Jocelyn J. Bélanger, Anna Sheveland, Malkanthi Hetiarachchi, and Rohan Gunaratna. The psychology of radicalization and deradicalization: How significance quest impacts violent extremism. *Political Psychology*, 35(S1):69–93, January 2014.

[103] Emily L. Wang, Luca Luceri, Francesco Pierri, and Emilio Ferrara. Identifying and characterizing behavioral classes of radicalization within the qanon conspiracy on twitter. *Proceedings of the International AAAI Conference on Web and Social Media*, 17(1):890–901, Jun. 2023.

[104] Justin E. Lane, Kevin McCaffree, and F. LeRon Shults. Is radicalization reinforced by social media censorship? *arXiv preprint arXiv:2103.12842*, 2021.

[105] Sue Curry Jansen and Brian Martin. The streisand effect and censorship backfire. *International Journal of Communication*, 2015.

[106] Mattia Samory and Tanushree Mitra. Conspiracies online: User discussions in a conspiracy community following dramatic events. *Proceedings of the International AAAI Conference on Web and Social Media*, 12(1), June 2018.

[107] Ahmed Ajil. *"Rebellious underdogs fighting for justice": A theory of politico-ideological mobilisation and violence*. PhD thesis, 06 2021.

[108] Jamie Bartlett and Carl Miller. the power of unreason conspiracy theories, extremism and counter-terrorism. *Demos*, 2010.

[109] Miriam Fernandez, Antonio Gonzalez-Pardo, and Harith Alani. Radicalisation influence in social media. *Journal of Web Science*, 6, 2019.

[110] Ghayda Hassan, Sébastien Brouillette-Alarie, Séraphin Alava, Divina Frau-Meigs, Lysiane Lavoie, Arber Fetiu, Wynnpaul Varela, Evgueni Borokhovski, Vivek Venkatesh, Cécile Rousseau, et al. Exposure to extremist online content could lead to violent radicalization: A systematic review of empirical evidence. *International journal of developmental science*, 12(1-2):71–88, 2018.

[111] Mehmet F Bastug, Aziz Douai, and Davut Akca. Exploring the "demand side" of online radicalization: Evidence from the canadian context. *Studies in Conflict & Terrorism*, 43(7):616–637, 2020.

[112] Derek O'Callaghan, Derek Greene, Maura Conway, Joe Carthy, and Pádraig Cunningham. An analysis of interactions within and between extreme right communities in social media. In *Ubiquitous Social Media Analysis: Third International Workshops, MUSE 2012, Bristol, UK, September 24, 2012, and MSM 2012, Milwaukee, WI, USA, June 25, 2012, Revised Selected Papers*, pages 88–107. Springer, 2013.

[113] Imran Awan. Cyber-extremism: Isis and the power of social media. *Society*, 54(2):138–149, 2017.

[114] Carsten Schwemmer. Social media strategies of right-wing movements-the radicalization of pegida. *Computational Methods for the Social Sciences: Applications to the Study of Ethnic Minorities*, page 141, 2018.

[115] Veronika Möller and Antonia Mischler. The soundtrack of the extreme: Nasheeds and right-wing extremist music as a "gateway drug" into the radical scene? *International Annals of Criminology*, 58(2):291–334, 2020.

[116] Cynthia Miller-Idriss. Hate in the homeland: The new global far right. *Princeton University Press*, 2022.

[117] Manoel Horta Ribeiro, Raphael Ottoni, Robert West, Virgílio AF Almeida, and Wagner Meira Jr. Auditing radicalization pathways on youtube. In *Proceedings of the 2020 conference on fairness, accountability, and transparency*, pages 131–141, 2020.

[118] Muhammad Haroon, Anshuman Chhabra, Xin Liu, Prasant Mohapatra, Zubair Shafiq, and Magdalena Wojcieszak. Youtube, the great radicalizer? auditing and mitigating ideological biases in youtube recommendations. *arXiv preprint arXiv:2203.10666*, 2022.

[119] Olivia Bamsey and Reza Montasari. The role of the internet in radicalisation to violent extremism. In *Digital Transformation in Policing: The Promise, Perils and Solutions*, pages 119–135. Springer, 2023.

[120] Oxford Analytica. 'alt-tech'will diffuse online extremist presence. *Emerald Expert Briefings*, (oxan-db), 2020.

[121] William Birdsall. Web 2.0 as a social movement. *Webology*, 4, 06 2007.

[122] Brett Rolfe. The impact of web 2.0 on social activism. In *7th Annual Conference of the Association of Internet Researchers*, 2006.

[123] Camilla MacTier. Who online cares?: Web 2.0, social capital and the self-responsibilisation of environmental impact. *Communication, Politics & Culture*, 41(1):99–113, 2008.

[124] Adam Klein. Slipping racism into the mainstream: A theory of information laundering. *Communication Theory*, 22(4):427–448, 2012.

[125] Daniela Peterka-Benton, Bond Benton, and Joel Penney. From conspiracy to normalcy: The mainstreaming of qanon in "disney grooming" messages online. *Crime, Media, Culture*, 2024.

[126] Marcus A. Brooks. It's okay to be white: laundering white supremacy through a colorblind victimized white race-consciousness raising campaign. *Sociological Spectrum*, 40(6):400–416, 2020.

[127] Jessie Daniels. The algorithmic rise of the "alt-right". *Contexts*, 17(1):60–65, 2018.

[128] Justin Ward. Day of the trope: White nationalist memes thrive on reddit's r/the_donald. https://www.splcenter.org/hatewatch/2018/04/19/day-trope-white-nationalist-memes-thrive-reddits-rthedonald Accessed 3 May, 2024, 2018.

[129] Colin Klein, Peter Clutton, and Adam G. Dunn. Pathways to conspiracy: The social and linguistic precursors of involvement in reddit's conspiracy theory forum. *PLOS ONE*, 14(11):e0225098, November 2019.

[130] Amin Mekacher and Antonis Papasavva. "i can't keep it up." a dataset from the defunct voat.co news aggregator. *Proceedings of the International AAAI Conference on Web and Social Media*, 16(1):1302–1311, May 2022.

[131] Amin Mekacher. Preparing for the boogaloo: How far-right communities rallied on discord for the unite the right rally. https://gnet-research.org/2024/03/25/preparing-for-the-boogaloo-how-far-right-communities-rallied-on-discord-for-the-unite-the-right-rally/ Accessed 3 May, 2024, 2024.

[132] Tom Wilson and Kate Starbird. Cross-platform information operations: Mobilizing narratives & building resilience through both 'big' & 'alt' tech. *Proc. ACM Hum.-Comput. Interact.*, 5(CSCW2), oct 2021.

[133] Eugenia Siapera. Alt tech and the public sphere: Exploring bitchute as a political media infrastructure. *European Journal of Communication*, 38(5):446–465, 2023.

[134] Ali Swenson. Rnc's livestreaming partner for the gop debate is a haven for disinformation and extremism. https://apnews.com/article/republican-debate-livestream-rumble-disinformation-extremism-a6e627ac88463f9f83ada062ea83c6db Accessed 13 June, 2024, 2023.

[135] Yini Zhang, Josephine Lukito, Jiyoun Suk, and Ryan McGrady. Trump, twitter, and truth social: how trump used both mainstream and alt-tech social media to drive news media attention. *Journal of Information Technology & Politics*, pages 1–14, 2024.

[136] Mason Youngblood. Extremist ideology as a complex contagion: the spread of far-right radicalization in the united states between 2005 and 2017. *Humanities and Social Sciences Communications*, 7(1):1–10, 2020.

[137] David Cunningham. Extremism's mainstream appeal: Slate's slow burn: David duke. *The American Historical Review*, 127(1):420–422, 2022.

[138] Southern Law Poverty Center. Knights of the ku klux klan. https://www.splcenter.org/fighting-hate/extremist-files/group/knights-ku-klux-klan Accessed 13 June, 2024, 2016.

[139] Glenn Beck. *The Overton Window*. Simon and Schuster, 2010.

[140] Marco Bastos and Raquel Recuero. The insurrectionist playbook: Jair bolsonaro and the national congress of brazil. *Social Media + Society*, 9(4):20563051231211881, 2023.

[141] Amin Mekacher. The koo dataset: An indian microblogging platform with global ambitions. *Zenodo*, May 2024.

[142] Sahana Udupa, Antonis Maronikolakis, and Axel Wisiorek. Ethical scaling for content moderation: Extreme speech and the (in)significance of artificial intelligence. *Big Data & Society*, 10(1):20539517231172424, 2023.

[143] Shruti Menon. Koo v twitter: Why india's government is favouring a social media newcomer. https://www.bbc.co.uk/news/world-asia-india-56037901 Accessed 30 April, 2024, 2021.

[144] Varsha Bansal. India's government wants total control of the internet. https://www.wired.com/story/indias-government-wants-total-control-of-the-internet/ Accessed 30 April, 2024, 2023.

[145] Jyotsna Singh and Benjamin Parker. 'curly tales' and 'beerbiceps': Narendra modi's bjp taps influencers ahead of india election. https://www.ft.com/content/dc383dfa-6d94-4a21-b2b7-0bb49daedfde Accessed 29 April, 2024, 2024.

[146] Aria Thaker. Indian politicians are now flocking to an unlikely "no english" social network. https://qz.com/india/1414241/sorry-facebook-indias-bjp-and-congress-flock-to-sharechat Accessed 29 April, 2024, 2018.

[147] Punyajoy Saha, Binny Mathew, Kiran Garimella, and Animesh Mukherjee. "short is the road that leads from fear to hate": Fear speech in indian whatsapp groups. WWW '21, New York, NY, USA, 2021. Association for Computing Machinery.

[148] Simon Chauchard and Kiran Garimella. What circulates on partisan WhatsApp in india? insights from an unusual dataset. *Journal of Quantitative Description: Digital Media*, 2, February 2022.

[149] Chloe Cornish. India's homegrown 'twitter' courts political parties to shed nationalist tag. https://www.ft.com/content/bd296702-18ad-4303-8192-bbb1268e3b56 Accessed 14 September, 2023, 2022.

[150] Amin Mekacher, Max Falkenberg, and Andrea Baronchelli. The koo dataset: An indian microblogging platform with global ambitions. *Proceedings of the International AAAI Conference on Web and Social Media*, 18(1):1991–2002, May 2024.

[151] Esteban Ortiz-Ospina. The rise of social media. https://ourworldindata.org/rise-of-social-media Accessed 27 June, 2024, 2019.

[152] Meltem Odabaş). 5 facts about twitter "lurkers'. https://www.pewresearch.org/short-reads/2022/03/16/5-facts-about-twitter-lurkers/ Accessed 14 May, 2024, 2022.

[153] Jakob Nielsen. The 90-9-1 rule for participation inequality in social media and online communities. https://www.nngroup.com/articles/participation-inequality/ Accessed 14 May, 2024, 2006.

[154] Rocío Galarza Molina. From lurkers to listeners: Introducing the concept of online listening into political communication studies. *Global Media Journal México*, 14(27), December 2017.

[155] Christina Neumayer and Judith Schoßböck. Political lurkers? In *Conference for e-democracy and open government*, page 131, 2011.

[156] Andrea Tagarelli and Roberto Interdonato. *Mining lurkers in online social networks: principles, models, and computational methods*. Springer, 2018.

[157] Andrea Tagarelli and Roberto Interdonato. "who's out there?": identifying and ranking lurkers in social networks. In *Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, ASONAM '13. ACM, August 2013.

[158] Benjamin Lee and Kim Knott. Fascist aspirants: Fascist forge and ideological learning in the extreme-right online milieu. *Behavioral Sciences of Terrorism and Political Aggression*, 14(3):216–240, January 2021.

[159] Tiana Gaudette Ryan Scrivens, Garth Davies and Richard Frank. Comparing online posting typologies among violent and nonviolent right-wing extremists. *Studies in Conflict & Terrorism*, 0(0):1–23, 2022.

[160] Adrienne Massanari. gamergate and the fappening: How reddit's algorithm, governance, and culture support toxic technocultures. *New Media & Society*, 19(3):329–346, July 2016.

[161] Mai ElSherief, Shirin Nilizadeh, Dana Nguyen, Giovanni Vigna, and Elizabeth Belding. Peer to peer hate: Hate speech instigators and their targets. *Proceedings of the International AAAI Conference on Web and Social Media*, 12(1), June 2018.

[162] Matthew Wade, Stephanie A Baker, and Michael J Walsh. Crowdfunding platforms as conduits for ideological struggle and extremism: On the need for greater regulation

and digital constitutionalism. *Policy & Internet*, 2023.

[163] Rosalie Gillett and Nicolas Suzor. Incels on reddit: A study in social norms and decentralised moderation. *First Monday*, 27(6), 2022.

[164] Hind Almerekhi, Supervised by Bernard J Jansen, and co-supervised by Haewoon Kwak. Investigating toxicity across multiple reddit communities, users, and moderators. In *WWW*, pages 294–298, 2020.

[165] Eshwar Chandrasekharan, Umashanthi Pavalanathan, Anirudh Srinivasan, Adam Glynn, Jacob Eisenstein, and Eric Gilbert. You can't stay here: The efficacy of reddit's 2015 ban examined through hate speech. *Proceedings of the ACM on human-computer interaction*, 1(CSCW):1–22, 2017.

[166] Amaury Trujillo and Stefano Cresci. Make reddit great again: assessing community effects of moderation interventions on r/the_donald. *Proceedings of the ACM on Human-computer Interaction*, 6(CSCW2):1–28, 2022.

[167] Adi Robertson. Welcome to Voat: Reddit killer, troll haven, and the strange face of internet free speech. https://www.theverge.com/2015/7/10/8924415/voat-reddit-competitor-free-speech Accessed 6 November, 2023, 2015.

[168] Antonis Papasavva, Jeremy Blackburn, Gianluca Stringhini, Savvas Zannettou, and Emiliano De Cristofaro. "Is it a Qoincidence?": An Exploratory Study of QAnon on Voat. In *WWW*, 2021.

[169] Antonis Papasavva, Max Aliapoulios, Cameron Ballard, Emiliano De Cristofaro, Gianluca Stringhini, Savvas Zannettou, and Jeremy Blackburn. The gospel according to q: Understanding the qanon conspiracy from the perspective of canonical information. In *ICWSM*, 2022.

[170] Eshwar Chandrasekharan, Shagun Jhaver, Amy Bruckman, and Eric Gilbert. Quarantined! examining the effects of a community-wide moderation intervention on reddit. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 29(4):1–26, 2022.

[171] Giuseppe Russo, Luca Verginer, Manoel Horta Ribeiro, and Giona Casiraghi. Spillover of antisocial behavior from fringe platforms: The unintended consequences of community banning. *arXiv*, 2022.

[172] C. Thi Nguyen. Echo chambers and epistemic bubbles. *Episteme*, 17(2):141–161, September 2018.

[173] James N Druckman, Samara Klar, Yanna Krupnikov, Matthew Levendusky, and John Barry Ryan. How affective polarization shapes americans' political beliefs: A study of response to the covid-19 pandemic. *Journal of Experimental Political Science*, 8(3):223–234, 2021.

[174] PGL Bewermeier. Gunning for office: Right-wing populism and gender in the congressional campaigns of lauren boebert and marjorie taylor greene. 2021.

[175] Mikhaila N Calice, Luye Bao, Isabelle Freiling, Emily Howell, Michael A Xenos, Shiyu Yang, Dominique Brossard, Todd P Newman, and Dietram A Scheufele. Polarized platforms? how partisanship shapes perceptions of "algorithmic news bias". *new media & society*, 25(11):2833–2854, 2023.

[176] Cristiano Lima. Gettr, parler, parler find a fanbase with brazil's far-right. https://www.washingtonpost.com/politics/2021/11/09/gettr-parler-parler-find-fanbase-with-brazils-far-right/ Accessed 13 January, 2023, 2021.

[177] Caleb Ecarma. Joe rogan got himself a twitter-ban insurance policy. https://www.vanityfair.com/news/2022/01/01/joe-rogan-twitter-ban-insurance Accessed 26 April, 2024, 2022.

[178] Damien Leloup. L'opération séduction du réseau social gettr envers l'extrême droite française. https://www.lemonde.fr/pixels/article/2022/03/28/l-operation-seduction-du-reseau-social-gettr-envers-l-extreme-droite-francaise_6119543_4408996.html Accessed 13 February, 2023, 2022.

[179] Pujan Paudel, Jeremy Blackburn, Emiliano De Cristofaro, Savvas Zannettou, and Gianluca Stringhini. A longitudinal study of the gettr social network. In *Workshop Proceedings of the 16th International AAAI Conference on Web and Social Media. Retrieved from https://doi. org/10.36190*, 2022.

[180] Alice Maciel. How the trump universe is backing bolsonaro's reelection bid in brazil. https://worldcrunch.com/world-affairs/brazil-bolsonaro-trump-gettr, Accessed 20 January, 2023, 2022.

[181] Mike Wendling. How trump's allies stoked brazil congress attack. https://www.bbc.co.uk/news/world-us-canada-64206484, Accessed 26 June, 2023, 2023.

[182] Meredith Deliso. Did the jan. 6 attack lay the blueprint for brazil's insurrection? https://abcnews.go.com/International/jan-6-attack-lay-blueprint-brazils-insurrection/story?id=96312273 Accessed 26 April, 2024, 2023.

[183] Dan Friedman. With its top investor in jail, a pro-trump social media site suffers mass layoffs. https://www.motherjones.com/politics/2024/02/gettr-guo-layoffs/ Accessed 26 April, 2024, 2024.

[184] Rishi Iyengar. Whatsapp has been linked to lynchings in india. facebook is trying to contain the crisis. https://edition.cnn.com/2018/09/30/tech/facebook-whatsapp-india-misinformation/index.html Accessed 26 April, 2024, 2018.

[185] Christina Fink. Dangerous speech, anti-muslim violence, and facebook in myanmar. *Journal of International Affairs*, 71(1.5):43–52, 2018.

[186] Gabriel Nicholas and Aliya Bhatia. Lost in translation: Large language models in non-english content analysis. *arXiv*, 2023.

[187] PIB Delhi. Mygov announces winners of 'aatmanirbhar bharat app innovation challenge'; encourages the spirit of self-reliance & innovation. https://pib.gov.in/PressReleasePage.aspx?PRID=1644229 Accessed 30 October, 2023, 2020.

[188] Anam Ajmal. 'we have over 4.7 million users on koo ... our vision was to build something catering to hundreds of languages'. https://timesofindia.indiatimes.com/blogs/toi-edit-page/we-have-over-4-7-million-users-on-koo-our-vision-was-to-build-something-catering-to-hundreds-of-languages/ Accessed 29 April, 2024, 2021.

[189] Indian Television. #kookiyakya asks koo in its first-ever tvc. https://www.indiantelevision.com/mam/marketing/mam/kookiyakya-asks-koo-in-its-first-ever-tvc-211025 Accessed 4 December, 2023, 2021.

[190] Christopher Nilesh. How koo became india's hindu nationalist–approved twitter alternative. https://restofworld.org/2021/how-koo-became-a-right-wing-darling-in-india/ Accessed 15 August, 2023, 2021.

[191] Vittoria Elliott. Hate speech proliferates on youtube in india, research finds. hhttps://www.wired.com/story/youtube-hate-speech-india-elections/ Accessed 11 September, 2024, 2024.

[192] Danny D'Cruze. Koo, india's twitter alternative, lays off 30 https://www.businesstoday.in/technology/news/story/koo-indias-twitter-alternative-lays-off-30-of-workforce-report-378155-2023-04-20 Accessed 29 April, 2024, 2023.

[193] StartupStory. Koo social media halts employee salaries amid funding crisis. https://startupstorymedia.com/insights-koo-social-media-halts-employee-salaries-amid-funding-crisis/ Accessed 29 April, 2024, 2024.

[194] Zoya Mateen. India's x alternative koo to shut down services. https://www.bbc.co.uk/news/articles/ced3144v98do Accessed 11 September, 2024, 2024.

[195] Justin Caffier. Here Are Reddit's Whiniest, Most Low-Key Toxic Subreddits. https://bit.ly/3LxQOy6 Accessed March 1, 2022, 2017.

[196] Billy Perrigo. Twitter Offers More Transparency on Racist Abuse by Its Users, but Few Solutions. https://bit.ly/3LDZMtJ Accessed June 2, 2022, 2021.

[197] Karen Hao. How Facebook and Google fund global misinformation. https://bit.ly/3r4WNTk Accessed March 1, 2022, 2021.

[198] James Vincent. Reddit reports 18 percent reduction in hateful content after banning nearly 7,000 subreddits. https://www.theverge.com/2020/8/20/21376957/reddit

-hate-speech-content-policies-subreddit-bans-reduction Accessed January 12, 2022, 2020.

[199] Fernando Alfonso. Reddit bans infamous forum about beating women. https://bit.ly/3LGWaqR Accessed June 1, 2022, 2014.

[200] 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. In *CSCW*, 2021.

[201] Abby Ohlheiser. Fearing yet another witch hunt, Reddit bans 'Pizzagate'. https://wapo.st/3FrdXij Accessed January 1, 2022, 2016.

[202] Jay Hathaway. Why Reddit finally banned one of its most misogynistic forums. https://www.dailydot.com/unclick/reddit-incels-ban/ Accessed 14 January, 2022, 2017.

[203] Adi Robertson. Reddit has banned the QAnon conspiracy subreddit r/GreatAwakening. https://fxn.ws/3IiXYor Accessed January 1, 2022, 2018.

[204] Eshwar Chandrasekharan, Mattia Samory, Anirudh Srinivasan, and Eric Gilbert. The bag of communities: Identifying abusive behavior online with preexisting internet data. In *ACM SIGCHI*, 2017.

[205] H Saleem, K Dillon, S Benesch, and D Ruths. A web of hate: Tackling hateful speech in online social spaces. *TACOS*, 2017.

[206] Zenodo. Dataset: "I Can't Keep It Up." A Dataset from the Defunct Voat.co News Aggregator. https://zenodo.org/record/5841668 Accessed January 1, 2022, 2021.

[207] Antonis Papasavva, Savvas Zannettou, Emiliano De Cristofaro, Gianluca Stringhini, and Jeremy Blackburn. Raiders of the lost kek: 3.5 years of augmented 4chan posts from the politically incorrect board. In *ICWSM*, 2020.

[208] Abigail Tracy. Web host drops voat after disgruntled redditors flock to the platform. https://bit.ly/32KhieP Accessed January 4, 2022, 2015.

[209] Paul Sawers. Amid censorship brouhaha, reddit clone voat has its servers closed by hosting provider. https://bit.ly/3sSXO2C Accessed January 2, 2022, 2015.

[210] Rachel Pick. PayPal cuts off reddit clone voat over obscenity. https://bit.ly/32PD4xz Accessed January 11, 2022, 2015.

[211] Caitlin Dewey. This is what happens when you create an online community without any rules, part 2. https://wapo.st/3qH9aUn Accessed January 3, 2022, 2015.

[212] Therese Poletti. Creator of surging Reddit rival Voat: We will avoid same mistakes. https://on.mktw.net/3sVWc81 Accessed January 8, 2022, 2015.

[213] Jeff Roberts. New reddit rival voat hit by ddos attack. https://bit.ly/3J8kJv9 Accessed January 10, 2022, 2015.

[214] Sarah Emerson. Founder of voat, the 'censorship-free' reddit, begs users to stop making death threats. https://bit.ly/3mT77vr Accessed January 5, 2022, 2019.

[215] Caitlin Dewey. The 'Reddit exodus' is a perfect illustration of the state of free speech on the Web. https://wapo.st/33zyZOc Accessed January 3, 2022, 2015.

[216] Archive Team. Archive Team: Voat. https://archive.org/details/archiveteam_voat Accessed January 1, 2022, 2020.

[217] Digital Preservation. Sustainability of Digital Formats: Planning for Library of Congress Collections. https://www.loc.gov/preservation/digital/formats/fdd/fdd000236.shtml Accessed January 1, 2022, 2020.

[218] Caitlin M Rivers and Bryan L Lewis. Ethical research standards in a world of big data. *F1000Research*, 2014.

[219] Andrew Griffin. Reddit alternative breaks because so many people leave site after harassment scandal. https://www.independent.co.uk/life-style/gadgets-and-tech/news/reddit-alternative-breaks-because-so-many-people-leave-site-after-harassment-scandal-10321474.html Accessed January 13, 2022, 2015.

[220] Abby Ohlheiser. Reddit bans r/greatawakening, the main subreddit for qanon conspiracy theorists. https://wapo.st/3GsSHtA Accessed January 7, 2022, 2018.

[221] Jon Kleinberg. Bursty and hierarchical structure in streams. *Data Mining and Knowledge Discovery*, 7(4):373–397, 2003.

[222] Adi Robertson. Was reddit always about free speech? yes, and no. https://bit.ly/3rjEh8K Accessed February 1, 2022, 2015.

[223] Savvas Zannettou, Jeremy Blackburn, Emiliano De Cristofaro, Michael Sirivianos, and Gianluca Stringhini. Understanding web archiving services and their (mis) use on social media. In *ICWSM*, 2018.

[224] Edward Newell, David Jurgens, Haji Mohammad Saleem, Hardik Vala, Jad Sassine, Caitrin Armstrong, and Derek Ruths. User migration in online social networks: A case study on reddit during a period of community unrest. In *ICWSM*, 2016.

[225] Milena Popova. Reading out of context: pornographic deepfakes, celebrity and intimacy. *Porn Studies*, 2019.

[226] Adi Robertson. Reddit Bans 'deepfakes' AI Porn Communities. https://bit.ly/35hS1Wy, 2018.

[227] Osama Khalid and Padmini Srinivasan. Style Matters! Investigating Linguistic Style in Online Communities. In *ICWSM*, 2020.

[228] Jason Baumgartner, Savvas Zannettou, Brian Keegan, Megan Squire, and Jeremy Blackburn. The pushshift reddit dataset. In *ICWSM*, 2020.

[229] Jason Baumgartner, Savvas Zannettou, Megan Squire, and Jeremy Blackburn. The pushshift telegram dataset. In *ICWSM*, 2020.

[230] Max Aliapoulios, Emmi Bevensee, Jeremy Blackburn, Barry Bradlyn, Emiliano De Cristofaro, Gianluca Stringhini, and Savvas Zannettou. An early look at the parler online social network. In *ICWSM*, 2021.

[231] Adrian Rauchfleisch and Jonas Kaiser. Deplatforming the far-right: An analysis of YouTube and BitChute. *SSRN Electronic Journal*, 2021.

[232] Charlie Winter, Peter Neumann, Alexander Meleagrou-Hitchens, Magnus Ranstorp, Lorenzo Vidino, and Johanna Fürst. Online extremism: Research trends in internet activism, radicalization, and counter-strategies. *International Journal of Conflict and Violence (IJCV)*, page Vol 14 No 2 (2020), 2020.

[233] Alexandre Bovet and Peter Grindrod. Organization and evolution of the UK far-right network on telegram. *Applied Network Science*, 7(1), November 2022.

[234] Zeynep Tufekci. How social media took us from tahrir square to donald trump. https://www.technologyreview.com/2018/08/14/240325/how-social-media-took-us-from-tahrir-square-to-donald-trump/ Accessed 26 January, 2023, 2018.

[235] Matteo Cinelli, Andraž Pelicon, Igor Mozetič, Walter Quattrociocchi, Petra Kralj Novak, and Fabiana Zollo. Dynamics of online hate and misinformation. *Scientific Reports*, 11(1), November 2021.

[236] Mahmoud Naffakh. How pro-terrorism accounts are circumventing moderation on social media. https://observers.france24.com/en/middle-east/20221125-social-media-propaganda-islamic-state-terrorism Accessed 26 January, 2023, 2022.

[237] Aja Romano. What we still haven't learned from gamergate. https://www.vox.com/culture/2020/1/20/20808875/gamergate-lessons-cultural-impact-changes-harassment-laws Accessed 26 January, 2023, 2021.

[238] Katanga Johnson. Twitter permanently bans u.s. representative marjorie taylor greene. https://www.reuters.com/world/us/twitter-permanently-bans-us-representative-marjorie-taylor-greene-2022-01-02/ Accessed 17 March, 2023, 2022.

[239] Tom McKay. Joe rogan joined gettr 10 days ago and already thinks it sucks. https://gizmodo.com/joe-rogan-joined-gettr-10-days-ago-and-already-thinks-i-1848346452 Accessed 17 March, 2023, 2022.

[240] Matthew R. DeVerna, Rachith Aiyappa, Diogo Pacheco, John Bryden, and Filippo Menczer. Identification and characterization of misinformation superspreaders on social

media. *arXiv*, 2022.

[241] Shiza Ali, Mohammad Hammas Saeed, Esraa Aldreabi, Jeremy Blackburn, Emiliano De Cristofaro, Savvas Zannettou, and Gianluca Stringhini. Understanding the effect of deplatforming on social networks. In *13th ACM Web Science Conference 2021*, New York, NY, USA, 2021. Association for Computing Machinery.

[242] Dan Friedman and Ali Breland. Leaked messages show gettr in crisis mode over joe rogan criticism. https://www.motherjones.com/politics/2022/01/leaked-messages-show-gettr-crisis-mode-joe-rogan-jason-miller-guo-wengui/ Accessed 20 January, 2023, 2022.

[243] David Thiel and Miles McCain. Topologies and tribulations of gettr: A month in the life of a new alt-network. *Stanford Internet Observatory*, 2021.

[244] Carsten Schwemmer. The limited influence of right-wing movements on social media user engagement. *Social Media + Society*, 7(3):20563051211041650, 2021.

[245] Maarten Grootendorst. Bertopic: Neural topic modeling with a class-based tf-idf procedure. *ArXiv preprint 2203.05794*, 2022.

[246] Erin Woo. Twitter will stiffen moderation policies in response to the war in ukraine. *The New York Times*, 2022.

[247] Twitter Transparency Team. Covid-19 misinformation transparency report. https://transparency.twitter.com/en/reports/covid19.html#2021-jul-dec, 2022. Accessed 7 February, 2023.

[248] Kate Conger and Mike Isaac. Twitter permanently bans trump, capping online revolt. *The New York Times*, 2021.

[249] Pablo Barberá. Birds of the same feather tweet together: Bayesian ideal point estimation using twitter data. *Political analysis*, 23(1):76–91, 2015.

[250] James Flamino, Alessandro Galeazzi, Stuart Feldman, Michael W. Macy, Brendan Cross, Zhenkun Zhou, Matteo Serafino, Alexandre Bovet, Hernán A. Makse, and Boleslaw K. Szymanski. Political polarization of news media and influencers on Twitter in the 2016 and 2020 US presidential elections. *Nature Human Behaviour*, March 2023.

[251] Max Falkenberg, Alessandro Galeazzi, Maddalena Torricelli, Niccolò Di Marco, Francesca Larosa, Madalina Sas, Amin Mekacher, Warren Pearce, Fabiana Zollo, Walter Quattrociocchi, et al. Growing polarization around climate change on social media. *Nature Climate Change*, 12:1114–1121, 2022.

[252] John A Hartigan and Pamela M Hartigan. The dip test of unimodality. *The annals of Statistics*, 13(1):70–84, 1985.

[253] Sandra González-Bailón and Manlio De Domenico. Bots are less central than verified accounts during contentious political events. *Proceedings of the National Academy of Sciences*, 118(11):e2013443118, 2021.

[254] Alexandre Bovet and Hernán A. Makse. Influence of fake news in twitter during the 2016 US presidential election. *Nature Communications*, 10(1):1–14, January 2019.

[255] Matteo Cinelli, Gabriele Etta, Michele Avalle, Alessandro Quattrociocchi, Niccolo Di Marco, Carlo Valensise, Alessandro Galeazzi, and Walter Quattrociocchi. Conspiracy theories and social media platforms. *Current Opinion in Psychology*, page 101407, 2022.

[256] Kathleen Hall Jamieson and Joseph N Cappella. *Echo chamber: Rush Limbaugh and the conservative media establishment*. Oxford University Press, 2008.

[257] Pablo Barberá. Social media, echo chambers, and political polarization. *Social media and democracy: The state of the field, prospects for reform*, 34, 2020.

[258] Michael Conover, Jacob Ratkiewicz, Matthew Francisco, Bruno Gonçalves, Filippo Menczer, and Alessandro Flammini. Political polarization on twitter. In *Proceedings of the international aaai conference on web and social media*, volume 5, pages 89–96, 2011.

[259] Wesley Cota, Silvio C Ferreira, Romualdo Pastor-Satorras, and Michele Starnini. Quantifying echo chamber effects in information spreading over political communication networks. *EPJ Data Science*, 8(1):35, 2019.

[260] Ryan J Gallagher, Andrew J Reagan, Christopher M Danforth, and Peter Sheridan Dodds. Divergent discourse between protests and counter-protests:# blacklivesmatter and# alllivesmatter. *PloS one*, 13(4):e0195644, 2018.

[261] Robin Mamié, Manoel Horta Ribeiro, and Robert West. Are anti-feminist communities gateways to the far right? evidence from reddit and youtube. In *13th ACM Web Science Conference 2021*, WebSci '21, page 139–147, New York, NY, USA, 2021. Association for Computing Machinery.

[262] Mona Lena Krook. Violence against women in politics: A rising global trend. *Politics & Gender*, 14(4):673–675, 2018.

[263] Panagiotis Metaxas, Eni Mustafaraj, Kily Wong, Laura Zeng, Megan O'Keefe, and Samantha Finn. What do retweets indicate? results from user survey and meta-review of research. In *Proceedings of the international AAAI conference on web and social media*, volume 9, pages 658–661, 2015.

[264] Pedro Calais Guerra, Roberto Nalon, Renato Assunçao, and Wagner Meira Jr. Antagonism also flows through retweets: The impact of out-of-context quotes in opinion

polarization analysis. In *Eleventh international AAAI conference on web and social media*, 2017.

[265] Massimo Stella, Emilio Ferrara, and Manlio De Domenico. Bots increase exposure to negative and inflammatory content in online social systems. *Proceedings of the National Academy of Sciences*, 115(49):12435–12440, 2018.

[266] Livia Van Vliet, Petter Törnberg, and Justus Uitermark. The twitter parliamentarian database: Analyzing twitter politics across 26 countries. *PLoS one*, 15(9):e0237073, 2020.

[267] Brittany Shepherd and Tal Axelrod. How liz cheney went from rising republican star to primary underdog after jan. 6. https://abcnews.go.com/Politics/liz-cheney-rising-republican-star-enters-primary-underdog/story?id=88415555 Accessed 21 February, 2023, 2022.

[268] Lauren Fedor and Andrew Edgecliffe-Johnson. Adam kinzinger: the 'rino' leading the charge for a post-trump gop. https://www.ft.com/content/7984b8e0-7c8b-4527-95ec-4e46b6048499 Accessed 21 February, 2023, 2021.

[269] Hause Lin, Jana Lasser, Stephan Lewandowsky, Rocky Cole, Andrew Gully, David Rand, and Gordon Pennycook. High level of agreement across different news domain quality ratings. *PsyArXiv*, 2022.

[270] Mohsen Mosleh, Qi Yang, Tauhid Zaman, Gordon Pennycook, and David Gertler Rand. Trade-offs between reducing misinformation and politically-balanced enforcement on social media. *PsyArXiv Preprint*, April 2022.

[271] Hong Fan, Wu Du, Abdelghani Dahou, Ahmed A. Ewees, Dalia Yousri, Mohamed Abd Elaziz, Ammar H. Elsheikh, Laith Abualigah, and Mohammed A. A. Al-qaness. Social media toxicity classification using deep learning: Real-world application UK brexit. *Electronics*, 10(11):1332, June 2021.

[272] Md Rabiul Awal, Rui Cao, Sandra Mitrovic, and Roy Ka-Wei Lee. On analyzing antisocial behaviors amid covid-19 pandemic. *arXiv*, 2020.

[273] Zachary Petrizzo. Jason miller's 'free speech' social media platform gettr boots white nationalist. https://www.thedailybeast.com/jason-millers-free-speech-social-media-platform-gettr-boots-white-nationalist-nicholas-fuentes Accessed 16 March, 2023, 2021.

[274] Meridith McGraw. Bannon on brazil riots: 'i'm not backing off 1 inch'. https://www.politico.com/news/2023/01/09/bannon-brazil-riots-trump-00077155 Accessed 24 July, 2023, 2023.

[275] Joshua A Tucker, Yannis Theocharis, Margaret E Roberts, and Pablo Barberá. From liberation to turmoil: Social media and democracy. *Journal of democracy*, 28(4):46–59,

2017.

[276] Nathaniel Persily and Joshua A Tucker. Social media and democracy: The state of the field, prospects for reform. *Cambridge University Press*, 2020.

[277] Stanford Internet Observatory. Gogettr. https://github.com/stanfordio/gogettr, 2021. Accessed 6 February, 2023.

[278] Jack Morse. Gettr, that site for twitter rejects, is mad twitter won't let it import tweets. https://www.nbcnews.com/think/opinion/twitter-lacked-something-important-political-discourse-joe-rogan-found-it-ncna1287285 Accessed 27 January, 2023, 2021.

[279] Samuel S. Guimarães, Julio C. S. Reis, Filipe N. Ribeiro, and Fabrício Benevenuto. Characterizing toxicity on facebook comments in brazil. In *Proceedings of the Brazilian Symposium on Multimedia and the Web*, WebMedia '20, page 253–260, New York, NY, USA, 2020. Association for Computing Machinery.

[280] Deepak Kumar, Jeff Hancock, Kurt Thomas, and Zakir Durumeric. Understanding longitudinal behaviors of toxic accounts on reddit. *arXiv*, 2022.

[281] Google perspective api: Attributes and languages. https://developers.perspectiveapi.com/s/about-the-api-attributes-and-languages?language=en_US Accessed 16 March, 2023, 2023.

[282] Sangam Singh. Koo said to be the second largest microblogging platform in the world now. https://www.cnbctv18.com/business/companies/koo-becomes-second-largest-microblogging-platform-in-the-world-after-twitter-15179881.htm Accessed 1 December, 2023, 2022.

[283] Prashanth Bhat. Platform politics: The emergence of alternative social media in india. *Asia Pacific Media Educator*, 31(2):269–276, December 2021.

[284] C.J. Werleman. Rising violence against muslims in india under modi and BJP rule. *Resurgence of Anti Islam in the World*, 23(Spring 2021):39–49, June 2021.

[285] Sofia Ammassari, Diego Fossati, and Duncan McDonnell. Supporters of india's BJP: Distinctly populist and nativist. *Government and Opposition*, 58(4):807–823, June 2022.

[286] Francisco Paulo Jamil Marques, Camila Mont'Alverne, and Isabele Mitozo. Editorial journalism and political interests: Comparing the coverage of dilma rousseff's impeachment in brazilian newspapers. *Journalism*, 22(11):2816–2835, December 2019.

[287] Francis P. Barclay, C. Pichandy, and Anusha Venkat. Indian elections, 2014: Political orientation of english newspapers. *Asia Pacific Media Educator*, 24(1):7–22, June 2014.

[288] Sheila Dang. Twitter rival koo integrates chatgpt to help users create content. https://www.reuters.com/technology/twitter-rival-koo-integrates-chatgpt-help-users-create-content-2023-03-13/ Accessed 6 November, 2023, 2023.

[289] HT Tech. Koo app yellow tick of verification: Here is how to get it. https://tech.hindustantimes.com/how-to/koo-app-yellow-tick-of-verification-here-is-how-to-get-it-71637228876715.html Accessed 2 November, 2023, 2022.

[290] Sankalp Phartiyal. Explainer: The indian twitter rival staging a koo. https://www.reuters.com/business/media-telecom/indian-twitter-rival-staging-koo-2021-02-10/ Accessed 30 October, 2023, 2021.

[291] Roshni Majumdar. What is the koo app and why are government officials joining it? https://www.indiatoday.in/india-today-insight/story/what-is-the-koo-app-and-why-are-government-officials-joining-it-1769895-2021-02-16 Accessed 6 December, 2023, 2021.

[292] Precog Research Group. Challenging big social media, falling to partisanship: The koo conundrum. https://precog.iiit.ac.in/blog/2021/08/30/the-koo-conundrum/ Accessed 15 November, 2023, 2022.

[293] Danielle Paquette. Nigeria suspends twitter after the social media platform freezes president's account. https://www.washingtonpost.com/world/2021/06/04/nigeria-suspends-twitter-buhari/ Accessed 30 October, 2023, 2021.

[294] Emmanuel Akinwotu. Nigeria lifts twitter ban seven months after site deleted president's post. https://www.theguardian.com/world/2022/jan/13/nigeria-lifts-twitter-ban-seven-months-after-site-deleted-presidents-post Accessed 2 November, 2023, 2022.

[295] Damilare Dosunmu. When nigeria banned x, koo had a golden opportunity, and squandered it. https://restofworld.org/2023/nigeria-koo-twitter-rival-flop/ Accessed 2 November, 2023, 2023.

[296] Regina Mihindukulasuriya. Brazilian president lula joins india's twitter rival koo, gets over 50,000 followers in 4 hours. https://theprint.in/world/brazilian-president-lula-joins-indias-twitter-rival-koo-gets-over-50000-followers-in-4-hours/1245516/ Accessed 23 October, 2023, 2022.

[297] Danny D'Cruze. Made in india twitter-rival koo enters brazil with support for portuguese language. https://www.businesstoday.in/technology/news/story/made-in-india-twitter-rival-koo-enters-brazil-with-support-for-portuguese-language-353635-2022-11-21 Accessed 29 November, 2023, 2022.

[298] Max Falkenberg, Fabiana Zollo, Walter Quattrociocchi, Jürgen Pfeffer, and Andrea Baronchelli. Affective and interactional polarization align across countries. *ArXiv Preprint*, page 2311.18535, 2023.

[299] Emily Sullivan, Max Sondag, Ignaz Rutter, Wouter Meulemans, Scott Cunningham, Bettina Speckmann, and Mark Alfano. Vulnerability in social epistemic networks. *International Journal of Philosophical Studies*, 28(5):731–753, June 2020.

[300] Toby Handfield. Regulating social media as a public good: Limiting epistemic segregation. *Social Epistemology*, page 1–16, February 2023.

[301] Elizabeth Anderson. Epistemic bubbles and authoritarian politics. *Political epistemology*, pages 11–30, 2021.

[302] Ashwini Ramesh. Media and religion: Media framing of significant religious issues in english newspapers of india. *Mass Communicator: International Journal of Communication Studies*, 16(1):12–20, 2022.

[303] Pradip Thomas. Religion, media and culture in india: Hindutva and hinduism. *Media and religion: The global view*, 74:205–218, 2021.

[304] Media Ownership Monitor. Media ownership matters. https://www.mom-gmr.org/en/ Accessed 22 November, 2023, 2019.

[305] Samson Adenekan. Police raid peoples gazette office, arrest five staffers. https://www.premiumtimesng.com/news/top-news/544341-police-raid-peoples-gazette-office-arrest-five-staffers.html?tztc=1 Accessed 4 December, 2023, 2022.

[306] Premium Times. Presidency attacks punch newspaper. https://www.premiumtimesng.com/news/more-news/282609-presidency-attacks-punch-newspaper.html Accessed 4 December, 2023, 2018.

[307] Nic Newman, Richard Fletcher, Anne Schulz, Simge Andi, Craig T Robertson, and Rasmus Kleis Nielsen. Reuters institute digital news report 2021. *Reuters Institute for the study of Journalism*, 2021.

[308] Teun A van Dijk. How globo media manipulated the impeachment of brazilian president dilma rousseff. *Discourse & Communication*, 11(2):199–229, February 2017.

[309] Letícia Paiva. Ciro gomes deverá indenizar dcm por dizer que é 'tudo picareta' contratado pelo pt. https://www.jota.info/justica/ciro-gomes-devera-indenizar-dcm-por-dizer-que-e-tudo-picareta-contratado-pelo-pt-06022023 Accessed 11 December, 2023, 2023.

[310] Perry Keller. *Democracy, Pluralism, and the Media*, page 405–448. Oxford University Press, 2011.

[311] Glen Joris, Frederik De Grove, Kristin Van Damme, and Lieven De Marez. News diversity reconsidered: A systematic literature review unraveling the diversity in conceptualizations. *Journalism Studies*, 21(13):1893–1912, 2020.

[312] William Allchorn. Far-right extremist exploitation of ai and alt-tech: The need for p/cve responses to an emerging technological trend. https://gnet-research.org/2023/10/09/far-right-extremist-exploitation-of-ai-and-alt-tech-the-need-for-p-cve-responses-to-an-emerging-technological-trend/ Accessed 14 December, 2023, 2023.

[313] Elizabeth Losh. Hashtag feminism and twitter activism in india. *Social Epistemology Review and Reply Collective*, 3(3):11–22, 2014.

[314] Anmol Panda, Ramaravind Kommiya Mothilal, Monojit Choudhury, Kalika Bali, and Joyojeet Pal. Topical focus of political campaigns and its impact: Findings from politicians' hashtag use during the 2019 indian elections. *Proceedings of the ACM on Human-Computer Interaction*, 4(CSCW1):1–14, 2020.

[315] Tunde Opeibi. The twittersphere as political engagement space: A study of social media usage in election campaigns in nigeria. *Digital Studies/Le champ numérique*, 9(1), 2019.

[316] Ufuoma Akpojivi. I won't be silent anymore: Hashtag activism in nigeria. *Communicatio: South African Journal of Communication Theory and Research*, 45(4):19–43, 2019.

[317] Heloisa Buarque de Almeida. From shame to visibility: hashtag feminism and sexual violence in brazil. *Sexualidad, Salud y Sociedad (Rio de Janeiro)*, pages 19–41, 2020.

[318] Felipe Bonow Soares and Raquel Recuero. Hashtag wars: Political disinformation and discursive struggles on twitter conversations during the 2018 brazilian presidential campaign. *Social Media+ Society*, 7(2):20563051211009073, 2021.

[319] Munish Chandra Pandey. Godman rampal sentenced to life in prison till death in murder case. https://www.indiatoday.in/india/story/sant-rampal-life-imprisonment-murder-1368904-2018-10-16 Accessed 6 December, 2023, 2018.

[320] Ayodele Oluwafemi. Group to hold "osinbajo day" on sunday, describes him as "nigeria's future for 2023"'. https://www.thecable.ng/group-to-hold-osinbajo-day-on-sunday-describes-him-as-nigerias-future-for-2023 Accessed 12 December, 2023, 2021.

[321] Business Day. How buhari's train to nowhere worsens nigeria's debt woes. https://businessday.ng/business-economy/article/how-buharis-train-to-nowhere-worsens-nigerias-debt-woes/ Accessed 4 December, 2023, 2023.

[322] Timmy Broderick. Russia is trying to leave the internet and build its own. https://www.scientificamerican.com/article/russia-is-trying-to-leave-the-internet-and-build-its-own/ Accessed 11 January, 2024, 2023.

[323] Tiffany Hsu, Stuart Thompson, and Steven Lee Myers. 'stakes are really high': misin-

formation researcher changes tack for 2024 us election. https://www.nytimes.com/2024/01/09/business/media/election-disinformation-2024.html Accessed 11 January, 2024, 2024.

[324] Global Witness. Letting hate flourish: Youtube and koo's lax response to the reporting of hate speech against women in india and the us. https://www.globalwitness.org/en/campaigns/digital-threats/letting-hate-flourish-youtube-and-koos-lax-response-to-the-reporting-of-hate-speech-against-women-in-india-and-the-us/, Accessed 14 September, 2024, 2024.

[325] Richard Wike, Laura Silver, Janell Fetterolf, Christine Huang, Sarah Austin, Laura Clancy, and Sneha Gubbala. Social media seen as mostly good for democracy across many nations, but u.s. is a major outlier. https://www.pewresearch.org/global/2022/12/06/social-media-seen-as-mostly-good-for-democracy-across-many-nations-but-u-s-is-a-major-outlier/ Accessed 8 February, 2024, 2022.

[326] Jacob Poushter, Caldwell Bishop, and Hanyu Chwe. Social media use continues to rise in developing countries but plateaus across developed ones. https://www.pewresearch.org/global/2018/06/19/social-media-use-continues-to-rise-in-developing-countries-but-plateaus-across-developed-ones/ Accessed 2 February, 2024, 2018.

[327] Megha Mandavia and Alnoor Peermohamed. Facebook disputes claims of inadequate flagging of vernacular content. https://economictimes.indiatimes.com/tech/technology/facebook-disputes-claims-of-inadequate-flagging-of-vernacular-content/articleshow/86814892.cms?from=mdr Accessed 2 February, 2024, 2021.

[328] Pauline Leong. Tech giants grapple with content moderation in south-east asia. https://www.straitstimes.com/opinion/tech-giants-grapple-with-content-moderation-in-south-east-asia Accessed 7 February, 2024, 2022.

[329] Dan Milmo. Facebook revelations: what is in cache of internal documents? https://www.theguardian.com/technology/2021/oct/25/facebook-revelations-from-misinformation-to-mental-health Accessed 12 February, 2024, 2021.

[330] Statista. Leading countries based on instagram audience size as of january 2023. https://www.statista.com/statistics/578364/countries-with-most-instagram-users/ Accessed 18 August, 2023, 2023.

[331] Music Business Worldwide. India is home to the largest number of youtube viewers in the world. but the platform's music chart isn't without flaws. https://www.musicbusinessworldwide.com/india-is-home-to-the-largest-number-of-youtube-viewers-in-the-world-but-the-platforms/ Accessed 31 October, 2023, 2022.

[332] Munsif Vengattil and Aditya Kalra. Facebook's growth woes in india: too much nudity, not enough women. https://www.reuters.com/article/meta-india-facebook-i

nsight-idCAKBN2OW0FG Accessed 31 October, 2023, 2022.

[333] Aprameya Radhakrishna. Multilingual social media could catalyse digital inclusion in india. https://www.livemint.com/opinion/online-views/multilingual-social-media-could-catalyse-digital-inclusion-in-india-11656950158510.html Accessed 22 January, 2024, 2022.

[334] X Developer Platform. Supported languages and browsers. https://developer.twitter.com/en/docs/twitter-for-websites/supported-languages Accessed 11 March, 2024, 2024.

[335] Nikhil Inamdar. Koo: India's twitter alternative with global ambitions. https://www.bbc.co.uk/news/world-asia-india-60194920 Accessed 14 September, 2023, 2022.

[336] Antonis Papasavva and Enrico Mariconti. Waiting for q: An exploration of qanon users' online migration to poal in the wake of voat's demise. *arXiv*, 2023.

[337] Lichan Hong, Gregorio Convertino, and Ed Chi. Language matters in twitter: A large scale study. *Proceedings of the International AAAI Conference on Web and Social Media*, 5(1):518–521, August 2021.

[338] Delia Mocanu, Andrea Baronchelli, Nicola Perra, Bruno Gonçalves, Qian Zhang, and Alessandro Vespignani. The twitter of babel: Mapping world languages through microblogging platforms. *PLoS ONE*, 8(4):e61981, April 2013.

[339] Balachander Krishnamurthy, Phillipa Gill, and Martin Arlitt. A few chirps about twitter. In *Proceedings of the first workshop on Online social networks*. ACM, August 2008.

[340] Akshay Java, Xiaodan Song, Tim Finin, and Belle Tseng. Why we twitter. In *Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 workshop on Web mining and social network analysis*. ACM, August 2007.

[341] Wouter Weerkamp, Simon Carter, and Manos Tsagkias. How people use twitter in different languages. *Interdisciplinary Journal for the Study of Discourse*, 01 2011.

[342] Yuri Takhteyev, Anatoliy Gruzd, and Barry Wellman. Geography of twitter networks. *Social Networks*, 34(1):73–81, January 2012.

[343] Suin Kim, Ingmar Weber, Li Wei, and Alice Oh. Sociolinguistic analysis of twitter in multilingual societies. In *Proceedings of the 25th ACM conference on Hypertext and social media*, HT '14. ACM, 2014.

[344] Jannis Androutsopoulos. *Language Choice and Code Switching in German-Based Diasporic Web Forums*, page 340–361. Oxford University Press, June 2007.

[345] George Bailey, Joseph Goggins, and Thomas Ingham. What can twitter tell us about the language diversity of greater manchester. *Report by Multilingual Manchester. School of Languages, Linguistics and Cultures at the University of Manchester. http://bit. ly/2kG42Qf*, 2013.

[346] Yu-Ru Lin, Shaomei Wu, and Winter Mason. Mapping language literacy at scale: a case study on facebook. *EPJ Data Science*, 12(1), May 2023.

[347] Mahadev L Apte. Multilingualism in india and its socio-political implications: An overview,'. *Language and Politics*, pages 141–164, 1976.

[348] B Mallikarjun. Patterns of indian multilingualism. *Language in India*, 10:1–18, 2010.

[349] Debi Prasanna Pattanayak. *Multilingualism in India*. Number 61. Multilingual Matters, 1990.

[350] Ethelbert E Kari. Multilingualism in nigeria: the example of rivers state. In *A Paper Presented at the Seminar on Multilingual Situation and Related Local Cultures in Asia and Africa Institute for the Study of Languages and Cultures of Asia and Africa, Tokyo University of Foreign Studies, Tokyo*, volume 25, 2002.

[351] Ayeomoni Moses Omoniyi. The languages in nigerian socio-political domains: Features and functions. *English Language Teaching*, 5(10), August 2012.

[352] Hannah Ellis-Petersen. India threatened to shut twitter down, co-founder jack dorsey says. https://www.theguardian.com/world/2023/jun/13/india-threatened-to-shut-twitter-down-co-founder-jack-dorsey-says Accessed 20 October, 2023, 2023.

[353] Gerry Shih. In india, a government-friendly social media network challenges twitter. https://www.washingtonpost.com/world/2021/11/16/india-twitter-koo-social-network/ Accessed 15 August, 2023, 2021.

[354] Agbaje Ayomide. Inside indian app koo, where nigerians are migrating to after a twitter ban. https://qz.com/africa/2024107/nigerian-government-moves-to-indian-app-koo-after-twitter-ban Accessed 15 August, 2023, 2022.

[355] Nimi Princewill and Stephanie Busari. Nigeria bans twitter after company deletes president buhari's tweet. https://edition.cnn.com/2021/06/04/africa/nigeria-suspends-twitter-operations-intl/index.html/ Accessed 31 October, 2023, 2021.

[356] Mehul Reuben Das. India's koo launched in brazil in portuguese, becomes the top downloaded app in 48 hours. https://www.firstpost.com/tech/news-analysis/koo-launched-in-brazil-in-portuguese-becomes-the-top-downloaded-app-in-48-hours-11671621.html Accessed 24 January, 2024, 2022.

[357] LiveMint. Not competing with twitter or threads, focusing on regional audiences: Koo co-founder. https://www.livemint.com/companies/not-competing-with-twitter-or-threads-focusing-on-regional-audiences-koo-cofounder-11692425534832.html Accessed 10 November, 2023, 2023.

[358] Munmun De Choudhury, Hari Sundaram, Ajita John, Doree Duncan Seligmann, and Aisling Kelliher. "birds of a feather": Does user homophily impact information diffusion in social media? *arXiv*, 2010.

[359] Morteza Dehghani, Kate Johnson, Joe Hoover, Eyal Sagi, Justin Garten, Niki Jitendra Parmar, Stephen Vaisey, Rumen Iliev, and Jesse Graham. Purity homophily in social networks. *Journal of Experimental Psychology: General*, 145(3):366–375, March 2016.

[360] Mohammed Ali Al-garadi, Kasturi Dewi Varathan, and Sri Devi Ravana. Identification of influential spreaders in online social networks using interaction weighted k-core decomposition method. *Physica A: Statistical Mechanics and its Applications*, 468:278–288, February 2017.

[361] Harold Schiffman. *Linguistic culture and language policy*. The Politics of Language. Routledge, London, England, March 1998.

[362] Keyi Kang, Yumeng Xiao, Hanxiang Yu, Michele T. Diaz, and Haoyun Zhang. Multilingual language diversity protects native language production under different control demands. *Brain Sciences*, 13(11):1587, November 2023.

[363] Olga Kepinska, Jocelyn Caballero, Myriam Oliver, Rebecca A. Marks, Stephanie L. Haft, Leo Zekelman, Ioulia Kovelman, Yuuko Uchikoshi, and Fumiko Hoeft. Language combinations of multilinguals are reflected in their first-language knowledge and processing. *Scientific Reports*, 13(1), February 2023.

[364] M. E. J. Newman. Assortative mixing in networks. *Physical Review Letters*, 89(20), October 2002.

[365] Craig T. Nagoshi, Ronald C. Johnson, and George P. Danko. Assortative mating for cultural identification as indicated by language use. *Behavior Genetics*, 20(1):23–31, January 1990.

[366] Lars Leszczensky and Sebastian Pink. What drives ethnic homophily? a relational approach on how ethnic identification moderates preferences for same-ethnic friends. *American Sociological Review*, 84(3):394–419, May 2019.

[367] Ana Sierra Leonard, Ajay Mehra, and Ralph Katerberg. The social identity and social networks of ethnic minority groups in organizations: a crucial test of distinctiveness theory. *Journal of Organizational Behavior*, 29(5):573–589, September 2007.

[368] A. Mehra, M. Kilduff, and D. J. Brass. At the margins: a distinctiveness approach to the social identity and social networks of underrepresented groups. *Academy of*

Management Journal, 41(4):441–452, August 1998.

[369] Roberta Amato, Lucas Lacasa, Albert Díaz-Guilera, and Andrea Baronchelli. The dynamics of norm change in the cultural evolution of language. *Proceedings of the National Academy of Sciences*, 115(33):8260–8265, August 2018.

[370] D.G. Scragg. *A History of English Spelling*. Mont Follick series. Manchester University Press, 1974.

[371] John H Vivian. Spelling an end to orthographical reforms: Newspaper response to the 1906 roosevelt simplifications. *American Speech*, 54(3):163–174, 1979.

[372] Daniel M. Abrams and Steven H. Strogatz. Modelling the dynamics of language death. *Nature*, 424(6951):900–900, August 2003.

[373] James S. Coleman. Relational analysis: The study of social organizations with survey methods. *Human Organization*, 17(4):28–36, 1958.

[374] Robert Warton, Chris Volny, and Kevin S. Xu. Counteracting filter bubbles with homophily-aware link recommendations. In *Social, Cultural, and Behavioral Modeling*, pages 155–164. Springer International Publishing, 2022.

[375] Elisha B. Are, Kiffer G. Card, and Caroline Colijn. The role of vaccine status homophily in the COVID-19 pandemic: A cross-sectional survey with modeling. *Cold Spring Harbor Laboratory*, June 2023.

[376] Prakash Karat. The role of the english-educated in indian politics. *Social Scientist*, 1(4):25, November 1972.

[377] Shahar Ronen, Bruno Gonçalves, Kevin Z. Hu, Alessandro Vespignani, Steven Pinker, and César A. Hidalgo. Links that speak: The global language network and its association with global fame. *Proceedings of the National Academy of Sciences*, 111(52):E5616–E5622, 2014.

[378] Office of the Registrar General & Census Commissioner, India (ORGI) . Language atlas of india 2011. https://censusindia.gov.in/nada/index.php/catalog/42561 Accessed 5 February, 2024, 2011.

[379] Aimei Yang, Ian Myoungsu Choi, Andrés Abeliuk, and Adam Saffer. The influence of interdependence in networked publics spheres: How community-level interactions affect the evolution of topics in online discourse. *Journal of Computer-Mediated Communication*, 26(3):148–166, May 2021.

[380] Gabriel Medina, Benno Pokorny, and Jes Weigelt. The power of discourse: Hard lessons for traditional forest communities in the amazon. *Forest Policy and Economics*, 11(5-6):392–397, October 2009.

[381] Amy D. Willis. Rarefaction, alpha diversity, and statistics. *Frontiers in Microbiology*, 10, October 2019.

[382] Xiaoquan Su. Elucidating the beta-diversity of the microbiome: from global alignment to local alignment. *mSystems*, 6(4), August 2021.

[383] François Grin and Guillaume Fürst. Measuring linguistic diversity: A multi-level metric. *Social Indicators Research*, 164(2):601–621, July 2022.

[384] Iman M. Mahfouz. The linguistic characteristics and functions of hashtags: #is it a new language? *Arab World English Journal*, 6:84–101, July 2020.

[385] Xuan-Phi Nguyen, Sharifah Mahani Aljunied, Shafiq Joty, and Lidong Bing. Democratizing llms for low-resource languages by leveraging their english dominant abilities with linguistically-diverse prompts. *arXiv*, 2023.

[386] Chun-Huo Chiu, Yi-Ting Wang, Bruno A. Walther, and Anne Chao. An improved nonparametric lower bound of species richness via a modified good–turing frequency formula. *Biometrics*, 70(3):671–682, 2014.

[387] Shaila Seshia. Divide and rule in indian party politics: The rise of the bharatiya janata party. *Asian Survey*, 38(11):1036–1050, November 1998.

[388] R. B. Bhagat. Cencus and the construction of communalism in india. *Economic and Political Weekly*, 36(46/47):4352–4356, 2001.

[389] Minhui Hao, J. Javier Corral-Rivas, M. Socorro González-Elizondo, K. Narayanagowda Ganeshaiah, M. Guadalupe Nava-Miranda, Chunyu Zhang, Xiuhai Zhao, and Klaus von Gadow. Assessing biological dissimilarities between five forest communities. *Forest Ecosystems*, 6(1), June 2019.

[390] M. B. Emeneau. India as a lingustic area. *Language*, 32(1):3–16, 1956.

[391] Sahith Aula. The problem with the english language in india. https://www.forbes.com/sites/realspin/2014/11/06/the-problem-with-the-english-language-in-india/ Accessed 12 November, 2023, 2014.

[392] Shonakshi Chakravarty. Koo solves everything that annoys twitter users today, says koo ceo. https://www.outlookindia.com/business/koo-solves-everything-that-annoys-twitter-users-today-says-koo-ceo--news-256090 Accessed 29 September, 2023, 2023.

[393] Benjamin Cedric Larsen. The geopolitics of ai and the rise of digital sovereignty. https://www.brookings.edu/articles/the-geopolitics-of-ai-and-the-rise-of-digital-sovereignty/ Accessed 22 January, 2024, 2022.

[394] Daniel Hershcovich, Stella Frank, Heather Lent, Miryam de Lhoneux, Mostafa Abdou, Stephanie Brandl, Emanuele Bugliarello, Laura Cabello Piqueras, Ilias Chalkidis, Ruix-

iang Cui, Constanza Fierro, Katerina Margatina, Phillip Rust, and Anders Søgaard. Challenges and strategies in cross-cultural nlp. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Association for Computational Linguistics, 2022.

[395] Sam Levin. Us capitol attack: is the government's expanded online surveillance effective? https://www.theguardian.com/us-news/2022/jan/07/us-capitol-attack-government-online-surveillance Accessed 2 May, 2024, 2022.

[396] Jessica Sciarone. Radicalization pathways among women in us far-right extremist networks and implications for deradicalization. *Journal for Deradicalization*, (38):81–121, 2024.

[397] Matthew Katsaros, Kathy Yang, and Lauren Fratamico. Reconsidering tweets: Intervening during tweet creation decreases offensive content. *Proceedings of the International AAAI Conference on Web and Social Media*, 16:477–487, may 2022.

[398] Jaihyun Park, JungHwan Yang, Amanda Tolbert, and Katherine Bunsold. You change the way you talk: Examining the network, toxicity and discourse of cross-platform users on twitter and parler during the 2020 us presidential election. *Journal of Information Science*, April 2024.

[399] Pablo Jost and Harald Sick. Cash for incitement: The monetisation of digital hate in germany. https://gnet-research.org/2024/05/01/cash-for-incitement-the-monetisation-of-digital-hate-in-germany/ Accessed 3 May, 2024, 2024.

[400] Vivian Ferrillo. r/the_donald had a forum: How socialization in far-right social media communities shapes identity and spreads extreme rhetoric. *American Politics Research*, April 2024.

[401] Srayan Datta and Eytan Adar. Extracting inter-community conflicts in reddit. In *Proceedings of the international AAAI conference on Web and Social Media*, volume 13, pages 146–157, 2019.

[402] Nicholas Reimann and Robert Hart. Elon musk says he's granting 'amnesty' for nearly all banned twitter accounts. https://www.forbes.com/sites/nicholasreimann/2022/11/24/elon-musk-says-hes-granting-amnesty-for-nearly-all-banned-twitter-accounts/?sh=5d3674e72e95 Accessed 26 January, 2023, 2022.

[403] Odilon Caldeira Neto. The brazilian far-right and the path to january 8th. https://gnet-research.org/2023/01/23/the-brazilian-far-right-and-the-path-to-january-8th/, Accessed 24 January, 2023, 2023.

[404] Caleb Ecarma. The right-wing media's coverage of brazil's insurrection is a rerun of january 6. https://www.vanityfair.com/news/2023/01/right-wing-media-brazils-insurrection Accessed 24 January, 2023, 2023.

[405] Linhong Zhu and Kristina Lerman. Attention inequality in social media. *arXiv*, 2016.

[406] Benjamin Guinaudeau, Fabio Vottax, and Kevin Munger. Fifteen seconds of fame: Tiktok and the democratization of mobile video on social media. *Unpublished paper. Available at: https://osf. io/f7ehq/download*, 2020.

[407] Max Falkenberg. Heterogeneous node copying from hidden network structure. *Communications Physics*, 4(1):200, 2021.

[408] Keziban Orman, Vincent Labatut, and Hocine Cherifi. An empirical study of the relation between community structure and transitivity. *Complex Networks*, pages 99–110, 2013.

[409] Murillo Camarotto. Despite efforts to fight falsehoods, brazil's tight election is threatened by dangerous lies. https://reutersinstitute.politics.ox.ac.uk/news/despite-efforts-fight-falsehoods-brazils-tight-election-threatened-dangerous-lies Accessed 17 March, 2023, 2022.

[410] Ricardo Brito and Carolina Pulice. Bolsonaro challenges brazil election he lost to lula. https://www.reuters.com/world/americas/brazils-bolsonaro-files-complaint-challenge-election-results-2022-11-22/ Accessed 17 March, 2023, 2022.