

The online competition between pro- and anti-vaccination views

<https://doi.org/10.1038/s41586-020-2281-1>

Received: 8 January 2020

Accepted: 7 April 2020

Published online: 13 May 2020

 Check for updates

Neil F. Johnson^{1,2✉}, Nicolas Velásquez², Nicholas Johnson Restrepo², Rhys Leahy², Nicholas Gabriel¹, Sara El Oud¹, Minzhang Zheng³, Pedro Manrique⁴, Stefan Wuchty⁵ & Yonatan Lupu⁶

Distrust in scientific expertise^{1–14} is dangerous. Opposition to vaccination with a future vaccine against SARS-CoV-2, the causal agent of COVID-19, for example, could amplify outbreaks^{2–4}, as happened for measles in 2019^{5,6}. Homemade remedies^{7,8} and falsehoods are being shared widely on the Internet, as well as dismissals of expert advice^{9–11}. There is a lack of understanding about how this distrust evolves at the system level^{13,14}. Here we provide a map of the contention surrounding vaccines that has emerged from the global pool of around three billion Facebook users. Its core reveals a multi-sided landscape of unprecedented intricacy that involves nearly 100 million individuals partitioned into highly dynamic, interconnected clusters across cities, countries, continents and languages. Although smaller in overall size, anti-vaccination clusters manage to become highly entangled with undecided clusters in the main online network, whereas pro-vaccination clusters are more peripheral. Our theoretical framework reproduces the recent explosive growth in anti-vaccination views, and predicts that these views will dominate in a decade. Insights provided by this framework can inform new policies and approaches to interrupt this shift to negative views. Our results challenge the conventional thinking about undecided individuals in issues of contention surrounding health, shed light on other issues of contention such as climate change¹¹, and highlight the key role of network cluster dynamics in multi-species ecologies¹⁵.

Social media companies are struggling to control online health dis- and misinformation, for example, during the COVID-19 pandemic in 2020⁸. Online narratives tend to be nurtured in in-built community spaces that are a specific feature of platforms such as Facebook (for example, fan pages) but not Twitter^{3,16–18}. Previous studies have pointed out that what is missing is a system-level understanding at the level of millions of people¹³, whereas another study¹⁴ has highlighted the need to understand the role of algorithms and bots in the amplification of risk among unwitting crowds.

Here we provide a system-level analysis of the multi-sided ecology of nearly 100 million individuals expressing views regarding vaccination, which are emerging from the approximately 3 billion users of Facebook from across countries, continents and languages (Figs. 1, 2). The segregation in Fig. 1a arises spontaneously. Individuals come together into interlinked clusters. Each cluster is a Facebook page and its members (that is, fans) who subscribe to, share and interact with the content and narratives of that Facebook page. A link from cluster A to B exists when A recommends B to all its members at the page level, as opposed to a page member simply mentioning a cluster. Each red node is a cluster of fans of a page with anti-vaccination content. Cluster size is given by the number of fans, for example, the page ‘RAGE Against the Vaccines’ has a size of approximately 40,000 members. Blue nodes are

clusters that support vaccinations, for example, the page ‘The Gates Foundation’ has a size (that is, number of fans) of more than 1 million. Each green node is a page focused around vaccines or another topic—for example, a school parent association—that has become linked to the vaccine debate but for which the stance is still undecided. Support and potential recruitment of these green clusters (crowds) is akin to a battle for the ‘hearts and minds’ of individuals in insurgent warfare.

Seven unexpected features of this cluster network (Fig. 1) and its evolution (Fig. 2) together explain why negative views have become so robust and resilient, despite a considerable number of news stories that supported vaccination and were against anti-vaccination views during the measles outbreak of 2019 and recent efforts against anti-vaccination views from pro-vaccination clusters and Facebook.

First, although anti-vaccination clusters are smaller numerically (that is, have a minority total size, Fig. 1d) and have ideologically fringe opinions, anti-vaccination clusters have become central in terms of the positioning within the network (Fig. 1a). Specifically, whereas pro-vaccination clusters are confined to the smallest two of the three network patches (Fig. 2a), anti-vaccination clusters dominate the main network patch in which they are heavily entangled with a very large presence of undecided clusters (more than 50 million undecided individuals). This means that the pro-vaccination clusters in the smaller

¹Physics Department, George Washington University, Washington, DC, USA. ²Institute for Data, Democracy and Politics, George Washington University, Washington, DC, USA. ³Institute for Quantitative Health Science and Engineering, Michigan State University, East Lansing, MI, USA. ⁴Theoretical Biology and Biophysics Group, Los Alamos National Laboratory, Los Alamos, NM, USA. ⁵Department of Computer Science, University of Miami, Coral Gables, FL, USA. ⁶Department of Political Science, George Washington University, Washington, DC, USA. ✉e-mail: neiljohnson@gwu.edu

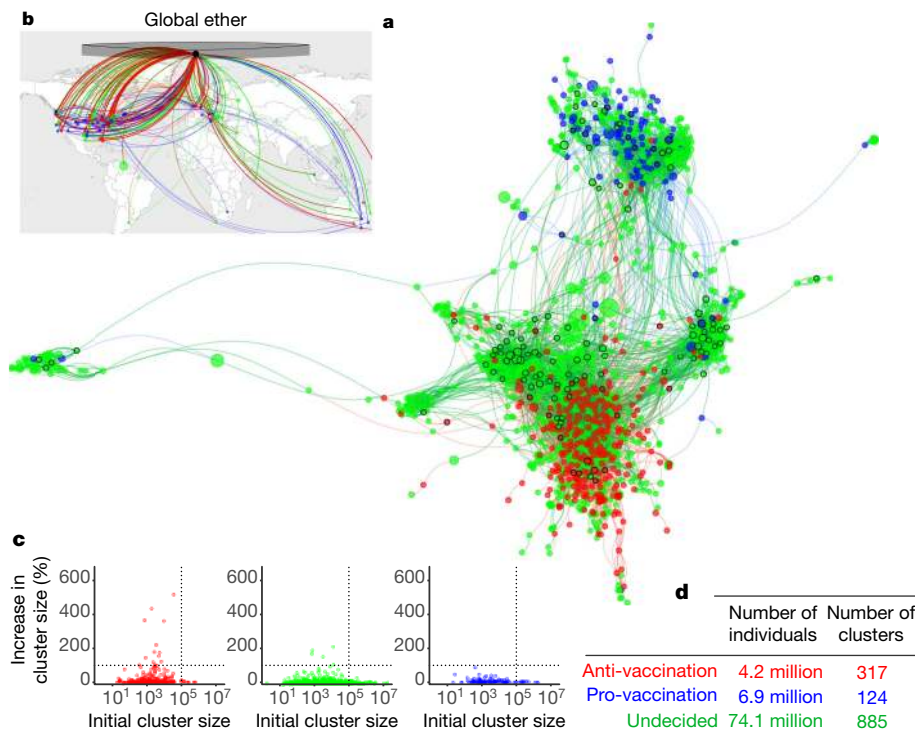


Fig. 1 | Online ecology of vaccine views. **a**, Snapshot from 15 October 2019 of the connected component in the complex ecology of undecided (green), anti-vaccination (red) and pro-vaccination (blue) views comprising nearly 100 million individuals in clusters (pages) associated with the vaccine topic on Facebook. The colour segregation is an emergent effect (that is, not imposed). Cluster sizes are determined by the number of members of the Facebook page. Black rings show clusters with more than 50% out-link growth. Each link

between nodes has the colour of the source node. **b**, Global spread of Fig. 1a for a small number of clusters. The ‘global ether’ represents clusters that remain global (grey). **c**, Anti-vaccination clusters have a stronger growth in cluster size. Each coloured dot is a node; data are from February–October 2019.

d, Anti-vaccination individuals are an overall numerical minority compared with pro-vaccination individuals; however, anti-vaccination individuals form more separate clusters.

network patches may remain ignorant of the main conflict and have the wrong impression that they are winning.

Second, instead of the undecided population being passively persuaded by the anti- or pro-vaccination populations, undecided individuals are highly active: the undecided clusters have the highest growth of new out-links (Fig. 1a), followed by anti-vaccination clusters. Moreover, it is the undecided clusters who are entangled with the anti-vaccination clusters in the main network patch that tend to show this high out-link growth. These findings challenge our current thinking that undecided individuals are a passive background population in the battle for ‘hearts and minds’.

Third, anti-vaccination individuals form more than twice as many clusters compared with pro-vaccination individuals by having a much smaller average cluster size. This means that the anti-vaccination population provides a larger number of sites for engagement than the pro-vaccination population. This enables anti-vaccination clusters to entangle themselves in the network in a way that pro-vaccination clusters cannot. As a result, many anti-vaccination clusters manage to increase their network centrality (Fig. 2b) more than pro-vaccination clusters despite the media ambience that was against anti-vaccination views during 2019, and manage to reach better across the entire network (Fig. 2a).

Fourth, our qualitative analysis of cluster content shows that anti-vaccination clusters offer a wide range of potentially attractive narratives that blend topics such as safety concerns, conspiracy theories and alternative health and medicine, and also now the cause and cure of the COVID-19 virus. This diversity in the anti-vaccination narratives is consistent with other reports in the literature⁴. By contrast, pro-vaccination views are far more monothematic. Using aggregation mathematics and a multi-agent model, we have reproduced the ability

of anti-vaccination support to form into an array of many smaller-sized clusters, each with its own nuanced opinion, from a population of individuals with diverse characteristics (Fig. 3b and Supplementary Information).

Fifth, anti-vaccination clusters show the highest growth during the measles outbreak of 2019, whereas pro-vaccination clusters show the lowest growth (Fig. 1c). Some anti-vaccination clusters grow by more than 300%, whereas no pro-vaccination cluster grows by more than 100% and most clusters grow by less than 50%. This is again consistent with the anti-vaccination population being able to attract more undecided individuals by offering many different types of cluster, each with its own type of negative narrative regarding vaccines.

Sixth, medium-sized anti-vaccination clusters grow most. Whereas larger anti-vaccination clusters take up the attention of the pro-vaccination population, these smaller clusters can expand without being noticed. This finding challenges a broader theoretical notion of population dynamics that claims that groups grow through preferential attachment (that is, a larger size attracts more recruits). Therefore, a different theory is needed that generalizes the notion of size-dependent growth to include heterogeneity (Fig. 3b).

Seventh, geography (Fig. 1b) is a favourable factor for the anti-vaccination population. Anti-vaccination clusters either self-locate within cities, states or countries, or remain global. Figure 1b shows a small sample of the connectivity between localized and global clusters. Any two local clusters (for example, two US states) are typically interconnected through an ether of global clusters and so feel part of both a local and global campaign.

The complex cluster dynamics between undecided, anti-vaccination and pro-vaccination individuals (Figs. 1, 2) mean that traditional mass-action modelling¹⁹ cannot be used reliably for

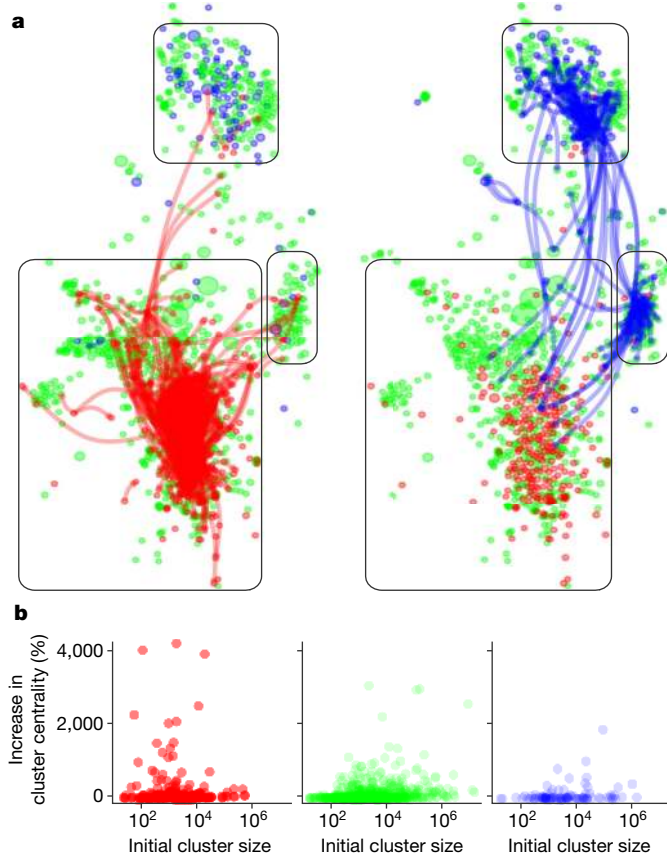


Fig. 2 | Temporal evolution of online ecology. **a**, Link growth during February–October 2019 for anti-vaccination (red; left) and pro-vaccination (blue; right) clusters. Anti-vaccination clusters successfully added many new links within the largest network patch and between network patches, despite the media ambience against anti-vaccination views during the measles outbreak in 2019. The underlying clusters are identical to Fig. 1a, that is, each network patch is a clustered region of clusters from Fig. 1a. **b**, Anti-vaccination clusters have a stronger growth in node eigencentality—which indicates the influence of a node in a network—than pro-vaccination clusters. Data are from February–October 2019.

predictions or policies. Mass-action models suggest that given the large pro-vaccination majority (Fig. 1d), the anti-vaccination clusters should shrink relative to pro-vaccination clusters under attrition, which is the opposite of what happened in 2019. Figure 3a shows the importance of these missing cluster dynamics using a simple computer simulation with mass-action interactions only between clusters, not populations. The simulation reproduces the increase in anti-vaccination support in 2019, and predicts that anti-vaccination views will dominate in approximately 10 years (Fig. 3a). These findings suggest a new theoretical framework to describe this ecology, and inform new policies that allow pro-vaccination entities, or the platform itself, to choose their preferred scale at which to intervene.

If the preferred intervention scale is at the scale of individual clusters (Fig. 3b), then Fig. 1a can identify and target the most central and potentially influential anti-vaccination clusters. Our clustering theory (see Supplementary Information) predicts that the growth rate of an influential anti-vaccination cluster can be reduced, and the onset time for future anti-vaccination (or connected undecided) clusters delayed, by increasing the heterogeneity within the cluster. This reduces parameter F of our theory, which captures the similarity of pairs of engaged individuals N in a particular narrative. The anti-vaccination (or connected undecided) cluster size $C(t)$ is reduced to $C(t) = N(1 - W[-2Ft/N] \exp[-2Ft/N]) / [-2Ft/N]$ where W is the Lambert function²⁰, and the

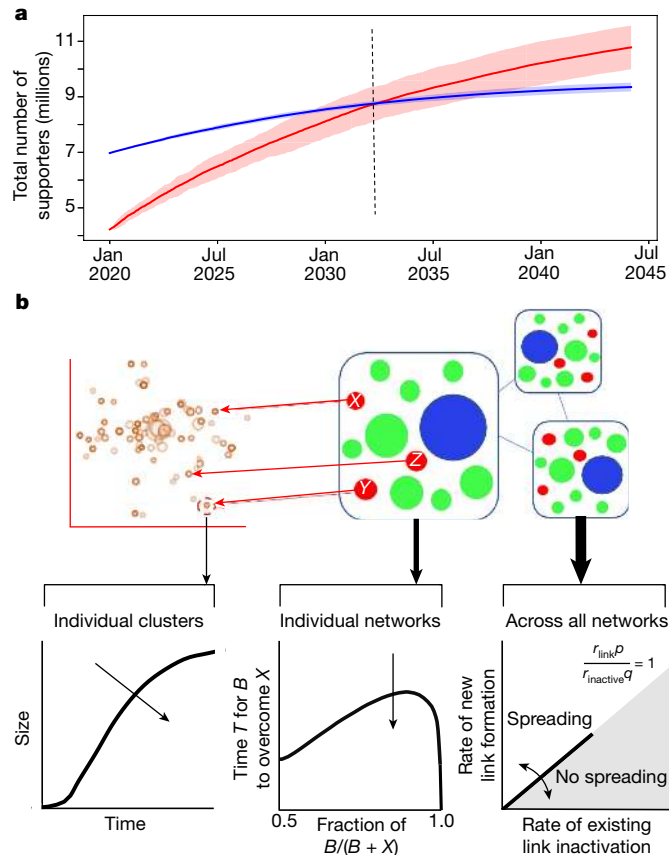


Fig. 3 | Predictions and interventions. **a**, Theoretical prediction for the future total size of anti-vaccination and pro-vaccination support without new interventions (coloured lines with 2σ bands from the simulation). Under the present conditions, it predicts that total anti-vaccination support reaches dominance in around 10 years. **b**, Top left, our theoretical model predicts that, as observed empirically, many smaller-sized anti-vaccination clusters form, with each cluster having its own nuanced type of narrative (for example, X, Y, Z) that surrounds a general topic (vaccines in this case). Bottom left, the predicted growth profile of individual clusters can be manipulated by altering the heterogeneity to delay the onset and decrease the growth. Bottom middle, pro-vaccination population B is predicted to overcome the anti-vaccination population, or persuade the undecided population, X , within a given network patch in time T by using Fig. 1a to identify and then engage with all the clusters. Bottom right, the link dynamics can be manipulated to prevent the spread of negative narratives. See Supplementary Information for all mathematical details.

delayed onset time for a future nascent anti-vaccination (or connected undecided) cluster is $t_{\text{onset}} = N/2F$. If instead the preferred intervention scale is at the scale of network patches (single or connected; Fig. 3b), our theoretical framework predicts that the pro-vaccination population (B) can beat the anti-vaccination population or persuade the undecided population (X) within a given network patch S over time T by using Fig. 1a to identify and then proactively engage with the other clusters in S , irrespective of whether they are linked or not:

$$T = \frac{1}{x_c(2 - d_b - d_x)} \left[4X + (B - X) \ln \frac{X(B - X + x_c)}{x_c B} \right] \quad (1)$$

where d_b and d_x are rates at which the activity of an average cluster becomes inactive (for example, no more posts in the cluster), and B and X are the current total sizes of the respective populations²¹. If instead the preferred intervention scale is the entire global ecology (Fig. 1a), this framework predicts the condition $r_{\text{link}}^D / r_{\text{inactive}}^Q < 1$ to prevent the

spreading of negative narratives²² (Fig. 3b), where r_{link} and r_{inactive} are the rates at which links are formed and become inactive between sets of clusters; p is the average rate at which a cluster shares material with another cluster and q is the average rate at which a cluster becomes inactive. Conversely, $r_{\text{link}}p/r_{\text{inactive}}q > 1$ predicts the condition for system-wide spreading of intentional counter-messaging. As p and q are properties related to a single average cluster and are probably more difficult to manipulate, the best intervention at this system-wide scale is to manipulate the rate at which links are created (r_{link}) and/or the rate at which links become inactive (r_{inactive}).

Finally, we note that our analysis is incomplete and that other channels of influence should be explored. However, similar behaviours should arise in any online setting in which clusters can form. Our mathematical formulae are approximations. We could define links differently, for example, as numbers of members that clusters have in common. However, such information is not publicly available on Facebook. Furthermore, our previous study of a Facebook-like platform for which such information was available showed that the absence or presence of such a link between pages acts as a proxy for low or high numbers of common members. How people react to intervention is ultimately an empirical question^{23,24}. One may also wonder about external agents or entities—however, clusters tend to police themselves for bot-like or troll behaviour. The crudely power law-like distribution of the cluster sizes of anti-vaccination clusters suggests that any top-down presence is not dominant.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41586-020-2281-1>.

1. Haerlin, B. & Parr, D. How to restore public trust in science. *Nature* **400**, 499 (1999).
2. Larson, H. A lack of information can become misinformation. *Nature* **580**, 306 (2020).
3. Martin, B. Texas anti-vaxxers fear mandatory COVID-19 vaccines more than the virus itself. *Texas Monthly* <http://www.texasmonthly.com/news/texas-anti-vaxxers-fear-mandatory-coronavirus-vaccines/> (18 March 2020).
4. Kata, A. A postmodern Pandora's box: anti-vaccination misinformation on the Internet. *Vaccine* **28**, 1709–1716 (2010).

5. Kennedy, M. Samoa arrests anti-vaccination activist as measles death toll rises. *NPR News* <http://www.npr.org/2019/12/06/785487606/samoa-arrests-anti-vaccination-activist-as-measles-death-toll-rises> (6 December 2019).
6. Givertash, L. Global measles cases surge amid stagnating vaccinations. *NBC News* <http://www.nbcnews.com/news/amp/ncna1096921> (6 December 2019).
7. CTCA. The dangers of fake medical news. *Cancer Treatment Centers of America* <http://www.cancercenter.com/discussions/blog/the-dangers-of-fake-medical-news/> (6 April 2017).
8. Frenkel, S., Alba, D. & Zhong, R. Surge of virus misinformation stumps Facebook and Twitter. *The New York Times* <https://www.nytimes.com/2020/03/08/technology/coronavirus-misinformation-social-media.html> (8 March 2020).
9. Shapiro, N. Vaccine proponents receive death threats. Again. *Forbes* <https://www.forbes.com/sites/ninashapiro/2019/07/22/vaccine-proponents-receive-death-threats-again/#5025011e2cfd> (22 July 2019).
10. Zadrozny, B. & Edwards, E. Anti-vaccine groups take dangerous online harassment into the real world. *NBC News* <https://www.nbcnews.com/news/amp/ncna1096461> (6 December 2019).
11. U.S. Government Accountability Project. Australian climate scientists targeted by death threats. (4 June 2018); www.whistleblower.org/general/climate-science-watch/australian-climate-scientists-targeted-by-death-threats/
12. Huang, X. et al. Examining patterns of influenza vaccination in social media. In *Proc. AAAI Joint Workshop on Health Intelligence (W3PHIAI)* 542–546 (2017).
13. DiResta, R. Of virality and viruses: the anti-vaccine movement and social media. *NAPSNet Special Reports* <https://nautilus.org/napsnet/napsnet-special-reports/of-virality-and-viruses-the-anti-vaccine-movement-and-social-media/> (8 November 2018).
14. Starbird, K. Disinformation's spread: bots, trolls and all of us. *Nature* **571**, 449 (2019).
15. Fryxell, J. M., Mosser, A., Sinclair, A. R. & Packer, C. Group formation stabilizes predator-prey dynamics. *Nature* **449**, 1041–1043 (2007).
16. Ammari, T. & Schoenebeck, S. “Thanks for your interest in our Facebook group, but it's only for dads”: social roles of stay-at-home dads. In *CSCW '16: Proc. 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing* 1363–1375 <https://doi.org/10.1145/2818048.2819927> (2016).
17. Johnson, N. F. et al. Hidden resilience and adaptive dynamics of the global online hate ecology. *Nature* **573**, 261–265 (2019).
18. Johnson, N. F. et al. New online ecology of adversarial aggregates: ISIS and beyond. *Science* **352**, 1459–1463 (2016).
19. Murray, J. D. *Mathematical Biology* (Springer, 2003).
20. Manrique, P. D., Zheng, M., Cao, Z., Restrepo, E. M. & Johnson, N. F. Generalized gelation theory describes onset of online extremist support. *Phys. Rev. Lett.* **121**, 048301 (2018).
21. Zhao, Z., Bohorquez, J. C., Dixon, A. & Johnson, N. F. Anomalous slow attrition times for asymmetric populations with internal group dynamics. *Phys. Rev. Lett.* **103**, 148701 (2009).
22. Zhao, Z. et al. Effect of social group dynamics on contagion. *Phys. Rev. E* **81**, 056107 (2010).
23. Frimer, J. A. et al. Liberals and conservatives are similarly motivated to avoid exposure to one another's opinions. *J. Exp. Soc. Psychol.* **72**, 1–12 (2017).
24. Bail, C. A. et al. Exposure to opposing views on social media can increase political polarization. *Proc. Natl Acad. Sci. USA* **115**, 9216–9221 (2018).

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

© The Author(s), under exclusive licence to Springer Nature Limited 2020

Methods

We used clusters (Facebook pages) as the unit for our analysis^{17,18}. Our cluster approach does not require any private information of individuals. The ForceAtlas2 layout of Gephi (Fig. 1a) simulates a physical system in which nodes (clusters) repel each other while links act as springs. It is colour-agnostic, that is, the colour segregation in Fig. 1a emerges spontaneously and is not in-built. Nodes that appear closer to each other have local environments that are more highly interconnected, whereas nodes that are far apart do not. Our data collection uses the same cluster snowballing methodology as described previously^{17,18}, that is, a combination of automated processes and human subject-matter analysis. Each cluster (Facebook page) directly receives the feed of narratives and other material from that page and all members (fans) can engage in the discussions and posting activity. Figure 1b uses the declared location of each cluster. Derivations of the equations are provided in the Supplementary Information; they build on published results^{20–22} and our approach complements other studies^{25–33}. Equation (1) is easily generalizable, but for simplicity we assume here a minimal model in which each pro-vaccination cluster has a narrative that persuades on average x_c members of each cluster X in each engagement, and the pro-vaccination cluster B picks a cluster X randomly within S . Equation (1) also applies to the full anti-vaccination–undecided ecology if we take the X -related quantities in equation (1) as weighted anti-vaccination–undecided values from Fig. 1a. The formula $r_{\text{link}D}/r_{\text{inactive}Q} < 1$, to prevent spreading, accounts for the key feature of cluster interconnections that change over time and can be applied to spreading between anti-vaccination clusters, between undecided clusters, or between both anti-vaccination and undecided clusters using weighted values. For the model in Fig. 3a, rates of cluster interaction are given to the first order by the relative number of links of each type with Y -undecided interactions that yield more recruits for Y when Y is anti-vaccination than when Y is pro-vaccination (see Supplementary Information). The fidelity of these predictions is affected by the approximations of the model. For Fig. 3b, all parameters can be extracted from data or estimated from simulations. In the top left graph of Fig. 3b, two dimensions are shown for simplicity, for example, the degree of belief in government conspiracy and the degree of belief in alternative health, but similar plots emerge for other numbers of dimensions. In the bottom middle graph of Fig. 3b, the total initial size B (pro-vaccination population) plus size X (for example, anti-vaccination population) is kept constant. Although this leaves open the details of the conversion process for each X cluster, a previous study³⁰ has shown that such conversion within an online cluster occurs and can be rapid. T for mass-action theory would tend to decrease monotonically as B

increases; however, our theory in equation (1) shows a counterintuitive dependence because smaller but finite numbers of X clusters take the pro-vaccination clusters longer to find. Only functional forms are shown (that is, no numbers) as the underlying formulae and models are not restricted by specific numerical choices of parameter values.

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

The dataset used to generate this paper is provided in the Supplementary Information.

Code availability

The computer code written by the authors is provided in the Supplementary Information. The open-source software packages Gephi and R were used to produce the figures.

25. Gavrillets, S. Collective action and the collaborative brain. *J. R. Soc. Interface* **12**, 20141067 (2015).
26. Havlin, S., Kenett, D. Y., Bashan, A., Gao, J. & Stanley, H. E. Vulnerability of network of networks. *Eur. Phys. J. Spec. Top.* **223**, 2087–2106 (2014).
27. Newman, M. E. J. Modularity and community structure in networks. *Proc. Natl Acad. Sci. USA* **103**, 8577–8582 (2006).
28. May, R. & McLean, A. R. *Theoretical Ecology Principles and Applications* (Oxford Univ. Press, 2007).
29. Palla, G., Barabási, A. L. & Vicsek, T. Quantifying social group evolution. *Nature* **446**, 664–667 (2007).
30. Centola, D., Becker, J., Brackbill, D. & Baronchelli, A. Experimental evidence for tipping points in social convention. *Science* **360**, 1116–1119 (2018).
31. Broniatowski, D. A. et al. Weaponized health communication: Twitter bots and Russian trolls amplify the vaccine debate. *Am. J. Public Health* **108**, 1378–1384 (2018).
32. Dredze, M., Broniatowski, D. A. & Hilyard, K. M. Zika vaccine misconceptions: a social media analysis. *Vaccine* **34**, 3441–3442 (2016).
33. Lama, Y. et al. Discordance between human papillomavirus Twitter images and disparities in human papillomavirus risk and disease in the United States: mixed-methods analysis. *J. Med. Internet Res.* **20**, e10244 (2018).

Acknowledgements We thank D. Broniatowski for related discussions.

Author contributions All authors contributed to the research design, the analysis and writing the paper.

Competing interests The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41586-020-2281-1>.

Correspondence and requests for materials should be addressed to N.F.J.

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

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see [Authors & Referees](#) and the [Editorial Policy Checklist](#).

Please do not complete any field with "not applicable" or n/a. Refer to the help text for what text to use if an item is not relevant to your study. For final submission, please carefully check your responses for accuracy; you will not be able to make changes later.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided
Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
Give P values as exact values whenever suitable.
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection The data was all collected from entirely publicly accessible information on the Internet. This entirely public data required to reproduce the empirical results in the paper is contained at the end of the SI document.

Data analysis The theoretical model output for figure 3 was obtained from a computer simulation by the authors, and the code for this simulation is given in the SI. The open-source software packages Gephi and R were used to analyze the network and to plot the figures from the data.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

This data required to reproduce the empirical results reported in the paper, is contained at the end of the SI document.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

- Life sciences Behavioural & social sciences Ecological, evolutionary & environmental sciences

Behavioural & social sciences study design

All studies must disclose on these points even when the disclosure is negative.

| | |
|-------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Study description | We analyzed freely available public information on the Internet |
| Research sample | As described in the paper, we collected information about clusters and kept moving to the next cluster until we returned to the same clusters. Hence our data is not technically a sample. |
| Sampling strategy | Our data is not a sample. We focus on Facebook, as stated in the paper. We collected information about clusters and kept moving to the next cluster until we returned to the same clusters. |
| Data collection | The data was all collected from entirely publicly accessible information on the Internet. This entirely public data required to reproduce the empirical results in the paper is contained at the end of the SI document. |
| Timing | The data were collected during 2019, as explained in the paper. |
| Data exclusions | We only collected public data from Facebook. |
| Non-participation | Not applicable. |
| Randomization | Not applicable. |

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

| n/a | Included in the study |
|-------------------------------------|------------------------------------------------------|
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Antibodies |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Eukaryotic cell lines |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Palaeontology |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Animals and other organisms |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Human research participants |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Clinical data |

Methods

| n/a | Included in the study |
|-------------------------------------|-------------------------------------------------|
| <input checked="" type="checkbox"/> | <input type="checkbox"/> ChIP-seq |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Flow cytometry |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> MRI-based neuroimaging |