SCIENTIFIC REPORTS
natureresearch

OPEN

Recursive patterns in online echo chambers

Emanuele Brugnoli [1*], Matteo Cinelli[1], Walter Quattrociocchi[2] & Antonio Scala[1,3]

Despite their entertainment oriented purpose, social media changed the way users access information, debate, and form their opinions. Recent studies, indeed, showed that users online tend to promote their favored narratives and thus to form polarized groups around a common system of beliefs. Confirmation bias helps to account for users' decisions about whether to spread content, thus creating informational cascades within identifiable communities. At the same time, aggregation of favored information within those communities reinforces selective exposure and group polarization. Along this path, through a thorough quantitative analysis we approach connectivity patterns of 1.2 M Facebook users engaged with two very conflicting narratives: scientific and conspiracy news. Analyzing such data, we quantitatively investigate the effect of two mechanisms (namely challenge avoidance and reinforcement seeking) behind confirmation bias, one of the major drivers of human behavior in social media. We find that challenge avoidance mechanism triggers the emergence of two distinct and polarized groups of users (i.e., echo chambers) who also tend to be surrounded by friends having similar systems of beliefs. Through a network based approach, we show how the reinforcement seeking mechanism limits the influence of neighbors and primarily drives the selection and diffusion of contents even among like-minded users, thus fostering the formation of highly polarized sub-clusters within the same echo chamber. Finally, we show that polarized users reinforce their preexisting beliefs by leveraging the activity of their like-minded neighbors, and this trend grows with the user engagement suggesting how peer influence acts as a support for reinforcement seeking.

Social media facilitated global communications all over the world, allowing information to spread faster and intensively. These changes led up to the formation of a disintermediated scenario, where contents flow directly from producers to consumers, without the mediation of journalists or experts in the field. Beyond its undoubted benefits, a hyper-connected world can foster confusion about causation, and thus encourage speculation, rumors, and mistrust[1–4]. Since 2013, indeed, the World Economic Forum (WEF) has been placing the global threat of massive digital misinformation at the core of other technological and geopolitical risks, ranging from terrorism, to cyber-attacks, up to the failure of global governance[5]. People are misinformed when they hold beliefs neglecting factual evidence, and misinformation may influence public opinion negatively. Empirical investigations have shown that, in general, people tend to resist facts, holding inaccurate factual beliefs confidently[6]. Moreover, corrections frequently fail to reduce misperceptions[7] and often act as a backfire effect[8].

Confirmation bias - i.e., the tendency to seek, select, and interpret information coherently with one's system of beliefs[9] - helps, indeed, to account for users' decisions about whether to promote content[2,10–12]. The action of this cognitive bias may lead to the emergence of homogeneous and polarized communities - i.e., echo-chambers[13–15], thus facilitating fake news and, more in general, misinformation cascades[3].

As previously described[16], two primary cognitive mechanisms are used to explain why people experience the confirmation bias[17]:

- Challenge avoidance - i.e., the fact that people do not want to find out that they are wrong,
- Reinforcement seeking - i.e., the fact that people want to find out that they are right.

As has been previously stated[16], though the two are strongly related, and though both behaviors revolve around people's attempt to minimize their cognitive dissonance - i.e., the psychological stress that people experience when they hold two or more contradictory beliefs simultaneously, challenge avoidance and reinforcement seeking are not inherently linked to each other, and they do not have to occur at the same time[18]. This distinction is important because the consequences of challenge avoidance are significantly more harmful to democratic

[1]Applico Lab, CNR-ISC, Rome, 00185, Italy. [2]Ca' Foscari University of Venice, Venice, 30123, Italy. [3]LIMS, the London Institute for Mathematical Sciences, London, United Kingdom. *e-mail: emanuele.brugnoli@roma1.infn.it

deliberation than those of reinforcement seekings[17]. Additionally, group membership has an interplay with the aforementioned cognitive biases. When individuals belong to a certain group, those outside the group are far less likely to influence them on both easy and hard questions[19].

In this work, by exploiting the social network of 1.2 M Facebook users engaged with very polarizing contents, we investigate the role of challenge avoidance and reinforcement seeking on the selection and spread of information, and the connection of such cognitive mechanisms with peer influence.

To our aim, with the help of very active debunking groups, we identified all the Italian Facebook pages supporting scientific and conspiracy news, and on a time span of five years (2010–2014) we downloaded all their public posts (with the related lists of likes and comments). On the one hand, conspiracy news simplify causation, reduce the complexity of reality, and are formulated in a way that is able to tolerate a certain level of uncertainty[20–22]. On the other hand, scientific news disseminates scientific advances and exhibits the process of scientific thinking. Notice that we do not focus on the quality of the information but rather on the possibility of verification. Indeed, the main difference between the two is content verifiability. The generators of scientific information and their data, methods, and outcomes are readily identifiable and available. The origins of conspiracy theories are often unknown and their content is strongly disengaged from mainstream society and sharply divergent from recommended practices[8], e.g., the belief that vaccines cause autism[23].

Our analyses show how challenge avoidance mechanism triggers the emergence, around the selected narratives, of two well-separated and polarized groups of users who also tend to surround themselves with friends having similar systems of beliefs.

Through a network based approach, we also prove that polarized users span their attention focus on a higher number of pages (and topics) supporting their beliefs (hereafter referred to as *community pages*) as their engagement grows, but they tend to remain confined within groups of very few pages even when the corresponding neighborhoods are active on several news sources. This suggests that the reinforcement seeking mechanism limits the influence of neighbors and primarily drives the selection and the diffusion of contents even among like-minded users, fostering the formation of highly polarized subclusters within the same echo chamber.

Finally, we investigate the effects of the joint action of confirmation bias and peer influence when the latter does not conflict the cognitive mechanisms of challenge avoidance and reinforcement seeking. Namely, we compare the liking activity of polarized users and the liking activity of their part of neighborhood likewise polarized, both with respect to size and time. Our findings reveal that polarized users reinforce their preexisting beliefs by leveraging the activity of their like-minded neighbors. Such a trend grows with the user engagement and suggests how peer influence acts as a support for reinforcement seeking. In such a context, also the positive role played by social influence - e.g., by enabling social learning[24–26], seems to lose its effectiveness in the effort of smoothing polarization and reducing both the risk and the consequences of misinformation. This makes it even more difficult to design efficient communication strategies to prevent rumors and mistrust. Individual choices more than algorithms[10] seem to characterize the consumption patterns of users and their friends. Therefore, working towards long-term solutions to polarization and misinformation online cannot be separated from a deep understanding of users' cognitive determinants behind these mechanisms.

Methods

Ethics statement.   Approval and informed consent were not needed because the data collection process has been carried out using the Facebook Graph application program interface (API), which is publicly available. For the analysis (according to the specification settings of the API) we only used publicly available data (thus users with privacy restrictions are not included in the dataset). The pages from which we download data are public Facebook entities and can be accessed by anyone. User content contributing to these pages is also public unless the user's privacy settings specify otherwise, and in that case it is not available to us.

Data collection.   Debate about social issues continues to expand across the Web, and unprecedented social phenomena such as the massive recruitment of people around common interests, ideas, and political visions are emerging. For our analysis, we identified two main categories of pages: conspiracy news – i.e., pages promoting contents neglected by main stream media – and science news. We defined the space of our investigation with the support of diverse Facebook groups that are very active in debunking conspiracy theses. As an additional control, we used the self-description of a page to determine its focus. The resulting dataset is composed by all the pages supporting the two distinct narratives in the Italian Facebook scenario: 39 about conspiracy theories and 33 about science news. For the two sets of pages we download all of the posts (and their respective user interactions) across a 5-y time span (2010–2014). We perform the data collection process by using the Facebook Graph API, which is publicly available and accessible through any personal Facebook user account. The exact breakdown of the data is presented in Table 1. Likes and comments have a different meaning from the user viewpoint. Most of the times, a like stands for a positive feedback to the post and a comment is the way in which online collective debates take form. Comments may contain negative or positive feedbacks with respect to the post.

Ego networks.   In addition, we collected the ego networks of users who liked at least one post on science or conspiracy pages - i.e., for each user we have collected her list of friends and the links between them (We used publicly available data, so we collected only data for which the users had the corresponding permissions open).

Preliminaries and definitions.   Let $\mathscr{P}$ be the set of all the pages in our collection, and $\mathscr{P}_{science}$ ($\mathscr{P}_{conspir}$) be the set of the 33 (39) Facebook pages about science (conspiracy) news. Let V be the set of all the 1.2 M users and E the edges representing their Facebook friendship connections; these sets define a graph $G = (V, E)$. Hence, the graph of likes on a post, $G^L = (V^L, E^L)$ is the subgraph of G whose users have liked a post. Thus, V^L is the set of users of V who have liked at least one post, and we set $E^L = \{(u, v) \in E; u, v \in V^L\}$. Following previous works[2,3,27], we

	Total	Science	Conspiracy
Pages	72	33	39
Posts	270,629	62,038	208,591
Likes	9,164,781	2,505,399	6,659,382
Comments	1,017,509	180,918	836,591
Likers	1,196,404	332,357	864,047
Commenters	279,972	53,438	226,534

Table 1. Breakdown of Facebook dataset. The number of pages, posts, likes and comments for science and conspiracy pages.

study the polarization of users - i.e., the tendency of users to interact with only a single type of information; in particular, we study the polarization towards science and conspiracy. Formally we define the polarization $\rho(u) \in [-1, 1]$ of user $u \in V^L$ as the ratio of likes that u has performed on conspiracy posts: assuming that u has performed x and y likes on conspiracy and science posts, respectively, we let $\rho(u) = (x - y)/(x + y)$. Thus, a user u for whom $\rho(u) = -1$ is totally polarized towards science, whereas a user with $\rho(u) = 1$ is totally polarized towards conspiracy. Note that we ignore the commenting activity since a comment may be an endorsement, a criticism, or even a response to a previous comment. Furthermore, we define the engagement $\psi(u)$ of a polarized user u as her liking activity normalized with respect to the number of likes of the most active user of her community. By defining $\theta(u)$ as the total number of likes that the user u has expressed in posts of $\mathscr{P}$, notice that the following condition holds: $\psi(u) = \frac{\theta(u)}{\max_v \theta(v)}$.

The degree of a node (here, user) u, $deg(u)$, is the number of neighbors (here, friends) of u. For any user u, we consider the partition $deg(u) = |N_c(u)| + |N_{ne}(u)| + |N_{np}(u)| + |N_s(u)|$ where $N_c(u)(N_s(u))$ denotes the neighborhood of u polarized towards conspiracy (science), $N_{ne}(u)$ denotes the neighborhood of u not engaged with science or conspiracy contents, $N_{np}(u)$ denotes the set of not polarized friends of u - i.e., friends who liked the same number of contents from science and conspiracy, respectively.

To understand the relationship between pages and user liking activity, we measure the polarization of users with respect to the pages of their own community. For a polarized user (or, more in general, a group of polarized users) u with $\sum_i \theta_i(u) = \theta(u)$ likes, where $\theta_i(u)$ counts the contents liked by u on the ith community page ($i = 1, \dots, N$, where N equals the number of community pages), the probability $\phi_i(u)$ that u belongs to the ith page of the community will then be $\phi_i(u) = \theta_i(u)/\theta(u)$. We can define the localization order parameter L as:

$$L(u) = \frac{(\sum_i \phi_i^2(u))^2}{\sum_i \phi_i^4(u)}$$

Thus, in the case in which u only has likes in one page, $L(u) = 1$. If u, on the other hand, interacts equally with all the community pages ($\phi_i(u) = 1/N$) then $L(u) = N$; hence, $L(u)$ counts the community pages where u fairly equally distributes her liking activity.

List of pages. In this section are listed pages of our dataset. Table 2 shows the list of scientific news and Table 3 shows the list of conspiracy pages.

Augmented Dickey–Fuller test. An augmented Dickey–Fuller test (ADF) tests the null hypothesis that a unit root is present in a time series[28,29]. The alternative hypothesis is stationarity. If we obtain a p-value less than the threshold value $\overline{\alpha} = 0.05$, the null hypothesis is rejected in favor of the alternative one. ADF is an augmented version of the Dickey–Fuller test[30] for a larger set of time series models. We use this test to investigate the stationarity of the time series given by the number of posts per day published by a community page during its lifetime. The general regression equation which incorporates a constant and a linear trend is used. The number of lags used in the regression corresponds to the upper bound on the rate at which the number of lags should be made to grow with the time series size T for the general ARMA(p, q) setup[31], and equals $T^{1/3}$.

Cosine similarity. Cosine similarity is a measure of similarity between two non-zero vectors $\mathbf{u} = (u_1, \dots, u_k)$ and $\mathbf{v} = (v_1, \dots, v_k)$ of a k-dimensional inner product space expressed by the cosine of the angle between them[32]. By means of the Euclidean dot product formula we obtain

$$\cos(\mathbf{u}, \mathbf{v}) = \frac{\mathbf{u} \cdot \mathbf{v}}{\|\mathbf{u}\| \|\mathbf{v}\|} = \frac{\sum_{i=1}^k u_i v_i}{\sqrt{\sum_{i=1}^k u_i^2} \sqrt{\sum_{i=1}^k v_i^2}}.$$

We use cosine similarity to evaluate whether a polarized user u and the part of her neighborhood with likewise polarization proportionally distribute their liking activity across her preferred community pages. Namely, for any user u polarized towards science (conspiracy), denoted with $\{P_{i_1}, \dots, P_{i_k}\} = \mathscr{P}^u_{\text{science}}$ ($\mathscr{P}^u_{\text{conspir}}$) the set of k science (conspiracy) pages where u distributes her liking activity, we compute the cosine between the vectors $(\theta_{i_1}(u), \dots, \theta_{i_k}(u))$ and $(\theta_{i_1}(N_s(u)), \dots, \theta_{i_k}(N_s(u)))$, both normalized with respect to the infinity norm. The space of

	Page name	Facebook ID
1	Scientificast.it	129133110517884
2	CICAP	32775139194
3	OggiScienza	106965734432
4	Query	128523133833337
5	Gravità Zero	138484279514358
6	COELUM Astronomia	81631306737
7	MedBunker	246240278737917
8	In Difesa della Sperimentazione Animale	365212740272738
9	Italia Unita per la Scienza	492924810790346
10	Scienza Live	227175397415634
11	La scienza come non l'avete mai vista	230542647135219
12	LIBERASCIENZA	301266998787
13	Scienze Naturali	134760945225
14	Perché vaccino	338627506257240
15	Le Scienze	146489812096483
16	Vera scienza	389493082245
17	Scienza in rete	84645527341
18	Galileo, giornale di scienza e problemi globali	94897729756
19	Scie Chimiche: Informazione Corretta	351626174626
20	Complottismo? No grazie	399888818975
21	INFN - Istituto Nazionale di Fisica Nucleare	45086217578
22	Signoraggio: informazione corretta	279217954594
23	Scetticamente	146529622080908
24	Vivisezione e Sperimentazione Animale, verità e menzogne	548684548518541
25	Medici Senza Frontiere	65737832194
26	Task Force Pandora	273189619499850
27	VaccinarSI	148150648573922
28	Lega Nerd	165086498710
29	Super Quark	47601641660
30	Curiosità Scientifiche	595492993822831
31	Minerva - Associazione di Divulgazione Scientifica	161460900714958
32	Pro-Test Italia	221292424664911
33	Uniti per la Ricerca	132734716745038

Table 2. Scientific news sources. List of Facebook pages diffusing main stream scientific news.

such versors is positive, then the cosine measure outcome is neatly bounded in $[0, \ 1]$: two versors are maximally similar if they are parallel and maximally dissimilar if they are orthogonal.

Akaike information criterion. The Akaike Information Criterion (AIC)[33–35] is an asymptotically unbiased estimator of the expected relative Kullback-Leibler distance (K-L)[36], which represents the amount of information lost when we use model *g* to approximate model *f*:

$$K - L = I(f, g) = \int f(x) \log \left(\frac{f(x)}{g(x|\mu)} \right) dx^2,$$

where $\mu = (\mu_1, \ldots, \mu_k)$ is the vector of *k* model parameters. The AIC for a given model is a function of its maximized log-likelihood (ℓ) and *k*:

$$\text{AIC} = -2\ell + 2k.$$

We use the AIC for selecting the optimal lag structure of a Granger causality test.

Granger causality and peer influence probability. The Granger causality test is a statistical hypothesis test for determining whether one time series is useful in forecasting another[37]. Roughly speaking, a time series *X* is said to Granger-cause (briefly, G-cause) the time series *Y* if the prediction of *Y* is improved when *X* is included in the prediction model of *Y*. Denoted with $\mathscr{I}^*(\tau)$ the set of all information in the universe up to time τ and with $\mathscr{I}^*_{-X}(\tau)$ the same information set except for the values of series *X* up to time τ, we write

$$Y_{\tau+1} \perp\!\!\!\perp \mathscr{I}^*(\tau) \big| \mathscr{I}^*_{-X}(\tau)$$

	Page name	Facebook ID
34	Scienza di Confine	188189217954979
35	CSSC - Cieli Senza Scie Chimiche	253520844711659
36	STOP ALLE SCIE CHIMICHE	199277020680
37	Vaccini Basta	233426770069342
38	Tanker Enemy	444154468988487
39	SCIE CHIMICHE	68091825232
40	MES Dittatore Europeo	194120424046954
41	Lo sai	126393880733870
42	AmbienteBio	109383485816534
43	Eco(R)esistenza	203737476337348
44	curarsialnaturale	159590407439801
45	La Resistenza	256612957830788
46	Radical Bio	124489267724876
47	Fuori da Matrix	123944574364433
48	Graviola Italia	130541730433071
49	Signoraggio.it	278440415537619
50	Informare Per Resistere	101748583911
51	Sul Nuovo Ordine Mondiale	340262489362734
52	Avvistamenti e Contatti	352513104826417
53	Umani in Divenire	195235103879949
54	Nikola Tesla - il SEGRETO	108255081924
55	Teletrasporto	100774912863
56	PNL e Ipnosi	150500394993159
57	HAARP - controllo climatico	117166361628599
58	Sezione Aurea, Studio di Energia Vibrazionale	113640815379825
59	PER UNA NUOVA MEDICINA	113933508706361
60	PSICOALIMENTARSI E CURARSI NATURALMENTE	119866258041409
61	La nostra ignoranza è la LORO forza.	520400687983468
62	HIV non causa AIDS	121365461259470
63	Sapere è un Dovere	444729718909881
64	V per Verità	223425924337104
65	Genitori veg	211328765641743
66	Operatori di luce	195636673927835
67	Coscienza Nuova	292747470828855
68	Aprite Gli Occhi	145389958854351
69	Neovitruvian	128660840526907
70	CoscienzaSveglia	158362357555710
71	Medicinenon	248246118546060
72	TERRA REAL TIME	208776375809817

Table 3. Conspiracy news sources. List of Facebook pages diffusing conspiracy news.

for indicating that X does not cause Y.

Let $t(u)$ be the time series given by the number of likes expressed by a user u polarized towards science on $\mathscr{P}^u_{\text{science}}$ every day of her lifetime - i.e., the temporal distance between its first and its last like. Let $t(N_s(u))$ be the time series of the number of likes expressed by $N_s(u)$ on the same pages every day in the same time window. We investigate a causal effect of $t(N_s(u))$ on $t(u)$ by testing the null hypothesis that the former does not Granger-cause the latter:

$$\mathbb{H}_0 := t(u)_{\tau+1} \perp\!\!\!\perp \mathscr{I}^*(\tau) | \mathscr{I}^*_{-t(N_s(u))}(\tau)$$

through a series of F-tests on lagged values of $t(u)$. The alternative hypothesis $\mathbb{H}_1$ is $t(N_s(u))$ G-cause $t(u)$. The number of lags to be included is chosen using AIC. If we obtain a p-value $\alpha(u)$ less than the threshold value $\overline{\alpha} = 0.05$, the null hypothesis $\mathbb{H}_0$ is rejected in favor of $\mathbb{H}_1$. The same analysis is carried out for testing a causal effect of $t(N_c(u))$ on $t(u)$ for any polarized user u towards conspiracy.

Furthermore, we define the peer influence probability $\text{PIP}^u_{\text{science}}$ of $N_s(u)$ on u as the rational number in the range $[0, 1]$ given by the complement of $\alpha(u)$ in the positive space of p-values, that is: $\text{PIP}^u_{\text{science}} = 1 - \alpha(u)$. Values close to 0 indicate low probability of peer influence, values close to 1 suggest high probability of peer
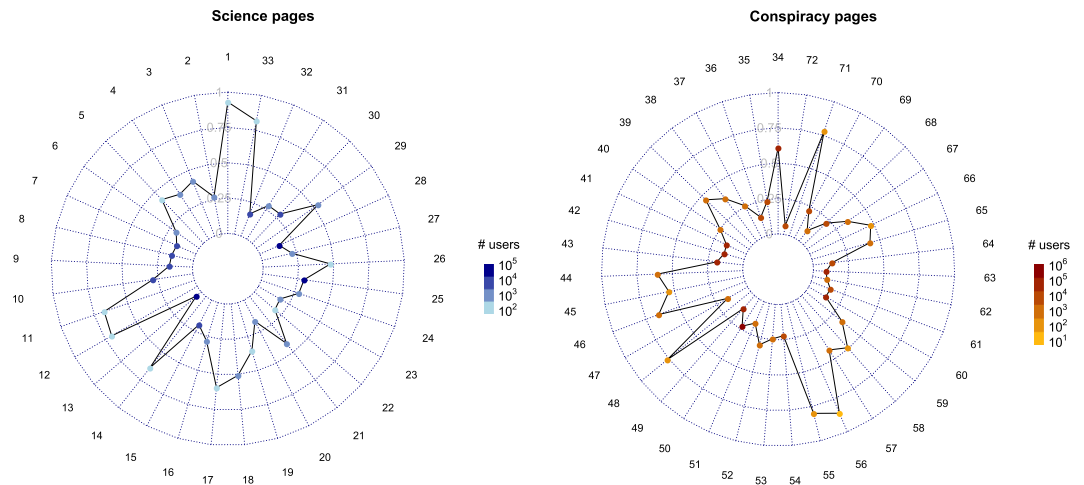
Figure 1. Distribution of likes and likers across the community pages. Plots shows the ratio likers/likes for any Science (left panel) and Conspiracy (right panel) page. Points are colored according to the number of users who liked contents published by the corresponding page (See Tables 2 and 3 for the list of scientific and conspiracy news sources, respectively).

influence. Analogously we define the peer influence probability $PIP^u_{conspir}$ of $N_c(u)$ on u, for any user u polarized towards conspiracy.

Dynamic time warping. Dynamic time warping (DTW) is an algorithm for measuring similarity between two time series X and Y which computes the optimal (least cumulative distance) alignment between points of X (also said query vector) and Y (also said reference vector). If X has size n and Y has size m, DTW produces an $n \times m$ cost matrix D whose (i, j)-element is the Euclidean distance $d(\overline{X}_i, \overline{Y}_j)$ where $\overline{X}_i$ and $\overline{Y}_j$ are obtained by stretching in time the vectors $(X[1], \ldots, X[i])$ and $(Y[1], \ldots, Y[j])$ to optimize the best alignment. The value $D(n, m)$ - i.e., the DTW distance between X and Y, is returned[38].

We use DTW distance for measuring the similarity between $t(u)$ and $t(N_s(u))$ $(t(N_c(u)))$ for any user u polarized towards science (conspiracy).

Results and Discussion

Anatomy of science and conspiracy pages. To ensure the robustness of our analysis about the online behavior of polarized users (i.e., if likes are not trivially distributed across pages and if data respect the assumptions of the tests described in *Methods*), we verify the eligibility of the space of our investigation. Namely we study how likers and their activity are distributed over pages and how pages' activity is distributed over time. Figure 1 shows the distribution of likes and likers across scientific and conspiracy news sources, respectively. Plots shows the ratio likers/likes for every science (left panel) and conspiracy (right panel) page. Points are colored according to the number of users who liked contents published by the corresponding page (See Tables 2 and 3 for the list of scientific and conspiracy news sources, respectively).

Points are mostly localized near the center of the radar chart and, in general, represent the pages with more likers (and more likes). Moreover, points far from the center correspond to pages with the lowest number of likers and likes. This ensures that a comparison between the normalized distributions of likes of two like-minded users (or groups of users) across the community pages is an unbiased estimator of their behavioral difference in terms of liking activity.

Furthermore, in order to investigate how scientific and conspiracy news sources distribute their posting activity over time, we compute the fraction of days with activity of any page with respect to its lifetime - i.e., the temporal distance between its first and its last post. Then we perform an augmented Dickey–Fuller (ADF) test for testing the null hypothesis that a unit root is present in the time series given by the number of posts per day published by a community page during its lifetime. The alternative hypothesis is stationarity (see *Methods* for further details). Figure 2 shows the PDF of the fraction of days with activity per page and the PDF of p-values obtained by performing ADF test for all the pages of science community (left panel) and all the pages of conspiracy community (right panel), respectively.

Plots indicate that the most pages of both the communities are active with a nearly constant number of posts almost every day of their lifetime.

Experiencing the confirmation bias: polarization and homophily. Users' liking activity across contents of the different categories[2,3,27] may be intended as the preferential attitude towards the one or the other type of information (documented or not). In Fig. 3 we show that the probability density function (PDF) for the polarization of all the users in V^L is a sharply peaked bimodal where the vast majority of users are polarized either towards science ($\rho(u) \sim -1$) or conspiracy ($\rho(u) \sim 1$). Hence, Fig. 3 shows that most of likers can be divided into
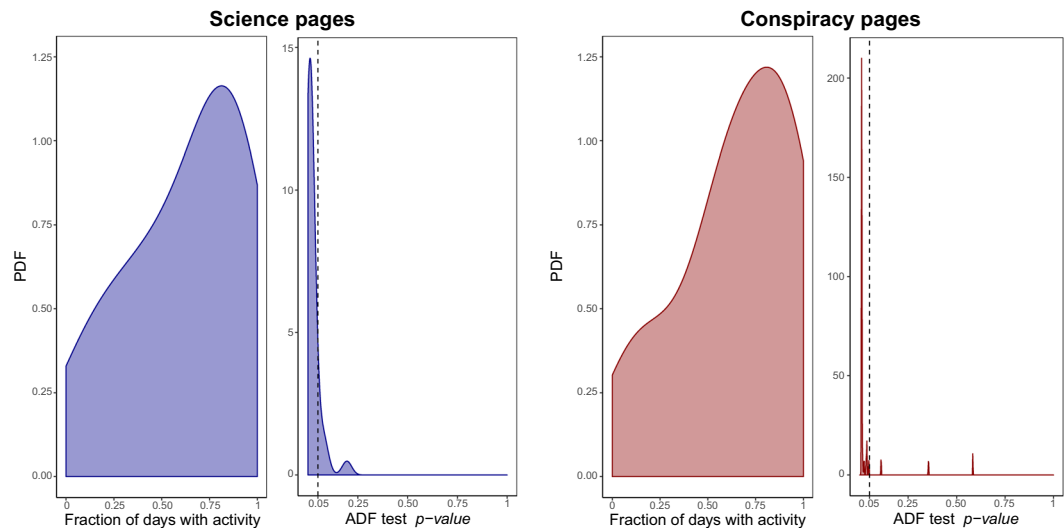
Figure 2. The most pages are active with a nearly constant number of posts almost every day of their lifetime, both in science and conspiracy community. Plots show the PDF of the fraction of days with activity per page and the PDF of p-values obtained by performing ADF stationarity test for all the pages, both in science (left panel) and conspiracy (right panel) community.
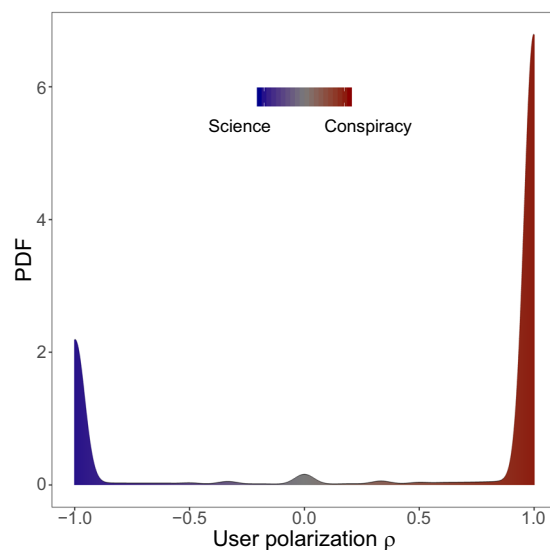

Figure 3. Polarization on contents. PDF of the frequency that a user has polarization ρ is remarkably concentrated in two peaks near the values $\rho = -1$ (science) and $\rho = 1$ (conspiracy). This indicates that users are clearly split into two distinct communities.

two groups of users, those polarized towards science and those polarized towards conspiracy. To better define the properties of these groups, we define the set V_{science}^L of users with polarization more than 95% towards science

$$V_{\text{science}}^L = \{u \in V^L; \ \rho(u) < -0.95\},$$

and the set V_{conspir}^L of users with polarization more than 95% towards conspiracy

$$V_{\text{conspir}}^L = \{u \in V^L; \ \rho(u) > 0.95\};$$

such sets corresponds to the two peaks of the bimodal distribution and show how the most users are highly polarized: $|V_{\text{science}}^L| = 243,977$ and $|V_{\text{conspir}}^L| = 758,673$.

Moreover, for a polarized users $u \in V_{\text{science}}^L$, in the left panel of Fig. 4, we show the log-linear plot of the average fraction of science pages where u is present with liking activity, respect given number of likes θ of the user u. In the right panel, we show the same quantities for polarized users in V_{conspir}^L. Figure 4 suggests in both cases a quadratic correlation
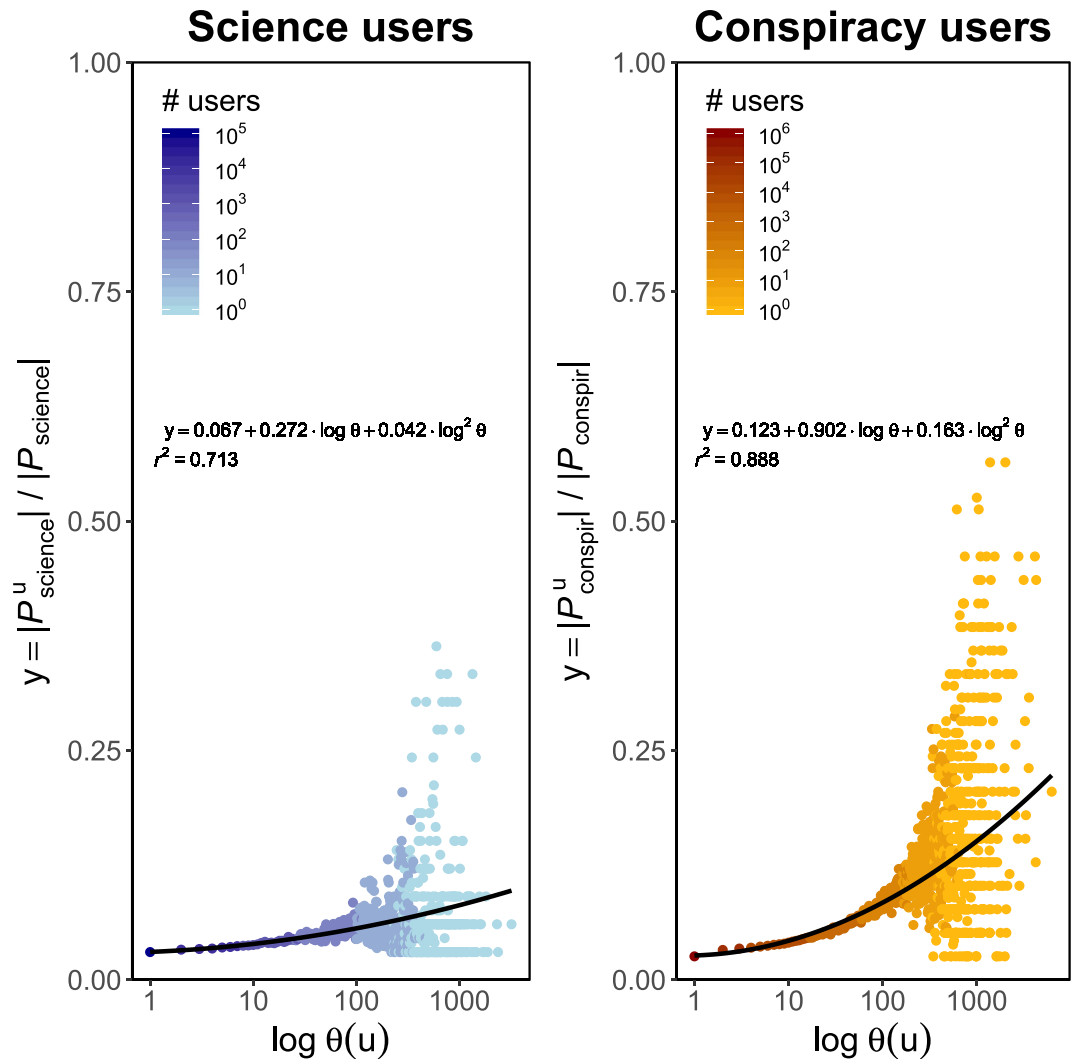
Figure 4. Polarized users span their attention focus on more news sources (and topics) as their engagement grows, but always keeping consistence with their way of thinking. Left panel: users polarized towards science. Right panel: users polarized towards conspiracy. In both panels we plot the average fraction of the total number of community pages where a polarized user u distributes her liking activity versus the number of likes $\log \theta(u)$ of u. Full lines are the results of a quadratic regression model $y = \beta_0 + \beta_1 \log \theta(u) + \beta_2 \log^2 \theta(u)$, where $y = |\mathscr{P}^u_{\text{science}}|/|\mathscr{P}_{\text{science}}|$ for $u \in V^L_{\text{science}}$ and $y = |\mathscr{P}^u_{\text{conspir}}|/|\mathscr{P}_{\text{conspir}}|$ for $u \in V^L_{\text{conspir}}$. Coefficients are estimated using weighted least squares with weights given by the total number of users per engagement value. In both cases, all the p-values are close to zero.

among the variables; thus, we check whether for a polarized user u, the fraction of community pages where u spans her liking activity, $y(u)$, can be predicted by means of a quadratic regression model where the explanatory variable is a logarithmic transformation of the number of likes $\theta(u)$, i.e. $y(u) = \beta_0 + \beta_1 \log \theta(u) + \beta_2 \log^2 \theta(u)$. Using the notation introduced in *Methods*, it is $y(u) = |\mathscr{P}^u_{\text{science}}|/|\mathscr{P}_{\text{science}}|$ for $u \in V^L_{\text{science}}$ and $y(u) = |\mathscr{P}^u_{\text{conspir}}|/|\mathscr{P}_{\text{conspir}}|$ for $u \in V^L_{\text{conspir}}$. Coefficients are estimated using weighted least squares with weights given by the total number of users per engagement value and they are – with the corresponding standard errors inside the round brackets - $\beta_0 = 0.0669(0.0011), \beta_1 = 0.2719(0.0137)$ and $\beta_2 = 0.0419(0.0040)$, with $r^2 = 0.7133$, for users polarized towards science, and $\beta_0 = 0.1229(0.0014), \beta_1 = 0.9023(0.0195)$ and $\beta_2 = 0.1629(0.0054)$, with $r^2 = 0.8876$, for users polarized towards conspiracy. All the p-values are close to zero.

Summarizing, we find that the consumption of polarizing contents is dominated by confirmation bias through the mechanism of challenge avoidance: users polarized towards a narrative tend to consume nearly exclusively content adhering to their system of beliefs, thereby minimizing their cognitive dissonance. Indeed, as their engagement grows, polarized users span their attention focus over a higher number of pages (and topics) keeping consistence with their behavioral attitude.

By exploiting the social network of polarized users and their friends, we investigate the role of reinforcement seeking mechanism in the homophily driven choice of friends on Facebook - i.e., the tendency of users to aggregate around common interests. Figure 5 shows the fraction of friends of polarized users as a function of their
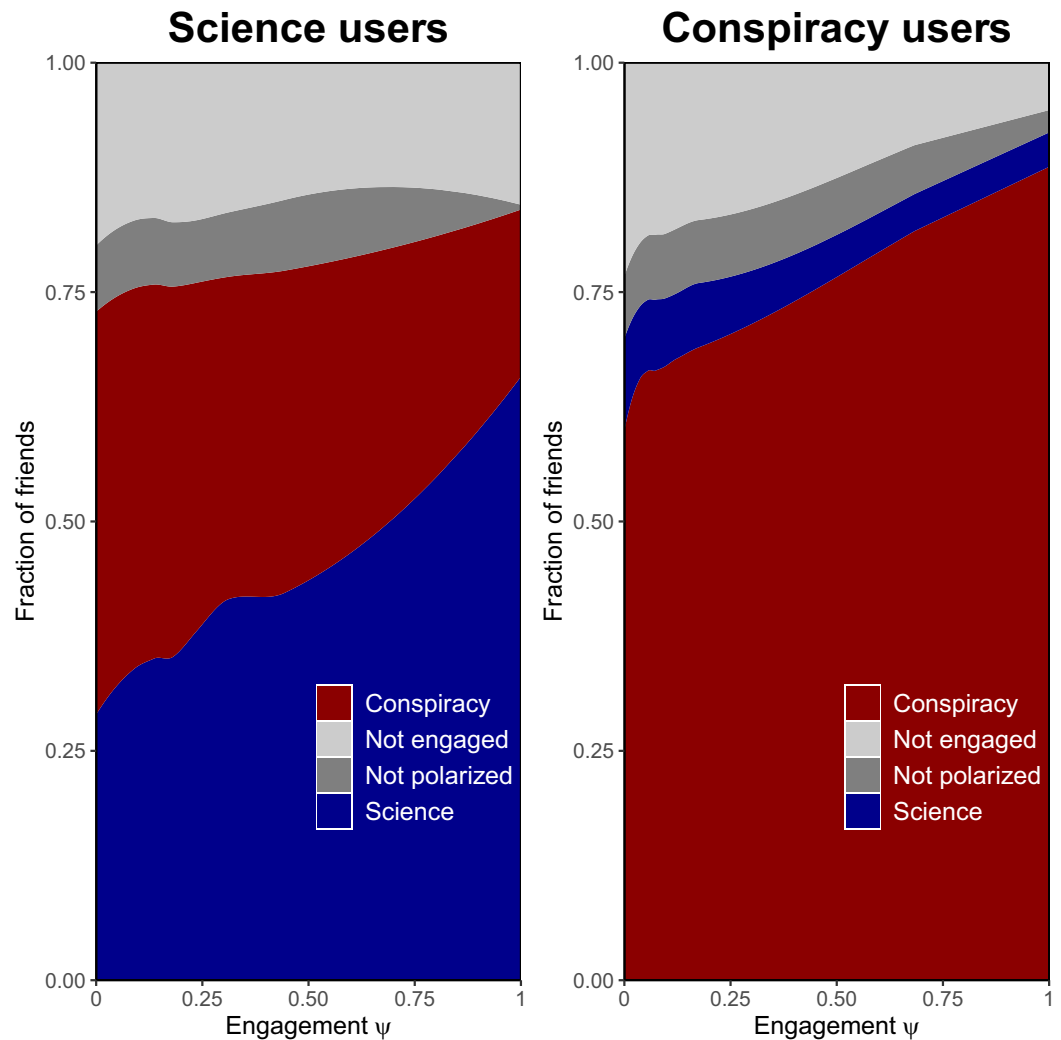
Figure 5. Users not only tend to be very polarized, but they also tend to be linked to users with similar preferences. Fraction of neighbors as a function of the engagement ψ. For a polarized science supporter u, the fraction of friends v with likewise polarization significantly grows with the engagement ψ from $\gtrsim 0.30$ to $\gtrsim 0.66$. For a polarized conspiracy supporter u, the fraction of friends v with likewise polarization is very high ($\gtrsim 0.62$) and grows with the engagement ψ up to $\gtrsim 0.87$ for the most engaged users. A user is labelled as "Not polarized" if she liked the same number of posts from Science pages and Conspiracy pages, respectively. A user is labelled as "Not engaged" if she has no liking activity on the pages of our dataset.

engagement $\psi(\cdot)$ both in the case of users in $V_{science}^{L}$ and in the case of users in $V_{conspir}^{L}$. Plots suggest that users not only tend to be very polarized, but they also tend to be linked to users with similar preferences. This is more evident among conspiracists where, for a polarized user u, the fraction of friends v with likewise polarization is very high ($\gtrsim 0.62$) and grows with the engagement ψ up to $\gtrsim 0.87$. The neighborhood of a polarized scientific user u tends to be more heterogeneous, but the fraction of friends with likewise polarization of u grows stronger with the engagement ψ (from $\gtrsim 0.30$ up to $\gtrsim 0.66$). Furthermore, Fig. 5 clearly indicates that the neighborhood of users engaged with polarizing contents (verified or not) is almost completely polarized as well (74–80% for science users and 72–90% of conspiracy users). The fact that highly polarized users have friends exhibiting an opposite polarization is a direct evidence of the challenge avoidance mechanism: contents promoted by friends which contrast one's worldview are ignored.

Summarizing, we find that the activity of a user on a polarizing content increases the probability to have friends with similar characteristics. Such information is a precious insight toward the understanding of information diffusion. Indeed, a previous work has shown that users usually exposed to undocumented claims (e.g., conspiracy stories) are the most likely to confuse intentional false information as usual conspiracy stories[3].

Engagement, friends and shared news sources. Looking at the self-description of the news sources, several distinct targets emerge both between science pages and between conspiracy pages (see Tables 2 and 3, respectively). This calls for a distinction between friends of a polarized user u who share with u a similar polarization resulting by liking contents of the same community and friends of u who actually like contents promoted
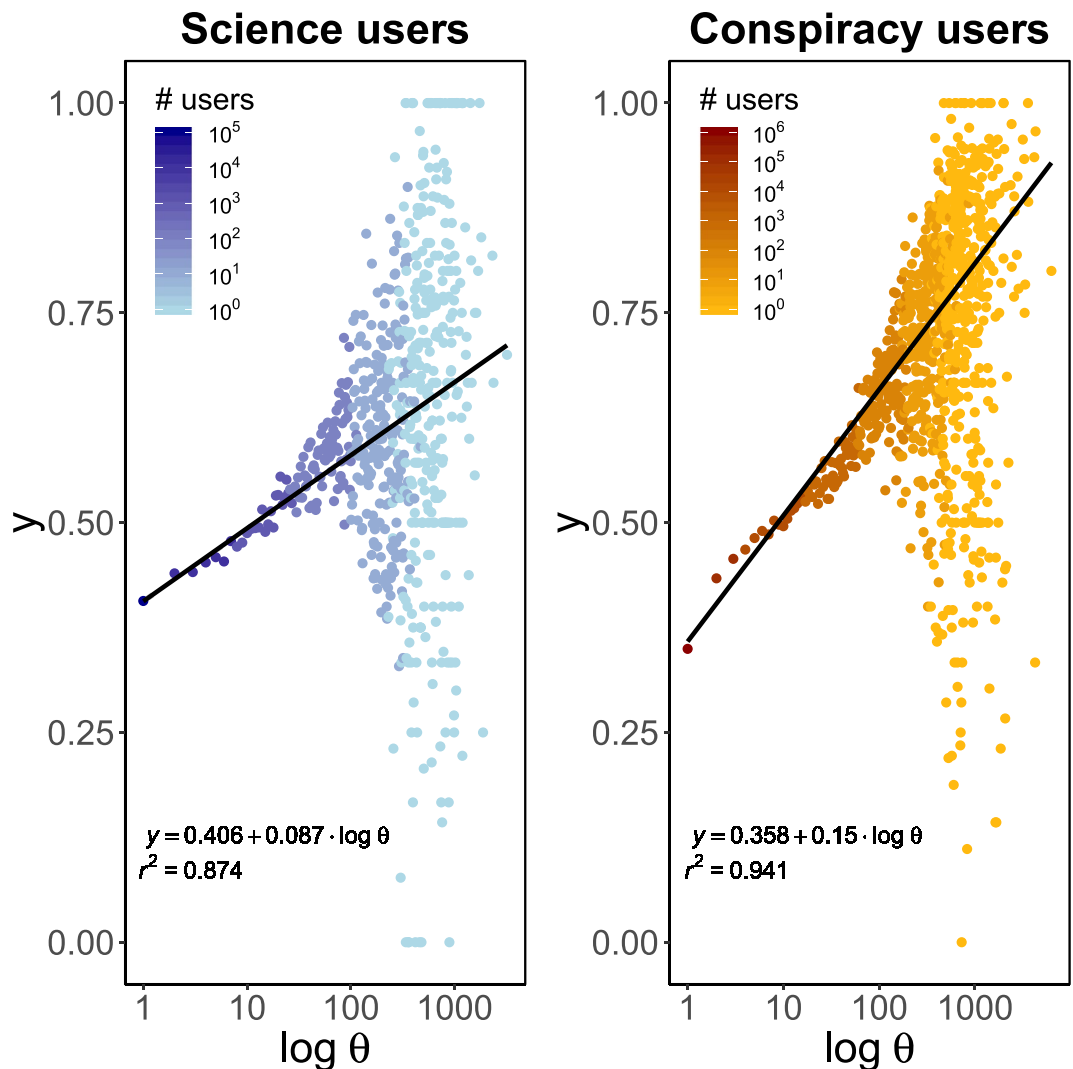
Figure 6. The fractions of science (conspiracy) friends with liking activity on the community pages liked by any given science (conspiracy) user. Left panel: users polarized towards science. Right panel: users polarized towards conspiracy. In both panels, for a polarized user u, we plot the average fraction of polarized friends with likewise polarization of u who like contents promoted by the same pages supported by u versus the number of likes $\log \theta(u)$ of u. Full lines are the results of a linear regression model $y(u) = \beta_0 + \beta_1 \log \theta(u)$. Coefficients are estimated using weighted least squares with weights given by the total number of users per engagement value. In both cases, all the p-values are close to zero.

by the same pages supported by u. In other words, in the first case the user u and her neighbourhood are grouped together at community-level (they have same/similar polarization but they like different pages); in the second case the user u and her neighbourhood are grouped together at page-level (they like not only pages in the same community but they are also somewhat active on the same set of pages).

For a polarized scientific user $u \in V_{\text{science}}^L$, in the left panel of Fig. 6, we show the log-linear plot of the average fraction y of friends $v \in V_{\text{science}}^L$ with liking activity on the community pages liked by u, respect given number of likes θ of the user u. In the right panel, we show the same quantities for polarized conspiracy users in V_{conspir}^L. Figure 6 suggests in both cases a linear correlation among the variables; thus, we check whether for a polarized user u, the fraction of friends in her category who like contents from the community pages preferred by u, $y(u)$, can be predicted by means of a linear regression model where the explanatory variable is a logarithmic transformation of the number of likes $\theta(u)$, i.e. $y(u) = \beta_0 + \beta_1 \log \theta(u)$. Coefficients are estimated using weighted least squares with weights given by the total number of users per engagement value and they are – with the corresponding standard errors inside the round brackets – $\beta_0 = 0.4062\,(0.0007)$ and $\beta_1 = 0.0869\,(0.0012)$, with $r^2 = 0.8744$, for users polarized towards science; $\beta_0 = 0.3582\,(0.0007)$ and $\beta_1 = 0.1501\,(0.0012)$, with $r^2 = 0.9413$, for users polarized towards conspiracy. All the p-values are close to zero. This suggests that polarized users not only tend to surround themselves with friends having similar systems of beliefs, but they actually share with them the involvement within the same community pages.
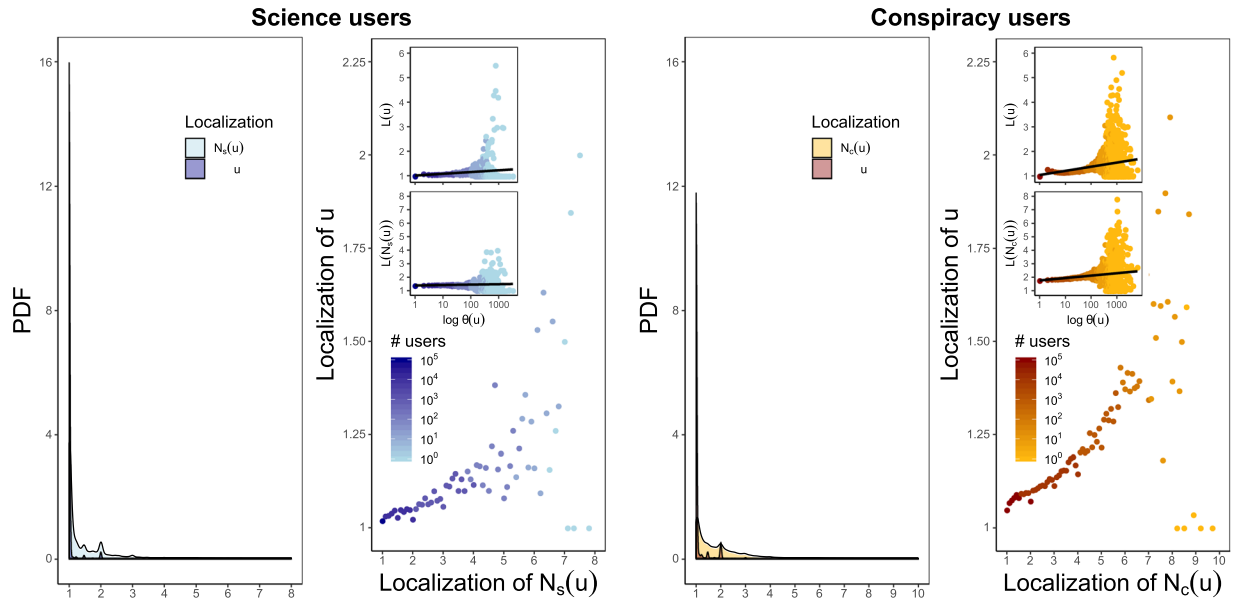
Figure 7. Users tend to remain confined within groups of very few pages even when the corresponding neighborhoods are active on several news sources. For both the communities, plots show the PDF of the localization of a polarized user u and the localization of her polarized neighborhood, and the relationship between these two order parameters. The inset plots show on a logarithmic x scale the relation of $\theta(u)$ with $L(u)$ and $L(N_s(u))$ $(L(N_c(u)))$, respectively, for any $u \in V_{\text{science}}^L$ (V_{conspir}^L). Full lines are the results of a linear regression model whose coefficients are estimated using weighted least squares with weights given by the total number of users per engagement value.

Confirmation bias as a filter to peer influence. Here we study the liking activity of polarized users in more detail by measuring how they span such activity across the various community pages. For science (conspiracy) community, Fig. 7 shows the probability distribution of the localization L along the user set and along the neighborhood set, and the relationship between $L(u)$ and $L(N_s(u))$ $(L(N_c(u)))$ for each science (conspiracy) user u.

For each polarized user u, we observe a positive correlation between these two order parameters: Pearson's correlation coefficient $r_{L(u),L(N_s(u))} \sim 0.5962$ with p-value $\sim 10^{-7}$ for science community, Pearson's correlation coefficient $r_{L(u),L(N_c(u))} \sim 0.5935$ with p-value $\sim 10^{-9}$ for conspiracy community. Nevertheless, the most users remain confined within groups of very few pages even with neighborhoods fairly active on several news sources. Moreover, the inset plots of Fig. 7 show on a logarithmic x scale the relation of $\theta(u)$ with $L(u)$ and $L(N_s(u))$ $(L(N_c(u)))$, respectively, for each $u \in V_{\text{science}}^L$ (V_{conspir}^L). Full lines are the results of a linear regression model whose coefficients are estimated using weighted least squares with weights given by the total number of users per engagement value.

By investigating the self-description of the news sources, we also find that the most users who decide to span their liking activity over a higher number of pages, choose pages dealing with very interlinked topics ($\gtrsim$76% of science users and $\gtrsim$69% of conspiracy users). Such an evidence suggests that the reinforcement seeking mechanism limits the influence of neighbors and primarily drives the selection and the diffusion of contents even within groups of like-minded people.

Peer support and reinforcement of preexisting beliefs. So far we have shown how confirmation bias acts as filter to peer influence. In this Section, we investigate the effects of the joint action of confirmation bias and peer influence when the latter does not conflict the cognitive mechanisms of challenge avoidance and reinforcement seeking. Namely, we first compare the liking activity of each polarized user across her preferred community pages with the liking activity expressed on the same pages by the part of her neighborhood with likewise polarization. Then we compare the daily time series given by the number of likes expressed by a polarized user and her like-minded neighborhood, respectively, and we investigate the existence of a causal effect of the latter on the former.

For any polarized user $u \in V_{\text{science}}^L$ we compute the cosine between the versors $\hat{\mathbf{u}} = \frac{\mathbf{u}}{|\mathbf{u}|}$ and $\widehat{\mathbf{N}_s}(\mathbf{u}) = \frac{\mathbf{N}_s(\mathbf{u})}{|\mathbf{N}_s(\mathbf{u})|}$ where $\mathbf{u}$ and $\mathbf{N}_s(u)$ are the vectors whose k^{th} component is the number of likes expressed by u and $N_s(u)$ on the kth page of $\mathscr{P}_{science}^u$, respectively (see *Methods* for further details). The same quantities are calculated for any polarized user $u \in V_{\text{conspir}}^L$. Figure 8 shows the level of proportionality between the distributions of liking activity of u and $N_s(u)$ $(N_c(u))$ across the pages of $\mathscr{P}_{science}^u$ $(\mathscr{P}_{conspir}^u)$, respectively, versus the number of likes $\log_2(\theta(u))$ of user u. Segments represent the average of the cosine measurements regarding users with a liking activity in the range of the corresponding bin (one of 1, 2, (2, 4], (4, 8], (8, 16], ...), and they are colored according to the total number of users belonging to such a range.
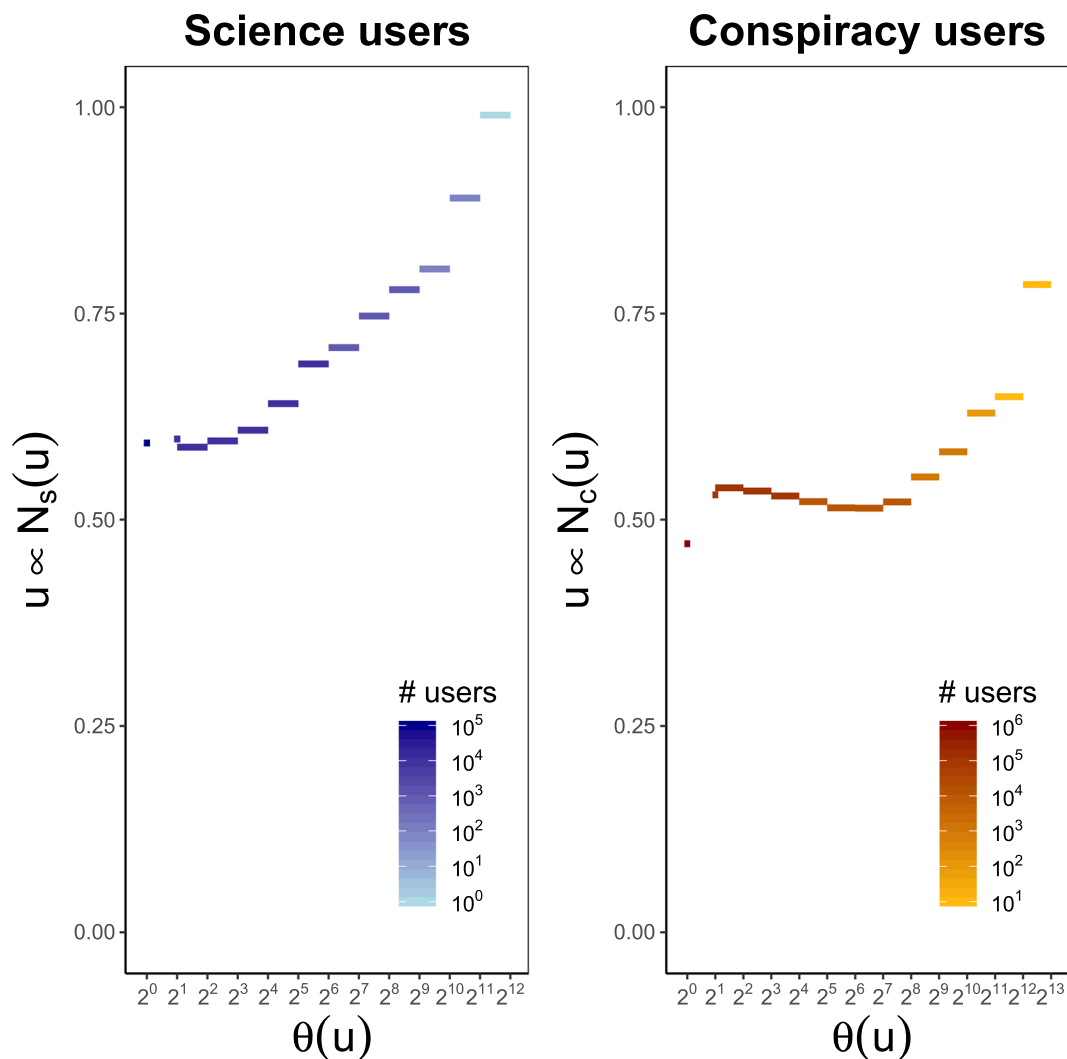
Figure 8. The distribution of likes of a polarized user across her community pages is proportional to the distribution of likes expressed on the same news sources by her neighborhood part with likewise polarization. In the left panel, for any user $u \in V_{\text{science}}^{L}$, we show the cosine similarity between the vectors $\mathbf{u}$ and $\mathbf{N}_s(\mathbf{u})$ whose k^{th} component is the number of likes expressed by u and $N_s(u)$ on the kth page of $\mathscr{P}_{\text{science}}^{u}$, respectively, versus the number of likes $\log_2(\theta(u))$. Segments represent the average of the cosine measurements regarding users with a liking activity in the range of the corresponding bin. The right panel shows the same quantities for polarized users in V_{conspir}^{L}.

The plots show that a polarized user and her likewise polarized neighborhood distribute their likes across her community pages in a similar way, both in science (left panel) and conspiracy (right panel) community. Moreover, except a nearly constant early pattern for conspiracy users, this trend grows with the user engagement suggesting how peer influence acts as a support for reinforcement seeking. Such an interpretation is pointed out more clearly by comparing the temporal evolution of the liking activity of a polarized user and her likewise polarized neighborhood, respectively.

In order to carry out such an analysis we restrict the observations to those polarized users u who exhibit a liking activity large enough to allow the comparison between the time series of likes per day expressed by u and her likewise polarized neighborhood, respectively. Namely we define

$$\overline{V}_{\text{science}}^{L} = \{u \in V_{\text{science}}^{L} | \theta(u) \geq \overline{\theta}_{\text{science}}\},$$

where $\overline{\theta}_{\text{science}} = 13$ is the average number of total likes expressed by a user of V_{science}^{L}, and

$$\overline{V}_{\text{conspir}}^{L} = \{u \in V_{\text{conspir}}^{L} | \theta(u) \geq \overline{\theta}_{\text{conspir}}\},$$

where $\overline{\theta}_{\text{conspir}} = 12$ is the average number of total likes expressed by a user of V_{conspir}^{L}. Furthermore, let $t(u)$ and $t(N_s(u))$ $(t(N_c(u)))$ be the time series of likes per day expressed over $\mathscr{P}_{\text{science}}^{u}$ $(\mathscr{P}_{\text{conspir}}^{u})$ by a user $u \in \overline{V}_{\text{science}}^{L}$ $(u \in \overline{V}_{\text{conspir}}^{L})$ and her likewise polarized neighborhood, respectively. We estimate the temporal similarity between

Figure 9. The most users produce a daily time series of likes very similar to the one produced on the same pages by the likes of their respective likewise neighborhood. Left panel: scientific polarized users. Right panel: conspiracy polarized users. PDF of Dynamic time warping (DTW) distance between the daily time series of likes expressed by a polarized user and her likewise polarized neighborhood, respectively. The inset plots show the almost perfect correlation between the difference of liking amount and the corresponding DTW distance, suggesting that extreme DTW distances are due to this factor more than to effective temporal dissimilarity between liking activities.

the liking activity of u and $N_s(u)$ $(N_c(u))$ by measuring the DTW distance $d(t(N_s(u)),\ t(u))\ (d(t(N_c(u)),\ t(u)))$ (see *Methods* for further details). Figure 9 shows the PDF of such distances for science users (left panel) and conspiracy users (right panel). In both cases we can observe that the most users produce a daily time series of likes very similar to that produced by the likes of their likewise neighborhood. Moreover, the inset plots show the strong positive correlation (Pearson's coefficient $\gtrsim 0.9887$ and $\gtrsim 0.9886$ for science and conspiracy, respectively, with both p-values close to zero) between difference in size of u liking activity compared to $N_s(u)$ $(N_c(u))$ and the corresponding DTW distance, suggesting that extreme DTW distances are due to the almost perfect uphill linear relationship more than to an effective temporal dissimilarity between liking activities.

For each science user in $\overline{V}_{\text{science}}^L$, we also investigate a causal effect of $t(N_s(u))$ on $t(u)$ by testing the null hypothesis that the former is Granger-noncausal for the latter, namely $\mathbb{H}_0$: $= t(u)_{\tau+1} \perp\!\!\!\perp \mathscr{I}^*(\tau) \,|\, \mathscr{I}^*_{-t(N_s(u))}(\tau)$. The alternative hypothesis $\mathbb{H}_1$ is predictive causality. The same analysis is repeated for each conspiracy user in $\overline{V}_{\text{conspir}}^L$ (see *Methods* for further details). In both panels of Fig. 10 we show the PDF of p-values obtained by performing such Granger causality tests. The inset plots show the cumulative distribution function (CDF) of the same quantities. Graphics show that the null hypothesis can be rejected as false: p-values less than the threshold $\overline{\alpha} = 0.05$ are more likely than the others in both the communities and represent ~29% and ~34% of the total in science and conspiracy, respectively.
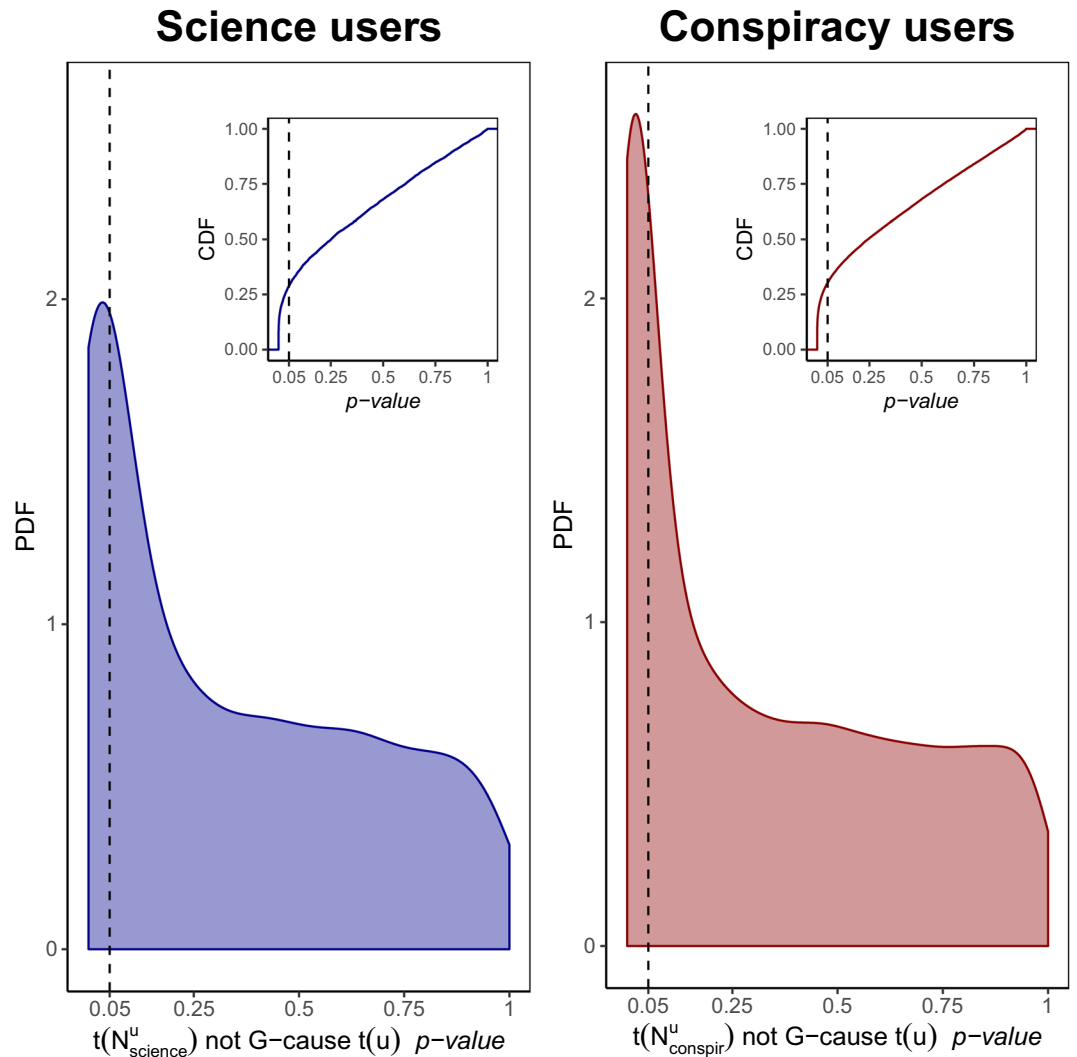
Figure 10. The causal effect of the polarized neighborhood liking activity on the temporal distribution of likes of a polarized user. PDF of p-values of Granger causality tests performed for investigating a causal effect of $t(N_s(u))$ $(t(N_c(u)))$ on $t(u)$ for any science user in $\overline{V}^L_{\text{science}}$ (left panel) and any conspiracy user in $\overline{V}^L_{\text{conspir}}$ (right panel). The inset plots show the cumulative distribution function (CDF) of the same quantities. Graphics show that the null hypothesis $\mathbb{H}_0$ can be rejected as false.

As an example, Fig. 11 shows the daily time series of a selected user $u \in \overline{V}^L_{\text{science}}$ with $\theta(u) = 767$ (left panel) and a selected user $v \in \overline{V}^L_{\text{conspir}}$ with $\theta(v) = 488$ (right panel) compared with the daily time series of their neighborhood part $N_s(u)$ and $N_c(v)$ who have expressed 779 and 919 likes, respectively. For the pair of time series $(t(N_s(u)), t(u))$, DTW returns a distance equal to 407 and the Granger causality test a p-value $\sim 10^{-4}$. For the pair of time series $(t(N_c(v)), t(v))$, DTW returns a distance equal to equal to 463 and the Granger causality test a p-value $\sim 10^{-5}$.

Finally, for each polarized user $u \in \overline{V}^L_{\text{science}}$, we study the relationship between predictive causality of $t(N_s(u))$ on $t(u)$ and the engagement of u. To this aim we use the peer influence probability $\text{PIP}^u_{\text{science}}$ (see *Methods* for further details) that provides a measure of neighbors influence effectiveness in reinforcing the system of beliefs of u. The same analysis is carried out for any polarized user $u \in \overline{V}^L_{\text{conspir}}$. Figure 12 shows the peer influence probability of u versus the number of likes $\log_2(\theta(u))$ of u both in science (left panel) and conspiracy (right panel) community. Segments represent the average of peer influence probabilities regarding users with a liking activity in the range of the corresponding bin, and they are colored according to the total number of users involved in such a range.

Plots show how, in both communities, polarized users reinforce their preexisting beliefs by leveraging the activity of their like-minded neighbors, and this trend grows with the user engagement suggesting how peer influence acts as a support for reinforcement seeking.
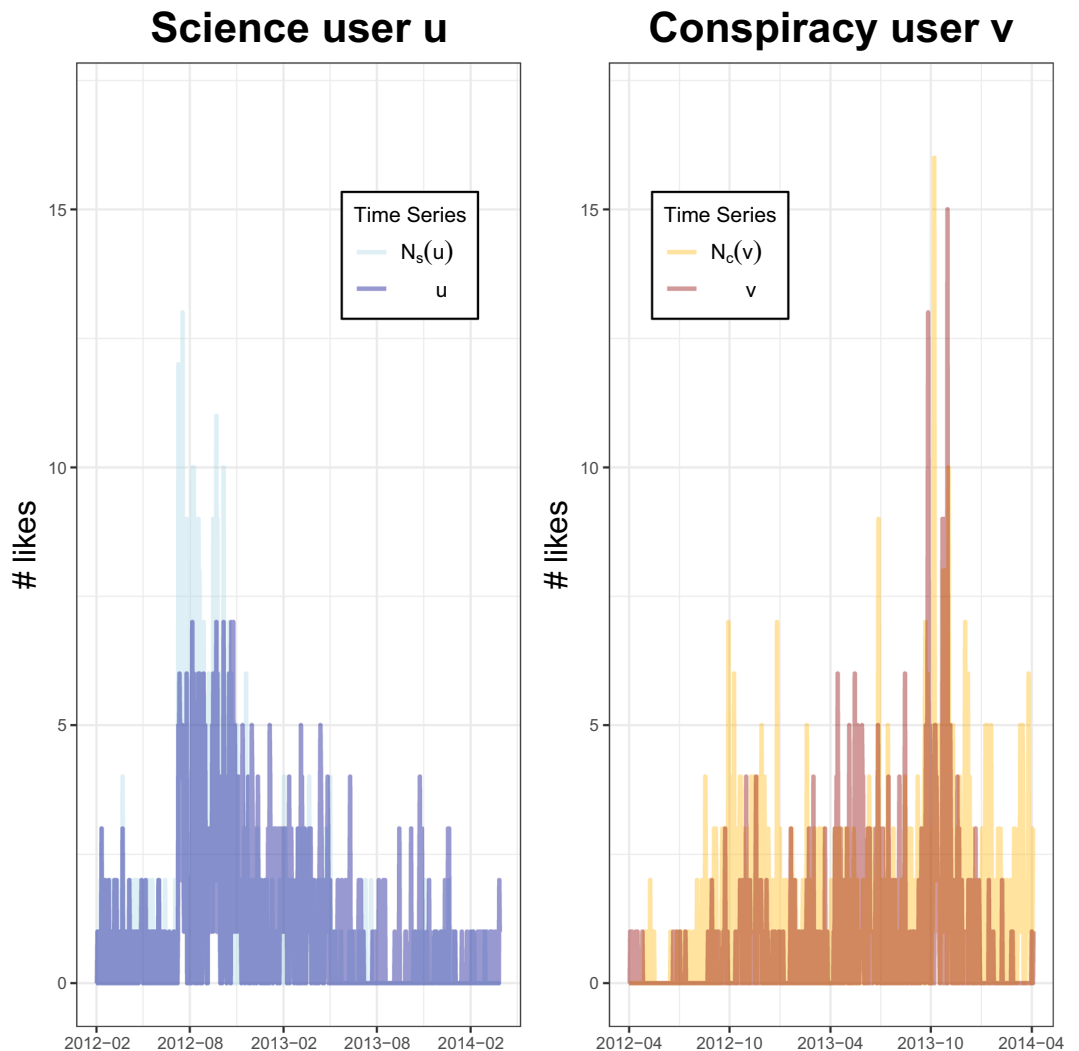
Figure 11. Time series of likes per day expressed by a selected science user *u* and by a selected conspiracy user *v* compared with the daily time series of likes of their neighborhood part $N_s(u)$ and $N_c(v)$, respectively. To the pair of time series $(t(N_s(u)), t(u))$ correspondes a DTW distance of 407 and a Granger p-value $\sim 10^{-4}$ (left panel). To the pair of time series $(t(N_c(v)), t(v))$ correspondes a DTW distance of 463 and a Granger p-value $\sim 10^{-5}$ (right panel).

Conclusions

In this paper we studied the effects of confirmation bias experience on the spreading of information in a social network of 1.2 M users engaged with two very distinct and conflicting narratives on Facebook.

Our analyses showed the action of challenge avoidance mechanism in the emergence, around the selected narratives, of two well-separated and polarized groups of users (i.e., echo chambers) who also tend to be surrounded by friends having similar systems of beliefs.

Furthermore, we explored the hypothesis that such a pattern is recursive within a single echo chamber. Despite a shared way of thinking, we proved how during social interactions the strength of confirmation bias is stronger than one could think, leading the action of peer influence into its service and fostering the formation of highly polarized subclusters within the same echo chamber. The fact that polarized users tend to remain confined within groups of very few pages even when the corresponding neighborhoods are active on several news sources, suggests that the reinforcement seeking mechanism limits the influence of neighbors and primarily drives the selection and the diffusion of contents even within groups of like-minded people.

Finally, we investigated the effects of the joint action of confirmation bias and peer influence when this latter does not conflict the cognitive mechanisms of challenge avoidance and reinforcement seeking. Namely, we compared the liking activity of polarized users and the liking activity of their likewise polarized neighborhood, and we test a causal effect of the latter on the former. Our findings revealed that polarized users reinforce their preexisting beliefs by leveraging the activity of their like-minded neighbors, and this trend grows with the user engagement suggesting how peer influence acts as a support for reinforcement seeking.
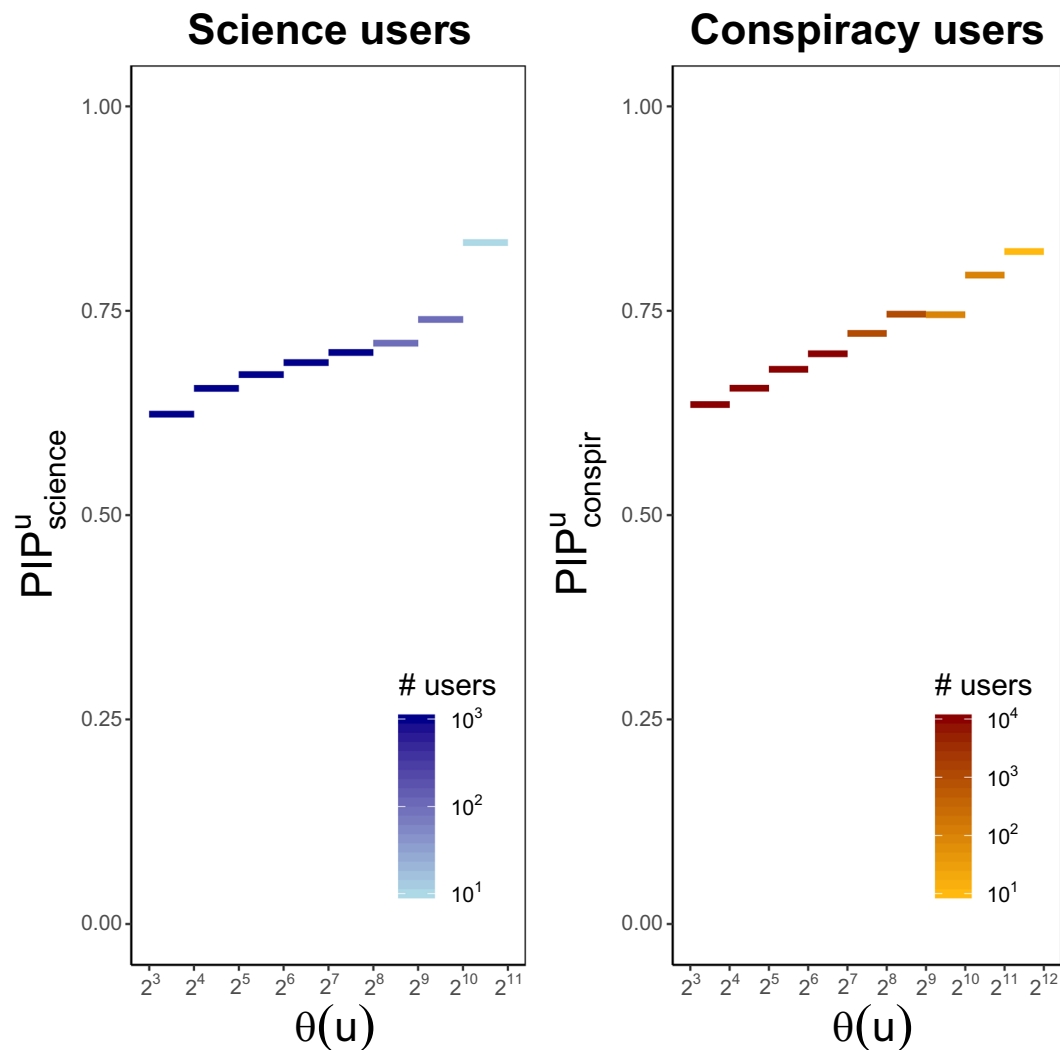
Figure 12. Polarized users reinforce their preexisting beliefs through the influence of their like-minded neighbors. In the left panel, for any user $u \in V^L_{\text{science}}$, we show the peer influence probability $\text{PIP}^u_{\text{science}}$ of $N_s(u)$ on u across the pages of $\mathscr{P}^u_{\text{science}}$, versus the number of likes $\log_2(\theta(u))$. Segments represent the average of peer influence probabilities regarding users with a liking activity in the range of the corresponding bin. The right panel shows the same quantities for polarized users in V^L_{conspir}.

In such a context, also the positive role played by social influence - e.g., by enabling social learning, seems to lose its effectiveness in the effort to smooth polarization and reduce misinformation risk and its consequences. This makes it even more difficult to design efficient communication strategies to prevent rumors and mistrust.

Internet and social media are the ideal ground for the spread of misinformation to speed up, but individual choices more than algorithms characterise the consumption patterns of users and their friends. Therefore, working towards long-term solutions for these challenges can not be separated from a deep understanding of users' cognitive determinants behind these phenomena.

References

1. Sunstein, C. R. & Vermeule, A. Conspiracy theories: Causes and cures. *Journal of Political Philosophy* **17**, 202–227, https://doi.org/10.1111/j.1467-9760.2008.00325.x (2009).
2. Bessi, A. *et al.* Science vs conspiracy: Collective narratives in the age of misinformation. *PLoS One* **10**(2), e0118093 (2015).
3. Mocanu, D., Rossi, L., Zhang, Q., Karsai, M. & Quattrociocchi, W. Collective attention in the age of (mis)information. *Computers in Human Behavior* **51**, 1198–1204, https://doi.org/10.1016/j.chb.2015.01.024, Computing for Human Learning, Behaviour and Collaboration in the Social and Mobile Networks Era (2015).
4. Bessi, A., Scala, A., Rossi, L., Zhang, Q. & Quattrociocchi, W. The economy of attention in the age of (mis)information. *Journal of Trust Management* **1**, 12, https://doi.org/10.1186/s40493-014-0012-y (2014).
5. Howell, W. L. Digital wildfires in a hyperconnected world (2013).

6. Kuklinski, J. H., Quirk, P. J., Jerit, J., Schwieder, D. & Rich, R. F. Misinformation and the currency of democratic citizenship. *Journal of Politics* **62**, 790–816, https://doi.org/10.1111/0022-3816.00033 (2000).
7. Nyhan, B. & Reifler, J. When corrections fail: The persistence of political misperceptions. *Political Behavior* **32**, 303–330, https://doi.org/10.1007/s11109-010-9112-2 (2010).
8. Zollo, F. *et al*. Debunking in a world of tribes. *PLoS One* **12**, e0181821, https://doi.org/10.1371/journal.pone.0181821 (2017).
9. Nickerson, R. S. Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology* **2**, 175–220, https://doi.org/10.1037/1089-2680.2.2.175 (1998).
10. Bakshy, E., Messing, S. & Adamic, L. A. Exposure to ideologically diverse news and opinion on facebook. *Science* **348**, 1130–1132, https://doi.org/10.1126/science.aaa1160, https://science.sciencemag.org/content/348/6239/1130.full.pdf (2015).
11. Del Vicario, M. *et al*. The spreading of misinformation. *Proceedings of the National Academy of Sciences*, https://doi.org/10.1073/pnas.1517441113, http://www.pnas.org/content/early/2016/01/02/1517441113.full.pdf (2016).
12. Cinelli, M. *et al*. Selective exposure shapes the facebook news diet. *arXiv e-prints* arXiv:1903.00699, 1903.00699 (2019).
13. Sunstein, C. *Echo Chambers* (Princeton University Press, 2001).
14. Bastos, M., Mercea, D. & Baronchelli, A. The geographic embedding of online echo chambers: Evidence from the brexit campaign. *PLoS One* **13**, 1–16, https://doi.org/10.1371/journal.pone.0206841 (2018).
15. Del Vicario, M. *et al*. Echo chambers: Emotional contagion and group polarization on facebook. *Scientific Reports* **6**, https://doi.org/10.1038/srep37825 (2016).
16. Shatz, I. The Confirmation Bias: Why People See What They Want to See. *Effectiviology* https://effectiviology.com/confirmation-bias/ (2018).
17. Garrett, R. K. Politically motivated reinforcement seeking: Reframing the selective exposure debate. *Journal of Communication* **59**, 676–699, https://doi.org/10.1111/j.1460-2466.2009.01452.x (2009).
18. Festinger, L. *A Theory of Cognitive Dissonance* (Stanford University Press, 1957).
19. Abrams, D., Wetherell, M., Cochrane, S., Hogg, M. A. & Turner, J. C. Knowing what to think by knowing who you are: Self-categorization and the nature of norm formation, conformity and group polarization. *British journal of social psychology* **29**, 97–119 (1990).
20. Byford, J. *Conspiracy Theories: A Critical Introduction*. (Palgrave Macmillan, London, 2011).
21. Fine, G., Campion-Vincent, V. & Heath, C. *Rumor Mills: The Social Impact of Rumor and Legend*. (Routledge, New York, 2005).
22. Hogg, M. A. & Blaylock, D. L. *Extremism and the Psychology of Uncertainty*. (John Wiley & Sons, Chichester, UK, 2011).
23. Betsch, C. & Sachse, K. Debunking vaccination myths: strong risk negations can increase perceived vaccination risks. *Health psychology: official journal of the Division of Health Psychology, American Psychological Association* **32**(2), 146–155 (2013).
24. Bikhchandani, S., Hirshleifer, D. & Welch, I. Learning from the behavior of others: Conformity, fads, and informational cascades. *The Journal of Economic Perspectives* **12**, 151–170 (1998).
25. Baddeley, M. Herding, social influence and expert opinion. *Journal of Economic Methodology* **20**, 35–44, https://doi.org/10.1080/1350178X.2013.774845 (2013).
26. Kassin, S. M., Dror, I. E. & Kukucka, J. The forensic confirmation bias: Problems, perspectives, and proposed solutions. *Journal of Applied Research in Memory and Cognition* **2**, 42–52, https://doi.org/10.1016/j.jarmac.2013.01.001 (2013).
27. Bessi, A., Caldarelli, G., Del Vicario, M., Scala, A. & Quattrociocchi, W. *Social Determinants of Content Selection in the Age of (Mis)Information*, chap. 18, 259–268, https://doi.org/10.1007/978-3-319-13734-6_18 (Springer International Publishing, Cham, 2014).
28. Said, S. E. & Dickey, D. A. Testing for unit roots in autoregressive-moving average models of unknown order. *Biometrika* **71**, 599–607 (1984).
29. Fuller, W. A. *Introduction to Statistical Time Series*, *Second Edition* (John Wiley & Sons, Inc, 1996).
30. Dickey, D. A. & Fuller, W. A. Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association* **74**, 427–431, https://doi.org/10.1080/01621459.1979.10482531 (1979).
31. Banerjee, A., Dolado, J., Galbraith, J. & Hendry, D. *Co-integration*, *Error Correction*, *and the Econometric Analysis of Non-Stationary Data* (Oxford University Press, 1993).
32. Newman, M. *Networks: An Introduction* (Oxford University Press, 2010).
33. Akaike, H. *Information Theory and an Extension of the Maximum Likelihood Principle* (pp. 199–213. Springer New York, New York, NY, 1998).
34. Akaike, H. A new look at the statistical model identification. *IEEE Transactions on Automatic Control* **19**, 716–723 (1974).
35. Sakamoto, Y., Ishiguro, M. & Kitagawa, G. *Akaike Information Criterion Statistics*. (Springer, Netherlands, 1986).
36. Kullback, S. & Leibler, R. A. On information and sufficiency. *Ann. Math. Statist.* **22**, 79–86, https://doi.org/10.1214/aoms/1177729694 (1951).
37. Granger, C. W. J. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica* **37**, 424–438 (1969).
38. Berndt, D. J. & Clifford, J. Using dynamic time warping to find patterns in time series. In *Proceedings of the 3rd International Conference on Knowledge Discovery and Data Mining*, AAAIWS'94, 359–370, (AAAI Press, 1994).

Acknowledgements

Author contributions

E.B., M.C. and A.S. conceived the experiments. E.B. conducted the experiments. E.B., M.C., W.Q. and A.S. analysed the results, wrote, reviewed and approved the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to E.B.

Reprints and permissions information is available at www.nature.com/reprints.

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