

<https://doi.org/10.1038/s44260-024-00018-8>

How U.S. Presidential elections strengthen global hate networks



Akshay Verma, Richard Sear & Neil Johnson

Local or national politics can be a catalyst for potentially dangerous hate speech. But with a third of the world's population eligible to vote in 2024 elections, we need an understanding of how individual-level hate multiplies up to the collective global scale. We show, based on the most recent U.S. presidential election, that offline events are associated with rapid adaptations of the global online hate universe that strengthens both its network-of-networks structure and the types of hate content that it collectively produces. Approximately 50 million accounts in hate communities are drawn closer to each other and to a broad mainstream of billions. The election triggered new hate content at scale around immigration, ethnicity, and antisemitism that aligns with conspiracy theories about Jewish-led replacement. Telegram acts as a key hardening agent; yet, it is overlooked by U.S. Congressional hearings and new E.U. legislation. Because the hate universe has remained robust since 2020, anti-hate messaging surrounding global events (e.g., upcoming elections or the war in Gaza) should pivot to blending multiple hate types while targeting previously untouched social media structures.

The Internet is a breeding ground for hate speech^{1,2}. Hateful content and its followers thrive in networks comprising an interconnected web of communities across multiple social media platforms^{3–6}. The communities on each platform (e.g., Telegram Channel, Gab Group, YouTube Channel) show an ideology rooted in hatred and discrimination, and can have anywhere from a few to a few million members. Any member of such a “hate community” (community A) can at any time cross-post content of interest from another community (community B) and hence create a link from A to B (see Fig. 1A). Other members of A are then alerted to B's existence, and can visit community B to share their hate^{7–11}. Reference¹² provides explicit examples of links and hate communities (nodes).

A “hate universe” is hence formed by the set of all such hate communities across social media platforms—together with the other communities that they directly link to (vulnerable mainstream, Fig. 1A). It is a novel network-of-networks¹³ in which both the network structure and its content co-evolve dynamically on similar timescales. We refer to the broad mainstream in Fig. 1A as “vulnerable” because, in network terms, these communities sit near the hate network and are therefore more susceptible to incoming hateful rhetoric from these known hate communities than a community selected at random from the entire internet. Prior results from collective social psychology show that hate speech is strongly predicted by contempt derived from feelings of in-group superiority¹⁴, i.e., feelings found in both hate and mainstream communities. However, thorough analysis of the composition of the vulnerable mainstream we identify is beyond the scope of this study and left for future work.

While the cross-platform linking behavior we study here is not necessarily unique to hate networks, it allows them to form collective resilience against moderator pressure³. Cross-platform studies are therefore particularly salient for networks that spread hate speech.

While there have been many in-depth studies on the relationship between polarizing real-world events and the evolution of online hate^{15–19}, our focus on the dynamics of this hate universe at scale enables us to fill a gap in the current understanding of how hate evolves globally around a local or national event such as an election²⁰. We focus here on the U.S. presidential election and choose the most recent case (2020) since it is the first in which many small and large platforms feature prominently, and its offline events included the Capitol attack on the day that Congress certified the president-elect²¹.

Results

Figure 2 shows there was a surge in the creation of hate links (i.e., links from hate communities) during both the election itself (November 2020) and the follow-up (January 2021). On November 3, the day of the election, the increase was 41.6% compared to November 1. On November 7 when Joe Biden was declared president-elect, this number spiked even further at 68%. An even larger spike surrounds January 6. Going forward in the digital age, the message of this for policymakers is that any event of seemingly only local or national interest has the capability to instantly trigger hate activity globally.

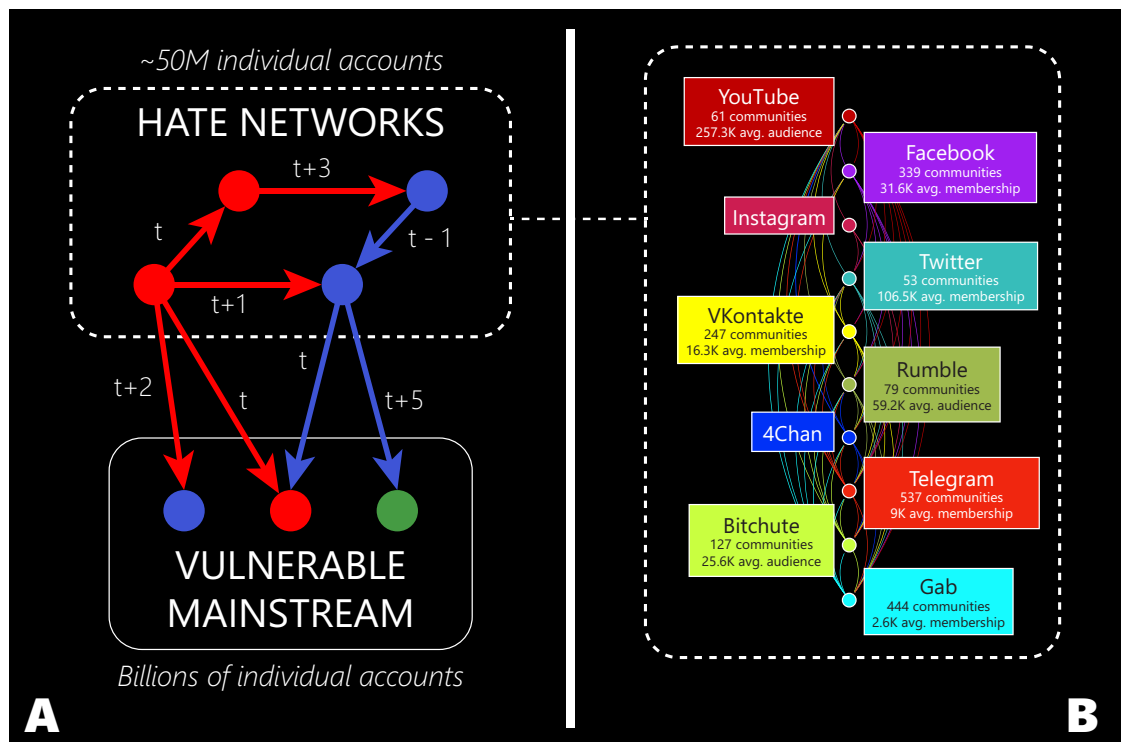


Fig. 1 | Hate universe infrastructure. **A** Schematic of how a hate community (node) on a given social media platform (given color) establishes a link (by sharing a URL for a piece of content) to another community (node) at a given time. A link from node A to node B means that members of community A are alerted to B's existence, and can visit community B to share hate. Each community is a platform-provided community: a Telegram Channel, a Gab Group, a YouTube Channel, etc. Top dotted box is the subset of communities that are identified as hate communities. These link

together over time to form hate networks-of-networks. Bottom box represents communities that are directly linked to by hate communities but which are not themselves in the list of hate communities; hence, we label these as vulnerable mainstream. **B** Empirically determined composition of the hate universe. Boxes show number of hate communities (nodes) in a given platform and the average number of members of each community. Edges shown are aggregated over time.

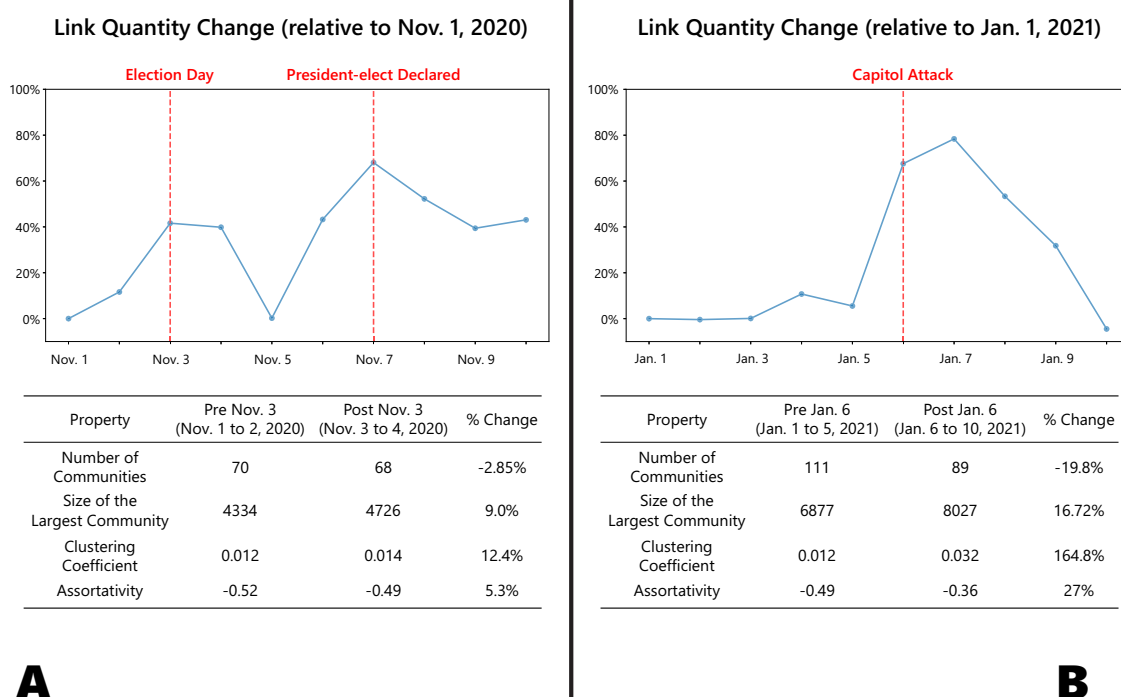


Fig. 2 | Changes in key network metrics surrounding events in the 2020 U.S. election period. **A** Actual date of the election (**B**) Day of Congress' confirmation which is when the attack took place on the U.S. Capitol.

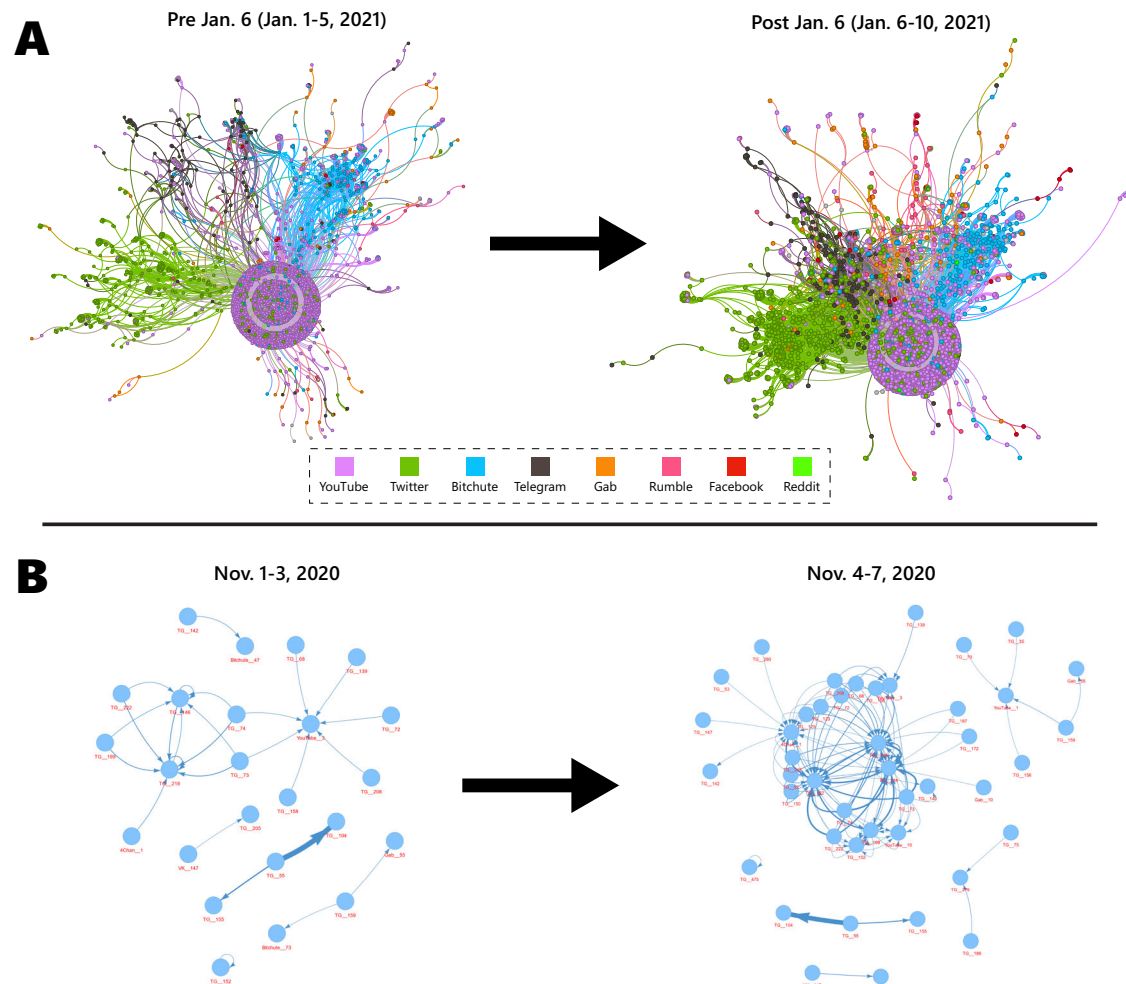


Fig. 3 | Hardening of hate universe's network-of-networks. **A** Gephi visualization of the hate universe before and after January 6, 2021. The partial rings comprise dense clouds of vulnerable mainstream communities (nodes) with successively

increasing numbers of links from the hate networks, hence creating orbital-like rings. **B** Subset of Telegram-connected networks before and after election day, showing its key role as a binding agent.

Hate Universe's Structure Hardens

Just because links get individually added does not automatically mean the hate universe as a whole will benefit in terms of hardening (strengthening) its structural cohesion. However, the lower panels of network measures in Fig. 2 show this is indeed exactly what happens.

The clustering coefficient, which is a standard measure that captures the tendency of nodes to form connected groups of three, jumped by 164.8% after January 6. Nodes previously on the periphery of the core formed new connections, leading to a denser network with well-defined clusters. Hence, individual nodes within the core became more interconnected on average¹³. The assortativity, which is a standard measure that captures the tendency of nodes to connect with other nodes that are similar in terms of their connectivity, also increased by 27%. This further supports the notion of increased cohesion: individual communities (nodes) preferentially connected with others who shared similar network characteristics. These suggest a strengthening of existing ideologies within the network, fostering a more homogeneous environment that is then more resilient to outside intervention^{22,23}.

The changes in the number of communities and the size of the largest community (Fig. 2 lower panels) provide further evidence of this network strengthening (hardening). The number of communities decreased by 19.8% at the same time as the size of the largest community grew by 16.72%, reflecting a convergence of smaller communities towards a single, more dominant one. This evolution toward a more cohesive hate universe with a more unified ideology at scale means it can better amplify the spread of hate

speech or coordinated actions at scale. Most importantly, this is all essentially organic and hence not the result of top-down control by a single bad actor.

These conclusions are further supported by visualizations employing the ForceAtlas2 layout²⁴. This layout algorithm assigns forces to nodes, attracting or repelling them based on their edges which act like springs: this means that the network's resulting visual appearance reflects its true structure at scale. The resulting visualizations (Fig. 3A) reveal a notable shift towards a more cohesive network-of-networks structure – in agreement with the metrics.

Hate Universe's Contents Harden

Just because the hate universe's structure becomes more cohesive does not automatically mean that the type of hate content that it contains should change – or if it does, how it would change. Yet Fig. 4A shows there was a significant uptick in hate speech targeting immigration, ethnicity, and antisemitism around November 7. In particular, there was a 269.5% surge in anti-immigration sentiments, a 98.7% rise in ethnically-based hatred, and a 117.57% escalation in expressions of antisemitism from November 7–11 compared to November 2–6. A comparable surge in anti-immigration sentiments occurred around January 6 (Fig. 4B). There was a 108.69% rise in anti-immigration content between January 6–10 compared to January 1–5, 2021.

These trends suggest an increase in immigration anxieties harbored by far-right communities, which often align with the Great Replacement conspiracy theory and attribution of perceived demographic shifts to Jewish influence²⁵.

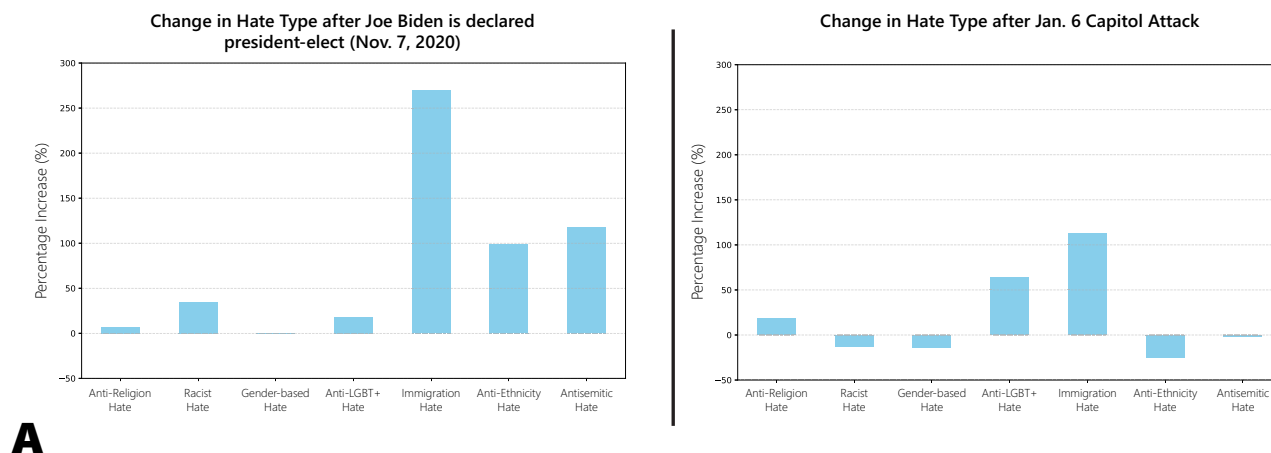


Fig. 4 | Hardening of hate universe's content around particular targets of hate narratives. A Results for the period around the declaration of Biden as president-elect. **B** Results for the period around January 6.

We also find strong correlations between changes in the hate universe's structure and changes in its content—specifically, correlations between daily link count originating from specific social media platforms and hate content over the period November 1, 2020 to January 10, 2021 (see Supplementary Fig. 1). For example, there is an increase in links originating from 4Chan, Gab, and Twitter, while Telegram exhibits a strong correlation with instances of hate speech that target immigration, race, and gender identity/sexual orientation (GISO).

Key role of Telegram

Historically, Telegram has never featured in U.S. Congressional hearings involving social media company representatives, nor does it feature in the E.U.'s flagship Digital Services Act²⁶ that began requiring some online platforms to comply with its policies in August 2023. This may soon change; recently, Telegram's founder was detained in France as a result of illegal activity on the platform and the E.U. is reportedly examining ways of regulating the platform using the Digital Services Act. We note that these developments do not change our conclusions in any way, but they highlight the increasing relevance of Telegram in the global conversation about online hate.

Figure 3B reveals Telegram's key role as a 'glue' in the hate network's hardening process, by showing the subset of the hate universe in which Telegram is a source or target node. During November 4–7, there was a remarkable surge in connectivity within Telegram, as evidenced by a substantial 299% increase in the number of connections involving Telegram compared to November 1–3 (from 592 to 2366). Telegram's role as a target node increased from 18.22% to 33.47%, while its role as a source node increased from 21.73% to 37.24%, as a percentage of all links present in the hate-to-hate sub-network. These numbers demonstrate Telegram's growing significance as a central platform for communication and coordination among hate communities²⁷ and align with growing concern among law enforcement agencies because of its association with 'Q-anon' and 'Pro-Trump Conspiracy theories'^{28–30}.

The Telegram community we call "TG_122" in our data (whose current name promotes a false narrative about voter fraud) is an example that highlights the emerging importance of Telegram in the network and its alignment with conspiracy theories. It was initially absent from the network during November 1–3, but then emerged and rapidly became one of the most crucial and interconnected nodes. "TG_122" exhibits strong associations with two other Telegram channels, "TG_218", a right-wing podcast, and "TG_146", a channel whose name promotes an antisemitic conspiracy theory, both of which were highly active before November 3.

Discussion

As well as showing how individual-level hate gets expressed collectively at scale, our findings reveal the following shortcomings in current practice and regulations:

- (1) Particular platforms can play different network roles in the hate universe, in particular smaller and/or less well-known ones such as Telegram. This suggests that current policies that focus only on particular platforms based on their popularity (e.g., Facebook, Twitter, or TikTok) or which treat all platforms the same, will not be effective in curbing hate and other online harms.
- (2) The type of hate content that surges around a real-world event is not necessarily the closest in theme to that event. Hence, the messaging of anti-hate campaigns ahead of an event should not focus exclusively on that event's theme.

In summary, the hate universe's rapid self-adaptation at scale, its wide array of target-ready mainstream communities, and the variety of hate content that it offers suggests that future major events will continue to provide it with a diverse set of new recruits globally while simultaneously strengthening (hardening) its structure and content at scale. These findings suggest that anti-hate policies ahead of nominally local or national events (such as elections) should mix multiple hate themes and use the multi-platform hate universe map to target their impact at the global scale.

Methods

We build the hate universe using the same methodology as prior publications^{12,31,32} but expanded to additional platforms. We define a hate community as one for which multiple subject matter experts determine two of its most recent twenty posts (at time of analysis) displays "hate". "Hate" is defined as content falling under the provisions listed in the FBI's definition of hate crimes, or content promoting extreme racial identitarianism/fascist regime types. Prior publications contain a full discussion and concrete examples of this methodology^{12,31,32}. We first extracted and classified an initial list of hate communities (see data provided) using terms and names in public resources such as the ADL Hate Symbols Database and Southern Poverty Law Center. We then monitored this initial set for cross-posts. This then led us to new communities and hence allowed us to grow a comprehensive candidate list of communities, which were then analyzed by subject matter experts using the same criteria to determine whether they are a hate community. Repeating this link-following process in a semi-automated way eventually produced a crude representation of the entire public-facing hate universe across social media platforms (Fig. 1B). In this directed network (i.e., comprising links from one community to another), each source node is a hate community (Fig. 1A top panel). If its target node is not on our hate community list, we label it as vulnerable mainstream: it is possible that it includes discriminatory content, but it had not at the time qualified to be on the hate community list.

We classify the text in the hate communities' posts using trained natural language processing (NLP) models to identify various types of hate. Seven were classified: race, gender, religion, antisemitism, gender identity/sexual

orientation (GISO), immigration, and ethnicity/identitarian/nationalism (EIN). The resulting NLP models each have high accuracy scores (at least 91%) and sample outputs have been validated by human subject matter experts using various reliability metrics, all of which are published in³¹.

Our social media data collection is, like others, necessarily imperfect. However, the resulting hate universe does encompass several billion individual accounts. Our focus is on changes in this hate universe's topology and content through the 2020 presidential election (November 3, 2020) and subsequent Capitol attack (January 6, 2021). We are not concerned here with the extent to which these changes have a causal effect on the offline events or are simply a response.

Data availability

The datasets used in this study are at <https://github.com/gwdonlab/data-access>. The original data contain sensitive information from social media platforms; to comply with data protection standards and avoid potential misuse, the raw data cannot be shared publicly. However, the processed derivative datasets can be used to reproduce the results in the study.

Code availability

Plots were created using Python (Matplotlib³³). The code is available at <https://github.com/gwdonlab/data-access>. The network visualization in Fig. 3a was created using Gephi and the network visualization in Fig. 3b was created using Python (PyVis). Together, this provides readers with access to the minimum dataset that is necessary to interpret, verify, and extend the research in the article. Cosmetic changes to figures were carried out in Adobe Illustrator, a commercially available piece of software.

Received: 7 June 2024; Accepted: 25 September 2024;

Published online: 29 October 2024

References

- Schackmuth, A. *Extremism, fake news and hate: effects of social media in the post-truth era*. Ph.D. thesis, DePaul University (2018).
- Starbird, K. Disinformation's spread: bots, trolls and all of us. *Nature* **571**, 449–449 (2019).
- Johnson, N. F. et al. Hidden resilience and adaptive dynamics of the global online hate ecology. *Nature* **573**, 261–265 (2019).
- Zannettou, S. et al. *On the origins of memes by means of fringe web communities*, IMC '18, 188–202 (Association for Computing Machinery, New York, NY, USA, 2018). <https://doi.org/10.1145/3278532.3278550>.
- Rogers, R. Deplatforming: following extreme internet celebrities to telegram and alternative social media. *Eur. J. Commun.* **35**, 213–229 (2020).
- Frischlich, L., Schatto-Eckrodt, T. & Völker, J. *Withdrawal to the shadows: dark social media as opportunity structures for extremism* Vol. 3 of *CoRE-NRW Forschungspapier*. Vol. 38 (Bonn International Centre for Conflict Studies (BICC) gGmbH, Bonn, 2022). <https://nbn-resolving.org/urn:nbn:de:0168-ssoar-88967-8>.
- Stress in America 2023: A nation recovering from collective trauma. <https://www.apa.org/news/press/releases/stress/2023/collective-trauma-recovery> (2023).
- Liu, Y.-Y. & Barabási, A.-L. Control principles of complex systems. *Rev. Modern Phys.* **88**, 035006 (2016).
- Cynthia Miller-Idriss. *Hate in the Homeland: The New Global Far Right* (Princeton University Press, 2020). <https://press.princeton.edu/books/hardcover/9780691203836/hate-in-the-homeland>.
- Gelfand, M. J. Cultural evolutionary mismatches in response to collective threat. *Curr. Directions Psychol. Sci.* **30**, 401–409 (2021).
- Douek, E. Content moderation as systems thinking. *Harv. L. Rev.* **136**, 526 (2022).
- Zheng, M., Sear, R. F., Illari, L., Restrepo, N. J. & Johnson, N. F. Adaptive link dynamics drive online hate networks and their mainstream influence. *npj Complexity* **1**, 2 (2024).
- Watts, D. J. *Small worlds: the dynamics of networks between order and randomness* Vol. 36 (Princeton university press, 2004).
- Bilewicz, M., Kamińska, O. K., Winiewski, M. & Soral, W. From disgust to contempt-speech: the nature of contempt on the map of prejudicial emotions. *Behavioral and Brain Sciences* **40**, e228 (2017).
- Rachel B. & Laura L. A new approach to assessing the role of technology in spurring and mitigating conflict: evidence from research and practice. <https://www.jstor.org/stable/26508121> (2018).
- Cinelli, M. et al. Dynamics of online hate and misinformation. *Sci. Rep.* **11**, 22083 (2021).
- Chen, X., Sin, S.-C. J., Theng, Y.-L. & Lee, C. S. Why students share misinformation on social media: motivation, gender, and study-level differences. *J. Acad. Librarianship* **41**, 583–592 (2015).
- Chen, X., Sin, S.-C. J., Theng, Y.-L. & Lee, C. S. *Why Do Social Media Users Share Misinformation?*, JCDL '15, 111–114 (Association for Computing Machinery, New York, NY, USA, 2015). <https://doi.org/10.1145/2756406.2756941>.
- Kim, Y. A. & Ahmad, M. A. Trust, distrust and lack of confidence of users in online social media-sharing communities. *Knowledge-Based Syst.* **37**, 438–450 (2013).
- Shahi, M. Protecting Democracy Online in 2024 and Beyond. Tech. Rep., The Center for American Progress. <https://www.americanprogress.org/article/protecting-democracy-online-in-2024-and-beyond/> (2023).
- Lee, C. S., Merizalde, J., Colautti, J. D., An, J. & Kwak, H. Storm the capitol: linking offline political speech and online twitter extra-representational participation on qanon and the january 6 insurrection. *Front. Sociol.* **7**, 876070 (2022).
- Leifeld, P. Polarization in the social sciences: Assortative mixing in social science collaboration networks is resilient to interventions. *Phys. A Statist. Mech. Appl.* **507**, 510–523 (2018).
- Oehlers, M. & Fabian, B. Graph metrics for network robustness—a survey. *Mathematics* **9**, 895 (2021).
- Jacomy, M., Venturini, T., Heymann, S. & Bastian, M. Forceatlas2, a continuous graph layout algorithm for handy network visualization designed for the gephi software. *PLoS ONE* **9**, e98679 (2014).
- Ekman, M. The great replacement: Strategic mainstreaming of far-right conspiracy claims. *Convergence* **28**, 1127–1143 <https://doi.org/10.1177/13548565221091983> (2022).
- Digital Services Act and Digital Markets Act – Stepping stones to a level playing field in Europe* (European Economic and Social Committee, 2021). <https://doi.org/10.2864/74293>.
- Schulze, H. et al. Far-right conspiracy groups on fringe platforms: a longitudinal analysis of radicalization dynamics on telegram. *Converg. Int. J. Res. New Media Technol.* **28**, 1103–1126 (2022).
- Jamie Gangel, D. O. Talk of overturning the 2020 election on new social media platforms used by qanon followers sparks fears of further violence. <https://www.cnn.com/2021/06/02/politics/telegram-qanon-trump-supporters/index.html> (2021).
- Guhl, J. & Davey, J. A safe space to hate: white supremacist mobilisation on telegram. *Institute for Strategic Dialogue* **26**, 1–11 (2020).
- Walther, S. & McCoy, A. Us extremism on telegram: fueling disinformation, conspiracy theories, and accelerationism. *Perspect. Terrorism* **15**, 100–124 (2021).
- Lupu, Y. et al. Offline events and online hate. *PLOS ONE* **18**, e0278511 (2023).
- Velásquez, N. et al. Online hate network spreads malicious COVID-19 content outside the control of individual social media platforms. *Sci. Rep.* **11**, 11549 (2021).
- Hunter, J. D. Matplotlib: A 2d graphics environment. *Comput. Sci. Eng.* **9**, 90–95 (2007).

Acknowledgements

N.J. is supported by US Air Force Office of Scientific Research awards FA9550-20-1-0382 and FA9550-20-1-0383, and by The John Templeton Foundation.

Author contributions

A.V. analyzed the data and generated Figs. 2, 3, and 4. R.S. contributed to obtaining, organizing, and preserving the data resources, as well as generated Fig. 1. N.J. supervised the project. All authors were involved in reviewing the final manuscript, and in the conceptualization, methodology, and validation. All correspondence and material requests should be addressed to N.J.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s44260-024-00018-8>.

Correspondence and requests for materials should be addressed to Neil Johnson.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2024