Building social capital to counter polarization and extremism? A comparative analysis of tech platforms' official blog posts

Watkin, Amy-Louise; Conway, Maura

Published in:
First Monday

DOI:
10.5210/fm.v27i5.12611

Published: 02/05/2022

Document Version
Peer reviewed version

Citation for published version (APA):
Watkin, A-L., & Conway, M. (2022). Building social capital to counter polarization and extremism? A comparative analysis of tech platforms' official blog posts. *First Monday*, *27*(5), [12611]. https://doi.org/10.5210/fm.v27i5.12611

Building Social Capital to Counter Polarization and Extremism?

A Comparative Analysis of Tech Platforms' Official Blog Posts

Amy-Louise Watkin and Maura Conway

Abstract

This research employs the concept of social capital to compare the efforts that a range of tech companies have claimed to take to counter polarization and extremism and build resilience on their platforms. The dataset on which our analysis is based is made-up of a purposive sample of official blog posts from three 'older' (*i.e.*, Facebook, Twitter, YouTube) and three 'newer' (*i.e.*, TikTok, Discord, and Telegram) technology platforms. The selected posts focused on companies' efforts to make their platform safer, build community resilience, counter violent extremism and/or polarization, or mentioned related topics such as countering hate organizations, radicalization, or misinformation. Revealed were seven themes incorporating, to a greater or lesser extent, the three main types of social capital (*i.e.*, bonding, bridging and linking). These themes were granting user powers, strengthening existing communities, provision of information and education, building community, enhancing user rights, keeping users safe, and building trust and relationships with users. Analysis of these showed that while creation of all three types of social capital was apparent, similar to previous studies, bridging capital dominated here too; while there were some discrepancies between social capital generating activities and their framings on 'older' versus 'newer' platforms, other factors, including platform size and company values are likely to be equally or more important; and, finally, that companies attempts at generating online social capital can have negative as well as positive impacts with regard to countering polarization and extremism.

1. Introduction

Social media's effects are complex and multifaceted but, despite Mark Zuckerberg's oft repeated refrain that Facebook is responsible for more good than harm in the world, its potential negatives are being scrutinized more closely today than ever before. Much of the available research on contemporary polarization and extremism — and even terrorism — implicates the Internet, especially social media platforms, to a greater or lesser extent. Discussions already underway amongst not just academics but also policy-makers, law enforcement, and journalists about tech companies' parts in encouraging polarization and extremism, were given added impetus by the myriad roles played by online platforms in the January 2021 storming of the U.S. Capitol. The aftermath of those events witnessed a flurry of deplatforming of polarizing and extremist users, groups, and movements, and their content (Conway, *et al.*, 2021). Rather than focusing on deplatforming however, this article employs the concept of social capital to examine the efforts that a range of tech companies have claimed to take to build resilience and discourage polarization and extremism on their platforms. In effect, we ask is social capital being generated by tech companies to address polarization and extremism and build resilience on their platforms? And, if so, how?

As regards data, we utilized an extensive and under-utilized primary resource: tech companies' official blog posts, which contain a range of valuable insights into tech companies' efforts to counter a range of bad actors. Specifically, we collected and analyzed blogs posted by three 'old' (*i.e.*, Facebook [established 2004], Twitter [established 2006], YouTube [established 2005]) and three 'new' companies (*i.e.*, Discord [established 2015], Telegram [established 2013], and TikTok [established 2016]) in the period September 2017 to August 2020 that addressed building resilience, countering polarization, and/or countering extremism. Together this amounted to 436 posts, which accounted for 30 percent of all the posts made by the six companies on their corporate blogs in the data collection period and resulted in over 300,000 words of text for analysis.

We are aware that one explanation for this data source being all but ignored to-date is its dismissal as 'spin' and 'PR.' We do not doubt that there is an element of these involved in the blog posts, along with the posts also going more to what the companies say than what they actually do. Worth pointing out however is that not only does their publicness ensure that these types of pronouncements matter, but also that many of the posts were announcing the release of relevant tools and services for building resilience and discouraging polarization and extremism. Having said this, the article is focused on the companies rather than their users and so does not, in the main, delve into the implementation and take-up of the described tools and services. Instead, our focus is on examining the efforts that the six tech platforms have taken to grow the three main types of social capital (*i.e.*, bonding, bridging, and linking) on their sites and the likely outcomes of these for resilience-building and countering polarization and extremism in the respective online communities.

This article is divided into five sections. In the second section, immediately below, we describe and discuss the three main types of social capital and follow-up with discussion of the bodies of research that explore social capital and the Internet and social capital and policy-making respectively. Section three explains our case selection choices, details our data, and describes the methodology for generating our social capital dictionary. The article's fourth and lengthiest section describe and discusses our findings. Specifically, this section opens with a brief recounting of the similarities and differences between the amount of attention given to building social capital by 'older' versus 'newer' platforms. It then describes and discusses the seven themes emerging from analysis of the tech company blog posts on preventing polarization, extremism, and radicalization and the types of social capital (*i.e.*, bonding, bridging, and linking) encompassed by each.

These seven themes are granting user powers, strengthening existing communities, provision of information and education, building community, enhancing user rights, keeping users safe, and building trust and relationships with users. This ‘deep dive’ includes provision of example blog text and treatment of similarities and differences in framings across platforms. The Conclusion underlines three key findings: that while creation of all three types of social capital was apparent, similar to previous studies, bridging capital dominated here too; while there were some discrepancies between social capital generating activities and their framings on ‘older’ versus ‘newer’ platforms, other factors, including platform size and company values are likely to be equally or more important; and, finally, and perhaps most importantly, that companies’ generation of online social capital can have negative as well as positive impacts with regard to countering polarization and extremism.

2. On social capital

2.1. What is social capital?

Social capital is a resource originating in social relations and the creation or maintenance of community- and/or organization-based social connections that may be acquired and/or mobilized by a wide array of social actors (*e.g.*, individuals, companies, countries) (Portes, 1998; Putnam 2000). Put another way, social capital is the ability to secure benefits or resources through one’s memberships and relationships in social networks (Portes, 1998). Review of the academic literature on social capital, a majority of which is focused on the role(s) of individuals, summarizes it as reliant on social norms and values, social networks, a culture of trust, norms of reciprocity, participation, and collective action (Putnam, 2001; Mignone and O’Neil, 2005b; Frank and Yasumoto, 1998). Social capital has been posited as crucial to the efficient functioning of liberal democracies (Fukuyama, 2001) and a prerequisite to resilience (Brisson, *et al.*, 2017). Resilience is defined in this research as “the capacity of individuals and groups to cope in adverse or challenging circumstances” and “facilitated by the interdependent individual, social, economic and political resources [individuals and groups] are able to access and mobilise” [1].

Social capital is often split into three types: bonding, bridging, and linking. Bonding social capital is the strongest of the three social capitals and refers to relationships with family, friends, and those sharing some other important characteristic with one (*e.g.*, ideology, religion, ethnicity) (Ali, *et al.*, 2019). It is formed when one feels anchored in one’s own cultural and other beliefs and practices (Brisson, *et al.*, 2017). Bonding capital can increase collective strength within a community, result in the provision of emotional support, and provide protection from external threats. It can also be used to compensate in situations where the state is unable or unwilling to assist or where there is a lack of trust for leaders or authority (Ali, *et al.*, 2019). However, a concern with bonding social capital is that it can create exclusion (Tolsma and Zevallos, 2009; Ali, *et al.*, 2019). An example of such a negative externality of bonding social capital is the creation and maintenance of hate groups that actively distrust and exclude minority groups (Fukuyama, 2001).

Bridging social capital refers to the building of relationships between heterogeneous groups; for example, connections between friends of friends or with other people from different social groups/situations (Ali, *et al.*, 2019). Bridging capital requires confidence in and support for those belonging to other social groups, engaging and making ties with those people, and valuing inter-community connections and harmony (Brisson, *et al.*, 2017; Mignone and O'Neil, 2005b). These bridging relationships, sometimes referred to as 'cross-cutting' or 'weak' ties (Granovetter, 1973; Narayan, 1999; Larsen, *et al.*, 2004), can extend the radius of trust (Fukuyama, 2001) thus widening people's access to resources that are not present in their own community or organization (Ali, *et al.*, 2019). It follows from this that where bridging social capital is absent, knowledge and resources are missed out upon. This is thought to lead to isolation and disenfranchisement among some social groups (Larson, *et al.*, 2004; Brisson, *et al.*, 2017). On the other hand, where this social capital exists, there is a risk that people could drift from their bonding social capital communities, thus creating an atomized society instead of an inclusive one (Marozzi, 2016). Overall, this social capital has the potential to build more powerful communities with wider networks creating opportunities and strengthening social inclusion. However, as with both the other kinds of social capital, there are risks of negative unintended consequences, such as previously disparate groups realizing they have shared commitments and connecting around these (*e.g.*, wellness influencers and QAnon; see Fitzgerald, 2022, in this special issue).

Finally, linking social capital is traditionally the relationship between citizens and authorities, such as the government (Ali, *et al.*, 2019). It involves trust and confidence in government and institutions, being able to access and make use of the knowledge and resources provided by such institutions, and being able to influence policy decisions that affect one's own community (Brisson, *et al.*, 2017). Linking capital can create ties between people with differing struggles and situations and it can lead to the generation of additional resources for all involved. Linking capital can also lead to the utilization of institutions and organizations outside of one's local community (Brisson, *et al.*, 2017). Without linking capital, some social groups will be at a disadvantage regarding policy decisions and interventions, which can result in a lack of trust in government and institutions

(Putnam, 2000). Linking capital cannot exist without a democratic environment (Ali, *et al.*, 2019).

As can be seen, social capital can have both positive and negative effects (Narayan, 1999). Coleman (1988) defines social capital as a “neutral resource” that creates action and argues that outcomes depend on the way in which it is used. For a community to be resilient, it is held that a balance of all three forms of social capital must be present (Brisson, *et al.*, 2017). This understanding of social capital highlights the importance of both in-group networks as well as wider networks involving other social groups and institutions, and the dynamics surrounding trust, reciprocity, collective participation, and access to resources (Kirmayer, *et al.*, 2009). Given that the valuing of one’s own culture and learning about the cultures of others increases social capital and resilience more broadly, where this is missing — where a social group tries to deny or remove the presence of minority cultures — polarization is at risk of emerging (Gunnestad, 2006). Such oppressive practices have been known to result in not just the emergence of hate and extremist groups, but also self-hatred and low self-esteem among the excluded groups (Sonn and Fisher, 1998). These are the circumstances that recruiters of violent extremist organizations have been known to exploit (Brisson, *et al.*, 2017; Pickering, *et al.*, 2007). It is therefore crucial that policy and practice utilize social capital to create positive outcomes and minimize negative outcomes as much as is possible.

2.2. Social capital and the Internet

There is a plethora of research investigating whether the internet, particularly social media, can affect social capital. De Zúñiga, *et al.* (2017) found evidence to suggest that social capital on social media is empirically distinct from off-line social capital. They posit that while social capital:

> “continues to be a robust benchmark on how strongly people connect in their communities, share values, and watch out for one another, the results of this study suggest that the platforms people use to connect with one another affect the nature of the value derived from those relationships” [2].

The study also found that social media social capital was a better predictor of off-line social capital than *vice versa*. For example, the way users made connections, fostered their values, and communicated community problems online predicted whether this would affect the continuity of such things off-line.

Social media has been found to create all three types of social capital — bonding, bridging and linking (Ellison, *et al.*, 2007; Hawkins and Mauren, 2010; Burke, *et al.*, 2011; Quinn, 2016; Kim and Kim, 2017; Steinfield, *et al.*, 2008; Chen and Li, 2017; Raza, *et al.*, 2017). Some of the most cited research in this field is survey research by Ellison, *et al.* (2007) and Ellison, *et al.* (2011a). This research, like much of the research in this area, focused on Facebook and found that although social media can create bonding social capital (*i.e.*, strong ties) it is more likely to grow bridging social capital (*i.e.*, weak ties). Social capital scores increased when using the platform to learn more about weak ties in one's network. Social capital also increased the more intensely the platform was used. These findings are thought to be because social media provides identity information and enables easy communication that can bring people with shared interests together (Ellison, *et al.*, 2011a). Additionally, the lowered costs of online communication can create more opportunities to communicate with weak ties than one may have off-line (Ellison, *et al.*, 2007). The same research found no impact on social capital when trying to connect with total strangers; also, that there is a point of diminishing returns, around 400–500 'friends,' where it becomes impossible to engage in the activities necessary to maintain relationships enough to benefit from weak ties (Ellison, *et al.*, 2011a).

Research by Burke, *et al.* (2010) supported the findings in the Ellison, *et al.* (2011a; 2007) studies. Burke, *et al.* found that the more active users were on Facebook, the higher their bonding and bridging social capital. A similar finding was made in survey research undertaken by Hwang and Kim (2015). Burke, *et al.'s* (2010) research also found that social capital was more likely to be gained when users actively contributed and engaged on the platform as opposed to just passively consuming content. Other Facebook-focused survey research by Burke, *et al.* (2011) found that exchanging messages on Facebook increased bridging social capital. Research by Steinfield, *et al.* (2008) further supported these findings. Again, survey research focused on Facebook found that use of Facebook interacted with self-esteem to influence bridging social capital. It is thought that those with lower self-esteem find it easier to interact with weak ties via social media — through methods such as messaging and tagging — than in 'real world' settings. Chen and Li (2017) also found, via surveys, that communication and self-disclosure of personal information positively related to both bonding and bridging social capital, but that 'friending' someone was only positively related to bridging capital.

Finally, research has highlighted that an increased focus on privacy concerns over the years has likely affected the ability to create all three types of social capital on social media. Social media users are likely to have connections online spanning several different dimensions of their life, for example, family, colleagues, and school friends, amongst others (boyd, 2009). Users are therefore likely to try to mold their connections, who they can interact with, and who can view certain content that they post using platforms' privacy settings (Ali, *et al.*, 2019). Although there are obvious safety benefits to the use of privacy settings that are likely to appeal to some users (Ellison, *et al.*, 2011b), given that the creation of bridging capital is reliant on being able to view identity information of weak ties, and bridging social capital is most likely to create social inclusion, the use of privacy settings may complicate or impede the creation of bridging social capital (Ellison, *et al.*, 2011a). Further, privacy settings and the ability to remove, mute, and block other users could result in greater exclusion, polarization, and marginalization (Ali, *et al.*, 2019).

To summarize: first, the majority of research to-date on the intersections of social capital and social media has been survey research of social media users; second, these surveys generally contain indexes specific to the context or domain being researched, with research in the area being dominated by Facebook; third, the bulk of the research has focused on whether social media can increase social capital and, if so, which kinds of social capital and what platform activities result in this; fourth, much of this research is arguably outdated given the speed at which platforms evolve. Nonetheless, one overarching conclusion from this research is that social media networks often take the form of large and heterogeneous collections of weak ties (Granovetter, 1973; Donath and boyd, 2004). Another core finding is that although social media can increase all three kinds of social capital, it tends to have the biggest impact on bridging capital. This suggests that social media can expand social networks across different communities and social groups albeit, as mentioned earlier, the consequences of this can be both positive and negative. Unfortunately, little-to-none of this research is explicit as to how these findings could be used to inform policy-making across platforms. In the next sub-section therefore, we introduce some work on policies for generating social capital albeit in ‘real world’ settings.

2.3. Social capital and policy-making

According to Narayan (1999), the generation of dense cross-cutting ties (*i.e.,* bridging social capital) among social groups, the kind the literature suggests social media is most conducive to generating, is also the best means of achieving social cohesion. Narayan (1999) suggests several ways that cross-cutting ties can be nurtured through policy, including free information flow, inclusive participation, deployment of conflict management mechanisms, educational access and shared values, governance and decentralization, and demand-driven service delivery. Information is a public good and free information flow is important for creating equal opportunities amongst social groups. Opportunities for citizen participation and ensuring all social groups are represented in big decisions will generate trust. Conflict resolution mechanisms are necessary to protect human rights, ensure fair treatment across social groups, and subsequently generate trust. Access to education across all social groups ensures fairness and a level-playing field regarding knowledge. The state and communities working together in partnership creates inclusion and participation.

Policy interventions should nonetheless consider the ways in which such interventions could be used by dominant social groups to undermine the public good. In order to avoid this, Woolcock and Narayan (2000) recommend that all relevant stakeholders and their inter-relations be identified in order to understand how policy interventions will affect them. A second recommendation is to consider how bridges can be built between social groups, particularly those that have been excluded from resources. Next is information disclosure to all affected for purposes of informed citizenship and to hold those in power accountable. There should also be improvement, they say, in the opportunities and ways in which information can be exchanged across groups.

We identify the ways in which a range of technology companies incorporate the three main types of social capital (*i.e.*, bonding, bridging, and linking) in their claimed efforts to counter polarization and extremism and build resilience on their platforms. Our research differs from much of the research on online social capital produced to date in that it is not Facebook-centred, user-focused, or survey-based, but instead focused on the public pronouncements of a range of broadly social media companies, including Facebook, but also Twitter, YouTube, Tik Tok, Telegram, and Discord. It is thus more structure- than agent-focused and thereby top-down rather than bottom-up in its overall orientation. Our analysis nevertheless draws on the just-reviewed literature where appropriate, including particularly the work on policymaking for social capital, where the policies are however those of the companies rather than of governments.

3. Methodology

3.1. Case selection

The data in this study was collected from official blog posts written and published by six technology platforms: Facebook, Twitter, YouTube, Tik Tok, Telegram, and Discord [3]. These six were chosen because they represent a diversity of types of online spaces and all host official blogs that provide insight into their efforts to build resilience and tackle polarization and violent extremism, in which they have each been implicated to greater or lesser extents (see, for example, Benigni, *et al.*, 2017; Gallagher, *et al.*, 2021; Kubin and Sikorski, 2021; O'Callaghan, *et al.*, 2015; Scrivens and Amarasingam, 2020; Walther and McCoy, 2021; Weimann and Nasri, 2020).

Facebook, Twitter, and YouTube are often referred to as the 'major platforms' and were founded much earlier than Tik Tok, Telegram, and Discord. In addition to the dates of their establishment, the platforms differ in numerous other ways including, but not limited to, size of user base, number of employees, main types of content they host (*e.g.*, video, chat, etc.) and the roles that they play in extremist ecosystems (Baele, Brace and Coan, 2020; Watkin, 2019). Tech companies, in general, have varying attitudes to and capacities regarding countering polarization and extremism on their platforms. Some companies are willing to proactively counter it and have the expertise and capacity to do so, while others share that willingness but struggle with knowledge and/or capacity issues. Some other companies are less willing to proactively counter polarization and extremism on their services, with some of these actively encouraging such activity (Lima, *et al.*, 2018; Watkins, 2019). Yet others (*e.g.*, decentralized platforms) are technologically unable to deal with these issues when they arise (Rochko, 2019).

Although official tech company blogs cover a variety of topics, one of the main ways that the blogs tend to be used is for the platforms to communicate new policy decisions or initiatives, including efforts to counter bad actors on their sites. The blogs differ across platforms, but they typically describe the decision or initiative and why it has been implemented. These blogs therefore contain a lot of information on the efforts that companies are making to grow social capital on their platforms. At time of writing, there does not appear to be any previous research analyzing companies' corporate blogs.

3.2. Data collection and analysis

For this research, all relevant blog posts published between September 2017 to August 2020 were collected from each of the six blog Web sites, unless a platform's blog was created after September 2017, in which case blog posts were collected from the date they began to be published (up until August 2020). All posts were manually reviewed and collected, with initial relevance determined by posts' titles. If the blog title appeared to be related to efforts to make the platform safer, build community resilience, counter violent extremism and/or polarization, or mentioned related topics such as countering hate organizations, radicalization, or misinformation, then the post was read to assess whether it was directly relevant and therefore should be collected. The details (*e.g.*, source company, blog title, date posted, author, URL, and full text) of all relevant blog posts were entered into a spreadsheet. See Table 1 for further details.

Two separate corpus files were then created, one that contained all blog posts for the 'older' three platforms (*i.e.*, Facebook, Twitter, and YouTube) and another containing all blog posts for the 'newer' three platforms (*i.e.*, TikTok, Discord, and Telegram). This was done for purposes of comparing the older platforms that have been around longer and are therefore sometimes assumed to have greater experience in and capacity to counter violent extremism and polarization with those platforms that were founded around a decade later and are sometimes assumed to be less responsive. Each corpus was then entered into AntConc, which is a freeware, multi-platform, multi-purpose corpus analysis toolkit [4]. Word lists were generated, and the stop words removed for each corpus.

The next stage of the analysis was to apply the social capital framework. This was done in three steps. In step one, these word lists were used to manually identify words that were employed by the platforms to communicate the main efforts, initiatives, and decisions that they claim to make regarding building resilience and countering violent extremism and polarization. AntConc's concordance tool was used to investigate the context of the words identified. If the context was deemed relevant to building resilience and countering violent extremism and polarization then the word was entered into a table, along with how many times it appeared in the corpus. All of these were then coded.

Coding "is the process by which a qualitative analyst links specific codes to specific data segments" and a code is "a textual description of the semantic boundaries of a theme or a component of a theme" [5]. Step two was to compile the social capital dictionary by associating each relevant word to the appropriate type(s) of social capital (*i.e.*, bonding, bridging, or linking), with some terms used in different contexts resulting in their being assigned to more than one social capital type. The final step was to identify themes in the data. A theme is "a unit of meaning that is observed (noticed) in the data by a reader of the text" [6]. This resulted in the identification of seven themes regarding what the platforms claimed to do to try to counter polarization and extremism on their sites. Each of the seven themes are discussed at some length in the next section. See Appendices A, B, and C for the final social capital dictionary tables that were produced at the conclusion of these three steps.

4. Findings and discussion

4.1. Quantity of relevant blog posts

In terms of raw numbers, Facebook had the highest number of blog posts in the period studied, the highest number of relevant posts, and the highest number of words per relevant blog posts. In contrast, Telegram had the fewest overall blog posts, the fewest relevant posts, and the fewest number of words of text in our dataset (see Table 1). Table 1 shows that fully 41 percent of Facebook and Twitter's blog posts during the period in which data was collected discussed their efforts to counter polarization and violent extremism, YouTube's attention to these issues was only just over half this (23 percent), however. Although TikTok had a large number of blogs posts overall, the percentage of these addressing countering polarization and extremism were similar to the other newer platforms, Discord and Telegram. Having said this, the total word count for relevant TikTok posts was just 683 words fewer than the total word count for relevant YouTube posts, even though the latter's blog is over a decade older than TikTok's [7].

Table 1: Numbers of blog posts and word counts: September 2017 to August 2020.

Company	**Year company established**	**Year of first blog post**	**Total numbers of blog posts during collection period**	**Numbers of relevant blog posts during collection period (percentage of overall posts)**	**Numbers of words per relevant blog posts**
Facebook	2004	2006	498	203 (41%)	168,303
YouTube	2005	2010	196	46 (23%)	39,130
Twitter	2006	2006	260	107 (41%)	74,414
Totals for older platforms			***954***	***356 (37%)***	***281,847***
Telegram	2013	2014	36	6 (17%)	2,991
Discord	2015	2015	60	8 (13%)	12,047
TikTok	2016	2018	410	66 (16%)	38,447
Totals for newer platforms			***506***	***80 (16%)***	***53,485***
Overall totals			**1,460**	**436 (30%)**	**335,332**

It’s probably relatively unsurprising that closing in on half (41 percent) of all blogs published by Facebook and Twitter in the period studied mention their efforts to counter polarization and extremism. They have had longer to learn about how their platforms are exploited, how to respond to this, and have greater capacity to do so. YouTube’s much lower level of attention to these issues may strike some as more surprising but aligns with the accusation by Douek that YouTube is “flying firmly under the radar” and taking a strategy of keeping “its head down and sort of let[ting] the other platforms take the heat” (Douek, 2020b). The newer platforms’ smaller percentages of blog posts regarding countering efforts may be expected given their much more recent origins and potentially also less knowledge and capacity to make such efforts; also, in the case of Telegram, less willingness to do so. Having said this, despite being newer, TikTok posted significantly more blogs than the two other newer platforms, with many of these being quite lengthy, communicating with their users their efforts to counter polarization and extremism.

This suggests that TikTok is at least keen to be seen as making efforts in this area; also, that platform size, rather than merely age, is an important factor.

While they have varying attitudes to and capacities regarding countering polarization and extremism on their platforms, all six of the platforms we studied appear to be ‘willing and working on it’ as regards generating social capital on their platforms to counter the latter based on the information provided in their blogs. They did however differ in terms of levels of commitment, knowledge, and capacity across the seven identified themes due to their already mentioned differences, and the functions and workings of their particular platforms.

4.2. Social capital themes

Seven themes emerged from analysis of the selected Blog posts. These were granting user powers; strengthening community; the provision of information and education; building community; enhancing user rights; keeping users safe; and building trust and relationships with users. While some of these themes closely align with just one category of social capital, other themes overlap different categories. Table 2 shows the type(s) of social capital associated with each theme and how frequently the words for each theme appeared in the social capital dictionary. Each of the seven themes is then addressed separately in some detail.

Table 2: Social capital dictionary word frequency per theme.

Theme	Social capital types encompassed	Word frequencies for ‘older’ platforms	Word frequencies for ‘newer’ platforms
User powers	Bonding	650 (0.2%)	521 (1.0%)
Strengthening community	Bonding, Bridging	465 (0.2%)	94 (0.2%)
Information and education	Bridging	1,796 (0.6%)	378 (0.7%)
Building community	Bridging	1,736 (0.6%)	401 (0.7%)
Enhancing user rights	Bridging	560 (0.2%)	203 (0.4%)
Keeping users safe	Linking	5,504 (2.0%)	713 (1.3%)
Building trust and	Linking	960 (0.3%)	127 (0.2%)

relationships with users			

4.2.1. Granting user powers

The theme of granting user powers was reflected in the use of terms — such as ‘controls’, ‘tools’, ‘block’, ‘report’, ‘mute’, and ‘settings’ — and falls under the category of bonding social capital. Upon investigation of the context of such words, it was identified that platforms were implementing tools that allow users to make more decisions for themselves regarding their user experience. For example, users can block or mute accounts that they do not wish to see or engage with. They can also restrict the kind of content and comments that they want to see. Overall, this theme was identified across all the platforms but was more prominent on the newer platforms, particularly TikTok, than the older platforms (see Table 2).

The following examples demonstrate that TikTok associated such user powers with increasing ‘safety’, ‘well-being’, and ‘privacy’:

> “This post is part of our Community Well-Being series that aims to educate users on how to customize their TikTok experience using the various safety, privacy, and well-being tools available to them” (TikTok, January 2019).

> “From filters to moderators, we do a lot on our end to minimize the opportunity for misuse on TikTok, but we’re also focused on building you tools and settings that let you take control of your own feed. For example, we understand that there are words some people may view as harmless but others perceive as harmful — so we created a ‘Filter Comments’ tool that allows you to make a custom list of keywords that will be automatically blocked from any comments on your videos” (TikTok, April 2019).

Facebook, on the other hand, tended to describe these powers as a way for users to be able to spend time focusing only on content relevant and of interest to them.

> “With features like Unfollow, Hide, Report, and See First, we’ve consistently worked towards helping people tailor their News Feed experience, so the time they spend on Facebook is time well spent” (Facebook, December 2017).

> "Even though we work to show you the most relevant posts on News Feed, we don't always get it right. That's why we've designed features like See First, Hide, Unfollow, Snooze, and now, Keyword Snooze. We hope that with additional options to help tailor your News Feed experience, you'll be able to spend more time focusing on the things that matter" (Facebook, June 2018).

Scholars have criticized users' exclusion from policy-making and moderation decisions (Klonick, 2017); these efforts appear to be one way that platforms are trying to provide users with greater autonomy. Basak, *et al.* (2019) argue that granting users these powers allows them to overcome the time lag in reporting abusive content and accounts to platforms and waiting for the platforms to respond. Muting or restricting content also has the benefit of shielding users from potentially harmful and harassing content that they would otherwise be exposed to (Basak, *et al.*, 2019). This gives users more control over the tone of a conversation and as regards stopping it from going off-topic (Elder, 2020). Allowing the user, rather than the platform, to decide what content or comments are removed from the users' view allows the user to see content and comments that they may want to respond to or rebut (Basak, *et al.*, 2019). Users therefore have more control over their social connections and can protect boundaries that are important to them (Elder, 2020).

Some users have certainly used these new powers. A study undertaken by the Pew Research Centre found that 31 percent of social media users reported changing their settings because they wanted to reduce their viewing of fellow users' politics-related content and 27 percent blocked or unfriended a user for the same reason (Duggan and Smith, 2016). When asked why they took these steps, 60 percent of users said it was because they found the other user's posted content offensive. Other survey responses to the same question were that the other user posted too much political content, the survey respondent disagreed with the other user's posted content, or the latter was abusive or harassing. The same research found that "social media users with high levels of political engagement take an active approach to curating the content they consume and the users they are connected to," with 42 percent of these highly politically engaged users changing their settings to see fewer posts from another user because of political disagreement [8]. On the other hand, 35 percent of social media users in this study reported that interactions with those with opposing political views were interesting and informative (Duggan and Smith, 2016). This is in keeping with the findings of Bozdag's (2020) interviews with social media users, which also found that some users choose not to implement such settings in order to try to understand alternatives views.

Such powers could potentially also reduce the volume of content that is reported to content moderators and thus reduce their enormous workload as well as relieve them of the responsibility of trying to decide what is best for the user (Basak, *et al.*, 2019). The following example from Twitter's official Blog supports this argument:

> "If you've reported an account or Tweet to us, it will take longer than normal for us to get back to you. We appreciate your patience as we continue to make adjustments. Because these automated systems don't have all of the context and insight our team has, we'll make mistakes. If you think we've made a mistake, you can let us know and appeal here. We appreciate your patience as we work to keep our teams safe, while also making sure we're protecting everyone on Twitter. You can always continue to use hide replies, mute, block, reply filters, and the other tools we offer you to control conversations on the service" (Twitter, March 2020).

These user powers are not without drawbacks, however. There is the possibility that users heavily reliant on muting, blocking, and similar may become incapable of interacting with users with opposing views raising concerns of isolation, echo chambers and filter bubbles (Elder, 2020). Elder (2020) points to the intrinsic value of interpersonal relationships and connections, but also that they:

> "require patience, steadfastness, and loyalty, even when an interaction is not immediately rewarding. Technologies that make disconnection easier and less visible — thereby sparing the person the social penalty that might be garnered by an evasion — can pose a moral hazard, a temptation to behave badly. And if acted on often enough, this sort of evasion can become habitual, gradually degrading individual character and interpersonal relationships" [9]

An alternative argument put forth by Elder (2020) is that, in some instances, dominant voices must be removed so that new arguments from new perspectives can be heard. Allowing users to block or mute dominant voices may therefore make it easier for them to be exposed to diverse other views.

So, while granting users powers has the potential to build bonding social capital, there are conflicting arguments when it comes to whether this has positive or negative consequences. On the one hand, user powers are important to allow users to protect themselves from abuse and harassment or anything else that may negatively affect their wellbeing. They may also be used to 'keep the peace' (Elder, 2020) and could allow new voices to be heard, voices that would otherwise be drowned out by those dominating the conversation. On the other hand, users can choose to use these powers to shield themselves from difference, even though those differences could be valuable and worth respecting, and ultimately result in lack of exposure to other views, social exclusion and polarization (Elder, 2019). Doing so could result in a particularly active and vocal group being able to dominate a public conversation which could give the impression that their views are more popular than they are (Elder, 2020). This is problematic because Elder (2020) claims that people are reluctant to voice their views when they perceive themselves to be in the minority, thereby skewing the conversation. Ultimately, user powers allow user experiences to be under constant construction, which may increase or decrease polarization (Bozdag, 2020) and extremism.

4.2.2. Strengthening communities

The strengthening communities theme focused on companies' efforts at strengthening existing communities on their platforms and included terms such as 'admins' and 'community leaders', the context of which was the platforms providing support, help, and guidance on how to manage and moderate safe communities. There was also a focus on the 'local': supporting local communities and local businesses/organizations, prioritizing local news, and following local laws. Finally, the platforms discussed their recommendation systems, informing readers that these are based on trying to show users content that is relevant to them. Some of these discussions did, however, include trying to diversify the recommendation systems to ensure that users see a diverse array of content. It is argued that all these efforts fall under the category of bonding social capital but could also affect bridging capital depending on whether the community consists of strong or weak ties. This theme was detectable across all six platforms and was consistent across older and newer platforms, including being the weakest of the seven themes across both (see Table 2).

It should be noted that just half of the platforms in this study (*i.e.*, Facebook, Discord, Telegram) have 'admins', that is users that have more powers over other users regarding what is posted in a specific group or server. Facebook and Discord appear to perceive 'admins' as partners who help them keep communities safe. Some examples are:

> "Building relationships with Group Admins — We loved getting to know group leaders at the Facebook Communities Summit, and we got to learn a lot more about the groups they manage ..." (Facebook, October 2017).
>
> "Online admin education resources: To help admins learn how to keep their communities safe and engaged, we've created an online learning destination. It includes tutorials, product demos, and case studies, all

> drawn from the experience and expertise of other admins ...” (Facebook, May 2018).
>
> “With group permissions, admins can now restrict all members from posting specific kinds of content. Or even restrict members from sending messages altogether, let the admins chat amongst themselves while everybody else witness their wisdom in silent awe.” (Telegram, January 2019).
>
> “We want to be your partners and help you build and manage your community. We want to make it easier for you to run successful servers, letting you spend more time connecting with and building your community. As we go forward, we want to work with you every step of the way.” (Discord, July 2020).

Leskovec, *et al.* [10] raise the point that “the overall behaviour of a social media site is generally driven by the collective activity of a large population, but in many cases these sites are also guided by a much smaller group of core participants who are strongly committed to the success of the site”. Scholars have argued that social media admins set the agenda for interaction on their pages (Poell, *et al.*, 2016) and have “a disproportionate degree of influence on movement communication, and thus also on the choreographing of its actions” [11]. The relationship between users and page admins is different to that of users and friends (Poell, *et al.*, 2016). As such, admins can be considered ‘connective leaders’ who “put content into context, turn information into communication, give sense and meaning to the chaotic richness brought by mass peer-production” [12]. Admin leadership consists of “triggering, shaping and incorporating user contributions” [13]. This raises similar pros and cons to those raised in respect to granting user powers.

Tech platforms provide tools and support to admins and community leaders to manage and moderate communities with the purpose of keeping those in the community safe from abusive and harassing users and harmful content. While some groups and admins may encourage and seek to build weak ties however, there is the risk that certain views will be favored and dominate the conversation. Such groups may only attract like-minded people or block entry to those wishing to join with alternative views. Therefore, there is the potential for these tools and supports to result in positive bonding social capital, and even bridging capital, but there is also the potential for exclusion and polarization. For example, admins and private groups proved important during the COVID-19 pandemic, which led not just to people spending more time than ever online but may also have led “to extended time spent in closed groups” (Centre for Resilient and Inclusive Societies, 2021), some of which were benign or even altruistic, but others of which were conspiratorial, hateful, and/or otherwise extremist (see, for example, Baker, 2022).

Also noticeable in the blog posts studied was the role of recommender algorithms. Two prominent topics were platforms’ efforts to prioritize posts from friends and other content thought to be directly relevant to individual users, including high-quality news sources. For example:

> “With this update, we will also prioritize posts that spark conversations and meaningful interactions between people. To do this, we will predict which posts you might want to interact with your friends about, and show these posts higher in feed. These are posts that inspire back-and-forth discussion in the comments and posts that you might want to share and react to — whether that’s a post from a friend seeking advice, a friend asking for recommendations for a trip, or a news article or video prompting lots of discussion” (Facebook, January 2018).

Some of the platforms did, however, acknowledge the need to ensure that users are seeing a diverse range of content:

> “Diversifying is essential to maintaining a thriving global community, and it brings the many corners of TikTok closer together. To that end, sometimes you may come across a video in your feed that doesn’t appear to be relevant to your expressed interests or have amassed a huge number of likes. This is an important and intentional component of our approach to recommendation: bringing a diversity of videos into your For You Feed gives you additional opportunities to stumble upon new content categories, discover new creators, and experience new perspectives and ideas as you scroll through your feed” (TikTok, June 2020).

Facebook and YouTube also discussed their efforts to try to safeguard their recommendation systems:

> “... we believe that limiting the recommendation of these types of [borderline and/or misinformational] videos will mean a better experience for the YouTube community. To be clear, this will only affect recommendations of what videos to watch, not whether a video is available on YouTube” (YouTube, January 2019).

> “Recommendations can help you discover things you love, but since recommended content doesn’t come from accounts you choose to follow, it’s important that we have certain standards for what we recommend.

> This helps ensure we don't recommend potentially sensitive content to those who don't explicitly indicate that they wish to see it. To be clear, this content is still allowed on our platforms, we just won't show it in places where we recommend content" (Facebook, August 2020).

In fact, Facebook's January 2018 update to its News Feed algorithm, aimed at increasing interactions with family and friends (*i.e.*, increasing bonding capital) resulted in increased polarization by amplifying the most divisive content (Hagey and Horowitz, 2021). Even before this, 2016 research — that became public only in 2020 — by a Facebook-employed sociologist "found extremist content thriving in more than one-third of large German political groups on the platform. Swamped with racist, conspiracy-minded and pro-Russian content, the groups were disproportionately influenced by a subset of hyperactive users." Most of the groups were private or secret and, the research found, Facebook's algorithms were responsible for their growth, with a presentation of the 2016 research stating "64% of all extremist group joins are due to our recommendation tools," with most of the activity arising from Facebook's "Groups You Should Join" and "Discover" algorithms. "Our recommendation systems grow the problem," the presentation read (Horwitz and Seetharaman, 2020). This points to the way in which attempts at increasing online social capital may have negative, not just positive, outcomes. In this case, Facebook, whatever their intentions, was complicit in strengthening hateful and extremist communities.

At the same time, the older tech platforms, in particular, discussed their efforts to support a variety of civic causes on and offline, provide training to local businesses to improve their digital skills and follow local laws. Facebook was especially active as regards making these kinds of posts, for example:

> "Over the next few days we'll be at the Columbus Athenaeum hosting training sessions, interactive workshops and speakers, all focused on helping local businesses and non-profits boost their digital skills. All of the courses are free and available to everyone in the community no matter what your skill level. You'll also get to hear from inspiring local entrepreneurs who will share what they learned about starting a new business" (Facebook, August 2018).

> "For the first time, we're inviting leaders of local business and nonprofit [sic] communities to the event, in addition to people leading communities on Facebook Groups, Pages and Fundraisers ... Attendees will learn from each other and hear announcements from Facebook about the latest tools and programs we're building to support them. Workshops and small

> group sessions will offer new skills to help leaders define and fulfill the purpose and vision of their communities" (Facebook, October 2018).

Whilst the described activity could contribute to increasing ties in existing communities it could also increase linking social capital because the tech platforms are providing users with knowledge and resources that they may not previously have had that may in turn increase trust and relationships between the platforms and their users.

4.2.3. Provision of information and education

Provision of information and education was signaled by the use of terms such as 'education', 'awareness', and 'digital literacy', which were used in the context of educating users on prohibited content, how to identify content that seeks to misinform, the importance of context when deciding what content should be removed, and how to stay safe online. The terms 'authoritative' and 'credible' were also identified as platforms discussed their efforts to promote authoritative and credible content and sources. Further, the platforms emphasized that they want their efforts to lead to content and conversations becoming more 'meaningful'. It is argued that these efforts fall under the category of bridging social capital because it seeks to provide users with new information and resources that they may not already have access to in their bonding social capital contexts. Again, this theme emerged consistently across older and newer platforms (see Table 2).

Digital literacy initiatives educate users to critically evaluate and understand the structures and syntax of content, manage new social norms, and recognize when strategies that aim to misinform and polarize are at play (Kidron, *et al.*, 2018). This could help spark users' interests in new social issues thus potentially inviting and encouraging the making or strengthening of weak ties. Examples from the blog posts include:

> "In EMEA, @TwitterDublin hosted UNICEF Ireland and 50 high school students for a special all-day event focused on digital literacy and active citizenship. Students were taught how to verify information sources, safeguard their online reputation and break down digital divides. Guest speakers talked about their learning experience with online platforms and how they've come to develop the knowledge required to effectively leverage Twitter to advocate on issues they're passionate about" (Twitter, November 2017).

> "Today we're launching our Digital Literacy Library, a collection of lessons to help young people think critically and share thoughtfully online. There are 830

> million young people online around the world, and this library is a resource for educators looking to address digital literacy and help these young people build the skills they need to safely enjoy digital technology" (Facebook, August 2018).
>
> "The 'Be Informed' series addresses an important building block for an informed online experience: media literacy. Being media literate means having the ability to access, analyze, evaluate, create, and act using all forms of communication. The "Be Informed" series builds on our previous safety videos ('You're in Control'), a series which highlighted TikTok's safety features, to now provide advice on how to evaluate content online and use those skills and TikTok's in-app features to help protect against the inadvertent spread of misleading information" (TikTok, July 2020).

One problem with some of these efforts is that they require users to actively read and engage with the resources provided as well as deem them credible and trustworthy, many of whom may not do so.

4.2.4. Building community

Terms combined into the building community theme were 'connecting' and 'participation' in 'global' communities. There was encouragement of 'discourse', 'dialogue', 'conversation', and user 'voices' on the platform. There was also discussion of 'partnerships' and 'collaborations' that platforms are involved in to try and make their sites a safer place for people to connect, including support and encouragement for 'diversity' and 'tolerance' on their sites. It is argued that these efforts fall under the category of bridging social capital because they seek to create a safe place for diverse global communities and, as with the previous theme, there were very similar levels of treatment of these issues across the six studied platforms (see Table 2).

In relevant blog posts, the platforms highlighted the efforts they are making to create collaborations and partnerships with various parties and bodies that they can learn from and work with (*e.g.*, other tech platforms, academia, NGOs, and CSOs, etc.) to make their platforms safer spaces for users to connect, voice their opinions and have conversations on a global scale about global issues. If done well, this could help increase and strengthen weak ties among users. This theme was apparent across the six studied platforms' blog posts:

> "Moreover, we spend a significant amount of time evaluating our policies within the Trust and Safety team, the company as a whole, and talking through

> potential changes with trusted NGOs to ensure our policies are fair.” (Discord, February 2019).
>
> “... we continue to meet with and learn from civil society who are intimately familiar with trends and tensions on the ground and are often on the front lines of complex crisis. To improve communication and better identify potentially harmful posts, we have built a new tool for our partners to flag content to us directly. We appreciate the burden and risk that this places on civil society organizations, which is why we’ve worked hard to streamline the reporting process and make it secure and safe” (Facebook, June 2019).
>
> “In the summer of 2017, Facebook, Microsoft, Twitter and YouTube came together to form the Global Internet Forum to Counter Terrorism (GIFCT). Since then, the organization has grown, with nine technology companies working together to disrupt terrorists’ and violent extremists’ abilities to promote themselves, share propaganda and exploit digital platforms to glorify real-world acts of violence ...” (Simultaneously across Facebook, Twitter and YouTube’s official blogs, December 2019).

Having said this, platforms are often criticized for these same efforts. For example, GIFCT has been condemned for increasing the power of member companies whilst lacking sufficient oversight, transparency and accountability (Douek, 2020a). Another example is Twitter’s Trust and Safety Council, which composed more than 40 experts and organizations, which reportedly only worked well for a short time before communication broke down and the company stopped consulting with it (Matsakis, 2019).

4.2.5. Enhancing user rights

The enhancing user rights theme included terms such as ‘free speech’, ‘free expression’, ‘civil rights’, and ‘human rights’ used in the context of the platforms valuing these and portraying them as at the root of their policy decisions and moderation efforts. The platforms sought to demonstrate their efforts in this area via examples of them supporting various groups in society, for example, the LGBTQ+ community, Black community, etc. It is argued that these efforts fall under the category of bridging social capital because it seeks to enhance the rights of all users and raise awareness of issues that are faced by different groups in society. The theme was slightly more prominent as regards the newer platforms than the older (see Table 2).

The platforms raised awareness of movements (*e.g.*, Black Lives Matter, Juneteenth, Pride, Hispanic Heritage Month, International Women's Day), which could provide spaces for education, connections, and conversation to flourish. Also falling into this category were Facebook's Human Rights Impact Assessment [14] and Civil Rights Audit [15] and efforts by platforms to nuance their moderation to respect difference:

> "Celebrate Black History Month with us by visiting youtube.com/spotlight on your phone and swiping over to the Reels tab. And after watching the Reels, tell us who inspire you with the hashtag #CreateBlackHistory" (YouTube, February 2018).
>
> "In May we accepted the call to undertake a civil rights audit. We asked Laura Murphy, a highly respected civil rights and civil liberties leader, to guide the audit. After speaking with more than 90 civil rights organizations, today Laura is providing an important update on our progress" (Facebook, December 2018).
>
> "As a platform used by hundreds of millions of people around the world, we have a responsibility to use our reach to help those who use their voices to advocate for change and support civic engagement and social justice. Starting next week, we'll begin to use our in-product screens and our blog to raise awareness of anti-racist causes and encourage you to take concrete action, such as calling on local officials to advocate for police reform" (Discord, June 2020).

It's worth noting too that, like with previous themes, this kind of activity by companies may be the sort that some users choose either not to view or can respond to with abuse and harassment.

Also, meriting mentioning here is that TikTok, despite rhetoric to the contrary, was criticized in June 2020 by Black creators and users for temporarily making it appear that videos using #BlackLivesMatter and #GeorgeFloyd hashtags had no views. The company put this down to a technical glitch, but later that summer again came under fire for flagging #BlackLivesMatter and #BlackSuccess-related hashtags as "inappropriate" but allowing hashtags such as #whitesupremacy and #whitesuccess, which they put down to a hate speech detection error (Spangler, 2020):

> "We're working to incorporate the evolution of expression into our policies and are training our moderation teams to better understand more nuanced content like cultural appropriation and slurs. If a

> member of a disenfranchised group, such as the LGBTQ+, Latinx, Asian American and Pacific Islander, Black and Indigenous communities, uses a slur as a term of empowerment, we want our moderators to understand the context behind it and not mistakenly take the content down ... ” (TikTok, August 2020).

4.2.6. Keeping users safe

Efforts to keep users safe were the most prominently signposted across both older and newer platforms’ blogs. Terms such as ‘community guidelines’, ‘community standards’, ‘policies’, and ‘rules’ were collected under this heading. Additionally, platforms discussed ‘technology’, ‘automation’, ‘artificial intelligence’, and ‘machine learning’. Further terms identified under this theme were ‘takedowns’, ‘remove’, ‘suspended’, ‘banned’, ‘labeled’, ‘reducing’, ‘ranking’, ‘blocking’, and ‘fact checking’. Table 2 shows that the frequency of these words was considerable higher than those for all other themes across both older and newer platforms blog posts, but also that the difference in levels of attention between the two types of platforms was also considerable here, with the older platforms paying closer attention to user safety than the newer.

It is argued that these efforts fall under the category of linking social capital because it is an attempt by the platforms to build trust and transparency with its users around trying to keep them safe. Its greater prominence across the older platforms (particularly in providing more specific technical details) may be because the creation and implementation of such technologies requires an enormous input of resources and expertise. TikTok and Discord also showed efforts in this area, however:

> “Today, we will start enforcing updates to the Twitter Rules announced last month to reduce hateful and abusive content on Twitter. Through our policy development process, we’ve taken a collaborative approach to develop and implement these changes, including working in close coordination with experts on our Trust and Safety Council” (Twitter, December 2017).

> “We are constantly working to balance aggressive policy enforcement with protections for users. And we see real gains as a result of this work: for example, prioritization powered by our new machine learning tools have been critical to reducing the amount of time terrorist content reported by our users stays on the

> platform from 43 hours in Q1 2018 to 18 hours in Q3 2018” (Facebook, November 2018).
>
> “YouTube has always had rules of the road, including a longstanding policy against hate speech. In 2017, we introduced a tougher stance towards videos with supremacist content, including limiting recommendations and features like comments and the ability to share the video. This step dramatically reduced view to these videos (on average 80%). Today we’re taking another step in our hate speech policy by specifically prohibiting videos alleging that a group is superior in order to justify discrimination, segregation or exclusion based on qualities like age, gender, race, caste, religion, sexual orientation or veteran status ...” (YouTube, June 2019).
>
> “If Trust and Safety can confirm a violation, the team takes steps to mitigate the harm. The following are actions that we may take on either users and/or servers: removing the content, warning users to educate them about their violation, temporarily banning for a mixed amount of time as a ‘cool-down’ period, permanently banning users from Discord and making it difficult for them to create another account, removing a server from Discord, disabling a server’s ability to invite new users ...” (Discord, August 2019).
>
> “We also actively work to learn and get feedback from experts, like those on our Content Advisory Council and civil society organizations. Our industry hasn’t always gotten these decisions right, but we are committed to learning from the mistakes of others’ and our own. We expect to be held accountable for any shortcomings and progress; by working together, we will continue to improve our policies, processes, and products that keep TikTok a place where everyone feels welcome” (TikTok, August 2020).

Telegram blog posts about these issues are conspicuously absent because the platform prioritizes speech over safety or, put another way, its founder and top decision-maker (Loucaides, 2022), Pavel Durov, believes free speech is the ultimate safety. Telegram received what Durov called, on his personal Telegram channel, “maybe the largest digital migration in human history” (Durov, 2021a) in January 2021 becoming the most downloaded mobile app in the world that month (Durov, 2021b). Contributing to this was an influx of far- and extreme right users in the wake of their deplatforming by other platforms in the wake of the 6 January events at the U.S. Capitol. The leader of the Proud Boys — which has since been proscribed as a terrorist organisation by Canada — Enrique Tarrio, sang Telegram's praises on his Telegram channel: "Welcome, newcomers, to the darkest part of the Web. You can be banned for spamming and porn. Everything else is fair game" (Schwirtz, 2021). Having said this, the company has taken steps against Islamic State content [16] and blocked hundreds of extreme right user posts calling for violence ahead of the U.S. presidential inauguration on 20 January 2021 (Schwirtz, 2021). At the end of 2021, Telegram nevertheless remained a preferred app of Islamic State supporters and extreme right users and the above-described deplatforming was not addressed on Telegram's corporate blog.

4.2.7. Building user trust and relationships

The final identified theme was platform efforts to build user trust and have a relationship with their users. Terms such as 'consultations', 'feedback', and 'transparency reports' were used in the context of communicating with users amongst other stakeholders about the platforms' policy-making and moderation decisions. Other terms were 'appeals' and 'accountability', in the context of users being able to appeal decisions that they think are erroneous. Finally, platforms also discussed 'requests' made by governments. It is argued that these efforts fall under the category of linking social capital because, once again, it is an attempt by the platforms to build trust and transparency with its users. Again too, this theme was fairly consistent across the older and newer platforms (see Table 2).

Discussed within relevant blog posts were companies' efforts to consult with and gain feedback from users, as well as an array of other stakeholders (*e.g.*, CSOs, academics, etc.):

> "More recently, people told us they were getting too many similar recommendations, like seeing endless cookie videos after watching just one recipe for snickerdoodles. We now pull in recommendations from a wider set of topics on any given day, more than 200 million videos are recommended on the homepage alone" (YouTube, January 2019).

> "We will undertake a third-party audit by an organization active in researching the spread of hate and racism to observe how Discord works, how we enforce our policies, and to make recommendations for us to be more effective. And we'll share what we learn so others in the industry can make use of their expertise" (Discord, June 2020).

> "We made updates to these tools recently, but we heard feedback from people that they can still be hard to understand and difficult to navigate. Today we're

> making two additional changes to address those concerns" (Facebook, July 2019).
>
> "We want our community to know that we're listening to their feedback, and we're working to increase transparency into the reasons content may be removed. For example, we recently released a feature that notifies users if they duet or react to a video that was removed for violating our Community Guidelines. This feature was built in response to feedback from users who made duets condemning other content; without clarity, they often felt betrayed to find their own video removed, which would happen because the original video they duetted with was taken down" (TikTok, August 2020).

The platforms' posts regarding publishing transparency reports, implementing appeal mechanisms and addressing how they respond to government requests may serve to build trust and stronger relationships with users. However, these efforts may not be effective if transparency reports omit information or are too vague. Further, these reports and mechanisms must be easily accessible, credible and trustworthy to users for this to work. Some companies' appeal mechanisms were limited during the COVID-19 pandemic (*e.g.*, Facebook, Instagram), which is problematic because this is how the majority of erroneously banned content gets reinstated. This has further importance because erroneous removals have been known to disproportionately affect Arab and Muslim communities and can therefore increase the risk of exclusion and polarization (Windwehr and York, 2020).

Facebook's Oversight Board [17] is a prominent attempt to change the course of tech platform governance. It is also an example of attempts to build linking social capital, particularly given the argument that the latter cannot exist without a democratic environment (Ali, *et al.*, 2019). Facebook has created a Board — the first of its kind — that consists of 40 members from a diverse array of backgrounds and expertise that will select and review some of Facebook's most contentious content moderation cases. The Board is funded and supported by an independent company instead of by Facebook itself. Facebook ran consultations to gain feedback from users, amongst a range of other stakeholders, during the development stage of the Board (Harris, 2019). Claims that the Board is independent, empowered, accessible, and transparent, are thought to be an attempt by the platform to build legitimacy, trust, and transparency with users:

> "We are also crafting bylaws which will provide greater operational detail on the board's institutional independence and rules of procedure. These bylaws will include accountability mechanisms, such as a code

> of conduct and board member disqualifications. They will also elaborate on the processes for assembling panels, developing case materials and implementing board decisions. While we are preparing these bylaws on the board's behalf, ultimately the board alone will have the ability to change them. As part of our overall transparency efforts, trust documents will be publicly released, and these will establish the formal relationship between the board, the trust and Facebook" (Facebook, September 2019).
>
> "The members announced today reflect a wide range of views and experiences. They have lived in over 27 countries, speak at least 29 languages and are all committed to the mission of the Oversight Board. We expect them to make some decisions that we, at Facebook, will not always agree with, but that's the point: they are truly autonomous in their exercise of independent judgement. We also expect that the board's membership itself will face criticism. But its long-term success depends on it having members who bring different perspectives and expertise to bear" (Facebook, May 2020).

There are, however, mixed reactions to the implementation of the Board, which in turn, could determine the success of its use in building linking social capital. On the one hand, it is argued that such a Board provides users more participation than they have had before (Klonick, 2019). Users will be able to band together to submit large numbers of appeals to increase their chances of getting their case heard, potentially leading to the Board recommending changes to Facebook's policies accordingly (Klonick, 2019; Douek, 2019). Such a board will also ensure that when decisions are made, these will have taken relevant contextual factors and any competing values into consideration, and the user will receive an explanation for the decision, a process that users are deprived of when their content is removed via automated means (Douek, 2019). Having this access to an independent appeal increases the power of user voices and platform transparency and accountability and thus linking capital.

On the other hand, there is a concern that Facebook could decide at any point to disband the board altogether for any number of reasons (Douek, 2019). Another concern is that the Board will simply create more of the same opaque censorship but via a form of independent governance that could try to block or avoid 'real' government regulation. A further critique is that this is an easy way for the platform to divert responsibility away from themselves and avoid having to take the blame for decisions (Clegg, 2019; Douek, 2019; Klonick, 2019). This adds a concern that the platform will be less incentivized to moderate content or will go in the opposite direction and leave up reported content to ensure it does not become the Board’s jurisdiction (Klonick, 2019).

Furthermore, although the aim of the Board is to make Facebook safer via independent oversight and accountability, the Board could be targeted by those seeking to exclude and polarize. For example, the system may be abused by trolls coordinating to report a specific type of content that they oppose and to overwhelm the system (Klonick, 2019). In fact, something akin to this happened in December 2021 following an announcement by Twitter that images or video of private individuals shared via the service without their permission (*i.e.*, ‘doxxing’) would be removed on request. In subsequent days, what Twitter described as a surge in “coordinated and malicious reports” from far-right activists against anti-extremism accounts monitoring neo-Nazis and white supremacists and documenting attendees at hate events occurred, seeking to get these accounts suspended and far right users’ images of themselves removed (BBC News, 2021).

Douek (2019) argues that one of the reasons why this kind of governance structure is likely to appeal to Facebook is because it will add an air of legitimacy to its content moderation decisions, which could in turn, aid user relations. However, it could be argued that for the Board to be seen as legitimate, real independence must be guaranteed as well as decisions in line with international legal standards for freedom of speech albeit there are concerns that the Board may not be well-equipped to do so (Clegg, 2019). Douek (2019) points out that, “Zuckerberg is not recreating liberal democratic governance. He is not subjecting himself or his role to democratic accountability. But the FOB initiative is in keeping with Zuckerberg’s long-standing pronouncements that Facebook is ‘more like a government than a traditional company’.” Klonick (2019) concurs, pointing out that users only have the right to access the Board’s system, not to be able to provide input. Thus “while users might expect *democratic* accountability, a more realistic outcome is *participatory* empowerment” [18].

Summing up, the Oversight Board could pave the way for governance with greater legitimacy, transparency, and accountability. If done poorly, the board could erode any trust that has been built with users and result in worse, rather than better, platform-user relations.

5. Conclusion

The aim of this research was to examine the efforts that a range of older and newer tech platforms (*i.e.*, Facebook, Twitter, YouTube, TikTok, Discord, and Telegram) take to grow bonding, bridging, and linking capital with the aim of building resilience and countering polarization and extremism in the communities on their sites. Previous research found that although social media can create all three types of social capital, it was most likely to grow bridging social capital. Much of this research is now around a decade old and highly user- and Facebook-focused, however. Tech platforms have evolved enormously since this time, with a whole ecosystem of diverse platforms now being exploited by those seeking to polarize and/or encourage extremism. This research sought to add to, and broaden, the research already undertaken in this area by, firstly, a cross-platform approach of a diverse sample of platforms, and secondly, by making the research platform-focused as opposed to user-focused.

Our compilation of a social capital dictionary based on posts on the selected platforms' corporate blogs identified seven main themes related to the companies' efforts to build social capital: granting user powers, strengthening existing communities, provision of information and education, building community, enhancing user rights, keeping users safe, and building trust and relationships with users. Analysis of these showed that, similar to previous studies, while creation of all three types of social capital (*i.e.*, bonding, bridging, linking) was apparent, bridging capital dominated here too. Also, while there were some differences between social capital generating activities and their framings on older versus newer platforms, other factors, including platform size and company values were found to be equally or more important.

Discrepancies in how the different platforms addressed the identified themes, both in terms of how they were understood and portrayed, as well as how frequently platforms posted about them are worth commenting on further here. An example of the former is 'granting user powers'. TikTok portrayed this as an effort to help users increase their safety, privacy, and well-being whereas Facebook implemented similar user powers but portrayed them as a way to increase users' ability to focus on what matters to them and save them time, rather having to scroll through content that they do not want to see. Then in the 'strengthening community' theme, there is the use of 'admins' by some platforms but not others. Also under this theme is the use of recommendation systems. As compared to the other platforms, Facebook and YouTube put a big focus on explaining how they safeguard their recommendations albeit there are increasing questions about the efficacy of these efforts.

In the building community theme, the older platforms' founding of the GIFCT creates a significant focus on collaborations and partnerships compared with the newer platforms. Regarding the 'enhancing user rights' theme, Facebook's efforts gain prominence due to their Civil Rights Audit and Human Rights Impact Assessments, both of which likely required considerable resources that may make equivalent processes less achievable for newer platforms. This theme also underscored TikTok's prominent focus on the work they are doing to support minorities. This may in part be due to the platform's user base being younger than some of the other platforms studied. Finally, regarding the 'keeping users safe' and 'building trust and relationships with users' themes, the older platforms have had a lot longer to generate the expertise and resources required to create and implement technologies and publish transparency reports than the newer platforms. This is particularly the case with Facebook’s Oversight Board. Also at play here however is Telegram’s inattention to user safety in their corporate blog posts.

Overall, we found that the platforms have different underlying priorities as regards where they believe their efforts should lie in regard to the identified themes and thus social capital production. This is most likely based on the platforms’ chosen values (*e.g.*, Telegram’s commitment to free speech), the audiences they are trying to attract (*e.g.*, TikTok’s younger audience) and their capacity (*e.g.*, resources, expertise, number of staff, financial turnover, etc.) to undertake such efforts. These differences, particularly the latter, are arguably not sufficiently considered in the context of the regulatory demands increasingly being made on platforms. This research provides some insight into the ways in which platforms are similar to but also different from one another and how this can feed into their responses to countering bad actors on their sites. It sheds light too on why a one-size-fits-all legislative approach is unlikely to be effective (see Watkin [2021] for further work in this area). Additionally, it could be expected that older platforms would, as a result of having had longer periods of time and a higher capacity to undertake such efforts, have been found to have published about these efforts more than newer platforms. This was only the case across some themes, however. This suggests there could be some benefit from older and newer platforms collaborating and sharing best practices in the realms of resilience-building and countering polarization and extremism.

The unintended negatives consequences of, on the face of them, good faith initiatives at building social capital are also worth noting. These include creating exclusion and isolation, as well as removing the need to be tolerant of other views (*e.g.*, by users simply removing them from their feeds). Further, there is the potential, particularly with linking social capital, that a failure to secure perceptions of legitimacy or a break of trust could result in worse rather than better user-platform relations. These unintended consequences have the potential to have quite negative outcomes given the earlier mentioned findings that these are circumstances that extremist groups and movements have been known to exploit (Brisson, *et al.*, 2017; Pickering, *et al.*, 2007). Platforms must therefore be more mindful of safeguarding their users from unintended consequences that may arise from any efforts they implement to try to counter polarization and extremism on their sites. Finally, it is recognized that while there is certainly a public-relations element to the blog posts, there is still much value in these large datasets from a researcher perspective.

About the authors

Amy-Louise Watkin is Lecturer in Criminal Justice at the University of the West of Scotland.
E-mail: amy-louise [dot] watkin [at] uws [dot] ac [dot] uk

Maura Conway is Paddy Moriarty Professor of Government and International Studies in the School of Law and Government, Dublin City University, Ireland; Visiting Professor of Cyber Threats at CYTREC, Swansea University, U.K.; and the Coordinator of VOX-Pol.
E-mail: maura [dot] conway [at] dcu [dot] ie

Acknowledgements

Thanks to the participants in the EU H2020-funded BRaVE project (Grant Agreement No. 822189) workshops in the context of which this research was originally devised and presented and those who contributed engaged in discussion and contributed suggestions in the seminars hosted by the School of Law and Government at Dublin City University and the Hilary Clinton School of Law at Swansea University at which we also presented this work.

Notes

[1] Brisson, *et al.*, 2017, p. 8.

[2] De Zúñiga, *et al.*, 2017, p. 61.

[3] https://about.fb.com/news/; https://blog.twitter.com/en_us.html; https://blog.youtube/; https://newsroom.tiktok.com/en-us/; https://telegram.org/blog; https://blog.discord.com/.

[4] For more on this, see https://www.laurenceanthony.net/software/antconc/.

[5] Guest, *et al.*, 2012, p. 3.

[6] *Ibid.*

[7] It should be noted that both Twitter and TikTok's official Blogs were navigated differently to the other platforms. While most of the platforms have one set of blogs for their global user base, Twitter and TikTok allow users to filter blog posts by country, with the United States set as the default. On Twitter, the majority of the U.S. blog posts were exact replicas of U.K. blog posts. On TikTok, U.S. and U.K. blog posts were not exact replicas, but similar to each other. We collected the relevant U.S. and U.K. blog posts for both Twitter and TikTok because both were

written in English, the U.S. was the default on both companies' blogs, and collecting U.K. posts ensured a wider data scope. This accounts for TikTok having so many blogs; no exact replicates were collected, however.

8. Duggan and Smith, 2016, p. 4.

9. Elder, 2020, p. 18.

10. Leskovec, *et al.*, 2010, p. 98.

11. Gerbaudo, 2012, p. 140.

12. Della Ratta and Valeriani, 2012, p. 56.

13. Poell, *et al.*, 2016, p. 1,009.

14. See https://about.fb.com/news/2018/11/myanmar-hria/.

15. See https://about.fb.com/news/2018/12/civil-rights-audit/.

16. Conway, 2020, pp. 19–20.

17. See https://oversightboard.com/.

18. Klonick, 2019, p. 2,490.

References

M. Ali, N. Azab, M.K. Sorour, and M. Dora, 2019. "Integration v. polarisation among social media users: Perspectives through social capital theory on the recent Egyptian political landscape," *Technological Forecasting and Social Change*, volume 145, pp. 461–473.
doi: https://doi.org/10.1016/j.techfore.2019.01.001, accessed 11 April 2022.

S.J. Baele, B. Lewys, and T.G. Coan, 2020. "Uncovering the far-right online ecosystem: An analytical framework and research agenda," *Studies in Conflict & Terrorism* (30 December).
doi: https://doi.org/10.1080/1057610X.2020.1862895, accessed 11 April 2022.

S.A. Baker, 2022. "Alt. Health Influencers: How wellness culture and Web culture have been weaponised to promote conspiracy theories and far-right extremism during the COVID-19 pandemic," *European Journal of Cultural Studies*, volume 25, number 1, pp. 3–24.
doi: https://doi.org/10.1177/13675494211062623, accessed 11 April 2022.

R. Basak, S. Sural, N. Ganguly, and S.K. Ghosh, 2019. “Online public shaming on twitter: Detection, analysis, and mitigation,” *IEEE Transactions on Computational Social Systems*, volume 6, number 2, pp. 208–220.
doi: https://doi.org/10.1109/TCSS.2019.2895734, accessed 11 April 2022.

BBC News, 2021. “Far-right target critics with Twitter’s new media policy” (6 December), at https://www.bbc.com/news/technology-59547353, accessed 11 April 2022.

M.C. Benigni, K. Joseph, K.M. Carley, 2017. “Online extremism and the communities that sustain it: Detecting the ISIS supporting community on Twitter,” *PLoS ONE*, volume 12, number 12, e0181405 (1 December).
doi: https://doi.org/10.1371/journal.pone.0181405, accessed 11 April 2022.

d. boyd, 2009. “Taken out of context: American teen sociality in networked publics,” *SSRN* (24 February).
doi: https://doi.org/10.2139/ssrn.1344756, accessed 11 April 2022.

C. Bozdag, 2020. “Managing diverse online networks in the context of polarization: Understanding how we grow apart on and through social media,” *Social Media + Society* (14 December).
doi: https://doi.org/10.1177/2056305120975713, accessed 11 April 2022.

J. Brisson, V. Gerrand, K. Hadfield, and P. Jefferies, 2017. *Understanding youth resilience to violent extremism: A standardised research measure: Final research report*. Burwood, Victoria, Australia: Alfred Deakin Institute for Citizenship and Globalisation, Deakin University.

M. Burke, R. Kraut, and C. Marlow, 2011. “Social capital on Facebook: Differentiating uses and users,” *CHI ’11: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 571–580.
doi: https://doi.org/10.1145/1978942.1979023, accessed 11 April 2022.

M. Burke, C. Marlow, and T. Lento, 2010. “Social network activity and social well-being,” *CHI ’10: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 1,909–1,912.
doi: https://doi.org/10.1145/1753326.1753613, accessed 11 April 2022.

Centre for Resilient and Inclusive Societies, 2021. “Inquiry into extremist movements and radicalism in Australia,”
at https://www.aph.gov.au/Parliamentary_Business/Committees/Joint/Intelligence_and_Security/ExtremistMovements, accessed 11 April 2022.

H.-T. Chen and X. Li, 2017. “The contribution of mobile social media to social capital and psychological well-being: Examining the role of communicative use,

friending and self-disclosure,” *Computers in Human Behavior*, number 75, pp. 958–965.
doi: https://doi.org/10.1016/j.chb.2017.06.011, accessed 11 April 2022.

N. Clegg, 2019. “Global feedback & input on the Facebook Oversight Board for content decisions,” at https://about.fb.com/wp-content/uploads/2019/06/oversight-board-consultation-report-2.pdf, accessed 11 April 2022.

J.S. Coleman, 1988. “Social capital in the creation of human capital,” *American Journal of Sociology*, number 94, supplement, pp. S95–S120.
doi: https://doi.org/10.1086/228943, accessed 11 April 2022.

M. Conway, 2020. “Violent extremism and terrorism online in 2019: The year in review,” *VOX-Pol*, at http://www.voxpol.eu/download/vox-pol_publication/Violent-Extremism-and-Terrorism-Online-in-2019-The-Year-in-Review.pdf, accessed 11 April 2022.

M. Conway, A.L. Watkin, and S. Looney, 2021. “Violent extremism and terrorism online in 2021: The year in review,” *VOX-Pol*,
at http://www.voxpol.eu/download/vox-pol_publication/RAN-Policy-Support_Prevent_Consolidated_Year-in-Review-2021.pdf, accessed 11 April 2022.

H.G. de Zúñiga, M. Barnidge, and A. Scherman, 2017. “Social media social capital, offline social capital, and citizenship: Exploring asymmetrical social capital effects,” *Political Communication*, volume 34, number 1, pp. 44–68.
doi: https://doi.org/10.1080/10584609.2016.1227000, accessed 11 April 2022.

D. Della Ratta and A. Valeriani, 2012. “Remixing the Spring!: Connective leadership and read-write practices in the 2011 Arab uprisings,” *CyberOrient*, volume 6, number 1, pp. 52–76, and
at https://cyberorient.net/2012/05/10/remixing-the-spring-connective-leadership-and-read-write-practices-in-the-2011-arab-uprisings/, accessed 11 April 2022.

J. Donath and d. boyd, 2004. “Public displays of connection,” BT Technology Journal, volume 22, number 4, pp. 71–82.
doi: https://doi.org/10.1023/B:BTTJ.0000047585.06264.cc, accessed 11 April 2022.

E. Douek, 2020a. “The rise of content cartels,” *Knight First Amendment Institute at Columbia University* (11 February), at https://knightcolumbia.org/content/the-rise-of-content-cartels, accessed 11 April 2022.

E. Douek, 2020b. “We hardly ever talk about YouTube and disinformation. Not anymore,” *Marketplace* (17 December), at https://www.marketplace.org/shows/marketplace-tech/we-hardly-ever-talk-about-youtube-and-disinformation-not-anymore/, accessed 11 April 2022.

E. Douek, 2019. “Facebook’s Oversight Board: Move fast with stable infrastructure and humility,” *North Carolina Journal of Law & Technology*, volume 21, number 1, pp. pp. 1–77, and at https://scholarship.law.unc.edu/ncjolt/vol21/iss1/, accessed 11 April 2022.

M. Duggan and A. Smith, 2016. “The political environment on social media,” *Pew Research Center* (25 October), at https://www.pewresearch.org/internet/2016/10/25/the-political-environment-on-social-media/, accessed 11 April 2022.

P. Durov, 2021a. “Since my last post, the already massive influx of new users to Telegram has only accelerated ...,” *Durov’s Channel, Telegram* (14 February), at https://t.me/s/durov?q=%E2%80%9Cthe+largest+digital+migration+in+human+history%E2%80%9D, accessed 11 April 2022.

P. Durov, 2021b. “Telegram became the most downloaded mobile app in the world in January 2021 ...,” *Durov’s Channel, Telegram* (8 February), at https://t.me/durov/152, accessed 11 April 2022.

A. Elder, 2020. “The interpersonal is political: Unfriending to promote civic discourse on social media,” *Ethics and Information Technology*, volume 22, pp. 15–24.
doi: https://doi.org/10.1007/s10676-019-09511-4, accessed 11 April 2022.

N.B. Ellison, C. Steinfield, and C. Lampe, 2011a. “Connection strategies: Social capital implications of Facebook-enabled communication practices,” *New Media & Society*, volume 13, number 6, pp. 873–892.
doi: https://doi.org/10.1177/1461444810385389, accessed 11 April 2022.

N.B. Ellison, J. Vitak, C. Steinfield, R. Gray, and C. Lampe, 2011b. “Negotiating privacy concerns and social capital needs in a social media environment,” In: S. Trepte and L. Reinecke (editors). *Privacy online: Perspectives on privacy and self-disclosure in the social Web*. Berlin: Springer, pp. 19–32.
doi: https://doi.org/10.1007/978-3-642-21521-6_3, accessed 11 April 2022.

N.B. Ellison, C. Steinfield, and C. Lampe, 2007. “The benefits of Facebook ‘friends’: Social capital and college students’ use of online social network sites,” *Journal of Computer-Mediated Communication*, volume 12, number 4, pp. 1,143–1,168.
doi: https://doi.org/10.1111/j.1083-6101.2007.00367.x, accessed 11 April 2022.

K.A. Frank and J.Y. Yasumoto, 1998. "Linking action to social structure within a system: Social capital within and between subgroups," *American Journal of Sociology*, volume 104, number 3, pp. 642–686.
doi: https://doi.org/10.1086/210083, accessed 11 April 2022.

F. Fukuyama, 2001. "Social capital, civil society and development," *Third World Quarterly*, volume 22, number 1, pp. 7–20.
doi: https://doi.org/10.1080/713701144, accessed 11 April 2022.

A. Gallagher, C. O'Connor, P. Vaux, E, Thomas, and J. Davey, 2021. "Gaming and extremism: The extreme right on Discord," *Institute for Strategic Dialogue*, at https://www.isdglobal.org/isd-publications/gaming-and-extremism-the-extreme-right-on-discord/, accessed 11 April 2022.

P. Gerbaudo, 2012. *Tweets and the streets: Social media and contemporary activism*. London: Pluto Press.

M.S. Granovetter, 1973. "The strength of weak ties," *American Journal of Sociology*, volume 78, number 6, pp. 1,360–1,380.
doi: https://doi.org/10.1086/225469, accessed 11 April 2022.

G. Guest, K.M. MacQueen, and E.E. Namey, 2012. "Themes and codes," In: G. Guest, K.M. MacQueen, and E.E. Namey. *Applied thematic analysis*. Los Angeles, Calif.: Sage.
doi: https://dx.doi.org/10.4135/9781483384436.n3, accessed 11 April 2022.

A. Gunnestad, 2006. "Resilience in a cross-cultural perspective: How resilience is generated in different cultures," *Journal of Intercultural Communication*, number 11, at http://hdl.handle.net/11250/2564077, accessed 11 April 2022.

K. Hagey and J. Horwitz, 2021. "Facebook tried to make its platform a healthier place. It got angrier instead,&edquo; *Wall Street Journal* (15 September), at https://www.wsj.com/articles/facebook-algorithm-change-zuckerberg-11631654215, accessed 11 April 2022.

B. Harris, 2019. "Getting input on an oversight board," *Facebook Newsroom* (1 April), at https://about.fb.com/news/2019/04/input-on-an-oversight-board/, accessed 11 April 2022.

R.L. Hawkins and K. Maurer, 2010. "Bonding, bridging and linking: How social capital operated in New Orleans following Hurricane Katrina," *British Journal of Social Work*, volume 40, number 6, pp. 1,777–1,793.
doi: https://doi.org/10.1093/bjsw/bcp087, accessed 11 April 2022.

J. Horwitz and D. Seetharaman, 2020. “Facebook executives shut down efforts to make the site less divisive,&edquo; *Wall Street Journal* (26 May), at https://www.wsj.com/articles/facebook-knows-it-encourages-division-top-executives-nixed-solutions-11590507499, accessed 11 April 2022.

H. Hwang and K.-O. Kim, 2015. “Social media as a tool for social movements: The effect of social media use and social capital on intention to participate in social movements,” *International Journal of Consumer Studies*, volume 39, number 5, pp. 478–488.
doi: https://doi.org/10.1111/IJCS.12221, accessed 11 April 2022.

B. Kidron, A. Evans, and J. Afia, 2018. “Disrupted childhood: The cost of persuasive design,” *5Rights Foundation*, at https://5rightsfoundation.com/static/5Rights-Disrupted-Childhood.pdf, accessed 11 April 2022.

B. Kim and Y. Kim, 2017. “College students’ social media use and communication network heterogeneity: Implications for social capital and subjective well-being,” *Computers in Human Behavior*, number 73, pp. 620–628.
doi: https://doi.org/10.1016/j.chb.2017.03.033, accessed 11 April 2022.

L.J. Kirmayer, M. Sehdev, R. Whitley, S.F. Dandeneau, and C. Isaac, 2009. “Community resilience: Models, metaphors and measures,” *Journal of Aboriginal Health*, volume 5, number 1, pp. 62–117, and at https://www.mcgill.ca/mhp/files/mhp/community_resilience.pdf, accessed 11 April 2022.

K. Klonick, 2019. “The Facebook Oversight Board: Creating an independent institution to adjudicate online free expression,” *Yale Law Journal*, volume 129, number 8, pp. 2,418–2,499, and at https://www.yalelawjournal.org/feature/the-facebook-oversight-board, accessed 11 April 2022.

K. Klonick, 2017. “The new governors: The people, rules, and processes governing online speech,” *Harvard Law Review*, volume 131, pp. 1,598–1,670, and at https://harvardlawreview.org/2018/04/the-new-governors-the-people-rules-and-processes-governing-online-speech/, accessed 11 April 2022.

E. Kubin and C. von Sikorski, 2021. “The role of (social) media in political polarization: A systematic review,” *Annals of the International Communication Association*, volume 45, number 3, pp. 188–206.
doi: https://doi.org/10.1080/23808985.2021.1976070, accessed 11 April 2022.

L. Larsen, S.L. Harlan, B. Bolin, E.J. Hackett, D. Hope, A. Kirby, A. Nelson, T.R. Rex, and S. Wolf, 2004. “Bonding and bridging: Understanding the relationship between social capital and civic action,” *Journal of Planning Education and*

Research, volume 24, number 1, pp. 64–77.
doi: https://doi.org/10.1177/0739456X04267181, accessed 11 April 2022.

J. Leskovec, D. Huttenlocher, and J. Kleinberg, 2010. "Governance in social media: A case study of the Wikipedia promotion process," *Proceedings of the Fourth International AAAI Conference on Web and Social Media*, volume 4, number 1, pp. 98–105, and at https://ojs.aaai.org/index.php/ICWSM/article/view/14013, accessed 11 April 2022.

L. Lima, J.C.S. Reis, P. Melo, F. Murai, L. Araujo, P. Vikatos, and F. Benevenuto, 2018. "Inside the right-leaning echo chambers: Characterizing Gab, an unmoderated social system," *ASONAM '18: Proceedings of the 2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, pp. 515–522, and at https://ieeexplore.ieee.org/document/8508809, accessed 11 April 2022.

D. Loucaides, 2022. "How Telegram became the anti-Facebook," *Wired* (8 February), at https://www.wired.com/story/how-telegram-became-anti-facebook/, accessed 11 April 2022.

M. Marozzi, 2016. "Construction, robustness assessment and application of an index of perceived level of socio-economic threat from immigrants: A study of 47 European countries and regions," *Social Indicators Research*, volume 128, number 1, pp. 413–437.
doi: https://doi.org/10.1007/s11205-015-1037-z, accessed 11 April 2022.

L. Matsakis, 2019. "Twitter Trust and Safety advisers say they're being ignored," *Wired* (23 August), at https://www.wired.com/story/twitter-trust-and-safety-council-letter/, accessed 11 April 2022.

J. Mignone and J. O'Neil, 2005b. "Conceptual understanding of social capital in First Nation communities: An illustrative description," *Pimatisiwin: A Journal of Aboriginal & Indigenous Community Health*, volume 3, number 2, pp. 7–44, at http://www.centroetnosalud.com/trabajos/ConceptualUnderstanding%20of%20Social%20Capital%20in%20Pimatisiwin.pdf, accessed 11 April 2022.

D. Narayan, 1999. "Bonds and bridges: Social capital and poverty," *World Bank, Poverty Reduction and Economic Management Network, Poverty Division*, at https://documents1.worldbank.org/curated/en/989601468766526606/107507322_20041117172515/additional/multi-page.pdf, accessed 11 April 2022.

D. O'Callaghan, D. Greene, M. Conway, J. Carthy, and P. Cunningham, 2015. "Down the (white) rabbit hole: The extreme right and online recommender systems," *Social Science Computer Review*, volume 33, number 4, pp. 459–478. doi: https://doi.org/10.1177/0894439314555329, accessed 11 April 2022.

S. Pickering, D. Wright-Neville, J. McCulloch, and P. Lentini, 2007. "Counter-terrorism policing and culturally diverse communities, Australian Research Council Linkage Project final report," *Monash University and Victoria Police*, at https://www.monash.edu/__data/assets/pdf_file/0011/1672805/counterterrorreport-07.pdf, accessed 11 April 2022.

T. Poell, R. Abdulla, B. Rieder, R. Woltering, and L. Zack, 2016. "Protest leadership in the age of social media," *Information, Communication & Society*, volume 19, number 7, pp. 994–1,014. doi: https://doi.org/10.1080/1369118X.2015.1088049, accessed 11 April 2022.

A. Portes, 1998. "Social capital: Its origins and applications in contemporary sociology," *Annual Review of Sociology*, volume 24, pp. 1–24. doi: https://doi.org/10.1146/annurev.soc.24.1.1, accessed 11 April 2022.

R. Putnam, 2001. "Social capital: Measurement and consequences," *Canadian Journal of Policy Research*, volume 2, number 1, pp. 41–51.

R. Putnam, 2000. *Bowling alone: The collapse and revival of American community*. New York: Simon & Schuster.

K. Quinn, 2016. "Contextual social capital: Linking the contexts of social media use to its outcomes," *Information, Communication & Society*, volume 19, number 5, pp. 582–600. doi: https://doi.org/10.1080/1369118X.2016.1139613, accessed 11 April 2022.

S.A. Raza, W. Qazi, and A. Umer, 2017. "Facebook is a source of social capital building among university students: Evidence from a developing country," *Journal of Educational Computing Research*, volume 55, number 3, pp. 295–322. doi: https://doi.org/10.1177/0735633116667357, accessed 11 April 2022.

E. Rochko, 2019. "Gab switches to Mastodon's code: Our statement," *Mastodon* (4 July), at https://blog.joinmastodon.org/2019/07/statement-on-gabs-fork-of-mastodon/, accessed 11 April 2022.

M. Schwirtz, 2021. "Telegram, pro-democracy tool, struggles over new fans from far right," *New York Times* (26 January), at https://www.nytimes.com/2021/01/26/world/europe/telegram-app-far-right.html, accessed 11 April 2022.

R. Scrivens and A. Amarasingam, 2020. "Haters gonna 'like': Exploring Canadian far-right extremism on Facebook," In M. Littler and B. Lee, (editors). *Digital*

extremisms: Readings in violence, radicalisation and extremism in the online space. Cham, Switzerland: Palgrave Macmillan, pp. 63–89.
doi: https://doi.org/10.1007/978-3-030-30138-5_4, accessed 11 April 2022.

T. Spangler, 2020. "TikTok blames 'technical glitch' for suppressing view counts on #BlackLivesMatter, #GeorgeFloyd videos," *Variety* (2 June),
at https://variety.com/2020/digital/news/tiktok-suppressed-view-counts-blacklivesmatter-georgefloyd-videos-1234622975/, accessed 11 April 2022.

C. Steinfield, N.B. Ellison, and C. Lampe, 2008. "Social capital, self-esteem, and use of online social network sites: A longitudinal analysis," *Journal of Applied Developmental Psychology*, volume 29, number 6, pp. 434–445.
doi: https://doi.org/10.1016/j.appdev.2008.07.002, accessed 11 April 2022.

L. Tolsma and Z. Zevallos, 2009. *Enhancing community development in Adelaide by building on the social capital of South Australian Muslims*. Melbourne, Australia: Institute for Social Research, Swinburne University of Technology.

S. Walther and A. McCoy, 2021. "US extremism on Telegram: Fueling disinformation, conspiracy theories, and accelerationism," *Perspectives on Terrorism*, volume 15, number 2, pp. 100–124, and
at https://www.universiteitleiden.nl/binaries/content/assets/customsites/perspectives-on-terrorism/2021/issue-2/walther-and-mccoy-.pdf, accessed 11 April 2022.

A.-L. Watkin, 2021. "Regulating terrorist content on tech platforms: A proposed framework based on social regulation," Ph.D. thesis, Swansea University.
doi: https://doi.org/10.23889/SUthesis.57985, accessed 11 April 2022.

A.-L. Watkin, 2019. "Considering the whole ecosystem in regulating terrorist content and hate online," *E-International Relations* (18 September),
at https://www.e-ir.info/2019/09/18/considering-the-whole-ecosystem-in-regulating-terrorist-content-and-hate-online/, accessed 11 April 2022.

G. Weimann and N. Masri, 2020. "Research note: Spreading hate on TikTok," *Studies in Conflict & Terrorism* (19 June).
doi: https://doi.org/10.1080/1057610X.2020.1780027, accessed 11 April 2022.

S. Windwehr and J. York, 2020. "Facebook's most recent transparency report demonstrates the pitfalls of automated content moderation," *Electronic Frontier Foundation* (8 October), at https://www.eff.org/deeplinks/2020/10/facebooks-most-recent-transparency-report-demonstrates-pitfalls-automated-content, accessed 14 January 2020.

M. Woolcock and D. Narayan, 2000. "Social capital: Implications for development theory, research, and policy," *World Bank Research Observer*, volume 15, number

2, pp. 225–249.
doi: https://doi.org/10.1093/wbro/15.2.225, accessed 11 April 2022.

Appendix A: Bonding social capital dictionary

	Older platforms	Newer platforms
Granting user powers	Tool/tools (81) Report/reported/reporting (300) Block/blocked/blocked (19) Control/controls (95) Settings (48) Restrictions/restricted/restrict (7) Flag/flagging/flagged/flagger (81) Digital well-being (3) Mute (10) Unfollow (6)	Community well-being series (8) Control/controls (33) Settings (64) Safety features (7) Safety, privacy, and well-being tools (9) Privacy and safety settings (5) Report/reporter/reported/reporting (181) You’re in control video/series (12) Privacy settings (14) Privacy controls (8) Tools (59) Digital well-being (14) Remove (9) Restrict/restrictions (19) Restricted mode (12) Customize (10) Filter/filters/filtered (25) Block/blocking/blocked (23) Mute (2) Unfollow (2) Flag (5)
Strengthening existing communities	Community leadership program (5) Community leader/s (25) Local community/communities (17) Admin/admins/administrator/administrators (119) Local business/businesses/organizations (11) Local law/s (24) Local news/publishers (41) Local non-profits (3) [Other?] Local (172) Recommendation/s (48)	Local (28) Local communities (8) Local culture (3) Local law/s (5) Local organizations (3) Recommendation/s (39) Admins (8)

Appendix B: Bridging social capital dictionary

	Older platforms	**Newer platforms**
The provision of education/information	Inform/Information (918) Top news and breaking news (4) Help Centre/Center (22) Media and information literacy (17) Media literacy (62) Digital literacy (41) Digital and media literacy (5) Context (163) Context button (16) Meaningful (69) Connect/connecting/connections (126) Authoritative (60) Educate/education/educational (100) Authentic/authentically/authenticity (86) Awareness (44) Safer Internet Day (22) Credible (41)	Safer Internet Day (8) World Health Organization (12) Digital literacy (3) Media literacy (9) Inform/Information (102) Authoritative (5) Credible (3) Authentic/authentically/authenticity (22) Be informed series (4) Safety videos (14) Internet Matters (14) Learn/learning (52) Educate/education/educational (68) Awareness (16) Context (15) Connect/connecting (18) #Bettermebetterinternet (2) Poynter Institute MediaWise program (2) Meaningful (9)
Building communities	Inform/Information (918) Top news and breaking news (4) Help Centre/Center (22) Media and information literacy (17) Media literacy (62) Digital literacy (41) Digital and media literacy (5) Context (163) Context button (16) Meaningful (69) Connect/connecting/connections (126) Authoritative (60) Educate/education/educational (100) Authentic/authentically/authenticity (86) Awareness (44) Safer Internet Day (22) Credible (41) Global community (4) Public conversation (101) Public discourse (15) Civic discourse (10)	Build/building (62) Global community (24) Partner/partnered/partnering/partnership (62) Teamed up with (2) Collaborate/collaboration/collaborating/collaboratively (21) Experts (37) Family Online Safety Institute (FOSI) (26) Law enforcement (15) Diverse/diversity (56) Promote/promoting (30) Voice (7) Inclusive (16) Dialogue (6) Tolerance (2) NGOs (7) Civil society (4) Academia (3) Inclusive/inclusion (21)

	Creators for change (22) Law enforcement (122) Global Network Initiative (4) Global Internet forum to counter terrorism (15) Experts (216) Family Online Safety Institute (2) YouTube partner program (8) Partner/partners/partnered/partnerships (477) Teamed up with (10) Civil society (70) NGOs/Non-governmental organization (60) Facebook communities summit (10) Healthy (57) Voice (100) Promote/promoting (119) Diverse/diversity (86) Collaborate/collaboration/collaborating/collaboratively (67) Participation/participate (64) Connectivity (25) Inclusive/inclusion (33) Tolerance (19) Dialogue (9) Institute for strategic dialogue (3) Academia (8)	
Enhancing user rights	Free speech (20) Freedom of speech (4) Freedom of expression (36) Free expression (59) Human right/s (91) Human rights impact assessment (8) Civil rights (68) Civil rights audit (6) Civil liberties (4) Black (50) Black community/communities (21) Black creators (11) Black history month (4) Black lives matter (7) Black-owned businesses (12) Black voices (7) Racial justice (22) Juneteenth (1) Women (57) International Women's Day (2) Hispanic/s (21)	Free speech (3) Freedom of speech (1) Black (31) Black community/communities (15) Black creators (8) Black history month (4) Black lives matter (2) Black voices (4) Juneteenth (10) LGBTQ+ (30) LGBTQ+ community (20) LGBTQ+ creators (4) Queer (5) Indigenous (2) Latinx (2) Native American (1) Women (16) International Women's Day (4) Women's History Month (3) Pride (25) Pride Month (3)

	Hispanic Heritage Month (2) Latino/latinx (15) LGBTQ (12) LGBTQ community (4) LGBTQ creators (3) Pride (9) Allyship (4)	Civic engagement (4) Underrepresented (6)

Appendix C: Linking social capital dictionary

	Older platforms	**Newer platforms**
Keeping users safe	Crisis protocols/Content Incident Protocol/s (8) Community Guidelines (50) Community Standards (154) Policy (920) World leader/s (15) Enforcement (288) Remove/removed/removals/removing (911) Technology/technologies (422) Review/reviews/reviewed/reviewing (392) Fact-check/checking/checker/checked (239) Rules (260) Tool/tools (155) Safety Center (5) Trust and Safety Council (8) Digital fingerprints/hashes (87) Rank/ranked/ranking (68) Detect/detection (261) Proactively/proactive (193) Machine learning (92) Automated (68) Artificial intelligence (42) Reduce/reducing (146) Reviewers (71) Suspend/suspended/suspension/suspensions (103) Label/labels/labeled/labeling (130) Block/blocking/blocked (63) Limit/limiting (52) Prohibit (40)	Community Guidelines (79) Terms of Service (15) User safety (8) Content Advisory Council (9) Moderation (64) Remove/Removals/removal/removed (62) Takedown/s (2) Trust and/& safety (53) Safety Center (22) Policy/policies (128) Ban/bans/banned/banning (62) Technology/technologies (34) Human moderation (6) Moderator/s (19) Enforcement (8) Proactively/proactive (21) Review/reviews/reviewed/reviewing (40) Fact check/checking (8) Detect/detection (11) Rules (10) Warning/s (19) Algorithm/algorithms (7) Delete/deleted (17) Prohibit (6) Suspend (1)

	Ban/bans/banned/banning (89) Moderation (24) Moderator/s (12) Warning/s (26) Takedown/s (23) Demote/demoting (3) Shutdowns (9) Restrictions/restricted/restrict (75)	
Building trust and relationships with users	Ad Library (71) Ads Transparency Center (15) Info and Ads (7) Consultation/consult (39) Data Transparency Advisory Group (6) Transparency Report/s (68) Transparent (62) Oversight Board (30) Request (200) Feedback (152) Ads for good (16) Disclose/disclosure/disclosing/disclosed (112) Accountable/accountability (85) Appeal/s (73) Restore/restored (24)	Transparency Report/s (28) Transparency Center (8) Transparency and Accountability Center (2) Accountable/accountability (10) Transparent (11) Feedback (25) Requests (25) Appeal/s (15) Disclose/disclosed (3)