

“More Polarized than Ever?” Evidence from Social Media

Veronica Cagle, Stephen Davies, Thomas J. Davies, Alexis Kochanski
(vcagle@umw.edu, stephen@umw.edu, daviestj@dukes.jmu.edu, alexis.kochanski@duke.edu)

Keywords: political polarization, social media, text classification, machine learning

Extended Abstract

For years, a common refrain in discussions about the state of contemporary U.S. politics is that the country is becoming “more politically polarized” (see, *e.g.*, [4,5,7,13,14]). Scholars have come to widely different conclusions about this claim, however, depending on who they are measuring — members of Congress[3], party “elites”[6], or the public at large[6,8] — and on how it is measured — *e.g.*, ideological consistency[13], party sorting[2,9], cultural subgroups[10], geographical clustering[12], or the perceptions of presidential candidates[11]. And even regarding specifically whether the ideological polarization of the masses has increased, experts have reached strikingly different verdicts (see, for example, [1] vs [9]).

We seek to shed light on this question by measuring the prevalence of politically polarized social media posts over an extended length of time. Online discourse is only one way polarization can manifest itself, of course, but it is an increasingly important one, and we believe that if markers of increased polarization are indeed increasing in our society, they could hardly fail to appear in the social media sphere.

Because of its popularity, anonymity of authorship, and easy delineation by subject matter, we chose to analyze Reddit, the American social news aggregation and discussion site. Reddit offers an open API through which one can download comment “threads”; that is, hierarchically-nested structures of user-generated replies to posts. We selected for analysis 25 politically-themed “subreddits” (forums devoted to specific topics) and collected the text of comment threads from each subreddit’s inception to the present day. This comprised 6,263,110 threads in total.

To judge the level of polarization of each of these threads, we trained and implemented a machine learning classifier — specifically, a multilayer perceptron (MLP) with two 64-unit layers, using as features the most informative 1,000 unigrams and bigrams present in the text. For training data, we first outlined a target definition of “polarized” for a team of eight undergraduates; this involved the presence of characteristics such as a combative style, an unconvincible mindset, and/or a purely argumentative purpose. This team then individually analyzed a set of threads chosen randomly from the 25 political subreddits, and rated each one as polarized, not polarized, or other. We threw out the “other” ratings, and then discarded any thread that did not have at least four unanimous ratings. The result was a training set of 522 hand-tagged threads, each of which was unanimously judged as either polarized or non-polarized by at least four raters.

Depending on parameter choices, the classifier performed with an accuracy of between 75% and 80% on an independent test set. Although we continue to explore improving the classifier using word embeddings, data bootstrapping, and hand-prioritization of features, we consider this a rather auspicious performance, considering that political polarization is a very nuanced phenomenon and human raters often disagree on whether it is present.

Running the classifier on all 6+ million threads, and computing the percentage of threads in each month that are classified as polarized, produces the plot in Figure 1. There does not appear to be a strong pattern indicating a rise in polarization overall on Reddit during the period 2008 to the present. (The higher variance on the left side of the plot is most likely a product of having fewer threads to rate in the early years, not a higher volatility of polarization.) The modest spike in the most recent year — when average polarization jumped from about 16% to about 24% — is interesting and warrants further analysis.

Figure 2 presents the same data, but broken down by subreddit. (As before, the wild swings in some of the subreddits, especially during their early years, are most likely the result of a paucity of data during that time period.) There are a few noteworthy takeaways from this plot. One is that although Figure 1 did not show a consistent rising trend, certain subreddits seem to do so. (Consider the subreddit “liberal,” shown as the red line in the upper middle pane.) Another is that although the state-oriented subreddits (bottom row) seem too noisy to draw firm conclusions about, the group of large national-oriented subreddits in the upper-left do seem to display a small but noticeable increase in about 2015. Finally, and perhaps most intriguing, is the prominent downturn in polarization in the “republican” subreddit in 2016. After two or three years far below its general average, the polarization of this subreddit regresses to its mean only in the recent past. The possible causes of this are interesting to contemplate.

References

- [1] A. I. Abramowitz and K. L. Saunders, “Is Polarization a Myth?,” *The Journal of Politics*, vol. 70, no. 2, pp. 542–555, 2008.
- [2] S. J. Abrams and M. P. Fiorina, “Party sorting: The foundation of polarized politics,” *American gridlock: the sources, character, and impact of polarization*, pp. 113–29, 2015.
- [3] C. Andris, D. Lee, M. J. Hamilton, M. Martino, C. E. Gunning, and J. A. Selden, “The Rise of Partisanship and Super-Cooperators in the U.S. House of Representatives,” *PLOS ONE*, vol. 10, no. 4, p. e0123507, 2015.
- [4] D. Baldassarri and P. Bearman, “Dynamics of political polarization,” *American sociological review*, vol. 72, no. 5, pp. 784–811, 2007.
- [5] L. Boxell, M. Gentzkow, and J. M. Shapiro, “Cross-Country Trends in Affective Polarization,” National Bureau of Economic Research, Working Paper 26669, Jan. 2020.
- [6] R. L. Claassen and B. Highton, “Policy Polarization among Party Elites and the Significance of Political Awareness in the Mass Public,” *Political Research Quarterly*, vol. 62, no. 3, pp. 538–551, 2009.
- [7] M. Dimock and R. Wike, “America is exceptional in the nature of its political divide,” *Pew Research Center*, 2020. <https://www.pewresearch.org/fact-tank/2020/11/13/america-is-exceptional-in-the-nature-of-its-political-divide/>
- [8] J. H. Evans, “Have Americans’ Attitudes Become More Polarized?—An Update,” *Social science quarterly*, vol. 84, no. 1, pp. 71–90, 2003.
- [9] M. P. Fiorina and S. J. Abrams, “Political polarization in the American public,” *Annu. Rev. Polit. Sci.*, vol. 11, pp. 563–588, 2008.
- [10] A. Hartman, *A War for the Soul of America, Second Edition: A History of the Culture Wars*, Second edition. University of Chicago Press, 2019.
- [11] M. J. Hetherington, M. T. Long, and T. J. Rudolph, “Revisiting the Myth: New Evidence of a Polarized Electorate,” *Public opinion quarterly*, vol. 80, no. S1, pp. 321–350, 2016.
- [12] P. A. Klinkner and A. Hapanowicz, “Red and Blue Déjà Vu: Measuring Political Polarization in the 2004 Election,” *The forum : a journal of applied research in contemporary politics*, vol. 3, no. 2, 2005, doi: [10.2202/1540-8884.1079](https://doi.org/10.2202/1540-8884.1079).
- [13] Pew Research Center, “Political Polarization in the American Public,” *Pew Research Center - U.S. Politics & Policy*, 2014. <https://www.pewresearch.org/politics/2014/06/12/political-polarization-in-the-american-public/>.
- [14] Washington Post, *Why is the U.S. so polarized? A neuroscientist and political scientist explain why*, (Dec. 26, 2020). Available: https://www.youtube.com/watch?v=8O2X5kj_-Kk

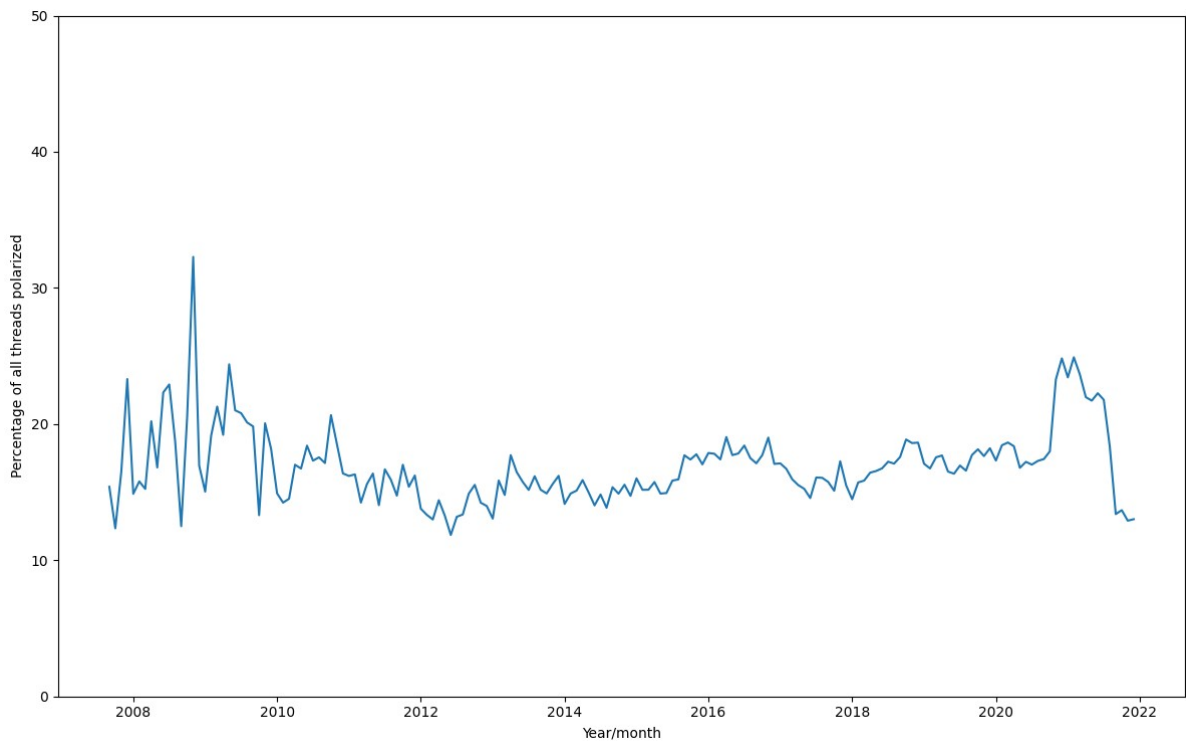


Figure 1. The overall percentage of threads polarized by month.

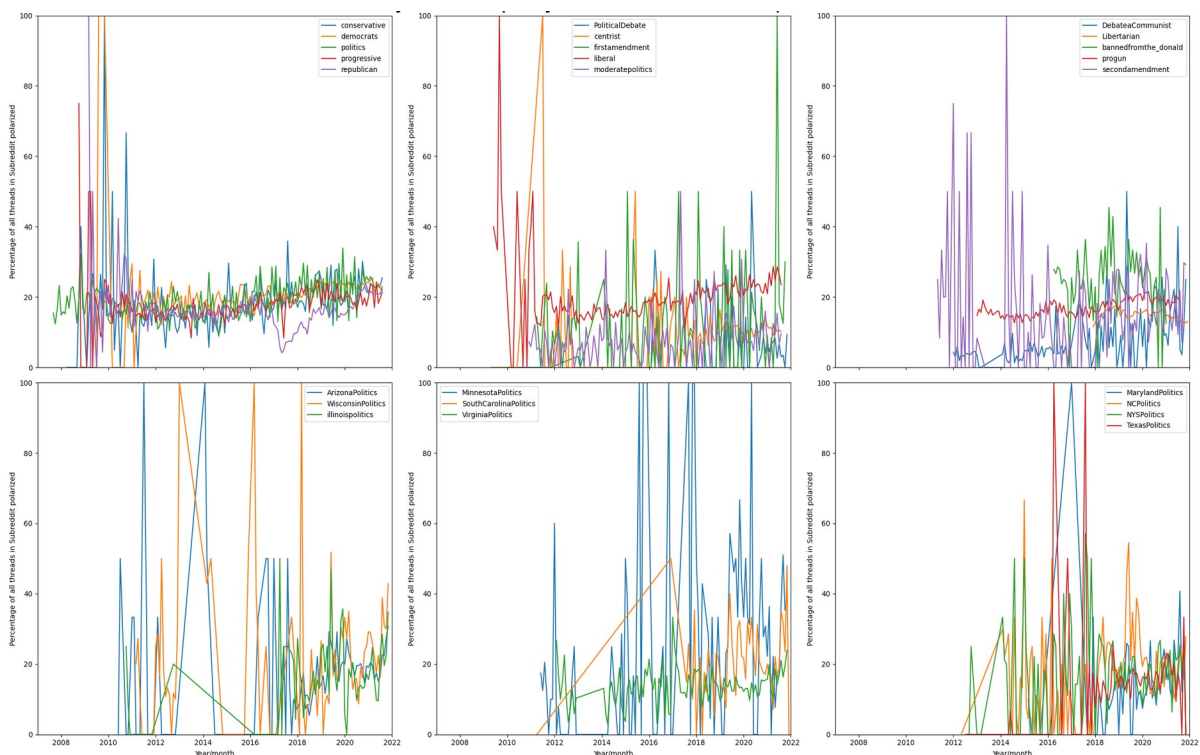


Figure 2. The percentage of threads polarized per month, by subreddit.