sociological science

The Diffusion and Reach of (Mis)Information on Facebook During the U.S. 2020 Election

Sandra González-Bailón,[a] David Lazer,[b] Pablo Barberá,[c] William Godel,[c] Hunt Allcott,[d] Taylor Brown,[c] Adriana Crespo-Tenorio,[c] Deen Freelon,[a] Matthew Gentzkow,[d] Andrew M. Guess,[e] Shanto Iyengar,[d] Young Mie Kim,[f] Neil Malhotra,[d] Devra Moehler,[c] Brendan Nyhan,[g] Jennifer Pan,[d] Carlos Velasco Rivera,[c] Jaime Settle,[h] Emily Thorson,[i] Rebekah Tromble,[j] Arjun Wilkins,[c] Magdalena Wojcieszak,[k,l] Chad Kiewiet de Jonge,[c] Annie Franco,[c] Winter Mason,[c] Natalie Jomini Stroud,[m] Joshua A. Tucker[n]

a) University of Pennsylvania; b) Northeastern University; c) Meta; d) Stanford University; e) Princeton University; f) University of Wisconsin-Madison; g) Dartmouth College; h) William & Mary; i) Syracuse University; j) The George Washington University; k) University of California, Davis; l) University of Warsaw; m) University of Texas at Austin; n) New York University

Abstract: Social media creates the possibility for rapid, viral spread of content, but how many posts actually reach millions? And is misinformation special in how it propagates? We answer these questions by analyzing the virality of and exposure to information on Facebook during the U.S. 2020 presidential election. We examine the diffusion trees of the approximately 1 B posts that were re-shared at least once by U.S.-based adults from July 1, 2020, to February 1, 2021. We differentiate misinformation from non-misinformation posts to show that (1) misinformation diffused more slowly, relying on a small number of active users that spread misinformation via long chains of peer-to-peer diffusion that reached millions; non-misinformation spread primarily through one-to-many affordances (mainly, Pages); (2) the relative importance of peer-to-peer spread for misinformation was likely due to an enforcement gap in content moderation policies designed to target mostly Pages and Groups; and (3) periods of aggressive content moderation proximate to the election coincide with dramatic drops in the spread and reach of misinformation and (to a lesser extent) political content.

Keywords: networks, social media, elections, misinformation, content moderation

Reproducibility Package: Deidentified data and analysis code from this study are deposited in the Social Media Archive at ICPSR, part of the University of Michigan Institute for Social Research. The data are available for university IRB-approved research on elections or to validate the findings of this study. ICPSR will receive and vet all applications for data access. Access through the ICPSR Archive ensures that the data and code are used only for the purposes for which they were created and collected. The code would also be more difficult to navigate separately from the data, which is why both are housed in the same space. Website: https://socialmediaarchive.org/collection/US2020.

Citation: González-Bailón, Sandra, David Lazer, Pablo Barberá et al. 2024. "The Diffusion and Reach of (Mis)Information on Facebook During the U.S. 2020 Election" Sociological Science 11: 1124-1146.

Received: September 9, 2024

Accepted: October 24, 2024

Published: December 11, 2024

Editor(s): Arnout van de Rijt, Cristobal Young

DOI: 10.15195/v11.a41

The U.S. 2020 presidential election took place amid heightened concern over the role social media would play in enabling the spread of misinformation. Of special concern was Facebook's role in this process, given its dominance as a platform. This article documents how misinformation spread on Facebook during the 2020 election period focusing on the interplay between affordances for sharing content and shifting content moderation policies. We find that misinformation spread more slowly than non-misinformation, relying far more on narrow, deep chains of user-to-user re-sharing in a context where background content moderation policies were targeting predominantly Pages and Groups, not users. We also document that there were significant variations in the number of large misinformation diffusion events during the election period, likely driven in part by *ad hoc* ("break the glass") content moderation policies instituted by Facebook.

We begin with the observation that platforms create an architecture that allows content to spread in specific ways. Platforms can also purposefully intervene to change how things spread – much in the same way as people sometimes intervene in other types of propagation. For instance, high-speed, long-distance travel fundamentally, if incidentally, changed how disease spreads; and interventions like mask wearing or social distancing became purposeful actions to slow the spread of viruses. In the case of information, affordances provide the technical and social possibilities for content to spread. Content moderation is a purposeful intervention to differentially affect what spreads, typically to slow down the diffusion of problematic content (like misinformation).

A platform like Facebook creates (and changes) the technical possibilities for information diffusion. The fact that Pages can have an unlimited number of followers is a design choice made intentionally. Social practices emerge from those design choices. For instance, the fact that celebrities use Pages to communicate with followers is a social practice built on top of a specific design choice. And content moderation—the set of decisions that a platform makes on which content to demote, remove, or label—is built on top of those design choices and social practices. The purpose of our analyses is to document how sharing dynamics (including the diffusion of misinformation) changed over time as content moderation policies were also shifting, given what we know about the policies that were being implemented by Facebook in the background during the U.S. 2020 election. In what follows, we discuss the technical possibility for diffusion that the design of Facebook circa 2020 created, and we then turn to a brief description of content moderation on Facebook during this time.

Mechanisms for Diffusion on Facebook

Diffusion through the re-sharing of content is intrinsic to social media, and it creates the possibility of person-to-person spreading. When users or Pages re-share content (in the case of users, through their Feed or in Groups), they create cascading events that can be represented as time-evolving networks. Each re-share creates a link in the network, yielding a diffusion tree with the initial post at the root. The shape of a diffusion tree reveals the nature of the spreading process (Nowell and Kleinberg 2008; Friggeri et al. 2014; Goel et al. 2016; Vosoughi, Roy, and Aral 2018). Following this literature, we view diffusion trees as describable along two dimensions: depth and breadth. A tree that is wide has a big burst of sharing at a single level. A tree that is deep results from a persistent pattern of re-sharing that cascades through the network. Thus, a wide but narrow tree would result from a diffusion event where a public figure shares content, and where the followers of that figure re-share that content, but the friends of those followers do not, in turn, re-share. In contrast, a tree that is deep but narrow results from a persistent pattern of a few re-shares cascading through the network, with a small fraction of the re-shares occurring at any given level in the tree. This is the case of most organic diffusion, where the initiator, a common user, elicits shares from a handful of friends, who may also inspire a few friends, and so on. An individual who sees content shared within a tree is not part of the tree but is counted as part of the ”reach” of the tree, which includes all individuals who end up being exposed to the re-shared post.

Social media platforms, among them Facebook, create architectures that may end up favoring certain types of diffusion (e.g., one-to-many) over others (e.g., one-to-few). A key component in that design is the set of affordances that allow or encourage certain behaviors. We define “affordances” as the conjunction of actions the platform makes technically possible for users to do, and users’ understandings of what the platform makes possible (Evans et al. 2016; Ronzhyn, Cardenal, and Rubio 2023). All platforms have a complex set of evolving affordances. Here we focus on the affordances that made it possible for content to spread on Facebook during the 2020 election. At that time, users (as Friends or followers), Pages, and Groups provided the technological possibilities for content to spread. At the time of our analyses, only content posted within a user’s social graph—that is, the set of users that were Friends or accounts that were followed by that user, including Pages and Groups—could appear in their Feed (excluding ads).

The broadcasting potential of these different accounts varies greatly, which introduces by design a source of heterogeneity in their outreach capabilities. Pages, for instance, are the typical mechanism for celebrities, politicians, and brands to share content, and they are unlimited in the number of followers they can accumulate. Unsurprisingly, they do garner very large numbers of followers. The CNN page, for instance, has $\sim$ 40 M followers at the time of writing; Breitbart has $\sim$ 5 M. Groups also have no hard limits on size, but they are still much smaller, on average, than Pages. For instance, one of the most popular Groups in the U.S. circa 2020, Pantsuit Nation, had fewer than 3 M members. When it comes to users, the number of friends they can have is capped at 5,000—which is still much higher than the networks an average person can maintain offline. These affordances, in other words, critically affect how content flows through Facebook. And the nature of these affordances, in turn, means that other affordances, like the re-share button, will have a very different impact depending on who clicks on it, e.g., a celebrity Page or a peripheral user.

Our analyses examine the structure of diffusion as it relates to these growth mechanisms. But, crucially, we also assess the overall impact of diffusion chains by looking at the number of views accumulated during the propagation. Most people online are lurkers (Amichai-Hamburger 2016), which means that most users view but rarely produce or re-share content. The implication is that if we only analyze the visible trails of re-sharing, we are likely substantially underestimating the actual reach of the propagation: Content that is seen by many but re-shared by few will appear as wrongly irrelevant if we only apply the narrow focus of diffusion metrics. None of the past studies that have analyzed diffusion through online networks have information on exposure to the diffusion events they track; they simply lack data on who viewed the content. As a result, prior research may be underestimating the impact of content that does not propagate but still has high reach through broadcasting, or it may be conflating peer-to-peer diffusion with higher reach (when in reality reach is a function of the underlying network size, not just the number of re-shares).

In summary, while there is quite a bit of work that examines diffusion through online networks (e.g., Onnela and Reed-Tsochas 2010; Bakshy, Rosenn et al. 2012; Vosoughi et al. 2018), there is far less research that directly examines how diffusion is related to platform affordances and virtually no research that measures exposure or the number of actual views accumulated by the content diffused. In addition, no prior research has analyzed both diffusion and reach for the full set of posts (including text, URLs, images, and videos) published and re-shared on a social media platform. Past work either examines narrow subsets of content, such as photos or text memes on Facebook (Cheng et al. 2014; Friggeri et al. 2014), or relies on Twitter (now called X) and is therefore forced to infer (rather than measure) the structure of diffusion trees (Goel et al. 2016; Vosoughi et al. 2018; Juul and Ugander 2021). In this work, each retweet could only be matched back to the root node, likely causing measurement and inference errors. In section S3.1 in the online supplement we offer a more extended discussion on these data considerations.

Here, we overcome past limitations while asking new questions about how content moderation interacts with platform affordances in shaping diffusion dynamics and reach. Even when platform design and moderation are not directly analyzed, their effects are always in the background. For example, Vosoughi et al. (2018) find that the predominant pattern of diffusion on Twitter takes the shape of trees that are broad and not deep. This finding of broadcasting likely reflects the heavy-tailed degree distribution that existed on Twitter, which is enabled by the absence of limits on the number of followers an account can have on the platform. This is not the pattern typically associated with friend-to-friend sharing on Facebook, as we discuss below, because there is a hard cap on the number of friends users can have. Social media platforms, in other words, do not offer a clean slate on which to analyze social behavior, including the spread of misinformation. Rather, platforms are complex systems of affordances, practices, policies, and *ad hoc* decisions that are constantly shifting. In such an environment, there is nothing intrinsic to misinformation that will make it spread in a particular way. While there are certainly psychological and sociological factors that will affect decisions on whether and how to share misinformation, misinformation can only spread if it is allowed to move through the channels an information system creates.

Content Moderation on Facebook

Content moderation refers to the set of decisions that a platform makes on which content to demote, remove, or label. It requires evaluating the compliance of content with those standards, as defined by the platform. Content moderation policies go beyond the algorithmic processes that determine content curation, that is, the automated selection of what goes into users' Feeds. Like other platforms, Facebook applies integrity processes to every post that gets published to determine how to rank it, that is, how visible to make it to users. Content moderation, however, refers to the broader policies that determine which content stays up, receives labels with contextual information, or is allowed to be re-shared. It also includes practices around punishment for accounts that violate those policies (for example, deplatforming, McCabe et al. 2024). These policies are ever-changing sets of broad principles and narrow practices, encompassing a multitude of announced and unannounced interventions that operate above and beyond the usual Feed ranking.

Facebook has a wide range of content moderation interventions, some of which are visible and some of which are not. As we discuss below, of particular importance is Meta's "misinformation repeat offender" policy, which at the time of the 2020 election stated that Pages that had received two misinformation "strikes" over a 90-day period and Groups that had received three "strikes" over the same period had the visibility of their subsequent content reduced. These strikes were accrued when a post by these sources was rated ″false″ or "altered" by a third-party fact-checker (3PFC) or contained content matching a post with this same rating. This policy is consequential because, as noted above, Pages and Groups offer the major capacity on the platform for creating one-to-many diffusion events, given that their average degree is many orders of magnitude greater than that of users.

There is only modest literature on the effects of content moderation on the quantity of misinformation circulating on Facebook and little work on the relationship between content moderation and the structure of diffusion events. For example, Broniatowski et al (2023) find that the impact of Facebook's content moderation efforts around vaccine misinformation was, at best, modest, perhaps because the flexibility of Facebook's dissemination architecture allowed evasion of content moderation measures. In Bandy and Diakopoulos (2023) the authors examine a specific "break the glass" measure (discussed in more detail below) instituted after the 2020 election, aimed at increasing the visibility of content from credible sources; their results suggest, again, minimal effects. Most relevant to this article, these authors find that Facebook's stated "repeat offender" policy (which, as mentioned, reduces the visibility of content from Pages and Groups that have shared multiple pieces of misinformation) aligned empirically with observational data suggesting that Groups and Pages experienced lower engagement after sharing multiple pieces of misinformation.

Around Election Day and in the aftermath of the storming of the Capitol on January 6, 2021, Facebook made a number of *ad hoc* decisions that adjusted the principles guiding its content moderation. These decisions have come to be known as "break the glass" measures because, as the name implies, they were interventions designed to respond to extreme circumstances and mitigate risks related to the election. One of these measures, for instance, was called the "virality circuit breaker," and it was specifically focused on reducing the viral spread of misinformation (see S4 in the online supplement for a timeline of interventions). But what do these measures tell us about the power that platforms have to control information flows?

In what follows, we answer this question by reconstructing the diffusion patterns of political content and misinformation and evaluating how those patterns changed during the election period. Our analyses pay special attention to shifts across content moderation regimes, that is, the different periods of low- and high-intensity interventions. The findings suggest that there were sizable decreases in misinformation diffusion (and, to a lesser extent, the diffusion of political content) during periods of high intensity content moderation. However, the data also suggest that the preexisting "repeat offender" policy—aimed at the broadcast affordances of the platform (Pages and Groups), not at the peer-to-peer affordances (primarily Friends)—created an enforcement gap that allowed users to activate the pathways through which misinformation could still flow. Overall, the vast majority of content posted on the platform did not propagate far, but the cumulative reach of all the misinformation diffusion trees, large and small, still amounted to millions.

Study Questions

Our analyses are designed to answer three empirical questions: (1) How prevalent were large diffusion events during the U.S. 2020 election and, within that set, how prevalent was misinformation? (2) How did the structure and rhythm of diffusion vary across mechanisms (affordances) for dissemination? And (3) were content moderation and the set of exceptional rules applied under the "break the glass" umbrella successful at reducing the flow of misinformation (within the context and idiosyncrasies of the historical moment)? These core research questions can be unpacked into more specific ones: (RQ1) How prevalent is broadcast versus viral diffusion? (RQ2) How concentrated are the re-sharing distributions in terms of users generating the diffusion events? (RQ3) Who, in terms of basic demographics, re-shares most political content and most misinformation? (RQ4) Does political content or misinformation generate, on average, larger diffusion trees? (RQ5) What affordances (user accounts, Pages, or Groups) are associated with the diffusion of political content or misinformation? (RQ6) Which affordances are associated with greater reach (number of views) of a given diffusion tree? And (RQ7) how much temporal variation do the data reveal in how diffusion unfolded, given the temporally bounded decisions on the type of posts that content moderation was to target more aggressively?

The answers to these questions allow us to illuminate what happened on one of the most influential communication platforms of our era during a very contentious election and its aftermath. They also provide critical insight into the role that platform design and content moderation play in spreading dynamics and the reach of the posts flowing through the network. Our focus is on political content and misinformation, but the broader implications of our analyses speak to how effective (or not) policies designed to control information flows can be.

Understanding how diffusion dynamics vary across content types and platform affordances is key to design interventions that can limit the effects of misinformation and other harmful content. All social media are subject to (often invisible) content moderation interventions that affect the nature and extent of information spreading. In 2020, Facebook chose to deploy a set of extraordinary interventions, of which the public knows some details only because of leaked reports and investigative journalism (as we detail in S4 in the online supplement; the academic co-authors of this paper requested, and did not receive from Meta, more details about these platform interventions). Here, we offer a rare, descriptive, glimpse into how the implementation of these interventions coincided with clear shifts in how content spread on the platform. In documenting these shifts, we cast light on the role of content moderation in shaping information flows.

Data and Measures

This study is part of a broader collaborative project between Meta and a team of external researchers. The collaboration was initiated in early 2020 to design and produce transparent and reproducible research on the political impact of Facebook and Instagram (see online supplement S1 for a more elaborate description of this collaboration). The data for this article draw from all U.S.-based monthly active adult users, and track exposures to and re-sharing of all posts published (publicly or privately) on the platform from July 1, 2020, to February 1, 2021. In other words, the results we report are based on the full set of user, Page, and Group posts created by U.S.-based adult accounts that were shared at least once ($N \sim 1,024,817,106$ posts, Fig. 1A). To protect privacy, we only have access to tree-level data for the posts that were shared (both privately and publicly) at least $k = 100$ times by U.S. users (1.2 percent or 12.1 M posts). In the online supplement, we offer additional aggregated analyses for the trees that do not meet this threshold to show that our main conclusions regarding overall diffusion patterns hold above and below the $k = 100$ cutting point (see online supplement S6 for additional details on these analyses). The set of large trees that meet the threshold account for 54.5 percent of all views accumulated by all diffusion trees.

We reconstruct the diffusion of these posts in the form of network trees. Tree data structures are hierarchical networks with a root node (i.e., the original post) and nested layers of re-sharing activity (if the post is re-shared). In Figure 1B, we offer a schematic representation of some simple tree structures that can emerge during diffusion events. Some posts are only re-shared by a few users one step removed from the original poster, so the trees are shallow and narrow (network 3); some other posts are re-shared by users a few steps removed from the original poster (e.g., friends of friends of friends), so these trees are deep (network 1); and still other posts are re-shared by many users in each layer of re-share activity, so these trees are wider (networks 2 and 4). We quantify these structural differences using measures of breadth and depth, in addition to tree size (i.e., number of re-shares). The breadth of a tree is the maximum number of re-shares over all depths. The depth is the maximum distance between the root post and all nodes. For instance, network 3 in Figure 1B has size 9 (we do not count the root node), breadth 9, and depth 1; network 1 has size 99, breadth 50, and depth 12.

In our construction of diffusion trees, we define an edge as occurring when a user, who we can call Bret, clicks on the re-share button of a post previously posted by another user, say Alice. As mentioned, this operationalization minimizes measurement error compared to prior studies that inferred re-shares based on exposure proxies (Vosoughi et al. 2018), but it does miss cases when Bret copies

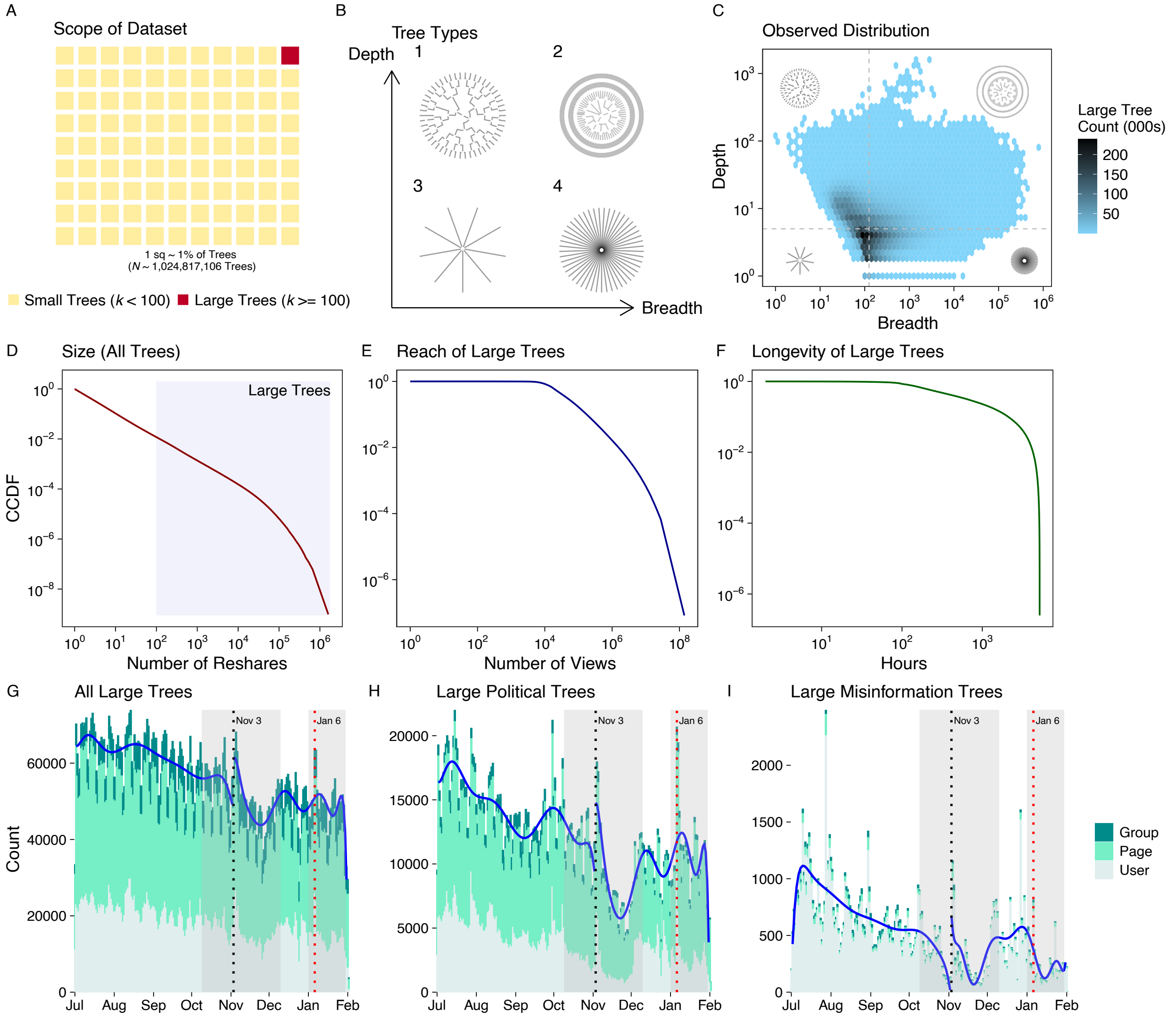

Figure 1: Description of the data. (A) Of all the posts that were re-shared at least once ($N \sim 1.02$ B), only 1.2 percent (or 12.1 M) reached the $k >= 100$ re-shares threshold. We refer to this subset of posts as "large trees" (posts under this threshold are labeled "small trees"). Large trees accumulate 54.5 percent of the views (section S6 in the online supplement shows additional analyses for small trees). (B) Diffusion trees can emerge with different structures that we summarize using their depth (maximum distance from a leaf node to the root of the tree); breadth (maximum number of nodes over all depths); and the measure of structural virality defined in Equation 1. (C) Empirical distribution of large trees (dashed lines indicate the median breadth/depth). (D) Size distribution of all trees; the subset of large trees amounts to $N \sim 12.1$ M (1.2 percent) of all trees. (E) Nearly all posts with $k => 100$ re-shares (99.6 percent) accumulated 1000+ views (1.6 percent of the trees gained 1 M+ views). (F) Most of the trees (67.8 percent) grew during a period of 7+ days. (G) Pages initiate most of the large trees in our data ($N \sim 6.5$ M), (H) including trees classified as political. (I) Misinformation trees, however, are predominantly initiated by user posts. Blue lines in panels G–I are 10th degree polynomials, fitted separately to data before and after Election Day. The shaded rectangles highlight the two periods of high-intensity interventions. There is a steady decline in the count of large misinformation trees, nearly reaching the 0 line just before Election Day (and again visibly accelerating in late November, partially overlapping with the rollback of "break the glass" measures late November to mid-December). The steep drop in late January (panels G and H) is an artifact of the right-censoring of the data.

and pastes content seen from Alice without clicking on the re-share button. It also requires attributing a specific source for the re-share when, in reality, a user may have seen the same content from multiple sources (even if they only clicked on the re-share button of a single post). The Feed algorithm also determines which posts a user sees first (and thus which content is more likely to be re-shared). In following the 'click on the re-share button' rule to build the diffusion trees, we reduce errors in inference, but we also acknowledge that the representation of dyadic spreading dynamics is a simplification of how diffusion processes likely unfold in the real world (for further discussion on these operationalization issues, see online supplement S3.1).

In addition to size, breadth, and depth, we calculate the structural virality index (Goel et al. 2016; Vosoughi et al. 2018), which measures the average distance d between all pairs of nodes in a diffusion tree T (with $n > 1$ nodes) according to this formula:

$$v(T) = \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j=1}^{n} d_{ij}. \tag{1}$$

The distance d_{ij} measures the length of the shortest path between any pair of nodes in the tree. The structural virality of tree 1 in Figure 1B is 7.7; it is 1.8 for tree number 3. This measure allows us to differentiate trees that rely on broadcasting to grow versus trees that grow virally via the accumulation of small re-shares extended through many layers of activity away from the original source. In Figure 1C, we show that high virality is the exception, not the rule.

The subset of large trees is characterized by a long tail in terms of size (Fig. 1D) and even longer tails in terms of reach and longevity (Fig. 1E,F). Long tails in this context mean that there is a very small number of trees with a disproportionate size, reach, or longevity (compared to most other trees). We also characterize the trees using their root source and composition. Figure 1G-I shows the number of trees initiated by users, Pages, and Groups, and the proportion that were classified as political and misinformation. We classify as 'misinformation' content (including text, URLs, images, and videos) that was rated "false" by 3PFCs, a network that includes organizations such as Snopes, Reuters, The Washington Post, Fact Checker, FactCheck.org, and PolitiFact, among others. Once a post is fact-checked as false, Meta propagates the label to similar content (e.g., other posts with the same URL). We note that we do not know why some content was fact-checked and other content not. We know that Facebook surfaces potential misinformation to fact-checkers using signals that are likely correlated with reach "How Meta's third-party fact-checking program works. Accessed May 2, 2024 (https://www.facebook.com/formedia/blog/third-party-fact-checking-how-it-works)." The significant advantage of using the 3PFC ratings is that they are a measure of content rather than content source. Much of the literature (e.g., Grinberg et al. 2019; Guess et al. 2019) uses also content-based measures, but here we use 3PFC ratings that also apply to multi-modal content (e.g., videos, memes), that is, to content that is not linked to specific sources (usually, domains).

To offer a few examples of what type of content was rated as misinformation by 3PFCs: one of the posts consists of the text "Tonight's voting shenanigans..." attached to a video recorded on a cell phone with a woman on the screen complaining about voting manipulation. In the video, she can be heard saying that the poll workers were distributing sharpies (instead of pens) to voters to invalidate their ballots. An off-screen male's voice (likely the voice of the person recording the video) can be heard repeating and emphasizing what the aggrieved woman says. This video has 115K re-shares. Another post labeled as misinformation was created by the Tucker Carlson Tonight show Page. The post consists of the text "Facebook has censored our video with a Chinese whistleblower. Big Tech wants control over the facts you see," added to a 13-minute clip from the Tonight Show with Carlson discussing the origins of the COVID-19 virus. This post was re-shared 107K times. And yet another post is just a block of text with the message: "Kamala Harris supports abortion right up to BIRTH. Full-term, 8-lb babies. Think about that (and you are ok with that?)." This post has 468K re-shares. In all three examples, the posts were visibly labeled as false by Meta (with an explanation of the rationale provided by 3PFCs), but the content could still be accessed and seen.

In total, $N \sim 114,000$ trees are labeled as misinformation. Note that only a very small number of fact-checked trees ($N \sim 2,000$) are labeled true. As we explain in S3.2 of the online supplement, this is likely due to the fact that there is a selection bias in the content that fact-checkers evaluate, a sample that contains more problematic content than truthful content. Likewise, the number of unlabeled trees is $N \sim 12$ M, which means that we are likely underestimating the prevalence of misinformation on the platform (to the degree that it goes unnoticed by the 3PFC program). Importantly, we also note that posts from the Pages of politicians containing the politicians' own claims or statements were ineligible to be fact-checked per Meta's policies (Fact-checking policies on Facebook, Instagram, and Threads. Accessed May 2, 2024 (https://www.facebook.com/business/help/315131736305613?id=673052479947730), see also the discussion in S3.2 in the online supplement). However, other content from politicians that matches content fact-checked when posted by other users would still be labeled as misinformation and demoted. Finally, we note that politicians' Pages were not subject to account-level penalties applied in cases of repeated sharing of misinformation (as identified by Meta's third-party fact-checking partners).

In the online supplement (S5.3), we offer additional analyses using an alternative approach to measuring potential misinformation that relies on identifying 'untrustworthy sources', i.e., sources that repeatedly post misinformation. This definition is inspired by Meta's Misinformation "repeat offender" policy, but its implementation is different. In particular, the 'untrustworthy' label is applied to Pages and Groups that have two or more posts rated 'false' by 3PFCs as well as to domains with two or more URLs rated 'false' by 3PFCs since Meta's "misinformation repeat offender" program began in 2018. Because of this operationalization, the vast majority of user-initiated trees with a 'false' root post are excluded from the analyses that use the 'untrustworthy' label: most user posts do not contain URLs, and users themselves do not receive the 'untrustworthy' label. Given these issues, the analysis of 'untrustworthy' sources is not a robustness test but rather an approximation to one key component of the content moderation policies that were in place during our observation window: systems that only monitor accounts identified as 'untrustworthy' allow a lot of problematic content to fly under the radar. As we explain in section S5.3, while users initiated 89 percent of the $N \sim 114,000$

trees rated 'false', only $N \sim 16,000$ of these trees receive also the label 'untrustworthy', with users accounting for only 3 percent of those trees. In short, a lot of misinformation activity is excluded from consideration when attention is only paid to 'untrustworthy' content as defined.

Back to Figure 1, there is a large decline (∼ 30 percent) in the number of large trees from July 1 to December 1 (panel G). The decline is bigger for political trees (∼ 70 percent, panel H) and even bigger for misinformation trees (∼ 90 percent, panel I). Events of a given day, no doubt, drove significant fluctuation, but the general decline of political and misinformation trees between July and Election Day seems unlikely to be driven by a reduced interest in politics during one of the most hotly contested Presidential elections in U.S. history. This, in turn, begs the question of how Facebook's content moderation policies relate to these shifts. As far as we can tell, the policies pertaining to the Feed integrity and ranking mechanisms did not significantly change during the election period. However, the "break the glass" measures introduced many *ad hoc* content moderation decisions that were taken at specific periods. Based on the timeline of interventions outlined in S4 of the online supplement, we identify two periods of high-intensity measures that were launched to reduce the amount of problematic content on the platform. During these periods, content moderation went far beyond what the Feed algorithm did to rank content.

We identify the cut points around the deployment/deprecation of the "break the glass" interventions following the timeline reconstructed using publicly available sources. We designate the periods before October 9 and between December 10 and before January 1 as "low intensity" and the periods from October 9 to December 10 and after January 1 as "high intensity." We note that these cutoffs are necessarily fuzzy in that: (1) there are many interventions that were launched simultaneously and in a step-wise fashion; and (2) the interventions may affect the growth of trees that start before the cutoff date. The data do suggest temporal shifts that correspond roughly to these dates; however, there are other patterns in the data that suggest important but undocumented changes in practices by Facebook during this period, as we discuss below.

In general, Figure 1I illustrates the point made above when discussing 'untrustworthy' sources: during our observation window, most misinformation was posted by users. (In Figs. S8 to S10 in the online supplement, we show additional analyses confirming that users posting on their profiles are the source of most misinformation.) Most of the posts users publish on Facebook do not contain URLs, which means most user-generated content escapes the net of "repeat offender" policies (even if their specific posts are repeatedly labeled as "false" by fact-checkers). In other words, the fact that users dominate the trends depicted in Figure 1I possibly reflects the deterrent effect of content moderation policies that were predominantly targeting Pages and Groups.

Results

We first consider the relative prevalence of broadcast versus viral diffusion (RQ1). Most of the trees in our data have breadth < 100 and depth < 5 (Fig. 1C), but there is also a long tail of trees that grow significantly wider and deeper than the median tree. This distribution suggests that virality is very unusual: most diffusion chains activate a relatively small number of users close to the original source. The complementary cumulative distribution functions (CCDFs) tell us that there is a small set of trees that grow to accumulate millions of re-shares (Fig. 1D) and millions of views (Fig. 1E). Most of the trees grow during a week of activity or more (Fig. 1F), although, as we discuss in section S5.10 in the online supplement, they grow the most during the first 24 hours.

RQ2 asked about concentration in the re-sharing distributions in terms of users generating the diffusion events. Since our platform data are aggregated at the tree level (a necessary constraint to preserve privacy), we do not know which users are involved in the diffusion events we track or whether the same set of users recurrently appear in the propagation trees. However, in section S5.15 in the online supplement, we leverage the broader project's panel data of recruited participants to estimate the individual-level cumulative distribution function necessary to assess the question of concentration. What these analyses suggest is that most of the re-shares and views accumulated on Facebook are generated by a very small percentage of users. This is especially the case for misinformation, where less than 1 percent of users generate most of the content labeled by 3PFCs as "false."

In Figure 2, we use the CCDFs of the diffusion tree statistics to address RQ4 (does misinformation generate, on average, larger diffusion trees?) and RQ7 (how much temporal variation do the data reveal?) We compare misinformation trees with the subset of trees classified as political and to all trees. Each row in this matrix plots the distributions for one of the four statistics of interest: size, depth, breadth, and virality. The columns plot the distributions for three sets of data: all trees (first column), political trees (second column), and misinformation trees (third column). Within each panel, the four lines refer to each of the four intervention periods we analyze. The goal of this plot is to compare the properties of diffusion across different types of posts and content moderation regimes.

The main important finding that comes out of this figure is that the average size of trees varies significantly over the observation period, especially for misinformation trees. The average size of misinformation trees with 100 or more re-shares is 682 for the full period, but this statistic fluctuates from 770 (prior to October 9) to 495 (during the first high intensity period) and from 572 (December-January) to 534 (second high intensity period). For political trees, the average size is 697 for the whole observation window, shifting from 735 to 674 (first high intensity period of content moderation) and from 583 to 645 (second high intensity period). For all trees, the average size is 919, showing a constant decrease across the four subperiods (951, 928, 885, 798). We show additional descriptive statistics in Table S5 in the online supplement, including the percentiles in tree depth and virality scores.

In interpreting our comparisons between misinformation and non-misinformation trees, we note that the structural measures we use to characterize the trees (virality, depth, and breadth) are all correlated with tree size. Because of this, we also run a set of comparisons holding the size of trees constant, following prior research (Juul and Ugander 2021). In other words, we complement the analyses reported in Figure 2 with a comparison of distributions for only size-matched trees (see sections S5.5 to S5.9 in the online supplement for more details on these comparisons). These additional analyses confirm that, even after matching exactly on tree size, trees

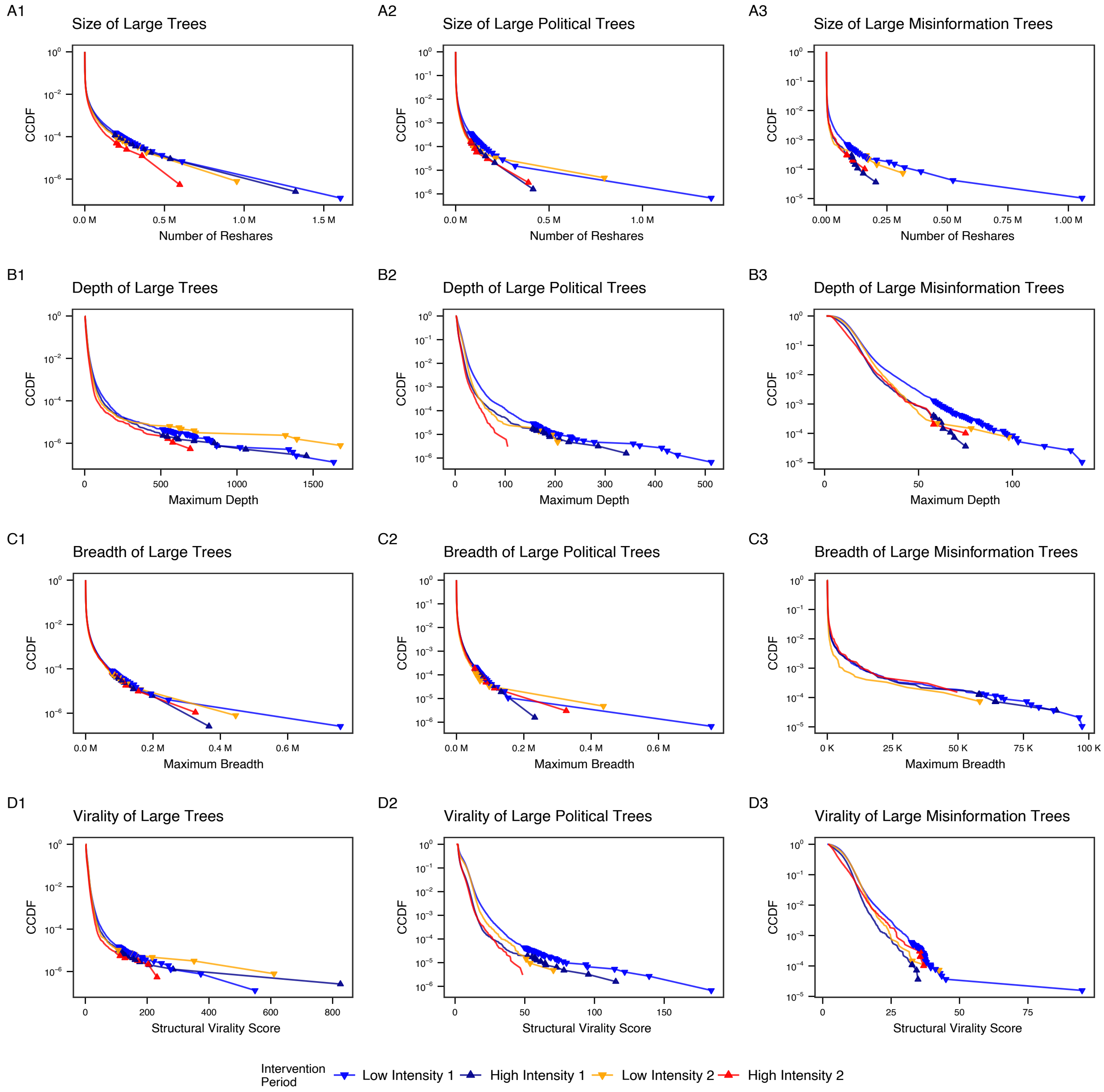

Figure 2: Structural differences between all large trees, large political trees, and large misinformation trees across intervention periods. (A1–D1) Structural properties of all large trees for the four subsets of the data that correspond to the low intensity and high intensity intervention periods before and after Election Day. (A2–D2) Structural properties for the subset of trees classified as political. (A3–D3) Structural properties of misinformation trees (per 3PFC ratings). Across all plots, the data points overlaid on the lines locate the 50 top trees, ordered by size, depth, breadth, and virality (respectively). The structure of trees and the tails of the distributions vary substantially across intervention periods, especially so for misinformation and political trees. (See section S5 in the online supplement for log–log versions of the plots and Kolmogorov–Smirnov tests that show the differences in these distributions are statistically significant.)

identified as misinformation are, on average, deeper and have higher virality (note, however, that the distribution of these metrics also has shorter tails, as Fig. 2 suggests). In Figures S22 and S23 in the online supplement, we also show that misinformation trees gather fewer re-shares at each step of the diffusion process and crucially that they grow more slowly, contrary to what past research has claimed about misinformation (Vosoughi et al. 2018). This suggests that, when growing above the $k > 100$ re-shares threshold, misinformation relies less on broadcasting and more on peer-to-peer diffusion through long and narrow paths. We also tested that this result is not an artifact of the size threshold: as we show in Figure S48 in the online supplement, the patterns also hold for small trees with $k < 100$ re-shares.

In addition to keeping track of how a tree is initiated (by users, on Pages, or by users in Groups), we have data on the average age of the users involved in each tree, their ideological composition, the percentage of users in each tree that are classified as having high political interest, and the percentage that reside in a swing state. Finally, we also have data on whether the root post of trees is boosted content (i.e., posts published by creators that paid Facebook to increase their distribution via paid views). More details on these variables can be found in section S3.2 in the online supplement.

In Figure 3, we show how these tree composition characteristics correlate with their structural features, which allows us to address RQ3 (who re-shares most misinformation?) and RQ5 (what affordances are associated with the diffusion of misinformation?). The analyses suggest that content posted by Pages generates the largest trees. This is not entirely surprising given that, as discussed, Pages have on average the largest potential audiences: the mean follower (or fan) count for Pages is $N = 17,956$ ($sd = 263,439$, $median = 859$), whereas the mean friend count for U.S. adult active users is $N = 495$ ($sd = 697$, $median = 273$). The mean member count for Groups is $N = 4,153$ ($sd = 22,628$, $median = 379$). However, controlling for whether the tree root is a Page or a Group post and whether the content diffused is classified as political and as news (as well as the rest of the other tree attributes), the second most important factor positively associated with tree size is whether the content diffused is labeled as misinformation.

This finding seems to contradict the patterns identified in Figure 2, but in fact it suggests that, compared to trees with similar attribute composition, misinformation trees are, on average, larger (attaining their size through depth, not breadth). In Figures S37 to S39 in the online supplement, we show that the two most important factors associated with the growth of misinformation trees (other than being initiated by Pages) are (a) whether the content diffused is classified as political news and (b) the average user age: larger misinformation trees are generated by older users (who are also more conservative, as we show in Fig. S28 in the online supplement). We show additional regression results for different quantiles and different time periods in section S5.14 in the online supplement. Across all these different specifications, the results are consistent with what we show here.

In Figure 4, we address RQ6 (which affordances are associated with greater reach?) and again RQ7 (how much temporal variation do the data reveal?) We plot the CCDF for views across intervention periods, which is the main measure of exposure we employ to contextualize our measures of diffusion (panels A-C).

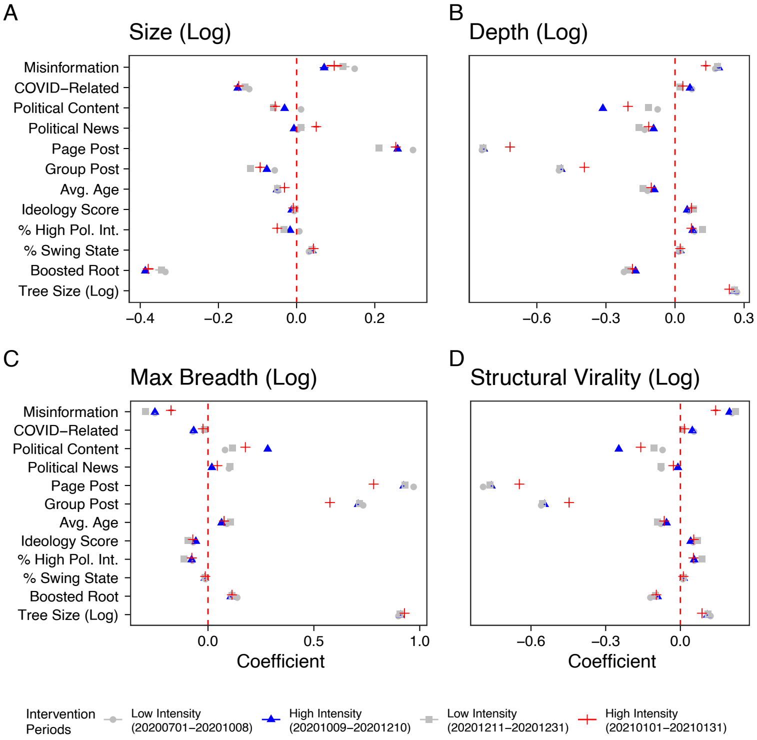

Figure 3: Correlates of large diffusion trees across intervention periods. (A–D) These panels show the results of OLS regressions with four dependent variables (tree size, depth, maximum breadth, and structural virality) for the four subsets of data. To aid interpretability, all continuous variables have been standardized (with the exception of tree size, which is logged). The models also include daily fixed effects to control for underlying but unobservable factors that may be changing in time. Across intervention periods, misinformation trees have systematically larger size, depth, and structural virality, even after being controlled by other tree-level characteristics. During high intervention periods, the magnitude of the estimates decreases slightly, but they still suggest a clear positive association.

Each row corresponds to types of content, with misinformation in the top panels, political posts in the middle panels, and all posts in the lower panels. Each column refers to a different type of root poster source: users posting in Groups in the first column, Pages in the second column, and users in the third. Reach results for trees with $k < 100$ re-shares can be found in Figure S47 in the online supplement. In Figure 4, we also plot temporal changes in the views of posts published by users, on Pages, and in Groups (panels D-F).

What Figure 4 tells us is that, overall, content posted by Pages is viewed by a larger number of people: 24.4 percent of Page posts reach 100K views, whereas only 8.2 percent of the user trees reach 100K views (the percentage is even lower when the source is a Group). However, the differences in the upper tail of the distributions suggest that, although extremely rare, large diffusion trees initiated by users can outmatch those of Pages: 1.8 percent of Page trees reach 1 M views, whereas for user posts the percentage is 1.9 percent. This is particularly true for the

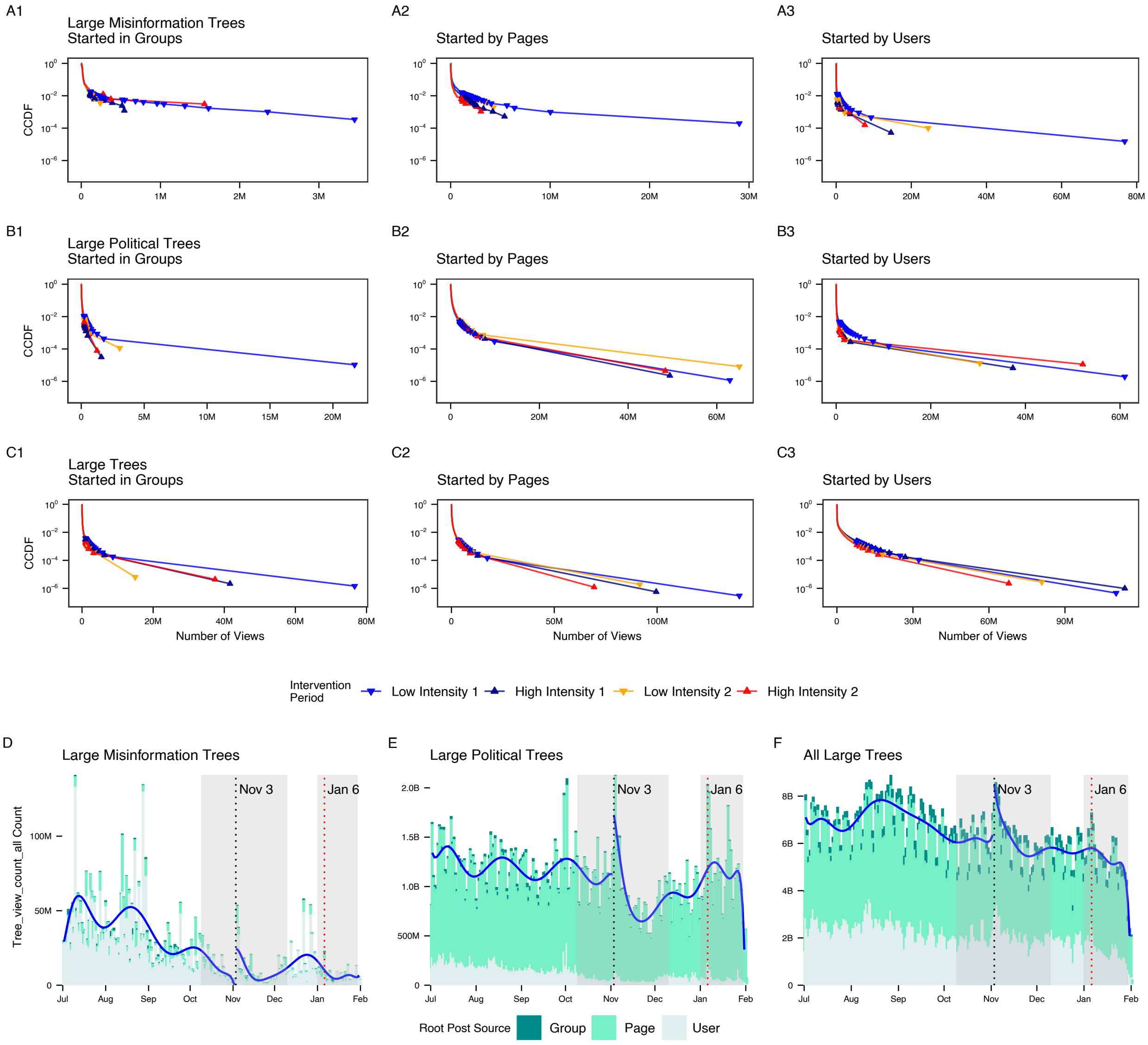

Figure 4: Distribution of views for content posted across intervention periods. (A1–C3) These CCDFs suggest that the reach of misinformation trees was substantially reduced during high intervention periods, especially compared to the baseline prior to October 9. Across all plots, the data points overlaid on the lines locate the 50 top trees, ordered by the number of views. (See section S5 in the online supplement for log–log versions of the plots and Kolmogorov–Smirnov tests that show the differences in these distributions are statistically significant.) (D–F) Changes in the reach of large trees, where blue lines are 10th degree polynomials, fitted separately to data before and after Election Day. There is a steady decline in the reach of large misinformation trees, going from $\sim$ 50 M at the start of the observation period to close to 0 just before November 3. Misinformation views visibly increase over the weeks the "break the glass" measures were rolled back after the election (end of first shaded rectangle). The increase is mostly driven by a surge in the number of large trees initiated by users, as seen in Figure 1I. The steep drop in late January is an artifact of the right-censoring of the data.

low intervention periods, when the most successful users compete with the most successful Pages and Groups in terms of reach.

When we focus specifically on misinformation, 10.6 percent of Page posts reach 100K views and 1.4 percent reach 1 M views; however, only 2.7 percent of user posts reach 100K views and 0.4 percent reach 1 M views. (Groups reach, on average, significantly fewer people in this category of content too.) In other words, users may be the source for most posts labeled as misinformation (Fig. 1I), but it is Pages that accumulate, per tree, more of the misinformation views; users lack, for the most part, the broadcasting potential that Pages have. However, these aggregate statistics again hide significant temporal variability. The daily counts show a clear and steady decline in views as the election grew closer: the view counts for large misinformation trees for the week before the election were lower than at any day prior in the data and much lower than the average daily count during the first low intensity period, dropping to values that border the zero line. Strikingly, the views for misinformation spiked the day after the election relative to the day before, gradually declining again to very low levels over the second half of November and then increasing once more during the second low intensity period. Relatively speaking, there were smaller fluctuations for large political trees and less dramatic fluctuations in all large trees across periods.

Again, users responsible for most misinformation posts, re-shares, and views amount to a very small minority. As discussed above, we estimate that ~ 1 percent of users account for most misinformation re-shares, which is consistent with patterns identified in prior research (e.g., Grinberg et al. 2019; Guess et al. 2019). Posting and re-sharing of misinformation is, therefore, concentrated in the hands of a very small number of users. This highly committed minority has a disproportionate influence on the information circulating on Facebook: millions of other users gained exposure to misinformation through the peer-to-peer diffusion channels this minority repeatedly activated. Moreover, this highly committed minority of users was able to maintain their misinformation activity under the radar of content moderation policies that especially targeted Pages and Groups.

A definitive interpretation of the temporal variations is impossible, given the lack of access to information on the supply of and demand for misinformation or a complete record of interventions by Facebook during this period (other than those cited in online supplement S4). Even if we had been provided this information, our analyses would still be by necessity descriptive: it is hard to conceive of a system-level randomized control trial, given the complex nature of the "break the glass" measures and how they were being implemented (i.e., as a response to exogenous events). Evaluating the likely effect of content moderation measures—"break the glass" and otherwise—depends on reasoning about an unobserved counterfactual, which is a tricky exercise in the context of the historical moment we consider here. It is akin to evaluating the question, 'Would Mr. Smith be alive today had he not been pushed off of the skyscraper?' We can still evaluate how healthy Mr. Smith was before the push and his likelihood to survive the day, absent the push, even if we cannot randomly push 100 people off skyscrapers (luckily). In our case, we can reason that the supply and demand of political information, generally, and misinformation, specifically, should have increased as the election neared. The analysis of these dynamics on other platforms, for example Twitter (Grinberg et al. 2019), does find a surge of misinformation prior to the 2016 election (although McCabe et al. 2024, using similar methods, finds less of a surge in 2020). The broader literature on information and elections, in fact, does not offer reasons to expect that the quantity of political information and misinformation should dramatically decline as an election nears. What we observe in our data is, therefore, surprising: the quantity of large diffusion trees steadily decreased from July until Election Day, and the views of large misinformation trees during the two weeks prior to the election (i.e., about the time that the "break the glass" measures were being instituted) strikingly plummeted to near zero. Furthermore, the number of large misinformation trees soared by an order of magnitude in early/mid December (not a traditional time for increased political information and misinformation prevalence).

Our most parsimonious explanation for these patterns is that content moderation efforts generally intensified as the election grew closer, with the "break the glass" measures accelerating the process. The retraction of these measures in early/mid December facilitated, in turn, the increase in the number of large diffusion events. In addition, the asymmetric drop in user-initiated diffusion trees during the peak periods of the "break the glass" measures also suggests that these interventions were aimed at viral, user-to-user diffusion. In turn, the dramatic increase in the number of misinformation trees and their reach the day after the election either means that the interventions were a poor match for the misinformation circulating on the platform or that the enforcement of the content moderation policies could not keep up with a surge in the supply of misinformation.

We would also argue that the simplest explanation for why misinformation trees rely far more heavily on peer-to-peer spread than political information does is that there were serious penalties aimed at the broadcast mechanisms on Facebook (Pages and Groups). Users were not subject to these penalties to the same extent. In particular, as noted above, Pages and Groups had the visibility of their content sharply reduced if they shared multiple pieces of misinformation; users did not. Here, we have somewhat less leverage in making this inference as compared to assessing "break the glass" measures because we cannot take advantage of temporal variation in the "repeat offender" policy. There may be reasons why sharers of misinformation would prefer to share content with Friends rather than through Pages and Groups. However, there is again little justification in the literature that sharers of misinformation would generally be averse to using the most effective means of sharing their content (broadcasting). Indeed, the literature on sharing misinformation suggests that most misinformation comes from a small number of "supersharers" who seem intent on maximizing the spread of misinformation (Guess et al. 2019; Baribi-Bartov et al. 2024).

Finally, we note here that Facebook, soon after the data collection for this research was completed, extended their "repeat offender" policy to users "Taking Action Against People Who Repeatedly Share Misinformation. Accessed May 2, 2024 (https://about.fb.com/news/2021/05/taking-action-against-people-who-repeatedly-share-misinformation/)." A powerful test of whether this particular content moderation policy caused misinformation trees to be deep and not broad would be to evaluate whether this pattern disappeared after this new policy was instituted.

Discussion

The results presented here support two statements that are simultaneously true. The first is that Facebook creates a broadcasting (rather than viral) mode of exposure when it comes to the overall reach of information, with Pages (not Groups) acting as the key engine for this type of dissemination. The second is that misinformation, as a subset of content, reverses those trends: Our analyses provide clear evidence that misinformation relies much more on viral spread, powered by a tiny minority of users who tend to be older and more conservative. Some of these findings run counter to what prior research found on other platforms (e.g., Vosoughi et al. 2018): the average misinformation post in our data is more structurally viral, but it spreads more slowly and, crucially, receives fewer views overall—even if the views it accumulates still amount to millions.

Importantly, our findings also reflect a medium that was heavily moderated in ways that almost certainly affected patterns of diffusion and virality. This is an aspect of how platforms mediate the flow of information that no prior research has considered while analyzing the full set of posts being disseminated, as we do here. Our results provide unambiguous evidence that diffusion dynamics are contingent on platform affordances (like the support for Pages) and the less stable set of principles encoded in the form of content moderation policies (in this case, interventions that more aggressively target misinformation posted by Pages and in Groups, which allowed peer-to-peer spread to evade the moderation system). We also show that what is true on average was not true at specific periods of collective vulnerability. The misinformation generated right after the election increased in terms of volume and reach at a time when the Stop the Steal campaign was on the rise and the "break the glass" measures were being deprecated. The observational nature of our data does not allow us to identify strong causal mechanisms. But the dynamics we document are clearly indicative of the influence that affordances and a reduction in content moderation efforts may have at crucial junctures.

The findings we provide here rely on the most ambitious analysis to date on how content propagates on social media. However, the analyses we discuss still have some limitations. The first is that, for privacy reasons, we were not granted access to the social network underlying the diffusion dynamics we analyze, so we cannot identify the network location (and embeddedness) of the very small minority of users spreading most misinformation. We also cannot analyze how the audiences of Pages and Groups overlap, potentially creating bridges for diffusion. Having access to these data could illuminate avenues for intervention, like creating friction in the streams of diffusion repeatedly activated by the minority of users committed to misinformation.

A second limitation is that we can only identify misinformation that has been labeled as such. Like most existing research on the diffusion of misinformation, we rely on the ratings of fact-checking organizations (as we explain in section S3.2 in the online supplement). But if misleading content goes unlabeled by these organizations, we simply underestimate the prevalence of this type of posts on the platform. Furthermore, what complicates the analysis of misinformation in the context of our data is that the 3PFC ratings were coupled with platform interventions that varied in time. This means that we cannot confidently claim that our findings generalize to unlabeled misinformation or other time periods. This same limitation affects all other studies published to date on the prevalence of misinformation on social media platforms (whether they acknowledge it or not).

The third limitation is that we could only gather information about specific content moderation policies and the extraordinary "break the glass" measures, through information in the public domain. This imposes some opacity and lack of granularity in our understanding of the content moderation policies in place that are affecting the data generation process. Fourth, our measurement of diffusion trees is a very accurate representation of re-sharing behavior, but it may still provide a noisy assessment of the causal pathways through which diffusion occurs. For example, Alice may influence the posting content of Bret even if Bret never re-shares anything posted by Alice. To the extent that our diffusion structures only track re-sharing behavior, those possible avenues of influence are outside of the scope of our data (and likely outside of the scope of any behavioral data).

Finally, a fifth limitation is that the patterns we analyze are the aggregate result of the muddled interactions between source following distributions (or network sizes), posting activity, algorithmic ranking, and content moderation policies and interventions. We cannot causally parse these interactions with the data we have. However, even if we cannot disentangle all these mechanisms, we still measure the resulting diffusion patterns with much broader data than those used in past work. Our analyses of temporal variation also help us uncover a source of heterogeneity that has been mostly disregarded by prior research.

We also note that our results reflect a certain moment in the evolution of social media and of Facebook as a platform. Some of the patterns we find may be quite robust across platforms and time, but others may be driven by the nature of the ecosystem and broader society at a particular moment. Indeed, some of these differences (e.g., in the array of platform-specific affordances on Facebook in 2020 or the enforcement of policies around misinformation) might account for the divergence of our findings from prior literature. We may in fact be on the cusp of a transition away from the types of socially driven diffusion that still dominated Facebook in 2020. Technologies today allow other forms of algorithmic curation that rely less on social networks and more on content affinities (like the recommendation engine employed by TikTok, which Facebook is now trying to emulate, Heath 2022); and the addition of AI technologies to search functions (like the integration of large language models in the search of web resources) may drastically change how information spreads online and finds an audience.

Broadcasting may become even more salient than viral diffusion on emerging platforms, but these two forms of dissemination may also become more intertwined in the amplification dynamics that online communication facilitates. Crucially, the need for content moderation (either in the form of policy interventions or guardrails for content generation) will not go away. This points to the need to replicate this type of research across time, societies, and platforms. The ability to control information flows should not be exercised outside of public scrutiny. We urge more publicly transparent research in this area, especially as it relates to the diffusion of misinformation and other problematic content.

References

Amichai-Hamburger, Yair, Tali Gazit, Judit Bar-Ilan, Oren Perez, Noa Aharony, Jenny Bronstein, and Talia Sarah Dyne. 2016. "Psychological Factors behind the Lack of Participation in Online Discussions." *Computers in Human Behavior* 55:268–77. https://doi.org/10.1016/j.chb.2015.09.009

Bakshy, Eytan, Itamar Rosenn, Cameron Marlow, and Lada A. Adamic. 2012. "The Role of Social Networks in Information Diffusion." *Proceedings of the 21st International Conference on World Wide Web*. 519–28. https://doi.org/10.1145/2187836.2187907

Bandy, Jack and Nicholas Diakopoulos. 2023. "Facebook's News Feed Algorithm and The 2020 US Election." *Social Media + Society* 9:20563051231196898. https://doi.org/10.1177/20563051231196898

Baribi-Bartov, Sahar, Briony Swire-Thompson, and Nir Grinberg. "Supersharers of Fake News on Twitter." *Science* 384:979–82. https://doi.org/10.1126/science.adl4435

Broniatowski, D. A., Simons, J. R., Gu, J., Jamison, A. M. & Abroms, L. C. The efficacy of Facebook's vaccine misinformation policies and architecture during the COVID-19 pandemic. Sci. Adv. 9, eadh2132 (2023).

Cheng, Justin, Lada Adamic, P. Alex Dow, Jon M. Kleinberg, Jure Leskovec. 2014. "Can Cascades Be Predicted?. *Proceedings of the 23rd International Conference on World Wide Web*. 925–36. https://doi.org/10.1145/2566486.2567997

Evans, Sandra K., Katy E. Pearce, Jessica Vitak, Jeffrey W. Treem. 2016. "Explicating Affordances: A Conceptual Framework for Understanding Affordances in Communication Research." *Journal of Computer-Mediated Communication* 22(1):35–52. https://doi.org/10.1111/jcc4.12180

Fact-checking policies on Facebook, Instagram, and Threads. Accessed May 2, 2024 (https://www.facebook.com/business/help/315131736305613?id=673052479947730).

Friggeri, Adrien, Lada Adamic, Dean Eckles, Justin Cheng. 2014. "Rumor Cascades." *Eighth International AAAI Conference on Weblogs and Social Media* 8. https://doi.org/10.1609/icwsm.v8i1.14559

Goel, Sharad, Ashton Anderson, Jake Hofman, Duncan J. Watts. 2016. "The Structural Virality of Online Diffusion." *Management Science* 62:180. https://doi.org/10.1287/mnsc.2015.2158

Grinberg, Nir, Kenneth Joseph, Lisa Friedland, Briony Swire-Thompson, and David Lazer. 2019. "Fake News on Twitter during the 2016 U.S. Presidential Election." *Science* 363:374–8. https://doi.org/10.1126/science.aau2706

Guess, Andrew, Jonathan Nagler, and Joshua Tucker. 2019. "Less Than You Think: Prevalence and Predictors of Fake News Dissemination on Facebook." *Science Advances* 5:eaau4586. https://doi.org/10.1126/sciadv.aau4586

Heath, Alex. 2022. "Facebook is changing its algorithm to take on TikTok, leaked memo reveals." The Verge. Accessed November 14, 2024 (https://www.theverge.com/2022/6/15/23168887/facebook-discovery-engine-redesign-tiktok).

How Meta's third-party fact-checking program works. Accessed May 2, 2024 (https://www.facebook.com/formedia/blog/third-party-fact-checking-how-it-works).

Juul, Jonas L. and Johan Ugander. 2021. “Comparing Information Diffusion Mechanisms by Matching on Cascade Size.” *Proceedings of the National Academy of Sciences* 118:e2100786118. https://doi.org/10.1073/pnas.2100786118

Liben-Nowell, David and Jon Kleinberg. 2008 “Tracing Information Flow on a Global Scale Using Internet Chain-Letter Data.” *PNAS* 105:4633. https://doi.org/10.1073/pnas.0708471105

McCabe, Stefan D., Diogo Ferrari, Jon Green, David M. J. Lazer, and Kevin M. Esterling. 2014. “Post-January 6th Deplatforming Reduced the Reach of Misinformation on Twitter.” *Nature* 630:132–40. https://doi.org/10.1038/s41586-024-07524-8

Onnela, Jukka-Pekka and Felix Reed-Tsochas. 2010. “Spontaneous Emergence of Social Influence in Online Systems.” *Proceedings of the National Academy of Sciences* 107:18375–80. https://doi.org/10.1073/pnas.0914572107

Ronzhyn, Alexander, Ana S. Cardenal, Albert B. Rubio. 2023. “Defining Affordances in Social Media Research: A Literature Review.” *New Media & Society* 25(11):3165. https://doi.org/10.1177/14614448221135187

Taking Action Against People Who Repeatedly Share Misinformation. Accessed May 2, 2024 (https://about.fb.com/news/2021/05/taking-action- against-people-who-repeatedly-share-misinformation/).

Théro, Héloïse and Emmanuel M. Vincent. 2022. “Investigating Facebook’s Interventions against Accounts that Repeatedly Share Misinformation.” *Information Processing & Management* 59:102804. https://doi.org/10.1016/j.ipm.2021.102804

Vosoughi, Soroush, Deb Roy, and Sinan Aral. 2018. “The Spread of True and False News Online.” *Science* 359:1146. https://doi.org/10.1126/science.aap9559

Acknowledgments: The Facebook Open Research and Transparency (FORT) team provided substantial support in executing the overall project. We are grateful for support on various aspects of project management from Chaya Nayak, Sadaf Zahedi, Lama Ahmad, Akshay Bhalla, Clarice Chan, Andrew Gruen, Bennet Hillenbrand, Pamela McLeod, and Dáire Rice; engineering and research management from Da Li and Itamar Rosenn; engineering from Yuxi Chen, Shiyang Chen, Tegan Lohman, Robert Pyke, and Yixin Wan; data engineering from Suchi Chintha, John Cronin, Devanshu Desai, Vikas Janardhanan, Yann Kiraly, Xinyi Liu, Anastasiia Molchanov, Sandesh Pellakuru, Akshay Tiwari, Chen Xie, and Beixian Xiong; data science and research from Hannah Connolly-Sporing; academic partnerships from Rachel Mersey, Michael Zoorob, Lauren Harrison, Simone Aisiks, Yair Rubinstein, and Cindy Qiao; privacy and legal assessment from Kamila Benzina, Frank Fatigato, John Hassett, Subodh Iyengar, Payman Mohassel, Ali Muzaffar, Ananth Raghunathan and Annie Sun; and content design from Caroline Bernard, Jeanne Breneman, Denise Leto, and Melanie Jennings. NORC at the University of Chicago partnered with Meta on this project to conduct the fieldwork with the survey participants. We are particularly grateful for the partnership of NORC Principal Investigator J.M. Dennis and NORC Project Director Margrethe Montgomery.

Sandra González-Bailón: lead author with control rights; Annenberg School for Communication, University of Pennsylvania. E-mail: sandra.gonzalez.bailon@asc.upenn.edu.
David Lazer: lead author with control rights; Network Science Institute, Northeastern University. E-mail: d.lazer@northeastern.edu
Pablo Barberá: lead Meta author; Meta. E-mail: us2020research@meta.com
William Godel: lead Meta author; Meta. E-mail: us2020research@meta.com
Hunt Allcott: Environmental and Energy Policy Analysis Center, Stanford University.
E-mail: allcott@stanford.edu
Taylor Brown: Meta. E-mail: us2020research@meta.com
Adriana Crespo-Tenorio: Meta. E-mail: us2020research@meta.com
Deen Freelon: Annenberg School for Communication, University of Pennsylvania.
E-mail: dfreelon@upenn.edu
Matthew Gentzkow: Department of Economics, Stanford University. E-mail: gentzkow@stanford.edu
Andrew M. Guess: Department of Politics and School of Public and International Affairs, Princeton University. E-mail: aguess@princeton.edu
Shanto Iyengar: Department of Political Science, Stanford University. E-mail: siyengar@stanford.edu
Young Mie Kim: School of Journalism and Mass Communication, University of Wisconsin-Madison. E-mail: ymkim5@wisc.edu
Neil Malhotra: Graduate School of Business, Stanford University. E-mail: neilm@stanford.edu
Devra Moehler: Meta. E-mail: us2020research@meta.com
Brendan Nyhan: Department of Government, Dartmouth College. E-mail: nyhan@dartmouth.edu
Jennifer Pan: Department of Communication, Stanford University. E-mail: jp1@stanford.edu
Carlos Velasco Rivera: Meta. E-mail: us2020research@meta.com
Jaime Settle: Department of Government, William & Mary. E-mail: jsettle@wm.edu
Emily Thorson: Department of Political Science, Syracuse University. E-mail: ethorson@gmail.com
Rebekah Tromble: School of Media and Public Affairs and Institute for Data, Democracy, and Politics, The George Washington University. E-mail: rtromble@email.gwu.edu
Arjun Wilkins: Meta. E-mail: us2020research@meta.com
Magdalena Wojcieszak: Department of Communication, University of California, Davis
Center for Excellence in Social Science, University of Warsaw. E-mail: mwojcieszak@ucdavis.edu
Chad Kiewiet de Jonge: Meta research lead; Meta. E-mail: us2020research@meta.com
Annie Franco: Meta research lead; Meta. E-mail: us2020research@meta.com
Winter Mason: Meta research lead; Meta. E-mail: us2020research@meta.com
Natalie Jomini Stroud: co-last author and academic research lead; Moody College of Communication and Center for Media Engagement, University of Texas at Austin. E-mail: tstroud@austin.utexas.edu
Joshua A. Tucker: co-last author and academic research lead; Wilf Family Department of Politics and Center for Social Media and Politics, New York University. E-mail: joshua.tucker@nyu.edu