Article

Quantifying social organization and political polarization in online platforms

https://doi.org/10.1038/s41586-021-04167-x

Received: 30 September 2020

Accepted: 19 October 2021

Published online: 01 December 2021

Check for updates

Isaac Waller[1] & Ashton Anderson[1]✉

Mass selection into groups of like-minded individuals may be fragmenting and polarizing online society, particularly with respect to partisan differences[1–4]. However, our ability to measure the social makeup of online communities and in turn, to understand the social organization of online platforms, is limited by the pseudonymous, unstructured and large-scale nature of digital discussion. Here we develop a neural-embedding methodology to quantify the positioning of online communities along social dimensions by leveraging large-scale patterns of aggregate behaviour. Applying our methodology to 5.1 billion comments made in 10,000 communities over 14 years on Reddit, we measure how the macroscale community structure is organized with respect to age, gender and US political partisanship. Examining political content, we find that Reddit underwent a significant polarization event around the 2016 US presidential election. Contrary to conventional wisdom, however, individual-level polarization is rare; the system-level shift in 2016 was disproportionately driven by the arrival of new users. Political polarization on Reddit is unrelated to previous activity on the platform and is instead temporally aligned with external events. We also observe a stark ideological asymmetry, with the sharp increase in polarization in 2016 being entirely attributable to changes in right-wing activity. This methodology is broadly applicable to the study of online interaction, and our findings have implications for the design of online platforms, understanding the social contexts of online behaviour, and quantifying the dynamics and mechanisms of online polarization.

In 1962, Marshall McLuhan proclaimed that "The new electronic interdependence recreates the world in the image of a global village"[5]. In the decades since, there has been fierce debate about the internet's dual forces of social integration, as the world becomes increasingly interconnected, and social fragmentation, as people can more easily select to join like-minded communities[1,3,4]. Twenty years into the widespread adoption of online social media platforms, it remains unclear how online communities are socially organized. Of particular concern is whether online populations increasingly sort into homogeneous 'echo chambers' and whether social media platforms tend to shift users towards ideological extremes[6–8]. However, since these platforms consist of massive amounts of unstructured and pseudonymous data, empirically quantifying the social makeup of online communities and, in turn, the social organization of online platforms, poses a unique challenge.

Here we develop and validate a methodology using neural community embeddings[9], which represent similarities in community membership as relationships between vectors in a high-dimensional space, to quantify the positioning of online communities along social dimensions. Focusing on traditional notions of identity—age, gender and political orientation—and leveraging the complete set of 5.1 billion comments made in 10,000 communities over a 14-year period on Reddit, one of the world's largest social platforms, we produce an accurate and high-resolution picture of how the platform's macroscale structure is organized along social lines. We then apply our methodology to quantify the dynamics and mechanisms of political polarization on Reddit, and investigate three related questions: (1) To what extent does platform-level political polarization change over time? (2) Do individual users become more polarized in their political activity over time, and if so, do these changes drive platform-level polarization? and (3) Are the dynamics of polarization ideologically symmetric?

Our approach differs from prior work examining social organization and political polarization in online platforms in three main ways. First, our methodology avoids the biases that result from using self-reported data, expert labels and survey-based methods by quantifying the social makeup of communities in a purely behavioural fashion. Communities are similar only if their user bases are similar; by computing this similarity along a social dimension (for example, US political partisanship), we can recover an accurate estimate of whether a particular community's user base is more behaviourally aligned with the left or right end of the spectrum (for example, the left or right wing of US politics). Users 'vote with their feet' to decide the social orientation of communities: only action across large numbers of people matters. Previous work has used word embeddings—high-dimensional representations of text—to study cultural stereotypes[10–12] and the cultural markers of class[13]. Although our dataset comprises billions of comments, we do not use the text in

[1]Department of Computer Science, University of Toronto, Toronto, Ontario, Canada. ✉e-mail: ashton@cs.toronto.edu

our methodology. Differences in identity are reflected in the words people use, but this relationship is relatively weak for our focus on measuring the social orientation of underlying community populations. Communities that use similar language may be socially distinct, and communities with distinct language may be socially similar.

Second, previous analyses have studied platforms such as Facebook, Twitter and Amazon, on which users are guided by algorithmic curation and personalized recommendations[7,14,15]. Traces of user activity on these platforms reflect not only natural human choices but also the influence of algorithms. A recent focus has been on examining the effects of algorithmic curation on shaping online social organization—for example, measuring the prevalence of algorithmic 'filter bubbles' of homogeneous content and groups[16,17]—but user choices may have an even larger role in shaping this structure[18]. Thus, although our methodology is generally applicable to many online platforms, we apply it here to Reddit, which has maintained a minimalist approach to personalized algorithmic recommendation throughout its history. The patterns of community memberships we observe are thus more likely to be reflective of the social organization induced by natural online behaviour.

Finally, we expand the study of political polarization in social media. Polarization is understood as both a state and a process[19], but existing empirical research is largely limited to static analyses of incomplete and non-representative snapshots of activity on a platform. As such, although there is evidence that online platforms exist in states of partisan fragmentation[7,20,21], important questions about the dynamics and mechanisms of polarization processes remain unanswered. In particular, the measurement of platform-level polarization with incomplete and non-representative datasets is difficult, and tracking it over time with static analyses is impossible. Furthermore, any observed platform-level polarization could be due to two separate mechanisms with different policy implications: individual users could move towards ideological extremes in their activity over time, or relatively moderate populations could be replaced by new, more extreme populations as the user base turns over. Applying our methodology, we conduct dynamic analyses of complete platform activity to measure both platform- and individual-level polarization, and compare these for the left and right wings, over the entire history of Reddit.

Social dimensions in community embeddings

We analysed the complete set of comments from Reddit, one of the world's largest online social platforms. Reddit comprises thousands of discussion-based communities, or 'subreddits', which are typically centred around a single topic (Methods, 'Data'). To quantify the macroscale structure of the platform, we used and extended community embeddings[9], which position communities in a high-dimensional space such that communities with similar memberships are close together in the space. We embedded the largest 10,006 communities, which account for 95.4% of all comments, into a 150-dimensional space (Fig. 1a) and optimized the embedding with community analogies (Methods, 'Creating the community embedding').

Analogously to how previous research uncovered axes in word embeddings that correspond to gender, class and affluence[10,11,13], we developed a methodology to find dimensions in community embeddings that correspond to social constructs. To do so, we first identified a seed pair of communities that differ in the target construct, but are similar in other respects. For example, we seeded our partisan dimension with r/democrats and r/Conservative, two partisan American political communities (see Supplementary Table 1 for descriptions of every community we reference). To robustly capture social differences along these dimensions as they are expressed on the platform, we algorithmically augmented these seeds with similar pairs of communities. For each dimension, we selected the nine pairs with the most similar vector difference from the set of all pairs of very similar communities (see Extended Data Table 1 for a list of selected pairs). The resulting set of ten seed vector differences were then averaged together to generate the final dimensions corresponding to each target concept (Fig. 1b). The method generalizes to more concepts than we study here (Methods, 'Finding social dimensions').

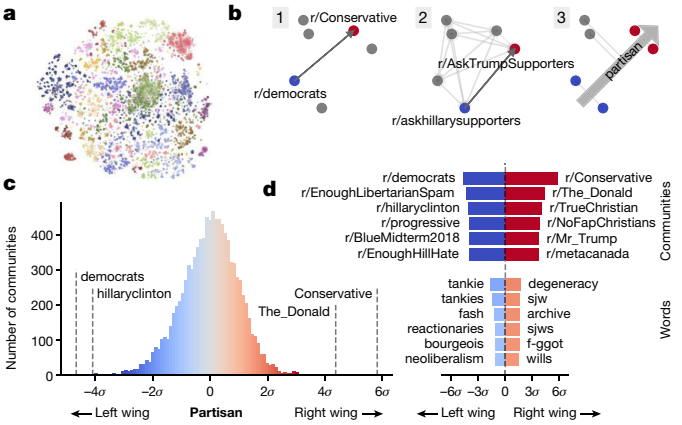


Fig. 1 | Quantifying social dimensions on Reddit. a, A two-dimensional *t*-distributed stochastic neighbour embedding (*t*-SNE) projection of the 10,006 subreddits in our Reddit community embedding, with points coloured by clusters found by hierarchical clustering. **b**, An illustration of our methodology to generate social dimensions. **c**, The distribution of partisan scores for the 10,006 most popular Reddit communities. The *x*-axis shows the number of standard deviations from the mean partisan score (*z*-score). Communities vary from far-left wing to far-right wing and are coloured by *z*-score. **d**, Top, communities most associated with the left-wing and right-wing ends of the dimension (for community descriptions, see Supplementary Table 1). Bottom, words most associated with the left-wing and right-wing ends of the dimension, considering only word usage in political communities in 2017 as quantified by the partisan-ness dimension (Extended Data Fig. 6).

Every community can then be positioned along a social dimension by projecting the community's vector representation onto the dimension. This is equal to the focal community's average similarity with communities on the right side of the seed pairs minus its average similarity with communities on the left. Communities with memberships that are more similar to one pole end up close to that pole, whereas communities that are equally similar to both ends of the spectrum fall in the middle. The distribution of community scores along the partisan dimension varies between the extreme left-wing and extreme right-wing on Reddit (Fig. 1c). The words most associated with the left and right poles illustrate how political discussion differs across the partisan spectrum (Fig. 1d). Community and word scores along the age and gender dimensions also demonstrate substantial variation (Extended Data Fig. 1). We validated these dimensions by demonstrating that scores are highly correlated with internal and external measures (Extended Data Fig. 2). While our validations suggest that the dimensions are correlated with real-world identities, we emphasize that they are measures of social associations, not individual characteristics. A community's position on the gender dimension, for example, should not be interpreted as a direct measure of the gender identity of the community's members, but instead reflects its association with the social constructs of masculinity and femininity as expressed on Reddit.

We also generated secondary dimensions that represent the strength of association with each primary dimension. For example, partisan-ness corresponds to how political a community is, whereas partisan corresponds to a community's position along the left–right political axis. These were calculated by taking the sum of the seed pairs' vectors instead of the difference, and measuring similarity to both ends of the primary dimension. We validated the partisan-ness dimension by showing that explicitly labelled partisan communities have far higher partisan-ness scores than communities in general (Extended Data Fig. 3).

The social organization of Reddit

We first applied hierarchical clustering to the embedding to obtain a grouping of communities that reflects the primary similarities and differences in their membership activity, then applied our social dimensions methodology to score every Reddit community along the age, gender and partisan axes. The distributions of Reddit communities along these social dimensions reveal significant inter- and intra-cluster diversity (Fig. 2). Entire top-level clusters of communities skew strongly towards the poles of the dimensions, significantly departing from the null hypothesis of a uniform distribution over community score percentiles. Since the clustering is based on all behavioural relationships in the original community embedding, the top-level clusters could have differed primarily in topic while remaining socially undifferentiated. Instead, the stratification along social dimensions demonstrates the importance of age, gender, and US partisanship to the high-level organization of activity on Reddit. Furthermore, the fine-grained distributions in Fig. 2 show how the platform is socially organized. For example, programming communities skew towards the masculine (36% are below the 20th percentile) and old (50% are above the 80th percentile) poles, and personal matters communities skew towards the feminine (77% above 80th percentile) and left-wing (36% below 20th percentile) poles. Hobbies 1 communities skew towards the old (52% above 80th percentile) and masculine (53% below 20th percentile) poles, whereas hobbies 2 communities skew towards the old (39% above 80th percentile) and feminine (73% above 80th percentile) poles. Politics communities exhibit a bimodal distribution on the partisan axis (77% below 20th percentile or above 80th percentile). Additionally, there is substantial diversity within each cluster of communities. Every group has communities that fall on both sides of the global mean of each dimension, and most groups have an outlier community (more than 2× s.d. from the mean) on both sides (Extended Data Fig. 4). Other dimensions exhibit similar diversity (Extended Data Fig. 5). The community scores derived from our social dimension methodology offer a high-resolution and large-scale picture of the social makeup of online communities.

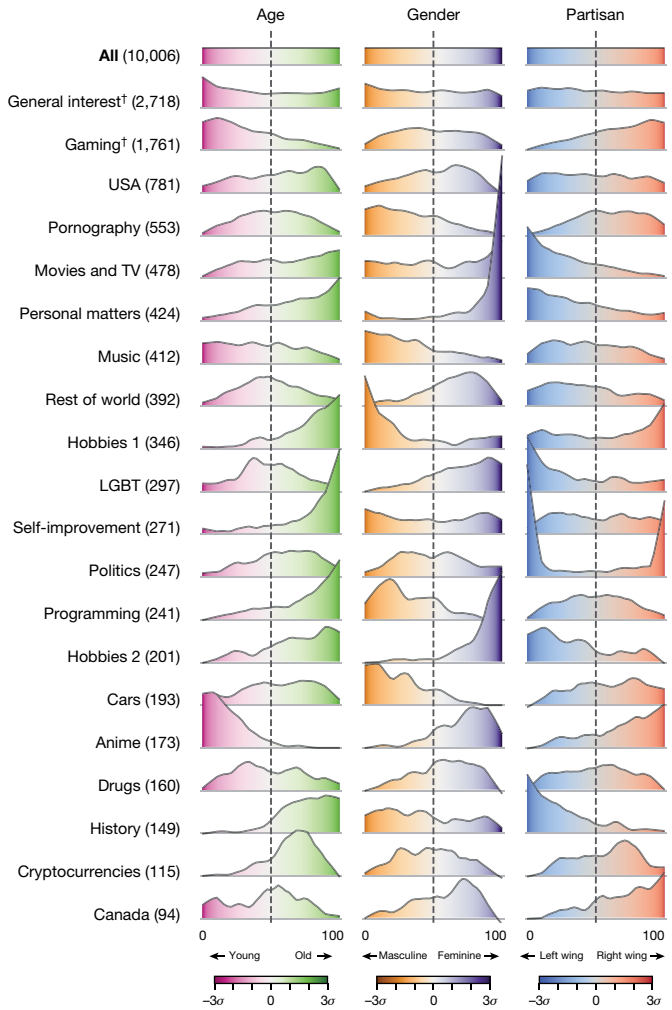


Fig. 2 | Macroscale social organization of Reddit communities. Distributions of communities along the age, gender and partisan dimensions, grouped into behavioural clusters found by hierarchical clustering. The *x*-axis represents community scores transformed into percentiles (for example, a community with age score greater than 76% of other communities would be positioned at percentile 76), and colour corresponds to z-score. As a result, the distribution for all communities (top row) is simply the uniform distribution $U(0, 100)$, while the distributions for individual clusters illustrate which percentiles are over- or under-represented within the cluster. Raw score (non-percentile) distributions are shown in Extended Data Fig. 4. The dashed line indicates the 50th percentile. Rows annotated with † comprise two or more clusters (Methods, 'Creating the community embedding').

To further clarify the nature of Reddit's social organization, we demonstrated that the online expressions of social constructs may differ from their traditional meanings in offline contexts. Focusing on the partisan axis, we quantified how it relates with the gender and age axes (Extended Data Fig. 6a, b). There is a significant monotonic relationship between the partisan and gender dimensions (Extended Data Fig. 6a), with masculine-leaning communities also skewing right wing ($r = -0.29$, $n = 10{,}006$, two-sided $P < 10^{-10}$). At the community level, the political poles on Reddit are almost completely segregated by gender; the most left-wing communities are 44.0% feminine-leaning and only 1.4% masculine-leaning, whereas the most right-wing communities are 23.3% masculine-leaning and only 2.9% feminine-leaning. The direction of this relationship is consistent with the American electorate; in the 2016 US presidential election, men voted for Donald Trump by a margin of 52% to 41%, and women voted for Hilary Clinton[22] by a margin of 54% to 39%. We also find a relationship between the partisan and age dimensions (Extended Data Fig. 6b); older communities skew towards the left-wing pole, whereas younger communities skew towards the right-wing pole ($r = -0.37$, $n = 10{,}006$, two-sided $P < 10^{-10}$). Among left-wing communities, 38.5% are older but only 2.1% are younger, while among right-wing communities, 26.1% are younger but only 2.9% are older. Notably, the direction of this relationship is the opposite of what is traditionally found in offline contexts—in 2016, the 18–29 age group voted for Hilary Clinton by a margin of 58% to 28%, whereas the 65+ age group voted for Donald Trump by 53% to 44%—but is consistent with previous observations of the relative youth of the online alt-right movement[23]. We repeated these analyses on dimensions generated with slightly different seeds to verify the robustness of our method and found similar results (Methods, 'Measuring relationships between dimensions').

Political polarization on Reddit

Next, we applied our methodology to study individual- and platform-level political polarization over time. Political activity on Reddit spans the ideological spectrum, with 35% of activity taking place left of centre, 22% of activity taking place right of centre, and 43% taking place in the centre (Fig. 3a, top). Despite this overall breadth, user activity is considerably more narrow. In line with the echo chamber hypothesis, the political activity contributed by a community's members is heavily skewed towards communities with similar partisan scores (Fig. 3a, bottom). For example, only 8% of political discussion occurs in the most left-wing communities, but among users who contribute to left-wing communities, an average of 44% of their activity takes place in left-wing communities. Similarly, only 16% of political discussion occurs in the most right-wing communities, but right-wing communities account for on average 62% of right-wing commenters' political activity. If users' distributions of activity were not skewed along

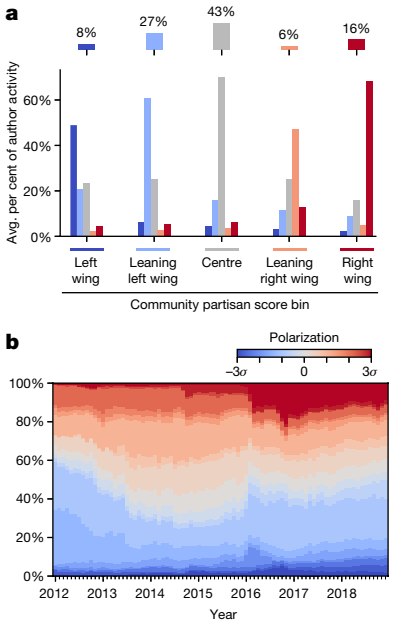


Fig. 3 | Distribution of political activity on Reddit. a, Average (avg.) distributions of political activity contributed by users of different partisan community bins. The top distribution shows the average distribution for all users—that is, independent of partisan activity—while each of the five bottom distributions shows the average distribution of political activity contributed by users who commented in the corresponding partisan category. **b**, The distribution of political activity on Reddit over time by partisan score. Each bar represents one month of comment activity in political communities on Reddit and is coloured according to the distribution of partisan scores of comments posted during the month (where the partisan score of a comment is the partisan score of the community in which it was posted.).

partisan lines, these average percentages would be approximately equal to the overall platform partisan distribution (Extended Data Fig. 7c). This pattern of selective partisan activity is also clearly apparent at the individual community level (Extended Data Fig. 7d). Consistent with previous studies of political activity on social media, a static analysis of complete Reddit activity shows that users selectively participate in ideologically homogeneous communities[7,20,21].

However, this style of analysis cannot address whether selection into partisan communities changes over time. To understand whether political activity on Reddit became more polarized throughout the platform's history, we tracked the distribution of political activity from 2012 to 2018 (Fig. 3b, Extended Data Fig. 8). While Reddit has always supported a wide range of political activity, the platform became substantially more polarized around the 2016 US presidential election. The polarization of discussion, measured by the mean absolute value partisan *z*-score of political comments (that is, mean absolute number of s.d. from the mean), remained consistently within a narrow band between 1.08 and 1.28 from 2012 until the end of 2015; it then increased sharply during 2016 and peaked at 1.86 in November 2016 (Extended Data Fig. 9a). The percentage of political activity that took place in far-left and far-right communities was only 2.8% in January 2015, but peaked at 24.8% in November 2016 (Extended Data Fig. 7a). The platform never returned to pre-2016 polarization levels, maintaining values greater than 1.44 until the end of the data time window.

A central concern is whether individual users become more polarized in their activity over time. Overall increases in platform-level polarization could be driven either by individual-level change, with existing users moving towards the partisan extremes, or by population-level turnover, with new users entering the platform in more extreme communities. To quantify this, we grouped users into cohorts on the basis of the date of their first comment in a political community and measured the average polarization of each cohort over time (Fig. 4a). This analysis reveals several insights about the dynamics and mechanisms of polarization on Reddit. First, with the exception of 2016, users generally do not polarize over time; within-cohort polarization levels usually either remained unchanged or decreased from one year to the next. We directly measured individual-level polarization by computing the fraction of users whose activity moved by at least one standard deviation towards the partisan poles. This fraction is consistently low; comparing user scores 12 months apart, it was between 1.9 and 3.3% prior to 2016, and peaked at 11.3% in November 2016 (Fig. 4b). Second, during 2016 every active cohort polarized at the same time. The month-to-month polarization trends in 2016 were remarkably synchronized across cohorts. Third, the intense increase in polarization in 2016 was disproportionately driven by new and newly political users. The change in platform-level polarization was 2.17 times what it would have been if the 2016 cohort had arrived at the average 2015 polarization level, despite only accounting for 38% of political activity during 2016 (Fig. 4c). Furthermore, a cohort's increase in polarization was directly related to its age, with newer cohorts polarizing more than older cohorts. Finally, individual polarization level is unrelated to previous activity on the platform, when measured either by calendar months since first activity or by active months spent on the platform (Fig. 4a, insets). Changes in polarization over time on Reddit are not

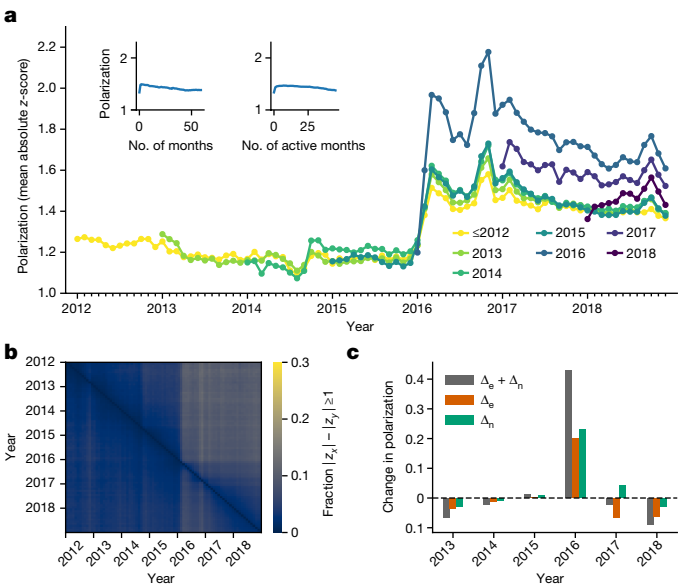


Fig. 4 | Political polarization of new and existing users. a, The average polarization of political activity on Reddit, broken down into seven user cohorts by year of the author's first political activity. Polarization is measured by the absolute *z*-score of the community—that is, the absolute number of s.d. from the mean partisan score. Inset, the relationship between polarization and number of total (left) and active (right) months since a user's first political activity. **b**, Observed within-user polarization. Each (x, y) cell represents the proportion of users whose average polarization (average absolute *z*-score) increased by one s.d. between months t_x and t_y (minor ticks on the *x*-axis indicate month), for all users active during both months. A user is considered active if they make at least 10 comments in a month. Results are robust for other choices of these two thresholds (Extended Data Fig. 7f, g). **c**, Annual change in polarization, decomposed into the change attributable to new (Δ_n) and existing (Δ_e) users. The grey bar represents the actual observed year-over-year change in polarization; Δ_e represents the change that would be observed had new users not changed at all (that is, they were only as polarized as the overall polarization in the previous year); Δ_n represents the change that would be observed had existing users not changed at all (that is, they remained only as polarized as in the previous year).

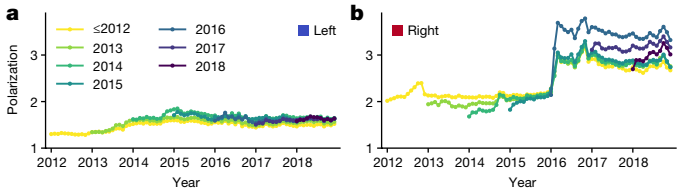


Fig. 5 | Ideological asymmetry in online polarization. a, **b**, Average polarization of activity in the left (**a**) and right (**b**) wings, decomposed into seven cohorts by year of first political activity.

associated with previous activity on the platform but rather are synchronously aligned with external events, and are disproportionately driven by new users.

Examining polarization over time separately for left-wing and right-wing communities reveals a stark ideological asymmetry. Activity on the right was substantially more polarized than activity on the left in every month between 2012–2018 (Extended Data Fig. 9a). In 2016, discussion on the right shifted significantly rightward, with polarization increasing from an average of 2.12 in November 2015 to a peak of 3.55 in November 2016. During the same period, discussion on the left and in the centre did not polarize at all (average polarization changed from 1.60 to 1.57 on the left, and from 0.58 to 0.57 in the center). The overall shift in polarization on the platform in 2016 was thus driven entirely by the change in activity on the right, despite the fact that the right was the smallest group by discussion volume (Extended Data Fig. 9b). Similar to the analogous findings for overall polarization, new users on the right in 2016 were significantly more polarized than all previous cohorts and disproportionately drove the observed polarization of the right-wing on Reddit (Fig. 5b, Extended Data Fig. 9d), consistent with the rise of large right-wing communities such as r/The_Donald. Changes in polarization on the left were small by comparison (Fig. 5a, Extended Data Fig. 9c).

Although instances of individual users becoming more polarized in their partisan score over time are rare, it is still possible that newly political users move from implicit 'gateway communities' to explicitly partisan communities. Some communities have a highly partisan user base but are not themselves explicitly political, and thus have extreme partisan scores but low partisan-ness scores (Extended Data Fig. 6c). If engagement with implicitly partisan communities is related to an increased propensity to subsequently engage with explicitly partisan communities, this could be evidence of an implicit process of polarization occurring on the platform. However, for users who were active in an explicitly left-wing or right-wing community, in any given month at most 27% had contributed in a previous month to an implicitly left-wing community and 27% had contributed to an implicitly right-wing community, restricting the population for whom such an effect could apply (Extended Data Fig. 10, top). Users tend to become active in both implicitly and explicitly partisan communities in the same month, further indicating that such a polarization effect is limited in its possible impact (Extended Data Fig. 10, bottom).

There are limitations in our approach. For example, by representing each community by a single vector in a common embedding, we measure community relationships aggregated over the entire time period of our dataset. This implicitly assumes that community similarities and community scores on social dimensions do not change. Although it is plausible that some communities change significantly in the partisan orientation of their membership, we expect these to be exceptional cases. Our method also relies on examples of the same user being a member of several communities. If large numbers of people use 'throwaway' user accounts for certain communities, thereby splitting their activity over several accounts, the relationships between these communities and the rest of the platform could be distorted.

This study introduces a new model for the analysis of online platforms. Sociologists dating back to Simmel, who pioneered the notion of 'the web of group affiliations', have used complex characterizations of group membership to understand social identity[24–27]. We have shown that by harnessing mass co-membership data, we can use high-dimensional representations of online communities to produce valid, fine-grained and semantically meaningful measurements of their social alignment. Furthermore, aggregating these measurements provides a macro-scale description of how platforms are organized along key social dimensions. Our methodology can be applied generally to quantify the social organization of online discussion, to situate important content and behaviours in the context of the platform, and to understand the nature of individual- and platform-level online polarization and the mechanisms that drive it.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-021-04167-x.

1. Sunstein, C. *#Republic: Divided Democracy in the Age of Social Media* (Princeton Univ. Press, 2018).
2. Iyengar, S. & Hahn, K. S. Red media, blue media: evidence of ideological selectivity in media use. *J. Commun.* **59**, 19–39 (2009).
3. van Alstyne, M. & Brynjolfsson, E. Electronic communities: global villages or cyberbalkanization? In *Proc. International Conference on Information Systems* 5 https://aisel.aisnet.org/icis1996/5 (1996).
4. van Dijck, J. *The Culture of Connectivity: A Critical History of Social Media* (Oxford Univ. Press, 2013).
5. McLuhan, M. *The Gutenberg Galaxy: The Making of Typographic Man* (Univ. of Toronto Press, 1962).
6. Farrell, H. The consequences of the internet for politics. *Ann. Rev. Pol. Sci.* **15**, 35–52 (2012).
7. Conover, M. D. et al. Political polarization on Twitter. *Proc. Intl AAAI Conf. Web Soc. Media* **133**, 89–96 (2011).
8. Bail, C. A. et al. Exposure to opposing views on social media can increase political polarization. *Proc. Natl Acad. Sci. USA* **115**, 9216–9221 (2018).
9. Martin, T. community2vec: vector representations of online communities encode semantic relationships. In *Proc. 2nd Workshop on NLP and Computational Social Science* 27–31 (2017).
10. Garg, N., Schiebinger, L., Jurafsky, D. & Zou, J. Word embeddings quantify 100 years of gender and ethnic stereotypes. *Proc. Natl Acad. Sci. USA* **115**, E3635–E3644 (2018).
11. Bolukbasi, T., Chang, K.-W., Zou, J. Y., Saligrama, V. & Kalai, A. T. Man is to computer programmer as woman is to homemaker? Debiasing word embeddings. *Adv. Neural Inf. Process. Syst.* **29**, 4349–4357 (2016).
12. Caliskan, A., Bryson, J. J. & Narayanan, A. Semantics derived automatically from language corpora contain human-like biases. *Science* **356**, 183–186 (2017).
13. Kozlowski, A. C., Taddy, M. & Evans, J. A. The geometry of culture: analyzing the meanings of class through word embeddings. *Am. Soc. Rev.* **84**, 905–949 (2019).
14. Shi, F., Shi, Y., Dokshin, F. A., Evans, J. A. & Macy, M. W. Millions of online book co-purchases reveal partisan differences in the consumption of science. *Nat. Hum. Behav.* **1**, 0079 (2017).
15. Del Vicario, M. et al. Echo chambers: emotional contagion and group polarization on Facebook. *Sci. Rep.* **6**, 37825 (2016).
16. Pariser, E. *The Filter Bubble: What the Internet is Hiding from You* (Penguin, 2011).
17. Flaxman, S., Goel, S. & Rao, J. M. Filter bubbles, echo chambers, and online news consumption. *Public Opin. Q.* **80**, 298–320 (2016).
18. Bakshy, E., Messing, S. & Adamic, L. A. Exposure to ideologically diverse news and opinion on Facebook. *Science* **348**, 1130–1132 (2015).
19. DiMaggio, P., Evans, J. & Bryson, B. Have American's social attitudes become more polarized? *Am. J. Sociol.* **102**, 690–755 (1996).
20. Barberá, P., Jost, J. T., Nagler, J., Tucker, J. A. & Bonneau, R. Tweeting from left to right: is online political communication more than an echo chamber? *Psychol. Sci.* **26**, 1531–1542 (2015).
21. Adamic, L. A. & Glance, N. The political blogosphere and the 2004 US election: divided they blog. In *Proc. 3rd International Workshop on Link Discovery* 36–43 (2005).
22. *An Examination of the 2016 Electorate, Based on Validated Voters* https://www.pewresearch.org/politics/2018/08/09/an-examination-of-the-2016-electorate-based-on-validated-voters/ (Pew Research Center, 2018).
23. Hawley, G. *Making Sense of the Alt-Right* (Columbia Univ. Press, 2017).
24. Simmel, G. *Conflict and the Web of Group Affiliations* (Free Press, 1955).
25. Breiger, R. L. The duality of persons and groups. *Social Forces* **53**, 181–190 (1974).
26. Bourdieu, P. *Distinction: A Social Critique of the Judgement of Taste* (Routledge, 1984).
27. Crenshaw, K. W. *On Intersectionality: Essential Writings* (The New Press, 2017).

Methods

Data

For our analysis, we used the complete set of 5.1 billion comments made on Reddit posts since comments were introduced in 2005 up to and including 2018. The dataset is publicly available and was downloaded from the pushshift.io Reddit archive[28] at http://files.pushshift.io/reddit/. All Reddit comments are public, and by posting on Reddit users consent to making their data freely available[29] ('By using the Services, you are directing us to share this information publicly and freely.'). For all 34.7 million Reddit commenters, our dataset contains their complete public commenting history, the communities (subreddits) their comments appeared in, and the timestamps associated with each comment. Over our entire study period, 52.9% of users commented in more than one subreddit, and the mean number of subreddits commented in by a user is 9.6, demonstrating that many users engage in the multi-community aspect of the platform. This activity provides crucial information about the behavioural similarity of subreddits, which we harnessed to create community embeddings and social dimensions.

Creating the community embedding

We used this Reddit commenting data to represent communities in a behavioural space using community embeddings, which were first proposed by Martin[9] and were subsequently refined by Kumar et al.[30] and Waller and Anderson[31]. Much like how word embeddings position words in a high-dimensional space such that similar words are nearby, community embeddings position communities in a high-dimensional space such that communities with similar memberships are close together in the space. The key difference is that community embeddings are learned solely from interaction data—high similarity between a pair of communities requires not a similarity in language but a similarity in the users who comment in them. Communities are then similar if and only if many similar users have the time and interest to comment in them both.

We created a community embedding from the Reddit data set using the open source software word2vecf (https://bitbucket.org/yoavgo/word2vecf/src), a modification of the original word2vec software to allow the usage of arbitrary contexts[32]. To generate our embedding, we applied the word2vec algorithm to interaction data by treating communities as 'words' and users as 'contexts'—every instance of a user commenting in a community becomes a word–context pair. For example, if user u_i commented in community c_j 10 times, the pair (u_i, c_j) would appear ten times in the training data. The model is then trained using the skip-gram with negative sampling (SGNS) method. To remove extremely small subreddits for which there are insufficient data to generate a meaningful vector representation, we restricted the analysis to the top 10,006 subreddits by number of comments, which accounts for 95.4% of all comments and 93.2% of all users. Since our training data are generated without using a context window or intermediate documents, in contrast with traditional word2vec, all word–context (community–user) pairs are included in our training data without restriction (analogous to using an infinite-sized context window).

The word2vec model has numerous hyperparameters that affect the training process and resulting embedding. To tune the model for the community embedding use case, we performed a grid search of the hyperparameter space, optimizing for performance on a set of community analogies. The hyperparameters we varied are: sample, the down-sampling threshold; negative, the number of negative examples; alpha, the starting learning rate; and size, the dimensionality of the resulting embedding. We added an additional parameter shuffled, a Boolean parameter which indicates whether the training data should be randomly shuffled prior to training. We assessed the model's performance on three sets of analogies: university subreddits to their corresponding cities; sports teams to their corresponding cities; and sports teams to their corresponding sport. By performing a grid search of the hyperparameter space, we found an embedding that solves 72% of the 4,392 analogies perfectly, and 96% of them nearly perfectly (correct answer in the top 5 communities). The resulting parameters from this process are alpha 0.18, negative 35, sample 0.0043, size 150, shuffled true. We believe that shuffling the data set prior to training prevents the model from over-fitting on temporal trends.

SGNS learns not only a vector for each word (in this case, each community) but a vector for each context as well (in this case, each user). While we only used the word (community) vectors in this paper, the context vectors play an important role in the training process. The training objective of the SGNS training procedure maximizes the dot product of word–context pairs that frequently co-occur, and minimize the dot product of randomly generated word-context pairs (negative examples). Intuitively, this suggests that communities with 'similar' users will end up with similar vectors, and users who participate in 'similar' communities will end up with similar vectors. However, this circular definition does not provide a concrete interpretation for the dot product of two community vectors. Levy and Goldberg[33] show that the SGNS objective is optimized by a factorization of the word–context pointwise mutual information (PMI) matrix (shifted by a constant). PMI is a measure of association between a word and a context, or, in our context, a measure of association between a community c and a user u, where 'Count' is the count of all matching comments:

$$\mathrm{PMI}(c,u) = log\frac{P(c,u)}{P(c)P(u)} = log\frac{\mathrm{Count}(c,u)\cdot \mathrm{Count}_{\mathrm{total}}}{\mathrm{Count}(c)\cdot \mathrm{Count}(u)}$$

Note that this matrix is dense, and in the common case where $\mathrm{Count}(c,u) = 0$, $\mathrm{PMI}(c,u) = -\infty$. In such a PMI matrix, the dot product of two community vectors is related to the similarity of their PMI values over all users:

$$\mathbf{c}_1 \cdot \mathbf{c}_2 = \sum_u \mathrm{PMI}(\mathbf{c}_1, u)\cdot \mathrm{PMI}(\mathbf{c}_2, u)$$

If SGNS was truly a pure factorization of the word-context PMI matrix, it would follow that this approximately holds in a community embedding as well. However, the iterative nature of the training procedure means that SGNS captures not only literal user overlap between communities but higher-order similarities as well. For example, if the two communities r/trucks and r/golf had no users in common, but both had a high overlap with the r/AskMen community, their vectors might end up somewhat close to each other despite no users being members of both communities. Indeed, empirical tests of SGNS and PMI demonstrate that SGNS is extremely capable of preserving second-order context overlap—even weighting this higher than first-order context overlap—while PMI is completely incapable of capturing it at all. In a simulation experiment performed by Schlechtweg et al.[34], the average cosine distance between words with first- and second- order context overlap were 0.11 and 0.00 respectively using SGNS and 0.51 and 1.0 using PMI. While matrix factorization of the PMI matrix is also able to capture such higher-order effects, Levy and Goldberg establish that in practice SGNS arrives at a different result than factorization of the PMI matrix, and that pure factorization does not perform well on many NLP tasks[33]. Thus, while deriving a closed-form equation that relates the cosine similarity of communities to their actual user overlap is still an unsolved problem, the architecture of the training process and empirical evidence suggests that cosine similarity of two community vectors is a strong measure of the similarity of the user-bases of the two communities.

We performed a clustering of the community embedding to understand Reddit's macroscale community structure. We used agglomerative clustering based on Euclidean distance to partition all communities into 30 clusters. We then manually labelled the clusters based on their dominant topic, for example, Movies and TV (n = 478), Music (n = 412), and Politics (n = 247). When more than one cluster has the same topical theme, we label them in descending order of size, for example, Hobbies

1 (n = 346) and Hobbies 2 (n = 201). Six clusters consist of communities with no clear theme, which we label General interest (1 through 6). To conserve space in Fig. 2, we merge the six General interest clusters into a single General interest row and the five Gaming clusters into a single Gaming row.

Finding social dimensions

Our methodological contribution is the idea and technique of finding social dimensions in community embeddings that correspond to social constructs. These dimensions allow us to compute scores that represent the social makeup of online communities. We first describe the generic algorithm for constructing social dimensions, then discuss the particular choices we made in our analyses. In the following sub-sections, we describe the computation of community scores and validate them against both internal and external sources.

To generate a social dimension that corresponds to a social construct, the analyst first identifies a seed pair of communities that differ primarily in the target construct. An ideal choice of seed is a pair of communities that are extremely similar except for a difference in the target social dimension. Note that the seed pair communities do not need to be at the extreme ends of the target dimension; they only need to differ primarily in the social construct.

Second, to ensure that the dimension is not overly tied to idiosyncrasies of the two seed communities, the seed pair (s_1, s_2) is algorithmically augmented with additional similar pairs of communities. Let k denote the desired total number of pairs, chosen by the analyst. We generated the set of all pairs of communities (c_1, c_2) such that $c_1 \neq c_2$ and c_2 is one of the 10 nearest neighbours to c_1. This is based on the aforementioned idea that we are looking for pairs of communities that are very similar, but differ only in the target concept. All pairs are ranked based on the cosine similarity of their vector difference with the vector difference of the seed pair $cos\,(\mathbf{s_2} - \mathbf{s_1}, \mathbf{c_2} - \mathbf{c_1})$. Additional pairs are then selected greedily. The most similar pair to the original seed pair that has no overlap in communities with the seed pair or any of the previously selected pairs is selected, and this process is repeated until $k-1$ additional pairs are selected, which results in the k pairs used to create the dimension.

Third, the vector differences of all k pairs are averaged together to obtain a single vector that robustly represents the desired social dimension. We also computed a complementary -ness version of the dimension by averaging the vector sums of all k pairs. This dimension represents similarity to the communities on both sides of the pairs.

In our analysis, we chose k = 10, which implies that $k-1$ = 9additional pairs are chosen to augment each seed pair. We tested with more and fewer than 10 pairs; fewer and axes appeared to be less robust, and more produced extremely similar axes (by cosine similarity and correlation between scores.) Using fewer pairs allows for conclusions to be drawn about more communities, so we opted for the fewest pairs with good robustness. We generated the set of all 100,060 non-trivial pairs of communities (c_1, c_2) with their 10 nearest neighbours and build dimensions as described above. For our gender dimension, we chose r/AskMen and r/AskWomen, personal discussion forums for men and women; for our age dimension, we chose r/teenagers and r/RedditForGrownups, personal discussion forums for teenagers and adults; and for our partisan dimension, we chose r/democrats and r/Conservative, two partisan American political communities. While we focused here on traditional forms of identity, the method is not inherently constrained to one-dimensional representations. For example, multiple gender dimensions could be generated to build a more complete analysis of gender. Extended Data Table 1 contains the 9 similar pairs automatically found for all the dimensions.

While the choice of seed is important, our dimension generation method is robust, as similar seed choices generate similar dimensions. To demonstrate this, we also generated a gender B dimension with r/Daddit and r/Mommit, parenting discussion forums for men and women; an age B dimension with r/AskMen and r/AskMenOver30, Q&A communities for men of all ages and men over 30; and a partisan B dimension with r/hillaryclinton and r/The_Donald, two partisan American political communities.

As an additional notion of identity, we generated an affluence dimension, choosing as seeds r/vagabond, a forum for houseless travellers, and r/backpacking, a more general interest travel community. We also generated three dimensions for concepts not necessarily related to traditional identity but relevant to Reddit as a platform: time, representing actual time from 2005 to the present; sociality, representing how discussion- and meetup-focused a community is; and edgy, representing provocation and antagonism (seeds can be found in Extended Data Table 1).

Computing community scores

Once a vector for a dimension has been obtained, all communities can be assigned a score on that dimension by simply projecting the normalized community vector $\mathbf{c}$ onto that vector: $\mathbf{c} \cdot \mathbf{d}$. The score of a community on a dimension is proportional to its average similarity with the right side minus its average similarity with the left side. This can be seen by noticing that the cosine similarity of a normalized community vector $\mathbf{c}$ with a social dimension with n normalized seed pairs $(A_1, B_1) \ldots (A_n, B_n)$ defined as $\mathbf{d} = \frac{1}{n}\sum (B_i - A_i)$ is the following:

$$cos\,(\mathbf{c}, \mathbf{d}) = \frac{\mathbf{c} \cdot \sum (B_i - A_i)}{n\|\mathbf{d}\|} = \frac{1}{n\|\mathbf{d}\|} \sum (\mathbf{c} \cdot B_i - \mathbf{c} \cdot A_i)$$

As the cosine similarity of two communities is related to the similarity between their memberships, a community's score on a dimension is reflective of how similar its membership is with the seeds at either pole. A community much more similar to one seed than the other will have a score at the poles, while a community equidistant between each of the seeds would receive a score of 0.

We calculated the scores for all 10,006 communities on all dimensions. The distributions of community scores for age, gender, partisan, and affluence can be found in Extended Data Fig. 1. Distributions of community scores transformed into percentiles on the age, gender, and partisan dimensions are provided in Fig. 2 (percentile score distributions are smoothed using LOESS with a smoothing span of 0.2 to reduce visual noise; raw score distributions are unsmoothed.) Distributions broken down by semantic cluster for age, gender, and partisan can be found in Extended Data Fig. 4 and for affluence, time, sociality, and edgy in Extended Data Fig. 5.

Scores for the aforementioned -ness dimensions are computed in the same fashion. Note that as a -ness dimension is formed from the sums of all pairs, the resulting scores simply reflect average similarity with the seeds on both sides. As a result, scores on -ness dimensions reflect association in general with a dimension. For example, both r/progressive, a community centred on the 'Modern Political and Social Progressive Movement', and r/LesbianGamers, a community for 'women who love women, who love gaming', are close to the left pole of the partisan axis (z = − 4.0, z = − 2.2), since they tend to have similar memberships as other communities on the left. However, r/progressive scores high on the partisan-ness axis (z = 4.4) whereas r/LesbianGamers scores low (z = − 1.2).

To demonstrate the robustness of the dimension generation method, we compared each of the age, gender, partisan axes with their B version. The age dimension is correlated with age B at r = 0.90; gender is correlated with gender B at r = 0.86; and partisan is correlated with partisan B at r = 0.55 ($P < 10^{-10}$ and n = 10,006 for all three correlations). These results demonstrate that community scores are robust to small changes in the input seeds. The partisan B dimension has a more moderate correlation than the other two. This is because partisan and partisan B capture slightly different concepts. For example, Trump was an outsider candidate and online Trump supporters displayed significantly different behaviour than the traditional online Republican

base. Therefore, using r/The_Donald as a seed generates a dimension that is more specific to Trump and his online supporters' interests, in contrast with using r/Conservative as a seed, which generates a dimension that more closely captures Republicanism in general. This emphasizes the importance of validating community scores using external constructs, as we do in the next section.

Validating community scores

We validated each of age, gender, partisan, and partisan-ness against the external concepts they represent. To validate the gender dimension, we compared the gender scores of occupation communities to the actual gender makeup of those occupations. We used gender makeup data from the 2018 American Community Survey, and manually match occupation descriptions to subreddit names (Supplementary Table 2.) We find there is a $r = 0.89$correlation between the percentage of women in an occupation and its communities' gender score ($n = 23$, two-sided $P < 10^{-8}$; Extended Data Fig. 2.) The gender dimension well represents the proportion of women in an occupation even for occupations at the extremes and in the middle. To validate the age dimension, we compared communities for universities and the communities for the respective cities, as universities tend to have a much younger population than a city as a whole. We find a very strong relationship between age and whether a community is associated with a university or a city ($r = 0.91$, two-sided $P < 10^{-58}$, $n = 150$, Cohen's $d = 4.37$). As shown in Extended Data Fig. 3, university communities skew far younger and city communities skew far older.

To validate the partisan dimension, we manually coded communities as left or right wing, and verify that the partisan score distinguishes between them. We selected communities that contain in their description either one of the left-wing terms 'democrat', 'clinton', 'left' or 'progressive', or one of the right-wing terms 'republican', 'trump', 'right' or 'conservative'. We then manually coded these communities based on their description into one of two categories: left-wing (or anti-right) and right-wing (or anti-left). Coding is performed strictly using these words and whether the description is supportive or against them. We coded 125 communities which contain one of these words and find 32 left-wing and 18 right-wing communities. The remainder were not labelled as there was no clear association in the description. We find that this label is strongly associated with the partisan score ($r = 0.92$, two-sided $P < 10^{-21}$, $n = 50$, Cohen's $d = 4.89$). We also used this labelling to validate the partisan-ness dimension. We compared the distribution of partisan-ness scores for the labelled left or right communities and find it is substantially different than that of all other communities (Cohen's $d = 3.27$).

We performed an additional validation using 2016 US Census data for the affluence and partisan dimensions. Reddit communities are matched to US Census metropolitan statistical areas (MSAs) by manual coding. We find that the median household income in a MSA is associated with the affluence score of MSA communities ($r = 0.42$, $n = 130$, two-sided $P < 10^{-7}$), and the Republican–Democrat vote differential in the 2016 presidential election (calculated for each MSA by combining county-level results from the MIT Election Lab) is associated with the partisan score of MSA communities ($r = 0.39$, $n = 112$, two-sided $P < 10^{-5}$; Extended Data Fig. 2). The presence of this correlation indicates that our method captures online partisanship, although see Extended Data Fig. 6 and the related discussion in the main text for more on how the online expression of US partisanship differs from its traditional offline analogue, including voting patterns in presidential elections.

Measuring relationships between dimensions

After validating the social scores, we measured the relationships between these dimensions as they exist on Reddit. We find a weak correlation exists between age and gender ($r = 0.10$); a moderate correlation exists between gender and partisan ($r = -0.29$); and a moderate correlation exists between age and partisan ($r = -0.37$; two-sided $P < 10^{-10}$ and $n = 10006$ for all dimension correlations). We repeated this analysis on the alternate B axes for robustness. We find similar relationships between partisan B and gender ($r = -0.34$), between partisan B and age ($r = -0.13$), between partisan and gender B ($r = -0.26$), and between partisan and age B ($r = -0.33$; two-sided $P < 10^{-10}$ and $n = 10{,}006$ for all dimension correlations). Extended Data Fig. 6 illustrates the relationships between partisan and age, partisan and gender, and partisan and partisan-ness.

Computing word scores

We additionally computed scores for words along all dimensions to provide context to our primary analyses. Word scores are weighted averages of community scores weighted by the number of times the word was used in a community in 2017. We excluded infrequent words, those that occur fewer than 10,000 times, and community-specific words, those with fewer than 5 bits of entropy in their distribution of usage over subreddits. To avoid distortion introduced by bots that re-use the same word over and over again in automated postings, we capped the number of usages of a word in a subreddit by one commenter that are counted at 100. Word scores represent the types of communities in which that word is likely to be observed. The words with the most extreme scores on each of our primary axes are available in Extended Data Fig. 1.

Measuring political polarization

To quantify political polarization on Reddit, we first restricted our focus only to 'explicitly political activity'—comments in political communities as defined by the partisan-ness axis. We chose a cut-off on the partisan-ness axis such that it is the highest value that includes 80% of the 'Politics' cluster. Using this cut-off to categorize communities as explicitly political, we label 553 (5.53%) of communities as political, and we find that it correctly categorizes 92% of the communities manually coded as 'political' by us based on their description in the previously described validation for the partisan dimension. For each political community, we calculated its partisan z-score z from its partisan score $\mathbf{c} \cdot \mathbf{d}$ and the mean and s.d. of the entire partisan distribution (including non-political communities): $z = \frac{(\mathbf{c} \cdot \mathbf{d}) - \mu}{\sigma}$. z represents the partisan association of a community, with a z of 0 indicating that a community has a partisan score equal to the overall mean (that is, it is in the centre), negative scores indicating a left-wing association, and positive scores indicating a right-wing association. The partisan z-score of a comment is equal to the partisan z-score of the community it was posted in.

We further restricted our attention to the 88.8% of political comments which have not been deleted. Deleted comments on Reddit are still visible, but their author is hidden. As we lack author data for these comments, we are unable to tell whether they were made by a new user or an existing user. Since one of our aims is to attribute changes in activity based on the prior political activity of users, we excluded these deleted comments from our political analyses. While deleted comments account for a small fraction of overall political activity, it is possible that deleted comments differ from non-deleted comments to such an extent that it affects our main findings. To assess whether such a difference exists, we compared the distribution of partisan scores of deleted comments Q to the distribution of partisan scores of non-deleted comments P. The distributions are extremely similar (Extended Data Fig. 7a); they have a difference of means of only −0.01 and a Kullback–Leibler divergence of $D_{\mathrm{KL}}(P||Q) = 0.033$ bits. We conclude that it is reasonable to exclude deleted comments from our analyses.

To measure the extent to which users self-select into partisan groups, we assigned all political communities one of five bins $B \in \{-2, -1, 0, 1, 2\}$ by z-score on the partisan axis (left wing (−2): $z < -2$, leaning left (−1): $-2 < z < -1$, center (0): $-1 < z < 1$, leaning right (1): $1 < z < 2$, right wing (2): $z > 2$). The proportion of all political activity that falls in each of these bins yields a discrete distribution of political activity on Reddit

(Fig. 3a, top). Within each bin b_1, we measured the likelihood that, if one randomly draws a comment in bin b_1, and then randomly draws one of its author's comments, the latter drawn comment falls in bin b_2. This measure is designed to give an idea of how much users that contribute to one bin contribute to the same or other bins and is equivalent to the average proportion of activity by authors in b_1 that takes place in the bin b_2. When $b_1 = b_2$, this can be interpreted as the average proportion of activity by authors in b_1 that takes place in the same bin. Let A denote the set of all authors. Let $c_{a,b}$ denote the number of comments made by author a in bin b. The average proportion of activity that takes place in bin b_2 by authors in b_1 is therefore:

$$f(b_1, b_2) = \frac{1}{\sum_{a \in A} c_{a,b_1}} \sum_{a \in A} c_{a,b_1} \frac{c_{a,b_2}}{\sum_{b \in B} c_{a,b}}$$

Notice that this quantity is weighted by the number of comments an author makes in a bin. Were authors not weighted by their number of comments, authors that make many comments in one bin and a non-zero but small number of comments in other bins would influence all distributions equally, making all distributions look artificially similar. We also computed this on the community level, where an individual community is substituted in the place of b_1. Results of the community-level analysis are shown in Extended Data Fig. 7c.

If each users' individual distribution over partisan was equal to the overall distribution, that is, there was no self-selection into partisan groupings, each bin's distribution would be approximately equal to the overall activity distribution (Fig. 3a, top). In such a scenario, where all users had the same likelihood to contribute to a bin, we would still expect to observe slightly more average activity in the 'same bin' in the above analysis due to two factors: one, in order to be included in the calculation for a bin a user must have contributed to it and therefore that users with no activity in a bin are excluded from its calculation, and two, since we chose the bins based on score on the partisan axis, communities within a bin are more similar to each other than average communities, and similarity in the embedding is correlated with user overlap. To show these effects are negligible, we repeated this analysis on a random dataset, generated by randomly shuffling all of the authors of Reddit comments. Since the userbases of all communities are similar in this random dataset, community vectors tend to be similar in the resulting embedding. As a result, there is far less variation in partisan score among political communities in this embedding, making it impossible to use the previous method of labelling communities by partisan affiliation by standard deviations from the mean. We instead created a best approximation to the conditions in the real embedding by selecting the same number of political communities (that is, we took the 553 communities with the highest partisan-ness scores as 'political') and then dividing these communities into five bins of the same number of communities as in the actual embedding by choosing the appropriate thresholds on the partisan axis in the random embedding. This accomplishes the goal of selecting bins with similar partisan scores to put an upper bound on the possible effects of the aforementioned confounds. The results in this random dataset show that all bin distributions are extremely similar to the overall distribution with a small (less than 0.85%) increase in the average percent for the same bin, showing that the overall activity distribution is an accurate reference point for what bin distributions would look like were there no self-selection into partisan groupings.

Measuring dynamics of polarization

To measure how platform-level polarization has changed over time, we measured the distribution of political activity on the partisan axis over time. Again focusing on only the subset of non-deleted comments in explicitly political communities, we quantified the distribution of partisan scores each month. Fig. 3b displays the distribution of the partisan scores of comments each month. As a direct measure of the partisan polarization of the distribution, we also computed the average absolute partisan z-score $|z|$ of activity in each month, that is, the average number of standard deviations from the mean partisan score, for each month (Fig. 5a). Note that we used the average absolute z-score and not the absolute average z-score. Using the absolute average z-score, equal amounts of activity in the far left and far right would average out to zero and be considered non-polarized. As we wished to capture the extent to which activity takes place in polarized communities regardless of polarity, we used the average absolute z-score. As an alternate metric, Extended Data Fig. 7b displays the proportion of activity that takes place in very left- and right-wing communities in any given month; very left-wing communities are those with a z-score less than −3 (42 communities), while very right-wing communities are those with a z-score greater than 3 (24 communities).

To measure the extent to which individuals have moved towards partisan extremes as they act on the platform, and the extent to which this has contributed to the overall platform polarization observed, we analysed the distribution of political activity of users with different levels of past activity. We divided all Reddit users active in political communities in seven cohorts by the year they made their first comment in political communities. To measure the average polarization of a cohort's activity, we used the average absolute partisan z-score $|z|$. Fig. 4a illustrates the average absolute z-score of each cohort's activity over time. As an alternate way to visualize the relationship between users' past and present activity, we plot a version of Fig. 3b broken down by users' prior political activity in Extended Data Fig. 8.

We computed two alternate measures of a user's time on the platform to provide a comparison point for the above analysis. For each comment in a political community, we computed the number of calendar months since the author's account was created, a, and the number of distinct calendar months the author has been active in political communities up to the point of the comment's posting, b. We grouped political activity by a and b and calculate the average absolute z-score for these comments. Insets in Fig. 4a display the relationship between a (left) or b (right) and the average absolute z-score of political comments.

To determine how common it is for users to significantly polarize in activity, we compared the same user's political activity in different calendar months. For each month and each user, we calculated the average partisan score of their activity in that month (that is, the average partisan score of the communities they participated in, weighted by the number of comments they made in each community). We computed these scores only for user–month pairs with at least ten comments to minimize noise; results for other choices of threshold are similar and can be found in Extended Data Fig. 7f. Extended Data Fig. 7d shows the Pearson correlation coefficient between the average partisan scores of a user in any pair of months. Figure 4b shows the proportion of users whose average absolute partisan score increased by 1× s.d. for any pair of months. Results for other choices of threshold can be found in Extended Data Fig. 7e.

To measure the effect of individual-level patterns on overall platform polarization, we calculated the average absolute z-score of political activity of new and existing users, and compare these levels of polarization year-over-year. A user is considered 'new' at the time of posting a comment if they have no prior activity in political communities 12 months ago or prior. Let C_t denote the set of all comments in time t. Let E_t denote the set of comments made by existing users in time t, where a comment c is considered to be made by an existing user if the author of the comment made their first comment in any political community at least 12 months prior to posting c. Let N_t denote the set of comments made by new users in time t, that is, all comments not made by existing users ($N_t = C_t \setminus N_t$). Let $\overline{z}(C) = \frac{1}{|C|} \sum_{c \in C} |z(c)|$ denote the average absolute z-score of comments in set C. The change in average polarization of activity from time $t-1$ to t is equal to $\overline{z}(C_t) - \overline{z}(C_{t-1})$. A natural metric to examine the change in average polarization of, for example, new

users would use a similar quantity, like $\bar{z}(N_t) - \bar{z}(N_{t-1})$. Such a quantity, however, does not itself say anything about platform-level change in polarization, as it does not take into account what proportion of overall activity is made by new or existing users, and whether that proportion itself changed between time periods. In addition, it is an awkward comparison as the new users at time t can be existing users at time $t+1$. We instead used $\Delta n_t = \frac{|N_t|}{|C_t|}(\bar{z}(N_t) - \bar{z}(C_{t-1}))$ to measure the change in overall polarization attributable to new users. This represents what the overall change in polarization would have been had the activity of existing users been at the same average level of polarization as that of the platform 12 months prior to when their comment was made. The remaining change is attributable to new user activity and is therefore termed Δn. Similarly, $\Delta e_t = \frac{|N_t|}{|C_t|}(\bar{z}(E_t) - \bar{z}(C_{t-1}))$ measures the change in polarization attributable to existing users, that is, what the overall change in polarization would have been had the activity of new users been at the same average level of polarization as that of the platform 12 months prior. This definition also has the desirable property that Δn and Δe add up to the overall change in polarization, that is, $\Delta n_t + \Delta e_t = \bar{z}(C_t) - \bar{z}(C_{t-1})$:

$$\begin{aligned}\Delta n + \Delta e &= \frac{|N_t|}{|C_t|}(\bar{z}(N_t) - \bar{z}(C_{t-1})) + \frac{|E_t|}{|C_t|}(\bar{z}(E_t) - \bar{z}(C_{t-1}))\\ &= \frac{|N_t|}{|C_t|}\bar{z}(N_t) + \frac{|E_t|}{|C_t|}\bar{z}(E_t) - \frac{|N_t| + |E_t|}{|C_t|}\bar{z}(C_{t-1})\\ &= \bar{z}(C_t) - \bar{z}(C_{t-1})\end{aligned}$$

Fig. 4c illustrates the values of Δe and Δn for each year in our data.

To measure whether polarization patterns differ between left-wing and right-wing activity on the platform, we repeated some of the above analyses on two subsets of our data: left-wing activity (including only comments in communities with $z \leq -1$) and right-wing activity (including only comments in communities with $z \geq 1$). We repeated the above change in polarization analysis on the two subsets of data; results can be found in Fig. 5c, d. We repeated the author year-of-first-political-comment analysis on the two subsets of data; results can be found in Fig. 5e, f.

To measure the possible effect of an 'implicit polarization' process, by which users are influenced by implicitly political subreddits that rank low on the partisan-ness axis but are highly polarized on the partisan axis, we performed an analysis of the relationship between explicitly partisan and implicitly partisan activity. Examining the 9,453 non-explicitly political communities, we labelled communities as 'implicitly political' if they have a partisan-ness score below our cut-off but a partisan score at least 2 standard deviations to the left or right of the global mean in a similar manner to the partisan bins B defined earlier. We used the sets of explicitly and implicitly partisan communities to examine the relationship between the time users become active in either of them. Let $m_I(u)$ denote the month that a user u was first active in any implicitly partisan community. Let $m_E(u)$ denote the month that a user u was first active in any explicitly partisan community. Extended Data Fig. 10 shows the relationship between m_I and m_E considering both only left-wing activity (left) and right-wing activity (right). Of users who were first active in an explicitly partisan community at time m_E, the proportion of them who were first active in an implicitly partisan community at time m_I is denoted by the colour in cell (m_E, m_I). The line graphs at the top show the total proportion of users who were active in implicitly partisan communities in a calendar month prior to when they were active in an explicitly partisan community (that is, the proportion of users for whom $m_I < m_E$). This corresponds to the proportion of users for which it would be possible for an 'implicit polarization' effect to apply (as it is not possible for an implicit polarization effect to apply if implicitly political activity did not precede explicitly political activity), given that a time granularity of one month is used.

Reporting summary

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

Data availability

All data are available from the pushshift.io Reddit archive[28] at http://files.pushshift.io/reddit/. Source data are provided with this paper. Reddit community embedding, social dimension vectors and community scores are available at https://github.com/CSSLab/social-dimensions.

Code availability

All code is available at https://github.com/CSSLab/social-dimensions. Analyses were performed with Python v3.7, pandas v1.3.3 and Spark v3.0.

28. Baumgartner, J., Zannettou, S., Keegan, B., Squire, M. & Blackburn, J. The Pushshift Reddit dataset. In *Proc. International AAAI Conference on Web and Social Media* **14**, 830–839 (2020).
29. Reddit privacy policy *Reddit* https://www.redditinc.com/policies/privacy-policy (2021).
30. Kumar, S., Hamilton, W. L., Leskovec, J. & Jurafsky, D. Community interaction and conflict on the web. In *Proc. 2018 World Wide Web Conference* 933–943 (2018).
31. Waller, I. & Anderson, A. Generalists and specialists: using community embeddings to quantify activity diversity in online platforms. In *Proc. 2019 World Wide Web Conference* 1954–1964 (2019).
32. Levy, O. & Goldberg, Y. Dependency-based word embeddings. In *Proc. 52nd Annual Meeting of the Association for Computational Linguistics* **2**, 302–308 (2014).
33. Levy, O. & Goldberg, Y. Neural word embedding as implicit matrix factorization. *Adv. Neural Inf. Process. Syst.* **27**, 2177–2185 (2014).
34. Schlechtweg, D., Oguz, C. & im Walde, S. S., Second-order co-occurrence sensitivity of skip-gram with negative sampling. Preprint at https://arxiv.org/abs/1906.02479 (2019).

Acknowledgements This research was supported by the National Sciences and Engineering Research Council of Canada (NSERC), the Canada Foundation for Innovation (CFI) and the Ontario Research Fund (ORF).

Author contributions I.W. performed the computational analysis. A.A. and I.W. designed the research, analysed the results and wrote the paper.

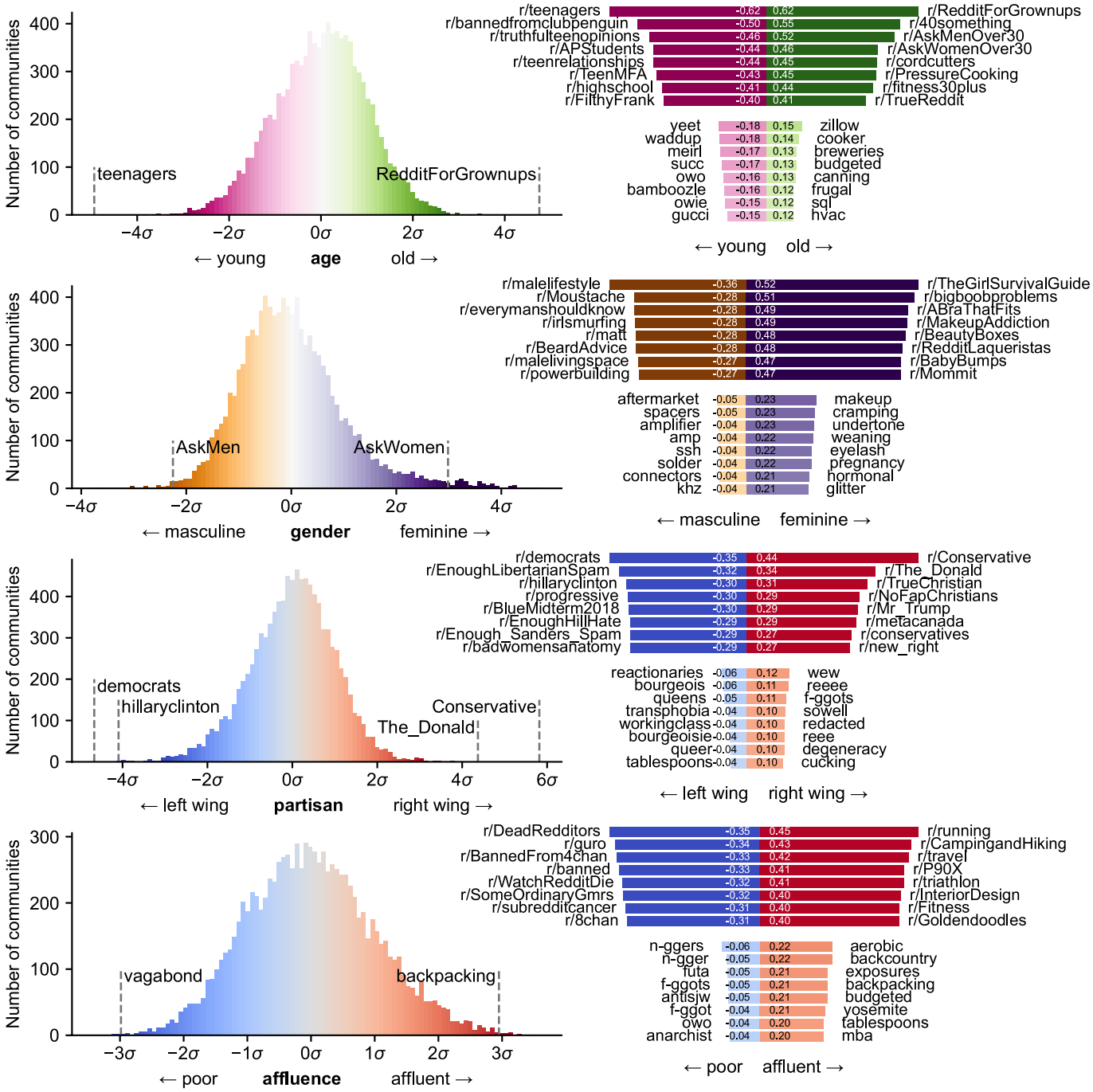
Competing interests The authors declare no competing interests.

Additional information

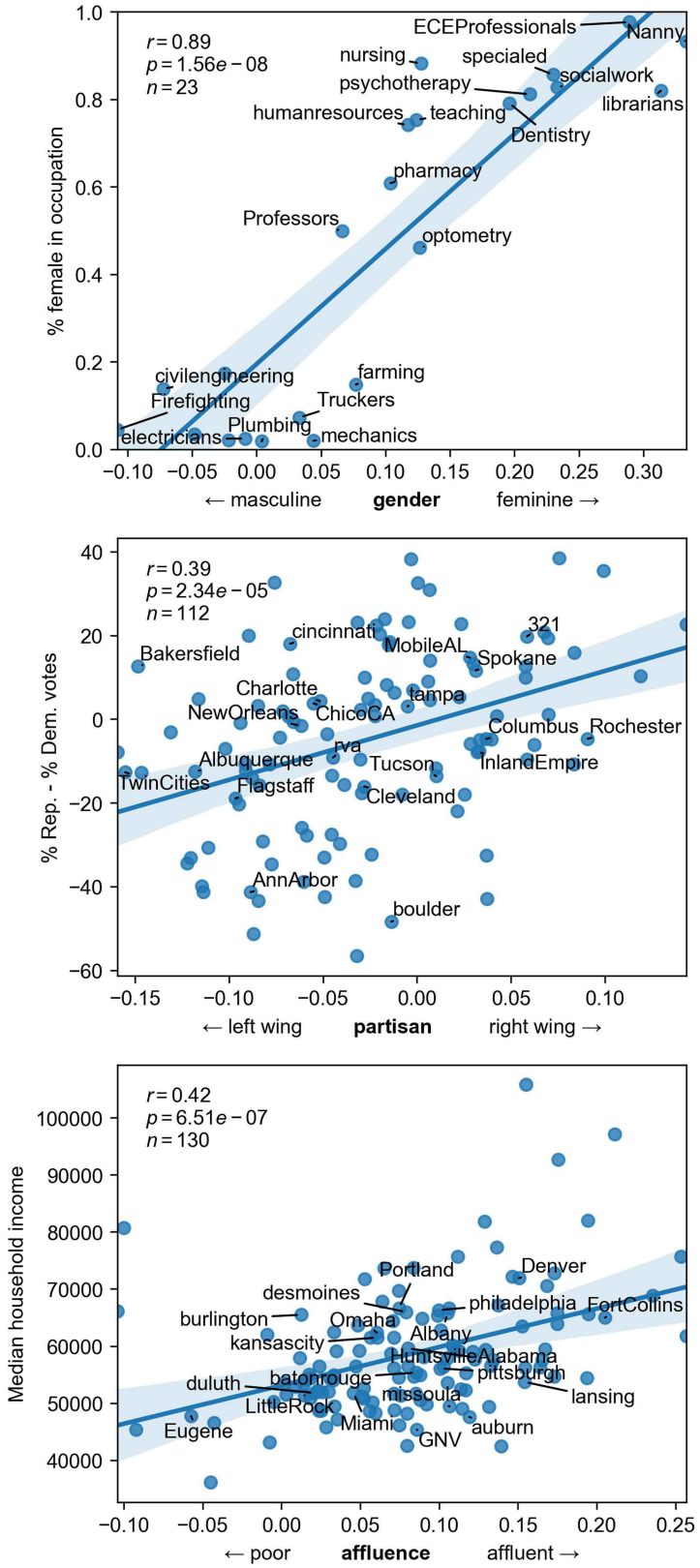
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41586-021-04167-x.

Correspondence and requests for materials should be addressed to Ashton Anderson.

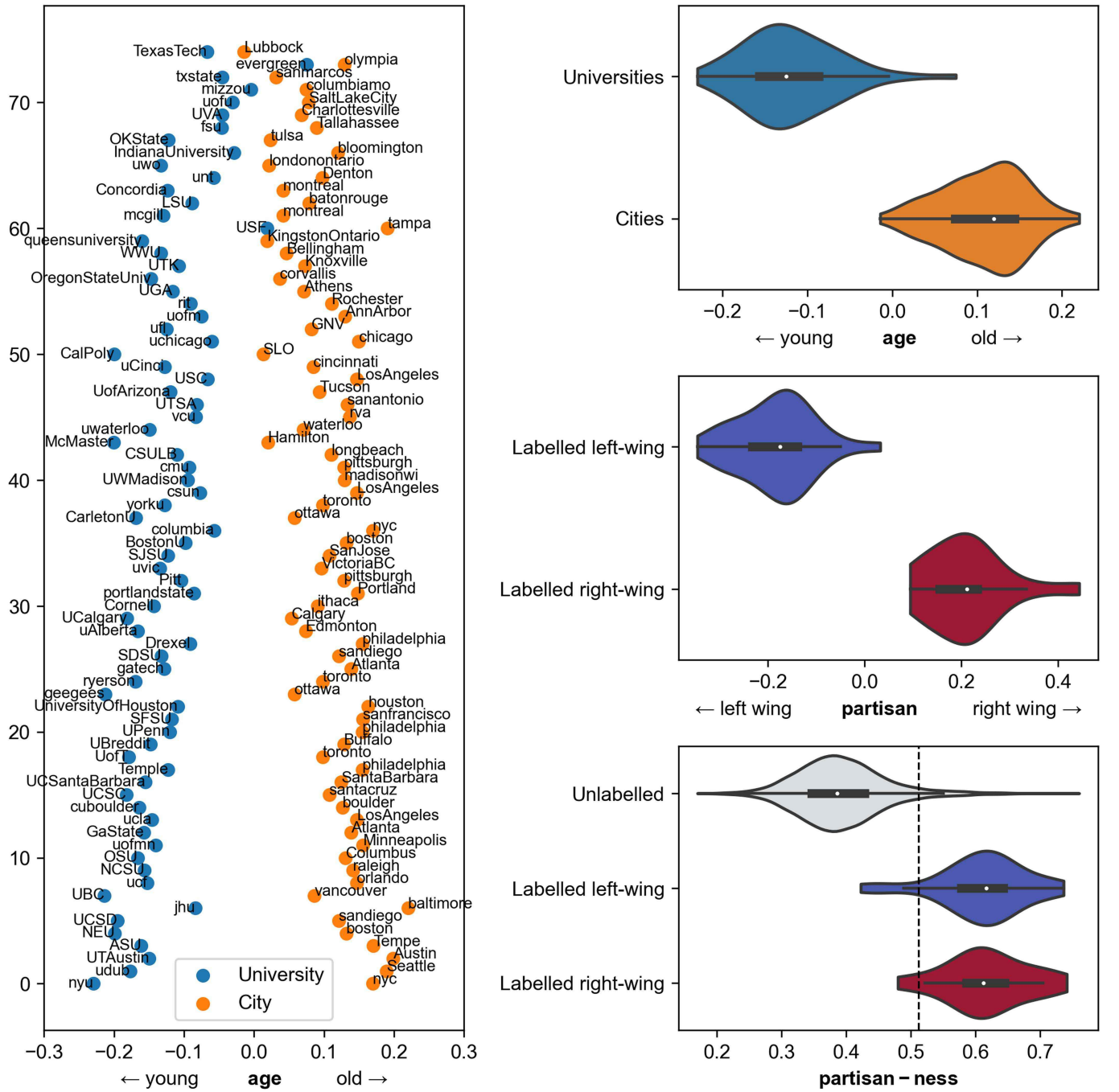
Peer review information *Nature* thanks Kenneth Benoit, Kate Starbird and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.



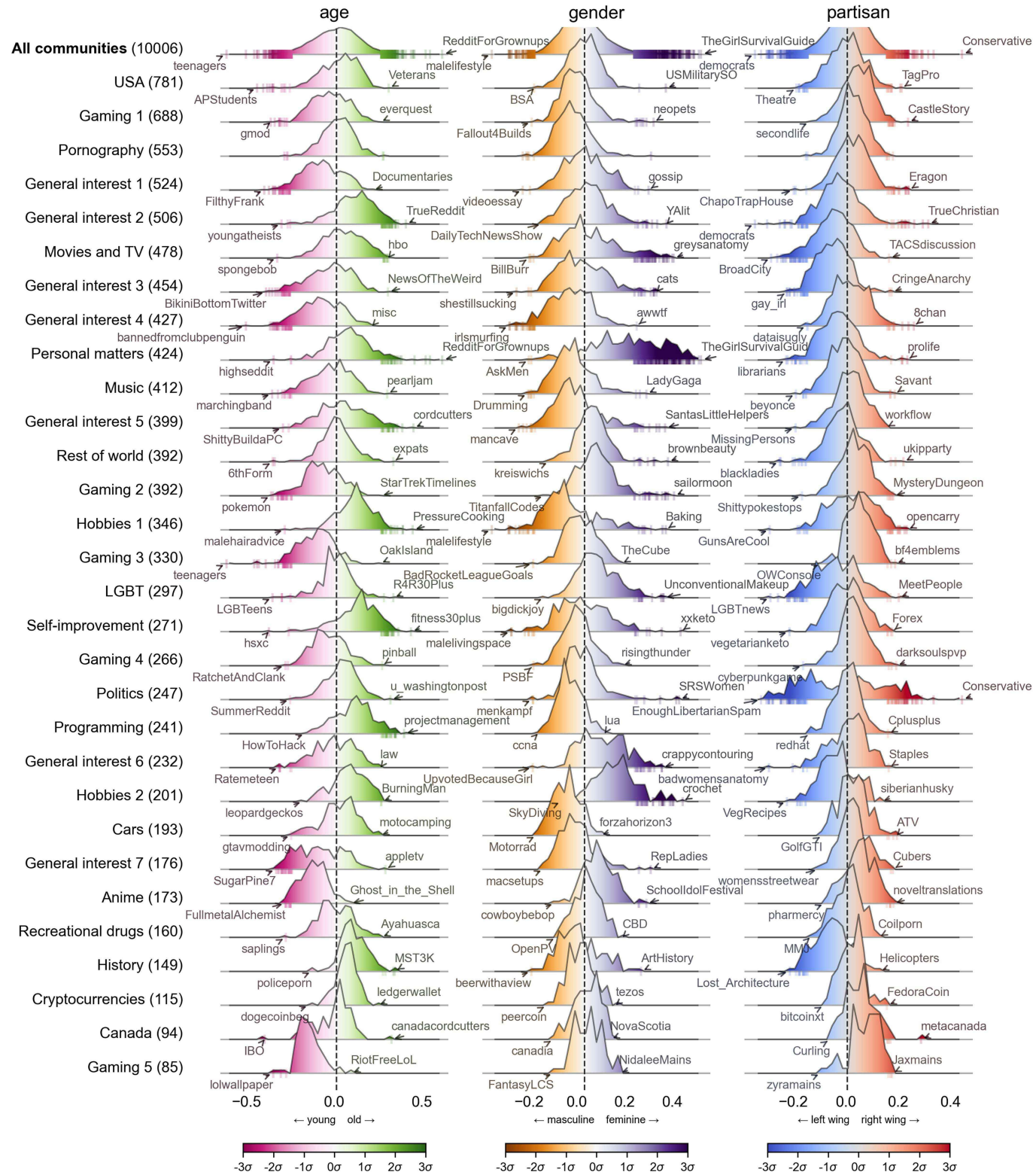
Extended Data Fig. 1 | Distribution of community scores. Left: distributions of communities on the age, gender, partisan, and affluence dimensions. Right: the most extreme communities and words on those dimensions. Word scores are calculated by averaging community scores weighted by the number of occurrences of the word in the community in 2017. Community descriptions can be found in the glossary (Supplementary Table 1).



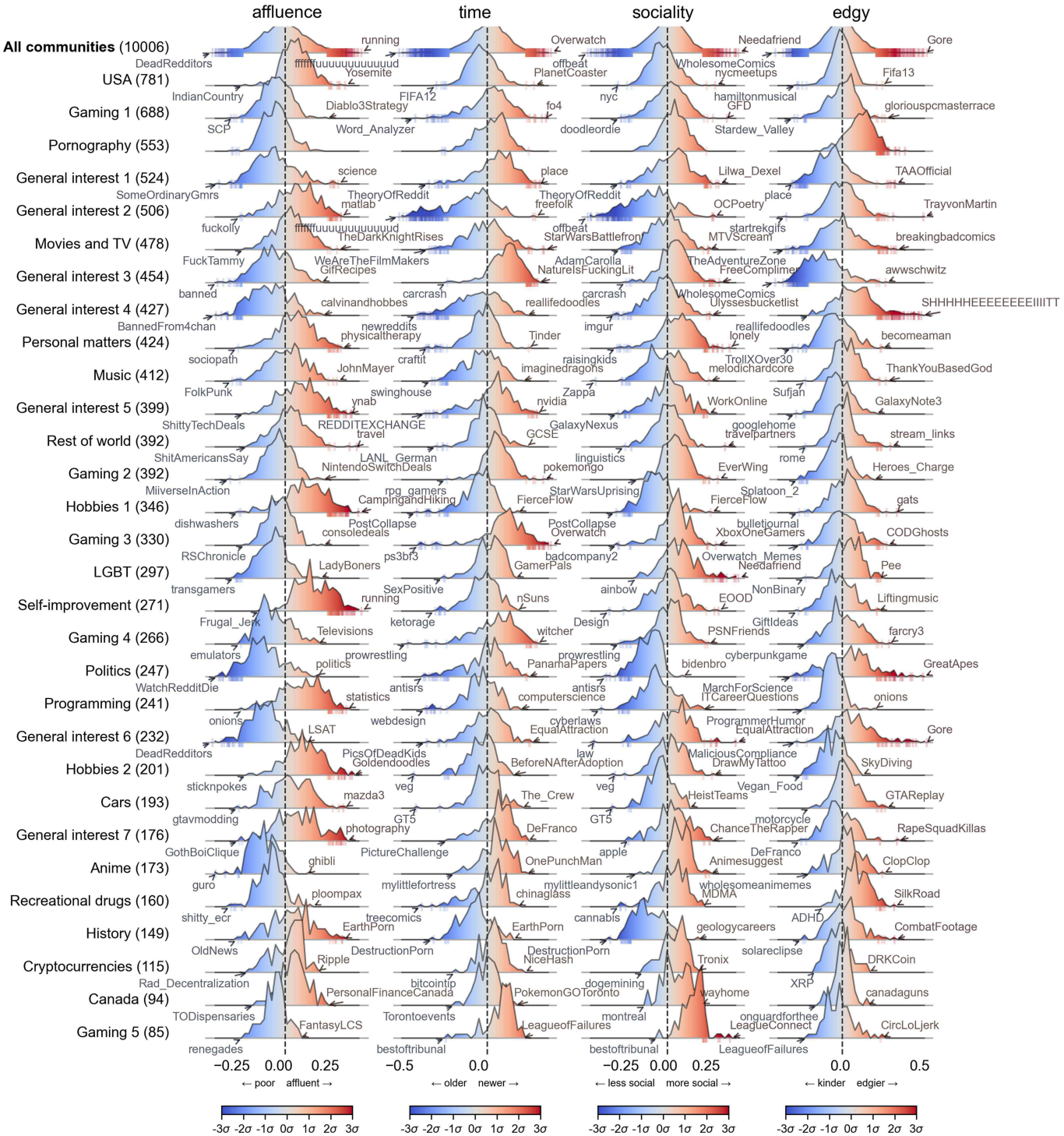
Extended Data Fig. 2 | External validations of social dimensions. Scatter plots of the external validations of the gender, partisan, and affluence axes. The gender scores for occupational communities are plotted against the percentage of women in that occupation from the 2018 American Community Survey. The partisan scores for city communities are plotted against the Republican vote differential for that metropolitan area in the 2016 presidential election. The affluence scores of city communities are plotted against the median household income for that metropolitan area from the 2016 US Census. The blue line is the best-fit linear regression for the data; the shaded area represents a 95% confidence interval for the regression estimated using a bootstrap. *p*-values for correlation coefficients computed using two-sided test of Pearson correlation assuming joint normality.



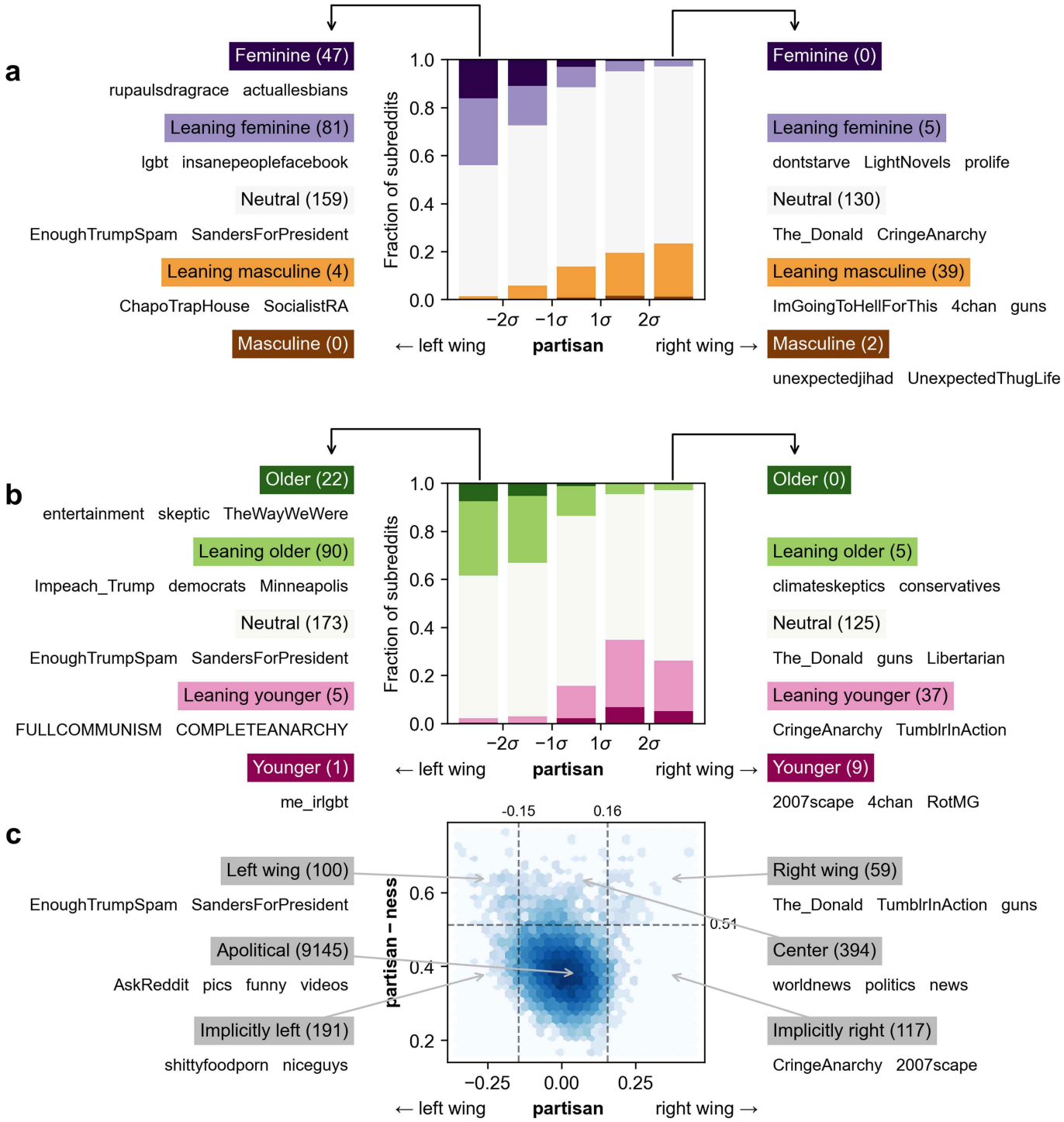
Extended Data Fig. 3 | Further validations of social dimensions. Clockwise from left: The gap between university and city communities on the age dimension. The distribution of university and city communities on the age dimension; age is strongly related to label ($r = 0.91$, two-sided $p < 10^{-58}$, $n = 150$, Cohen's $d = 4.37$). The distribution of left and right wing labelled communities on the partisan dimension; partisan is strongly related to label ($r = 0.92$, two-sided $p < 10^{-21}$, $n = 50$, Cohen's $d = 4.89$). The distribution of explicitly labelled left- and right-wing communities on the partisan-ness axis as compared to the general distribution; there is a large difference in their means (Cohen's $d = 3.27$). For violin plots, white dot represents median; box represents 25th to 75th percentile; whiskers represent 1.5 times the inter-quartile range; and density estimate ('violin') extends to the minima and maxima of the data. p-values for correlation coefficients computed using two-sided test of Pearson correlation assuming joint normality.



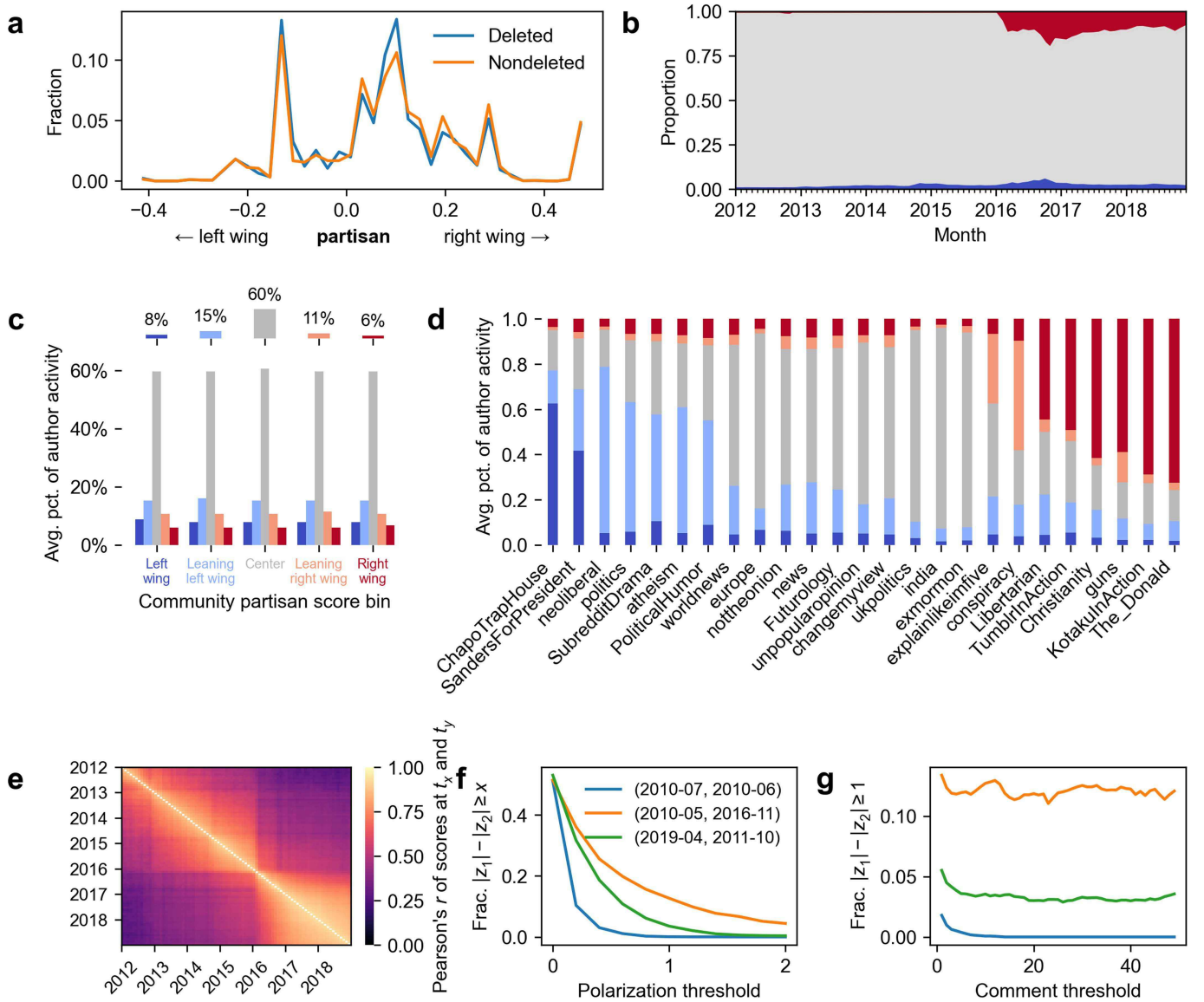
Extended Data Fig. 4 | Distributions of age, gender and partisan scores by cluster. Distributions of raw age, gender and partisan scores, separated by cluster. Outlier communities that lie more than two standard deviations from the mean are annotated. Dashed lines represent the global mean on each dimension. Community descriptions can be found in the glossary (Supplementary Table 1).



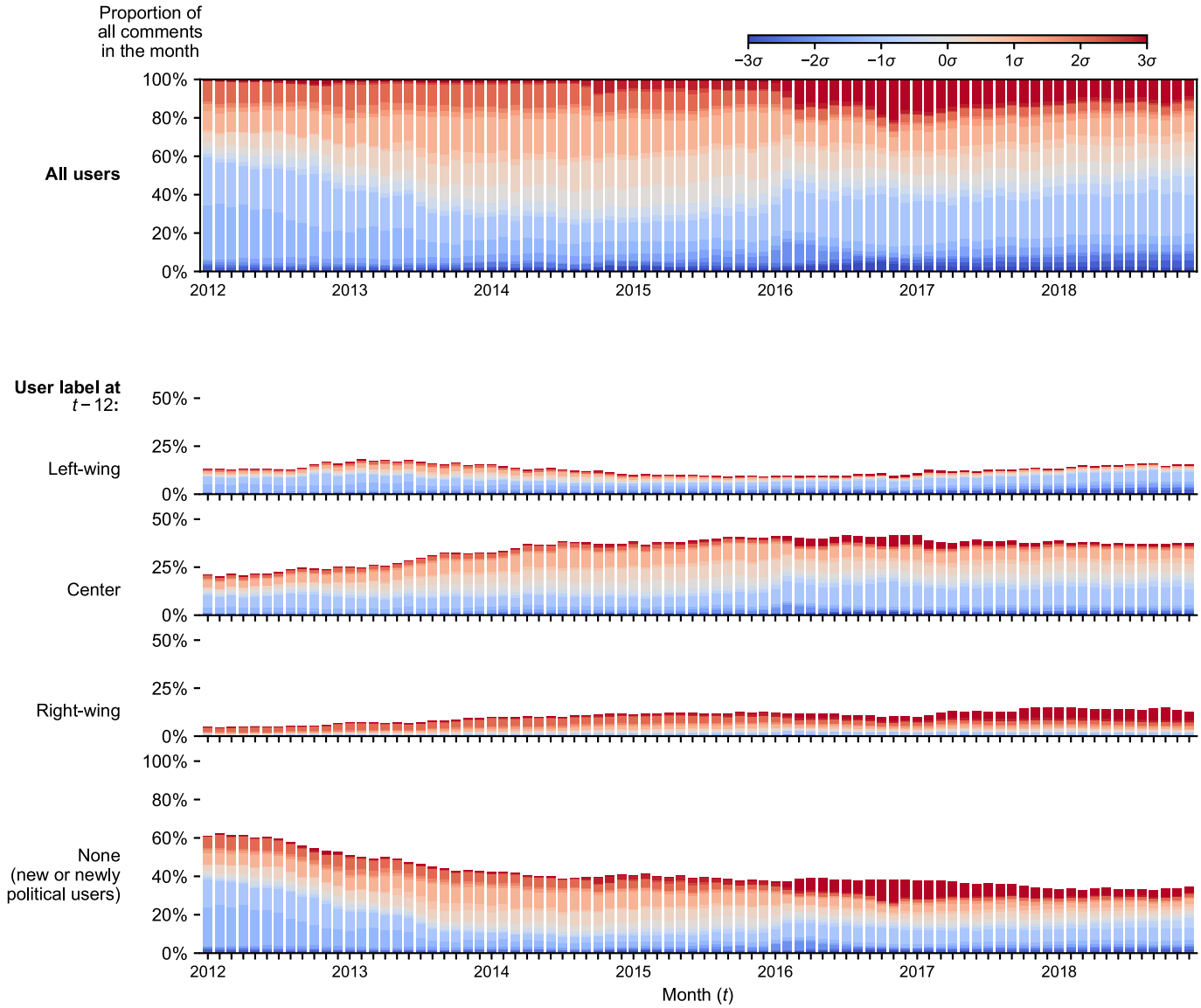
Extended Data Fig. 5 | Distributions of affluence, time, sociality and edgy scores by cluster. Outlier communities that lie more than two standard deviations from the mean are annotated. Dashed lines represent the global mean on each dimension. Community descriptions can be found in the glossary (Supplementary Table 1).



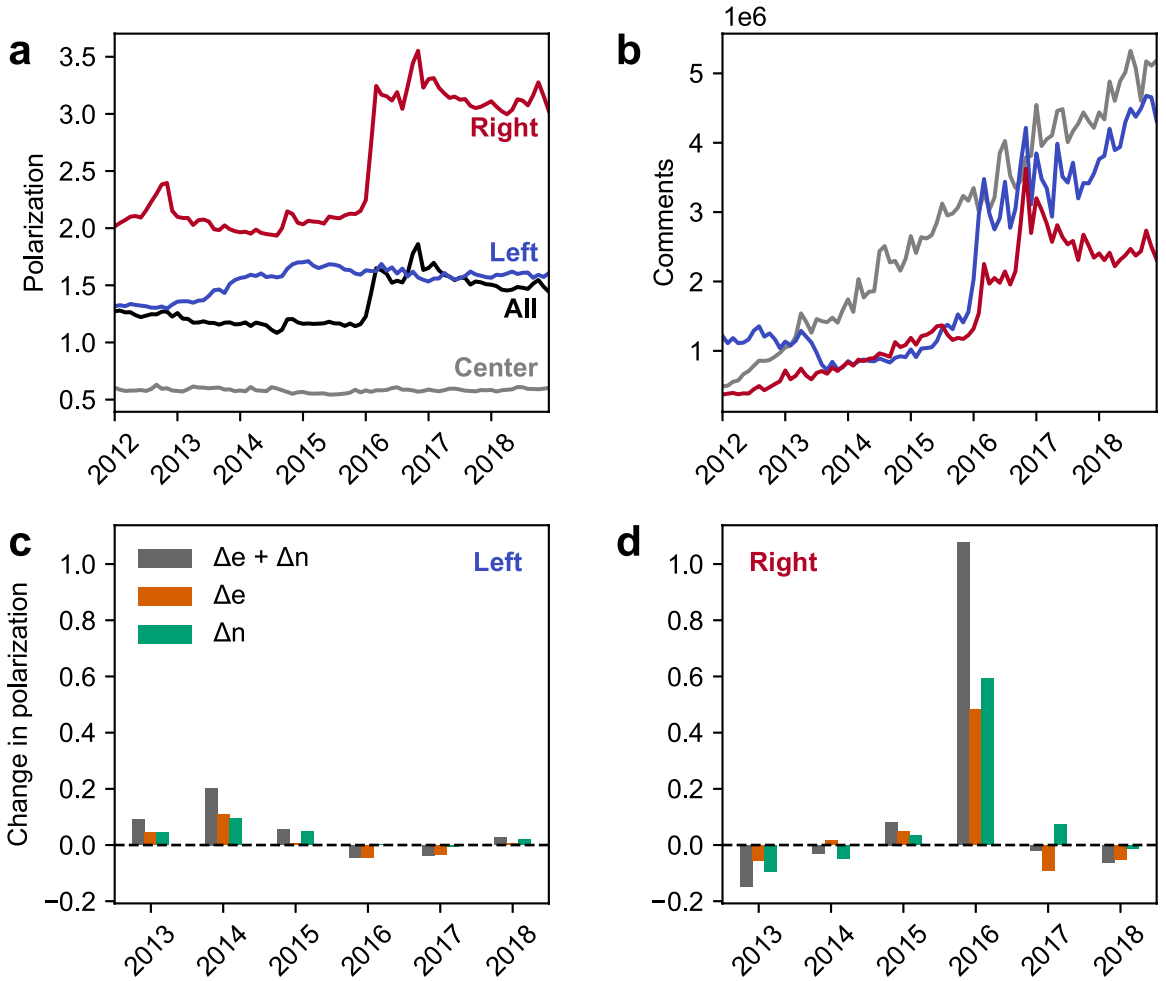
Extended Data Fig. 6 | Relationships between online social dimensions. The relationships between the partisan dimension and (a) gender, (b) age, (c) partisan-ness. Every bar represents a bin of communities with partisan scores a given number of standard deviations from the mean, and the distribution illustrates the scores on the secondary dimension (e.g. gender in (a)). From left to right, the bars represent highly left-wing, leaning left-wing, center, leaning right-wing, highly right-wing communities. The leftmost and rightmost bars are annotated with the number of communities, and examples of the largest communities, in each group. The hex-plot in (c) illustrates the joint distribution of partisan and partisan-ness scores. Labels correspond to the categorizations used in the polarization analysis.



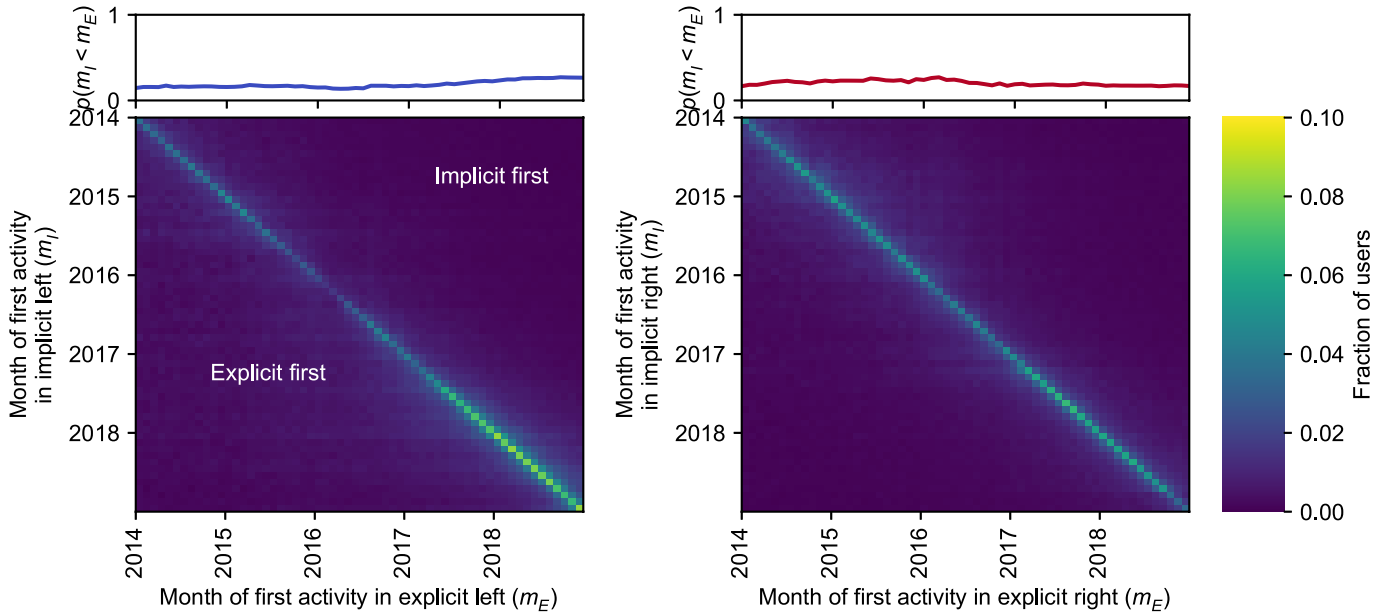
Extended Data Fig. 7 | Polarization robustness checks. (a) The partisan distribution of deleted and non-deleted comments in political communities. **(b)** The proportion of activity that took place in very left-wing ($z < -3$) and very right-wing ($z > 3$) communities over time. **(c)** Alternate version of Fig. 3a generated using a dataset in which the authorship of all comments was randomly shuffled. Each individual bin distribution is extremely similar to the overall activity distribution, showing that the overall activity distribution is a useful reference point for what bin distributions would look like if there were no tendency for users to comment in ideologically homogeneous communities. **(d)** Average distributions of political activity for authors of comments in the 25 largest political communities on Reddit (by number of comments). **(e)** Correlation of users' average partisan scores over time. Each (x, y) cell represents the correlation between scores of a user in month t_x and that same user in month t_y, for all users active in both time periods. A user is only considered active if they make at least 10 comments in a month. **(f)** The relationship between the proportion of users who polarize and the polarization threshold. The polarization threshold is the number of standard deviations a user must increase in polarization to be considered polarized. Three lines are plotted corresponding to three pairs of months; the pairs of months with the minimum (blue), maximum (orange), and median (green) proportion of users polarized when using a threshold of 1. A threshold of 1 is used in all other calculations. **(g)** The relationship between the proportion of users who polarize and the comment threshold. The comment threshold is the value used to filter inactive users from the calculation. Users must have at least x comments in each of the two months to be included in the calculation of the proportion of users who polarize. The same three month pairs are plotted as in part (e). There are minimal differences between different thresholds. A threshold of 10 is used in all other calculations.



Extended Data Fig. 8 | Distribution of political activity by user group. The distribution of political activity on Reddit over time by partisan score. Each bar represents one month of comment activity in political communities on Reddit, and is coloured according to the distribution of partisan scores of comments posted during the month (the partisan score of a comment is simply the partisan score of the community in which it was posted.) The top plot includes all activity as in Fig. 3b, while the four following plots decompose this into the subsets of activity authored by particular groups of users. Users are classified based on the average partisan score of their activity in the month 12 months prior–into left-wing (having a score at least one standard deviation to the left), right-wing (one standard deviation to the right), or center. Users with no political activity in the month 12 months prior use the label of the most recent month more than 12 months prior in which they had political activity; if they have never had political activity before, they fall into the new / newly political category (bottom).



Extended Data Fig. 9 | Additional measures of ideological asymmetry. **(a)** Average polarization (absolute *z*-score) of activity in different ideological categories over time. **(b)** Volume of activity (number of comments) in different ideological categories over time. **(c, d)** Annual change in polarization in the two partisan activity categories, decomposed into the change attributable to new (Δn) and existing (Δe) users as done in Fig. 4.



Extended Data Fig. 10 | Implicit polarization. The relationship between explicitly partisan and implicitly partisan activity (left: left-wing activity; right: right-wing activity.) Of users who were first active in an explicitly partisan community at time m_E, the proportion of them who were first active in an implicitly partisan community at time m_I is denoted by the colour in cell (m_E, m_I). The line graphs at the top show the total proportion of users who were active in implicitly partisan communities before they were active in an explicitly partisan community (i.e. the sum of each column below the diagonal back to 2005, or the total proportion of users for whom $m_I < m_E$).

Extended Data Table 1 | Social dimension seeds

age						teenagers	→	RedditForGrownups
youngatheists	→	TrueAtheism	teenrelationships	→	relationship_advice	AskMen	→	AskMenOver30
saplings	→	eldertrees	hsxc	→	running	trackandfield	→	trailrunning
TeenMFA	→	MaleFashionMarket	bapccanada	→	canadacordcutters	RedHotChiliPeppers	→	pearljam
gender						AskMen	→	AskWomen
TrollYChromosome	→	CraftyTrolls	AskMenOver30	→	AskWomenOver30	OneY	→	women
TallMeetTall	→	bigboobproblems	daddit	→	Mommit	ROTC	→	USMilitarySO
FierceFlow	→	HaircareScience	malelivingspace	→	InteriorDesign	predaddit	→	BabyBumps
partisan						democrats	→	Conservative
GunsAreCool	→	progun	OpenChristian	→	TrueChristian	GamerGhazi	→	KotakuInAction
excatholic	→	Catholicism	EnoughLibertarianSpam	→	ShitRConservativeSays	AskAnAmerican	→	askaconservative
askhillarysupporters	→	AskTrumpSupporters	liberalgunowners	→	Firearms	lastweektonight	→	CGPGrey
affluence						vagabond	→	backpacking
hitchhiking	→	hiking	DumpsterDiving	→	Frugal	almosthomeless	→	personalfinance
AskACountry	→	travel	KitchenConfidential	→	Cooking	Nightshift	→	fitbit
alaska	→	CampingandHiking	fuckolly	→	gameofthrones	FolkPunk	→	IndieFolk
age B						AskMen	→	AskMenOver30
AskWomen	→	AskWomenOver30	AskAnAmerican	→	RedditForGrownups	androidthemes	→	googleplaydeals
cringepics	→	ghettoglamourshots	windmobile	→	canadacordcutters	geegees	→	ontario
waterpolo	→	Yosemite	gatech	→	OMSCS	saplings	→	eldertrees
gender B						daddit	→	Mommit
predaddit	→	BabyBumps	TallMeetTall	→	bigboobproblems	parentsofmultiples	→	breastfeeding
BeardAdvice	→	NoPoo	freemasonry	→	pagan	matt	→	DrunkOrAKid
Leathercraft	→	sewing	ketodrunk	→	xxketo	techwearclothing	→	womensstreetwear
partisan B						hillaryclinton	→	The_Donald
GamerGhazi	→	KotakuInAction	SandersForPresident	→	HillaryForPrison	askhillarysupporters	→	AskThe_Donald
BlueMidterm2018	→	PoliticalHumor	badwomensanatomy	→	ChoosingBeggars	PoliticalVideo	→	uncensorednews
liberalgunowners	→	Firearms	GrassrootsSelect	→	DNCleaks	GunsAreCool	→	dgu
sociality						nyc	→	nycmeetups
law	→	LSAT	paris	→	travelpartners	sanfrancisco	→	SFr4r
boston	→	bostonhousing	Zappa	→	stonerrock	conan	→	NewGirl
ClashOfClans	→	EverWing	answers	→	findareddit	xbox360	→	XboxOneGamers
edgy						memes	→	ImGoingToHellForThis
watchpeoplesurvive	→	watchpeopledie	MissingPersons	→	MorbidReality	twinpeaks	→	TrueDetective
pickuplines	→	MeanJokes	texts	→	FiftyFifty	startrekgifs	→	DaystromInstitute
subredditoftheday	→	SRSsucks	peeling	→	Gore	rapbattles	→	bestofworldstar
time						PS3	→	PS4
xbox360	→	xboxone	battlefield3	→	Battlefield	blackops2	→	blackops3
deadisland	→	dyinglight	ps3bf3	→	battlefield_4	prowrestling	→	WWE
fo3	→	fo4	wii	→	wiiu	counter_strike	→	GlobalOffensive

Community pairs used to calculate social dimensions. The blue highlighted pair is the initial seed provided to the algorithm. The rest of the pairs are algorithmically found as described in Methods.

nature portfolio

Corresponding author(s): Waller

Last updated by author(s): Sep 28, 2021

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

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

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

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	No software was used for data collection as we used the pushshift.io data archive.
Data analysis	Analyses were performed with Python 3.7, pandas 1.3.3, and Spark 3.0, along with custom code available at https://github.com/CSSLab/socialdimensions

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

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

Our analysis uses the pushshift.io data archive of publicly accessible Reddit comment data. All data used can be obtained directly from the pushshift.io data archive at http://files.pushshift.io/reddit/.

Field-specific reporting

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

☐ Life sciences ☒ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Behavioural & social sciences study design

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

Study description	A quantitative observational analysis of community commenting patterns on Reddit. User-community comment frequencies are analyzed to quantify community relationships and changes in commenting patterns over time.
Research sample	We use a large data trace consisting of the commenting activity of the entire population of Reddit users, so our sample is the complete set of Reddit users. This data is representative of the population of Reddit as it is complete. This data is chosen to lend insight into overall platform-wide dynamics of Reddit, which is of interest due to its status as a major social media platform.
Sampling strategy	No sampling strategy was used, as we used complete data for the entire population of Reddit users.
Data collection	Data was collected programmatically by Pushshift, which was then collected in the Reddit archive files that we used. For each comment made during the study period, we use the username of the author of the comment, the name of the subreddit (community) in which it was posted, and the time it was posted.
Timing	The archive contains all comments posted on Reddit from June 2005 (when comments were introduced) to the end of 2018. The data was collected by Pushshift on a continuous basis from 2015 onwards.
Data exclusions	No data was excluded from the analysis.
Non-participation	Not relevant to our observational data set.
Randomization	Not relevant to our observational data set.

Reporting for specific materials, systems and methods

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

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☐	☒ Human research participants
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Human research participants

Policy information about studies involving human research participants

Population characteristics	See above.
Recruitment	Data was collected from anyone who contributed public comments to Reddit during the study period. All data is publicly available, and all users consent to the public sharing of contributed content. The Reddit Privacy Policy states "When you submit content (including a post, comment, chat message, or RPAN broadcast) to a public part of the Services, any visitors to and users of our Services will be able to see that content, the username associated with the content, and the date and time you originally submitted the content." [...] "By using the Services, you are directing us to share this information publicly and freely."
Ethics oversight	Not applicable.

Note that full information on the approval of the study protocol must also be provided in the manuscript.