Community embeddings reveal large-scale cultural organization of online platforms

Isaac Waller* and Ashton Anderson

Department of Computer Science, University of Toronto

{walleris,ashton}@cs.toronto.edu

* Corresponding author

September 2020

Abstract

Optimism about the Internet's potential to bring the world together has been tempered by concerns about its role in inflaming the "culture wars". Via mass selection into like-minded groups, online society may be becoming more fragmented and polarized, particularly with respect to partisan differences. However, our ability to measure the cultural makeup of online communities, and in turn understand the cultural structure 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 cultural dimensions by leveraging large-scale patterns of aggregate behaviour. Applying our methodology to 4.8B Reddit comments made in 10K communities over 14 years, we find that the macro-scale community structure is organized along cultural lines, and that relationships between online cultural concepts are more complex than simply reflecting their offline analogues. Examining political content, we show Reddit underwent a significant polarization event around the 2016 U.S. presidential election, and remained highly polarized for years afterward. Contrary to conventional wisdom, however, instances of individual users becoming more polarized over time are rare; the majority of platform-level polarization is driven by the arrival of new and newly political users. Our methodology is broadly applicable to the study of online culture, and our findings have implications for the design of online platforms, understanding the cultural contexts of online content, and quantifying cultural shifts in online behaviour.

1 Introduction

In 1962, Marshall McLuhan proclaimed "The new electronic interdependence recreates the world in the image of a global village." In the decades since, there has been fierce debate about the Internet's dual forces of cultural integration, as the world becomes increasingly interconnected, and cultural fragmentation, as people can more easily select into like-minded communities [28, 30, 12]. Twenty years into the widespread adoption of social platforms, it remains unclear how online communities are culturally organised. Of particular concern is whether online populations sort into homogeneous "echo chambers" in which discussion is becoming increasingly polarized, and whether online platforms tend to shift users towards cultural and ideological extremes [13, 9, 2].

Since online social platforms consist of massive amounts of unstructured, pseudonymous data, empirically quantifying the cultural makeup of online communities, and in turn the cultural structure of online platforms, poses a unique challenge. Traditional network approaches quantify structure internal to online platforms, such as viral cascades, but typically do not connect online activity with core notions of individual identity such as age or gender [31, 17]. Surveys and interviews can shed light on the makeup of certain groups, but don't scale to the thousands of communities and millions of users that comprise online platforms.

We develop and validate a methodology to quantify the positioning of online communities along cultural dimensions by leveraging large-scale patterns of aggregate behavior. Using neural community embeddings, which represent similarities in community membership as relationships between vectors in a high-dimensional space, we are able to accurately measure the cultural makeup of online communities. Focusing on traditional notions of identity—age, gender, and U.S. political affiliation—and applying our method to 4.8B comments in 10K communities on Reddit, one of the world's largest social platforms, we investigate three related questions: (i) To what extent are online communities organized along traditional cultural lines?, (ii) Do online cultural relationships reflect their offline analogues?, and (iii) Is political activity on Reddit becoming increasingly polarized, and if so, how?

Our approach differs from prior work examining cultural structure and polarization in online platforms in three main ways. Most importantly, our methodology quantifies the cultural makeup of communities in a purely behavioral fashion. Communities are similar to each other if and only if their user bases are surprisingly similar; by computing this similarity along a cultural dimension (e.g. U.S. 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 (e.g. the left or right wing of U.S. politics). Users "vote with their feet" to decide the cultural orientation of communities: only action, across large numbers of people, matters. In this way, we avoid biases that result from self-reported data, expert labels, and survey-based methods. Self-reported data can be subject to social desirability bias, where users may over-report or under-report community memberships according to their perceived social value [29], as well as response bias, as those who opt to report cultural affiliations may differ systematically from the population as a whole [1]. Expert labels of community affiliations are subject to labeler bias, and may be less accurate for communities with implicit cultural associations. Additionally, community leaders could be incentivized to manipulate the perception of their community's cultural alignment, and thus the community descriptions they write may not accurately represent the makeup of their membership. Leveraging the aggregate behavior of community members rather than relying on labeled data avoids this problem. Survey methods can suffer from researcher bias, as they may only test for associations that researchers are aware of or interested in [8], as well as social desirability bias in how participants choose to respond [20]. Our method quantifies the positions of all communities along our chosen cultural axes, enabling a complete analysis of cultural structure that could uncover previously unknown associations.

Relatedly, our work builds upon a line of research examining cultural relationships in word embeddings. High-dimensional representations of text, where similar words are positioned close together in the space, have been applied to the study of cultural stereotypes [16, 5, 6] and the cultural markers of class [19]. Although our dataset is composed of billions of comments, we do not use the text to create our embedding. Differences in cultural identity are reflected in the words people use, but this relationship is relatively weak for our focus on measuring the cultural orientation of underlying community populations. Communities using similar language may still be culturally distinct, e.g. if they discuss similar topics from different viewpoints, and communities with distinct language may still be culturally similar, e.g. if they have similar memberships but discuss different topics, or discuss even the same topic in a different language. This discrepancy between the cultural makeup of online communities and the language they use may be especially large for *implicit* associations. For example, our behavioural method scores the membership of the cycling community, dedicated to 'discussion of everything bicycle related', as left-wing, even though explicitly left-wing vocabulary is not particularly prominent in their discussions.

Finally, previous analyses have studied platforms such as Facebook, Twitter, and Amazon, on which users are guided by algorithmic curation and personalized recommendations [26, 11, 9]. As such, traces of user activity on these platforms reflect not only natural human choices but also the influence of underlying algorithms. A recent focus has been on examining the effects of algorithmic curation on shaping online cultural structure, e.g. measuring the prevalence of algorithmic "filter bubbles" of homogeneous content and groups [23, 15], but user choices may play even a larger role in shaping this structure [3]. 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. For many years, there were no algorithmic recommendations directing users to new communities on Reddit [24]. Even when introduced, they were mostly unpersonalized and relegated to less visible parts of the site design. By and large, when users discover and join communities, they do so through their own exploration—the content of what they see is not algorithmically adjusted to match their previous behaviour. Since the user experience on Reddit is relatively untouched by algorithmic personalization, the patterns of community memberships we observe are more likely the result of user choices, and thus reflective of the cultural structure induced by natural online behaviour.

2 Data and methods

Reddit. We analyze discussion forum data from Reddit, one of the world's largest online social platforms. Reddit is composed of tens of thousands of communities, or "subreddits", each of which is typically centered around a single topic or shared interest. Subreddits are collections of posts, each of which can either be a short piece of text initiating a discussion or an external link, and users can comment on others' posts. For our analysis, we use the complete set of 5.1B comments made on Reddit posts since comments were introduced in 2005 up to and including 2018 [4]. For all 34.7M Reddit commenters, our dataset contains their complete public commenting history, the communities their comments appeared in, and the timestamps associated with each comment.

To study the macro-scale structure of the platform, we use and extend *community embeddings*. 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 similar communities are close together in the space. The key difference is that our community embedding is 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. To generate our embedding, we applied the Word2Vec algorithm to interaction data by treating communities as "words" and commenters as "contexts"—every instance of a user commenting in a community becomes a word-context pair. Communities are then similar if and only if many similar users have the time and interest to comment in them both (a detailed discussion of what similarity entails in this context can be found in Appendix B.) We embedded the largest 10,006 communities by number of comments, which account for 95.4% of all Reddit comments, into a 150-dimensional space and optimized the embedding with community analogies. For visualization purposes, we apply hierarchical clustering to the cosine similarities between communities to group them into 18 coherent clusters (see Appendix B for more details).

Finding cultural axes. Analogously to how previous research uncovered axes in word embeddings that correspond to gender, class, and affluence [19, 5, 16], we develop a methodology to find dimensions in our community embedding that correspond to age, gender, and U.S. political partisanship. To do so, we first identify a seed pair of communities that differ in the target cultural concept, but are similar in other respects. We seed our `gender` axis with AskMen and AskWomen, question-and-answer forums for men and women; our `age` axis with teenagers and RedditForGrownups, personal discussion forums for teenagers and adults; and our `partisan` axis with democrats and Conservative, two partisan American political communities. (Descriptions of every community we reference can be found in

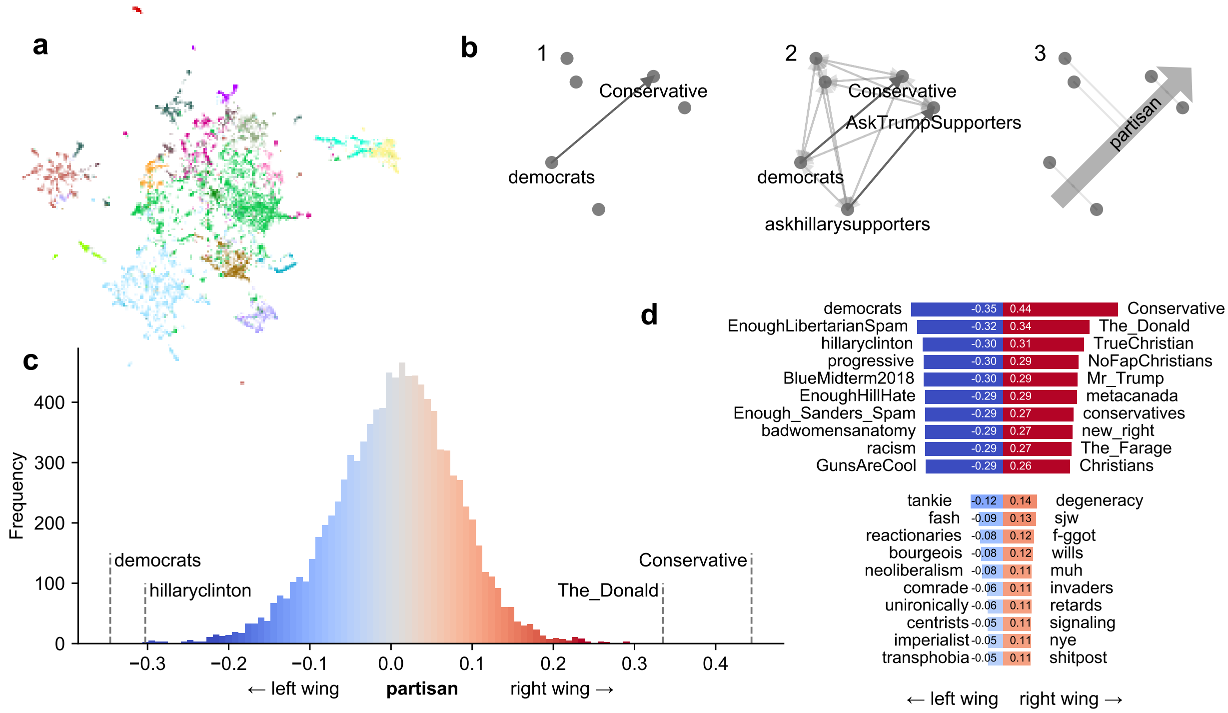

Figure 1: **Quantifying cultural dimensions on Reddit.** **a.** A two-dimensional UMAP projection of the 10,006 communities in our community embedding of Reddit, with points coloured by semantic clusters found by hierarchical clustering. **b.** An illustration of the cultural axis generation process. First, an input pair representing the desired cultural notion is provided to the algorithm; for example, democrats and Conservative are provided to obtain a vector that represents a difference in American partisan affiliation (1). All other pairs between similar subreddits are computed to produce a pool of candidate directions, and directions that are extremely similar to the input pair direction are selected to augment the seed pair; for example, askhillarysupporters and AskTrumpSupporters, two Q&A communities for supporters of the respective 2016 U.S. presidential candidates, is selected given this input pair (2). The average of the differences between all selected pairs becomes the single axis representing the desired cultural dimension (3). **c.** The distribution of the partisan scores for the 10,006 most popular Reddit communities. Communities vary from far left-wing to far right-wing, and are coloured by z-score. **d.** Top: communities most associated with the left- and right-wing ends of the axis (for community descriptions, see Appendix I.) Bottom: words most associated with the left- and right-wing ends of the axis, considering only word usages in political communities in 2017 as quantified by the `partisan-ness` axis (see Figure 3.)

Appendix I.) To more robustly capture cultural differences along these dimensions as they are expressed on the platform, we automatically augment these seeds with similar pairs of communities. For each dimension, we select the 9 pairs with the most similar vector difference from the set of all pairs of very similar communities (see Table A1 for a list of selected pairs.) The resulting set of 10 seed vector differences are then averaged together to generate the final dimensions corresponding to each target concept (Figure 1b). The method generalizes to more concepts than we study here; see Appendix C for details.

Every community can then be positioned along a cultural dimension by projecting the community's vector representation onto the dimension. This is equal to the average similarity of the focal community with communities on the right side of the seed pairs minus the 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. As the cosine similarity of two communities is related to the similarity between their memberships, a community's score on an axis is reflective of how similar its membership is with the seeds at either pole. The distribution of community scores along the `partisan` dimension is approximately Gaussian, and varies between the extreme left-wing and extreme right-wing on Reddit (Figure 1c). The words most associated with the left and right poles illustrate how political discussion differs across the `partisan` spectrum (Figure 1d).

We validate that scores on these axes closely correspond to their target cultural concepts by demonstrating that scores are highly correlated with external ground-truth estimates. For the `gender` axis, we show that the positions of occupation-based subreddits, such as pharmacy and Carpentry, are strongly correlated with the gender proportion of workers in those occupations in a national survey ($r = 0.89$). For the `partisan` dimension, we analyze communities that explicitly mention their partisan affiliation in their description and verify that their `partisan` scores reflect their affliation ($r = 0.92$; Cohen's $d = 4.89$). We also verify that the partisan scores of city-based subreddits are correlated with the Republican vote differential in the 2016 U.S. presidential election ($r = 0.39$). For the `age` axis, we observe that communities for universities, e.g. UofT, are consistently far younger than the community for the corresponding city, e.g. Toronto ($r = 0.91$, Cohen's $d = 4.37$). Plots and full methodology for these validations are available in Appendix E. While these validations suggest that the dimensions are correlated with real-world identities, we emphasize that they are measures of cultural associations, not individual characteristics. A community's position on the gender axis, for example, should not be interpreted as a direct measure of the gender makeup of the community, but instead reflects its association with the cultural construct of masculinity or femininity as expressed on Reddit.

We also generate secondary dimensions that represent the strength of association with each primary dimension, which we term `partisan-ness`, `gender-ness`, and `age-ness`. These secondary dimensions are calculated by taking the sum of the seed pairs' vectors, instead of the difference, and measuring similarity to *both* ends of the 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. Both progressive, a community centered on the “Modern Political and Social Progressive Movement”, and 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, progressive scores high on the `partisan-ness` axis ($z = 4.4$) whereas LesbianGamers scores low ($z = -1.2$), reflecting the difference in how overtly political the activity of their memberships are. We validate the `partisan-ness` dimension by examining the same explicitly-labeled communities as in the `partisan` validation, and we find that communities that label themselves as political have far higher `partisan-ness` scores than other communities (Cohen's $d = 3.27$; see Appendix E).

3 Results

The distributions of Reddit communities along the `age`, `gender`, and `partisan` axes reveal significant intra- and inter-topic diversity (Figure 2). Entire top-level groups of communities found by hierarchical clustering skew towards poles on the cultural dimensions, reflecting the aggregate behaviour of their members. For example, Programming communities skew male (mean z-score = -0.58) and older ($\bar{z} = 0.80$), Cars communities skew male ($\bar{z} = -0.72$) and right-wing ($\bar{z} = 0.30$), and Personal issues communities skew female ($\bar{z} = 1.72$) and left-wing ($\bar{z} = -0.45$). Additionally, there is substantial cultural diversity within each topical group 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 (> 2 std. dev. from the mean) on both sides of 0. The intra-group variation of a group corresponds to the degree of heterogeneity the group displays. For example, on the `gender` dimension, the Hobbies and Personal Issues groups have high variance, whereas the USA and Gaming groups have low variance. There are high- and low-variance groups in each dimension, and this variance is generally only weakly correlated with group size (age: $r = 0.65$, gender: $r = 0.10$, partisan: $r = -0.05$). The stratification of community clusters on these dimensions demonstrates that the macro-scale structure of Reddit is organized along cultural lines.

We find that relationships between age, gender, and partisan affiliation on Reddit do not simply reflect their offline analogues. Focusing on the `partisan` axis, we quantified its relationship with the `gender`, `age`, and `partisan-ness` axes (Figure 3). There is a strong and significant relationship between the `partisan` and `gender` dimensions (Figure 3a). Communities that skew towards the feminine pole also skew left-wing, and more masculine-associated communities skew right-wing ($r = -0.29$). This is consistent with the American electorate; in the 2016 U.S. presidential election, men voted for Trump by a share of 52 to 41, and women voted for Clinton by a share of 54 to 39 [7]. The relationship is monotonically increasing in `partisan` score. Every group of communities in a particular band of `partisan` scores has a significantly larger fraction of masculine communities than the one to the left of it. The relationship is also strongest at the `partisan` extremes. The most left-wing communities are 44.0% feminine leaning and 1.4% masculine leaning, whereas the most right-wing communities are 23.3% masculine leaning and 2.9%

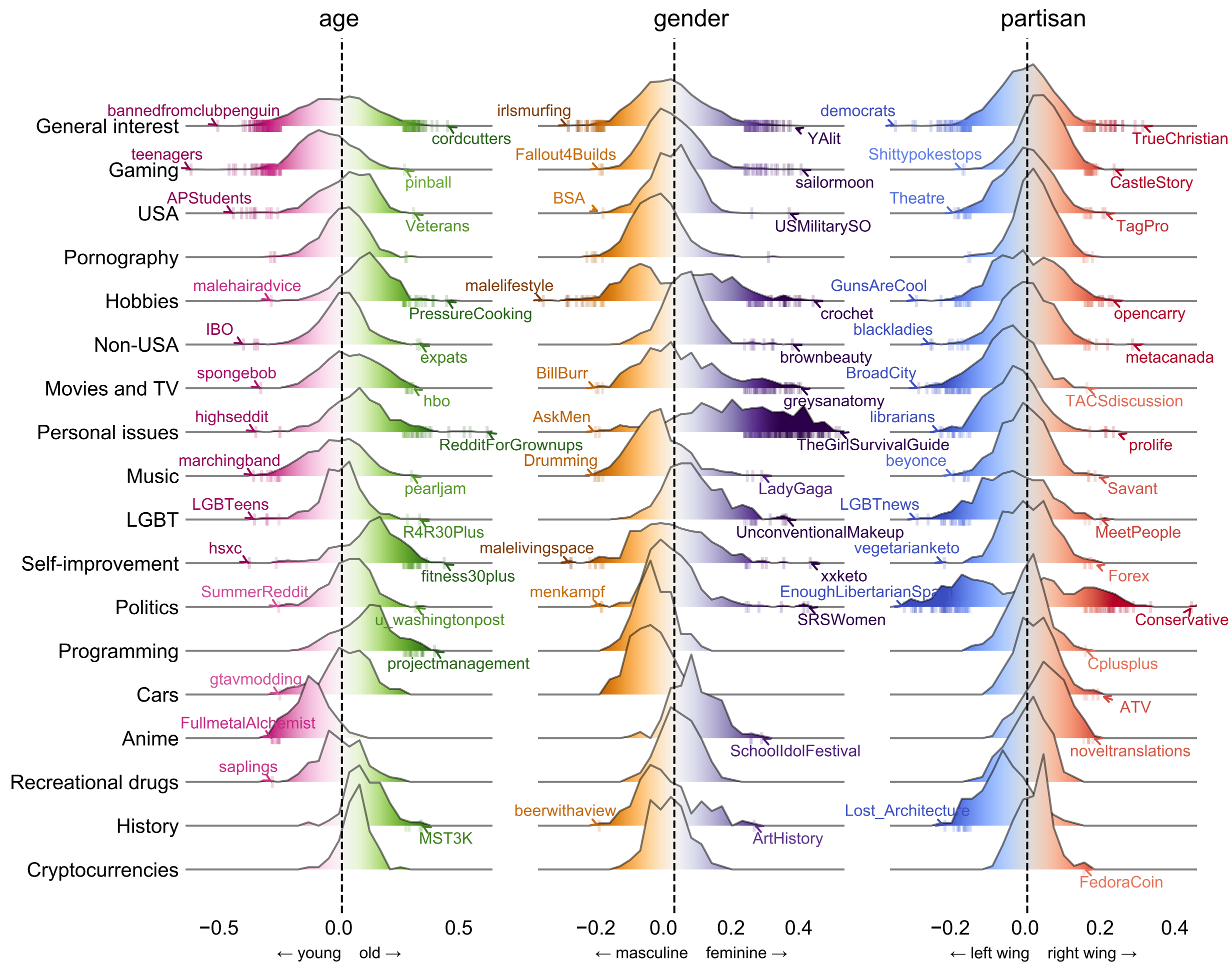

Figure 2: **Inter- and intra-topic diversity.** Distributions of communities along the `age`, `gender`, and `partisan` axes, divided by topical clusters found by hierarchical clustering. Colour corresponds to z-score. The dashed line represents the global mean on an axis. Communities that lie more than two standard deviations from the global axis mean are indicated with vertical bars, and names of the leftmost and rightmost outliers are annotated. Descriptions of all referenced subreddits are given in Appendix I.

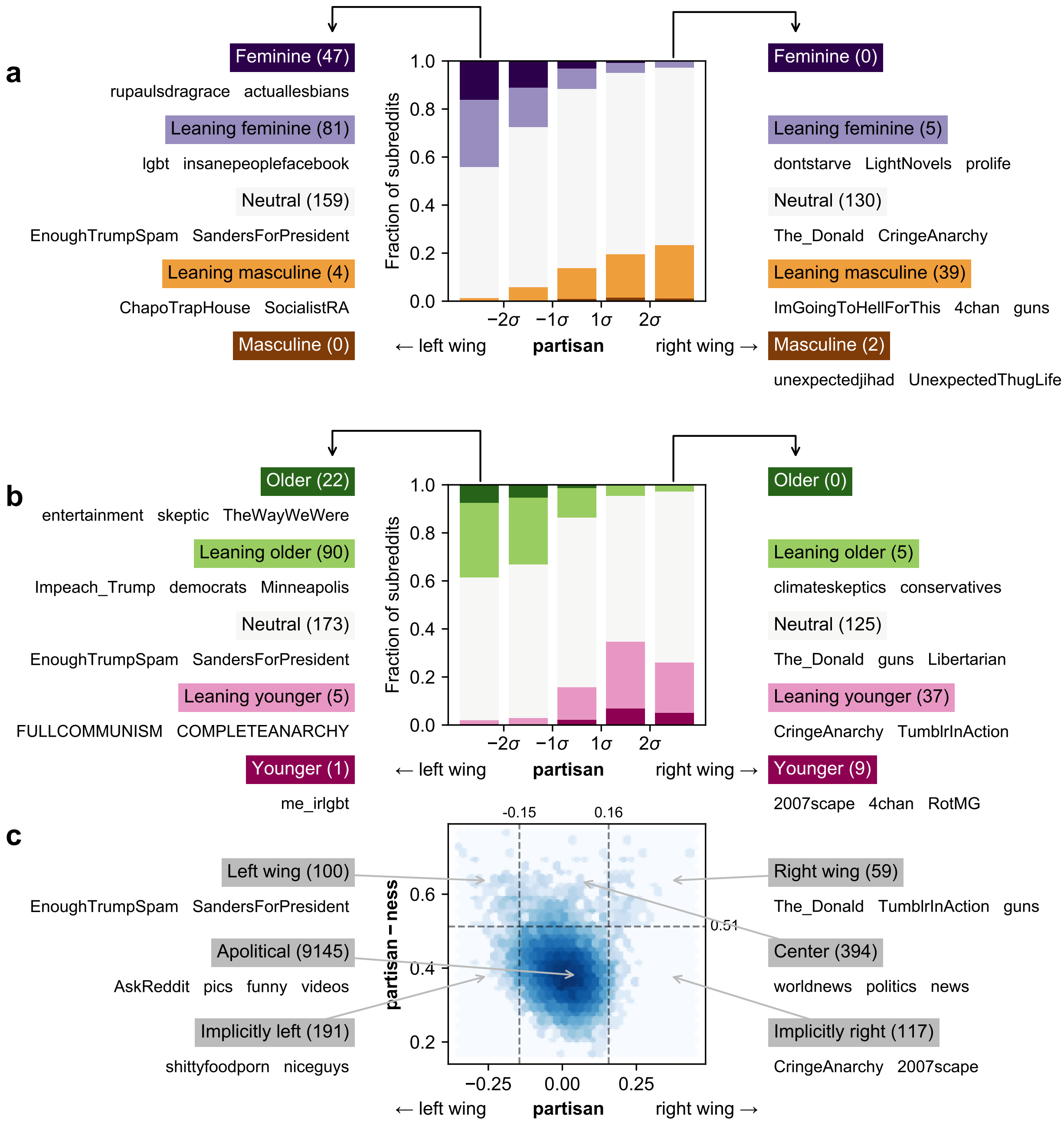

Figure 3: **Online cultural relationships.** The relationships between the `partisan` dimension and (a) `gender`, (b) `age`, (c) `partisan-ness`. Every bar represents a bin of communities with `partisan` scores a given number of standard deviations from the mean, and the distribution illustrates the 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.

feminine leaning. At the community level, the political poles on Reddit are almost completely segregated by gender. Despite this clear relationship, Reddit's cultural range is highlighted by a number of large, active communities that go against the dominant trend. For example, ChapoTrapHouse, a community associated with the 'dirtbag left', is highly left-wing and leans masculine, and prolife, an anti-abortion-rights community, is highly right-wing and leans feminine.

We also find a relationship between `partisan` and `age`; communities that skew older also skew left-wing, while communities that skew younger also skew right-wing ($r = -0.37$). The `partisan` extremes are segregated not only by gender but also by age: among left-wing communities, 38.5% are older but only 2.1% are younger; among right-wing communities, 26.1% are younger but only 2.9% are older. However, the direction of this relationship is the opposite of what is traditionally found in offline contexts—in the 2016 U.S. presidential election, the 18–29 age group voted for Clinton by a share of 58 to 28, while the 65+ age group voted for Trump 53 to 44—but is consistent with prior observations of the relative youth of the alt-right movement [18]. Despite this strong relationship, there are countervailing communities, e.g. me_irlgbt, a left-wing meme community that skews younger, and climateskeptics, a right-wing climate change denial community that skews older. We repeat this analysis on dimensions generated with slightly different seeds to verify robustness of our method and find similar results (see Appendix D for full details.)

Measuring political polarization on Reddit. In our central analysis, we apply our methodology to understand whether and how political activity on Reddit became more polarized over time. To do so, we track the distribution of political activity, as measured by the `partisan` axis, from 2012 to 2018 (Figure 4). We find that while Reddit has supported a wide range of political activity throughout its history, the platform became significantly more polarized around the 2016 U.S. presidential election (Figure 4a). The mean absolute value `partisan` z-score (i.e. absolute number of std. dev. from the mean) of political comments was consistently within a narrow band between 1.09 and 1.27 from 2012 until the end of 2015; it then rose sharply during 2016 and peaked at 1.85 in November 2016 (Figure A9). 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 (Figure A10). The platform maintained an elevated level of polarization (greater than 1.45) until the end of the data time window (end of 2018).

To investigate whether Reddit's platform-level changes in polarization are driven by changes within individuals or changes in the overall population, we decompose total activity by users' political affiliations 12 month prior (Figure 4b). We find that the intense increase in polarization in 2016 was largely driven by a drastic increase in extreme activity among new and newly political users. These users' political activity had an average score of 1.78 in 2016, 0.61 standard deviations more polarized than average activity in 2015, whereas existing users' political activity had an average score of 1.50 in 2016, 0.32 standard deviations more polarized than average activity in 2015. Despite the fact that new users only account for 38.2% of political activity in 2016, they accounted for 51.1% of the activity in far-left and far-right communities, and 53.8% of the overall platform-level increase in polarization is attributable to them (i.e. the overall

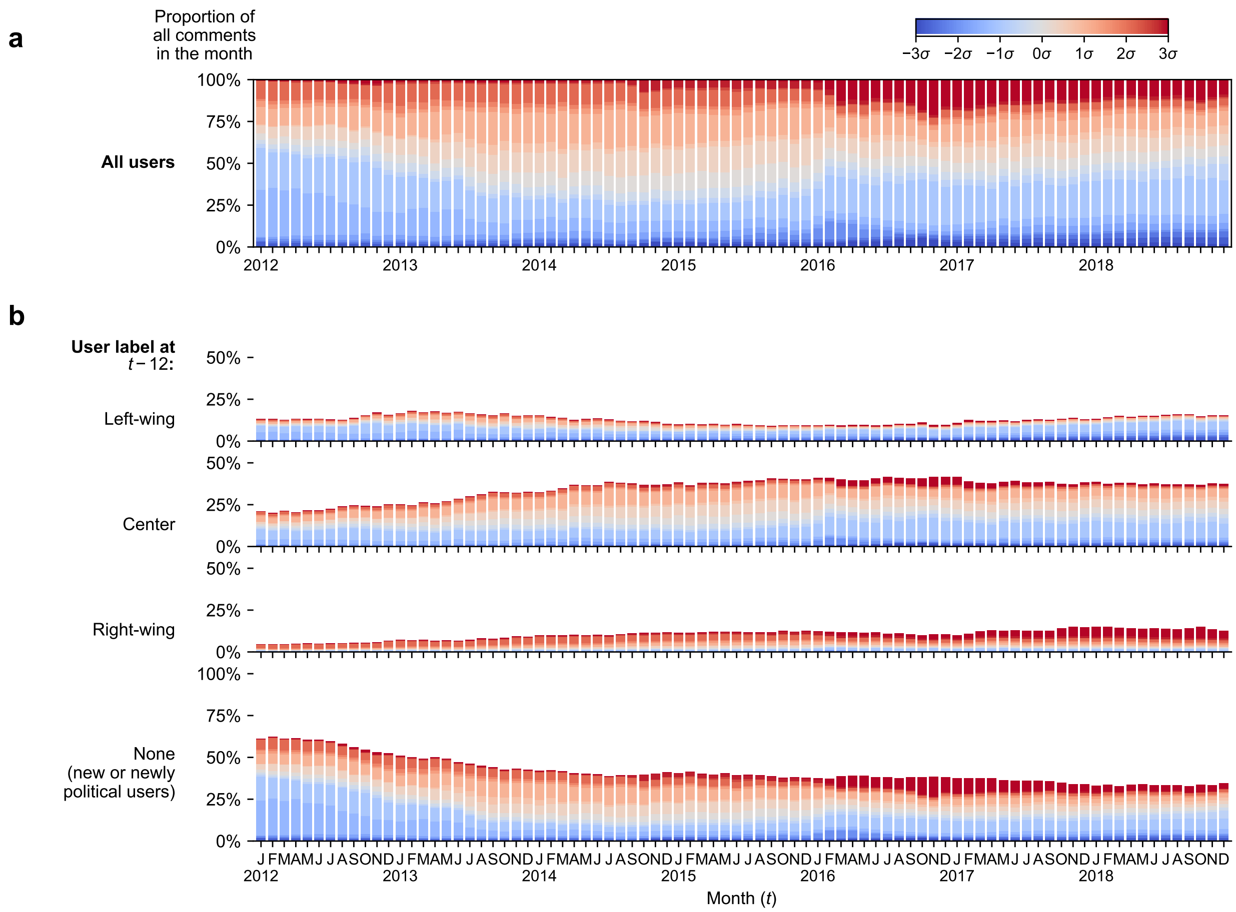

Figure 4: **Political polarization on Reddit.** The distribution of political activity on Reddit over time by `partisan` score. Each bar represents one month of comment activity in political communities on Reddit, and is coloured according to the distribution of partisan scores of comments posted during the month (the partisan score of a comment is simply the partisan score of the community in which it was posted.) Subplot (a) includes all activity, while (b) decomposes 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).

increase in polarization would have been 53.8% less if new user activity in 2016 was only as polarized as activity was in 2015; see Figure A11).

While new and newly political users account for the majority of platform-level polarization, changes in the activity of existing users are also significant. We find that activity of previously right-wing users became significantly more extreme than that of previously left-wing users in 2016. In January 2015, activity in far-right communities accounted for 2.1% of all activity by previously right-wing users; this proportion rose rapidly in 2016 and peaked at 43.5% in October 2018 (Figure A10). In contrast, activity in far-left communities accounted for 5.2% of activity by previously left-wing users in January 2015 and peaked at 10.8% in October 2016. Throughout the rest of Reddit's history, we observe that political activity is remarkably consistent. The distribution of left-wing users' activity remains mostly left-wing 12 months later, and the distribution of right-wing users' activity remains mostly right-wing 12 months later.

To further investigate the extent of within-user polarization, we directly measure it by computing the correlation of users' average `partisan` scores over time. We find `partisan` scores to be highly stable. The within-user correlation between `partisan` scores 12 months apart consistently remained above $r = 0.64$ since 2013, and exceeded $r = 0.83$ during the entire final year of our observation period (Figure 5a). We also find further evidence of the major polarization event in 2016; political activity on Reddit divides cleanly into two epochs on either side of 2016. User `partisan` scores are highly correlated when comparing pre-2016 time periods or post-2016 time periods, but are less correlated when comparing a pre-2016 time with a post-2016 time. After the burst of increased polarization in 2016, within-user `partisan` score correlations stabilized at a higher level of consistency than at any prior time. We measure the prevalence of significant individual-level polarization by computing the fraction of users whose activity moved by at least one standard deviation towards the partisan poles (i.e. their absolute average `partisan` z-score increased by at least 1). This fraction is consistently low; considering time periods 12 months apart, it was 3.0–5.5% prior to 2016, and peaked at 11.3% in November 2016 (Figure 5b). Choosing the two months where the fraction is highest over the entire observation period (April 2013 to September 2018), only 14.5% of individuals became significantly more polarized.

Although instances of individual users becoming more polarized in their `partisan` score over time are rare, it is still possible that newly political users moved from implicitly to explicitly partisan communities. For example, some communities, such as watchpeopledie, have a highly partisan user base but are not themselves explicitly political, and thus have extreme `partisan` scores but low `politicalness` scores (Figure 3c). If engagement with implicitly left- or right-wing communities is related to an increased propensity to subsequently engage with explicitly left- or right-wing 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% to an implicitly right-wing community, restricting the population for whom such an effect could apply (Figure A15, line plots). Users tend to become active

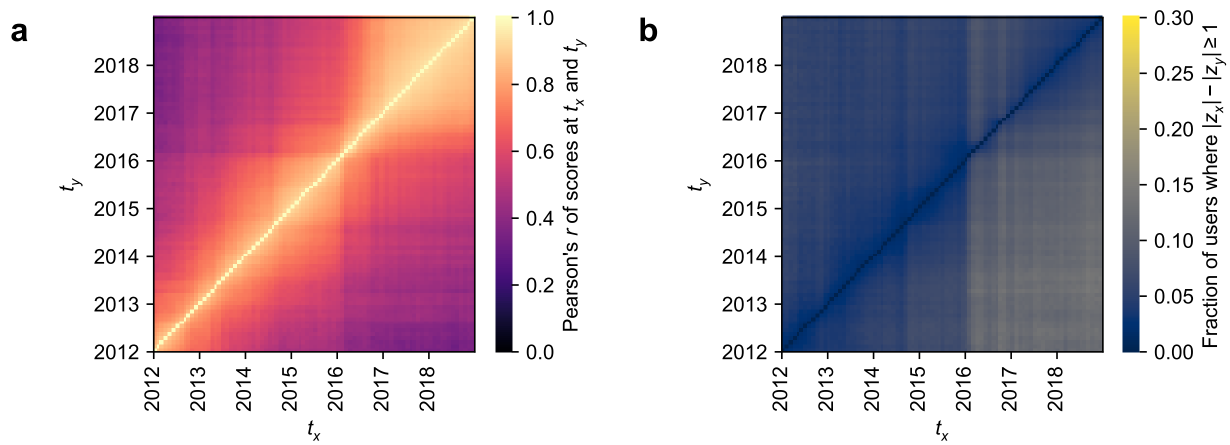

Figure 5: **User-level partisan affiliation. a.** Correlation of users' average `partisan` scores over time. Each (x, y) cell represents the correlation between scores of a user in month t_x and that same user in month t_y, for all users active in both time periods. A user is only considered active if they make at least 10 comments in a month. **b.** The prevalence of significant user-level polarization over time. The colour of the (x, y) cell corresponds to the observed fraction of users whose activity in months t_x is at least one standard deviation more polarized than their activity in month t_y.

in both implicitly and explicitly partisan communities in the same month, further indicating that such a polarization effect is limited in its possible impact (Figure A15, heatmaps).

4 Discussion

There are limitations in our methodological 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 cultural dimensions, do not change. Although it is plausible that some communities change significantly in the cultural makeup of their membership, we expect these to be exceptional cases as most communities are organized around a fixed topic or shared interest. Our method relies on co-membership data—examples of the same user being a member of several communities. If large numbers of people use "throwaway" user accounts for certain communities (e.g. those dedicated to fringe or controversial content), thereby splitting their activity over several accounts, the relationships between these communities and the rest of the platform could be distorted.

Group membership is fundamental to social identity. Sociologists dating back to Simmel, who pioneered the notion of "the web of group affiliations", have employed complex, high-dimensional characterizations to understand identity [27, 14, 10]. We have shown that by harnessing mass co-membership data, we can construct high-dimensional representations of online communities that reflect similarities and differences in their memberships along key cultural dimensions. The large-scale social structure of Reddit, despite its pseudonymity, is organized along cultural lines, and online cultural relationships are related to but meaningfully distinct from their offline analogues. Applying our methodology to political activity, we show that Reddit underwent a significant polarization event around the 2016 U.S. presidential election. However, individual-level polarization is rare. Individual users do not tend to become more extreme in their partisan affiliation, nor do they tend to move from implicitly to explicitly partisan communities. These findings illuminate the cultural relationships between communities that emerge from aggregate user choices, and the extent and nature of political polarization on Reddit.

This study introduces a new paradigm for the analysis of cultural structure in online platforms. Embedding communities into a high-dimensional space and projecting them onto dimensions that correspond to cultural concepts distills vast amounts of behavioural metadata into semantically meaningful, fine-grained measurements of cultural alignment. Our methodology can be generally applied to quantify the cultural organization of online discussion, to situate important content and behaviours, such as misinformation and toxic language, in the cultural context of the platform, and to measure the presence of online cultural shifts and the mechanisms that drive them.

References

[1] Hugh J Arnold and Daniel C Feldman. "Social desirability response bias in self-report choice situations". In: *Academy of Management Journal* 24.2 (1981), pp. 377–385.

[2] Christopher A Bail et al. "Exposure to opposing views on social media can increase political polarization". In: *Proceedings of the National Academy of Sciences* 115.37 (2018), pp. 9216–9221.

[3] Eytan Bakshy, Solomon Messing, and Lada A Adamic. "Exposure to ideologically diverse news and opinion on Facebook". In: *Science* 348.6239 (2015), pp. 1130–1132.

[4] Jason Baumgartner et al. "The pushshift reddit dataset". In: *Proceedings of the International AAAI Conference on Web and Social Media*. Vol. 14. 2020, pp. 830–839.

[5] Tolga Bolukbasi et al. "Man is to computer programmer as woman is to homemaker? debiasing word embeddings". In: *Advances in neural information processing systems*. 2016, pp. 4349–4357.

[6] Aylin Caliskan, Joanna J Bryson, and Arvind Narayanan. "Semantics derived automatically from language corpora contain human-like biases". In: *Science* 356.6334 (2017), pp. 183–186.

[7] Pew Research Center. *An examination of the 2016 electorate, based on validated voters*. 2018. URL: `https://www.pewresearch.org/politics/2018/08/09/an-examination-of-the-2016-electorate-based-on-validated-voters/`.

[8] Ronald J Chenail. "Interviewing the investigator: Strategies for addressing instrumentation and researcher bias concerns in qualitative research." In: *Qualitative Report* 16.1 (2011), pp. 255–262.

[9] Michael D Conover et al. “Political polarization on twitter.” In: *Icwsm* 133.2011 (2011), pp. 89–96.

[10] Kimberlé W Crenshaw. *On intersectionality: Essential writings*. The New Press, 2017.

[11] Michela Del Vicario et al. “Echo chambers: Emotional contagion and group polarization on facebook”. In: *Scientific reports* 6 (2016), p. 37825.

[12] Nicole B Ellison, Charles Steinfield, and Cliff Lampe. “The benefits of Facebook “friends:” Social capital and college students’ use of online social network sites”. In: *Journal of computer-mediated communication* 12.4 (2007), pp. 1143–1168.

[13] Henry Farrell. “The consequences of the internet for politics”. In: *Annual review of political science* 15 (2012).

[14] Scott L Feld. “The focused organization of social ties”. In: *American journal of sociology* 86.5 (1981), pp. 1015–1035.

[15] Seth Flaxman, Sharad Goel, and Justin M Rao. “Filter bubbles, echo chambers, and online news consumption”. In: *Public opinion quarterly* 80.S1 (2016), pp. 298–320.

[16] Nikhil Garg et al. “Word embeddings quantify 100 years of gender and ethnic stereotypes”. In: *Proceedings of the National Academy of Sciences* 115.16 (2018), E3635–E3644.

[17] Sharad Goel et al. “The structural virality of online diffusion”. In: *Management Science* 62.1 (2016), pp. 180–196.

[18] George Hawley. *Making sense of the alt-right*. Columbia University Press, 2017.

[19] Austin C Kozlowski, Matt Taddy, and James A Evans. “The geometry of culture: Analyzing the meanings of class through word embeddings”. In: *American Sociological Review* 84.5 (2019), pp. 905–949.

[20] Ivar Krumpal. “Determinants of social desirability bias in sensitive surveys: a literature review”. In: *Quality & Quantity* 47.4 (2013), pp. 2025–2047.

[21] Omer Levy and Yoav Goldberg. “Dependency-based word embeddings”. In: *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*. 2014, pp. 302–308.

[22] Omer Levy and Yoav Goldberg. “Neural word embedding as implicit matrix factorization”. In: *Advances in neural information processing systems*. 2014, pp. 2177–2185.

[23] Eli Pariser. *The filter bubble: What the Internet is hiding from you*. Penguin UK, 2011.

[24] *Reddit Historical Code Archive*. `https://github.com/reddit-archive/reddit`. Accessed: 2020-09-04.

[25] Dominik Schlechtweg, Cennet Oguz, and Sabine Schulte im Walde. “Second-order Co-occurrence Sensitivity of Skip-Gram with Negative Sampling”. In: *arXiv preprint arXiv:1906.02479* (2019).

[26] Feng Shi et al. “Millions of online book co-purchases reveal partisan differences in the consumption of science”. In: *Nature Human Behaviour* 1.4 (2017), pp. 1–9.

[27] George Simmel. *Conflict and the web of group affiliations*. 1955.

[28] Marshall Van Alstyne and Erik Brynjolfsson. “Electronic Communities: Global Villages or Cyberbalkanization?” In: *ICIS 1996 Proceedings* (1996), p. 5.

[29] Thea F Van de Mortel et al. “Faking it: social desirability response bias in self-report research”. In: *Australian Journal of Advanced Nursing, The* 25.4 (2008), p. 40.

[30] José Van Dijck. *The culture of connectivity: A critical history of social media*. Oxford University Press, 2013.

[31] Soroush Vosoughi, Deb Roy, and Sinan Aral. “The spread of true and false news online”. In: *Science* 359.6380 (2018), pp. 1146–1151.

Appendix

A Data

We use the complete set of 5.1 billion comments made on Reddit posts since comments were introduced in 2005, up to and including 2018. Data is publicly available and was downloaded from the `pushshift.io` Reddit archive [4] at `http://files.pushshift.io/reddit/`. Data for each comment includes the body, the time it was posted, the subreddit (community) in which it was posted, and the username of the commenter. There are 34.7 million distinct commenters in our dataset. Figure A1 shows the number of comments posted each month.

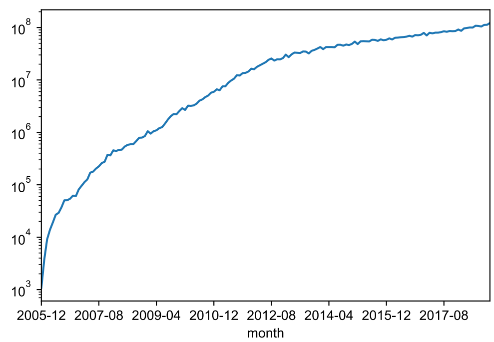

Figure A1: Total number of comments posted per month to Reddit.

Over our entire study period, 52.9% of users commented in more than one subreddit, indicating that many users engage in the multi-community aspect of the platform. Indeed, the mean number of subreddits commented in by a user is 9.6. This provides crucial information about the semantic similarity of subreddits, which we harness to create community embeddings.

B Creating the community embedding

We create a community embedding from this 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 [21]. To create a community embedding, we treat communities as "words" and users as "contexts". As input, we provide all pairs of users and communities present in the comment data set. For example, if user u_i commented in community c_j 10 times, the pair (u_i, c_j) would appear 10 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 is insufficient data to generate a meaningful vector representation, we restrict to the top 10,006 subreddits by number of comments, which accounts for 95.4% of all comments and 93.2% of all users.

The `word2vecf` model has numerous hyperparameters that affect the training process and resulting embedding. To tune the model for the community embedding use case, we perform a grid search of the hyperparameter space, optimizing for performance on a set of community analogies. The hyperparameters we vary 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 add an additional parameter `shuffled`, a boolean parameter which indicates whether the training data should be randomly shuffled prior to training. We assess 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 are able to find an embedding that solves 72% of 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 use 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 random 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 [22] show that the SGNS objective is optimized by a factorization of the word-context pointwise mutual information (PMI) matrix. 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 $\#$ is the count of all matching comments:

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

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

$$\vec{c_1} \cdot \vec{c_2} = \sum_u PMI(c_1,u) \cdot PMI(c_2,u)$$

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, 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 [22]. 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 trucks and golf had no users in common, but both had a high overlap with the AskMen community, their vectors might end up somewhat close to each other despite no users being members of both communities. As factorization of the PMI matrix is unable to capture any higher-order effects, it has been theorized that it is this ability which makes SGNS more capable of capturing semantic relationships in word embeddings. 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. [25], 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. 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 userbases of the two communities.

We perform a clustering of the community embedding to provide context to our visualizations and to get a basic overview of Reddit's community structure. We use agglomerative clustering based on Euclidean distance to cluster the embedding into 25 clusters. These clusters are then manually labeled based on dominant topic, and clusters with the same label are merged. We end up with 18 topical clusters representing major categories of community; for example, "Movies and TV" (n = 478), "Gaming" (n = 1761), and "Hobbies" (n = 547). The largest cluster consists of communities with no clear theme; we call this cluster "General interest" (n = 2178).

C Finding cultural axes

To find cultural axes in our community embedding, we first identify a seed pair of communities that differ primarily in the target cultural concept. An ideal choice of seed is a pair of communities that are extremely similar except for a difference in the target cultural concept, as that will prevent the resulting dimension from being confused with any other differences between the two communities. A common confound is a difference in the time when two communities were popular. If two communities which were popular at very different times are selected, there will be little user overlap between them (given Reddit's exponential growth), and the dimension will primarily represent a difference in real-world time. A good choice of seed is two similar subreddits with a single cultural difference and very similar age.

For our `gender` axis, we choose AskMen and AskWomen, personal discussion forums for men and women; for our `age` axis, we choose teenagers and RedditForGrownups, personal discussion forums for teenagers and adults; for our `partisan` axis, we choose democrats and Conservative, two partisan American political communities; and for our `affluence` axis, we choose vagabond, a forum for homeless travellers, and backpacking, a more general interest travel community. While we focus here on traditional gender identity, the method is not inherently constrained to

			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
			edginess			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

Table A1: Community pairs used to calculate axes. The blue highlighted pair is the initial seed provided to the algorithm. The rest of the pairs are algorithmically found as described in Appendix C.

one-dimensional conceptions of gender. Multiple gender axes could be generated to build a more complete analysis of gender.

While the choice of seed is important, our axis generation method is robust as similar seed choices generate similar dimensions. To demonstrate this, we also generate a `gender B` axis with Daddit and Mommit, parenting discussion forums for men and women; an `age B` axis with AskMen and AskMenOver30, Q&A communities for men of all ages and men over 30; and a `partisan B` axis with hillaryclinton and The_Donald, two partisan American political communities. We also generate three axes for concepts not necessarily related to offline culture 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 `edginess`, representing aggression and coarseness (seeds can be found in Table A1).

To make sure a dimension is not overfit to the two communities in the seed pair, we automatically augment the seed pair with 9 additional pairs of communities. We generate the set of all 100,060 non-trivial pairs of communities with their 10 nearest neighbours. 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 cultural concept. All 100,060 pairs are ranked based on the cosine similarity of their vector difference with the vector difference of the seed pair. The 9 most similar pairs are selected to end up with 10 pairs to create the axis. Table A1 contains all the similar pairs automatically found for all the axes. We tested with more and less 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.

The vector differences of all 10 pairs are averaged together to obtain a single vector for the axis that robustly represents the desired cultural concept. We also compute an alternate `-ness` version of each axis obtained by simply averaging the vector *sums* of all 10 pairs. This axis respresents similarity to the communities on both sides of the pairs. This can be used to measure association with an axis in general; for example, the `partisan-ness` dimension represents how explicitly political a community is.

D Computing community scores

Once a vector for an axis has been obtained, all communities can be assigned a score on that axis by simply projecting the normalized community vector $\vec{c}$ onto that vector: $\vec{c} \cdot \vec{d}$. The score of a community on an axis 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 $\vec{c}$ with a cultural dimension with n normalized seed pairs $(A_1, B_1) \ldots (A_n, B_n)$ defined as $\vec{d} = \frac{1}{n} \sum (B_i - A_i)$ is the following:

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

As previously explained, the dot product of two community vectors is a measure of the similarity of their members. Thus, 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 calculate the scores for all 10,006 communities on all axes. The distributions of community scores for `age`, `gender`, `partisan`, and `affluence` can be found in Figure A2 and the distributions for `age-ness`, `gender-ness`, `partisan-ness`, `time`, `sociality`, and `edginess` in Figure A3. Distributions broken down by semantic cluster for `age`, `gender`, and `partisan` can be found in Figure 2 and for `age-ness`, `gender-ness`, `partisan-ness`, `affluence`, `time`, `sociality`, and `edginess` in Figure A4.

To demonstrate the robustness of the axis generation method, we compare each of the `age`, `gender`, `partisan` axes with their `B` version. Scatter plots for each of these pairs can be found in Figure A5. 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 $r = 0.55$. 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 The_Donald as a seed generates a dimension that is more specific to Trump and his online supporters' interests, in contrast with using 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.

E Validating community scores

We validate each of `age`, `gender`, `partisan`, and `partisan-ness` against the external concepts they represent. To validate the `gender` axis, we compare the `gender` scores of occupation communities to the actual gender makeup of those occupations. We use gender makeup data from the 2018 American Community Survey, and manually match occupation descriptions to subreddit names (see Table A2 for a list.) We find there is a $r = 0.89$ correlation between the percentage of women in an occupation and its communities' gender score (scatter plot in Figure A6.) The `gender` axis well represents the proportion of women in an occupation even for occupations at the extremes and in the middle. To validate the `age` axis, we compare 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

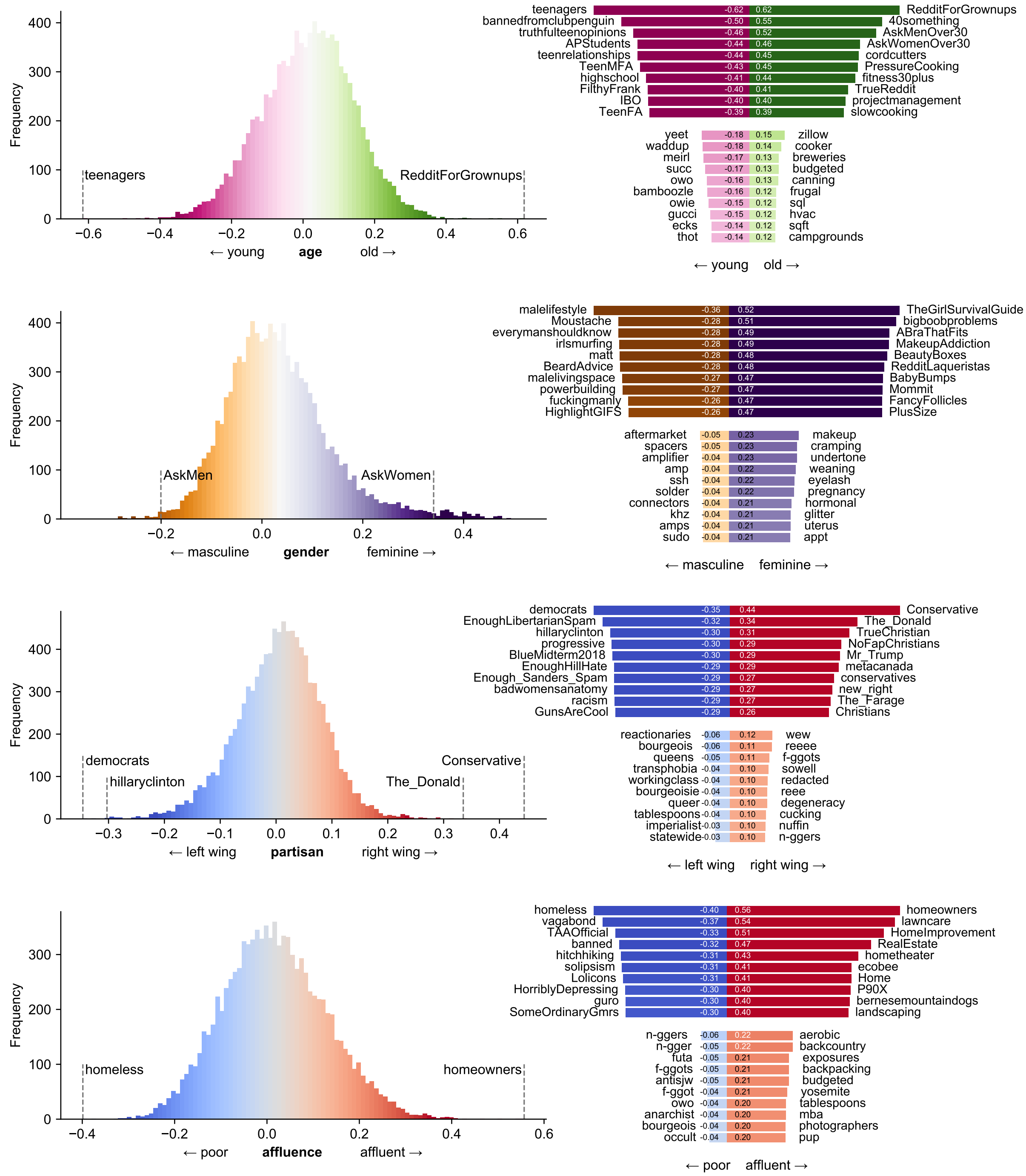

Figure A2: Left: distributions of communities on the age, gender, partisan, and affluence axes. Right: the most extreme communities and words on those axes. Word scores are calculated by averaging community scores weighted by the number of occurrences of the word in the community in 2017. Community descriptions can be found in Appendix I.

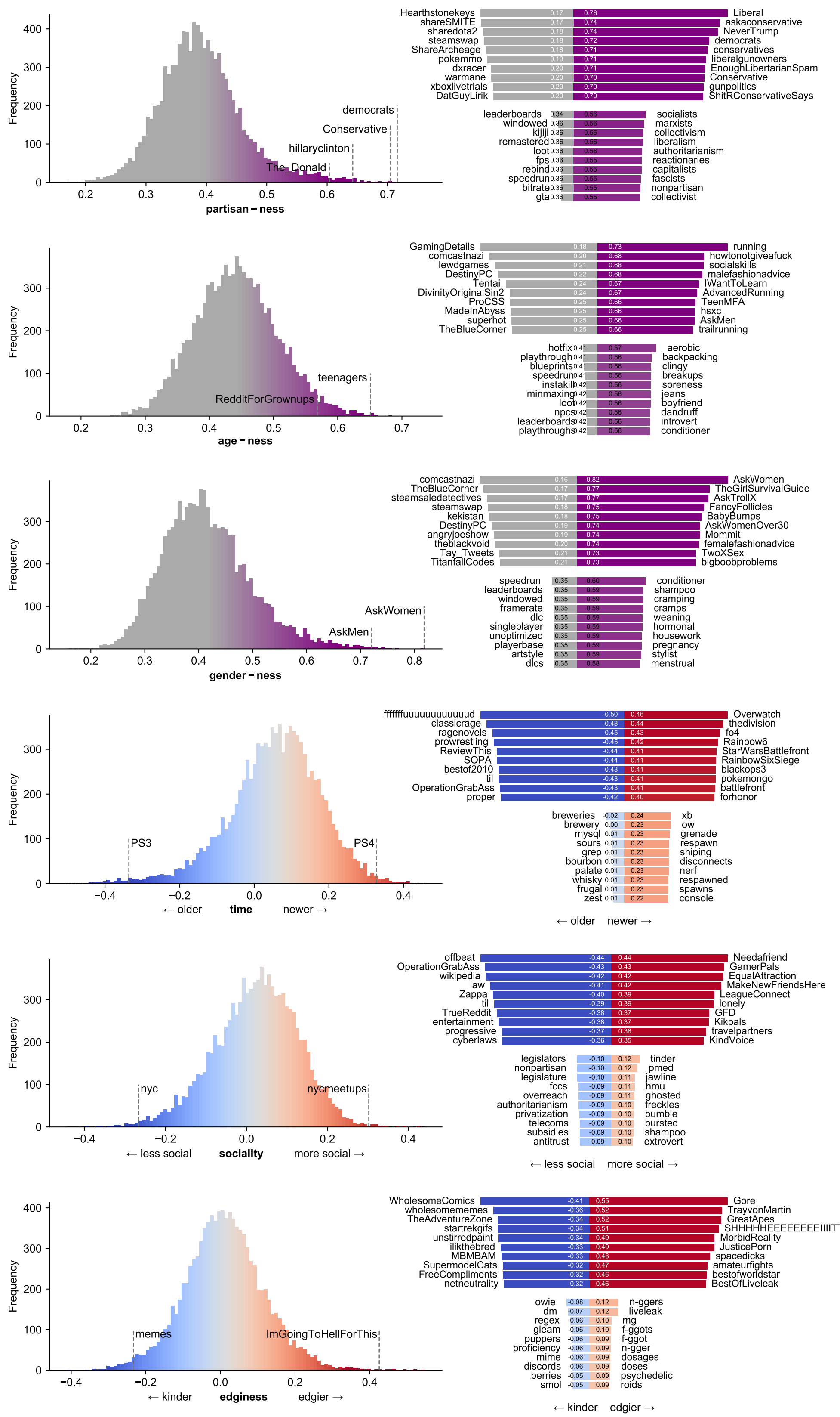

Figure A3: Left: distributions of communities on the partisan-ness, age-ness, gender-ness, time, sociality, and edginess axes. Right: the most extreme communities and words on those axes. Community descriptions can be found in Appendix I.

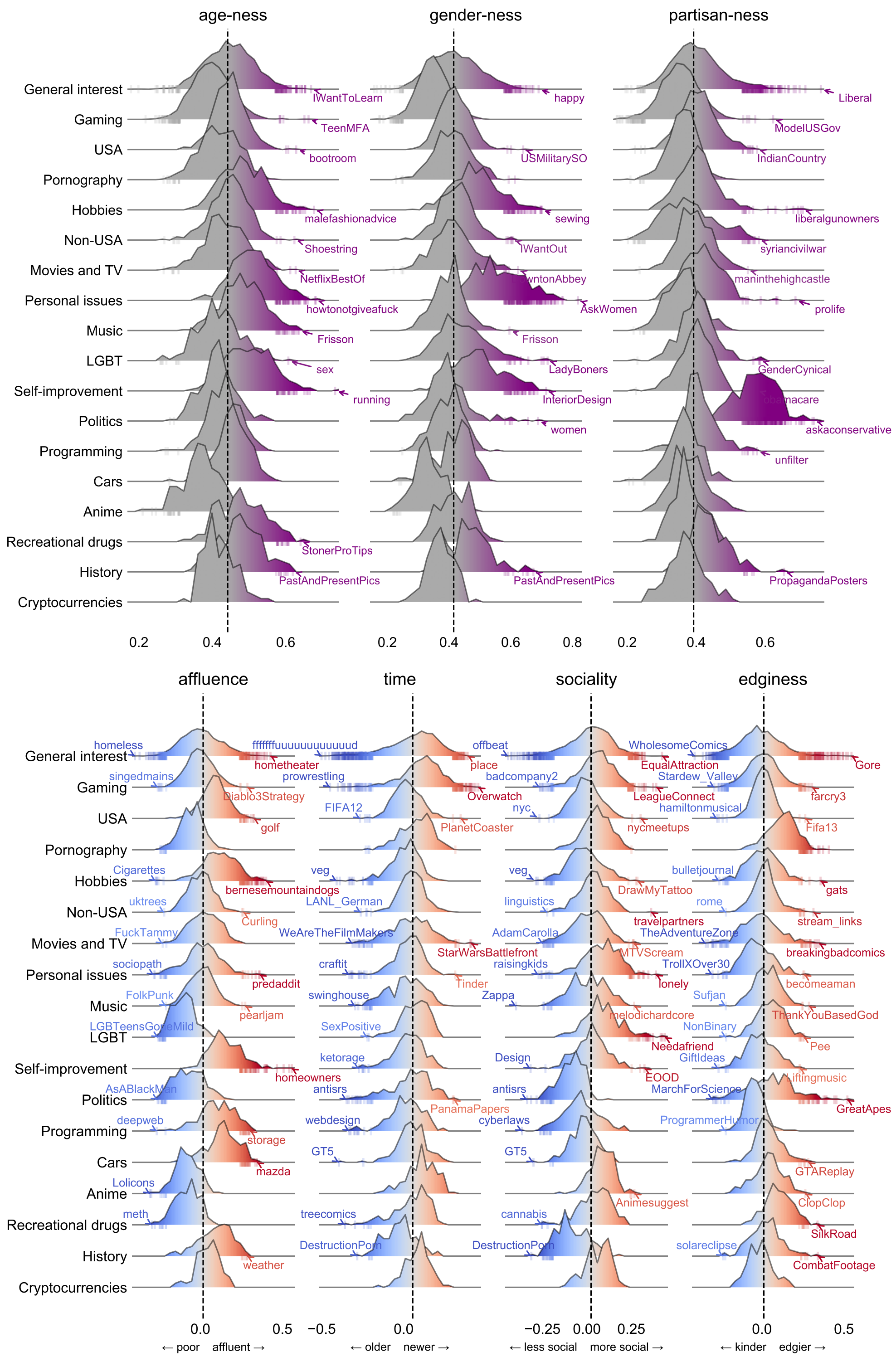

Figure A4: Distributions of communities on the age-ness, gender-ness, partisan-ness, affluence, time, sociality, and edginess axes, broken down by semantic clusters. Outliers which lie more than two standard deviations from the mean are labelled (on 'ness' axes, only those which are greater than the mean.) Community descriptions can be found in Appendix I.

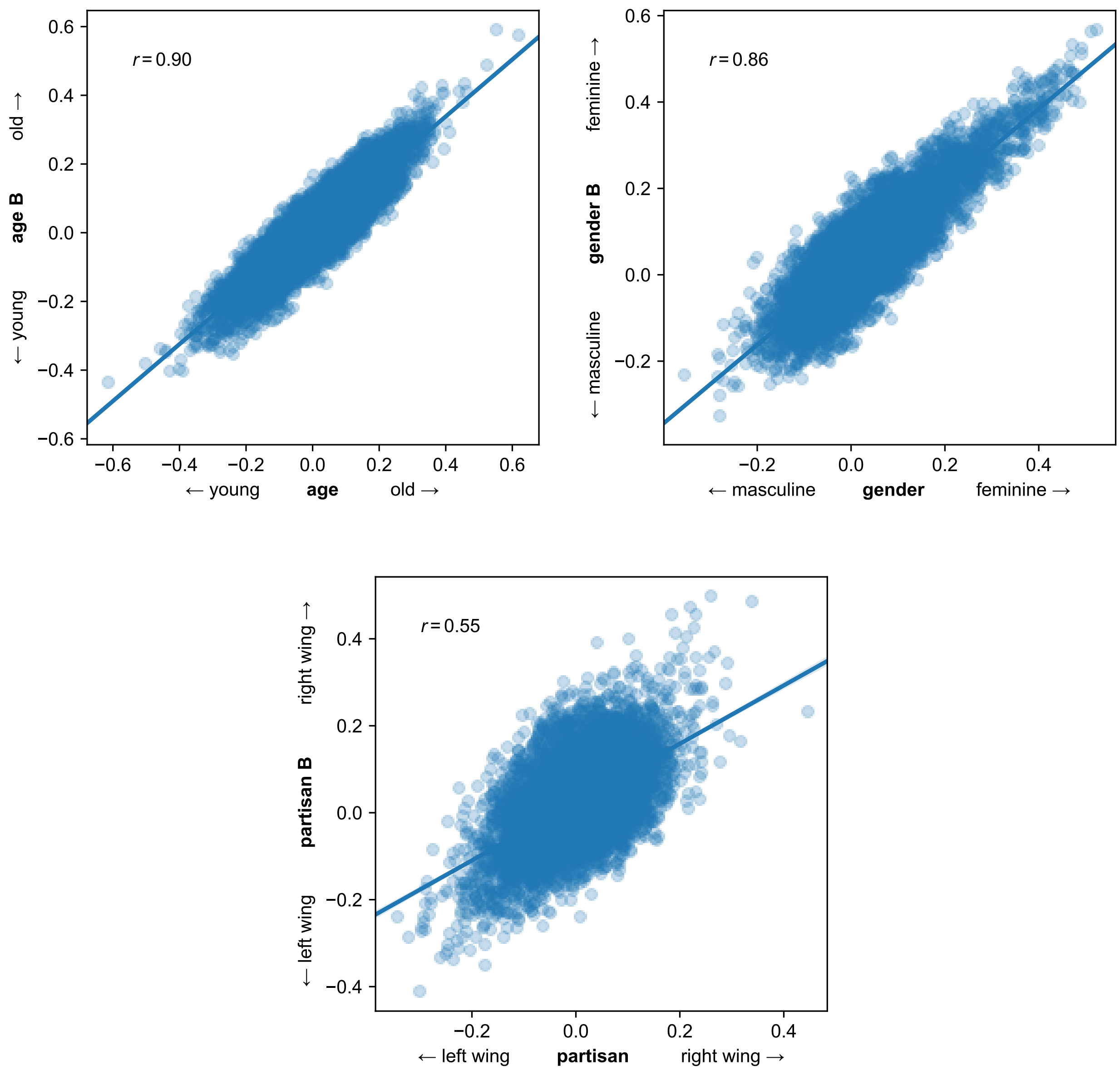

Figure A5: Joint distributions of the age, gender, and partisan axes with their alternate 'B' versions generated with different seeds, illustrating that axis generation is robust to seed choice.

Community	Occupation
"Firefighting"	"Firefighters"
"civilengineering"	"Civil engineers"
"Construction"	"Construction laborers"
"metalworking"	"Sheet metal workers" "Other metal workers and plastic workers"
"Carpentry"	"Carpenters"
"electricians"	"Electricians"
"Plumbing"	"Plumbers, pipefitters, and steamfitters"
"Truckers"	"Driver/sales workers and truck drivers"
"mechanics"	"Automotive service technicians and mechanics"
"farming"	"Farmers, ranchers, and other agricultural managers"
"humanresources"	"Human resources workers"
"teaching"	"Elementary and middle school teachers" "Secondary school teachers"
"ECEProfessionals"	"Preschool and kindergarten teachers"
"nursing"	"Registered nurses"
"Dentistry"	"Dentists" "Dental hygienists" "Dental assistants"
"psychotherapy"	"Marriage and family therapists" "Therapists, all other"
"specialed"	"Special education teachers"
"socialwork"	"Child, family, and school social workers" "Mental health and substance abuse social workers" "Social workers, all other"
"Nanny"	"Childcare workers"
"optometry"	"Optometrists"
"pharmacy"	"Pharmacists" "Pharmacy aides"
"librarians"	"Librarians and media collections specialists"
"Professors"	"Postsecondary teachers"

Table A2: Reddit community and American Community Survey occupation pairs used for the `gender` validation. Some subreddits are mapped to multiple occupations, in which case the average gender ratio is used.

`age` and whether a community is associated with a university or a city (r = 0.91, Cohen's d = 4.37). As shown in Figure A7, university communities skew far younger and city communities skew far older.

To validate the `partisan` axis, we manually code communities as left or right wing, and verify that the partisan score distinguishes between them. We select communities that contain in their description either one of the left-wing terms “democrat”, “clinton”, “left”, “progressive” or one of the right-wing terms “republican”, “trump”, “right”, “conservative”. We then manually code these communities based on their bio 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 bio is supportive or against them. We code 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 bio. We find that this label is strongly associated with the partisan score (r = 0.92, Cohen's d = 4.89). We also use this labelling to validate the `partisan-ness` axis. We compare 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 perform an additional validation using 2016 US Census data for the `affluence` and `partisan` axes. 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 ($r = 0.39$), 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 ($r = 0.42$). We view the moderate strength of these correlations as not an indicator of the weakness of our method but as a likely consequence of the unrepresentativeness of Reddit's population.

F Measuring relationships between axes

After validating the cultural scores, we measure the relationships between these axes as they exist on Reddit. Figure A8 shows the joint distributions between our three primary axes, `age`, `partisan`, and `gender`. 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` ($r = -0.37$). The correlation between `gender` and `partisan` is similar to the relationship that exists offline. The correlation between `age` and `partisan` is in the opposite direction as the offline relationship. We speculate this could be driven by the many successful right-wing meme subreddits that exist on Reddit, as memes skew younger.

We repeat 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 ($r = -0.33$).

G Computing user and word scores

We additionally compute scores for users and words along all axes to provide context to our primary analyses. User scores are weighted averages of community scores weighted by the number of times the user has commented in a community. Word scores are weighted averages of community scores weighted by the number of times the word was used in a community. To avoid distortion introduced by bots that re-use the same word over and over again in automated postings, we cap the number of usages of a word in a subreddit by one commenter that are counted at 100. User and word scores represent in which types of communities that word or user is likely to be observed. The words with the most extreme scores on the poles of each axis are available in Figures A2 and A3.

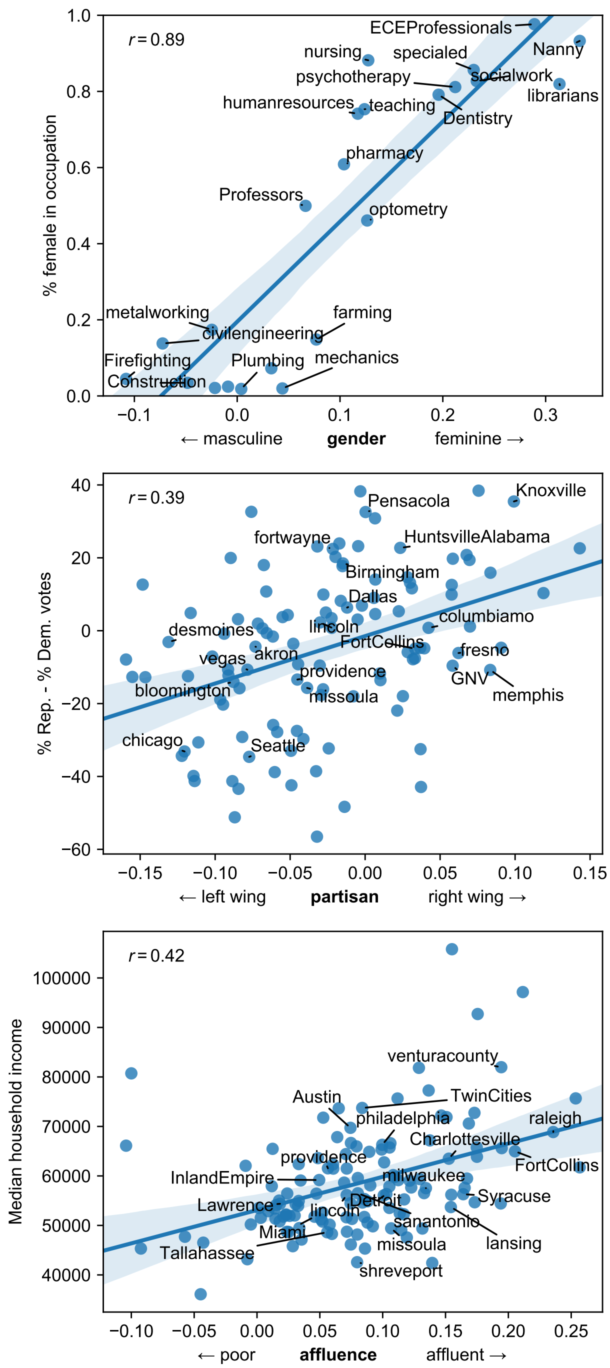

Figure A6: Scatter plots of the external validations of the gender, partisan, and affluence axes. Gender scores for occupational communities are plotted against the percentage of women in that occupation from the 2018 American Community Survey. Partisan scores for city communities are plotted against the Republican vote differential for that metropolitan area in the 2016 presidential election. Affluence scores of city communities are plotted against the median household income for that metropolitan area from the 2016 US Census.

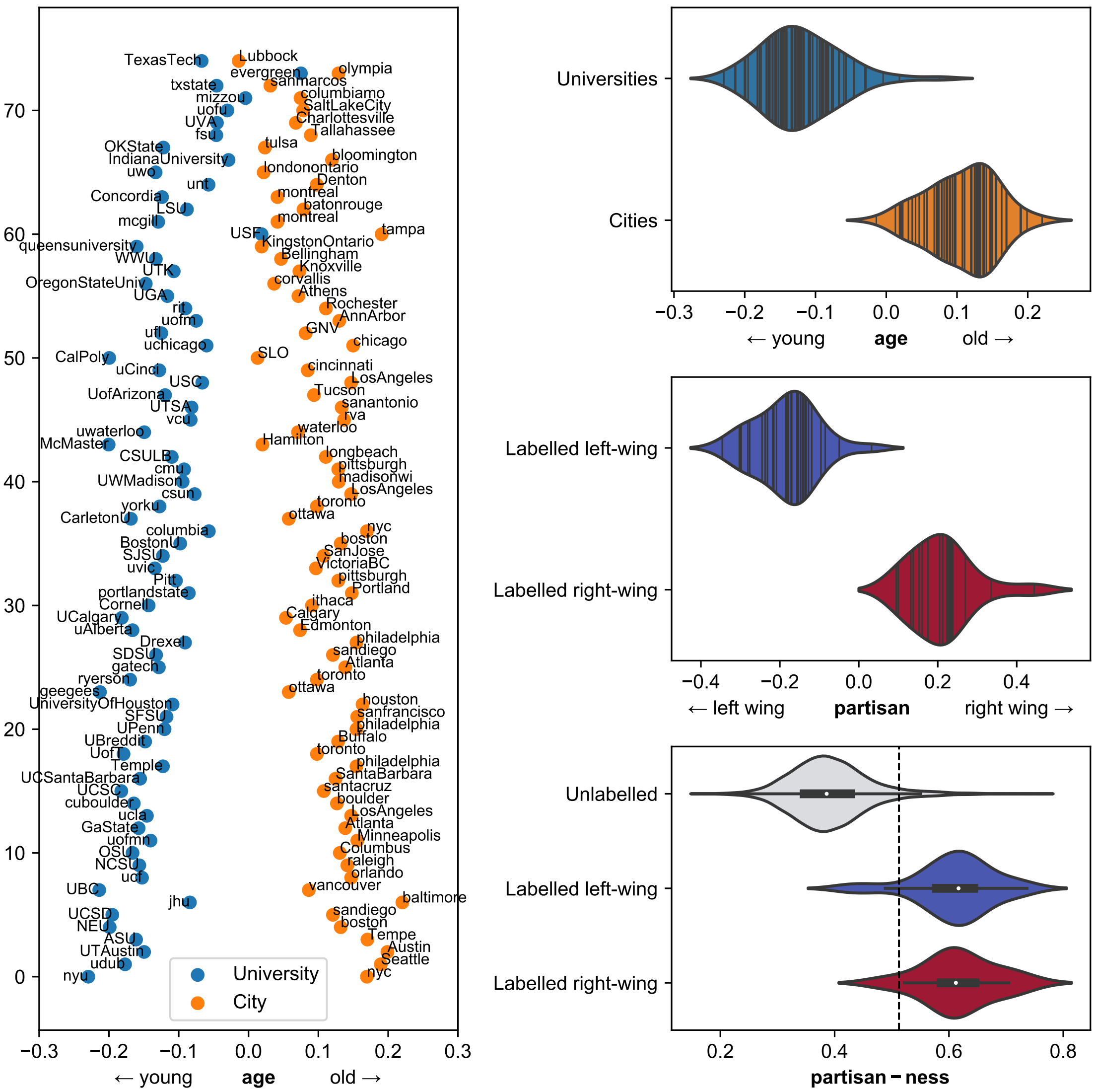

Figure A7: Clockwise from left: The gap between university and city communities on the age axis. The distribution of university and city communities on the age axis; age is strongly related to label ($r = 0.91$, Cohen's $d = 4.37$). The distribution of left and right wing labelled communities on the partisan axis; partisan is strongly related to label ($r = 0.92$, Cohen's $d = 4.89$.) The distribution of left and right wing labelled communities on the partisan-ness axis as compared to the general distribution; there is a large difference in their means (Cohen's $d = 3.27$).

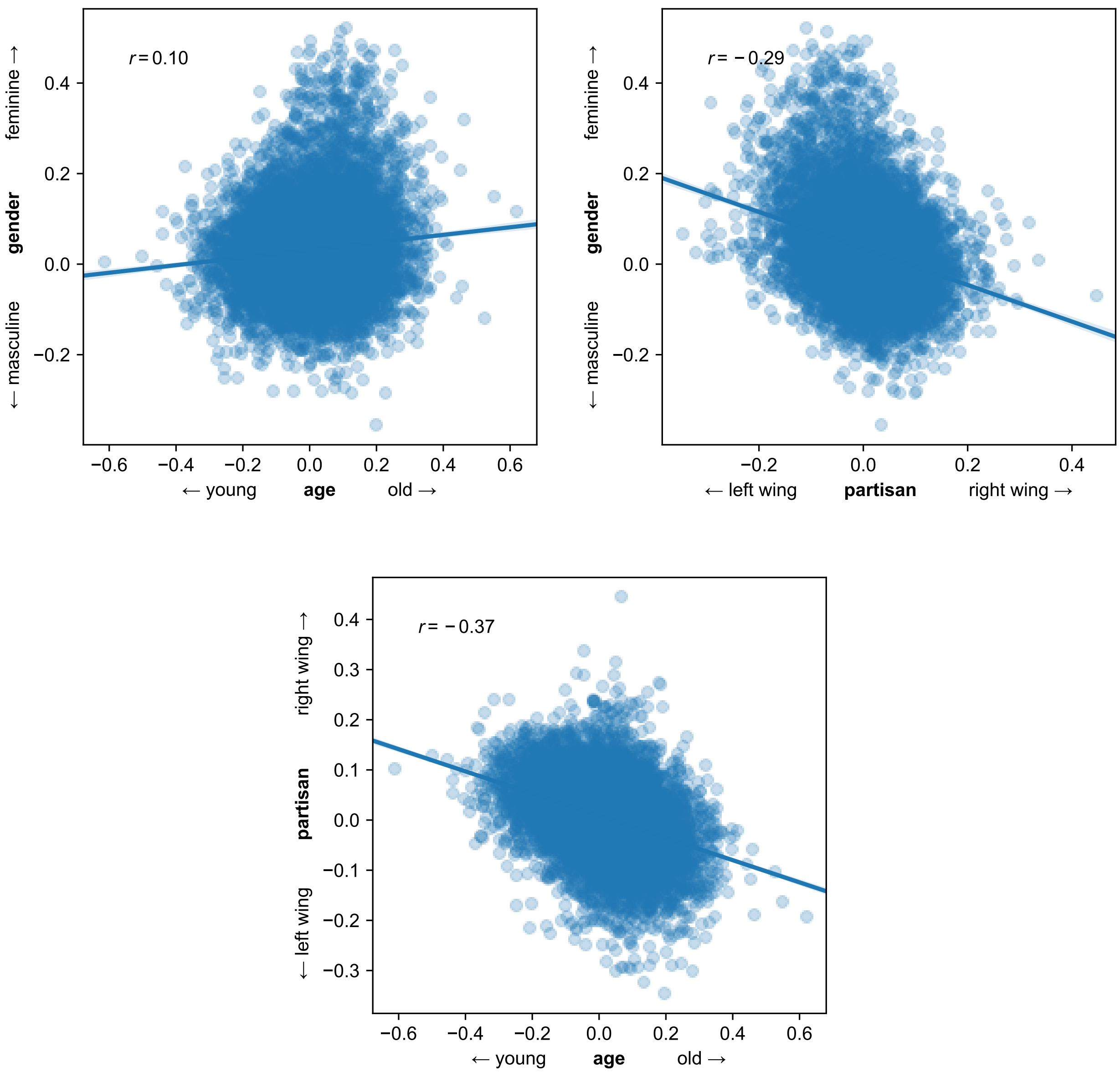

Figure A8: Joint distributions of the age, partisan, and gender dimensions.

H Measuring political activity

To quantify political polarization on the platform, we measure the extent to which political activity has gotten more extreme over time. We first restrict our focus only to "political activity"—comments in political communities as defined by the `partisan-ness` axis. We choose a cutoff on the `partisan-ness` axis such that it is the highest value that includes 80% of the "Politics" cluster. Using this cutoff 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 labelled as explicitly political by us in the previously described validation for `partisan`.

We further restrict 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 are lacking author data for these comments, we are unable to tell whether they were made by a new user or an existing user. Since our aim is to attribute changes in activity to new or existing users, we exclude 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 effect that it affects our main findings. To assess whether such a difference exists, we compare the distribution of deleted comments on the partisan axis Q to the distribution of non-deleted comments P. We find that the distributions are extremely similar (Figure A12); they have a difference of means of only -0.01 and a Kullback–Leibler divergence of $D_{KL}(P \parallel Q) = 0.033$ bits. We conclude that it is reasonable to exclude deleted comments from the following analyses.

Focusing on only this subset of comments in explicitly political communities, we quantify the distribution of `partisan` scores each month. Figure 4 displays the distribution of the `partisan` scores of comments each month and how this varies by a user's previous political activity. Figure A9 displays a single metric—average absolute value of z-score—to simplify this distribution and measure how polarized, on average, activity is in a period. Figure A10 displays the proportion of activity that takes place in very left- and right-wing communities as an alternate metric. Figure A11 shows how the average absolute z-score has changed from year to year and the proportion of the change that is attributable to new user activity or existing user activity.

We repeat the analysis shown in Figure 4 on the `partisan B` axis and find very similar results (see Figure A13.) The average absolute value z-score in January 2015 was 1.13, and it peaked in November 2016 at 1.98. The average absolute value z-score for activity of new and newly political users was 1.13 in January 2015, and peaked at 2.19 in November 2016.

We also repeat the implicit polarization analysis shown in Figure 5 on the `partisan B` axis and find similar results (Figure A14).

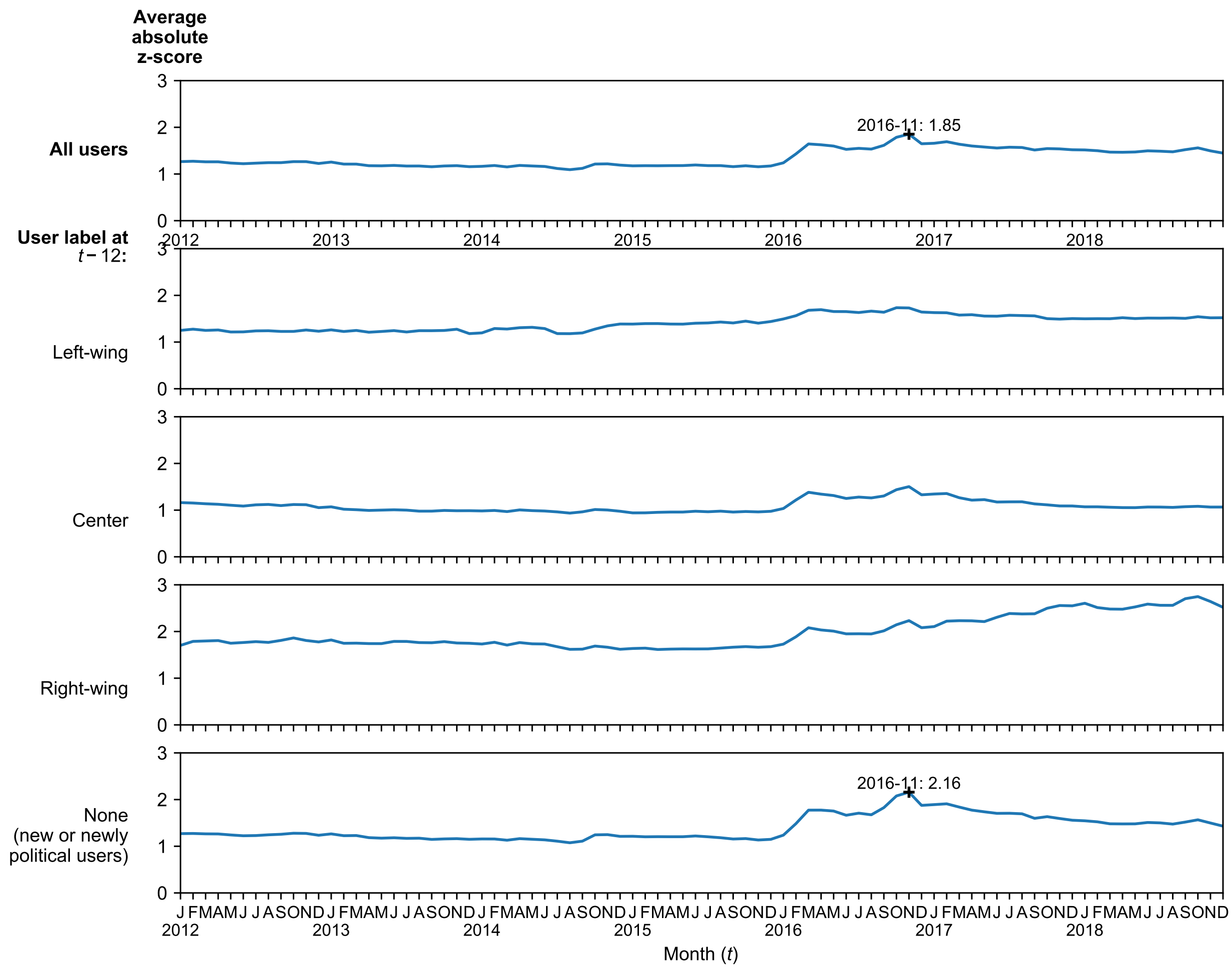

Figure A9: The average absolute z-score of political activity over time. Average absolute z-score is a measure of polarization and is the average number of standard deviations from the mean of comments during the month. Activity is broken up into user categories in the same manner as in Figure 4. November 2016 is annotated.

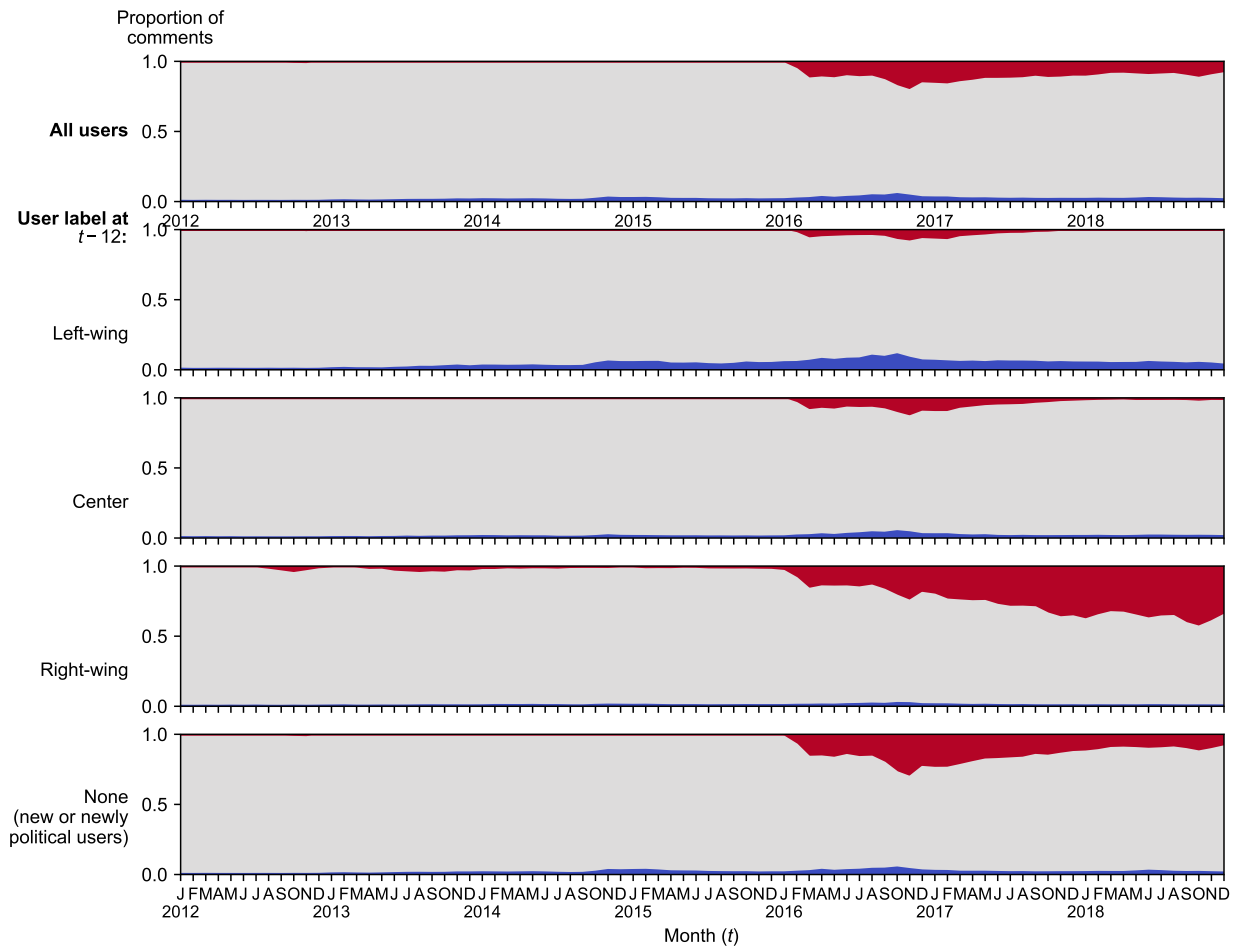

Figure A10: The proportions of activity in very left-wing (blue) and very right-wing (red) communities over time over time. Very left- and right-wing communities are defined as those communities which lie at least 3 standard deviations from the mean on either side. Activity is broken up into user categories in the same manner as in Figure 4.

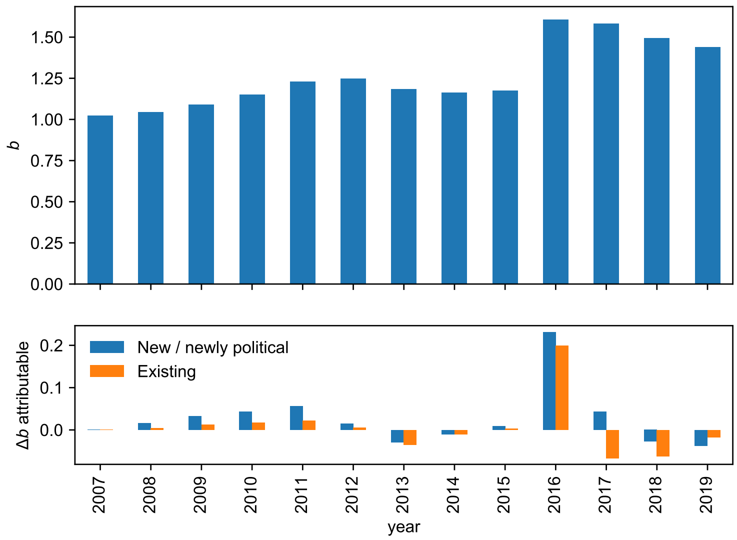

Figure A11: Top: Yearly average absolute partisan z-score of political comments (b). Bottom: The year-to-year change in b, broken down by user label. The change in the distribution of comments is calculated for each group of users by comparing their comments in a year y to the distribution of all comments made in the previous year $y-1$. This quantity is then multiplied by the fraction of comments in y made by users in that group to get the amount of overall change in b that is attributable to those comments. The sum of Δb attributable to both categories equals the overall change $\Delta b = b_y - b_{y-1}$.

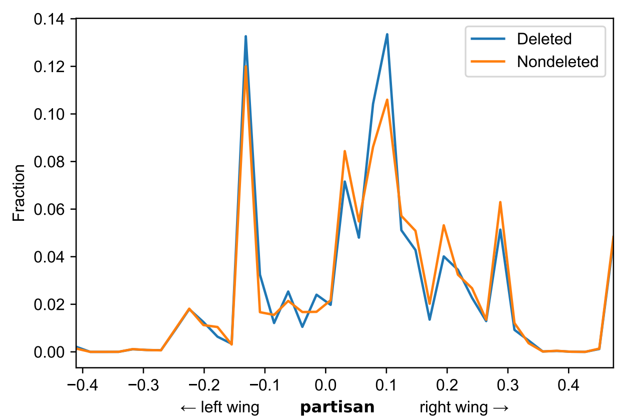

Figure A12: The distribution of deleted and non-deleted political comments on the `partisan` axis.

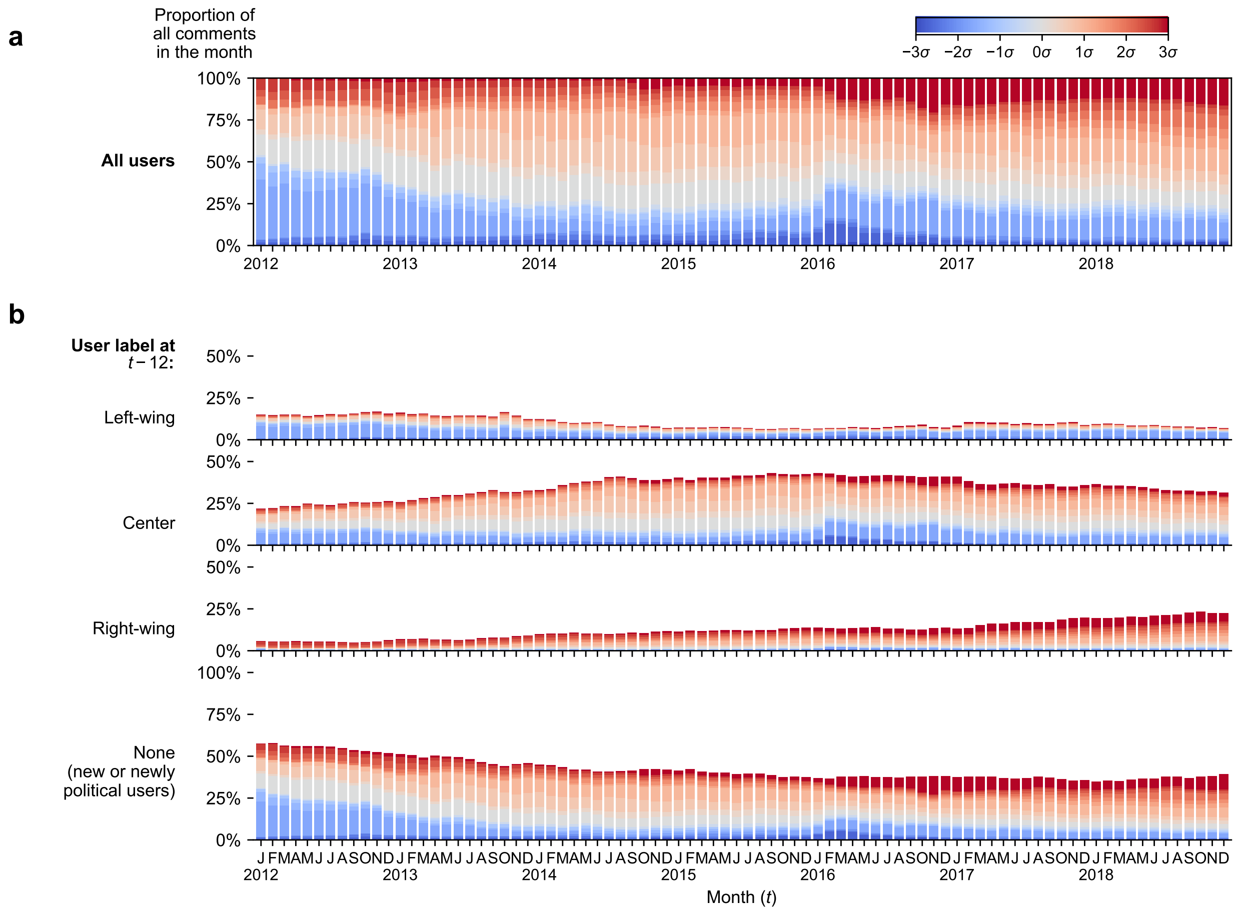

Figure A13: Figure 4 replicated using the partisan B dimension.

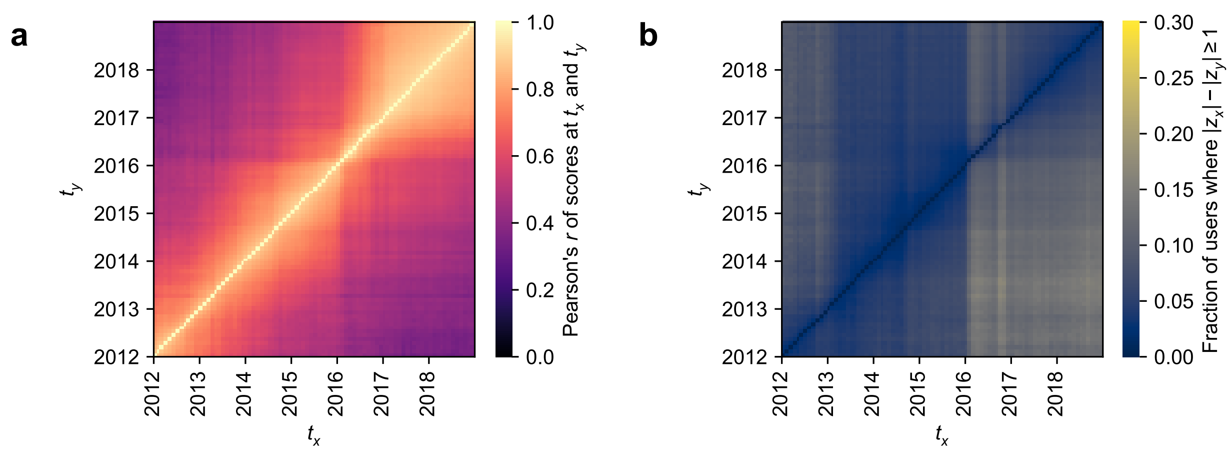

Figure A14: Figure 5 replicated using the partisan B dimension.

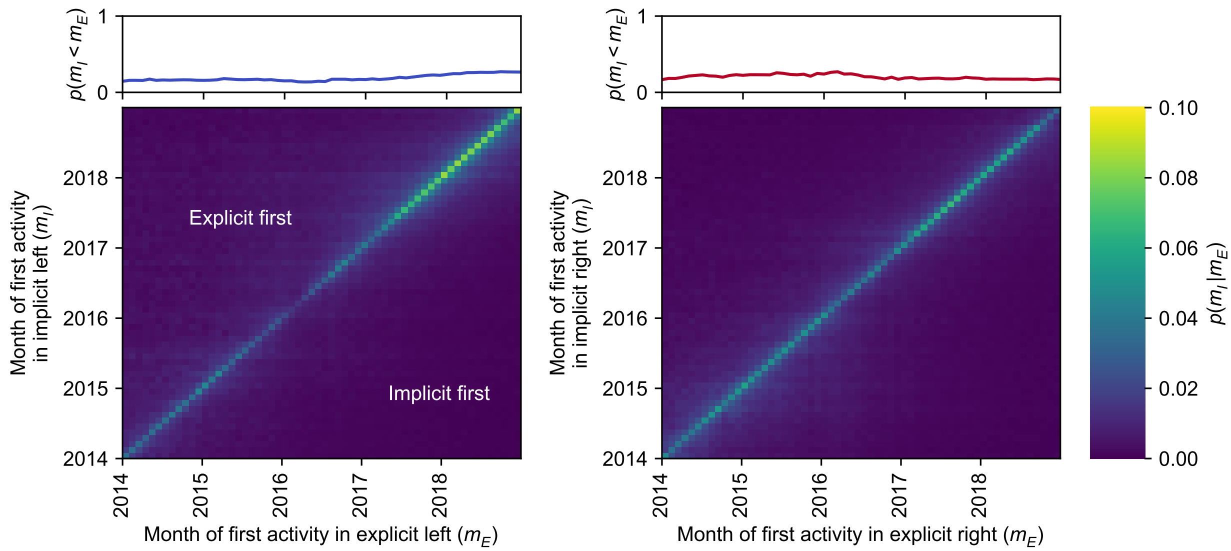

Figure A15: The relationship between explicitly partisan and implicitly partisan activity (left, left-wing; right, right-wing.) Of users who were first active in an explicitly partisan community at time x, the proportion of them who were first active in an implicitly partisan community at time y is denoted by the colour in cell (x, y). 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).

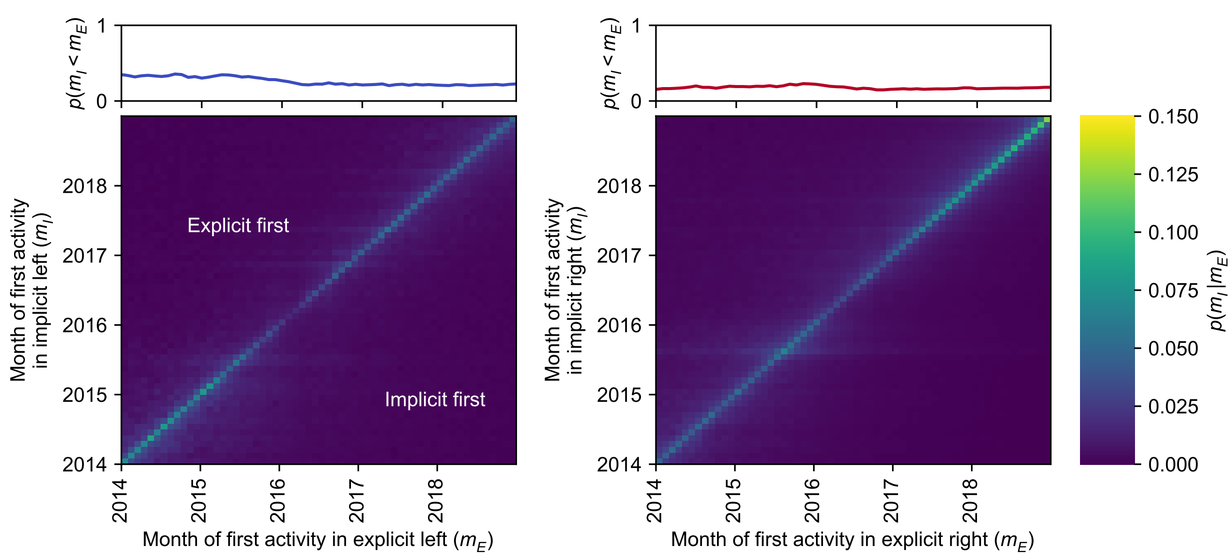

Figure A16: Figure A15 replicated using the partisan B dimension.

I Glossary of referenced communities

Each subreddit is listed along with its description, which is written by the community moderators of the subreddit. Subreddit descriptions were scraped in 2019 and 2020, and as a result, some descriptions are missing for communities which closed or were banned prior to scrape time.

Each community is labeled with its score on each of our main dimensions, by distance from the mean:

-3σ	A	Very young	G	Very masculine	P	Very left-wing
-2σ	A	Young	G	Masculine	P	Left-wing
$-\sigma$	A	Leaning young	G	Leaning masculine	P	Leaning left-wing
	A	Neutral	G	Neutral	P	Neutral
σ	A	Leaning old	G	Leaning feminine	P	Leaning right-wing
2σ	A	Old	G	Feminine	P	Right-wing
3σ	A	Very old	G	Very feminine	P	Very right-wing

0-9

4chan A G P
The 1 Ad-Free Reddit ""experience"" without the hassle of being quarantined

A

alaska A G P
Subscribed to by 5% of Alaska's Total Population!

androidthemes A G P
Showcasing our Android phones, one theme at a time!

APStudents A G P
No matter what course you are taking, we are a community that is forming in order to help students earn college credit!...

askaconservative A G P
Ask a Conservative: ask conservatives questions about conservative theory, values, policy and politics. For new conservatives,...

AskAnAmerican A G P
AskAnAmerican: Learn about America, straight from the mouth of Americans

AskMen A G P
r/AskMen is the premier place to ask random strangers for terrible dating advice, but preferably from the male perspective. A...

AskReddit A G P
/r/AskReddit is the place to ask and answer thought-provoking questions.

AskTrumpSupporters A G P
A Q&A Subreddit to help improve understanding of the views of Trump Supporters, and the reasons behind those views.

AskWomenOver30 A G P
A place for mature women redditors to discuss questions in a loosely moderated setting.

B

backpacking A G P
A subreddit for world traveling Backpackers. Often confused with the other type of (wilderness) backpacker in the US, this...

bannedfromclubpenguin A G P
Home to bans from Club Penguin, Club Penguin Island, and CPPSs....

Battlefield A G P
/r/battlefield

battlefield_4 A G P
The Official Battlefield 4 Subreddit.

2007scape A G P
The community for Old School RuneScape discussion on reddit. Join us for game discussions, weekly events and skilling...

actuallesbians A G P
A place for discussions for and by cis and trans lesbians, bisexual girls, chicks who like chicks, bi-curious folks, dykes,...

almosthomeless A G P
A place to ask for help, advice, or assistance for people who are imminently at-risk of becoming homeless.

answers A G P
Reference questions answered here.

ArtHistory A G P
This is a community of art enthusiasts interested in a vast range of movements, styles, media, and methodologies. Please feel free...

AskACountry A G P
Ask countries questions!

askhillarysupporters A G P
Ask supporters of Hillary Clinton for President

AskMenOver30 A G P
AskMenOver30 is a place for supportive and friendly conversations between over 30 adults....

AskThe_Donald A G P
A place to discuss Trump, his Administration, and MAGA related topics. Not a safe space.

AskWomen A G P
AskWomen: A subreddit dedicated to asking women questions about their thoughts, lives, and experiences; providing a place where...

ATV A G P
Reddit ATV United! Under one banner with mud, sand, mountains, and snow for all! A place to discuss ATVs, ATV news, and ATV...

BabyBumps A G P
A place for pregnant redditors, those who have been pregnant, those who wish to be in the future, and anyone who supports them.

badwomensanatomy A G P
Women are made of sugar and spice and all things nice. Except their vaginas which are sqwicky and attract bears.

bapccanada A G P
Inspired by a discussion on /r/bapcsalescanada the new /r/bapccanada is the /r/buildapc but for Canadians!

battlefield3 A G P
Battlefield3 discussion. Read the rules please. No Memes or blogspam....

BeardAdvice A G P
The definitive source for men who need answers to their bearded questions. Everyone is welcome to join in on the action....

beerwithaview A G P
Pictures contaning: Foreground - Beer Background - View

beyonce A G P
Bey, Queen, our R&B princess. Whatever you call her, this is the place to be on Reddit for everything Beyoncé.

BillBurr A G P
The Bill Burr subreddit. For fans of his stand up, cameos, and the Monday Morning Podcast.

blackops2 A G P
This Subreddit has moved to /r/CallofDuty!

BlueMidterm2018 A G P
A subreddit for Democrats to discuss the 2018 midterm elections - primaries, candidates, strategy, news, odds, funding,...

bostonhousing A G P
r/bostonhousing is a great resource for anyone looking for Boston apartments, rooms for rent in Boston, roommates in Boston,...

BroadCity A G P
Discuss and share Broad City-related stuff here! You can watch Broad City on Comedy Central's website or Hulu. Maybe it's even...

BSA A G P
This is a reddit for things relating to Scouts BSA and the Boy Scouts of America.

bestofworldstar A G P
The Best Of World Star Hip-Hop

bigboobproblems A G P
Vent in this judgment-free community that encourages discussion in a safe environment.

blackladies A G P
The face of Black women on Reddit!

blackops3 A G P
Call of Duty: Black Ops III is a military science fiction first-person shooter video game, developed by Treyarch and published by...

boston A G P
A reddit for the city of Boston, MA (featuring the cities of Cambridge, Somerville, Malden, Medford, Quincy, Braintree, Newton and...

breastfeeding A G P
This is a community to encourage, support, and educate mothers nursing babies/children through their breastfeeding journey....

brownbeauty A G P
This is a subreddit for makeup addicts of a darker coloring. As hard as it is to find quality products for darker skinned ladies...

C

canadacordcutters A G P
This subreddit is focused on educating Canadians on the legal, reasonably priced options, news and discussion in regards to...

Catholicism A G P
/r/Catholicism is a place to present new developments in the world of Catholicism, discuss theological teachings of the Catholic...

ChapoTrapHouse A G P
Chapo Trap House is a podcast by @willmenaker @cushbomb @amberaleefrost @virgiltexas @ByYourLogic, produced by @saywhatagain

Christians A G P
/r/Christians is a non-denominational community for Christianity that exists firstly for God's glory and secondly for...

climateskeptics A G P
Questioning climate related environmentalism.

conan A G P
Sub for Conan talk show on TBS

conservatives A G P
Conservatism (from conservare, "to preserve") is a political and social philosophy that promotes the maintenance of traditional...

cordcutters A G P
Are you tired of paying too much for cable television? Join us and become a cordcutter today. We offer advice on live streaming...

Cplusplus A G P
Ask questions, share news and projects about C++ here

CringeAnarchy A G P
The official sub Reddit room for all organized alt right trolls attacking everything Amy Schumer does.

crochet A G P
Crochet

CampingandHiking A G P
Hikers who bring Camping gear in their Backpack. Tips, trip reports, back-country gear reviews, safety and news....

CastleStory A G P
Castle Story is a mix between a real time strategy game and a voxel based sandbox building game. Created by a small company,...

CGPGrey A G P
Subreddit for CGP Grey stuff.

ChoosingBeggars A G P
NEXT!

ClashOfClans A G P
Welcome to the subreddit dedicated to the smartphone game Clash of Clans!...

COMPLETEANARCHY A G P
Just... The most Complete Anarchy.

Conservative A G P
The place for Conservatives on Reddit.

Cooking A G P
/r/Cooking is a place for the cooks of reddit and those who want to learn how to cook. Post anything related to cooking here,...

counter_strike A G P
For the Counter-Strike Gamer. Whether you are a seasoned veteran posting tips, trick and hints or new to the game and need a few...

CraftyTrolls A G P
Expanding the awesome TrollX and TrollY subreddit universe. Show us your skills! Ask about new ones! Make things!

cringepics A G P
An offshoot of /r/cringe, for those images that depict an awkward or embarrassing situation.

D

DaystromInstitute A G P
A subreddit for in-depth discussion about Star Trek .

democrats A G P
Offers daily news updates, policy analysis, links, and opportunities to participate in the political process. Feel free to discuss...

DNCleaks A G P
https://wikileaks.org/dnc-emails/ This subreddit was created to post details and significant finds from the DNC leak in a more...

Drumming A G P
A small but mighty subreddit that caters to all things drumming. Staffed by industry professionals, we hope that /r/Drumming...

DumpsterDiving A G P
Advice, information, and first-hand accounts about finding cool stuff in, or making cool stuff out of, trash.

daddit A G P
For geek, nerd or neuro-atypical dads.

deadisland A G P
First person zombie survival game by Techland....

dgu A G P
A subreddit dedicated to cataloging incidents in the United States where legally-owned or legally-possessed guns are used to deter...

dontstarve A G P
Everything about Don't Starve , a game by Klei Entertainment, creators of Mark of the Ninja, Shank and N+, among many others....

DrunkOrAKid A G P
This subreddit is for stories of the greatest stupidity. Inspired by How I Met Your Mother, this subreddit was created for the...

dyinglight A G P
Dying Light and Dying Light 2 are first person zombie survival games developed by Techland....

E

eldertrees A G P
A friendly haven for ents 18+.

Enough_Sanders_Spam A G P
Behold! /r/Enough_Sanders_Spam, Flame of the Establishment! Forged from the shills of /r/enoughsandersspam.

EnoughLibertarianSpam A G P
Sick of all the conspiracy theories, racism, anti-Semitism and general douchebaggery of libertarians? You are not alone! Award...

entertainment A G P
For news and discussion of the entertainment industry....

excatholic A G P
This subreddit is for any and all ex-Catholics to talk, educate, discuss and maybe even bitch about their experiences within the...

EnoughHillHate A G P
Community banned or deleted prior to publication

EnoughTrumpSpam A G P
Because the amount of Trump spam is too damn high! Enough Spam

EverWing A G P
Discussion on anything related to the Facebook Messenger in-app game 'EverWing'.

expats A G P
For anyone who has left their native country and moved to a new one.

F

FedoraCoin A G P
TiPS (a.k.a. FedoraCoin) is a new state of the art cryptocoin based on the Tips Fedora meme. Our objective is to become the...

FiftyFifty A G P
Risky Clicks the Subreddit

Firearms A G P
A place to discuss firearms and news relating to guns and other small arms. We value the freedom of speech as much as we do the...

fitness30plus A G P
A subreddit of /r/fitness for issues more relevant to individuals both male and female over the age of 30....

fo4 A G P
The Fallout 4 Subreddit. Talk about quests, gameplay mechanics, perks, story, characters, and more.

Forex A G P
Your Forex Trading community! /r/Forex is for traders who are serious about sharpening their skills and becoming consistently...

Frugal A G P
Frugality is the mental approach we each take when considering our resource allocations. It includes time, money, convenience, and...

FULLCOMMUNISM A G P
Welcome to the glorious subreddit of the glorious land of Lenin, Stalin, and Mao! Military service is compulsory and everyone...

funny A G P
Welcome to r/Funny: reddit's largest humour depository

Fallout4Builds A G P
Fallout 4 Builds

FierceFlow A G P
/r/FierceFlow is a subreddit for men with long (er) hair to share tips, progress pictures, anecdotes, or anything else.

findareddit A G P
Having trouble finding the reddit you need? Post here what you're looking for and someone can suggest a reddit for you!

fitbit A G P
A forum for discussion of all Fitbit-related products. Come ask questions, encourage/challenge others, and join a community making...

fo3 A G P
A community for Fallout 3 and everything related....

FolkPunk A G P
A community of punk folks, creating and enjoying folk punk music, and actively standing with Black Lives Matter.

freemasonry A G P
The main Reddit sub for Masonic Masons and Freemasonry

fuckolly A G P
Subreddit dedicated to fucking that piece of shit Olly from HBO's Game of Thrones series. ThanksObama'ed on May 8, but the party...

FullmetalAlchemist A G P
A subreddit for the Fullmetal Alchemist franchise, including the manga, the two anime, and all related content.

G

GamerGhazi A G P
Diversity and geek culture collide

geegees A G P
University of Ottawa/Université d'Ottawa

GlobalOffensive A G P
/r/GlobalOffensive is a home for the Counter-Strike: Global Offensive community and a hub for the discussion and sharing of...

Gore A G P
Community banned or deleted prior to publication

greysanatomy A G P
The subreddit for all your Grey's Anatomy and Private Practice Discussion! The show was created by Shonda Rhimes when it premiered...

guns A G P
A place for responsible gun owners and enthusiasts to talk about guns without the politics.

gameofthrones A G P
This is a place to enjoy and discuss the HBO series, book series ASOIAF, and GRRM works in general. It is a safe place regardless...

gatech A G P
A subreddit for my dear Georgia Tech Yellow Jackets.

ghettoglamourshots A G P
Where Faith in Humanity Comes to Die

googleplaydeals A G P
The best deals on the Google Play store!

GrassrootsSelect A G P
Grassroots Select has relocated to /r/Political_Revolution the Reddit branch of The Political Revolution a digital...

gtavmodding A G P
Gta V Modding

GunsAreCool A G P
The cost of 'cool'. Mass Shooter Tracker Data. Mass shootings. Tracking mass shootings via all guns, firearms, semi-automatics,...

HaircareScience A G P
This subreddit aims to provide resources for achieving better hair quality through scientific research in trichology, physiology,...

highseddit A G P
Self improvement and seduction relevant to high school.

H

hbo A G P
The official subreddit for HBO, discover full episodes of original series, movies, schedule information, exclusive video content,...

hiking A G P
The hikers' subreddit.

HillaryForPrison A G P
Hillary Clinton for Prison – LOCK HER UP!! IT'S WHERE SHE BELONGS. We believe Hillary Clinton should should be in federal prison...

hsxc A G P
Talk about the season, post race results, and discuss Cross Country in general!

hillaryclinton A G P
/r/hillaryclinton is a pro-Hillary Clinton forum to support Hillary Clinton. Join other Hillary Clinton supporters on Reddit! We...

hitchhiking A G P
Good information for the nomadic vagabonds out there. Not just limited to hitchhiking. Trainhopping, destinations, stories, etc....

I

IBO A G P
The subreddit for the International Baccalaureate Diploma Programme (IBDP). This subreddit encourages questions, constructive...

ImGoingToHellForThis A G P
Looking for a Japanese twink getting railed by two hunky latino men in a hot tub? Trying to figure out the name of the man getting...

IndieFolk A G P
A Subreddit for Indie & Folk music. Feel free to post away!

InteriorDesign A G P
Interior Design is the art and science of understanding people's behavior to create functional spaces within a building. It is a...

K

KitchenConfidential A G P
Home to the largest community of restaurant and kitchen workers on the internet.

L

lastweektonight A G P
Last Week Tonight with John Oliver is an American late-night talk show airing Sundays on HBO in the United States and HBO Canada,...

Leathercraft A G P
A subreddit for people interested in working with leather, sharing tips, and tricks. Learn more about leatherworking and share...

LGBTeens A G P
A place where LGBTeens and LGBT allies can hang out, get advice, and share content!...

liberalgunowners A G P
Gun-ownership through a liberal lens. This is a place for liberal gun-owners (this means leftist to you "classical liberals")...

librarians A G P
Discuss optimal filing methods, book rejuvenation, humidity regulation, and dusting practices....

Lost_Architecture A G P
r/Lost_Architecture, is a subreddit devoted to images and discussion of interesting buildings that no longer exists.

M

malehairadvice A G P
A subreddit dedicated to helping men improve their hairstyles....

malelivingspace A G P
MaleLivingSpace is dedicated to places where men can live. Here you can find posts discussing, showing, improving, and...

matt A G P
Come ye merry Matts and Matthews!

MeanJokes A G P
A collection of the cruelest, most offensive jokes you can think of.

memes A G P
Memes!

metacanada A G P
Canada's only not-retarded subreddit

MissingPersons A G P
A subreddit for all things related to missing people and their cases.

MorbidReality A G P
Welcome to /r/MorbidReality, a subreddit devoted to the most disturbing content the internet has to offer. Here, we study and...

MST3K A G P
A place for fans of Mystery Science Theater 3000 to share all things MST3K-related.

N

NewGirl A G P
A subreddit for fans of the show New Girl. Discussion of, pictures from, and anything else New Girl related.

niceguys A G P
For all the self proclaimed "nice guys" who are actually manchildren or douches, or mistake their hilarious spinelessness for...

NoFapChristians A G P
NoFapChristians is a safe place for Christian fapstronauts to discuss the process of abstaining from pornography and masturbation....

noveltranslations A G P
A place for Japanese light novels, and web novels that are translated from Japanese, Chinese, and Korean to English. Original...

Impeach_Trump A G P
For Americans Against Trump /r/Impeach_Trump

insanepeoplefacebook A G P
Do you have any insane people on your social media feeds? Post screenshots here!

irlsmurfing A G P
A celebrity or professional pretending to be amateur usually under disguise. The video has to be an activity that the person is...

ketodrunk A G P
A subreddit devoted to the careful craft of the low-carb drunk. Too many sugary cocktails and carb-laden beer finding their way...

KotakuInAction A G P
KotakuInAction is the main hub for GamerGate on Reddit and welcomes discussion of community, industry and media issues in gaming...

LadyGaga A G P
A sub-reddit for fans of Lady Gaga

law A G P
A place to discuss developments in the law and the legal profession.

lgbt A G P
A safe space for GSRM (Gender, Sexual, and Romantic Minority) folk to discuss their lives, issues, interests, and passions. LGBT...

LGBTnews A G P
/r/LGBTnews is for sharing links to recent news about lesbian, gay, bisexual, transgender, and all queer issues from a variety of...

Libertarian A G P
This subreddit is about the political philosophy of libertarianism, broadly speaking. We are in no way affiliated or associated...

LightNovels A G P
A community for those interested in the Light Novel medium, as well as Japanese Novels and Web Novels.

LSAT A G P
The best place on Reddit for LSAT advice. The Law School Admission Test (LSAT) is offered four times per year, and you must...

MaleFashionMarket A G P
A place for redditors to sell, buy or trade their previously-owned clothes, shoes and accessories.

malelifestyle A G P
The community of interest for man at his best....

marchingband A G P
A place for all of us marching band geeks to get together and share spicy memes, help each other out, or just spread the love....

me_irlgbt A G P
gay selfies of the gay

MeetPeople A G P
Here at /r/MeetPeople you can find people from everywhere to talk to, or hang out with!

menkampf A G P
In the post you're about to make, replace cis/white/hetero/male people with the Jews and if the result sounds like something that...

Minneapolis A G P
Minneapolis, Minnesota (MN)

Mommit A G P
We are people. Mucking through the ickier parts of child raising. It may not always be pretty, fun and awesome, but we do it....

Mr_Trump A G P
Community banned or deleted prior to publication

new_right A G P
New Right, Alternative Right, Traditionalist, Neoreaction and Dark Enlightenment: new right news and discussion for...

news A G P
/r/news is: real news articles, primarily but not exclusively, news relating to the United States. /r/news isn't: editorials,...

Nightshift A G P
For the people who work during the night

NoPoo A G P
"No Shampoo" - A place to discuss natural hair care and alternatives to shampoo. Girls & Guys welcome! 'No Poo'

nyc A G P
r/nyc, the subreddit about new york city

nycmeetups A G P
The subreddit for discussing New York City reddit meetups....

O

OneY A G P
A place to thoughtfully discuss issues that affect men of the world today. Everyone is welcome but intolerance is not.

opencarry A G P
/r/opencarry is a reddit community which focuses on issues which face advocates of the open carry of firearms in America.

OMSCS A G P
A place for discussion for people participating in GT's OMS CS

ontario A G P
A subreddit to discuss all the news and events taking place in the province of Ontario, Canada.

OpenChristian A G P
OpenChristian is a community dedicated to Progressive Christianity. This is a space where progressive Christians can support each...

P

pagan A G P
Contemporary Paganism

paris A G P
La Ville-Lumière. Posts en Anglais et en Français.

parentsofmultiples A G P
A place for parents of twins, triplets, and beyond to discuss the unique challenges of raising and parenting multiples.

pearljam A G P
A subreddit about all things Pearl Jam. Of course we're talking about the best band from the 1990s featuring Eddie Vedder, Mike...

peeling A G P
peeling

personalfinance A G P
Learn about budgeting, saving, getting out of debt, credit, investing, and retirement planning. Join our community, read the PF...

pickuplines A G P
A subreddit for all your pick up line needs. Yes, our icon is a line drawing of a pickup....

pics A G P
A place to share photographs and pictures.

pinball A G P
The subreddit for pinball lovers, collectors, and competitors.

PoliticalHumor A G P
A subreddit for political humor (particularly US politics), such as political cartoons and satire.

PoliticalVideo A G P
A great place for video content of the political kind. Politics through videos.

politics A G P
/r/Politics is for news and discussion about U.S. politics.

predaddit A G P
For men about to become fathers

PressureCooking A G P
Pressure Cooking!

progressive A G P
A community to share stories related to the growing Modern Political and Social Progressive Movement. The Modern Progressive...

progun A G P
For pro-gun advocacy!...

projectmanagement A G P
This reddit is dedicated to project management with a focus on the software development area. No commercial promotion allowed.

prolife A G P
A place for Pro-Lifers of all religious, secular and political views to gather on Reddit....

prowrestling A G P
Y our arena for the enjoyment of the performance art and pseudo-sport aspects of pro wrestling. Great wrestling from around...

PS3 A G P
The PlayStation 3 Subreddit (PS3, PlayStation3, Sony PlayStation 3) The Official FAQ Your place for: News Reviews...

ps3bf3 A G P
Battlefield 3 on PS3

PS4 A G P
The largest PlayStation 4 community on the internet. Your hub for everything related to PS4 including games, news, reviews,...

R

R4R30Plus A G P
Come in and meet people over 30! r4r30plus - Redditors for Redditors over 30 Whether you're looking for friends, partners,...

racism A G P
Reddit's anti-racism community, a safe(r) space for People of Color and their supporters. All discussions are expected to be from...

rapbattles A G P
Rap Battles and anything to do with the Battlerap movement around the world.

RedditForGrownups A G P
This is a community for Redditors that are starting to get that "get off my lawn" feeling whenever they check their front page. So...

RedHotChiliPeppers A G P
A community for RHCP fans to share music videos, personal stories, pictures, documentaries, Frusciante solo material, Ataxia, Dot...

relationship_advice A G P
Need help with your relationship? Whether it's romance, friendship, family, coworkers, or basic human interaction: we're here to...

ROTC A G P
Reserve Officers' Training Corps

RotMG A G P
A community-driven subreddit for the online Flash-player bullet-hell perma-death game, Realm of the Mad God.

running A G P
Runners welcome....

rupaulsdragrace A G P
Do you have what it takes? Only those with Charisma, Uniqueness, Nerve and Talent will make it to the top! Start your...

S

sailormoon A G P
A subreddit for fans of the Sailor Moon franchise. Please remember to read in the sidebar, and please read The Sailor Moon FAQ...

SandersForPresident A G P
Midterm Bern & Bernie 2020! Supporting Senator Bernie Sanders, Our Revolution, National Nurses United, Justice Democrats,...

sanfrancisco A G P
Cold summers, thick fog, and beautiful views. Welcome to the subreddit for the gorgeous City by the Bay! San Francisco,...

saplings A G P
r/saplings: a place to learn about cannabis use and culture

Savant A G P
This subreddit is dedicated to the norwegian music producer Aleksander Vinter, known otherwise as Savant, Blanco, Vinter in Vegas...

SchoolIdolFestival A G P
A subreddit made for the mobile rhythm game Love Live! School Idol Festival. All SFW LL!SIF content welcome!...

sewing A G P
This is a community specifically for sewing including, but not limited to: machine sewing, embroidery, quilting, hand sewing,...

SFr4r A G P
SFr4r: An R4R fork for the SF Bay Area

shittyfoodporn A G P
shittyfoodporn

skeptic A G P
skeptic

spongebob A G P
The subreddit about Spongebob Squarepants. The show, characters, and other spongebob stuff.

SRSWomen A G P
Community banned or deleted prior to publication

stonerrock A G P
Stoner Rock

SummerReddit A G P
A place to chronicle the posts that are popular in summertime....

T

TagPro A G P
Suggestions and discussion for the free-to-play webgame TagPro.

techwearclothing A G P
Practical, functional and utilitarian clothing.

TeenMFA A G P
Community banned or deleted prior to publication

texts A G P
/r/texts - a subreddit to submit your funny, weird, or random text messages from your mobile or cell phone....

The_Farage A G P
/r/The_Farage is the best subreddit of the british god man himself.

TheGirlSurvivalGuide A G P
This subreddit was created for girls to request tips and share discoveries to aid others in daily life. A survival guide of "life...

trackandfield A G P
Track and Field

travel A G P
r/travel is a community about exploring the world. Your pictures, questions, stories, or any good content is welcome. Clickbait,...

TrollYChromosome A G P
/r/TrollYChromosome is going private to protest against Reddit continuing to provide a platform for racism is hate. 75 Things...

TrueChristian A G P
A subreddit for Christians of all sorts. We exist to be a safe place for discussion between believers on all sides of the fence;...

TumblrInAction A G P
A true force for good!

U

uncensorednews A G P
Community banned or deleted prior to publication

unexpectedjihad A G P
ayy...

USMilitarySO A G P
Subreddit for sharing advice, support and information for the significant others of current and past members of the United States...

V

vegetarianketo A G P
We're a small community of like-minded people attempting the Ketogenic diet while avoiding the typical staple food group, meat....

videos A G P
The best place for video content of all kinds. Please read the sidebar below for our rules.

W

ShitRConservativeSays A G P
Have you ever suspected r/conservative is just liberal satire and most of the moderators are trying to make conservatives look...

Shittypokestops A G P
A subreddit for the shittiest Pokéstops in the world of Pokémon Go. The shittier the better. What denotes a shitty Pokéstop you...

SocialistRA A G P
Never disarm the working class!

SRSsucks A G P
Remember, do not misgender a SRSter, ever. Please refer to them as "It" or "THNG" when "BRD" wont suffice.

startrekgifs A G P
A sub to post gifs that are all things Trek

subredditoftheday A G P
Subreddit of the Day ... is a celebration of the interesting communities on reddit.com. Once a day we shine a spotlight on the...

TACSdiscussion A G P
/r/TACSdiscussion has moved to /r/CompoundMedia Peckahs.

TallMeetTall A G P
For tall people who want to meet other tall people

teenagers A G P
r/teenagers is the biggest community forum run by teenagers for teenagers. Our subreddit is primarily for discussions and memes...

teenrelationships A G P
A subreddit with the goal is helping teens work through their relationships....

The_Donald A G P
The_Donald is a never-ending rally dedicated to the 45th President of the United States, Donald J. Trump.

Theatre A G P
Theatre; theory, design, news and community.

TheWayWeWere A G P
What was normal everyday life like for people living 50, 100, or more years ago? Featuring old photos, scanned documents,...

trailrunning A G P
The fun begins where the road ends....

travelpartners A G P
Meet up with a fellow traveller or travel buddy to visit the world

TrueAtheism A G P
A subreddit dedicated to insightful posts and thoughtful, balanced discussion about atheism specifically and related topics...

TrueDetective A G P
We get the world we deserve.

twinpeaks A G P
A subreddit for fans of David Lynch's and Mark Frost's wonderful and strange television series. We live inside a dream...

u_washingtonpost A G P
Democracy Dies in Dankness. Official account. Our award-winning journalists have covered Washington and the world since 1877....

UnconventionalMakeup A G P
This is for those of us that don't always fit into the meta looks of MUA. For those that like to do SFX makeup, goth makeup,...

UnexpectedThugLife A G P
The subreddit for videos where there's some unexpected thug behavior showing when least expected....

vagabond A G P
A digital community created by vagabonds, for vagabonds! Hitchhikers, Trainhoppers, Rubbertramps, Backpackers, and more! Feel...

Veterans A G P
This is a subreddit for news, sites, information and events that may interest veterans. We are here to support one another and...

watchpeopledie A G P
Welcome to watchpeopledie. This community is intended to observe and contemplate the very real reality of death. We are attempting...

watchpeoplesurvive A G P
r/watchpeoplesurvive is like r/watchpeopledie, but dedicated to people surviving near misses, eg: - Car accidents - Plane crashes...

wii A G P
Wii games and scene news

windmobile A G P
A discussion of WIND Mobile products and services for existing users, those thinking of making the switch, or people just wanting...

womensstreetwear A G P
Welcome to Women's Streetwear! Feel free to post any questions, inspirations, pick ups, fit pics, and news regarding women's...

WWE A G P
A subreddit for fans of World Wrestling Entertainment. This includes WCW, ECW, NXT and whatnot.

waterpolo A G P
A sports subreddit dedicated to everything water polo related.

wiiu A G P
Reddit's source for news, pictures, reviews, videos, community insight, & anything related to Nintendo's 8th-generation console,...

women A G P
A safe, respectful space to discuss the lives and stories of women of all backgrounds, and the current events which affect us....

worldnews A G P
A place for major news from around the world, excluding US-internal news....

X

xboxone A G P
Everything and anything related to the Xbox One. News, reviews, previews, rumors, screenshots, videos and more!

xxketo A G P
/r/xxketo is a subreddit dedicated to discussing a ketogenic diet from a female-identifying perspective

xbox360 A G P
Everything and anything related to the Xbox 360. News, reviews, previews, rumors, screenshots, videos and more! Note: We are not...

XboxOneGamers A G P
With many people not upgrading consoles or switching to the other side, friends are hard to find, this subreddit will help make...

Y

Yosemite A G P
Yosemite

YAlit A G P
Young Adult Literature...

youngatheists A G P
A place for young atheists to share their experiences and questions. We welcome all for a nice slice of skepticism and a cold...

Z

Zappa A G P
All that is Frank Zappa (fl. 1940-1993)