Science Advances
AAAS

Supplementary Materials for

Losing the battle over best-science guidance early in a crisis: COVID-19 and beyond

Lucia Illari *et al.*

Corresponding author: Neil F. Johnson, neiljohnson@gwu.edu

Sci. Adv. **8**, eabo8017 (2022)
DOI: 10.1126/sciadv.abo8017

The PDF file includes:

Other Supplementary Material for this manuscript includes the following:

Replication programs and data

Supplementary Material (SM)

Contents:

Section 1: Methodology for collecting data, building network, and analysis in the paper

A list of communities (nodes) and links, together with full analysis files for the mathematical model and data fitting in this study, will be provided with the Supplementary Material (SM) on publication.

Our methodology for data collection and classification follows the 2020 paper referenced in the main text, i.e., Johnson, N.F., N. Velasquez, N. Johnson Restrepo, R. Leahy, N. Gabriel, S. El Oud, M. Zheng, P. Manrique, S. Wuchty, Y. Lupu. The online competition between pro- and anti-vaccination views. Nature 582, 230–233 (2020), henceforth called "our 2020 vaccine ecology paper". In the present paper, we go further with the classification by further categorizing the neutral communities (the 'greens' in our 2020 vaccine ecology paper; here, the ‘golds’) according to their page's declared interest, e.g. parenting community. This is relatively easy since community pages are organized around a particular stated interest: see Sec. 2 for examples. As for the original classifications in our 2020 vaccine ecology paper, this entire process is carried out with the help of three subject matter experts working independently to classify the neutrals ('golds') by interest, and then cross-checking between them for any differences, as discussed below. There were no remaining contentious cases following this process. Here we refer to our research team, including the subject matter experts, simply as "we".

Covid-19 guidance (where "guidance" is defined in the Oxford Dictionary as "advice or information aimed at resolving a problem or difficulty") is advice or information that appears at a given time on a given page. We log the timestep at which it appears. We ignore irrelevant mentions where someone simply says an expletive word about Covid-19. Again, this approach could be improved to classify different types of information or guidance, but this goes beyond the scope of the current paper—particularly given the complexity already in this exposure dynamics as shown in Fig. 2(a). We state explicitly in the main paper that given the complexity already in Fig. 2(a), we needed to simplify other aspects of this paper's analysis. We do not determine an absolute fraction of scientific truth in each piece of Covid-19 guidance: this is in any case challenging since even the wildest anti content can contain truthful fragments, e.g., it has been shown experimentally that quantum dots can act as vaccine markers. Instead, manual analysis of all the community content confirms our expectation that pro communities do indeed promote best-science Covid-19 guidance, e.g., from Centers for Disease Control (CDC); anti communities' guidance opposes this; and neutral communities' guidance sits in between and gets further downgraded by additional non-scientific comments and opinions posted by page members. Even if a significant fraction of our classifications is wrong, the main conclusions are unchanged since they only depend on relative numbers: we checked this explicitly by introducing a 15% error into our classifications. We recognize that best-science guidance can change over time and may eventually be proven wrong—but that is rare.

Though the methodology for collecting nodes and links is written out explicitly in our 2020 vaccine ecology paper, we review the process for obtaining the network here. We start with a seed of manually identified pages discussing either vaccines, public policies about vaccination, or the pro-vs-anti vaccination debate. These were obtained by searching on Facebook’s search engine in 2018 and 2019 for key words and phrases involving vaccines. Then we captured the list of outbound links from these pages (i.e., which pages did each of the seed pages recommend to its members at the page level, see below) using Facebook’s GraphAPI endpoint for fanned pages. Some of these links went back to other pages already in the list, some did not. We repeated this process two times to obtain a final list of candidate nodes and the links between them. In the summer of 2021, Facebook removed this feature from its web interface, but as of Fall 2021 the page-level likes remain available in the background as an endpoint through Facebook’s GraphAPI. We then prune this list of nodes by classifying the pages by their content and description: specifically, we only included pages—and hence the links between them—when (1) the pages were talking about vaccines, or (2) the pages weren't talking about vaccines themselves but self-identified as a cause, community, or NGO and were connected within one step to pages active in the vaccine debate. This pruning of nodes in our set of nodes, created a smaller set of nodes, and it served to automatically create a smaller set of links since the set of links only comprises the ones connecting the nodes in the set, i.e., a pruned node set yields a pruned link set. A link from

A to B means page A explicitly lists page B as one of the pages to which it links, not necessarily because it agrees with page B's content but because page B's content is of interest to page A. Such a link could have appeared because page A's users noticed it and then recommended it to page A's managers who then established the link. Such a link creates an information conduit feeding content from page B to page A, likely exposing A's users to B's content. A list of communities (nodes) and links, together with full analysis files for the mathematical model and data fitting in this study, will be provided with the Supplementary Material (SM) on publication.

We then take this list of nodes and further classify them—based on their content—as being pro-vaccine (blue), neutral (gold), or anti-vaccine (red). To do this, we reviewed the fan page's posts, about, and self-described category. Pro and anti-vaccine classifications required that either (a) at least 2 of the most recent 25 posts dealt with the pro or anti vaccine debate, or (b) the fan page's title or about section described it as a pro or anti vaccine page. The neutral (gold) classification required that (c) either 1 of the most recent 25 posts referred to the vaccine debate, but that they did not explicitly took a side pro or against, or (d) the about section clearly classified the page as a neutral in the pro-vs-anti vaccine debate, or (e) none of the most recent 25 posts dealt with vaccines but the page self-classified as either an NGO, a cause, a community, or a grass roots organization. Most of the neutral nodes (gold) appeared in the second iteration of fanned pages as many nodes in the first iteration were more explicitly focused on the vaccine debate. This makes sense since they are more removed from the debate, i.e. they do not actively engage in it as a focus but rather are pages that have become interconnected with the pros and the antis who do. The subject matter experts in our team classified each node independently, and when two disagreed on their suggested classification (which happened approximately 15% of the time, largely since the material was ambiguous) all three reviewers discussed these cases in more depth. Agreement was reached in every case. Our team collecting and classifying the communities (Facebook pages) consisted of analysts who have several years experience in analyzing and classifying online community content on Facebook and other platforms, being trained through prior work on the content of their online communities in association with establishment health guidance as well as various types of extremism and hate (*69*, *70*).

We also comment for completeness on the nomenclature of likes, follows and feed. The information in what follows is taken directly from online material which is so common, with similar descriptions in many places, that we do not cite any particular source. In what follows, "you" and "your" can refers to the entire page. Facebook itself has essentially the same following text. A (Facebook) like is a person or page who has chosen to attach their name to your page as a 'fan' though this does not necessarily mean they agree with what is being said, rather it is that they have a keen interest in it and want to see content from it. The page will show up in the about section of your account under likes. A (Facebook) follow is a person or page who has chosen to see updates posted by a page on its news feed. If someone or another page follows a Facebook page, it means the content from the Facebook page will show up on their news feed. By default, when you like a page, you will automatically follow it and this means the content from the page will show up on your newsfeed. The numbers of likes and follows are very close to each other. They are not exactly the same because at some stage, some people who have liked your page may have manually unfollowed you. Your content won't show up on their timeline. We have never seen a significant difference between Likes & Followers, and the difference is typically a factor of about 5%. So, this technicality does not affect our conclusions in any way since we have tested adding up to 15% noise to our data.

In the rest of this section (Section 1) we provide answers to potential questions and critiques that may be raised about our methodology and research in the paper:

We could define nodes and links in another way, and we recognize that our dataset is ultimately an imperfect sample of some larger 'correct' network. Any definition of nodes and links—including ours in the present work—can of course be criticized since we are reducing the many attributes of a real system down to the few necessary to build a network. Instead, we could, for example, analyze the content of posts on the pages, identifying shares and URL links of posts from other pages to build a weighted, directed network that captures how users of one page are actually

exposed to the content of another page and driven to it. This can also be done without analyzing any individual-level data and preserving all privacy considerations and could be done in a scalable and automatic way using the graph API or CrowdTangle. This would also get round the fact that it is difficult now to obtain links. However, it also comes with its own downsides: the content may be in different languages and hence not so easy to identify as being shared, and the terminology can evolve quickly (e.g., the use of slang to avoid attracting the attention of Facebook moderators). This issue is worthy of a study to establish comparative advantages of the two approaches, but such a study is beyond the scope of the present paper. Ultimately, the 'best' choice of links and nodes will depend on the questions being asked about the system, since the network is less useful if it has too few nodes and if it is either very sparse in terms of links or too dense with every node essentially connected to most others. The best choice will also depend crucially on the available data and the level of granularity at which this data is reliable. Even with this, the complexity of the data will require some form of simplification in order to make the analysis tractable and understandable. Fortunately, the simplification that we make of each node being a community (page) does bring some advantages: for example, it avoids the need for accessing individual-level information and makes the definition of each node unique since each page has its own unique identification number. Also, because the process that we follow yields of order 1000 nodes (communities), but each contains of order 100,000 users, the network produced by our definition is visually manageable and yet interpretable at scale, as opposed to being overwhelmed with links or being too sparse. This means that the open-source software Gephi and its ForceAtlas2 algorithm that we use, which follows the principle of energy minimization, provides an uncluttered spatial representation. It also means the network is scalable to the population level, since 1000 nodes with 100,000 users each means we are potentially tapping into the behavior of 100 million individuals.

As we state and reviewed earlier in this SM Sec. 1, we use the same methodology for obtaining and filtering links as in our published 2020 vaccine ecology paper. It follows an objective recipe. The selection rule for links is that they must connect pages where (1) the pages were talking about vaccines, or (2) the pages weren't talking about vaccines themselves but self-identified as a cause, community, or NGO and were connected within one step to pages active in the vaccine debate. In terms of the rating agreement between subject matter experts, as for the original classifications in our 2020 vaccine ecology paper, this entire process is carried out with the help of three subject matter experts working independently to cross-checking the links and nodes and comparing differences between them. They disagreed in about 15% of cases and after discussion between them, there were no remaining contentious cases. In terms of stability, they tried varying the classification procedure but found that the overall segregation structure remained (Figs. 2(c)(d)) and the main conclusions of this paper are untouched. Including links with a more tenuous relevance added to the density of the network and made the network noisier, i.e., it obscured the structures observed. The meaningfulness of the resulting set of links, can only be tested empirically. In that respect, support for our approach comes from the fact that the null model (see SM Sec. 5) gives poor fit to the empirical data, whereas the actual network gives a good fit.

We also note that the set of nodes and links is obtained without regard for the specific classifications of the nodes (i.e., pro vs. anti. vs. neutral parenting etc.). So, the fact that the modular structures of such subpopulations emerge spontaneously in Figs. 2(c)(d), also lends support that the links that we identify are meaningful. Taking a devil's advocate position, if the link methodology were so subjective that the links are not meaningful in a scientific sense, then one would expect similar outcomes to a null model, which is not the case.

We have investigated that instead of using human expert coders, we train a supervised language model of the recent posts of each page, running it over each newly discovered communities (page) and including it in the network only if it can be considered relevant by the model. This gives a crudely similar list—however there are some glaring anomalies that are easily caught by a human expert, showing that while such automation improves volume and speed of analysis, it can also introduce glaring anomalies that a human would be very unlikely to let through. Also, it is hard to have the supervised language model treat material in Spanish, French, Russian etc. on an equal footing to English. So although such machine automation sounds desirable because it scales and promises to be less subjective

perhaps, it is clearly a trade-off. Overall, the fact that we had already published and had critiqued a similar list of nodes and edges in our 2020 vaccine ecology paper, and the present list—obtained from scratch—was similar, and the fact that number of nodes and links is not too large, and the fact that the network gives results which are very different from a null network model, give us confidence that our study though imperfect, is capturing significant features of the actual online system.

Another approach would be for us to list a series of terms that we consider relevant, and automatically include communities with sufficient posts using those terms. One might think that this could help ensure that the sampling process is more systematic, transparent, and reproducible. However, it also has a downside, in that we see the terms evolving in time - in part to avoid attracting the attention of Facebook and in part since the topics are evolving (e.g. the use of bleach then turned to other hot topics) and so such a list could also be forever playing catchup. Also, there is the prior question of terms translating between different languages. Ultimately whether this approach would perform well would need to be carefully studied and would represent a research project and paper in its own right.

We do *not* assume that the linked pages from a given page immediately influences the browsing and information sharing of the users subscribed to the linked pages, and hence that strong correlations in activity levels should accompany links between pages. The neutrals hardly ever discuss vaccines—but the fact that they have material from antivax pages appearing in their feed, will be noticed by many of them, and so when some later decision needs to be made such as taking their children to be vaccinated, they may think twice. So, it is not that the links carrying material automatically give rise to higher activity at that moment, but that they represent an influence. As referenced explicitly in the main text, (*60*) showed experimentally and theoretically that an online community can suddenly tip to an alternate stance in a reproducible way if there is a committed minority of around 25%. No amount of prior analysis of levels of activity in the experiment of (*60*) would have predicted this. Having said this, one might wonder if there is some evidence of elevated activity levels. Though this is not important to our paper, we note that we do indeed see this, as shown by the following earlier work of Nicholas Gabriel in our group: Consider two Pages, i and j, among a collection of N Facebook Pages, and suppose that Page i likes Page j so in our analysis there is a link from Page i to Page j. We calculate a measure of the correlation between Pages for unlinked and linked Pages, assigning 1 for a link and 0 for no link. The mean for the unlinked Pages is 0.115 and the mean of the linked Pages is 0.255. Since these samples are relatively large, and their means relatively far apart, this gives an extremely small p-value of $p = 0.0000037$ for the hypothesis that the means are the same. The reason the correlation is not much larger is likely that (1) the links are not fully utilized by Facebook's algorithms to "forward" posts to the users feeds. e.g., all posts may not be shown from a linked Page if they are not interesting, or historically they haven't had success sharing along a particular link. In that way the links may be present, and hence the feed read and understood by users of the other community, but not necessarily responded to. (2) People (e.g., parents) likely read and digest the content, but do not then go and post new activity in immediate response. In other words, the link had influence on the reader—but this influence likely sits passive until a later time. To extend this analysis, which again goes beyond the requirements and scope of the present paper, one could perform a Granger test of activity time series in pairs of linked and unlinked communities, including the link presence as a variable in the time series models and assessing whether it can be considered significantly positive. A second approach would be to retrieve URL links and shares from the posts of all communities and run the test that the frequency of these links from one community to another is higher if they are linked as in the studied network.

One might be concerned about any analysis of the angle and distance of network clusters when visualizing them with the ForceAtlas layout in Gephi. It is certainly true that layout algorithms have stochastic components and are not designed to reliably test hypotheses. However, as we show in Sec. 4 of this SM, angles and lengths can be used as a crude guide for changes in the network—and it is clear even visually from Figs. 2(c)(d) that there has been a strengthening of bonds, however one chooses to measure this. In the future, again out of the scope of the present paper, we would operationalize this by developing an optimal network science measure of indirect bonding strength

between communities, then performing a statistical analysis versus a null model that shows that there was indeed a reliable structural change between years.

We only assess one type of Facebook intervention (i.e., banner message in selected pages)—however this is an important one, since it is the banner that users see at the top of their page. If the banner is absent on a neutral page that features Covid-19 guidance in Fig. 2(d), and yet is present on other pages, it might erroneously be inferred by users that the Covid-19 guidance on that neutral page is automatically best science guidance. We have no doubt that Facebook was carrying out other interventions, however these are the ones that were visible to us and likely also visible to the millions of users involved. We cannot test Facebook's other hidden policies without involving internal knowledge that is only available at Facebook as a private company. In the end, the fact remains that large numbers of users in neutral communities in Fig. 2(d) did not experience this Facebook intervention.

We now discuss possible critiques of the mathematical model. First, though it assumes—quite reasonably—that the various types of community may in principle influence each other (e.g., antis coupled with pros and neutrals) there is no need for the activity to be directly correlated. Nor are these couplings constant in time since they depend on the variable values (e.g., $G(t)$) as they change in time. Indeed, in the model plots shown later in Sec. 6, it can be seen that the activity (i.e., curve such as $G(t)$ for one subpopulation) can be very different from the activity for another curve such as $R(t)$ for another subpopulation, even though they are coupled. In short, it would be a mistake to think that the activities should mirror each other and hence that their curves are the same. The mathematical model does not assume this, and the output clearly shows that they are not. This is also shown in the results in Figs. 4 and 5(a)–(d) in the main paper. Second, it is only for the prediction of the future behavior in Fig. 5(a)–(d) that we assume that any of the subpopulations is in steady-state—and we do that just for the pros ($B(t)$) because the pros already seem to have maximized their activity in terms of promoting best-science guidance during Covid-19. Indeed, the media now regularly features stories suggesting that the health agencies have tried everything to try to convince those who have not yet been vaccinated. A recent example is the reported lack of correlation between proposed prizes and vaccination uptake. Third, we freely acknowledge that our model is of a type that is very common in systems biology, which has the advantage that it opens up future analysis to the tools and techniques developed there.

One might also wonder if there is in fact a handful of people/organizations who are responsible for most of the misinformation about vaccines and Covid-19 on Facebook. While there are certainly some that post more than others, we find that the ones who tend to pass on the most material are not necessarily these loudest ones. Elsewhere we have started to analyze this in a working paper online. Whether there is a form of master core or not, and which set of people are best removed from Facebook or not, is a fascinating topic but is one that does not fundamentally change the conclusions in our paper. It deserves a study in its own right.

Section 2: Color scheme for neutral nodes in plots. Breakdown of page admin country locations. The system before Covid, one year later, and two years later. Classification of neutral nodes

Color scheme for neutral node plots.

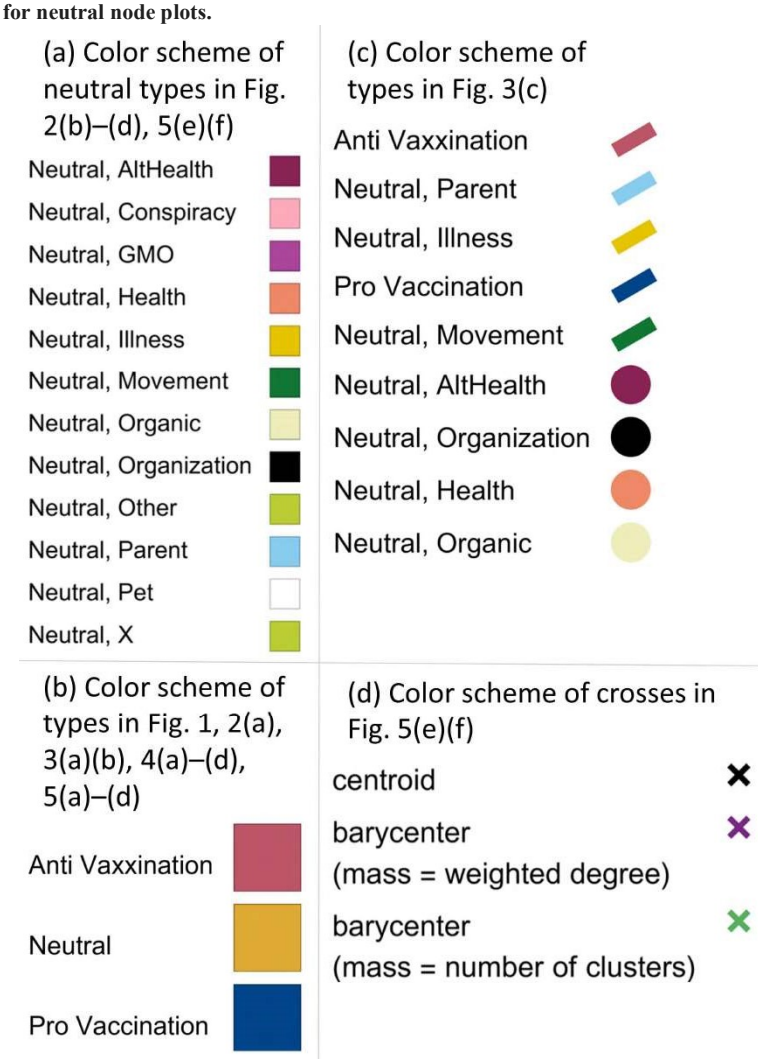

Figure S1: Color legends for figures in paper

Breakdown of page admin country locations

Additionally, it is possible to scrape a list of admin top countries (i.e., the country in which the largest number of page admins reside) for all the pages collected. This information was retrieved for 834 of the pages in the dataset, or 62.17% of all the pages, and the top 10 countries are as follows:

Table S1: Top 10 countries in which the largest number of a Facebook page's admins reside

Country	Number of pages where country is where the largest number of admins reside
United States	352
Australia	33
Canada	27
United Kingdom	20
Italy	9
France	8
New Zealand	6
Sweden	4
Germany	4
Belgium	4

Other countries such as: India, Ireland, Norway, Switzerland, Mexico, Uruguay, South Africa, the Netherlands, Slovenia, Thailand, Malaysia, Brazil, the United Arab Emirates, Croatia, Israel, Austria, Slovakia, Bulgaria, the Philippines, Singapore, Romania, Belize, Denmark, Serbia, Czechia, Portugal, Pakistan, Poland, and Costa Rica also appear in these results, though with diminishing quantities.

These results, however, can only give a partial or incomplete understanding of the true breakdown of admin locations, which a page's Transparency section can shed further light on. The Transparency section is part of Facebook's effort to provide more information regarding the page and the people who manage it as an effort to increase the accountability and transparency of pages and can be read bout in further detail on the Facebook website. Importantly, this information includes the primary country locations where the page is managed and provides a more complete breakdown of where page admins are located. Some examples of what the Transparency section contains are provided here, with identification information blocked out, and admin information boxed in red. The page in the upper left-hand corner, for example, would be considered exactly the same as the two, smaller pages on the bottom left and right, which only have admins from one country. Thus, the dataset contains pages where the admins might all come from the same country (e.g., the pages on the right-side of the image and the lower left-hand), but it also contains pages where there is a mix of countries, and that includes countries where the dominant or official language is not English, even if the page's contents are primarily or only in English.

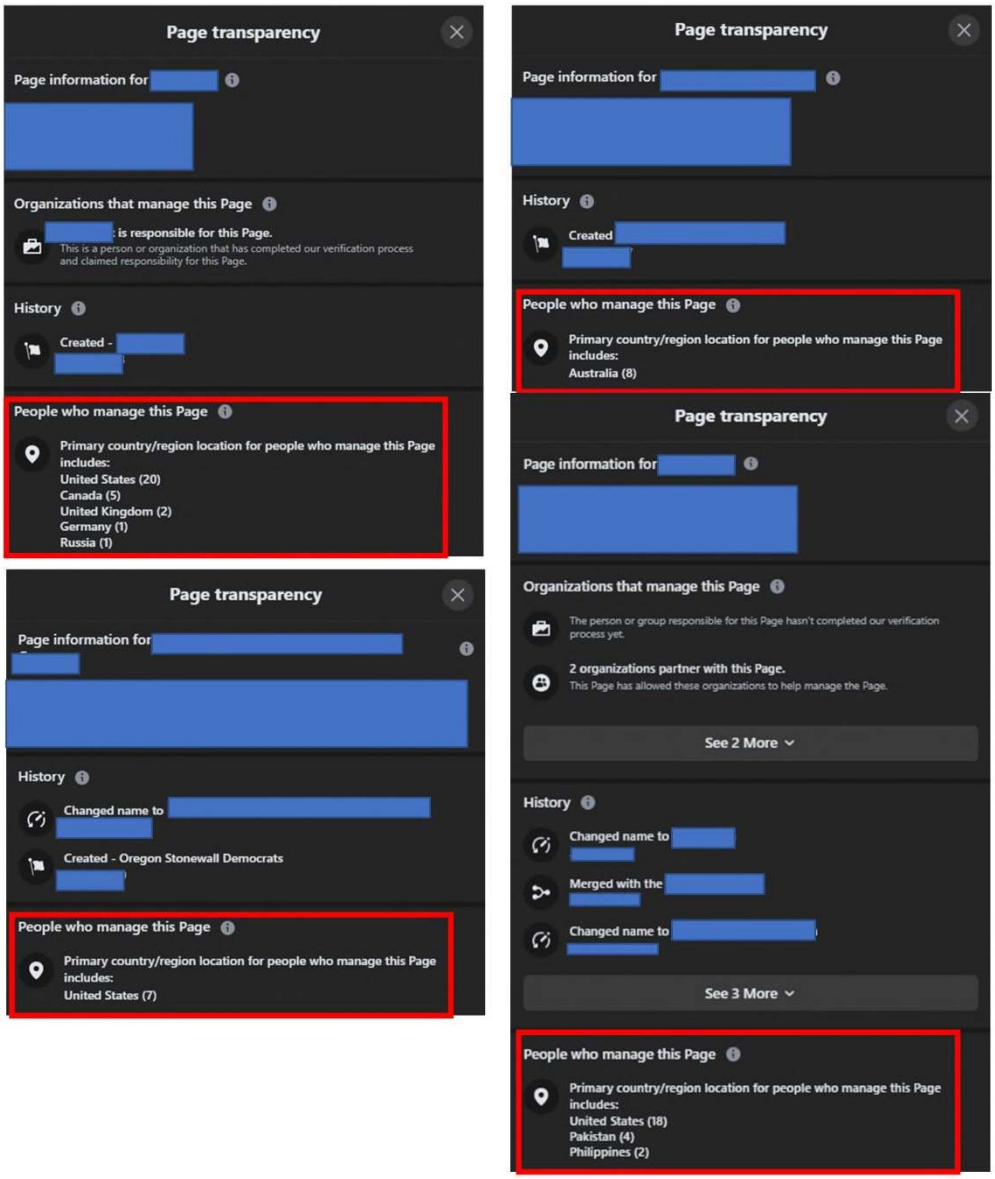

Figure S2: Examples of the Transparency section for several Pages in the dataset. The Pages on the left and bottom right would all be classified as having top number of moderators in the U.S., obscuring in data scraping the true moderation team composition

The system before Covid, one year later (13 months), and two years later (27 months)

Similar to Fig. 2 in the paper, this figure includes what the system looked like post-vaccine in February 2022. 200 nodes have been removed from this system (deleted) and 31 one pages have gone private. The rings provide visual aid in seeing the increase in node bonding, and how similar the system looks post-vaccine as it did pre-vaccine.

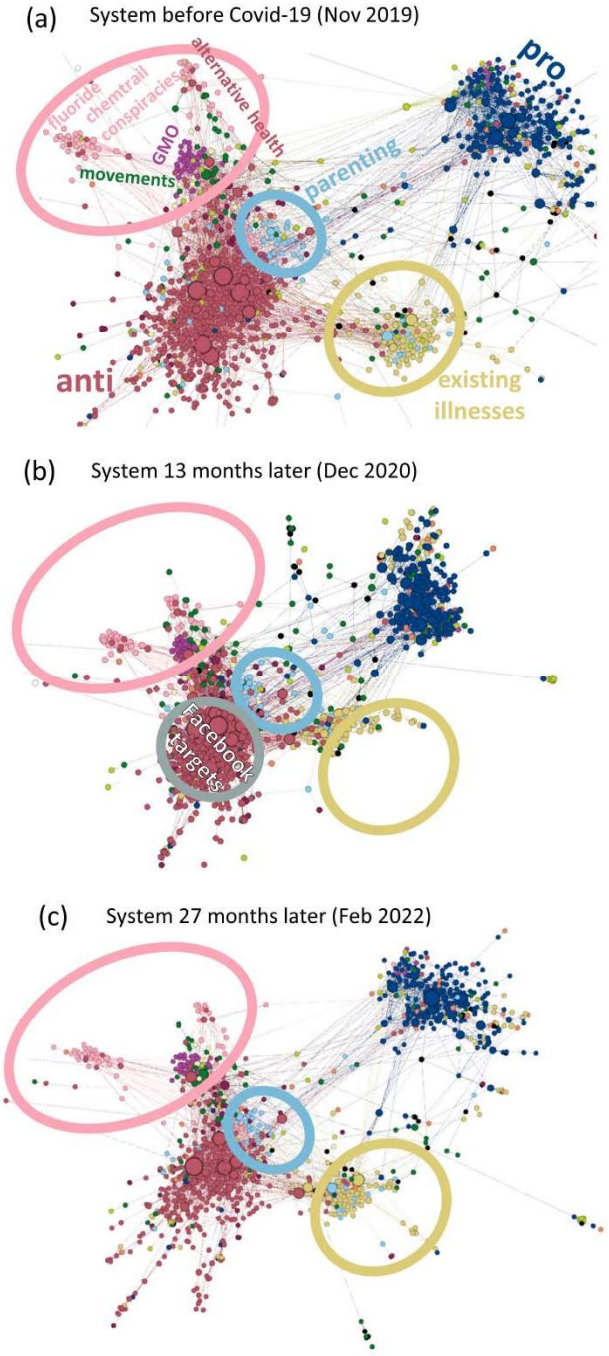

Figure S3: The system before Covid, one year later, and two years later

Classification of neutral nodes

The following pages provide an explanation of the 12 categories of neutral node used in the main paper.

Categories of neutrals

The term ‘cluster’ is a generic name for a cluster of individuals who have formed a community structure on a platform, e.g. Facebook page

AltHealth

An Alternative Health (AltHealth) cluster is a Facebook Page that promotes, discusses, or features content centered around alternative cures and practices, as opposed to traditional medical practical. This includes homeopathy, naturopathy, and spiritual healing. These clusters focus on anything from more common conditions such as headaches, indigestion, and general wellness, up to serious illnesses/conditions such as cancer and genetic disorders.

Some of these clusters promote and market “remedies” such as essential oils, herbal supplements, or unconventional medicines. They sometimes do this by addressing these products in their posts/pictures and/or sharing links to websites where they can be purchased. Other clusters might share anecdotal remedies: this includes recommending alternative diets and practices such as eating “raw”.

While not the largest category in terms of number of clusters, this is by far the largest in terms of total user size. This can be attributed to the large size of the top few clusters in this category. For example, the largest green cluster in the network, “Sun Gazing”, is an AltHealth cluster.

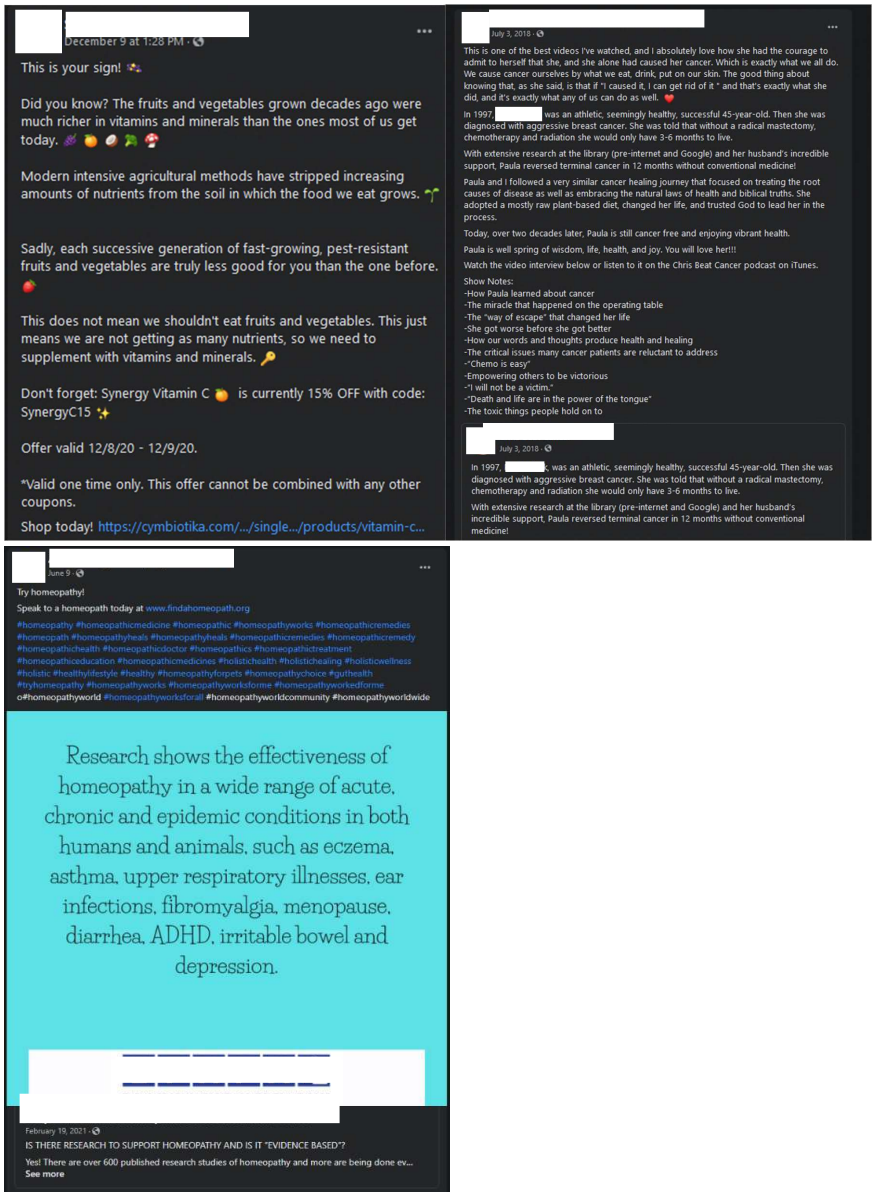

Figure S4: Examples of AltHealth Pages

Conspiracy

A Conspiracy cluster is a Facebook Page that promotes or discusses fringe or extreme theories based on unfounded claims that covert actors are responsible for events or circumstances.

This category is dominated by two main conspiracies: fluoride in water and chemtrails. The fluoride conspiracy theory claims governments use it to control the population, lower individuals' intelligence levels, affect fertility levels, and cause health problems. The nomenclature tends to include terms such as "fluoride free" or "clean water". Chemtrail conspiracy clusters share the idea that planes/aircraft under the direction of the government and shadowy organizations are spraying chemicals in the sky to affect the health and mental capacity of the population below. Although these are two different conspiracies, they share a similar theme of malign actors controlling the population through chemical poisoning.

While there are many Conspiracy clusters, they tend to be small in terms of users. The names of these clusters tend to include geographical references, such as "Fluoride Free Kansas" or "Chemtrails Global Skywatch of Oklahoma", which perhaps limit the potential user base. Nonetheless, these users tend to hold extreme views, which makes them susceptible to other conspiracies. These Pages might also serve as a starting point for radicalizing users toward more extreme views.

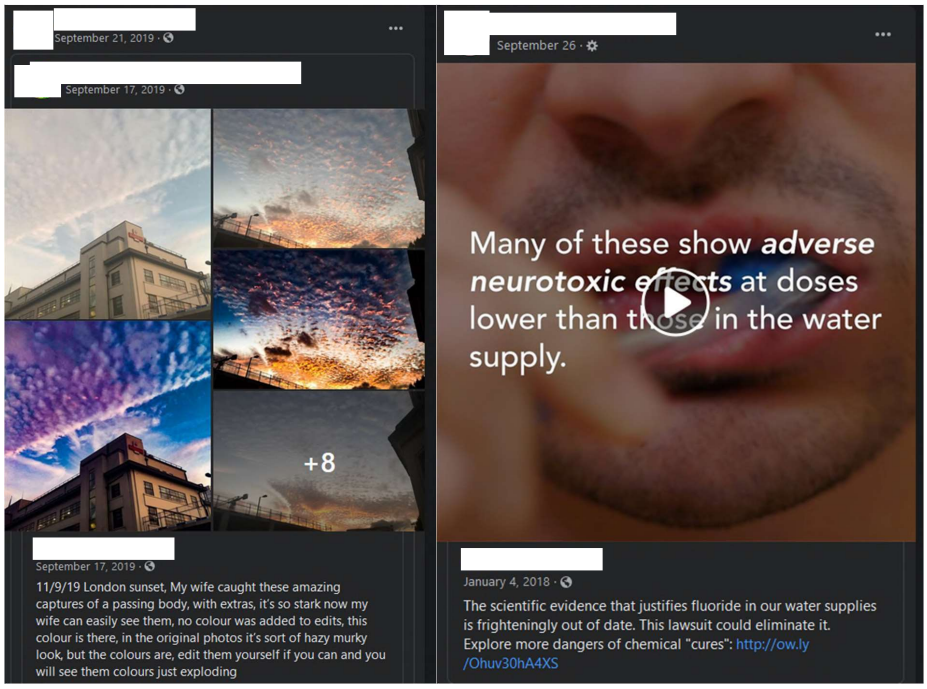

Figure S5: Examples of Conspiracy Pages

GMO

A GMO cluster is a Facebook Page that debates or is against the use of genetically modified organisms in food and medicines. Posts generally attempt to raise awareness of products that contain GMOs and argue that these cause harmful effects. These clusters often call for a boycott of certain brands or products and/or a requirement to label GMOs. Users often focus on Monsanto as the main antagonist; the company is often named in posts and the names of clusters.

There are relatively few GMO clusters, and they tend to be small in terms of users. The names of these clusters tend to include geographical references, such as “March Against Monsanto Fort Meyers” or “GMO Free Canada”, which perhaps limit the potential user base.

Figure S6: Examples of GMO Pages

Health

A Health cluster is a Facebook Page that discusses general health matters and medical institutions. In contrast to the AltHealth clusters, the Health clusters focus on traditional medicine, practitioners, and institutions. These include clinics, pharmacies, mental health services, medical staff, and health initiatives. The average Health cluster has relatively few users.

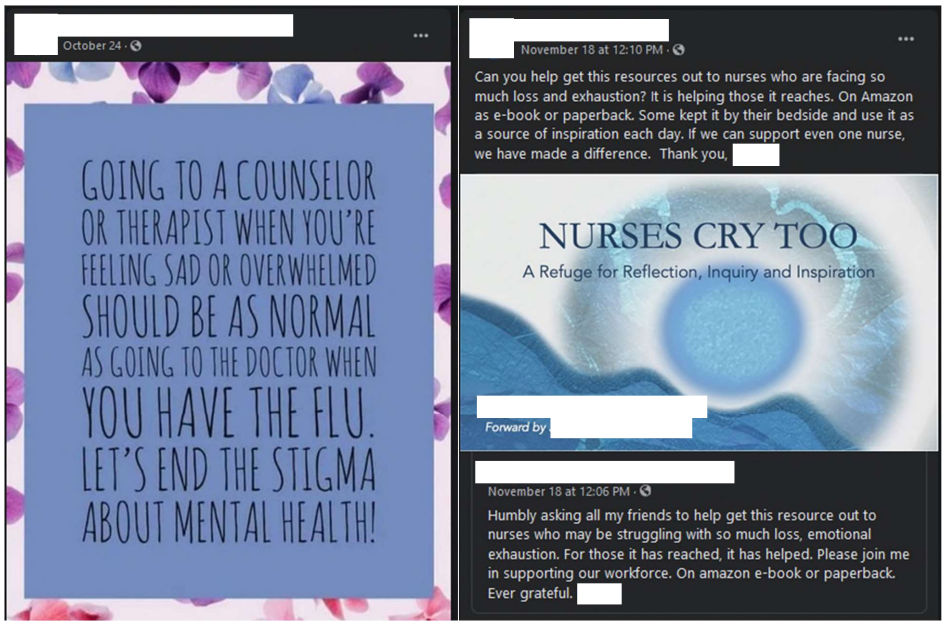

Figure S7: Examples of Health Pages

Illness

An Illness cluster is a Facebook Page that aims to raise awareness of, discuss, or serve as a support group for certain medical and mental illnesses/conditions. Fibromyalgia, cancer, and HIV/AIDS play a significant role in this category, but by far the most discussed condition is autism.

Due to the serious nature of these medical conditions, the role of these pages as support groups leads to an interesting dynamic. This is the most common category of green cluster in the network, perhaps because they tend to narrowly focus on a specific illness. Nonetheless, the total number of users in this category is toward the middle of the pack, which means that the average number of users per Illness cluster is relatively small. These are cohesive clusters, featuring discussions that are highly salient to a small number of users. They have relatively active discussions and are made up of users who have the condition or who have family and friends who are affected.

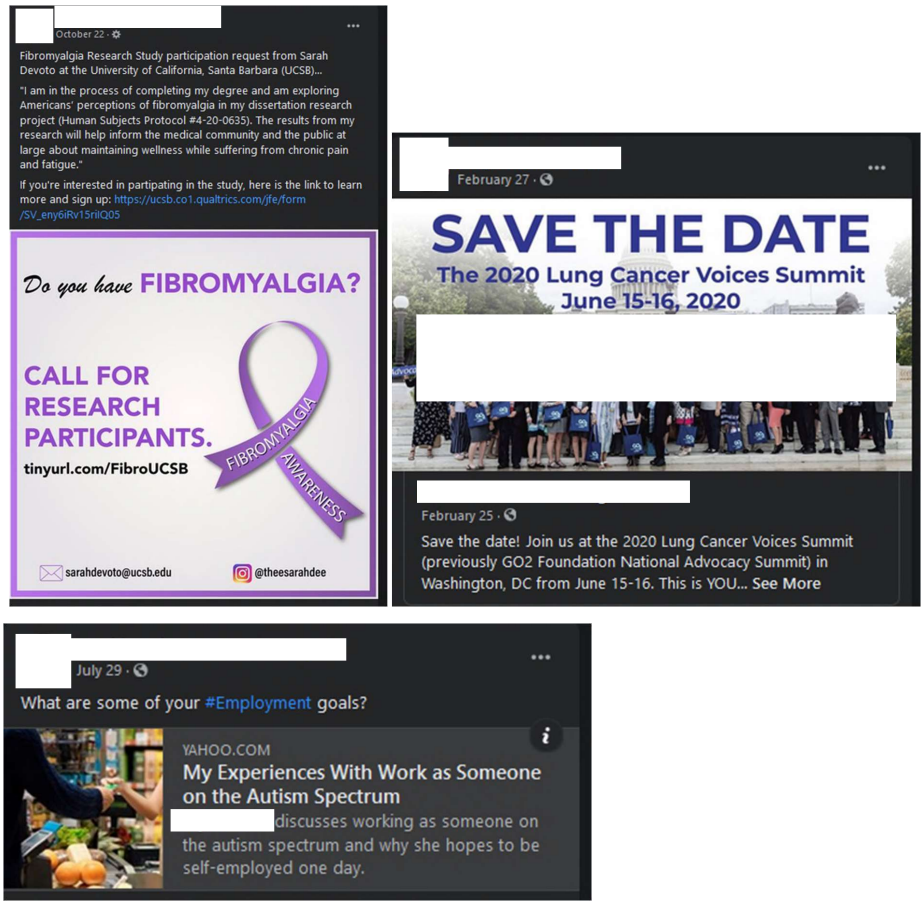

Figure S8: Examples of Illness Pages

Movement

A Movement cluster is a Facebook Page that advocates for a specific cause or political objective. The main topics include cannabis legalization, universal equality, domestic violence/human trafficking victim protection, the environment, and mental health awareness.

In terms of both total number of nodes and user size, Movement is the second largest. The wide variety of causes tends to increase the total user size, but not all clusters agree with each other. Some even hold opposing views on specific issues. Two of the three largest clusters advocate for the legalization of cannabis. Victim protection and awareness is also a common theme among many in this category, including victims of bullying, domestic violence, human trafficking, child abuse, and mental health issues.

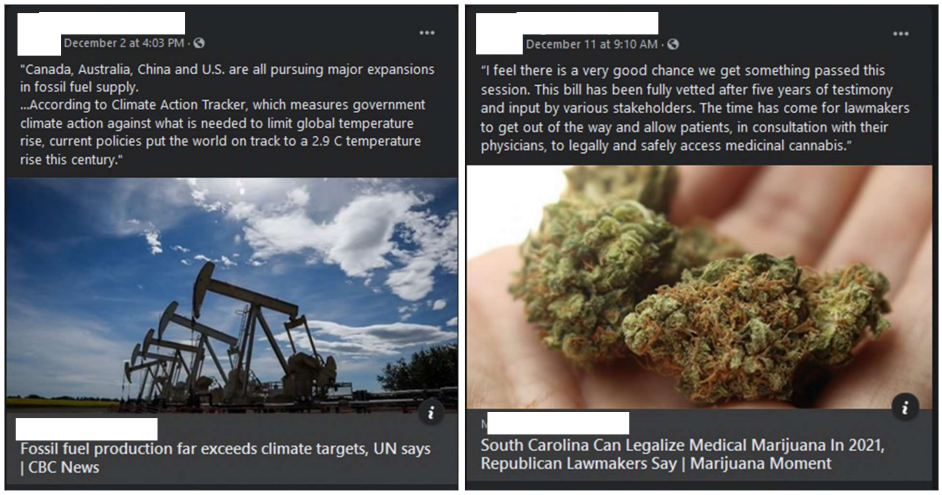

Figure S9: Examp es of Movement Pages

Organic

An Organic cluster is a Facebook Page that promotes an organic diet and lifestyle. Users tend to share recipes, diet plans, and information about particular organic ingredients. Growing your own food is an important activity that most of these pages advocate.

While this category is the smallest in terms of the number of clusters, in terms of total users it is twice as large as the GMO and Health categories, indicating that the Organic clusters are fairly popular.

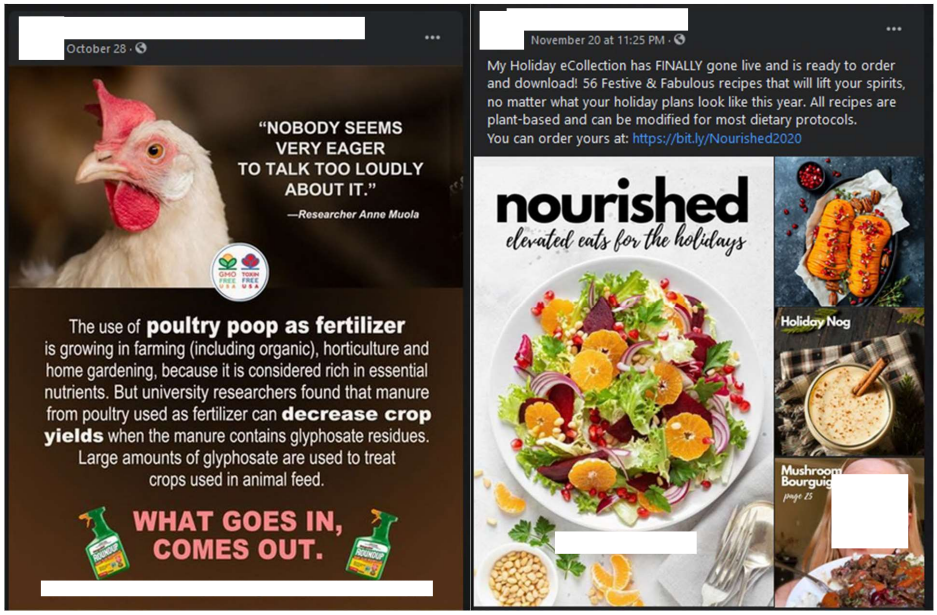

Figure S10: Examples of Organic Pages

Organization

An Organization cluster is a Facebook Page focused on a formal institution, including both government and non-governmental institutions. Clusters include the U.S. government's Stop Bullying program, foreign government institutions, assistance projects, TEDx Change, and county initiatives such as “Get Healthy Knox County” and "Pathways Connect Central". The largest Organization clusters are news organizations, such as the American Independent and LGBT News.

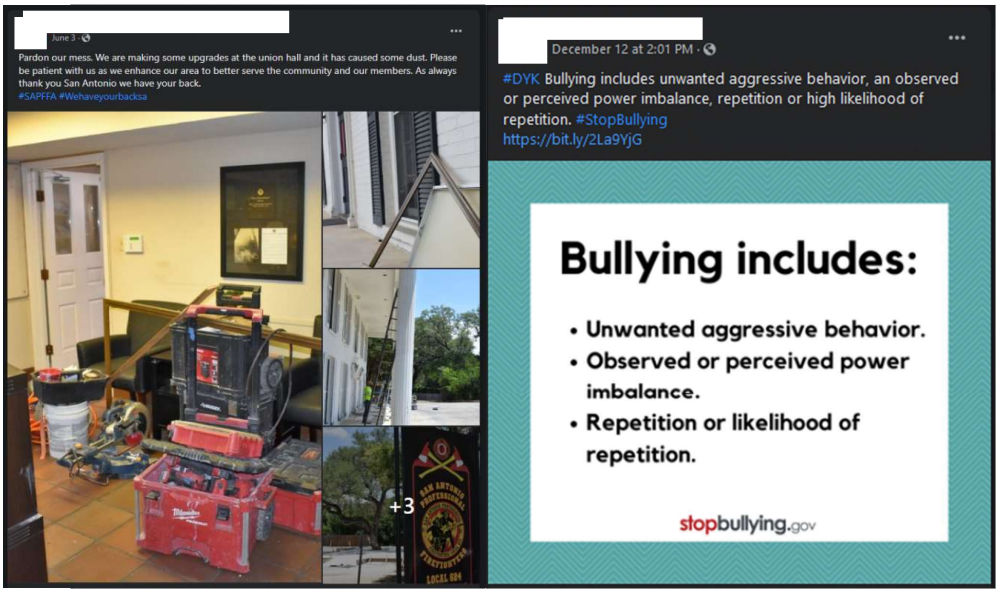

Figure S11: Examples of Organization Pages

Parent = “Parenting”

A Parent cluster is a Facebook Page that discusses or offers advice and support for parenthood. The most commonly discussed issues include parental rights, breastfeeding, homeschooling and birthing, and raising children with special needs. Although there are clusters focused on issues particular to the role of fatherhood, most of these clusters seem to be focused on motherhood and frequented mostly by women.

There are relatively many Parent clusters, and they tend to be relatively large, making this one of the most important substantive categories in the network.

Figure S12: Examples of Parenting Pages

Pet

A Pet cluster is a Facebook Page that is centered around pets (typically dogs and/or cats). The category is almost exclusively made up by shelters, pet rescue/adoption, and lost-and-found organizations. A substantial share of these clusters are based in Texas.

There are relatively many Pet clusters, and they tend to be relatively large, making this one of the most important substantive categories in the network.

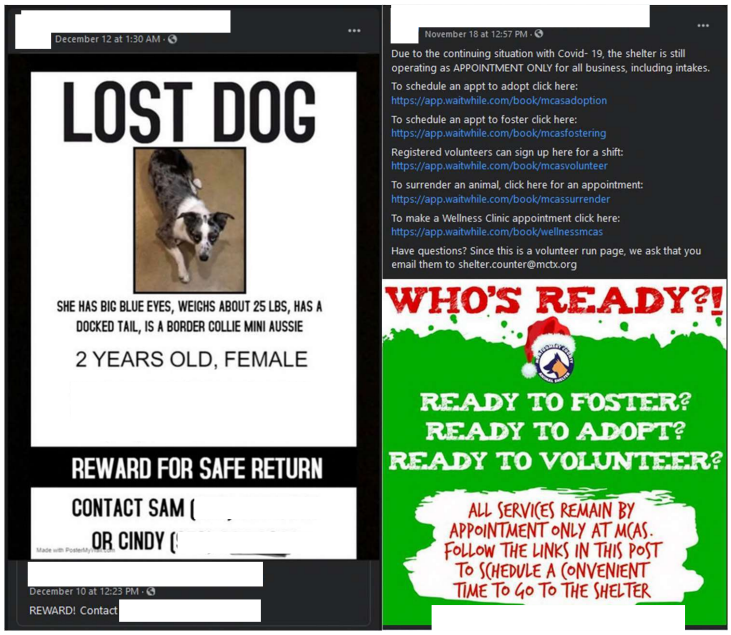

Figure S13: Examples of Pet Pages

Other

Some clusters did not fit in the categories we created, and thus we classify them as Other. These include clusters focused on specific churches, spiritual issues, farming, meme sharing, and community-building. The largest of these is an “earth lover” page, which shares photos of beautiful landscapes and natural features. There are also two large spiritualism clusters that tend to share inspirational sayings and general spiritual content.

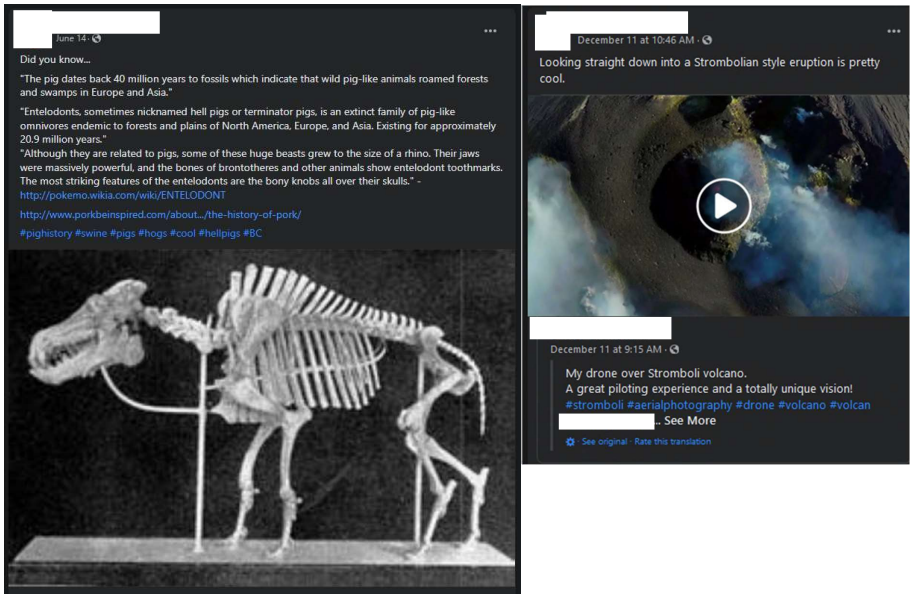

Figure S14: Examples of Other Pages

X are clusters that did not have an interest focus, or it was too ambiguous. They are a minor, unimportant category in our study

Section 3: Example of Facebook banners promoting best-science Covid-19 guidance. Positions in network of the nodes that receive Facebook banners promoting best-science Covid-19 guidance

Figure S15: Example of Facebook banner promoting best-science Covid guidance

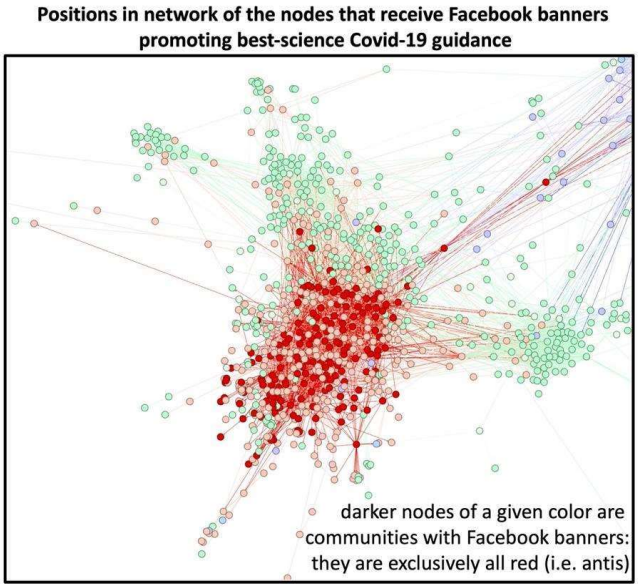

Figure S16: Category and location of nodes (i.e., Pages) that received Facebook banner promoting best-science Covid guidance

Section 4: ForceAtlas2 layout and analysis showing dependence of layout on strength of bonding

The ForceAtlas2 layout of Gephi simulates a physical system in which nodes repel each other while links act as springs. It is color-agnostic, that is, the color segregation and hence segregation of community types in Fig. 2 and Fig. 5(e)(f) emerges spontaneously and is not in-built. The network nodes and links were built without regard for this color segregation. Nodes that appear closer to each other have local environments that are more highly interconnected, whereas nodes that are far apart do not. Each node (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. On could measure bond lengths as the distance between the centers of the respective community types and the bond angle as the angle in between—which can be done in practice by arranging the networks, as we do in Figs. 2(c)(d), for the two times on the same scale, so that the distance between the centers of the red and blue communities remained the same. Changes can then be measured relative to this, e.g., manually or using the freely available software Aequo.

We now prove that changes in the positions, and in particular visual strengthening of bonding, is associated with changes in the effective link strengths—and hence that sets of communities (nodes) appearing closer together are more interconnected and hence likely have more shared content and users. We use the simple 3-body setup and notation shown in the diagram below:

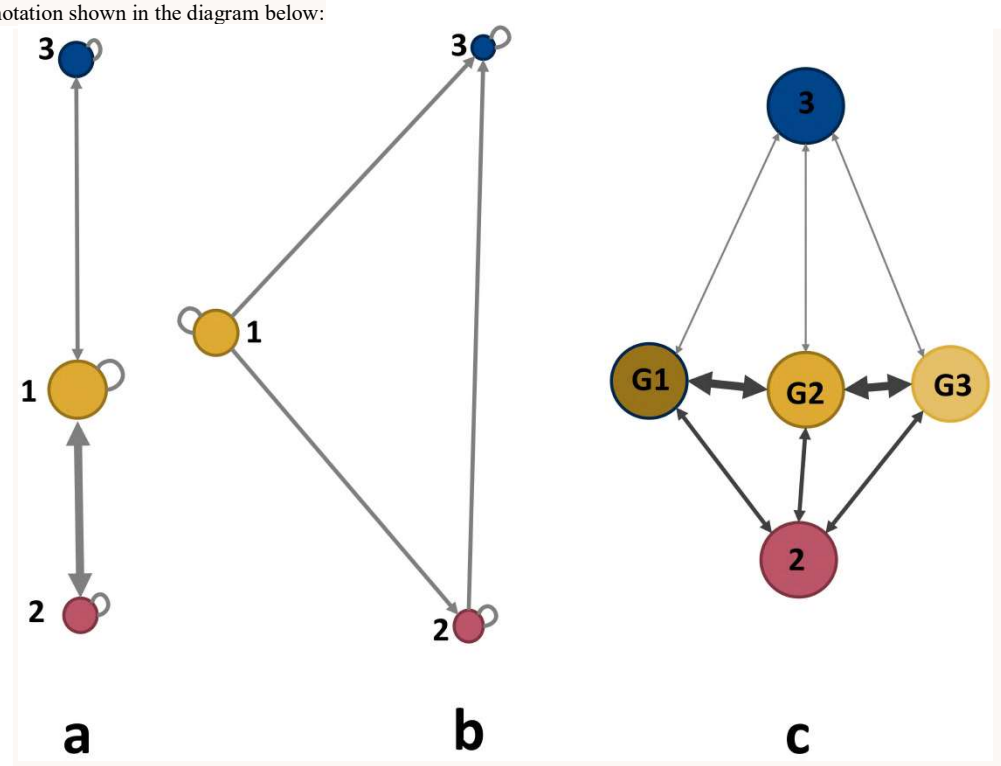

Figure S17: 3-body diagram of systems where nodes represent clusters of communities

Each ball (which can be thought of as a super-node) can represent a collection of reasonably tightly bound balls (i.e., nodes as in the main paper) of the same color, and hence represents clusters of communities of the same type as observed in Figs. 2–5 in the main paper. In this way, the 3-body analysis that we give below, can represent the much

more complex multi-node system in Figs. 2–5, simply by renormalizing what a ball is and hence means. For example, the parent communities roughly correspond to position 1, as does the center of mass of the gold communities from our 2020 vaccine paper, while 2 represents the anti (i.e., red) communities and 3 represents the pro (i.e., blue) communities. By inputting relative values for the links (equivalently for ForceAtlas2 algorithm, this can be the number of links) between 1, 2 and 3, and letting the ForceAtlas2 algorithm relax the network through energy minimization as in Fig. 2, we explore below what relative weights are needed in the links in order to obtain different arrangements. Our findings are as follows. We consider for simplicity unidirectional links, i.e., 1-2 is a link from 1 to 2, but the results are the same if we use bidirectional links:

- with link weight (or equivalently, the relative number of links) from 1-2 taken as being similar to 1-3, and both much larger than 2-3 (e.g., 10,10,1) then the layout is similar to case **a** above with bond length 1-3 similar to 1-2
- with link weight (or equivalently, the relative number of links) from 1-2 taken as being similar to 1-3, and both somewhat larger than 2-3 (e.g., 3,3,1) then the layout is an isosceles triangle as in case **b** with bond length 1-3 similar to 1-2 and bond length 2-3 larger
- with all 3 link weights (or equivalently, the relative number of links) taken as being similar, i.e., 1-2 similar to 1-3 and 2-3, then the layout is an equilateral triangle with all 3 bond lengths similar
- with link weight (or equivalently, the relative number of links) from 1-2 the largest, and 1-3 similar to 2-3 (e.g., 3,1,1) then the layout is an isosceles triangle but now with bond length 1-3 similar to 2-3 and both larger than bond length 1-2

This can of course be extended to include different categories of neutrals separately, as shown in case **c**.

Based on this, we can now understand quantitatively why the networks in Figs. 2–5 of the main paper have clusters of communities in the particular layout positions that they do. To demonstrate this, we take the total number of links from the data between red communities and blue communities, between gold communities and blue communities, and between gold communities and red communities, and convert them to a relative weighting which is given by the ratio 1: 6.7 : 14.1. Putting this into the ForceAtlas2 algorithm produces a layout as in case **a** with bond length 1-2 = 3.8 and bond length 1-3 = 6.0 which compares favorably to the empirical results for the layout and distances between the center of masses of the red communities, the gold communities, and the blue communities, which also approximates case **a** and has bond length 1-2 (i.e., Gold-Red) = 3.4 and bond length 1-3 (Gold-Blue) = 7.3. Obviously the many-body nature of the actual network, with many nodes and complex link arrangements, introduces the relatively small differences. This analysis also allows us to predict the tipping point where the entire ecology will flip into an arrangement like an equilateral triangle. Though obviously crude, it predicts that this will occur if the total number of links between gold communities and red communities decreases by 53% or the total number of links between gold communities and blue communities increases by 112%. It also explains why the conspiracy communities sit to one side, based on the same approach of summing the relevant number of links. We note that in all cases, the number of self-links, i.e., gold community to another gold community etc., are so large, that gold communities act as an approximately single entity with its own center of mass (like molecules in a gold ball), and the same for the red communities separately, and for the blue communities separately.

Statistics of the fit in Fig. 4(c)(d). *K*-fold cross-validation with the model. Null model analysis throughout the paper

Statistics of the fit in Fig. 4(c)(d).

Good fits can be obtained to the empirical curves in Figs. 4(a)(b) using the simplest possible implementation of our mathematical model (shown in Fig. 1 and see derivation in Sec. 6, this SM) even though it sets all the gels as having the same onset time of zero and includes no decay terms (i.e., no loss of interest and hence no fragmentation of the gels). The equations in Fig. 1 represent an Occam's razor minimal model, yet the fit is still good as shown by the visual fit in the SM and the statistics in the table below for the receivers (Fig. 4(a)(c)):

Table S2: Goodness-of-fit metrics for the model of the number of clusters listening with no onset or delay terms

Goodness-of-fit metric	Blue (pro) theoretical model curve	Red (anti) theoretical model curve	Gold (neutral) theoretical model curve
R^2	0.944711	0.956236	0.866737
Adjusted R^2	0.937577	0.948942	0.839165
AIC	-103.313	-70.4654	-65.6173
BIC	-95.536	-61.1333	-54.7298

The fits are slightly less good but qualitatively similar to the fit statistics for the more sophisticated version in which onsets do not occur at the same time, as indeed predicted by the full mathematical theory (Sec. 6, SM). The resulting curves are the theoretical ones in Figs. 4(c)(d). One could go further and get an even better fit, by adding the ability for fragmentation in order to mimic users leaving or losing interest in a page or set of pages. We stress that as shown explicitly in Sec. 6 of this SM, none of these extra features is essential to reproduce the features of the empirical graphs in Fig. 4. They just make the fits even closer as shown in the table below. Instead, the only crucial ingredient in our mathematical theory (Sec. 6) is that the equations contain the core feature that the dominant online couplings in the health debate involving $R(t)$ (antis), $B(t)$ (pros) and $G(t)$ (neutrals), are (i) $B(t)$ couples to neither of the others, just to itself and hence includes a possible decay, (ii) $R(t)$ couples with $B(t)$ and itself hence the decay, and (iii) $G(t)$ couples with both $R(t)$ and $B(t)$ and itself hence the decay. Neither the specific forms nor the inclusion of different onset times and decays are important in terms of capturing the general shapes of the exposure dynamics curves, and hence they are not important in capturing the core gel-formation-and-interaction mechanism driving the exposure dynamics: they just serve to improve further the fit.

To prepare the activity data for fitting, the dates on which counts were obtained were converted into an integer day count starting from Jan. 1st. Since about 30 weeks of data was inspected, the time data was rescaled from 0 to half that, or 15. The measure of the activity was also renormalized, so that the maximum activity value was now equal to 1. The coupled differential equations from Section 6 of this SM were then implemented in Mathematica using NDSolve, and the initial conditions were set around the activity value at time $t = t_0$, so that the model curves would start off at around the same initial activity values as the empirical data. With the simplistic assumption that Blue is independent of both Red and Gold, the parameter values for the B-R and B-G interaction terms can be set to 0, which allows us to fit the Blue curve first with four parameters using Mathematica's NonlinearModelFit. Once the Blue activity was fit, implementing the assumption that Red is impacted by only Blue, we can set the parameter value for the R-G interaction term as 0 as well, and then proceed to fit the Red activity with 5 parameters. Lastly, on the assumption that Gold is impacted by Blue and Red, we fit the Gold activity with six parameters. Initial parameter suggestions were fed into the function, determined from inspection of the data, to prevent the risk of falling into some local minimum before finding the global optimum.

For measures of the goodness-of-fit, we turn to the coefficient of determination R^2, the R^2 adjusted for the number of model parameters, the Akaike Information Criterion, and the Bayesian Information Criterion. The values of these properties for each curve are denoted in the table below:

Table S3: Goodness-of-fit metrics for the model of the number of clusters listening (Fig. 4(c)), i.e., receivers

Goodness-of-fit metric	Blue (pro) theoretical model curve	Red (anti) theoretical model curve	Gold (neutral) theoretical model curve
R^2	0.994664	0.997709	0.986172
Adjusted R^2	0.993976	0.997413	0.983311
AIC	-144.524	-124.773	-57.3219
BIC	-136.747	-116.996	-46.4345

Table S4: Goodness-of-fit metrics for the model of the number of clusters speaking (Fig. 4(d)), i.e., emitters

Goodness-of-fit metric	Blue (pro) theoretical model curve	Red (anti) theoretical model curve	Gold (neutral) theoretical model curve
R^2	0.973629	0.976358	0.992076
Adjusted R^2	0.970227	0.972418	0.990436
AIC	-129.224	-92.0186	-164.4
BIC	-121.447	-82.6865	-153.513

***K*-fold cross-validation with the model**

In *k*-fold cross-validation, the data is randomly partitioned into *k* equally sized subsamples, in this case 3, making each subsection 12 data points. One of the *k*-subsamples is retained as the validation dataset for testing the resulting model while the remaining *k*-1 subsamples are used for training the model. The process is repeated *k* times so that each of the subsamples is used as the validation dataset once. The *k* results are then averaged to produce the final parameter estimations.

However, there's a risk one runs into with smaller datasets, and that is that using hold-out sets for validation and testing are most practical with large datasets. While many Covid-related posts were scrapped over the 8-month period, the data was collected in weekly increments, which when processed to analyze the emitter-receiver dynamics of the system results in 35 datapoints per curve, i.e., one point per week of data collection. The outcome of running *k*-fold cross validation on a smaller dataset is that the validation score is generally poorer, and there is the risk of over-fitting because the model may get less information to go off in the training dataset. Thus, the model may be inclined to generate inferences specific to the training set and be dependent on which points are sampled for the folds.

We include here tables that show the results of the estimated parameters when employing 3-fold cross-validation, compared against using the full dataset, as well as what the goodness-of-fit metrics look between the two methods.

Table S5: Comparison of model parameters when fitting on full dataset versus employing 3-fold cross-validation

		R_0	B_0	G_0	a_R	a_B	a_G	h_R	h_B	h_G	d_R	d_B	d_G	r_B	g_B	g_R
Receiver curves	**Full dataset**	.7	.9	1.3	.9	.4	-.2	1.3	2.1	1.7	0	.4	.2	-.4	-.9	1.6

	Cross-validation	2.2	.7	3.2	.3	.5	-.3	1.3	2.1	2.3	.6	.3	1.0	-.2	-3.9	3.4
	Absolute difference	1.5	.2	1.9	.6	.1	.1	0	0	.6	.6	.1	.8	.2	3	1.8
Emitter curves	**Full dataset**	1.0	.7	.5	.2	.2	4.5	2.1	2.4	5.3	-.1	.4	1.0	.7	.7	-5.4
	Cross-validation	1.0	.8	1.0	.3	.2	1.8	2.1	2.5	5.1	-.1	.3	4.3	1.2	-3.8	-1.6
	Absolute difference	0	.1	.5	.1	0	2.7	0	0.1	.2	0	.1	3.3	.5	4.5	3.8

Table S6: Goodness-of fit metrics when fitting on full dataset versus employing 3-fold cross-validation

			R^2	Adjusted R^2	AIC	BIC
Receiver curves	**Full dataset**	**Red curve**	0.99	0.99	-124.77	-116.99
		Blue curve	0.99	0.99	-144.52	-136.75
		Gold curve	0.99	0.98	-57.32	-46.43
	Cross-validation	**Red curve**	0.97	0.96	-31.08	25.88
		Blue curve	0.95	0.93	-35.83	23.61
		Gold curve	0.89	0.83	-5.47	21.40
Emitter curves	**Full dataset**	**Red curve**	0.98	0.97	-92.02	-82.69
		Blue curve	0.97	0.97	-129.22	-121.45
		Gold curve	0.99	0.99	-164.4	-153.51
	Cross-validation	**Red curve**	0.94	0.92	-29.12	22.38
		Blue curve	0.84	0.80	-33.58	20.21
		Gold curve	0.97	0.96	-46.98	28.32

In general, the models produced by 3-fold cross-validation had poorer metric values, as expected, and while sometimes the estimated parameters were very similar to those from fitting the entire dataset, at times they might diverge by up to 4.5 points. By visual inspection, however, the different parameter estimations appear exceptionally similar:

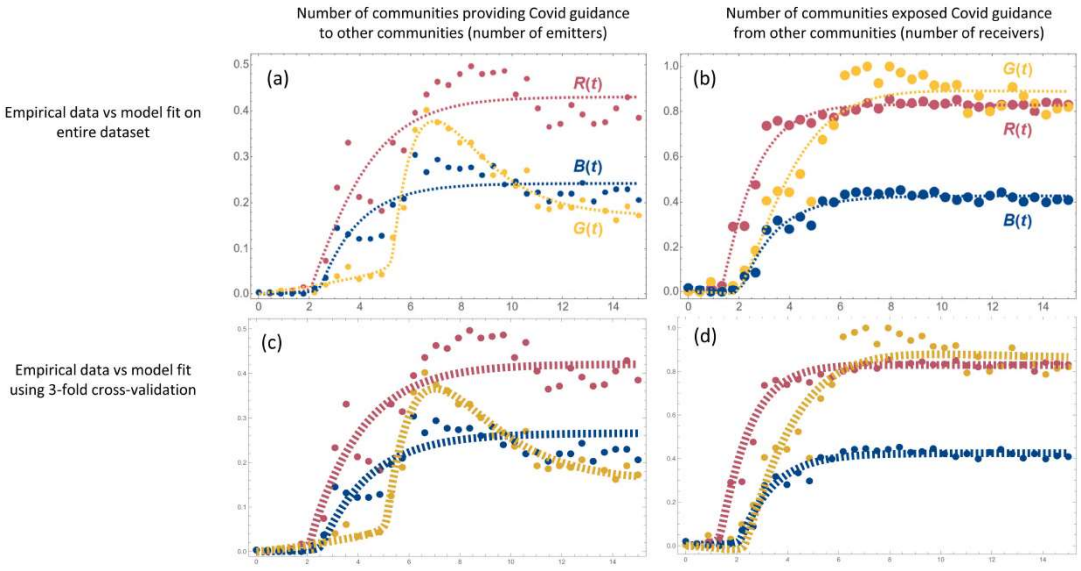

Figure S18: Difference in parameter estimates from using the entire dataset versus 3-fold cross-validation

Here, the empirical data is represented by circular markers, while the model is represented by dashed lines. While some of the parameter estimates may vary by a large degree between the two methods, comparing the top row, which is parameters were estimated using the entire dataset, to the bottom row, which is where parameters were averaged over 3-folds, there don't appear to be major deviations. This implies that there are a large range of values which the parameters can occupy which will produce similar results.

Null model analysis throughout the paper

Below we show plots of the null model results obtained from randomizing nodes/links labels. These provide evidence against the null hypothesis that the results in the main paper could have been achieved by chance:

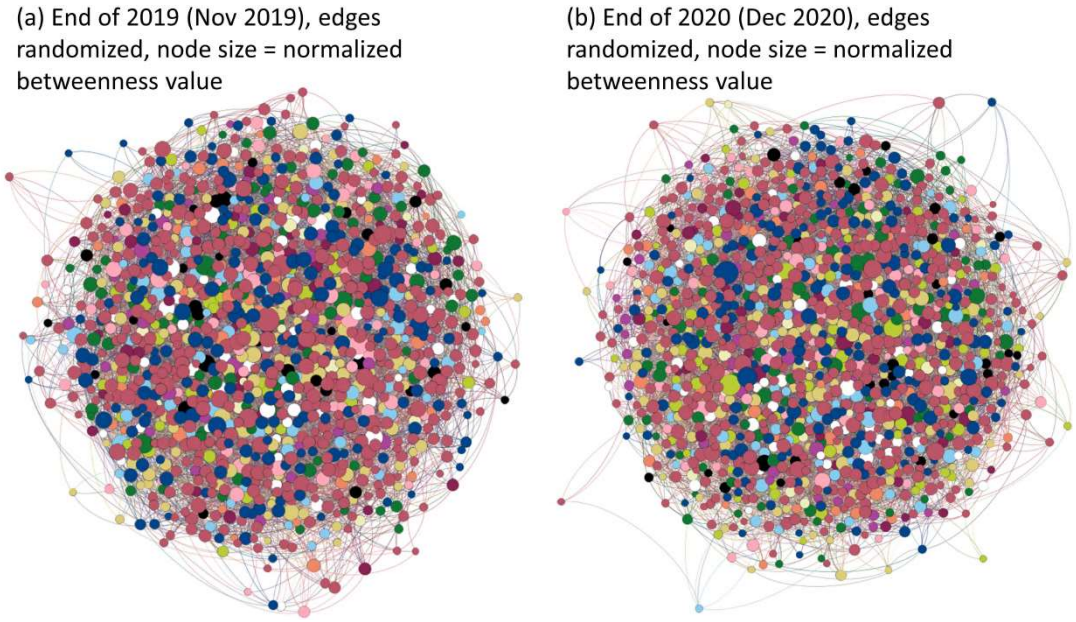

Figure S19: Null model versions of the networks in Figs. 2(c)(d)

To generate these networks, we counted the number of emitters and receivers at time *t*, and randomly sampled the list of nodes present at time *t* (to prevent introducing any nodes that were not present/active) to get a new list of emitters and receivers. This allows us to maintain the total number of nodes and links present to be identical with the actual data. In this way we are able to generate a null model where the links have been randomized, and can visually compare the (lack of) structure seen here against the structures present in the actual data.

The null model that we demonstrate in Fig. 4(a) and (b) was also built for Fig. 3(c). We investigated to check to see if the results were a by-product of the proportion of pages in each vaccination category, and to that end we randomized which pages were generating COVID-related posts at every time step *t*. We left the pre-existing links (i.e., when some Page *i* fans another Page *j*) untouched, because randomizing these would imply that every page is equally likely to have fanned another page, which isn't the case with the actual data. We were thus investigating if our results could be a product of the proportion of Facebook pages of different vaccination types generating COVID-related material, and the influence the specific pages generating these posts had on the results. With that goal in mind, in order to randomize the emitters at a time *t*, we tallied the number of emitters of each type (anti, pro, the neutral subcategories) and then randomly sampled the full node list that many times. We repeated this process 1000 times, which generated 1000 new lists of emitters, after which we can employ the same process as with the

actual data to find which pages fanned these new emitters. Then, at each time *t*, we grouped the Facebook pages by vaccination category to find the new total number of emitters and receivers, which allowed us to calculate the mean and one standard deviation across the 1000 iterations.

Both plots track pages from which Parent pages received COVID-related posts – that is to say, we investigated when Parenting pages were receivers, and who was emitting to this group. In the top image, generated in R, the actual data values are represented by the lines with circular point markers, whereas the bands with triangular point markers represent the mean value and one sigma deviation of the 1000 iterations of randomized data. We can see, for example, that Parenting pages appearing to be discussing with each other on a scale you would expect but are disproportionately interacting with anti Facebook pages on a level far that far exceeds that predicted by the null model. Interactions with Illness-related pages also do not conform with the null model predictions. The bottom image was generated using Mathematica, and the actual data is only presented with circular point markers. The randomized data is represented here with three joined lines and a shaded band, where the top and bottom lines represent one sigma and the center line represents the average, to aid in visualizing the actual data points against the randomized values.

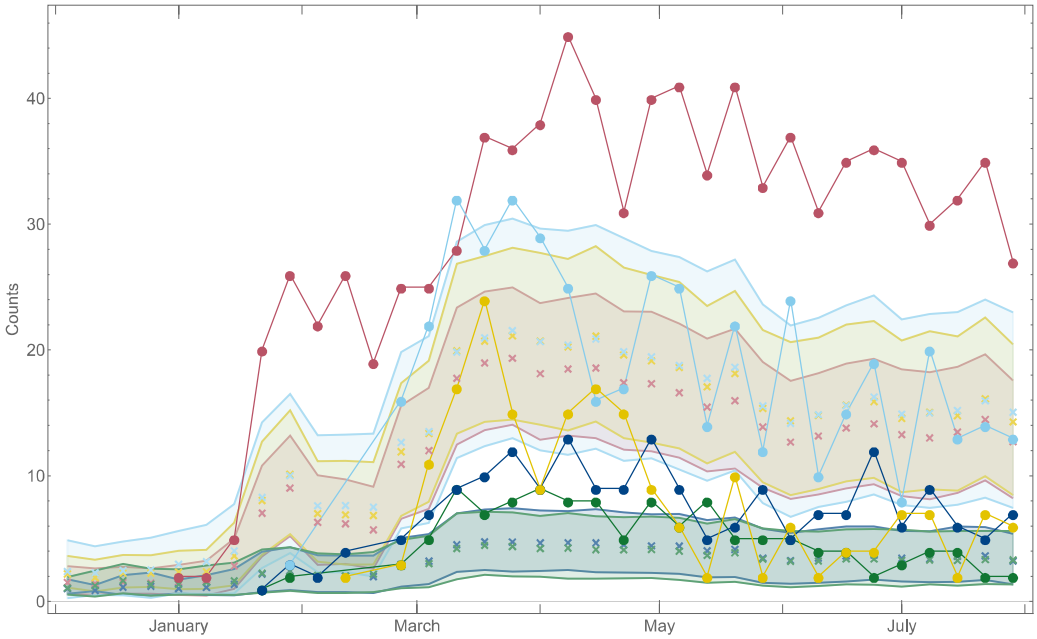

Figure S20: Null model version of Fig. 3(c)

and here it is shown using different symbols to aid visual analysis:

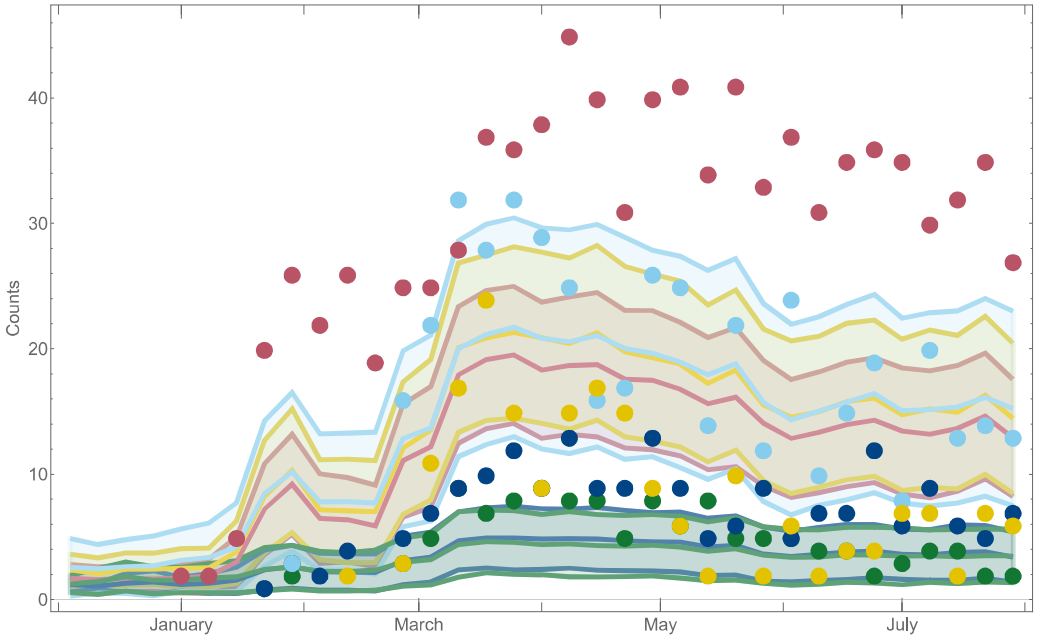

Figure S21: Null model version of Fig. 3(c), alternate symbols

Random deletion of Covid-guidance links and model robustness

We acknowledge that our dataset is ultimately an imperfect sample of some larger 'correct' network, hence we performed simulations to investigate the effects of noise in the data, by checking the effects of if links had been missed during the data collection process, or simply did not exist. To that end, we randomly selected 1% to 15% (in 1% increments) of the Covid-guidance links from the entire network to be deleted to explore what system-wide effects this would have, particularly on the emitter and receiver dynamics. The biggest impact is on the magnitude (which one would naturally expect to change), and one can see the general robustness of the curve shapes. There do however appear to be increased fluctuations from March onwards, where the effect of link deletions seems to be slightly more impactful, but the general shape is still discernable.

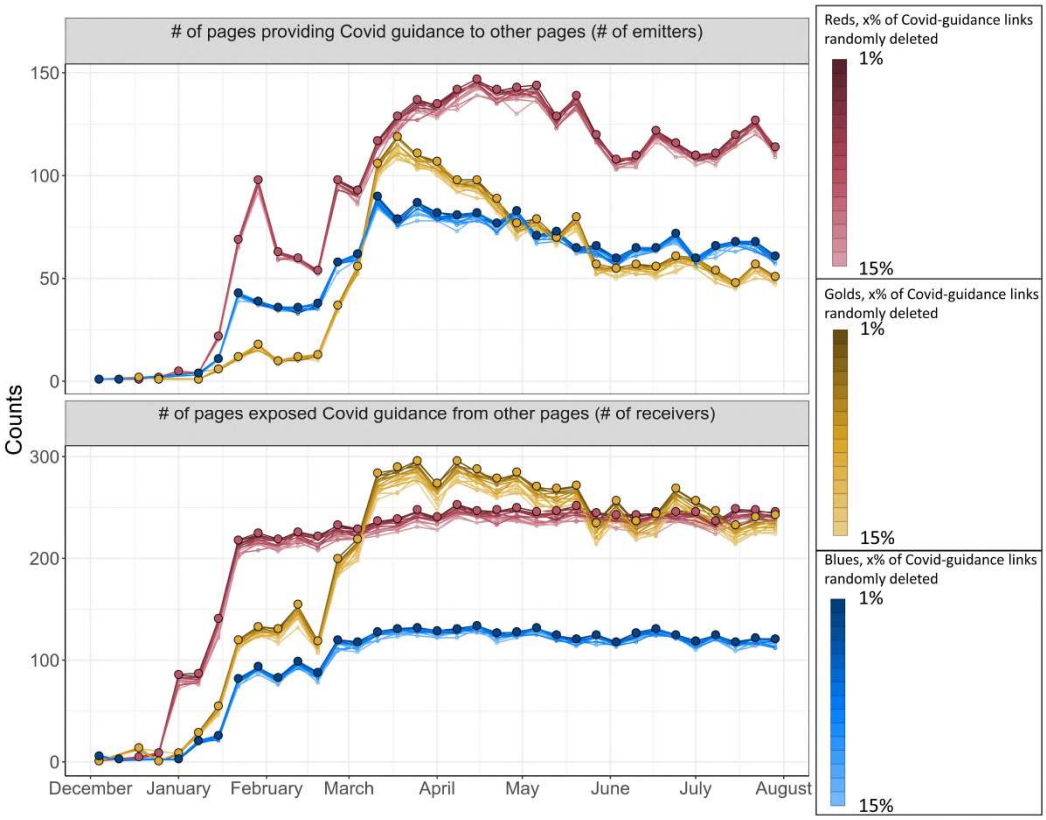

Figure S22: Effects of randomly deleting 1% to 15% of Covid guidance links from the system on the emitter-receiver dynamics

Further, we can investigate the robustness of our model to variations in parameter values. We saw some of this in a previous section, with how in the 3-fold cross-validation parameters varied from the full dataset fit by up to 4.5 points, and the fits look visually the same. We can investigate the parameter space and see what effects uniform changes in values have on the model:

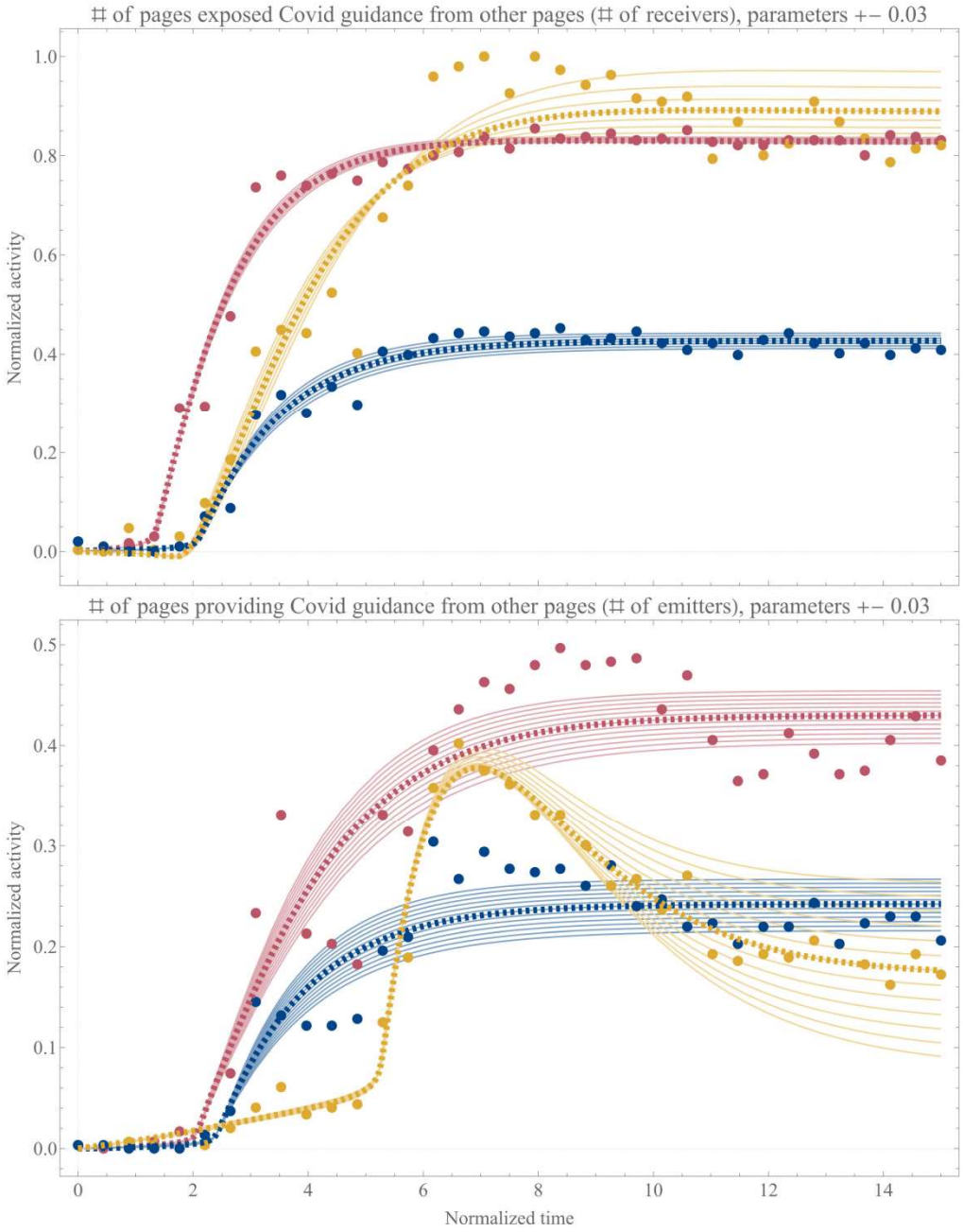

Figure S23: Effects of uniform changes in parameters values

By changing the parameter values in increments, we can see that the overall shape of the curves are preserved, and in some cases the changes have a very minimal effect (primarily in the region before normalized time 4, which is

before March). When investigating the effects of k-fold cross-validation, we saw that large fluctuations in parameter values, when combined, could produce very similar shapes, and here we further see that the general shape is also robust against uniform changes in parameter values.

Section 6: Derivation of our mathematical model and its use in Fig. 4, and predictions from our mathematical model shown in Fig. 5(a)-(d)

In the following pages, we lay out the full mathematical derivation that leads to the model results presented in the main paper, in conjunction with the discussion in Section 5 of this SM.

We are very grateful to Pedro Manrique for collaboration in the following formalism discussion, which is drawn and developed from joint published works and an ongoing collaboration together. A portion of this is also from joint research carried out with Om Kant as part of his PhD thesis under the supervision of one of us.

The key mechanism in the real online world that our mathematical model captures, and then carries in an approximate way across scales – using an approximate or 'dirty' renormalization using physics terminology – is as follows:

Individuals aggregate (i.e. they 'gel') online into individual communities (e.g. page such as one of the antis, pros or neutrals); and these communities of a given type (i.e. anti, pro or neutral, or category of neutral depending on the level of renormalization) can then be aggregated into the respective subpopulations of anti, pro or neutral given by $R(t)$, $B(t)$ and $G(t)$ respectively. The aggregation process at each scale may involve mostly coalescing, i.e. size growth, but also could involve leaving and hence fragmentation, i.e. size shrinks or, if $R(t)$, $B(t)$ and $G(t)$ represents online activity, a loss of interest either as individuals or collectively. This means that the precise mechanisms of aggregation and hence coalescence and fragmentation, can be changed in the formalism to reflect the most common observed processes online. Each subpopulation $R(t)$, $B(t)$ and $G(t)$ can be thought of as a 'gel' at this next scale, like a community-of-communities, which is the key idea of renormalization in physics. It can also apply across platforms, using a community-of-communities-of-communities and so on. Since this can be applied at different scales depending on the definition of $G(t)$ etc., it can also apply to categories of the neutral communities in the analysis of online health debate: in this way, it can be extended to any number of types: $R(t)$, $B(t)$ and $G_1(t)$, $G_2(t)$, $G_3(t)$, . . . etc. for any topics and hence domains of interest (health, finance, hate etc.) at any scale within or across platforms.

The resulting equations that we derive and discuss, can be simplified or left in general form, and solved as we now show below. In the context of the present paper, specifically Fig. 3, the resulting 3 'gels' for the entire subpopulations $R(t)$, $B(t)$ and $G(t)$ respectively, then interact with each other because of the links between their constituent communities, which hence yields a set of coupled differential equations for $R(t)$, $B(t)$ and $G(t)$ respectively. Specifically, the mathematical model used in Figs. 3 and 4 is a simple set of coupled differential equations for the separate anti, pro and neutral subpopulations $R(t)$, $B(t)$ and $G(t)$ respectively. These quantities provide approximate aggregate descriptions of the coupled activity related to exposure of the anti, pro and neutral subpopulations.

Before we give these equations, we show how they are derived and hence justified, starting from the microscopic process of aggregation of individuals into communities. The full mathematical derivation starts with the online dynamics of clumps of individuals that can aggregate into an online gel. These clumps are equivalent to small pieces of a network that

are disconnected from the rest of the network, or pieces that are weakly connected to the rest of the network, since both can be described by the same approximate mean-field equations that we develop. These lead to a gel forming, or equivalently the giant connected component (GCC) of the network, which means it is a macroscopically large cluster.

We also need to comment on the potentially confusing terminology clump, cluster, community and group. Mathematically, there is no confusion since the issue simply concerns whether a collection of objects is microscopic (i.e. much smaller than the size of the population and hence not scaling with the size of the population) or macroscopic (i.e. a finite fraction of the population and hence scaling with the size of the population). Prior to the dynamical phase transition to a gel (or equivalently GCC), all clumps of objects are much smaller than the relevant population size and are hence microscopic at that scale. On the scale that the gel represents an individual community, the clumps are microsopically small clumps of correlated individuals (e.g. from WhatsApp) who come together to form a single Page or Group, and these clumps are not observable to us in our data collection. Under aggregation, following the dynamical phase transition, a clump becomes so large that it is macroscopic, i.e. a single large clump emerges that has a size comparable to the population size. This large clump is the gel, or a GCC in a network setting. In the end, the mathematical equations are exact and precise in terms of their form, while words such as 'group' can suffer from vagueness and alternative interpretations.

The mathematical aggregation theory that we present here can equivalently be viewed as applying to the linking together of objects in a network, or to the aggregation of objects in a more abstract setting since the coupled equations are equivalent in both cases at the mean-field level – and it can be applied at different scales. The words chosen to describe what these clumps, gels etc. are, then become a matter of choice and will be nuanced by academic discipline.

We could also add fragmentation to represent (a) the possible later shutting down of links by moderator action etc., and/or (b) loss of interest by individuals or collectively. However for simplicity we restrict our focus here to the aggregation process through which the movement is growing. In the language of networks, we note that the individual heterogeneity that we incorporate is an intrinsic heterogeneity of an individual node, i.e. it is not a node's degree of connectivity but rather an intrinsic property of that node, like a hidden set of attributes that could be referred to as a 'character'. More generally, the gel (or equivalently the GCC in a network) can emerge from some coupled combination of the axes corresponding to each dimension of the internal 'character' variable described below, however the same derivation applies.

We allow for heterogeneity among the interacting individuals, and consider the situation where this heterogeneity helps dictate the evolution of the aggregation process, i.e. it ultimately produces the distinct flavors of the individual Facebook Pages seen in the ecology online, and hence the distinction between subpopulations such as pros, antis and neutrals at the next scale of renormalization, and so on. A hidden variable x that we for simplicity call 'character', is randomly assigned to each individual taken from a given distribution $q(x)$. While this is undoubtedly a crude way of adding individual human heterogeneity, it is common practice in computational social science: moreover, x could be made a general D-dimensional vector $\vec{x} = (x_1, x_2, \ldots x_D)$ where D is an arbitrarily large number, without changing the analysis. A gel can hence form along each of these D dimensions, or combi-

nation of them, yielding a set of $\leq D$ gels, i.e. a set of $\leq D$ Facebook Pages or a set of $\leq D$ VKontakte Groups. Also, x could be made time-dependent at the cost of analytical complexity. The interaction between individuals is described in terms of their similarity or dissimilarity (diversity) and hence is some function of their respective x values. We incorporate this by first defining the similarity S_{ij} between individual i and individual j as $S_{ij} = 1 - |x_i - x_j|$, so that individuals with alike character have a high similarity and otherwise for a pair of individuals with unlike character. We consider that the probability of aggregation for any two individuals i and j is given by $\mathcal{C} = S_{ij}$. Our definition also recognizes the opposite mechanism of diversity (i.e. dissimilarity) which tends to form clumps of dissimilar individuals, where the aggregation probability between i and j is $\mathcal{C} = 1 - S_{ij}$. The random case is recovered in the limit where the aggregation probability is independent of x, which is $\mathcal{C} = 1$. Doing this in $D = 1, 2, 3...$ etc. dimensions leads to a set of diverse flavors of gel emerging, as observed empirically.

This heterogeneous aspect of the aggregation process is then transferred to the equations for the evolving population by means of a mean-field probability F for aggregation. Different calculations and values for F will follow according to the choice made for the initial composition of the population $q(x)$, and also for variants away from pure homophily and pure heterophily. In all cases, the resulting mean-field probability F determines the likelihood for any pair of individuals i and j to merge into a new clump at a given timestep t. To summarize so far, we have effectively bundled up all the subtleties of character and rules for grouping together, into a time- and population- average. For example, for a uniform character distribution $q(x)$, the probability density function (PDF) of the similarity $y = S_{ij}$, for homophily is $f(y) = 2y$ and hence the mean-field aggregation probability F, becomes:

$$F = \int_0^1 y f(y) dy = 2/3. \tag{1}$$

By contrast, for dissimilarity defining $z = 1 - S_{ij}$, the PDF $f(z) = 2(1 - z)$ resulting into a mean-field aggregation probability F, of:

$$F = \int_0^1 z f(z) dz = 1/3. \tag{2}$$

Future work will look at going beyond this mean-field description and hence go beyond this use of a single parameter F in the aggregation dynamical equations, e.g. include variation/spread in F so that both the mean and diversity of the population's heterogeneity are included. Future work will also allow the population's heterogeneity to evolve in time as individuals come and go, and also allow individuals to change their character over time. Both will impact the system by making, at the very least, F a function of time in the equations that follow.

We now develop a set of rate equations for the number of small clumps of individuals of size k ($k = 1, 2, ...$) over time. The mathematical material that we present here builds on from work in the physics, chemistry and mathematics literature, with the generalization that particles (individuals) that are typically treated as identical now have individual heterogeneity. Aggregation theories describe the interaction and growth of clumps traditionally by means of kernels that depend on the sizes of the interacting clumps. That is, two clumps of sizes i and

j merge forming a new clump of size $i+j$ at a rate given by the kernel K_{ij}. When the aggregation rate increases sufficiently rapidly with the size of the clumps, the system experiences a large-scale transition where a non-negligible fraction of the total population gather into the largest clump (i.e. a gel or equivalently a GCC forms). Its distance independence reflects the global reach of online interactions, and serves as a mean-field approximation in other settings. Studies have shown empirically that human grouping phenomena follow closely the traditional aggregation theory for the case in which the kernel is proportional to the sizes of the interacting clumps, i.e., $K_{ij} \propto ij$. It has been empirically verified for communities in human communication and collaboration networks. With this in mind we can write a set of equations for the number of clumps of size k (n_k), for the heterogeneous system as:

$$\dot{n}_k(t) = -2F\frac{kn_k}{N^2}\sum_{r=1}^{\infty} rn_r + \frac{F}{N^2}\sum_{r=1}^{k} rn_r(k-r)n_{k-r}, \quad k \geq 2 \tag{3}$$

$$\dot{n}_1(t) = -2F\frac{n_1}{N^2}\sum_{r=1}^{\infty} rn_r, \quad k = 1, \tag{4}$$

where N is the subpopulation from which a particular flavor of future Facebook Page or VKontakte Group (gel) might emerge if gelation occurs. The first term of Eqs. (1) and (2) represents the population of clumps of size k that merge with other clumps, while the second term in Eq. (1) is the population of smaller clumps that merge to form clumps of size k, consisting of the well-known product kernel. By considering $N = \sum_{r=1}^{\infty} kn_k$, Eq. (2) can be immediately solved and the expression for the number of isolated individuals is:

$$n_1(t) = Ne^{-\frac{2F}{N}t}, \tag{5}$$

where we have assumed that initially the system is comprised by individuals only, ($n_1(0) = N$). Equation (3) can be used to solve equation (1) for the case of $k = 2$. The result is:

$$n_2(t) = Fte^{-\frac{4Ft}{N}}. \tag{6}$$

Similarly, the found expressions for n_1 and n_2 are used in equation (1) to solve for $k = 3$ (n_3) resulting in:

$$n_3(t) = \frac{2F^2t^2}{N}e^{-\frac{6Ft}{N}}. \tag{7}$$

This recursive process is repeated to solve for higher k values. The general expression for any $k \geq 2$ is found to be:

$$n_k(t) = \frac{1}{k!}\left(\frac{k}{N}\right)^{k-2}(2Ft)^{k-1}\,e^{-\frac{2kF}{N}t}. \tag{8}$$

At some later point into the dynamics, a finite non-negligible fraction of the total population may condense into a single large cluster called a 'gel', or equivalently, a giant connected component GCC in a network system. This phenomena is known as gelation and divides the dynamics of the system. After the gel is formed, the moments of the size distribution become decomposed into the small clumps (or *solution*) and the gel in the following way:

$$M_j = \sum_{k\geq 1} k^j n_k = \sum_{sol} k^j n_k + (k^j n_k)_{gel}. \tag{9}$$

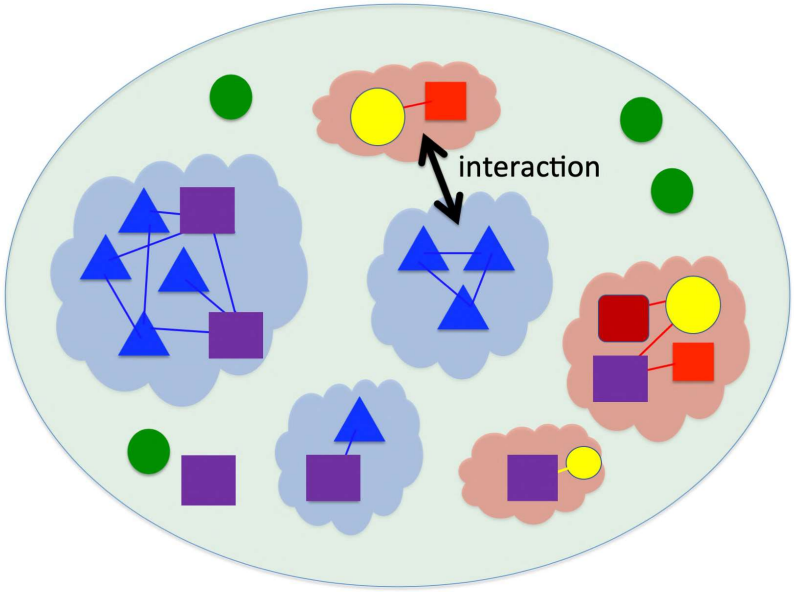

Figure S24: Overview of our model of the system of interest. It comprises a population of interacting, heterogeneous individuals (i.e. objects/nodes) and can be represented as a partially connected network of nodes, or in a more abstract way as pockets of coupled or correlated entities. The gel – or equivalently the giant connected component GCC – can then emerge from these as a result of aggregation.

The importance of this decomposition becomes evident when analyzing the zeroth moment, $M_0 = \sum_{k\geq 1} n_k$, which provides the number of clumps of any size. By looking at its first derivative we find:

$$\begin{aligned}
\frac{dM_0}{dt} &= \sum_{k\geq 1} \frac{dn_k}{dt} \\
&= -2F \sum_k \frac{kn_k}{N} + \frac{F}{N^2} \sum_k \sum_{i+j}^{k} (in_i)(jn_j) \\
&= -F, \quad \Rightarrow \quad M_0(t) = N - Ft. \qquad (10)
\end{aligned}$$

The solution for the zeroth moment becomes negative when $t > N/F$ which is problematic since M_0 gives the total number of clumps present. This problem is solved by using, above the gel point, $\sum_{k\geq 1} kn_k = N - G$, where G is the size of the gel. With this correction the derivative of the zeroth moment becomes:

$$\begin{aligned}
\frac{dM_0}{dt} &= -\frac{2F}{N}(N-G) + \frac{F}{N^2}(N-G)^2 \\
&= \frac{F}{N}(N-G)\left(\frac{N-G-2N}{N}\right) = -\frac{F}{N^2}(N-G)(N+G) \\
&= \frac{F}{N^2}\left(G^2 - N^2\right). \qquad (11)
\end{aligned}$$

Note that equation (9) indicates that the number of clumps stops decreasing when the gel reaches the system size N. The appearance of the gel is mathematically manifested as a singularity in the second moment of the size distribution. The evolution of the second moment is given by:

$$\frac{dM_2}{dt} = \sum_{k\geq 1} k^2 \frac{dn_k}{dt}. \qquad (12)$$

Using equation (1), we can work out this expression as follows:

$$\begin{aligned}
\frac{dM_2}{dt} &= \sum_{k\geq 1}\left(\frac{k}{N}\right)^2 F \sum_{i+j=k} i n_i j n_j - \sum_{k\geq 1} k^3 n_k \frac{2F}{N} \\
&= \sum_{\substack{i\geq 1\\ j\geq 1}} (i+j)^2 \frac{F}{N^2} i n_i j n_j - \sum_{k\geq 1} k^3 n_k \frac{2F}{N} \\
&= \sum_{\substack{i\geq 1\\ j\geq 1}} \left(i^3 j n_i n_j + 2i^2 j^2 n_i n_j + i j^3 n_i n_j\right) \frac{F}{N^2} - \sum_{k\geq 1} k^3 n_k \frac{2F}{N} \\
&= \sum_{\substack{i\geq 1\\ j\geq 1}} \left(i^2 n_i\right)\left(j^2 n_j\right) \frac{2F}{N^2} + \left(i^3 n_i j n_j + i n_i j^3 n_j\right) \frac{F}{N^2} - \sum_{k\geq 1} k^3 n_k \frac{2F}{N} \\
&= \sum_{\substack{i\geq 1\\ j\geq 1}} \left(i^2 n_i\right)\left(j^2 n_j\right) \frac{2F}{N^2} + \sum_{k\geq 1} k^3 n_k \frac{2F}{N} - \sum_{k\geq 1} k^3 n_k \frac{2F}{N} \\
&= \sum_{\substack{i\geq 1\\ j\geq 1}} \left(i^2 n_i\right)\left(j^2 n_j\right) \frac{2F}{N^2} \\
&= M_2^2 \frac{2F}{N}, \qquad (13)
\end{aligned}$$

which gives a closed differential equation for the second moment of the size distribution. The solution for the initial condition where all clumps are of size one ($M_2(0) = N$), is:

$$M_2(t) = \left(\frac{1}{N} - \frac{2Ft}{N^2}\right)^{-1}, \qquad (14)$$

which has a singularity at the time

$$t_{onset} = N/2F \qquad (15)$$

which is the onset time at which the transition to having a gel takes place.

This critical time for the onset of the gel or equivalently GCC, $t_{onset} = N/2F$, depends on (a) the mean-field aggregation probability F and hence on the nature of the aggregation process as well as the initial character distribution of the population $q(x)$, (b) the size of pool of potential recruits online N. For a uniform character distribution, unlike clumps (dissimilarity, or diversity) are slower to be formed and hence the transition occurs at a later time than alike cluster formation (homophily). Random clumps are the quickest to form since they have the maximum mean-field aggregation probability per timestep ($F = 1$).

This result is important, since it means that the time at which a gel (and hence page, or movement involving a collection of pages) appears is not some fixed number, but rather it is an effect that emerges from the conditions under which the aggregation occurs. Such a statement may superficially but erroneously be classed as 'obvious' to someone discussing the problem of online collective action verbally: in reality, it emerges from the mathematics

in a complex way that can be quantified and even used to predict this onset time. Instead of something collective appearing online 'out of nothing', this provides a way of exploring this onset and predicting it, then comparing to real data. This is what we ourselves did in looking at the emergence of pages and movements online surrounding extremism in the 2021 Scientific Reports paper referenced earlier.

The expression for the evolution of the gel size is obtained by means of the exponential generating function $\mathcal{E}(y,t) \equiv \sum_{k\geq 1} k n_k e^{yk}$. Hence:

$$\begin{aligned}
\frac{\partial \mathcal{E}}{\partial t} &= \sum_{k\geq 1} k \frac{\partial n_k}{\partial t} e^{yk} \\
&= -\frac{2F}{N} \sum_k k^2 n_k e^{yk} + \frac{F}{N^2} \sum_{i\geq 1} \sum_{j\geq 1} (i+j) i n_i j n_j e^{yk} \\
&= -\frac{2F}{N} \sum_k k^2 n_k e^{yk} + \frac{F}{N^2} \sum_i i^2 n_i e^{yi} \sum_j j n_j e^{yj} + \frac{F}{N^2} \sum_i i n_i e^{yi} \sum_j j^2 n_j e^{yj} \\
&= \frac{2F}{N^2} \mathcal{E} \frac{\partial \mathcal{E}}{\partial y} - \frac{2F}{N} \frac{\partial \mathcal{E}}{\partial y} \\
&= \frac{\partial \mathcal{E}}{\partial y} \frac{2F}{N} \left(\frac{\mathcal{E}}{N} - 1 \right). \qquad (16)
\end{aligned}$$

Equation (16) is known as the inviscid Burgers equation which is the simplest nonlinear hyperbolic equation and can be solved by the method of characteristics. For this type of partial differential equation, the characteristics are straight lines in the y-t plane where $\mathcal{E}$ is constant and have slope $\alpha(1-\mathcal{E}')$, where for simplicity we have defined $\alpha = 2F/N$ and $\mathcal{E}' = \mathcal{E}/N$. The equation of motion for y along the characteristic is therefore:

$$\frac{dy}{dt} = \alpha(1-\mathcal{E}'). \qquad (17)$$

Since $\mathcal{E}$ (and hence $\mathcal{E}'$) is constant, the solution for $y(t)$ along the characteristic is:

$$y(t) = \alpha(1-\mathcal{E}')t + f(\mathcal{E}), \qquad (18)$$

where $f(\mathcal{E})$ depends on the initial conditions which for the generating function we find it to be $\mathcal{E}(y, t=0) = Ne^y$ which yield $y(t=0) = \ln \mathcal{E}'$. The derivation moves forward as follows:

$$\begin{aligned}
y &= \ln \mathcal{E}' + \alpha t(1-\mathcal{E}) \\
e^y &= \mathcal{E}' e^{\alpha t(1-\mathcal{E}')} \\
e^{y-\alpha t} &= \mathcal{E}' e^{-\alpha t \mathcal{E}'}. \qquad (19)
\end{aligned}$$

Now note that the generating function for $y=0$ yields $\mathcal{E}(0,t) = N - G$ above the gel point and the following expression for the largest clump (cluster, i.e. gel or equivalently GCC) is found:

$$\frac{G}{N} = 1 - e^{-\frac{2Ft}{N^2} G}. \qquad (20)$$

The solution of equation (20) can by written by means of the W-Lambert function as:

$$G = N(1 - W\left(ze^z\right)/z), \quad z = -2Ft/N\ . \qquad (21)$$

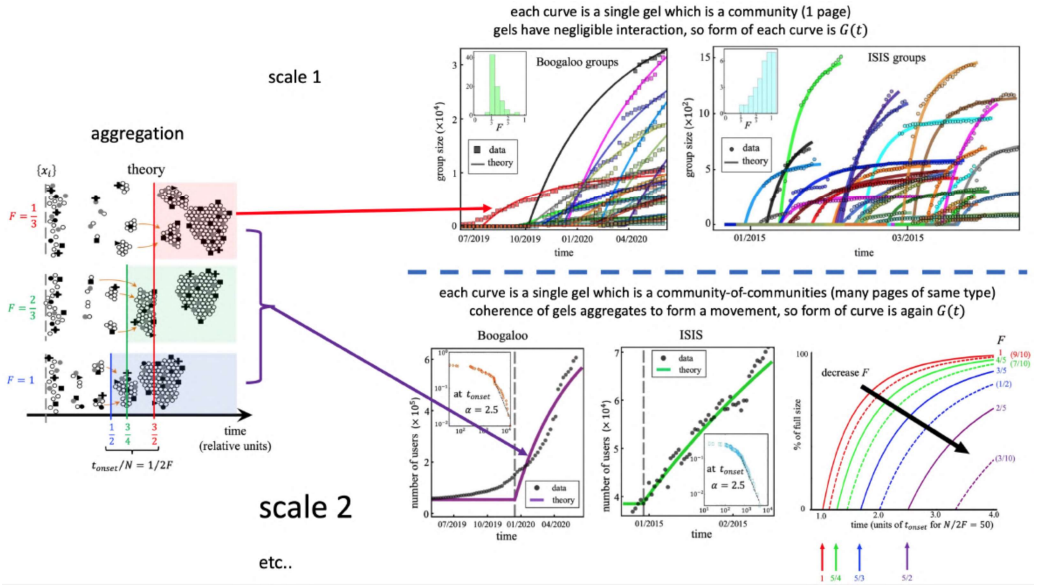

Figure S25: Demonstration of good agreement between our mathematical theory of online aggregation ($G(t)$ growth curve and onsets) at two scales: Top, the scale of individual pages (i.e. each gel is a community, i.e. a single page (Facebook for Boogaloos) or group (VKontakte for ISIS). Each curve is a gel $G_1(t)$, $G_1(t)$, $G_1(t)$ etc. for extremist pages 1, 2, 3 etc. The inferred values of F are shown for each, as a distribution in the inset. Bottom: the next scale level, across pages of a given type (community-of-communities, i.e. Facebook online pages for Boogaloo movement as a whole, or VKontakte online groups for ISIS movement as a whole). A group on VKontakte is often called a club, but is a community in any case. Each curve at this scale is a gel for the aggregated extremist pages 1, 2, 3 etc. from the lower scale.

This comparison also shows a non-obvious prediction of our theory that is borne out by the empirical data: the $5/2 = 2.5$ power-law scaling of sizes at the onset of the emergence at the next scale. We now give the analytical prediction of this result. The theoretical model predicts an approximate power-law (PL) size distribution at the transition point with exponent $-5/2 = -2.5$. This can be seen in equation (8) when we look at large k. Using

the Stirling approximation this equation can be written as:

$$\begin{aligned} n_k(t) &\approx \left(\frac{e}{k}\right)^k \left(\frac{k}{N}\right)^{k-2} \frac{1}{\sqrt{2\pi k}} (2Ft)^{k-1} e^{-2kFt/N} \\ &= \frac{N^2}{2Ft} \left(\frac{2Ft}{N}\right)^k \frac{1}{\sqrt{2\pi}} e^{-\frac{2F}{N}k\left(t-\frac{N}{2F}\right)} k^{-5/2} \\ &= \frac{N}{\sqrt{2\pi}} \left(\frac{t}{t_c}\right)^{k-1} e^{-\frac{k}{t_{onset}}(t-t_{onset})} k^{-5/2} \end{aligned} \tag{22}$$

which around $t \approx t_{onset}$, we can approximate $e^{-k(\tau-\ln\tau-1)} \to e^{-k(1-\tau)^2/2}$, which yields:

$$n_k(t) \to \frac{N}{\sqrt{2\pi}} e^{-\frac{k}{2}(1-\frac{t}{t_{onset}})^2} k^{-5/2}. \tag{23}$$

For $t = t_{onset}$ it yields a PL distribution with exponent $-5/2 = -2.5$ that also serves as a signal of the gel transition.

We could now go on and develop a full mathematical theory of interacting gels, as shown in Fig. S26. However we can greatly simplify this task here for the purposes of this project. To do this, we need to develop an approximate form for the mathematical expression for $G(t)$. We turn to this next.

There is no closed-form solution of the shape of the gel size $G(t)$ for a gel labelled i as a function of time. But we can develop a very similar and much simpler closed form approximation. Consider the differential equation for an entity (e.g. gel in our case) $\dot{R}_i(t) = r_R(R_{0,i} - R_i(t))$ where $R_i(t)$ is the size of gel i as a function of time. It turns out (see Fig. S27) that if we modify this to have time start from the gel onset time t_c, then it is a very good approximation to the actual gel solution developed above. The equation becomes:

$$\dot{R}_i(t) = H(t - t_{c,R_i}) r_R(R_{0,i} - R_i(t)) \tag{24}$$

and the solution of this is

$$R_i(t) = R_{0,i}(1 - e^{-r_R(t-t_{c,R_i})}) \tag{25}$$

for $t \geq t_{c,R_i}$ and $R_i(t) = 0$ for $t < t_{c,R_i}$. Figure S27 shows that this agrees very well with the gel shape $G(t)$, at least for times that are earlier on in the development which is any case the main interest.

If we then put multiple of these approximate gel equations together, with their own onsets t_c, and we let them interact, we get the types of deviations-from-perfect gel shapes that we observe on closer inspection empirically for the size of in-built communities (pages etc.) over time. We could also add a simple decay term with rate constant d_R,d_B, d_G to capture the loss of interest or activity and hence reduction in size, or from moderators removing users. This gives the following class of equations for the 3 gel case, in which there are linear couplings possible between the various gel types and also a decay term for each gel. These, and similar forms with decay terms set to zero etc., yield good agreement with the empirical data, and

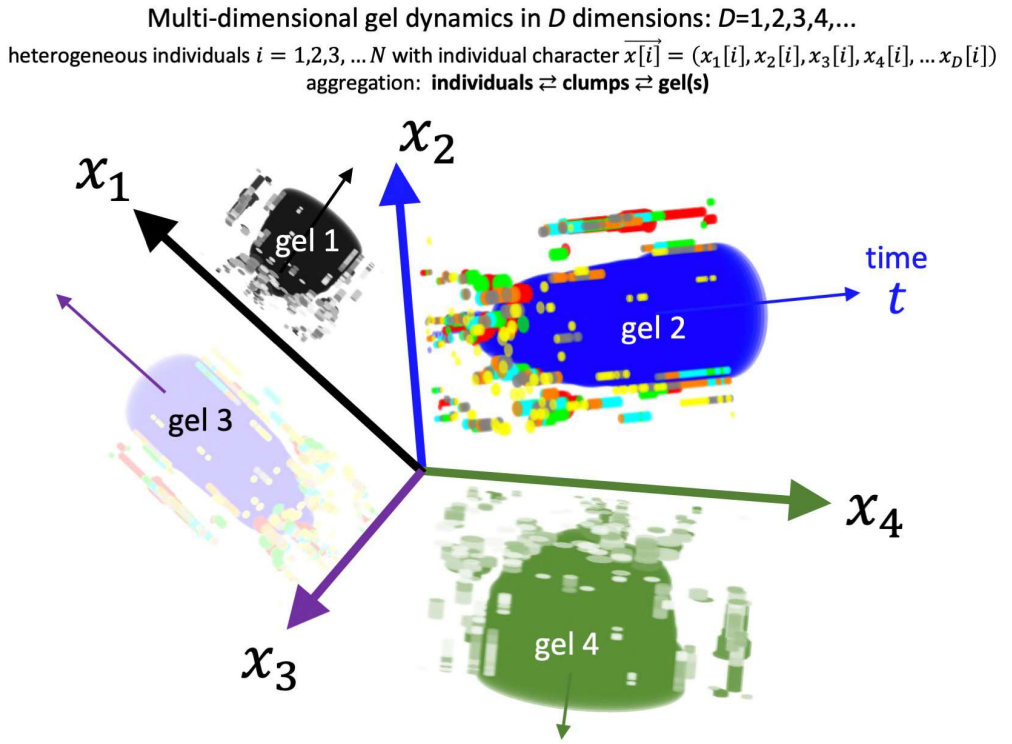

Figure S26: A gel – or equivalently a giant connected component GCC in a network – can emerge along any of the character dimensions. More generally, gels could form that combine character dimensions, and essentially the same mathematics applies since this only affects F at the mean-field level, but we keep the story simpler for the purposes of this discussion. We can apply this aggregation theory at multiple scales.

are the ones used in the theoretical plots in Fig. 3 of the main paper. Similar forms of equations with different parameter values are also illustrated in Fig. S28:

$$\begin{aligned}
\dot{R} &= H(t-t_{c,R})[r_R\,(R_0-R)+r_B(B-R)-d_R R] \\
\dot{B} &= H(t-t_{c,B})[b_B\,(B_0-B)-d_B B] \\
\dot{G} &= H(t-t_{c,G})[g_G\,(G_0-G)+g_R(R-G)+g_B(B-G)-d_G G]
\end{aligned} \tag{26}$$

This form of the equations contains the core feature that the dominant online couplings in the health debate involving $R(t)$ (antis), $B(t)$ (pros) and $G(t)$ (neutrals), are (i) $B(t)$ couples to neither of the others, just to itself and hence includes a possible decay, (ii) $R(t)$ couples with $B(t)$ and itself hence the decay, and (iii) $G(t)$ couples with both $R(t)$ and $B(t)$ and itself hence the decay. Again we emphasize, and show below explicitly, neither the specific forms nor the inclusion of different onset times and decays are important in terms of capturing the general shapes of the exposure dynamics curves and hence capturing the core gel-formation-and-interaction mechanism driving the exposure dynamics – they just serve to improve further the fit. As we show below, good fits can also be obtained with even simpler forms in which the onset times are the same and there are no decay terms. We show this explicitly, with statistics, in Sec. 5 of this SI.

This simple implementation of the model be extended to any number of categories, such as 2 Greens and 1 Red and 1 Blue and a wide range of uncoventional growth shapes can easily be produced. Examples are given in Figs. S29 and S30. Figures S31, S32, S33, S34 show some example growth curves in the online vaccine debate. Some of these empirical curves of online growth – whether at the level of individual pages or aggregates over pages of a given type, e.g. antis – look like the pure growth of the model for a single gel (or GCC). Others look like the irregular shapes that emerge from the coupled gels (GCCs). A longer term goal is to fit these by including the gel onset time, as we do in Fig. 3 of the main paper for aggregated Red, Blue and Green.

We now demonstrate the robustness of our theory, by demonstrating how the empirical data in Fig. 3 of the main paper can also be fit reasonably well using an even simpler version of our model in which all onset times are set to the same value (and hence all gel curves start from the same relative zero of time) and there are no decay terms, i.e. no fragmentation and hence no natural decay in size of the gel. Though the fit is not quite as good quantitatively as in Fig. 3 of the main paper, the advantage of this simpler version is that the analysis can all be done entirely analytically and yet still leads to interesting features that mimic the empirical data and capture visually the evolution of the dynamical curves. We now show this for the case of Red, Blue and Green. The expressions start from a general number of types, and then focus down to the Red, Blue and Green with the specific couplings discussed in the main paper. As shown, this still reproduces the general shapes observed in the exposure dynamics, which means that the fits that are shown in Fig. 3 are not the result of cherry-picking of parameter values in some complex model with many parameters, i.e. we are capturing core features of the exposure dynamics in Fig. 3 and hence lends further support to our claim of interacting gels being the core mechanism of online dynamics.

To illustrate how this enables analytical analysis, we now consider Red (R), Blue (B) and Green (G) subpopulations for the simplest implementation of the coupled gel dynamics. We start with a general discussion about couplings. In general, the dynamics coupling these subpopulations is described by the following time-dependent equation $\dot{x}_i = f(x_i) + \sum_j A_{ij} C_{ij}(x_i, x_j)$ where A_{ij} has components a_{ij} to represent the net links from j to i and $C_{ij}(x_i, x_j)$ is the dynamical coupling. Taking the Occam's razor approach of first trying the simplest model, we take the linear expansion of the coupling terms to first order, i.e. $C_{ij}(x_i, x_j) \sim g_{ij}(x_j - x_i) + \ldots$ where g_{ij} is the coupling constant. We set many of the couplings to zero because this set of influences shown in Fig. 2 of the main paper are key in our view. Internally, each subpopulation is a partially connected set of nodes where each node is an online community such as a Facebook Page. This internal structure of each subpopulation does not concern us here since we are analyzing a model of interacting subpopulations where the unit of analysis is the subpopulation. This system has the directed network structure as shown in Fig. 2 of the main paper. The coupling of G to the rest of the world is some outside term which does not concern us, since it can be taken as a term that feeds back to G like a growth term (i.e. it is included in g_G). Hence the equations become:

$$\begin{aligned} \dot{R} &= r_R (R_0 - R) + r_B (B - R) \\ \dot{B} &= b_B (B_0 - B) \\ \dot{G} &= g_G (G_0 - G) + g_R (R - G) + g_B (B - G) \end{aligned} \tag{27}$$

It makes sense that the coupling term, e.g. from Red into Green, will depend in some way on the difference between Red and Green activity levels $R(t)$ and $G(t)$. Positive couplings r_B, g_R and g_B imply positive feedback, e.g. $r_B > 0$ means that excess activity in B compared to R will increase the activity in R by increasing the value of rate of change of $R(t)$ and so on. Likewise, $r_B < 0$ implies negative feedback so that excess activity in B compared to R will decrease the activity in R by decreasing the value of rate of change of $R(t)$. The fixed point here is (R^*, B^*, G^*). The expressions are:

$$\begin{aligned} R^* &= \frac{r_R R_0 + r_B B_0}{(r_R + r_B)} \\ B^* &= B_0 \\ G^* &= \frac{g_G G_0 + g_B B_0}{(g_G + g_R + g_B)} + \frac{g_R (r_R R_0 + r_B B_0)}{(r_R + r_B)(g_G + g_R + g_B)} \, . \end{aligned} \tag{28}$$

Taking $\boldsymbol{\epsilon} = (\epsilon_R, \epsilon_B, \epsilon_G) = (R - R^*, B - B^*, G - G^*)$ as the deviation from the stable point (R^*, B^*, G^*), we obtain:

$$\begin{aligned} \frac{\mathrm{d}\epsilon_R}{\mathrm{d}t} &= -r_R \epsilon_R + r_B (\epsilon_B - \epsilon_R) \\ \frac{\mathrm{d}\epsilon_B}{\mathrm{d}t} &= -b_B \epsilon_B \\ \frac{\mathrm{d}\epsilon_G}{\mathrm{d}t} &= -g_G \epsilon_G + g_R (\epsilon_R - \epsilon_G) + g_B (\epsilon_B - \epsilon_G) \end{aligned} \tag{29}$$

This equation can be written:

$$\frac{\mathrm{d}\boldsymbol{\epsilon}}{\mathrm{d}t} = \mathcal{M}\boldsymbol{\epsilon} \tag{30}$$

$$\mathcal{M} = \begin{pmatrix} -(r_R + r_B) & r_B & 0 \\ 0 & -b_B & 0 \\ g_R & g_B & -(g_G + g_R + g_B) \end{pmatrix} \tag{31}$$

and this matrix has eigenvalues:

$$\begin{aligned} \lambda_1 &= -b_B \\ \lambda_2 &= -(r_R + r_B) \\ \lambda_3 &= -(g_G + g_R + g_B) \end{aligned} \tag{32}$$

Figure S36 shows that this 3 gel system ($B(t)$ not constant) with this simplification of identical onset times, reproduces the general shapes observed in Fig. 3 of the main paper. This is a simpler model than in the main paper since it does not have different possible onset times, yet it still presents an acceptable reproduction of the observed empirical trends – hence it demonstrates the robustness of our overall aggregation theory.

The predictions in Fig. 5(a)-(d) of the main paper consider the scenario that B remains at its fixed point at all times ($B(t)$ constant and hence ($\epsilon_B = 0$)) because the pros have reached their maximum capacity in terms of production of best-science guidance. The dynamics then reduce to a two dimensional case. From the initial equations,

$$\begin{aligned} \frac{\mathrm{d}\epsilon_R}{\mathrm{d}t} &= \lambda_2 \epsilon_R \\ \frac{\mathrm{d}\epsilon_G}{\mathrm{d}t} &= \lambda_3 \epsilon_G + g_R \epsilon_R \\ \text{hence in 2x2 matrix form} \quad \frac{\mathrm{d}\boldsymbol{\epsilon}}{\mathrm{d}t} &= \mathcal{M}\boldsymbol{\epsilon} \end{aligned} \tag{33}$$

The eigenvalues and the corresponding eigenvectors are λ_2, λ_3 and $\begin{pmatrix} \lambda_2 - \lambda_3 \\ g_R \end{pmatrix}$, $\begin{pmatrix} 0 \\ 1 \end{pmatrix}$ respectively. Writing $\boldsymbol{\epsilon}$ as a linear combination of the eigenvectors of $\mathcal{M}$, we have

$$\boldsymbol{\epsilon}(t) = \mathcal{C}_1 e^{\lambda_2 t} \begin{pmatrix} \lambda_2 - \lambda_3 \\ g_R \end{pmatrix} + \mathcal{C}_2 e^{\lambda_3 t} \begin{pmatrix} 0 \\ 1 \end{pmatrix} \tag{34}$$

The corresponding solutions for the deviation from the fixed points R^* and G^* are thus,

$$\begin{aligned} \epsilon_R(t) &= \epsilon_R(0)\, e^{\lambda_2 t} \\ \epsilon_G(t) &= \frac{g_R \epsilon_R(0)}{\lambda_2 - \lambda_3} e^{\lambda_2 t} + \left[\epsilon_G(0) - \frac{g_R \epsilon_R(0)}{\lambda_2 - \lambda_3} \right] e^{\lambda_3 t} \end{aligned} \tag{35}$$

Hence

$$\begin{aligned} R(t) &= \frac{r_R R_0 + r_B B_0}{(r_R + r_B)} + \epsilon_R(0)\, e^{\lambda_2 t} \\ G(t) &= \frac{g_G G_0 + g_B B_0}{(g_G + g_R + g_B)} + \frac{g_R(r_R R_0 + r_B B_0)}{(r_R + r_B)(g_G + g_R + g_B)} + \frac{g_R \epsilon_R(0)}{\lambda_2 - \lambda_3} e^{\lambda_2 t} + \left[\epsilon_G(0) - \frac{g_R \epsilon_R(0)}{\lambda_2 - \lambda_3}\right] e^{\lambda_3 t} \end{aligned} \tag{36}$$

These equations generate the predictions in Fig. 5(a)-(d) of the main paper. These resulting predictions in Fig. 5(a)-(d), and the following figures here (lower panels of Fig. S36, and Figs. S37-S40 and S41) consider this scenario that B remains at its fixed point at all times ($B(t)$ constant and hence ($\epsilon_B = 0$)) because the pros have reached their maximum capacity in terms of production of best-science guidance. Specifically, Figs. S37-S40 verifies the parameters and results in Fig. 5(a)-(d) of the main paper. Figure S41 explains the curves in Fig. 5(a)-(d) by showing the corresponding trajectories in phase space. Since $B(t)$ is constant, the space is two dimensional and spanned by $R(t)$ and $G(t)$.

The parameter values used in Fig. 3 for the fits shown, are for the receivers (Fig. 3(b)): $R_0 = 0.662$, $B_0 = 0.887$, $G_0 = 1.28$, $r_R = 0.871$, $b_B = 0.351$, $g_G = -0.244$, $t_{c,R} = 1.34$, $t_{c,B} = 2.08$, $t_{c,G} = 1.73$, $d_R = 0$, $d_B = 0.378$, $d_G = 0.247$, $r_B = -0.361$, $r_G = 0$, $b_R = 0$, $b_G = 0$, $g_B = -0.895$, $g_R = 1.61$. The parameter values used in Fig. 3 for the fits shown, are for the emitters (Fig. 3(d)): $R_0 = 0.985$, $B_0 = 0.656$, $G_0 = 0.511$, $r_R = 0.202$, $b_B = 0.245$, $g_G = 4.52$, $t_{c,R} = 2.05$, $t_{c,B} = 2.38$, $t_{c,G} = 5.26$, $d_R = -0.0384$, $d_B = 0.419$, $d_G = 1.04$, $r_B = 0.683$, $r_G = 0$, $b_R = 0$, $b_G = 0$, $g_B = 0.687$, $g_R = -5.43$. However we stress again that almost as good fits can be obtained with far fewer parameters (specifically all the onset times and decays equal to zero) as shown in Sec. 5 and this Sec. 6.

```
(* Comparison of Gel function form
(RED: PRL 2018 W-Lambert)
to simple RBG uncoupled node equation
dx/dt=a[x(infinity)-x]=a(100-x)
solution
x(t)=x(infinity)[1-exp(-at)] shifted to start at t_c (BLUE and GREEN curves)
with N=100  F=1.0 so t_c=N/2F=50 *)

Plot[
{100(1-ProductLog[(-2(1.0)t/(100)) Exp[-2(1.0)t/(100)]]/(-
2(1.0)t/(100))),100HeavisideTheta[t-50](1-Exp[-(t-50)/32]),100(1/2+1/2*Tanh[(t-3.0)/5])(1-
Exp[-(t-50)/32])},{t,0,200},PlotStyle-
>{{Red,Thick},{Blue,Dashed,Thick},{Black,Dotted,Thick}},
PlotRange->{0,100},
PlotTheme->"Business",
GridLines->{{50}, {-50}},
GridLinesStyle->Directive[Gray, Dashed, Thick],
Epilog->{
Inset[Style["time t",20,Bold],{195,5}],
Inset[Style["gel (or GCC)\nsize G(t)",20,Bold],{12,50}],
Inset[Framed[Style["Subscript[t, c]=50",20],Background->White],{50,40}],
Inset[Style[Rotate["exact (Lambert function)",40 Degree],20,Red],{80,85}],
{Red,Arrowheads[0.02],Arrow[{{81,84},{93,76}}]},
Inset[Style[Rotate["solution of approximate equation",40 Degree],20],{105,65}],
{Arrowheads[0.02],Arrow[{{104,66},{94,74}}]}
}
]
```

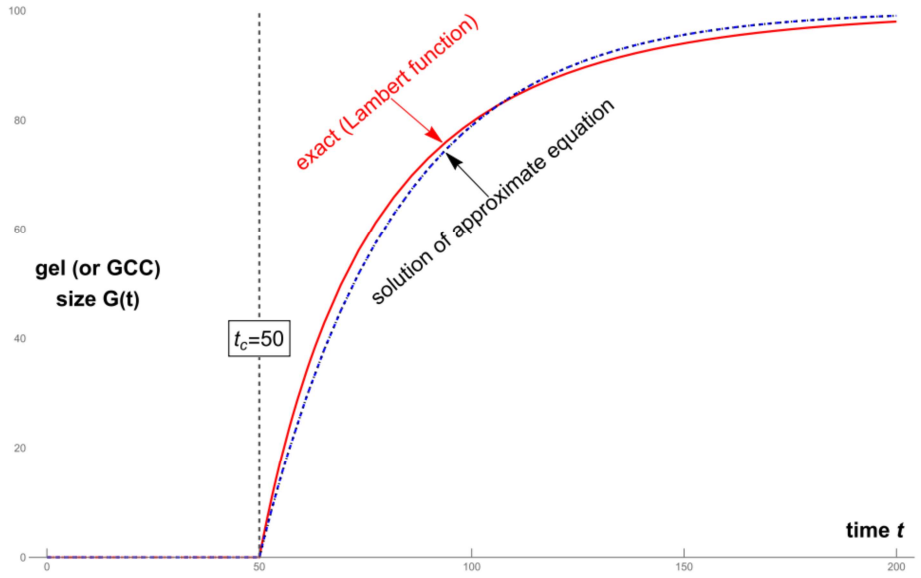

Figure S27: Comparison of actual gel (GCC) growth (red curve) to approximate solution (Eq. 25) of approximate differential equation Eq. 24 (blue and green curves are different approximations to the sudden onset, but agree so well that they lie on top of each other visually).

```
(* 3 BODY CASE: Red-Blue-Green *)
(* Code below is GENERAL: linear general model for Red-Blue-Green including intrinsic
   growth, couplings, and intrinsic decay *and* delayed onsets using ArcTan Heaviside onset
   approximate function*)
(* Parameters are for comparison to EMITTER plot *)
ClearAll["Global`*"];
ClearSystemCache[];
(* intrinsic growth rates *)
rR=2.0; (* growth rate R *)
bB=0.5; (* growth rate B *)
gG=1.0; (* growth rate G *)
(* intrinsic decay rates *)
dR=0.0; (* decay rates all ZERO *)
dB=0.0;
dG=0.0;
(* coupling of Red with X *)
rB=1.0;
rG=0.0;
(* coupling of Blue with X *)
bG=0.0;
bR=0.0;
(* coupling of Green with X *)
gR=0.5;
gB=-1.0;
(* X0 mimics capacity levels for X=Red, Blue and Green *)
R0=1.2;
B0=0.8;
G0=0.8;
NDSolve[{
R'[t]==(1/2+1/Pi*ArcTan[(t-0.2)/0.05])((rR(R0-R[t]) +rB(B[t]-R[t])+rG(G[t]-R[t])-dR R[t])),
B'[t]==(1/2+1/Pi*ArcTan[(t-1.0)/0.05])((bB(B0-B[t]) +bG(G[t]-B[t])+bR(R[t]-B[t])-dB B[t])),
G'[t]==(1/2+1/Pi*ArcTan[(t-1.0)/0.05])((gG(G0-G[t]) +gR(R[t]-G[t])+gB(B[t]-G[t])-dG G[t])),
(* initial conditions *)
R[0]==0.2,
B[0]==0.2,
G[0]==0.2
},{R,B,G},{t,0,100},MaxSteps->20000];
Plot[Evaluate[{R[t],B[t],G[t]}/.%],{t,0,12},PlotStyle->{Red,Blue,Green},PlotRange-
   >{{0,15.0},{0.0,1.5}},PlotPoints->100,Axes->True,AxesLabel->{time,activity}]
```

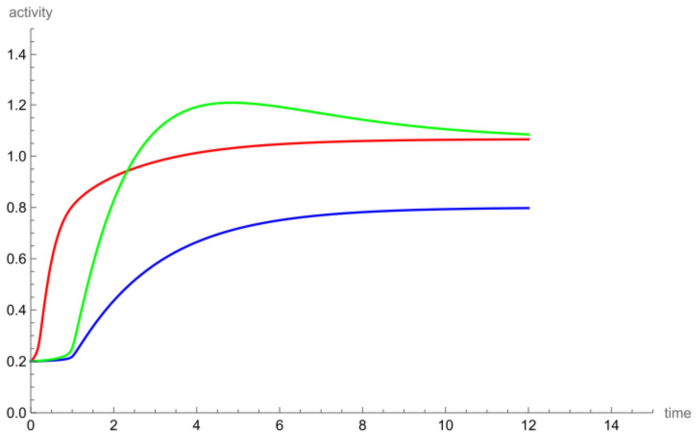

Figure S28: A simple example of 3 coupled gels each with own onset time.

```
(* 4 BODY CASE: Red-Blue-Green1-Green2 *)
(* Code is GENERAL: Linear but general model for Red-Blue-Green1-Green2.. including
   intrinsic growth, couplings, and intrinsic decay *and* delayed onsets using ArcTan Heaviside
   onset approximate function*)
(* Parameters are for comparison to SUBSCRIBER plot *)
ClearAll["Global`*"];
ClearSystemCache[];
(* intrinsic growth rates *)
rR=1.0; (* growth rate R *)
bB=1.0; (* growth rate B *)
g1G1=1.0; (* growth rate G1 - green curve *)
g2G2=1.0; (* growth rate G2 - black curve *)
(* intrinsic decay rates *)
dR=0.0; (* decay rates all ZERO *)
dB=0.0; dG1=0.0; dG2=0.0;
(* coupling of Red with X *)
rB=0.0; rG1=0.0; rG2=0.0;
(* coupling of Blue with X *)
bG1=0.0; bG2=0.0; bR=0.0;
(* coupling of Greens with X *)
g1R=1.0; g1B=1.0; g2R=1.0; g2B=1.0; g1G2=1.0; g2G1=-5.1;
(* X0 mimics capacity levels for X=Red, Blue and Greens *)
R0=1.0; B0=1.0; G10=1.0; G20=1.5;
NDSolve[{
R'[t]==(1/2+1/Pi*ArcTan[(t-1.0)/0.01])((rR(R0-R[t]) +rB(B[t]-R[t])+rG1(G1[t]-
   R[t])+rG2(G2[t]-R[t])-dR R[t])),
B'[t]==(1/2+1/Pi*ArcTan[(t-4.0)/0.01])((bB(B0-B[t]) +bR(R[t]-B[t])+bG1(G1[t]-
   B[t])+bG2(G2[t]-B[t])-dB B[t])),
G1'[t]==(1/2+1/Pi*ArcTan[(t-2.0)/0.01])((g1G1(G10-G1[t]) +g1R(R[t]-G1[t])+g1B(B[t]-
   G1[t])+g1G2(G2[t]-G1[t])-dG1 G1[t])),
G2'[t]==((1/2+1/Pi*ArcTan[(t-3.0)/0.01])(g2G2(G20-G2[t]) +g2R(R[t]-G2[t])+g2B(B[t]-
   G2[t])+g2G1(G1[t]-G2[t])-dG2 G2[t])),
(* initial conditions *)
R[0]==0.0,
B[0]==0.0,
G1[0]==0.0,
G2[0]==0.0
},{R,B,G1,G2},{t,0,100},MaxSteps->20000];
Plot[Evaluate[{R[t],B[t],G1[t],G2[t]}/.%],{t,0,10},PlotStyle-
   >{{Red,Dashed},{Blue,Dashed},{Green,Thick},{Black,Thick}},PlotRange-
   >{{0,6.0},{0.0,1.2}},PlotPoints->100,Axes->True,AxesLabel->{time,activity}]
```

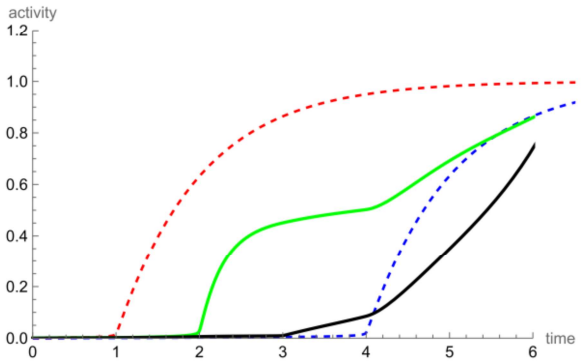

Figure S29: A simple example of 4 coupled gels (2 Greens, 1 Red, 1 Blue) each with own onset time.

```
(* 4 BODY CASE: Red-Blue-Green1-Green2 *)
(* Code is GENERAL: Linear but general model for Red-Blue-Green1-Green2.. including
   intrinsic growth, couplings, and intrinsic decay *and* delayed onsets using ArcTan Heaviside
   onset approximate function*)
(* Parameters are for comparison to SUBSCRIBER plot *)
ClearAll["Global`*"];
ClearSystemCache[];
(* intrinsic growth rates *)
rR=1.0; (* growth rate R *)
bB=1.0; (* growth rate B *)
g1G1=1.0; (* growth rate G1 - green curve *)
g2G2=1.0; (* growth rate G2 - black curve *)
(* intrinsic decay rates *)
dR=0.0; (* decay rates all ZERO *)
dB=0.0; dG1=0.0; dG2=0.0;
(* coupling of Red with X *)
rB=0.0; rG1=0.0; rG2=0.0;
(* coupling of Blue with X *)
bG1=0.0; bG2=0.0; bR=0.0;
(* coupling of Greens with X *)
g1R=1.0; g1B=1.0; g2R=1.0; g2B=1.0; g1G2=1.0; g2G1=-1.0;
(* X0 mimics capacity levels for X=Red, Blue and Greens *)
R0=1.0; B0=1.0; G10=1.0; G20=1.5;
NDSolve[{
R'[t]==(1/2+1/Pi*ArcTan[(t-1.0)/0.01])((rR(R0-R[t]) +rB(B[t]-R[t])+rG1(G1[t]-
   R[t])+rG2(G2[t]-R[t])-dR R[t])),
B'[t]==(1/2+1/Pi*ArcTan[(t-4.0)/0.01])((bB(B0-B[t]) +bR(R[t]-B[t])+bG1(G1[t]-
   B[t])+bG2(G2[t]-B[t])-dB B[t])),
G1'[t]==(1/2+1/Pi*ArcTan[(t-2.0)/0.01])((g1G1(G10-G1[t]) +g1R(R[t]-G1[t])+g1B(B[t]-
   G1[t])+g1G2(G2[t]-G1[t])-dG1 G1[t])),
G2'[t]==((1/2+1/Pi*ArcTan[(t-3.0)/0.01])(g2G2(G20-G2[t]) +g2R(R[t]-G2[t])+g2B(B[t]-
   G2[t])+g2G1(G1[t]-G2[t])-dG2 G2[t])),
(* initial conditions *)
R[0]==0.0,
B[0]==0.0,
G1[0]==0.0,
G2[0]==0.0
},{R,B,G1,G2},{t,0,100},MaxSteps->20000];
Plot[Evaluate[{R[t],B[t],G1[t],G2[t]}/.%],{t,0,10},PlotStyle-
   >{{Red,Dashed},{Blue,Dashed},{Green,Thick},{Black,Thick}},PlotRange-
   >{{0,6.0},{0.0,1.2}},PlotPoints->100,Axes->True,AxesLabel->{time,activity}]
```

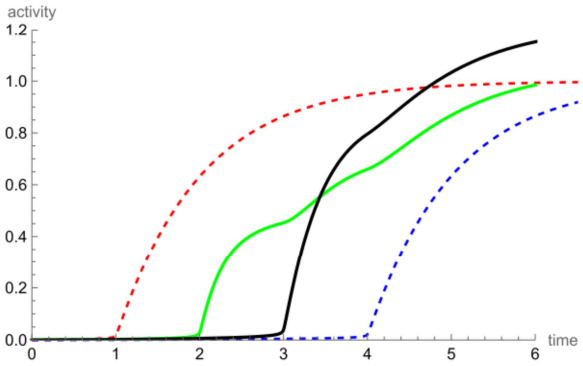

Figure S30: Another simple example of 4 coupled gels (2 Greens, 1 Red, 1 Blue) each with own onset time.

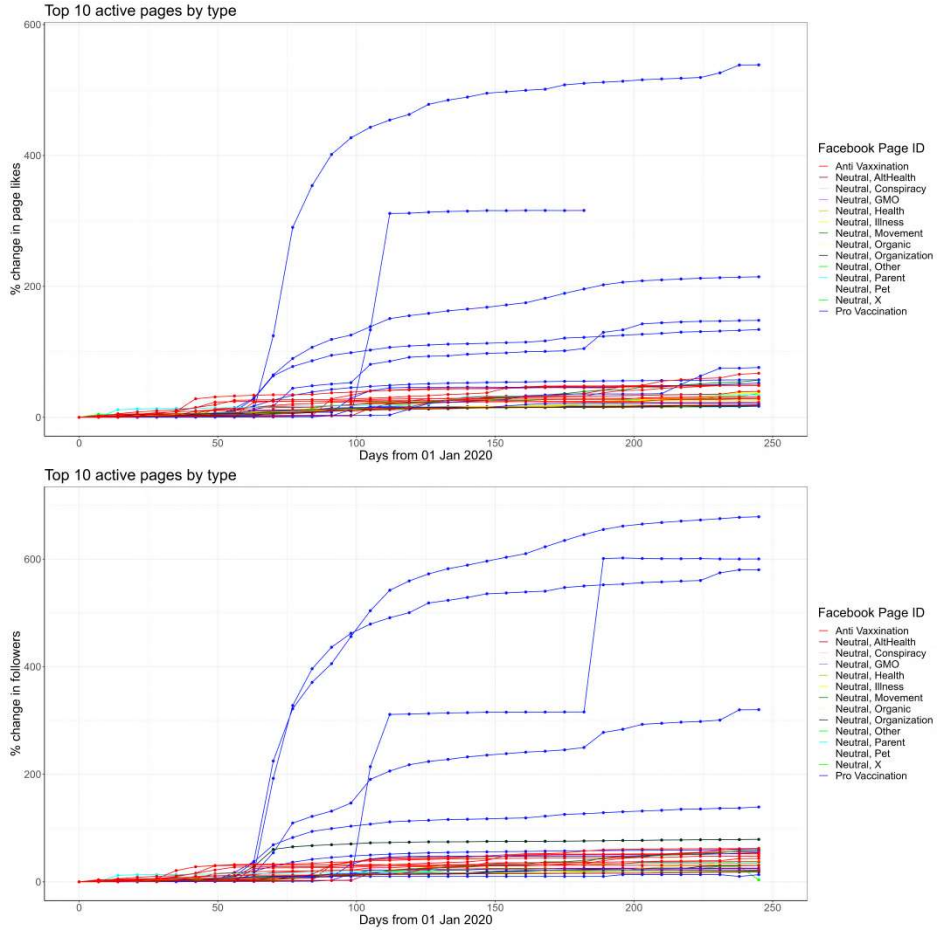

Figure S31: Empirical data for the three populations involved in the online vaccine debate: Red, Blue and Green (which has multiple subcategories). Each curve is the aggregate of all the in-built communities (pages) of a given type. The top ones are shown by likes (members).

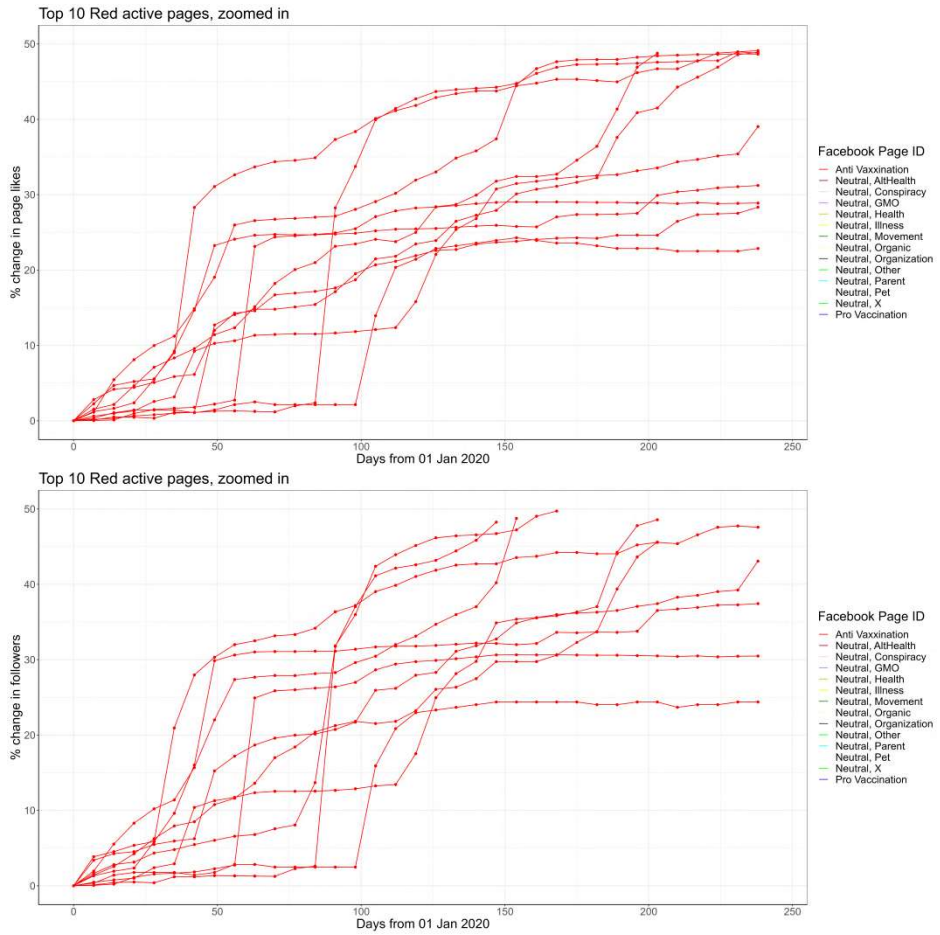

Figure S32: Each curve is an in-built community (page) in the online vaccine debate that we label as Red. The top ones are shown by likes (members).

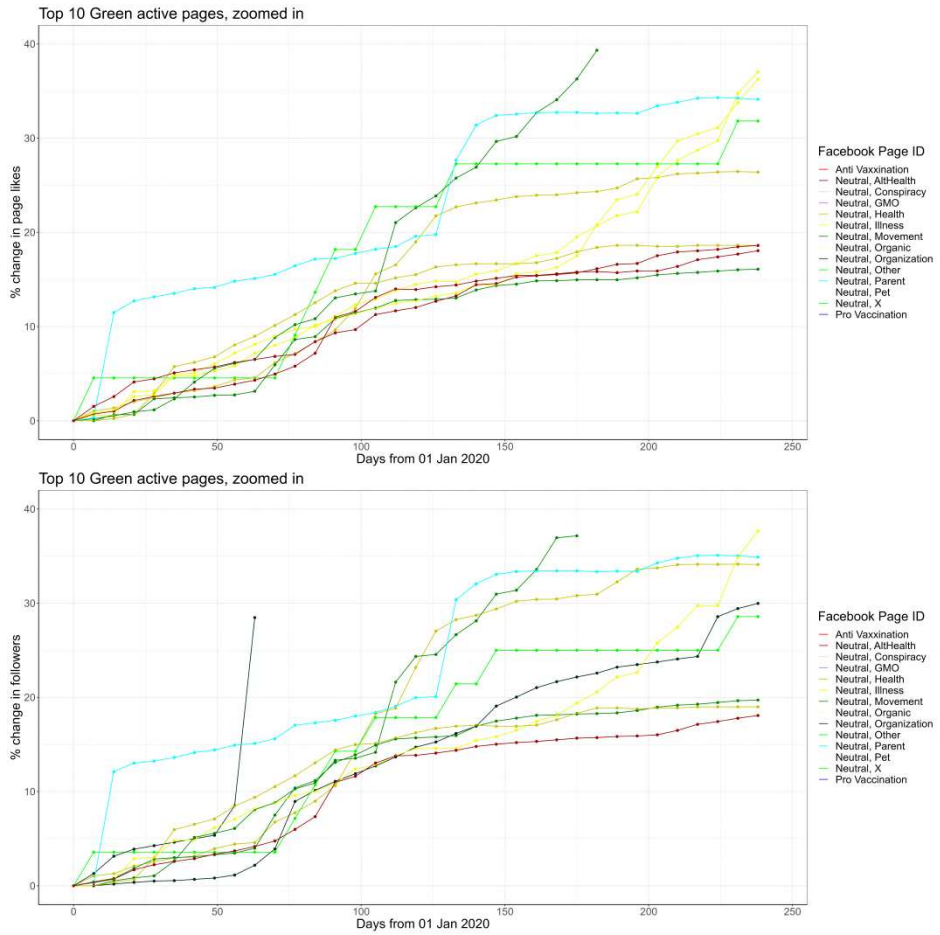

Figure S33: Each curve is an in-built community (page) in the online vaccine debate that we label as Green. The top ones are shown by likes (members).

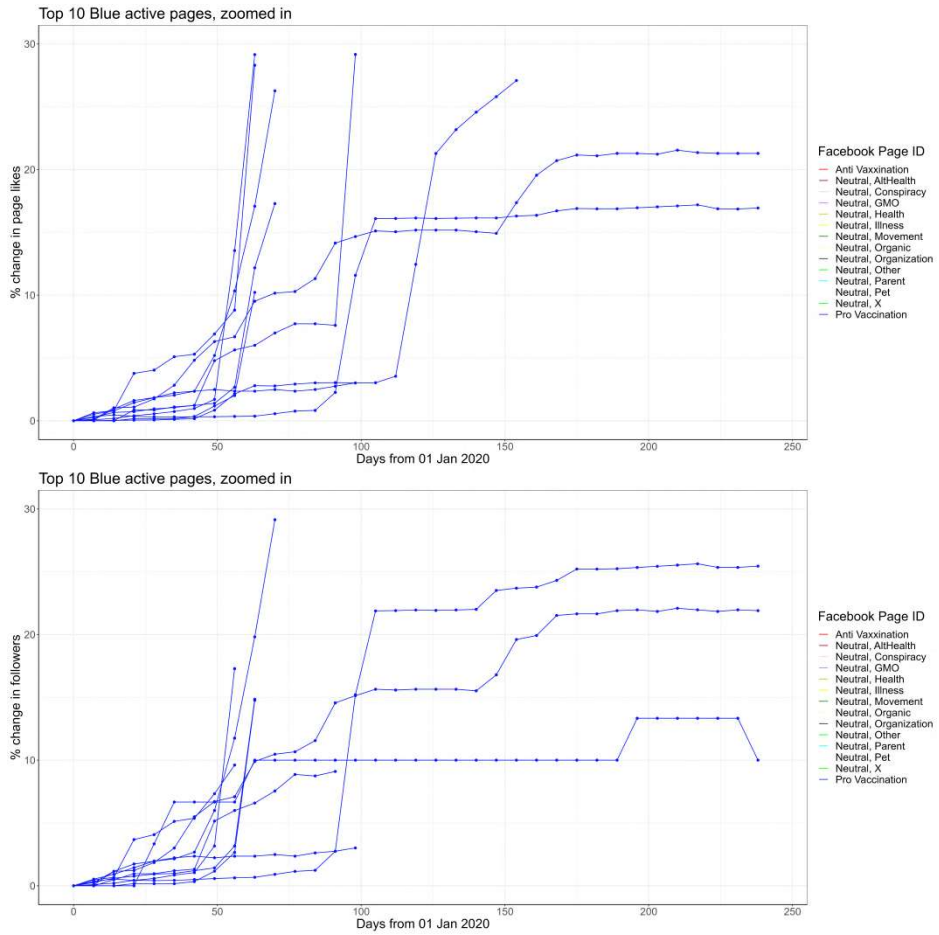

Figure S34: Each curve is an in-built community (page) in the online vaccine debate that we label as Blue. The top ones are shown by likes (members).

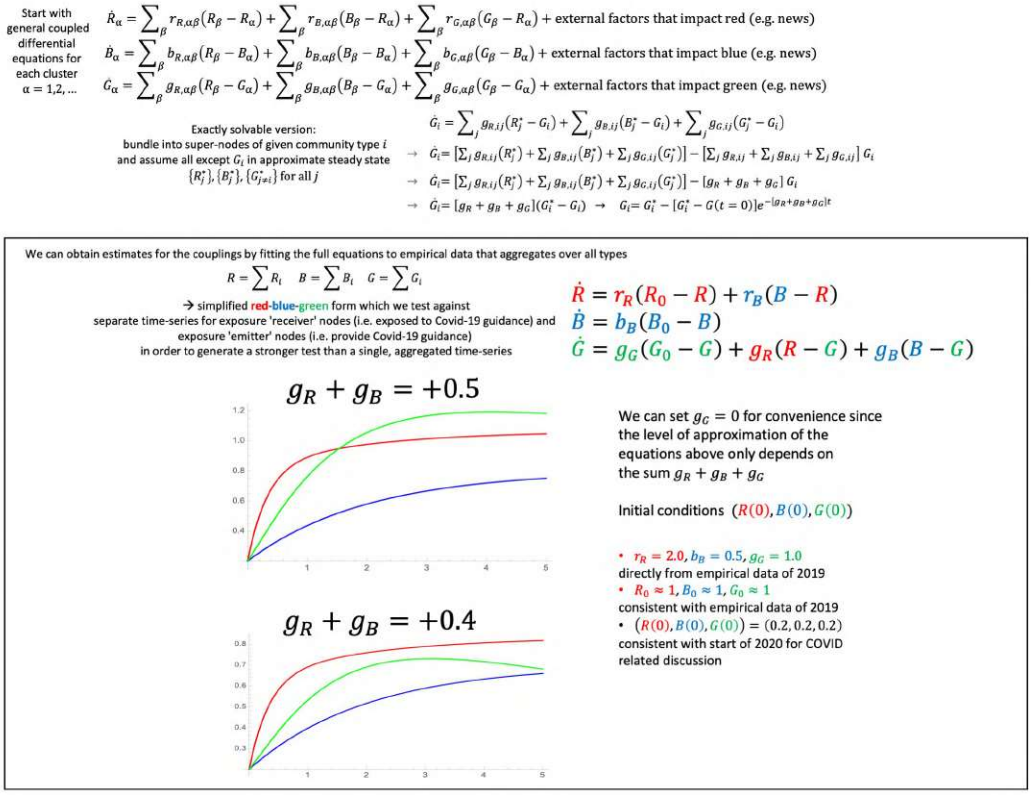

Figure S35: General equations for multiple gels (or GCCs) in the limit that all the onset times are set equal and hence can be taken as $t = 0$ on the axis. We also show an exactly solvable example of this in the case of steady state of all gels other than gel i. We then consider just 3 gels and we set some of the couplings to be zero, as in Fig. 2 of the main paper. This leads to a visually close reproduction of the empirical data in Fig. 3. The actual theoretical plots in Fig. 3 allow for different onset times, but the resulting difference in shapes is quite minor.

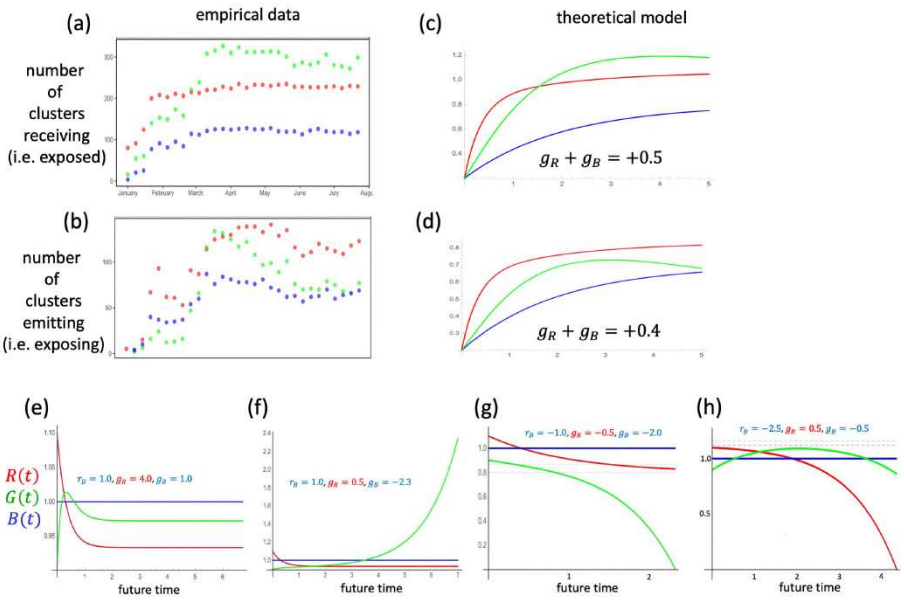

Figure S36: We show the output of the simple 3 gel system with the further simplification of identical onset times. As shown, this still reproduces the general shapes observed in the exposure dynamics separately as shown in Fig. 3 of the main paper, for number of communities exposed to Covid-19 guidance (i.e. number of 'receivers') and the number of communities exposing Covid-19 guidance (i.e. number of 'emitters'). The actual theoretical plots in Fig. 3 allow for different onset times, but the resulting difference in shapes is quite minor. The lower panels show predictions of this model, as in Fig. 5(a)-(d) of the main paper, for the situation of steady-state $B(t)$. The analysis is in the next figure shows the underlying Mathematica code for these prediction curves.

```
ClearAll["Global`*"];
(*colors={{RGBColor[187/244,85/244,102/244],Thickness[0.01]},{RGBColor[0/244,68/244,136/244]
,Thickness[0.01]},{RGBColor[221/224,170/224,51/224],Thickness[0.01]}};*)
col={{RGBColor[187/244,85/244,102/244],Thickness[0.01]},{RGBColor[0/244,68/244,136/244],Thic
kness[0.01]},{RGBColor[221/224,170/224,51/224],Thickness[0.01]},{RGBColor[187/244,85/244,102
/244],Thick,Dashed},{RGBColor[0/244,68/244,136/244],Thick,Dashed},{RGBColor[221/224,170/224,
51/224],Thick,Dashed}};
(* Same parameters as for 2020 fit, just slight excess of R and lower G; leads to switching
behavior in G between gB=-2.3 and gB=-2.4. *)
ClearSystemCache[];
aR=2.0; (* Growth rate R *)
aB=0.5; (* Growth rate B *)
aG=1.0; (* Growth rate G *)
rB=1; (* ve=copy, -ve=contra *)
(* rB0=4; (* rB=rBt*rBt t *)
rBt=0.0; (* rB=rB0*rBt t *)*)
gR=4.0; (* ve=copy, -ve=contra *)
gB=1.0; (* ve=copy, -ve=contra *)
Rfin=0.9; Bfin=1; Gfin=1.1;
Rstar:=(aR Rfin+rB Bfin)/(aR + rB)
Bstar := Bfin
Gstar := (aG Gfin+gB Bfin)/(aG+gR+gB)+(gR(aR Rfin +rB Bfin))/((aR+rB)(aG+gR+gB))
NDSolve[{
R'[t]==aR*(Rfin-R[t])+rB*(B[t]-R[t]),
B'[t]==aB*(Bfin-B[t]),
G'[t]==aG*(Gfin-G[t])+gR*(R[t]-G[t])+gB*(B[t]-G[t]),
R[0]==1.1, B[0]==1.0, G[0]==0.9
},{R,B,G},{t,0,100}];
p1=Plot[Evaluate[{R[t],B[t],G[t],Rstar,Bstar,Gstar}/.%],{t,0,7},PlotStyle->col,PlotRange-
>All,PlotPoints->100,
(*FrameLabel\[Rule]{Style["future time",Black,Bold,30],""},*)
Frame->{True,True,False,False},AspectRatio->0.75,
Epilog->{Inset[Framed[Style["R(t)",Bold,Italic,30,colors[[1,1]]],Background-
>White,FrameStyle-
>White],{6,0.935}],Inset[Framed[Style["B(t)",Bold,Italic,30,colors[[2,1]]],Background-
>White,FrameStyle-
>White],{6,1}],Inset[Framed[Style["G(t)",Bold,Italic,30,colors[[3,1]]],Background-
>White,FrameStyle->White],{6,0.972}]},FrameTicks->{{Automatic,None},{None,None}},
ImageSize->Large
]
```

0 -ve

$a_R+r_B=3.0$; $a_G+g_R+g_B=6.0$

$g_R>0$ and $g_B>0$

λ_1, $\lambda_2<0$ (stable); $\lambda_3<0$ (stable)

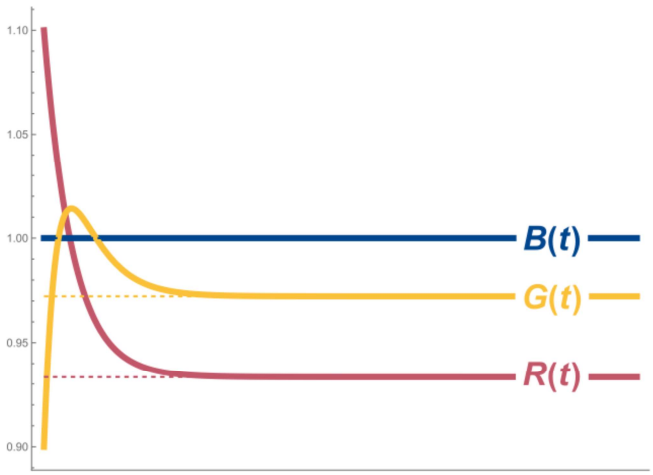

Figure S37: For convenience, we use here the notation, as in the Mathematica code, that $a_R \equiv r_R$, $a_B \equiv b_B$ and $a_G \equiv g_G$. This is Fig. 5(a).

```
ClearSystemCache[];
aR=2.0; (* Growth rate R *)
aB=0.5; (* Growth rate B *)
aG=1.0; (* Growth rate G *)
rB=1; (* ve=copy, -ve=contra *)
(* rB0=4; (* rB=rBt*rBt t *)
rBt=0.0; (* rB=rB0*rBt t *)*)
gR=0.5; (* ve=copy, -ve=contra *)
gB=-2.3; (* ve=copy, -ve=contra *)
Rfin=0.9;
Bfin=1;
Gfin=1.1;
Rstar:=(aR Rfin+rB Bfin)/(aR + rB)
Bstar := Bfin
Gstar := (aG Gfin+gB Bfin)/(aG+gR+gB)+(gR(aR Rfin +rB Bfin))/((aR+rB)(aG+gR+gB))
NDSolve[{
R'[t]==aR*(Rfin-R[t])+rB*(B[t]-R[t]),
B'[t]==aB*(Bfin-B[t]),
G'[t]==aG*(Gfin-G[t])+gR*(R[t]-G[t])+gB*(B[t]-G[t]),
R[0]==1.1,
B[0]==1.0,
G[0]==0.9
},{R,B,G},{t,0,100}];
p2=Plot[Evaluate[{R[t],B[t],G[t],Rstar,Bstar,Gstar}/.%],{t,0,7},PlotStyle->col,PlotRange->{0.89,1.55},PlotPoints->100,
(*FrameLabel\[Rule]{Style["future time",Black,Bold,30],""},*)
Frame->{True,True,False,False},AspectRatio->0.75,FrameTicks->{{Automatic,None},{None,None}},
Epilog->{Inset[Framed[Style["R(t)",Bold,Italic,30,colors[[1,1]]],Background->White,FrameStyle->White],{6,0.935}],Inset[Framed[Style["B(t)",Bold,Italic,30,colors[[2,1]]],Background->White,FrameStyle->White],{6,1}],Inset[Framed[Style[Rotate["G(t)",77 Degree],Bold,Italic,30,colors[[3,1]]],Background->White,FrameStyle->White,FrameMargins->-5,ContentPadding->False],{6,1.5}]},
ImageSize->Large
]
```

1 -ve

$a_R+r_B=3.0$; $a_G+g_R+g_B=-0.8$

$g_R>0$ and $g_B<0$

$\lambda_1, \lambda_2<0$ (stable); $\lambda_3>0$ (unstable)

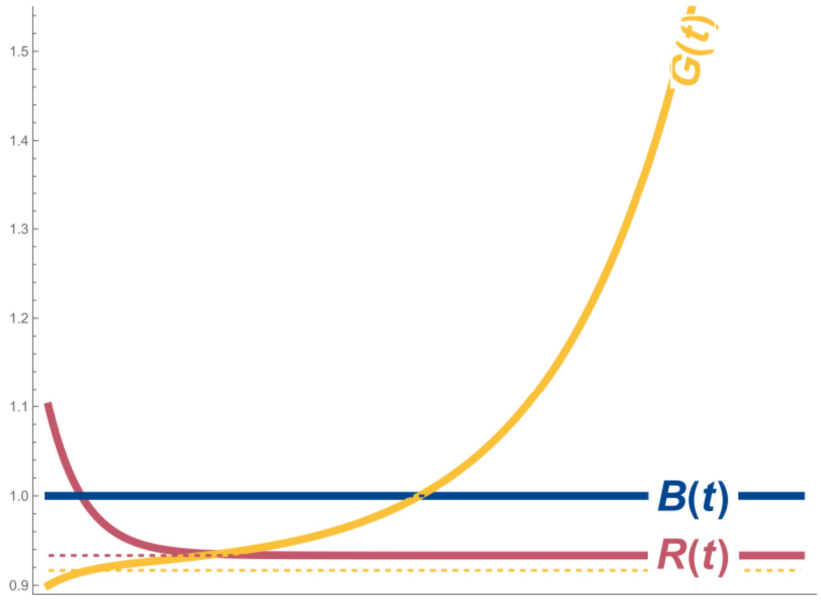

Figure S38: Fig. 5(b) in the main text.

```
1.  ClearSystemCache[];
2.  aR=2.8; (* Growth rate R *)
3.  aB=0.5; (* Growth rate B *)
4.  aG=1.0; (* Growth rate G *)
5.  rB=-1; (* ve=copy, -ve=contra *)
6.  (* rB0=4; (* rB=rBt*rBt t *)
7.  rBt=0.0; (* rB=rB0*rBt t *)*)
8.  gR=-0.5; (* ve=copy, -ve=contra *)
9.  gB=-2.0; (* ve=copy, -ve=contra *)
10. Rfin=0.9;
11. Bfin=1;
12. Gfin=1.1;
13. Rstar:=(aR Rfin+rB Bfin)/(aR + rB)
14. Bstar := Bfin
15. Gstar := (aG Gfin+gB Bfin)/(aG+gR+gB)+(gR(aR Rfin +rB Bfin))/((aR+rB)(aG+gR+gB))
16. NDSolve[{
17. R'[t]==aR*(Rfin-R[t])+rB*(B[t]-R[t]),
18. B'[t]==aB*(Bfin-B[t]),
19. G'[t]==aG*(Gfin-G[t])+gR*(R[t]-G[t])+gB*(B[t]-G[t]),
20. R[0]==1.1,
21. B[0]==1.0,
22. G[0]==0.9
23. },{R,B,G},{t,0,100}];
24. p3=Plot[Evaluate[{R[t],B[t],G[t],Rstar,Bstar,Gstar}/.%],{t,0,2.5},PlotStyle->col,PlotRange-
    >{0,1.15},PlotPoints->100,
25. (*FrameLabel\[Rule]{Style["future time",Black,Bold,30],""},*)
26. Frame->{True,True,False,False},AspectRatio->0.75,FrameTicks->{{Automatic,None},{None,None}},
27. Epilog->{Inset[Framed[Style[Rotate["R(t)",-
    5Degree],Bold,Italic,30,colors[[1,1]]],Background->White,FrameStyle-
    >White],{2.15,0.84}],Inset[Framed[Style["B(t)",Bold,Italic,30,colors[[2,1]]],Background-
    >White,FrameStyle->White],{2.15,1}],Inset[Framed[Style[Rotate["G(t)",-53
    Degree],Bold,Italic,30,colors[[3,1]]],Background->White,FrameStyle->White,FrameMargins-
    >0,ContentPadding->False],{2.3,0.22}]},
28. ImageSize->Large
29. ]
```

3 -ve

$a_R+r_B=1.0$; $a_G+g_R+g_B=-1.5$

$g_R<0$ and $g_B<0$

$\lambda_1, \lambda_2<0$ (stable); $\lambda_3>0$ (unstable)

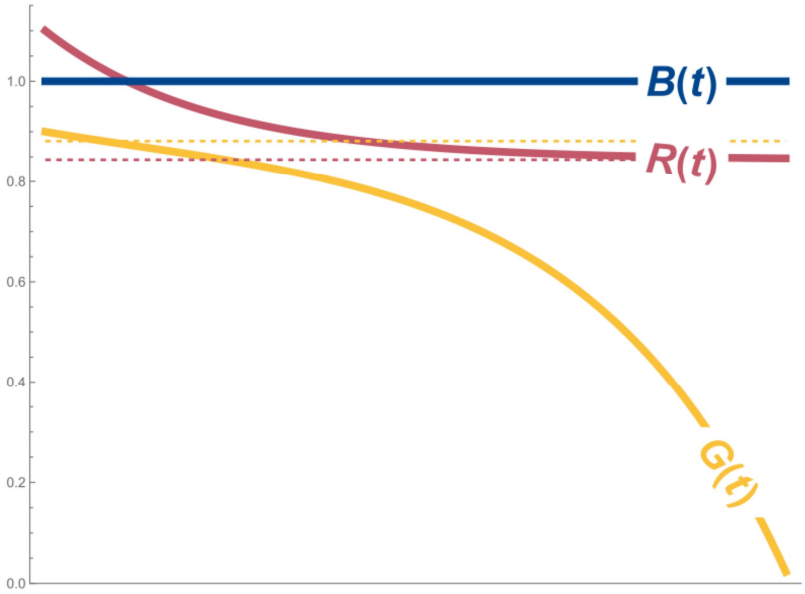

Figure S39: Fig. 5(c) in the main text.

```
1.  aR=2.0; (* Growth rate R *)
2.  aB=0.5; (* Growth rate B *)
3.  aG=1.0; (* Growth rate G *)
4.  rB=-2.5; (* ve=copy, -ve=contra *)
5.  gR=0.5; (* ve=copy, -ve=contra *)
6.  gB=-0.5; (* ve=copy, -ve=contra *)
7.  Rfin=0.9; Bfin=1; Gfin=1.1;
8.  Rstar:=(aR Rfin+rB Bfin)/(aR + rB)
9.  Bstar := Bfin
10. Gstar := (aG Gfin+gB Bfin)/(aG+gR+gB)+(gR(aR Rfin +rB Bfin))/((aR+rB)(aG+gR+gB))
11. NDSolve[{
12. R'[t]==aR*(Rfin-R[t])+rB*(B[t]-R[t]),
13. B'[t]==aB*(Bfin-B[t]),
14. G'[t]==aG*(Gfin-G[t])+gR*(R[t]-G[t])+gB*(B[t]-G[t]),
15. R[0]==1.1, B[0]==1.0, G[0]==0.9},{R,B,G},{t,0,100}];
16. p4=Plot[Evaluate[{R[t],B[t],G[t],Rstar,Bstar,Gstar}/.%],{t,0,3.5},PlotStyle->col,PlotRange-
    >{0,1.5},PlotPoints->100,Frame->{True,True,False,False},AspectRatio->0.75,FrameTicks-
    >{{Automatic,None},{None,None}}, Epilog->{Inset[Framed[Style[Rotate["R(t)",-
    50Degree],Bold,Italic,30,colors[[1,1]]],Background->White,FrameStyle->White,FrameMargins->-
    6,ContentPadding-
    >False],{3,0.075}],Inset[Framed[Style["B(t)",Bold,Italic,30,colors[[2,1]]],Background-
    >White,FrameStyle->White],{3,1}],Inset[Framed[Style[Rotate["G(t)",-30
    Degree],Bold,Italic,30,colors[[3,1]]],Background->White,FrameStyle->White,FrameMargins->-
    3,ContentPadding->False],{3,0.83}]},ImageSize->Large]
```

2 -ve

$a_R+r_B=-0.5$; $a_G+g_R+g_B=1.0$

$g_R>0$ and $g_B<0$

$\lambda_1, \lambda_2<0$ (stable); $\lambda_3>0$ (unstable)

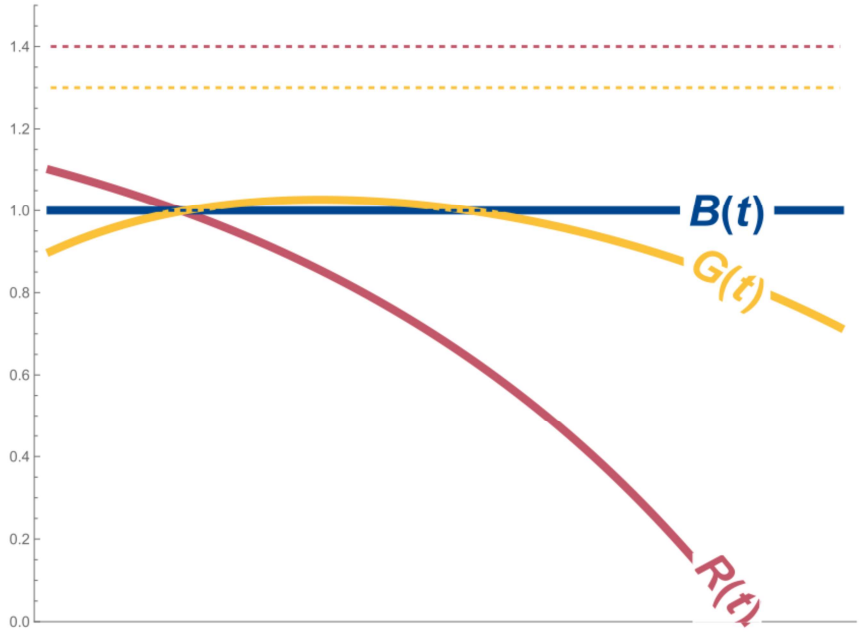

Figure S40: Fig. 5(d) in the main text. The next figure contains the underlying trajectories and fixed point behavior that explain these 4 panels of output curves.

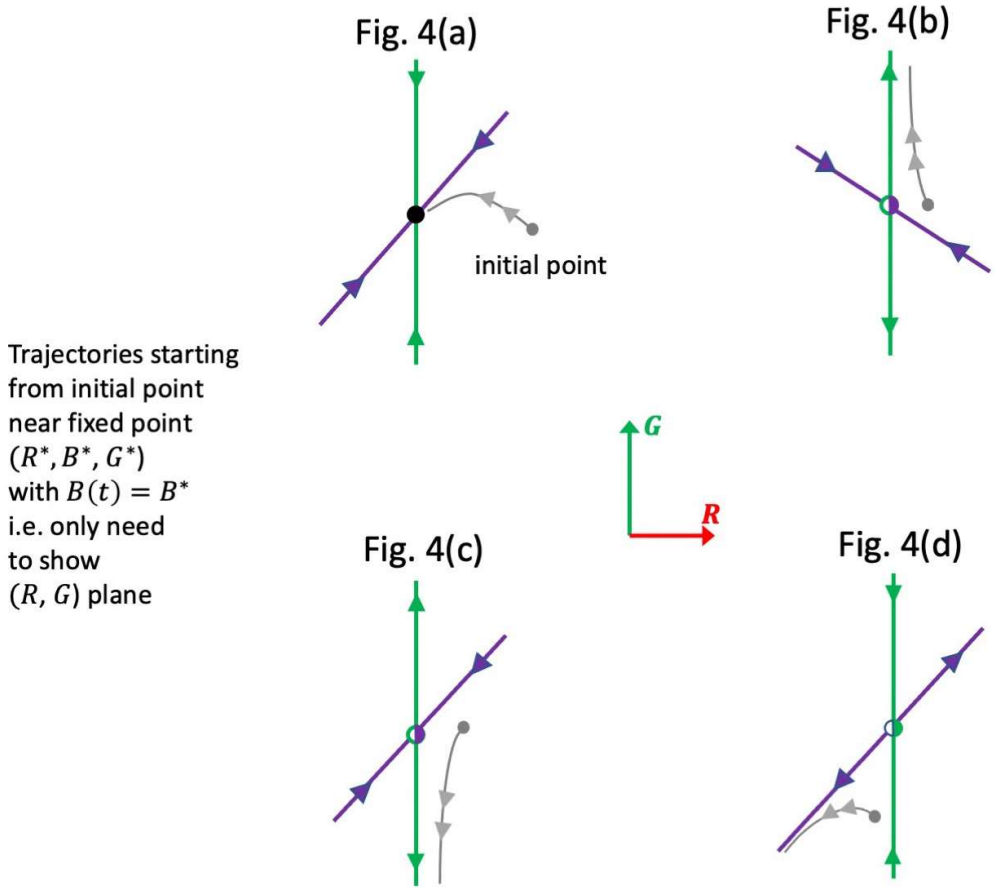

Figure S41: Underlying trajectories and fixed point behavior.

Replication programs and data

This set of files, replication programs and data, is provided online.

REFERENCES AND NOTES

1. R. Brown, Counteracting dangerous narratives in the time of covid-19, (Over Zero, 2022); www.projectoverzero.org/media-and-publications/counteracting-dangerous-narratives.

2. J. Zarocostas, How to fight an infodemic. *Lancet* **395**, 676 (2020).

3. H. J. Larson, Blocking information on COVID-19 can fuel the spread of misinformation. *Nature* **580**, 306–306 (2020).

4. S. van der Linden, A. Leiserowitz, S. Rosenthal, E. Maibach, Inoculating the public against misinformation about climate change. *Glob. Chall.* **1**, 1600008 (2017).

5. S. Gavrilets, Collective action and the collaborative brain. *J. R. Soc. Interface* **12**, 20141067 (2015).

6. M. J. Gelfand, J. R. Harrington, J. C. Jackson, The strength of social norms across human groups. *Perspect. Psychol. Sci.* **12**, 800–809 (2017).

7. B. Nogrady, 'I hope you die': How the covid pandemic unleashed attacks on scientists. *Nature* **598**, 250–253 (2021).

8. Trust science pledge calls for public to engage in scientific literacy, (News Direct, 2021); https://newsdirect.com/news/trust-science-pledge-calls-for-public-to-engage-in-scientific-literacy-737528151.

9. F. Slakey, American Physical Society (APS) message to members, (American Physical Society, 2021); https://aps.org/about/support/index.cfm.

10. P. Ball, Hot air by Peter Stott Review—The battle against climate change denial, (The Guardian, 2021); www.theguardian.com/books/2021/oct/09/hot-air-by-peter-stott-review-the-battle-against-climate-change-denial.

11. P. Stott, *Hot Air: The Inside story of the Battle Against Climate Change Denial* (Atlantic Books, 2021).

12. "guidance". Oxford Dictionaries. Oxford University Press. https://premium.oxforddictionaries.com/us/definition/american_english/guidance (accessed via Oxford Dictionaries Online on September 19, 2022).

13. S. Kemp, Hootsuite 2021, (Hootsuite, 2021); www.hootsuite.com/resources/digital-trends.

14. B. Hope, M. Ozdaglar, Parent assaults teacher over mask dispute at Amador County School, Superintendent says, (KCRA, 2021); www.kcra.com/article/sutter-creek-parent-assaults-teacher-over-mask-dispute-at-amador-county-school-superintendent/37295267?et_rid=934107578&s_campaign=fastforward%3Anewsletter.

15. Pfizer and BioNTech achieve first authorization in the world for a vaccine to combat COVID-19, (Pfizer, 2020); www.pfizer.com/news/press-release/press-release-detail/pfizer-and-biontech-achieve-first-authorization-world.

16. N. Calleja, A. AbdAllah, N. Abad, N. Ahmed, D. Albarracin, E. Altieri, J. N. Anoko, R. Arcos, A. A. Azlan, J. Bayer, A. Bechmann, S. Bezbaruah, S.C. Briand, I. Brooks, L. M. Bucci, S. Burzo, C. Czerniak, M. De Domenico, A. G. Dunn, U. K. H. Ecker, L. Espinosa, C. Francois, K. Gradon, A. Gruzd, B. S. Gülgün, R. Haydarov, C. Hurley, S. I. Astuti, A. Ishizumi, N. Johnson, D. Johnson Restrepo, M. Kajimoto, A. Koyuncu, S. Kulkarni, J. Lamichhane, R. Lewis, A. Mahajan, A. Mandil, E. McAweeney, M. Messer, W. Moy, P. Ndumbi Ngamala, T. Nguyen, M. Nunn, S. B. Omer, C. Pagliari, P. Patel, L. Phuong, D. Prybylski, A. Rashidian, E. Rempel, S. Rubinelli, P. Sacco, A. Schneider, K. Shu, M. Smith, H. Sufehmi, V. Tangcharoensathien, R. Terry, N. Thacker, T. Trewinnard, S. Turner, H. Tworek, S. Uakkas, E. Vraga, C. Wardle, H. Wasserman, E. Wilhelm, A. Würz, B. Yau, L. Zhou, T. D. Purnat, A public health research agenda for managing infodemics: Methods and results of the first who infodemiology conference. *JMIR Infodemiology* **1**, e30979 (2021).

17. P. S. Dodds, K. D. Harris, I. M. Kloumann, C. A. Bliss, C. M. Danforth, Temporal patterns of happiness and information in a global social network: Hedonometrics and Twitter. *PLOS ONE* **6**, e26752 (2011).

18. J.P. Onnela, J. Saramäki, J. Hyvönen, G. Szabó, D. Lazer, K. Kaski, J. Kertész, A.L. Barabási, Structure and tie strengths in mobile communication networks. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 7332–7336 (2007).

19. A. Bessi, M. Coletto, G.A. Davidescu, A. Scala, G. Caldarelli, W. Quattrociocchi, Science vs conspiracy: Collective narratives in the age of misinformation. *PLOS ONE* **10**, e0118093 (2015).

20. D. Lazer, M. A. Baum, Y. Benkler, A. J. Berinsky, K. M. Greenhill, F. Menczer, M. J. Metzger, B. Nyhan, G. Pennycook, D. Rothschild, M. Schudson, S. A. Sloman, C.R. Sunstein, E. A. Thorson, D. J. Watts, J. L. Zittrain, The science of fake news. *Science* **359**, 1094–1096 (2018).

21. E. Chen, K. Lerman, E. Ferrara, Tracking social media discourse about the COVID-19 pandemic: Development of a public coronavirus Twitter data set. *JMIR Public Health Surveill.* **6**, e19273 (2020).

22. S. Havlin, D. Y. Kenett, A. Bashan, J. Gao, H. E. Stanley, Vulnerability of network of networks. *Eur. Phys. J. Spec. Top.* **223**, 2087–2106 (2014).

23. B.L. Hoffman, E.M. Felter, K.H. Chu, A. Shensa, C. Hermann, T. Wolynn, D. Williams, B.A. Primack, It's not all about autism: The emerging landscape of anti-vaccination sentiment on Facebook. *Vaccine* **37**, 2216–2223 (2019).

24. N. Smith, T. Graham, Mapping the anti-vaccination movement on Facebook. *Inf. Commun. Soc.* **22**, 1310–1327 (2019).

25. K. Starbird, E. S. Spiro, K. Koltai, Misinformation, crisis, and public health—Reviewing the literature. *Soc. Sci. Res. Council* 10.35650/md.2063.d.2020 (2020).

26. C. Miller-Idriss, *Hate in the Homeland: The New Global Far Right* (Princeton Univ. Press, 2020).

27. R. DiResta, Of virality and viruses: The ANTI-VACCINE movement and Social Media: Nautilus Institute for Security and Sustainability, *Nautilus Institute for Security and Sustainability | We hold that it is possible to build peace, create security, and restore sustainability for all people in our time* (2018); https://nautilus.org/napsnet/napsnet-special-reports/of-virality-and-viruses-the-anti-vaccine-movement-and-social-media/.

28. K. Starbird, Disinformation's spread: Bots, trolls and all of us. *Nature* **571**, 449–449 (2019).

29. A. Bechmann, Tackling disinformation and Infodemics demands media policy changes. *Digit. Journal.* **8**, 855–863 (2020).

30. Y. Theocharis, A. Cardenal, S. Jin, T. Aalberg, D.N. Hopmann, J. Strömbäck, L. Castro, F. Esser, P. Van Aelst, C. de Vreese, N. Corbu, K. Koc-Michalska, J. Matthes, C. Schemer, T. Sheafer, S. Splendore, J. Stanyer, A. Stępińska, V. Štětka, Does the platform matter? Social media and COVID-19 conspiracy theory beliefs in 17 countries. *N. Med. Soc.* 146144482110456 (2021).

31. A. M. Guess, B. Nyhan, J. Reifler, Exposure to untrustworthy websites in the 2016 US election. *Nat. Hum. Behav.* **4**, 472–480 (2020).

32. J. D. West, C. T. Bergstrom, Misinformation in and about science. *Proc. Natl. Acad. Sci. U.S.A.* **118**, e1912444117 (2021).

33. D. A. Salmon, M. Z. Dudley, J. M. Glanz, S. B. Omer, Vaccine hesitancy. *Am. J. Prev. Med.* **49**, S391–S398 (2015).

34. S. Vosoughi, D. Roy, S. Aral, The spread of true and false news online. *Science* **359**, 1146–1151 (2018).

35. G. Pennycook, D. G. Rand, Fighting misinformation on social media using crowdsourced judgments of news source quality. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 2521–2526 (2019).

36. A. Gruzd, P. Mai, Inoculating against an Infodemic: A Canada-wide covid-19 news, social media, and misinformation survey. *SSRN Electron. J.* 10.2139/ssrn.3597462 (2020).

37. S. Lewandowsky, J. Cook, U. Ecker, D. Albarracin, M. Amazeen, P. Kendou, D. Lombardi, E. Newman, G. Pennycook, E. Porter, D. Rand, D. Rapp, J. Reifler, J. Roozenbeek, P. Schmid, C. Seifert, G. Sinatra, B. Swire-Thompson, S. van der Linden, E. Vraga, T. Wood, M. Zaragoza, *The Debunking Handbook 2020* (Boston Univ. Libraries OpenBU, 2020); https://open.bu.edu/handle/2144/43031.

38. J. Donovan, Social-media companies must flatten the curve of misinformation. *Nature* 10.1038/d41586-020-01107-z (2020).

39. T. Burki, Vaccine misinformation and social media. *Lancet Digital Health* **1**, e258–e259 (2019).

40. R. Smith, C. Wardle, S. Cubbon, Under the surface: COVID-19 vaccine narratives, misinformation and data deficits on social media, (First Draft, 2020); https://firstdraftnews.org/long-form-article/under-the-surface-covid-19-vaccine-narratives-misinformation-and-data-deficits-on-social-media/.

41. A. M. Guess, B. A. Lyons, Misinformation, disinformation, and online propaganda, in *Social Media and Democracy* (Cambridge Univ. Press, 2020), pp. 10–33.

42. H. Allcott, M. Gentzkow, C. Yu, Trends in the diffusion of misinformation on social media. *Res. Polit.* **6**, 205316801984855 (2019).

43. D. Allington, B. Duffy, S. Wessely, N. Dhavan, J. Rubin, Health-protective behaviour, social media usage and conspiracy belief during the COVID-19 public health emergency. *Psychol. Med.* **51**, 1763–1769 (2021).

44. E. Bakshy, S. Messing, L. A. Adamic, Exposure to ideologically diverse news and opinion on Facebook. *Science* **348**, 1130–1132 (2015).

45. A. J. Berinsky, Rumors and health care reform: Experiments in political misinformation. *Brit. J. Polit. Sci.* **47**, 241–262 (2017).

46. M. Cinelli, W. Quattrociocchi, A. Galeazzi, C.M. Valensise, E. Brugnoli, A.L. Schmidt, P. Zola, F. Zollo, A. Scala, The COVID-19 social media infodemic. *Sci. Rep.* **10**, 16598 (2020).

47. M. Del Vicario, A. Bessi, F. Zollo, F. Petroni, A. Scala, G. Caldarelli, H.E. Stanley, W. Quattrociocchi, The spreading of misinformation online. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 554–559 (2016).

48. D. Freeman, F. Waite, L. Rosebrock, A. Petit, C. Causier, A. East, L. Jenner, A.L. Teale, L. Carr, S. Mulhall, E. Bold, S. Lambe, Coronavirus conspiracy beliefs, mistrust, and compliance with government guidelines in England. *Psychol. Med.* **52**, 251–263 (2020).

49. S. González-Bailón, N. Wang, Networked discontent: The anatomy of protest campaigns in social media. *Soc. Networks* **44**, 95–104 (2016).

50. H. O.-Y. Li, A. Bailey, D. Huynh, J. Chan, YouTube as a source of information on COVID-19: A pandemic of misinformation? *BMJ Glob. Health* **5**, e002604 (2020).

51. J. E. Oliver, T. J. Wood, Conspiracy theories and the paranoid style(s) of mass opinion. *Am. J. Polit. Sci.* **58**, 952–966 (2014).

52. O. Papakyriakopoulos, J. C. Medina Serrano, S. Hegelich, *The Spread of COVID-19 Conspiracy Theories on Social Media and the Effect of Content Moderation* (Harvard Kennedy School Misinformation Review, 2020).

53. D. Romer, K. H. Jamieson, Conspiracy theories as barriers to controlling the spread of covid-19 in the U.S. *Soc. Sci. Med.* **263**, 113356 (2020).

54. J. Strömbäck, Y. Tsfati, H. Boomgaarden, A. Damstra, E. Lindgren, R. Vliegenthart, T. Lindholm, News Media Trust and its impact on media use: Toward a framework for future research. *Ann. Int. Commun. Assoc.* **44**, 139–156 (2020).

55. J. R. Zaller, *The Nature and Origins of Mass Opinion* (Cambridge Univ. Press, 2011).

56. V. Cosenza, *World Map of Social Networks* (VincosBlog, 2022); https://vincos.it/world-map-of-social-networks/.

57. T. Ammari, S. Schoenebeck, Thanks for your interest in our Facebook group, but it's only for dads, in *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing* (2016).

58. R. Y. Moon, A. Mathews, R. Oden, R. Carlin, Mothers' perceptions of the internet and social media as sources of parenting and Health Information: Qualitative study. *J. Med. Internet Res.* **21**, e14289 (2019).

59. R. Laws, A.D. Walsh, K.D. Hesketh, K.L. Downing, K. Kuswara, K.J. Campbell, Differences between mothers and fathers of young children in their use of the internet to support healthy family lifestyle behaviors: Cross-sectional study. *J. Med. Internet Res.* **21**, e11454 (2019).

60. D. Centola, J. Becker, D. Brackbill, A. Baronchelli, Experimental evidence for tipping points in social convention. *Science* **360**, 1116–1119 (2018).

61. N.F. Johnson, N. Velásquez, N.J. Restrepo, R. Leahy, N. Gabriel, S. El Oud, M. Zheng, P. Manrique, S. Wuchty, Y. Lupu, The online competition between pro- and anti-vaccination views. *Nature* **582**, 230–233 (2020).

62. What is the Page Transparency Section on Facebook Pages? (Facebook Help Center); www.facebook.com/help/323314944866264.

63. L. Gustafsson, M. Sternad, When can a deterministic model of a population system reveal what will happen on average? *Math. Biosci.* **243**, 28–45 (2013).

64. M. Chen, Y. Cao, L. T. Watson, Parameter estimation of stochastic models based on limited data. *ACM SIGBioinformatics Record* **7**, 1–3 (2018).

65. K. J. McHugh, L. Jing, S. Y. Severt, M. Cruz, M. Sarmadi, H. S. N. Jayawardena, C. F. Perkinson, F. Larusson, S. Rose, S. Tomasic, T. Graf, S. Y. Tzeng, J. L. Sugarman, D. Vlasic, M. Peters, N. Peterson, L. Wood, W. Tang, J. Yeom, J. Collins, P. A. Welkhoff, A. Karchin, M. Tse, M. Gao, M. G. Bawendi, R. Langer, A. Jaklenec, Biocompatible near-infrared quantum dots delivered to the skin by microneedle patches record vaccination. *Sci. Transl. Med.* **11**, eaay7162 (2019).

66. Weintraub, Invisible ink could reveal whether kids have been vaccinated, (Scientific American, 2019); www.scientificamerican.com/article/invisible-ink-could-reveal-whether-kids-have-been-vaccinated/.

67. M. Roser, H. Ritchie, E. Ortiz-Ospina, *Internet* (Our World in Data, 2015); https://ourworldindata.org/internet.

68. R. Sear, R. Leahy, N. Restrepo, Y. Lupu, N.F. Johnson, Machine learning language models: Achilles heel for social media platforms and a possible solution. *Adv. Artificial Intell. Mach. Learn.* **1**, 191–202 (2021).

69. N. F. Johnson, M. Zheng, Y. Vorobyeva, A. Gabriel, H. Qi, N. Valesquez, P. Manrique, D. Johnson, E. Restrepo, C. Song, S. Wuchty, New online ecology of adversarial aggregates: Isis and beyond. *Science* **352**, 1459–1463 (2016).

70. N. F. Johnson, R. Leahy, N. J. Restrepo, N. Valesquez, M. Zheng, P. Manrique, P. Devkota, S. Wuchty, Hidden resilience and adaptive dynamics of the Global Online Hate Ecology. *Nature* **573**, 261–265 (2019).