Multidimensional online American politics: Mining emergent social cleavages in social graphs

Pedro Ramaciotti Morales

▶ **To cite this version:**

Pedro Ramaciotti Morales. Multidimensional online American politics: Mining emergent social cleavages in social graphs. The 11th International Conference on Complex Networks and their Applications, Nov 2022, Palermo, Italy. hal-03721438

HAL Id: hal-03721438
https://hal.science/hal-03721438

Submitted on 12 Jul 2022

Multidimensional online American politics: Mining emergent social cleavages in social graphs

Pedro Ramaciotti Morales

médialab, Sciences Po & LPI, Université Paris Cité, Paris, France
pedro.ramaciottimorales@sciencespo.fr

Abstract. Dysfunctions in online social networks (*e.g.*, echo chambers or filter bubbles) are studied by characterizing the opinion of users, for example, as Democrat- or Republican-leaning, or in continuous scales ranging from most liberal to most conservative. Recent studies have stressed the need for studying these phenomena in complex social networks in additional dimensions of social cleavage, including anti-elite polarization and attitudes towards changing cultural issues. The study of social networks in high-dimensional opinion spaces remains challenging in settings such as that of the US, both because of the dominance of a principal liberal-conservative cleavage, and because two-party political systems structure preferences of users and the tools to measure them. This article builds on embedding of social graphs in multi-dimensional ideological spaces and NLP methods to identify additional cleavage dimensions linked to cultural, policy, social, and ideological groups and preferences. Using Twitter social graph data I infer the political stance of nearly 2 million users connected to the political debate in the US for several issue dimensions of public debate. The proposed method shows that it is possible to identify several dimensions structuring social graphs, non-aligned to liberal-conservative divides and related to new emergent social cleavages. These results also shed a new light on ideological scaling methods gaining attention in many disciplines, allowing to identify and test the nature of spatial dimensions mined on social graphs.

Keywords: social graphs, graph embedding, network homophily, ideological scaling, ideologies, polarization.

1 Introduction

The study of socio-political dysfunctions or disorders unfolding in digital social media and social networks [7] has raised to prominence in the past decade, including studies of algorithmic bias [3], extremism [21], or echo chambers [5]. These studies hinge on ontologies for the political positions or stances of online users. Bakshy et al. [3], for example, classified users and content on Facebook as Democrat- or Republican-leaning to analyze cross-cutting recommendations, and Barbera et al. [5] positioned Twitter users on liberal-to-conservative continuous scales to investigate the so-called echo chambers. These *categorical* (*e.g.*, user classification) or *geometrical* (*e.g.*, opinion scales) have been leveraged in several

studies. In many settings, including that of the US, categorical approaches are limited to binary classifications, while geometrical ones are often limited to one-dimensional opinion scales. This stems from the fact that the US is in essence a two-party system, and that the US has undergone several decades of *issue alignment* [23]: the alignment of issue positions on several matters, resulting in highly-correlated views on gun control, abortion, racial issues, among several others [18]. Methods for embedding social networks in higher dimensional opinion spaces (with dimensions standing for opinion indicators for different issues of social cleavage) have been developed only recently. These methods allow to tackle, for example, several European socio-political settings, known to require higher *dimensionality* [2] to account for observed social choice data, from roll call voting [10] to online social network activity [28]. Accordingly, recent methods have proposed to embed online social graphs in empirical geometrical spaces of several dimensions, where dimensions stand for indicators of opinions on immigration, income redistribution, or ideological stances such as left- and right-wing positions [27]. In settings such as that of the US and that display a main dominant dimension of social cleavage, there is comparatively less development on methods to mine additional dimensions. In this context, *cleavages* refers to historically-determined political, social or cultural dimensions dividing society in the sense that individuals into groups that structure political debate [15]. The study of online social networks in higher opinion dimensions calls nonetheless for renewed efforts, as recent research highlights the importance of new additional cleavages, in particular relating to anti-elite sentiment and populism, as driving forces in many relevant political and informational phenomena in the US [32].

This article builds on recent ideological scaling [14] and aforementioned graph embedding methods for spatializing social graphs in multi-dimensional ideological spaces [27]. Exploiting graph embedding and NLP methods, it proposes a method to mine cleavage dimensions linked to cultural, policy, social, and ideological groups and preferences in social graphs. The method is applied to Twitter social graph data of nearly 2 million users strongly connected to the online political debate in the US. The proposed method shows that several cleavage dimensions traditionally considered in social network analysis (*e.g.*, conservatism, gun control, patriotism, religion) are indeed strongly related, as most studies find, but further allows to quantitatively measure the relative alignment of these issues. More importantly, the method identifies and computes positions of large numbers of users in emerging, non-aligned dimensions identified in political science research as emerging cleavages, and that several works in social science have hypothesized and put forward in recent years. These results extend previous research works showing that it is possible to spatialize users in the US in several space dimensions that are not highly correlated. These results also shed light on the ubiquitous practice of analyzing social networks with ideological scaling methods, by showing ways to assess its validity and improving existing methods. While traditional ideological scaling methods yield a scale that is highly correlated with Democrat-Republican cleavages (resulting from an issue align-

ment process), party cleavage in specifically can be more accurately determined among many other aligned cleavages.

2 Related Work

Binary categorical classification of social media and network users counts numerous works, from exploitation of self-reporting and surveys [3] to sophisticated methods using neural networks on heterogeneous graphs [33]. More related to this article, ideological spatialization of relational behavioral data can be traced back at least to the Nominal Three-Step Estimation (NOMINATE) method [22]. The NOMINATE method is used to position parliamentarians and bills in one-dimensional liberal-conservative scales. It assumes 1) the existence of unobservable geometrical ideological parameters or positions, and 2) an underlying homophilic generative process (*i.e.*, parliamentarians vote to approve bills that are ideologically *close* to them in the unobservable space), and uses roll call data in Bayesian inference of the unobservable parameters. This process, called *ideological scaling*, has been used in numerous settings involving social choice: court rulings, campaign donations, parliamentary vote, and social network activity among others [14]. In its most wide-spread form, the homophilic probabilistic generative process is modeled as [4]:

$$P\left(i \leftarrow j | \alpha, \beta, \boldsymbol{\phi}_i, \boldsymbol{\phi}_j\right) = \text{logit}^{-1}\left(\alpha - \beta \|\boldsymbol{\phi}_i - \boldsymbol{\phi}_j\|^2\right), \qquad (1)$$

where the probability of observing user j interacting with user i (*i.e.*, $i \leftarrow j$) depends on position and scale parameters α and β, and, most importantly, on the distance between the unobservable positions $\boldsymbol{\phi}_i$ and $\boldsymbol{\phi}_j$ of users i and j. Social choice data (*i.e.*, pairs $i \leftarrow j$) can then be used to infer positions $\boldsymbol{\phi}_i$ for all users i. A large number of works assumes that one dimension is enough to retrieve the main social cleavage in the US, implicitly assuming it is the liberal-conservative one, and uses social network choice data (*e.g.*, Facebook likes [8] or Twitter following link [4]) to compute the position of users in this dimension. These works often rely on *ex post* validation using text cues to argue that the latent dimension indeed represents the political positions of users. Multi-dimensional inference for $\boldsymbol{\phi}_i$ can be achieved in a computationally-tractable manner with Correspondence Analysis [12] ,as it has been shown to approximate the inference of unobservable parameters of (1), both theoretically [16] and empirically [5].

Because (1) depends on unobservable parameters $\boldsymbol{\phi}$ through pairwise distances, their inference is invariant to isometric transformations. In particular rotation transformations mean that retrieved dimensions cannot be assured to be aligned with social cleavages that might be structuring social choice. In European settings, using the position of referential users such as politicians of known political parties, and party positions in reference issue spaces (provided, *e.g.*, by political polls or surveys), it has been shown how inferred dimensions display different levels of alignment with issues of public political debate [28]. This means that, in general, it cannot be assured that a single-dimensional ideological scaling model will yield a political opinion scale completely aligned with some

presumed main left-right or liberal-conservative cleavage. Using the position of several political parties, this fact has been leveraged in embedding large numbers of users in multi-dimensional space where dimensions stand for identifiable and separate political issues, not requiring *ex post* interpretation or validation [27]. Because the US is in essence a two-party system, and because of the difficulty in obtaining reference users that have both 1) known positions ϕ in latent space and 2) positions known through political surveys (in particular for issues that might not be highly correlated with liberal-conservative cleavages), similar methods have proved to be more challenging in this setting. This article thus proposes a method for constructing such referential positions for large numbers of users, relating to hypothesized cleavages, based on text descriptions written by users in their online profiles on social network sites. Using both, multidimensional scaling and embedding based on (1), and NLP methods for constructing groups of referential users, this article proposes a method to mine several spatial opinion dimensions, and most importantly, emergent opinions that are not highly-correlated with liberal-conservative cleavages.

3   Social Network Data

Following multidimensional ideological scaling works in Europe [28] and in the US [5], we select a bipartite sub-graph of the Twitter social graph. To capture online social choices that might be revealing of several social and political preferences we take as reference users members of the US Congress. We manually annotate the Twitter accounts of 550 members of the 116th United States Congress (looking for verified accounts corresponding to each congressperson), and collected their 17 952 824 followers (collection performed using Twitter's API in October 27th, 2020, see the Acknowledgements section for privacy-compliance information and references). To minimize the probability of followers being *bots*, we follow criteria adopted by several studies [28, 29, 19, 24] and further identified followers with more than 25 followers (7 325 940), and users that have posted more at least 100 tweets (7 471 365). See [4] for further details behind the rationale for these parameters. To identify users that are strongly connected to political debate and that follow spatial preference models we identify followers that follow at least three members of congress (3 846 925) [4]. We select the 1 821 272 unique followers that satisfy all three conditions.

To establish reference points in latent space, we collect the text self-descriptions made by users in their Twitter profiles (also on October 27th, 2020). Out of 1 821 272 users, 1 442 716 had written any text entry in their Twitter profiles. This collection, performed in the days leading to the 2020 US Presidential Election allows us to investigate cleavages in candidate preferences. The strategy presented in this article consists in finding space dimensions capable of dichotomizing pairs of groups of users identified with keywords used in their profiles, and revealing of social cleavages. We follow several studies to identify several possible cleavages of interest, including [22, 6, 27, 32]: party, candidate, racial cleavages, cleavages in regional politics (urban vs rural, or state vs federal po-

sitions), religious cleavages, cleavages relating to gun control, the long-standing issue of communism in the US, cleavages related to liberal "life-styles" [1] (*e.g.*, homosexuality, feminism), welfare, military, patriotism, globalization, conspirationism and mistrust (including mistrust of the establishment and experts), and cleavages around attitudes towards elites. Table 1 summarizes proposed issues of partition according to the surveyed literature, with the keywords identified for the classification of users in binary partitions. User classification relies on a keyword-based strategy of minimalist choice: keywords do not intend to portrait the diverse forms of expression of position on each issue, but rather to identify groups of users with low ambiguity of position. In addition to keywords, we rely on sentiment analysis of profile text to distinguish positive and negative mentions of keywords (using a pre-trained BERT base model for uncased words [11]), assigning to each profile text a sentiment from 1 (very negative) to 5 (very positive). We label text profiles as negative [−] if sentiment is equal to 1, and as positive [+] if sentiment is equal to 5. In Table 1 we also distinguish users whose profiles are not negative [$\not\diagup$]. This is needed, for example, to identify users that might use the word "republican" in their profiles, but in order to utter critique (*e.g.*, "I hate republicans!").

4 Homophily Network Embedding

To identify dimensions that might be revealing of cleavages, we first produce a multi-dimensional space embedding in which these dimensions might emerge as directions. For this, we take the bipartite social subgraph of the 550 members of congress and their 1 821 272 followers to produce an homophily embedding using Correspondence Analysis on the adjacency matrix to compute values ϕ of (1) following the procedure in [28, Section IV]. In this multi-dimensional space, dimensions δ_i ($i = 1, 2, ...$) are ranked according to the information they contain about choices represented in the bipartite social graph, as measured by the inertia. Fig. 1 shows the inertia of each dimension and their relative gain, showing that at most the three first dimensions are relatively more informative than the rest. Fig. 1 also shows the embedding positions of both, congressional members and followers, and the marginal density on these first three dimensions, estimated for the purposes of visualization with kernel density estimation. We can compute party positions as the mean position of congressional members. As expected, the first –most explicative– dimension, δ_1, stands as a good candidate of dimension for attitudes towards parties (or related cleavages, such as liberal-conservative). However, the question remains whether isometric transformations can improve the distinction made by any dimension between Democrat- and Republican-leaning followers. In particular, we do not know if a rotation might improve the ability of a classifier to distinguish between Democrat and Republicans on the first dimension. We know that δ_1 stands for a latent cleavage, and we know that it is highly aligned with Democrat-Republican cleavages, but we do not know if it is the best spatial direction for distinguishing these two groups. More broadly,

Partition	Values	Criteria: keywords and sentiment in profile text	Users
Party	Republican	"republican(s)" AND [✗]	7 895
	Democrat	"democrat(s)" AND [✗]	14 481
Candidate	Trump sup.	("Trump" OR "MAGA") AND [+]	50 159
	Biden sup.	("Biden" OR "Build Back Better") AND [+]	10 229
Black Lives Matter	Pro-BLM	("BLM" OR "black lives matter"] AND [+]	13 350
	Anti-BLM	("BLM" OR "black lives matter") AND [−]	12 017
Black/BlueLivesMatter	BlackLivesMatter	("black lives matter" OR "blm") AND [✗]	16 690
	BlueLivesMatter	"blue lives matter AND [✗]	240
Rural/Urban	Rural	" rural" AND [✗]	567
	Urban	" urban" AND [✗]	1 724
State/Federal	State	"state" AND [✗]	18 369
	Federal	"federal" AND [✗]	1 233
Religion	Christian	"christian" AND [✗]	22 735
	Atheist	"atheist" AND [✗]	1081
Guns	Pro-Gun	"pro gun" OR "gunner" OR ("guns" AND [✗])	605
	Anti-Gun	"anti gun" OR ("gun laws" OR "gun control" AND [✗])	496
Communism	Anti-Communism	("communist" OR "communism") AND [−]	1 160
	Pro-Communism	("communist" OR "communism") AND [+]	448
Liberal LifeStyle	Pro-Lib. LifeStyle	("gay" OR "feminist" OR "lgbt" OR "feminism") AND [+]	9 659
	Anti-Lib. LifeStyle	("gay" OR "feminist" OR "lgbt" OR "feminism") AND [−]	9 921
Welfare	Libertarian	"libertarian" AND [✗]	3 045
	Welfare	"welfare" AND [✗]	681
Police	Pro-Police	"police" AND [+]	1 811
	Anti-Police	("police" AND [−]) OR "defund the police" or "fuck the police"	1 686
Military	Pro-Military	("army" OR "navy" OR "air force" OR "military") AND [+]	17 721
	Anti-Military	("army" OR "navy" OR "air force" OR "military") AND [−]	8 187
Patriotism	Patriot	"patriot"	27 907
	Anti-imperialist	"imperialism"	120
Local/Global focus	Local focus	"local" AND [✗]	5 532
	Global focus	"global" AND [✗]	6 283
DeepState/Rationalist	DeepState	"deep state"	486
	Rationalist	"rationalist"	111
Anti-elitism	Elite	"ceo" OR "founder" OR "representative" OR "expert" OR "phd" OR "professor"	47 032
	Anti-Elite	"people" OR "elite" OR "bankers" OR "politicians" OR "millionaires" OR "experts"	38 086

Table 1. Summary of the proposed issue partitions of users into minimal groups for mining spatial cleavage dimensions capable of classifying them. For each issue we identify two disjoint groups defined by queries of the Twitter profile text descriptions, including keywords (case insensitive, here all written in lowercase), and sentiment: positive [+], negative [−], and non-negative [✗].

it is not trivial to attribute an inductive meaning to what δ_2 and δ_3 might stand for, or to any other space direction for that matter.
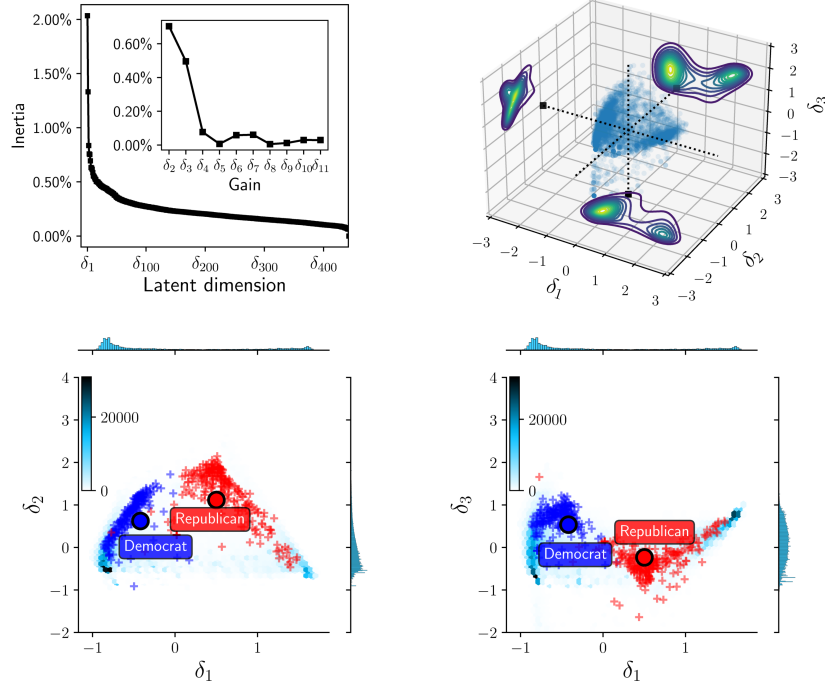

Fig. 1. Multi-dimensional homophily embedding of the collected Twitter network. Dimensions ranked by inertia, and relative gain of including each dimension (top left). Scatter plot and estimated density of the conditional probabilities for the position of users in first three dimensions (top right). Density of followers and positions of members of congress colored by party, and party positions as mean of groups shown in first three dimensions (bottom).

5   Mining Cleavage Dimensions

To extract cleavage dimensions as directions in this latent ideological space, we leverage our proposed set of binary labels from Section 3. We first illustrate this on party cleavages. Among users our users, 7 895 use the word "republican" and 14 481 the word "democrat" in their profile without negative sentiment. To measure the degree to which δ_1, δ_2 or δ_3 might be good candidate directions for distinguishing these two groups, we fit a logistic regression model on each dimension. With this model, we can compute users that might be rightly

or wrongly classified with it, and thus compute a precision, recall and F1-score metric. Fig. 2 shows these values and the distribution of these two groups along δ_1, δ_2 and δ_3, showing that δ_1 indeed is the only dimension among the three to produce a meaningful distinction. Alternatively, instead of using a given dimension, we can fit a multivariate logistic regression model, and identify the direction perpendicular to the decision hyper-plane boundary. In the case of our three-dimensional model, the decision boundary will be a plane and the direction a three-dimensional vector (see in Fig. 2). This discovered direction separating these two groups of users is well aligned with δ_1, but it does not produce an improvement in the F1-score. Still, ideological scaling cannot rely on the *a priori* assumption that this will always be the case as it is standard practice in many works in several disciplines, especially in light of research suggesting a decline in left-right cleavages structuring social choice [13, 28, 32].
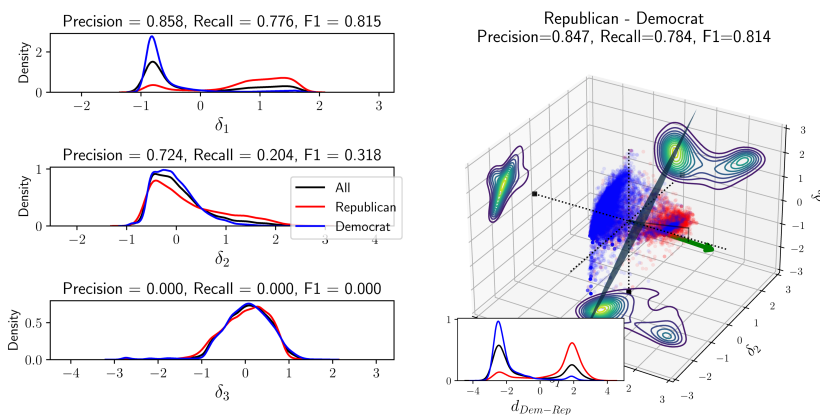

Fig. 2. Distribution Republican- or Democrat-leaning according to their Twitter text profile description, their distribution along the first 3 latent space dimensions, and the accuracy of logistic regression models fitted on each dimension (left). Conditional distributions and positions of labeled users in three-dimensions, and distributions along the direction perpendicular to the boundary of a multivariate logistic regression (right).

This method can be further used to test all proposed binary groups of Table 1, hypothesized to represent potentially relevant cleavages. We fit a multivariate logistic regression model for each pair of groups, and measure the classification accuracy of the model, reported in Table 2, highlighting the cases with F1-score accuracy equal or greater to 0.6. As some pairs are highly imbalanced (*e.g.*, for religious cleavages), we systematically sub-sample the majority group with a Near-Miss strategy [17]. Fig. 3 shows a selection of groups of pairs of labeled users, with the decision boundary and discovered orthogonal direction of the fitted multivariate decision model. This selection highlights the different qualities in the accuracy of the multivariate logistic regression classifier, corresponding to

different strengths of cleavages for the pairs in each labeled group, under the assumption that the chosen criteria identify a relevant group of users.

Label 1	Label 2	Dimension	Precision	Recall	F1-score
Democrat	Republican	$d_{Dem-Rep}$	0.814	0.823	**0.819**
Trump sup.	Biden sup.	$d_{Trump-Biden}$	0.944	0.874	**0.908**
Pro-BLM	Anti-BLM	d_{BLM}	0.488	0.205	0.289
BlackLivesMatter	BlueLivesMatter	d_{Police}	0.687	0.925	**0.789**
Urban	Rural	d_{Urban}	0.529	0.577	0.552
State	Federal	d_{St-Fed}	0.626	0.536	0.578
Christian	Atheist	$d_{Christian}$	0.706	0.559	0.624
Pro-Gun	Anti-Gun	$d_{Pro-Gun}$	0.713	0.544	**0.617**
Pro-Communism	Anti-Communism	$d_{Comm.}$	0.647	0.650	**0.648**
Anti-Liberal LifeStyle	Pro-Liberal LifeStyle	$d_{LifeStyle}$	0.537	0.867	**0.664**
Libertarian	Welfare	$d_{Welfare}$	0.695	0.720	**0.707**
Pro-Police	Anti-Police	d_{Police}	0.617	0.624	**0.620**
Pro-Military	Anti-Military	$d_{Military}$	0.599	0.474	0.529
Patriot	Anti-imperialist	$d_{Patriot}$	0.709	0.692	**0.700**
Global focus	Local focus	d_{Local}	0.547	0.676	**0.605**
DeepState	Rationalist	$d_{DeepState}$	0.569	0.667	**0.614**
Elite	Anti-Elite	$d_{AntiElite}$	0.549	0.439	0.488

Table 2. Groups of pairs of labeled users (according to criteria of Table 1), naming of the mined dimension perpendicular to the decision boundary of a multivariate logistic regression classification model, and the accuracy of the fitted model.

The identification of different plausible dimensions of social cleavage presents us with several interesting possibilities. First, the coordinate of each user along these new identified directions can be used to disentangle issue alignment: *e.g.*, we can produce separate and explicit dimensions for party cleavages and attitudes towards gun control. In contrast, δ_1 (usually reported in works using ideological scaling) is a proxy for party cleavages, but also for other positions on correlated issues (*e.g.*, racial or religious issues, see Fig. 3) in non-explicit ways. Similarly, by inspecting the alignment between different cleavage dimensions we can identify and quantify this issue polarization. Fig. 4 shows the mined cleavage dimensions (*i.e.*, with F1-score ≥ 0.6) and their pairwise angular distance, with metrical clusters computed with Un-weighted Pair Group Method with Arithmetic (UPGMA) mean [31]. The mined cleavages can be organized in four *ideologies* (in the sense of issue alignment): 1) a dominant ideology comprising party, candidate, and other stances correlated with δ_1, 2) an ideology separating people defining themselves using the words "local" and "global", 3) an ideology separating those defining themselves using the words "welfare" and "libertarian", and 4) an ideology separating those with positive and negative mentions of issues relating to sexual diversity and feminism, and the use of the word "communism". Four directions cannot be perfectly orthogonal in three-dimensions, but any two cleavages belonging to two different identified ideological groups will display enough angular distance, so as to not be considered as highly correlated. Being able to disentangle issues in separate dimensions, allows us to conduct different investigations against the map positions of actors in now identifiable axes. Because we can also measure the position of reference users (politicians) in
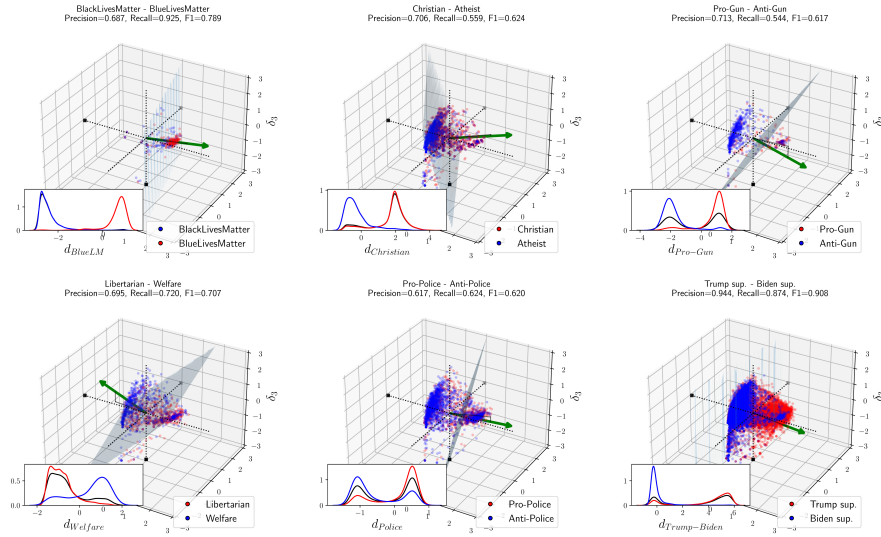
Fig. 3. Selection of groups of pairs of labeled users, with the decision boundary and orthogonal direction of the fitted multivariate decision model.

mined cleavage dimensions, we can investigate intra-party diversity on separate issues: *e.g.*, of support for their presidential candidate, or attitudes towards welfare, religious diversity, or diversity of views on racial issues. Fig. 5 shows, for example, that Republicans are more heterogeneous in their support for Donald Trump than the Democrats in their support for Joseph Biden, both the members of congress (in crosses in Fig. 5) and the followers (density shown in light blue in Fig. 5).

6   Discussion and conclusions

This article argued for the importance of re-examining the assumptions implicitly leveraged by studies that use ideological scaling, showing that the dominant latent dimensions cannot be assumed to be the direction that is most aligned with traditional Democrat-Republican cleavages in the US. The degree of alignment between latent dimension and different cleavages can and must be determined and measured. This article further presented a way of mining dimensions of social cleavage with explicit meaning using both network embedding and NLP methods. Furthermore, using this combination of methods, this article analyzed the case of a political Twitter friend network in the US, identifying the main dominant cleavage, but also additional ones hypothesized as relevant by recent studies in social sciences [32]. In particular, four ideologies, or bundled groups of cleavage dimensions were identified. These groups of dimensions are not highly aligned among them, and represent new cleavage dimensions that can be used
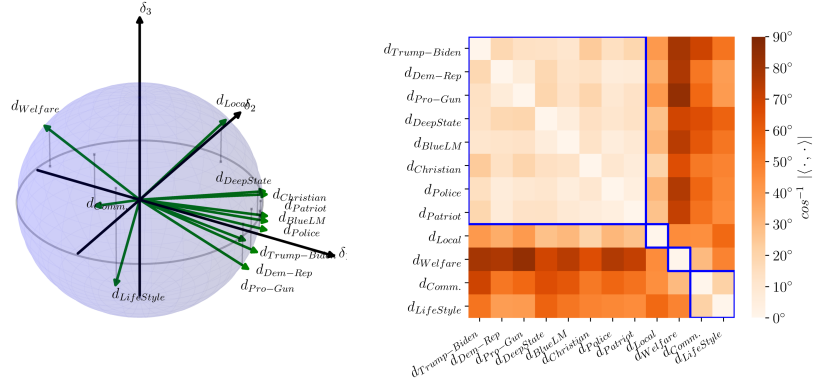
Fig. 4. Mined cleavage dimensions (left) can be organized in four *ideologies* (in the sense of issue polarization), shown in four groups in blue in the angular distance matrix (right).

in further studies. This method also offers the possibility of developing new applications for explicitly measuring issue polarization as the alignment of bundled social cleavages, as well as a method for projecting large numbers of users onto space dimensions with explicit meaning in terms of the issues to which it measures positive and negative views. This new possibility, it was argued, opens interesting paths for research, which was illustrated with a brief example. By measuring positions of Democrat and Republican congressional members on both, party cleavage and candidate cleavage dimensions with data collected in the days leading to the 2020 US presidential election, this article showed that, when compared with Democrats, it may be proved that Republicans display higher heterogeneity in their support for their candidate. Beyond this example, many others could leverage these results and methods. In particular, having multidimensional distributions of political attitudes could be leveraged in the study of social mobilization by analyzing the determinants and characteristics of political movements (see for example [10, 25, 26]). Additionally, by leveraging information consumption practices and media diets, attitudinal positions could be attributed to news media articles and outlets, allowing for the study of diversity, or lack thereof, in information consumption patterns [30, 20]. This, in turn, presents interesting possibilities for large-scale analysis of wide news and informational ecosystems [9].
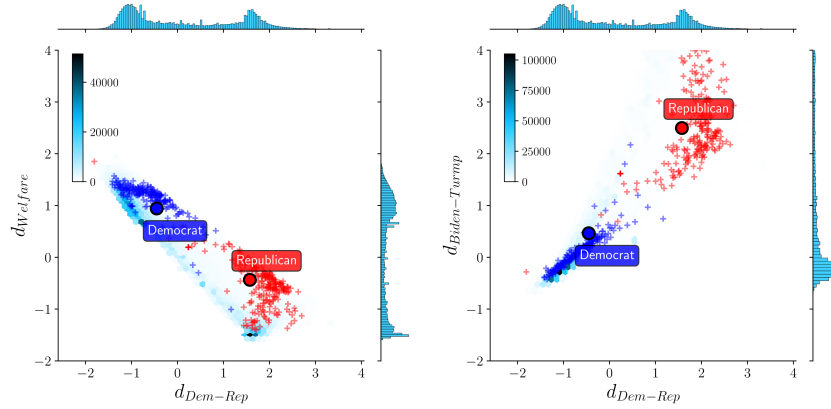
Acknowledgments

Fig. 5. Density of Twitter users and positions of members of congress along mined dimensions. The distribution of members of congress shows the intra-party diversity of stances towards presidential candidates and welfare.

References

1. Bakker, R., Hooghe, L., Jolly, S., Marks, G., Polk, J., Rovny, J., Steenbergen, M., Vachudova, M.A.: 2019 chapel hill expert survey. Chapel Hill (2020), www.chesdata.eu
2. Bakker, R., Jolly, S., Polk, J.: Complexity in the european party space: Exploring dimensionality with experts. European Union Politics 13(2), 219–245 (2012)
3. Bakshy, E., Messing, S., Adamic, L.A.: Exposure to ideologically diverse news and opinion on facebook. Science 348(6239), 1130–1132 (2015)
4. Barberá, P.: Birds of the same feather tweet together: Bayesian ideal point estimation using twitter data. Political analysis (2015)
5. Barberá, P., Jost, J.T., Nagler, J., Tucker, J.A., Bonneau, R.: Tweeting from left to right: Is online political communication more than an echo chamber? Psychological science 26(10), 1531–1542 (2015)
6. Baumann, F., Lorenz-Spreen, P., Sokolov, I.M., Starnini, M.: Modeling echo chambers and polarization dynamics in social networks. Physical Review Letters 124(4), 048301 (2020)
7. Benkler, Y., Faris, R., Roberts, H.: Network propaganda: Manipulation, disinformation, and radicalization in American politics. Oxford University Press (2018)

8. Bond, R., Messing, S.: Quantifying social media's political space: Estimating ideology from publicly revealed preferences on facebook. American Political Science Review 19 (2015)
9. Cointet, J.P., Cardon, D., Mogoutov, A., Ooghe-Tabanou, B., Plique, G., Ramaciotti Morales, P.: Uncovering the structure of the french media ecosystem. arXiv preprint arXiv:2107.12073 (2021)
10. Cointet, J.P., Ramaciotti Morales, P., Cardon, D., Froio, C., Mogoutov, A., Ooghe-Tabanou, B., Plique, G.: What colours are the yellow vests? an ideological scaling of facebook groups. Statistique et Société (2021)
11. Devlin, J., Chang, M.W., Lee, K., Toutanova, K.: Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805 (2018)
12. Greenacre, M.: Correspondence analysis in practice (2017)
13. Grossman, E., Sauger, N.: Economic internationalization and the decline of the left–right dimension. Party Politics (2019)
14. Imai, K., Lo, J., Olmsted, J.: Fast estimation of ideal points with massive data. American Political Science Review 110(4), 631–656 (2016)
15. Lipset, S.M., Rokkan, S., et al.: Cleavage structures, party systems, and voter alignments: an introduction (1967)
16. Lowe, W.: Understanding wordscores. Political Analysis (2008)
17. Mani, I., Zhang, I.: knn approach to unbalanced data distributions: a case study involving information extraction. In: Proceedings of workshop on learning from imbalanced datasets. vol. 126, pp. 1–7. ICML (2003)
18. Mason, L.: "i disrespectfully agree": The differential effects of partisan sorting on social and issue polarization. American journal of political science 59(1), 128–145 (2015)
19. Morales, P.R., Cointet, J.P., Laborde, J.: Your most telling friends: Propagating latent ideological features on twitter using neighborhood coherence. In: 2020 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM). pp. 217–221. IEEE (2020)
20. Morales, P.R., Lamarche-Perrin, R., Fournier-S'Niehotta, R., Poulain, R., Tabourier, L., Tarissan, F.: Measuring diversity in heterogeneous information networks. Theoretical Computer Science 859, 80–115 (2021)
21. O'Callaghan, D., Greene, D., Conway, M., Carthy, J., Cunningham, P.: Down the (white) rabbit hole: The extreme right and online recommender systems. Social Science Computer Review 33(4), 459–478 (2015)
22. Poole, K.T., Rosenthal, H.: A spatial model for legislative roll call analysis. American journal of political science pp. 357–384 (1985)
23. Poole, K.T., Rosenthal, H.: Congress: A political-economic history of roll call voting. Oxford University Press on Demand (2000)
24. Ramaciotti Morales, P., Cointet, J.P.: Auditing the effect of social network recommendations on polarization in geometrical ideological spaces. In: 15th ACM Conference on Recommender Systems, RecSys' 21 (2021)
25. Ramaciotti Morales, P., Cointet, J.P., Benbouzid, B., Cardon, D., Froio, C., Metin, O.F., Ooghe, B., Plique, G.: Atlas multi-plateformes d'un mouvement social: Le cas des gilets jaunes. Statistique et Société (2021)
26. Ramaciotti Morales, P., Cointet, J.P., Froio, C.: Posters and protesters. Journal of Computational Social Science pp. 1–29 (2022)
27. Ramaciotti Morales, P., Cointet, J.P., Muñoz Zolotoochin, G., Fernández Peralta, A., Iñiguez, G., Pournaki, A.: Inferring attitudinal spaces in social networks (2022)

28. Ramaciotti Morales, P., Cointet, J.P., Muñoz Zolotoochin, G.: Unfolding the dimensionality structure of social networks in ideological embeddings. In: Proceedings of the 2021 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining. pp. 333–338 (2021)
29. Ramaciotti Morales, P., Muñoz Zolotoochin, G.: Measuring the accuracy of social network ideological embeddings using language models. In: 2022 International Conference on Information Technology & Systems (ICITS22) (2022)
30. Ramaciotti Morales, P., Tabourier, L., Ung, S., Prieur, C.: Role of the website structure in the diversity of browsing behaviors. In: Proceedings of the 30th ACM Conference on Hypertext and Social Media. pp. 133–142 (2019)
31. Sokal, R.R.: A statistical method for evaluating systematic relationships. Univ. Kansas, Sci. Bull. 38, 1409–1438 (1958)
32. Uscinski, J.E., Enders, A.M., Seelig, M.I., Klofstad, C.A., Funchion, J.R., Everett, C., Wuchty, S., Premaratne, K., Murthi, M.N.: American politics in two dimensions: Partisan and ideological identities versus anti-establishment orientations. American Journal of Political Science 65(4), 877–895 (2021)
33. Xiao, Z., Song, W., Xu, H., Ren, Z., Sun, Y.: Timme: Twitter ideology-detection via multi-task multi-relational embedding. In: Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. pp. 2258–2268 (2020)