2022 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)

Embedding social graphs from multiple national settings in common empirical opinion spaces

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

Zografoula Vagena
zografoula.vagena@u-paris.fr
médialab, Sciences Po, LPI & LIPADE, Université Paris Cité
Paris, France

***Abstract*—Ideological scaling is an ubiquitous tool for inferring political opinions of users in social networks, allowing to position a large number of users in *left-right* or *liberal-conservative* scales. More recent methods address the need, highlighted by social science research, to infer positions in additional social dimensions. These dimensions allow for the analysis of emerging divisions such as anti-elite sentiment, or attitudes towards globalization, among others. These methods propose to embed social networks in multi-dimensional attitudinal spaces, where dimensions stand as indicators of positive or negative attitudes towards several and separate issues of public debate. So far, these methods have been validated in the context of individual national settings. In this article we propose a method to embed a large number of social media users in multi-dimensional attitudinal spaces that are common to several countries, allowing for large-scale comparative studies. Additionally, we propose novel statistical benchmark validations that show the accuracy of the estimated positions. We illustrate our method on Twitter friendship networks in France, Germany, Italy, and Spain.**

***Index Terms*—Graph embedding, latent space, opinion mining, ideological scaling, political attitude data, polarization.**

I. Introduction

Ideological scaling of social choice data [1] allows to infer ideological positions of individuals on opinion scales. In contrast with the categorization of individuals in political groups (*e.g.*, in Democrat- or Republican-leaning groups [2]), ideological scaling estimates positions in continuous scales; traditionally, *left-right* or *liberal-conservative* scales. As the study of hypothesized dysfunctions of large online social platforms (*e.g.*, echo chambers, filter bubbles) gains in importance, ideological scaling gains in popularity because these studies hinge on the characterization of individual opinions. Ideological scaling has been used to estimate positions of large numbers of users in liberal-conservative scales using *likes* on Facebook [3] and friendship (or *follower*) networks on Twitter [4]. The validity of these empirical opinion scales is, however, limited to the nodes present in the data used for inference, with no means for direct extrapolation or comparison. If a given Twitter follower network (a sub-graph of the Twitter graph) is used to infer a latent liberal-conservative scale on which to position users (*i.e.*, nodes), this scale is only valid for the nodes of the network with which it was computed. We cannot directly compare the position of existing nodes with that of new nodes coming to the network, or with nodes of a different network. For example, if we used ideological scaling to compute left-right positions in two countries (*e.g.*, Twitter networks in the US [5] and in Europe [6]), we cannot compare the positions of users in different countries: we cannot say, *e.g.*, whether a given user in one country is to the left or the right of a given user in the other country. This represents an important limitation in comparative studies of across countries that require opinion characterization, such as the studies of algorithmic biases and polarization [7] or news media consumption [8], to name a few.

IEEE/ACM ASONAM 2022, November 10-13, 2022
978-1-6654-5661-6/22/$31.00 © 2022 IEEE

A second limitation arises in the analysis of multi-dimensional empirical opinion spaces, capturing attitudes towards other social lines of division beyond left-right or liberal-conservative ones. Recent works have suggested the importance of including additional dimensions of analysis in the study of socio-political dysfunctions of large online social networks. Emerging social divides linked to anti-elite sentiments and attitudes towards globalization [9] have been shown to be related to phenomena such as the spread of misinformation [10], social protest [11], and lack of trust in institutions [12]. Accordingly, in recent works we have identified these emerging divisions in multi-dimensional ideological scaling in online social graphs [13]. More recent works have proposed a method for embedding social graphs in multi-dimensional spaces of political attitudes. In these attitudinal spaces, dimensions stand for indicators of individual positive or negative attitudes towards separate and identifiable issues of public political debate (including in particular anti-elite sentiments and globalization). This method, called *attitudinal embedding* [14], relies on the position of referential users in both 1) latent spaces in social networks (computed using multi-dimensional ideological scaling) and 2) political surveys, to embed large number of users in spaces with explainable issue dimensions.

In this article, we rely on referential attitudinal spaces provided by such political surveys, but conducted in several countries, and on political Twitter follower networks in the four most populous EU countries (France, Germany, Italy, and Spain) to embed users from these networks in common referential attitudinal spaces. We call these spaces *attitudinal reference frame* because dimensions are specified in a given poll or survey, associated with pre-defined issues (*e.g.*, immigration, taxation, left-right positions) onto which positions of users (or other entities, such as parties) are mapped. To achieve the proposed embedding in a common space, we tackle two main problems, the identification of both, 1) the dimensions of latent space of Twitter networks of each national setting that can be used to embed users, and 2) the dimensions of referential attitudinal frames given by surveys for each countries, onto which we can embed large numbers of users. Additionally, 3) we develop a method to test and benchmark the quality of the embedding of users from these countries into a common multi-dimensional attitudinal space. To test our method and the measurement of the accuracy of our embedding we choose two attitudinal dimensions of interest in measuring socio-political dysfunction: Left-Right positions and Anti-elite sentiment (or salience) of individual users.

II. Related Work

Many opinion mining methods rely on text data, making them language-dependent and limiting their use in comparative studies of social networks across countries. Ideological scaling, because it relies on relational behavioral traces (*e.g.*, who *follows* or *likes* whom or what), is a good candidate for estimating opinion positions in scales comparable across different national settings. Ideological scaling methods (see [1] for a comprehensive survey) can be traced back at least to the NOMINATE method by Poole and Rosenthal [15]. In its most widespread form, ideological scaling assumes that observed social choice data follows an homophilic [16] generative process [4]:

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

where the probability of observing user j interacting user i ($i \leftarrow j$) depends on bias α and scale β parameters, and on the distance $\|\vec{\vartheta}_i - \vec{\vartheta}_j\|$ between users i and j in some latent space in which the position of users might be explaining the formation of the observed social network $\mathcal{G} = (V, E)$, with $E \subseteq \{i \leftarrow j : i, j \in V, i \neq j\}$. The stochastic process underlying (1) is homophilic in the sense that, the closer users are in latent space (*i.e.*, the more *similar* they are, displaying *value homophily* in the context of political opinions [17]), the higher the probability of observing an interaction between them. Ideological scaling uses observed empirical data $\mathcal{G}$ and Bayesian inference to estimate positions $\vec{\vartheta}_i \in \mathbb{R}^N$, and it is called multi-dimensional ideological scaling if $N \geq 2$. For example, Barbera [4] uses (1) to estimate a single latent dimension (*i.e.*, $N = 1$) using a bipartite social sub-graph of the Twitter graph made of members of the US Congress and their followers, which is inductively demonstrated to stand for liberal-conservative positions. Similar single-dimensional methods have been used in several national settings, including, *e.g.*, bipartite Twitter networks in Spain [18], France [6], and Chile [19].

Motivated by numerous recent results highlighting the importance of additional emerging lines of political divisions [8], [20] and the declining importance of the left-right dimensions [9], recent works have taken a renewed interest in multi-dimensional ideological scaling (*i.e.*, $N \geq 2$). Using both, 1) latent ideological spaces computed with multi-dimensional ideological scaling, and 2) positions of referential users along tens of predefined issue dimensions available in political surveys, recent works have shown that, indeed, positions of individuals towards elites participate in their social choice in online networks [13], [21]. Subsequent methods, such as the cited *attitudinal embedding* (AE) method, have proposed to mine these relations to further embed ideological positions $\vec{\vartheta} \in \mathbb{R}^N$ onto the multi-dimensional space formed by M issue dimensions (*i.e.*, an M-dimensional space or *attitudinal reference frame*) of a given political survey [14].

In this article, we tackle several challenges in producing political multi-dimensional positions for users in several countries in a way that is comparable. In particular, we exploit political surveys available for several countries and we address the problem of the choice of the ideological dimensions N and the choice of attitudinal dimensions M that can be extracted for chosen countries. Additionally, we measure the quality of

TABLE I
TWITTER GRAPH DATA COLLECTED FOR EXPERIMENTS.

Country	References	Parties (survey)	Followers	Pol. followers
France	827	10 (8)	5 097 543	325 672
Germany	560	9 (7)	2 880 687	172 137
Italy	791	16 (7)	5 639 305	377 067
Spain	219	12 (9)	5 533 868	674 793

the inferred position of users in common multi-dimensional attitudinal spaces relying on text utterances produced by them, and that are linked to the political divisions that we intend to measure.

III. DATA

A. Political social graph data

To illustrate our method, we follow recent results in multi-dimensional ideological scaling in the US [4] and in Europe [13]. We collect bipartite sub-graphs of the Twitter graph in national settings, formed by Twitter followers of members of parliament (MPs). We consider national settings covered by a single multi-dimensional political survey: the Chapel Hill Expert Survey (CHES) [22], covering all EU countries on 51 attitudinal issue dimensions. We select the four most populous EU countries (France, Germany, Italy, and Spain), manually annotate the Twitter accounts of MPs, and collect their followers[1]. We filter some of the collected followers to keep those that have posted at least 100 tweets and that are followed by at least 25 other followers to minimize the ratio of bots in our sample. We also keep only followers that follow at least 3 MPs. This serves two purposes: 1) to filter out users that may follow an MP for reasons other than similarity or alignment in political views or stances, and 2) to maximize the degree of *political sophistication* [23] of our sample, *i.e.*, their knowledge in policy issues and thus the degree to which their political choices are meaningfully modeled by homophilic spatial models. Accordingly, we call the resulting samples *politicized followers*. Table I summarizes the countries on which we identify Twitter accounts of MPs, the number of parties to which they belong, the number of parties that are also included in the CHES, their followers per country, and the number of politicized followers (or simply *followers*, hereinafter).

B. Profile text data

For each follower in our dataset we collected the Twitter bio profile text written by them. Not all every user writes a profile. This resulted in the following number of text profiles: 255 794 (78.5% of politicized followers) in France, 142 880 (83%) in Germany, 265 227 (70,3%) in Italy, and 484 147 (71,7%) in Spain. With these texts we will propose validation benchmarks for the estimation of the accuracy of the estimated multi-dimensional political positions in the following sections. For this validation we will also rely on the estimated sentiment of each text description, which we compute using a pre-trained multi-lingual BERT base model [24][2].

C. Referential attitudinal survey data

To create a multi-dimensional ARF that is common to our selected countries, we rely on the 2019 Chapel Hill Expert Data (CHES) [22] (the 2019 wave being the closest in time to our time of collection[3]). The reference points available in the CHES dataset are political parties, which we can produce in the latent ideological space by taking the mean position of MPs for each party. The CHES data contain party positions in attitudinal scales ranging from *0* (*most opposed*) to *10*[4] (*most favorable*) and associated with selected explicit issues of public debate: *e.g.*, special rights for minorities, anti-elite sentiments, left or right economic policy, among others. The 2019 CHES data were compiled with the responses of 421 experts in European politics, in which they place European parties for 51 different issues. Some of these dimensions are of special interest for studying the aforementioned traditional cleavages and new emerging lines of division. In particular, left-right cleavages, and anti-elite sentiment (called *anti-elite salience* in the CHES data) [13]. Among the parties of the manually annotated MPs, 8 parties in France, and 7 in Germany and Italy, and 9 in Spain are present both in the our Twitter dataset and in the CHES data.

IV. MULTI-DIMENSIONAL POLITICAL OPINIONS

To produce and embedding of users from Table I in a common multi-dimensional ARF using the CHES data, we rely on the AE method [14] consisting of two stages (see Fig. 1). First, each social graph (*i.e.*, each bipartite sub-graph of the MPs and their followers for each country) is embedded in a latent ideological space using the ideological scaling (IS) method underlying (1). We compute a multi-dimensional IS on each national bipartite graph by computing a spatialization with Correspondence Analysis [25], which has been proved –theoretically [26], [27], and empirically [5]– to approximate

[1]Collection was carried out in October 2020; please refer to the Acknowledgments section to see our GDPR deposit and for a link to the legal notice.

[2]https://huggingface.co/nlptown/bert-base-multilingual-uncased-sentiment.
[3]https://www.chesdata.eu/.
[4]Except for attitudes towards European integration, ranging from 1 to 7.

values of the multi-dimensional parameters $\vec{\vartheta}$ in (1). The IS yields a spatialization in dimensions $\delta_1, \delta_2, \ldots, \delta_N$ of $\mathbb{R}^N$ for each reference user (MPs) and their followers. Second, using reference points present in both 1) the latent ideological space and 2) in some ARF $\mathbb{R}^M$ with explicit issue dimensions given, the method maps ideological space to ARF. We use as reference points the positions of the P political parties that, for each country, are available in both spaces. Party positions are given explicitly in the 51 dimensions of the CHES data, and are in the latent ideological space as the mean positions of the MPs grouped by party. AE hinges on the choice of the number of dimensions N of departure and M of arrival space to compute an affine transformation $T_{\text{aff}} : \mathbb{R}^N \to \mathbb{R}^M$ minimizing error in the position of the P parties in the arrival space. If $Y \in \mathbb{R}^{M\times P}$ is the position of parties in the CHES ARF, and $X \in \mathbb{R}^{N\times P}$ is position of parties in the latent ideological space, the optimal affine transformation T^*_{aff} minimizes the error between Y with $\hat{Y} = T^*_{\text{aff}}(X)$ being estimated party positions. The transformation equation can be recast as an augmented matrix problem (in *homogeneous coordinates*), for $\widetilde{T}_{\text{aff}} : \mathbb{R}^{N+1} \to \mathbb{R}^{M+1}$ (see [14]):

$$\underbrace{\begin{pmatrix} Y \\ \hline 1 \end{pmatrix}}_{\widetilde{Y}\in\mathbb{R}^{(M+1)\times P}} = \underbrace{\left[\begin{array}{c|c} A & B \\ \hline 0\cdots 0 & 1 \end{array}\right]}_{\widetilde{T}_{\text{aff}}\in\mathbb{R}^{(M+1)\times(N+1)}} \cdot \underbrace{\begin{pmatrix} X \\ \hline 1 \end{pmatrix}}_{\widetilde{X}\in\mathbb{R}^{(N+1)\times P}}, \qquad (2)$$

where $A \in \mathbb{R}^{M\times N}$ and $B \in \mathbb{R}^M$. If the error metric is chosen to be the Frobenius norm of $Y - \hat{Y}$ in the arrival space, *i.e.*,

$$\|Y - \hat{Y}\|_F = \sqrt{\sum_{d=1}^{M}\sum_{p=1}^{P} \left|Y_{dp} - \hat{Y}_{dp}\right|^2}, \qquad (3)$$

the error is minimized by the pseudo-inverse $\widetilde{T}^*_{\text{aff}} = \widetilde{Y}\widetilde{X}^T\left(\widetilde{X}\widetilde{X}^T\right)^{-1}$ [28] (see [29] for further details).

When IS is computed using Correspondence Analysis, we can rank dimensions by inertia (*i.e.*, total variance of the factor scores), ordering dimensions by contributions to explaining the variance in followed MPs. Fig. 2 shows the inertia of the first 100 dimensions computed for each bipartite social graph of each country, showing that at least the first few ideological dimensions have marginal contributions to the explanation of the variance. To estimate the two attitudinal dimensions of interest using the CHES data, *i.e.* Left – Right and Anti-elite salience, we conservatively chose the first two ideological dimensions (δ_1 and δ_2, *i.e.*, $N = 2$).

Next, we analyze the dimensionality of the CHES dataset, and the space spanned by the two attitudinal dimensions of interest Left – Right, Anti-elite salience. Fig 3 shows the position of the political parties available in the four selected countries, along three selected dimensions among the 51 available (for the purpose of illustration): Left – Right, Anti-elite salience, and Anti-immigration stance. To examine the dimensionality of party positions in each country, we computed a PCA and show the explained variance of the first components. No principal component (PC) beyond the 8th contributes to explain spatial variance in the 51 dimensions. Projecting the CHES Left-Right and the Anti-elite salience dimensions onto the first two PCs shows that these are both relevant for the first two PCs and independent; *i.e.*, they are strongly aligned. We select CHES Left-Right $d_{Left-Right}$ and the Anti-elite salience $d_{Anti-elite}$ (*i.e.*, $M = 2$) and compute the attitudinal position of followers for each country on this common CHES ARF. CHES data contains an *economic* and an *ideological* Left-Right attitudinal dimension, we use the former. Fig. 4 (top row) shows the positions of parties (computed as the mean position of MPs grouped by party) and the density of followers

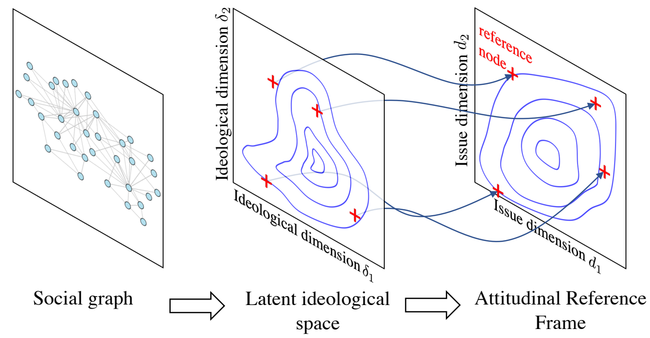

Fig. 1. Diagram of the two phases of attitudinal embedding: a social graph is embedded in a multi-dimensional space with ideological scaling, which is then mapped to an attitudinal reference frame given by an instrument such as political survey, using reference nodes present in both, 1) the social graph and thus the ideological space, and 2) the referential attitudinal frame.

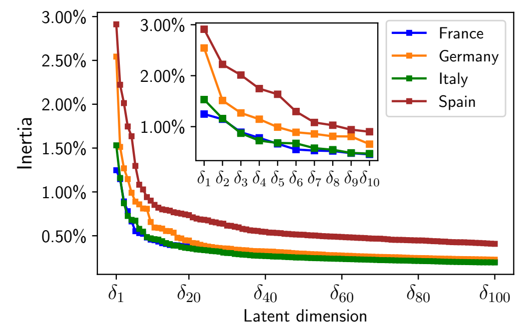

Fig. 2. Inertia of the first 100 latent ideological dimensions of the multi-dimensional ideological scaling computed for each country.

per country, both in the first two ideological dimensions of the latent ideological space computed with multi-dimensional IS, and (bottom row) in the CHES Left–Right and Anti-elite salience attitudinal dimensions of the CHES ARF computed with AE. Fig. 4 (bottom row) also shows the referential extreme party positions as boundary values of the CHES data, showing, unsurprisingly, that some users are more extreme than parties for some dimensions, falling outside the bounding box of the most extreme values considered by the survey.

V. Accuracy of Inferred Positions in ARF

A. Labeling users declaring sides in cleavages

To test the accuracy of the positions of the users embedded in this common ARF, we produce a text-based labeling and we measure the performance of a spatial classifier in distinguishing these labels. To label users we turn now to the text profile descriptions of section III-B. For each one of the two attitudinal dimensions, $d_{Left-Right}$ and $d_{Anti-elite}$, we propose minimalist criteria to label part of users based on their text profile descriptions, into groups that may reveal the cleavage of each dimension.

For $d_{Left-Right}$ we label users as describing themselves as having left- or right-leaning political sympathies. We label a user as being from the Left, if it uses the keyword “left” in its Twitter profile without negative sentiment. Similarly, we label a user as being from the Right, if it uses the keyword “right” in its Twitter profile without negative sentiment. We deem a text profile as being non-negative, if the sentiment score obtained with the multi-language pre-trained BERT model is 3 or higher (on a scale from 1 to 5). Filtering profiles with negative sentiment is needed to minimize the probability of, *e.g.*, labeling a user in the *Right* because of utterances of critique towards the right (*e.g.*, “I hate right-wingers!”). Keywords “left’ and “right” are included in the national languages for each country.

For $d_{Anti-elite}$ we label users that use words identified with anti-elitist and anti-establishment discourse [22]. We label a user as talking about *People & Elites* whenever its profile includes the keywords “people”, “elite” or “politicians” (in the corresponding national languages). “People” and “elite” are included in plural and singular, while “politicians” is included in plural so as to not include users defining themselves as politicians. Anti-elite sentiment does not intend to measure belonging or not to social social groups of given material, educational, cultural, or social wealth, but to a group that subscribes a worldview that opposes two supposedly homogeneous and antagonistic groups, “the elites” and “the people” [30]. Anti-elite sentiment is the degree of subscription to this worldview. Individuals of great wealth and influence, may potentially subscribe to anti-elite views to a high degree (*e.g.*, Donald Trump [31]). This is how the question is presented to CHES respondents [22]: “salience of anti-establishment and anti-elite rhetoric”, with *0* being “not at all” and *10* being “extremely important”. The selection of keywords defining this class captures anti-elite sentiments in that they are used themselves as a critique, and are not used by people with low anti-elite sentiments [13], [14].

The keywords defining these classes are not intended to capture the diversity of forms through which users can express sympathies towards political left- and right-stances, or “anti-establishment and anti-elite rhetoric”. Instead, their minimalist nature is intended to identify a small but sufficient number of users that have low probability of false positive and negative in classifications (*e.g.*, a left-leaning user identified as labeled *Right*). Table II presents a summary of the groups, the labels, the criteria used, and the number of users identified for each country.

B. Measuring accuracy of estimated positions

We now leverage our groups of labeled users to measure the accuracy of the position of users positioned in the common ARF. If their positions along dimensions $d_{Left-Right}$ and $d_{Anti-elite}$ are accurate, they should be good features for distinguishing labeled users. We chose a logistic regression (LR) classifier to show that a simple linear model is able to produce accurate classifications based on these dimensions. For $d_{Left-Right}$ we fit a LR model based on groups *Left* and *Right*. For $d_{Anti-elite}$ we fit a LR model based on groups *People & Elites* and the rest of users (*Other* in Table II). Because the *People & Elites* and *Other* are so imbalanced, we adopt a the Near-Miss sub-sampling strategy [32] to extract a subset of the group *Other* for regression. Fig. 5 illustrates the use of LR as classifiers on our dichotomous pairs of labels for the case of France and $d_{Left-Right}$.

Once a LR model is fitted on each dimension, we can evaluate the fitness of this model by using it as a classifier and computing the precision, recall, and F1-score. We also compute the accuracy of classification for logistic models fitted on the original ideological dimensions (δ_1 and δ_2) for comparison. Table III presents a summary of the accuracy metrics for these LR classifiers. The inferred positions computed with AE, now on a common ARF, have accuracy comparable with that achievable with IS alone in most cases. Our goal is not to show that AE improves accuracy with respect to

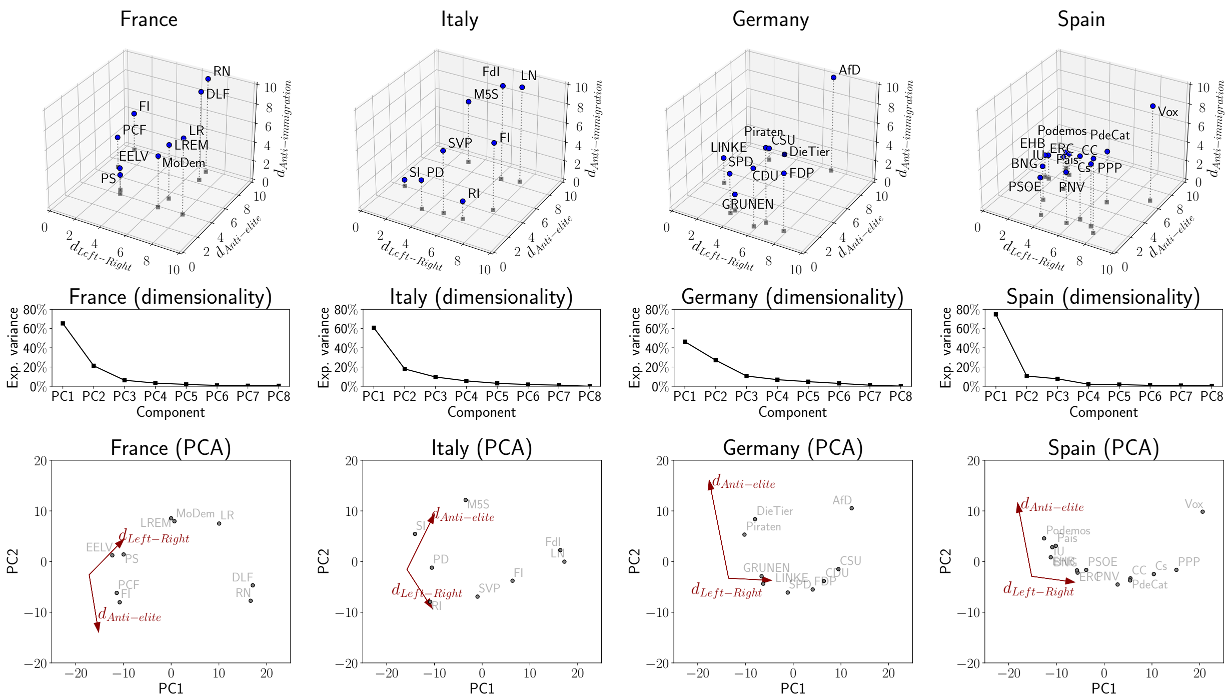

Fig. 3. Positions of parties in three selected attitudinal dimensions: Left–Right, Anti-elite salience, and Anti-immigration (top). Explained variance for the first principal components computed with PCA on all 51 dimensions of the CHES data (center). Left-Right and Anti-elite attitudinal dimensions projected onto the plane of the first two principal components (bottom).

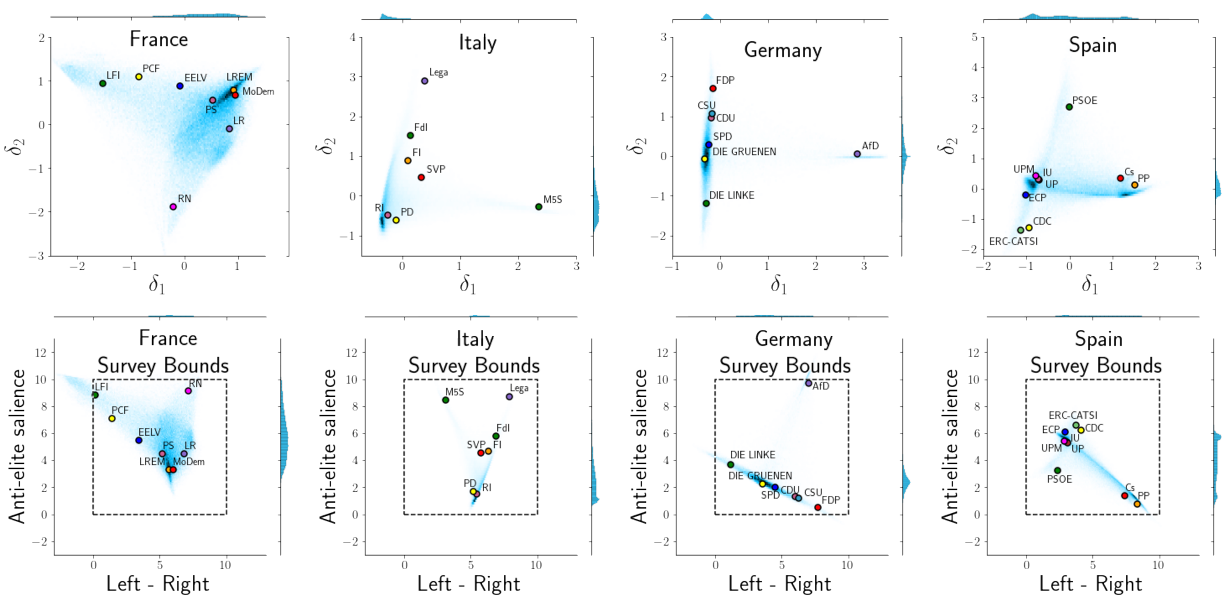

Fig. 4. Party positions and the density of Twitter users in the first 2 dimensions of the latent ideological space (top), and in the Left–Right and Anti-elite salience dimensions of the CHES attitudinal reference frame (bottom).

TABLE II

SUMMARY OF THE GROUPS, THE LABELS, THE CRITERIA USED, AND THE NUMBER OF USERS IDENTIFIED FOR EACH COUNTRY.

Partition name	Label	Criteria	France	Germany	Italy	Spain
Left/Right	*Left*	Profile includes keyword "left" in local language AND does not have negative sentiment.	1 616	943	788	2 472
	Right	Profile includes keyword "right" in local language AND does not have negative sentiment.	1 036	740	353	803
Anti-elitism	*People&Elites*	Profile includes keyword "people" OR "elite(s)" OR "politicians" in local language.	1 780	364	3 575	9 018
	Other	All the rest.	253 957	142 516	261 652	475 129

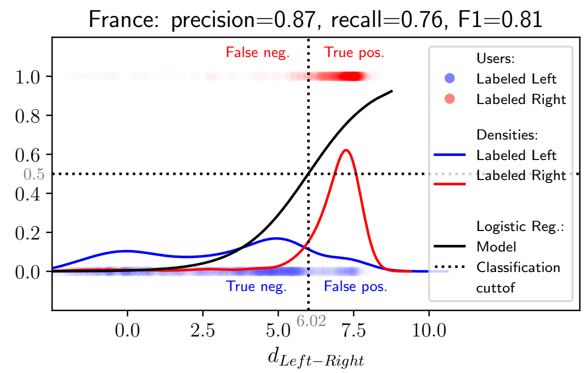

Fig. 5. Illustration of the use of logistic regression as clssifiers for the assessment of the capacity of attitudinal dimensions to distinguish pairs of dichotomous labels. Example for France on $d_{Left-Right}$ and users labeled *Left* and *Right*, achieving an F1-score of 0.81.

IS. AE on $\delta_{Left-Right}$ improves F1-score accuracy in all countries, except for Germany, where it achieves, nonetheless, a value of 0.66. AE on $\delta_{Anti-elite}$ improves the accuracy in Germany, nearly maintains the accuracy in France and Spain, and sees a slight decrease in Italy (in comparison to IS). These results prove that AE may have an accuracy comparable with that of IS, while providing the crucial advantage of yielding graph embedding in dimensions that do not require *ex post* interpretation or justification. Secondly, these results show that it is now possible to achieve this accuracy while producing embedding of networks in several countries in a common ARF. Third, in our example, δ_1 and δ_2 turned out to be good classifiers for labeled users, but this cannot be assured to be always the case, as no *a priori* interpretation can be given to ideological dimensions δ. (1) is invariant, for example, to rotations in $\mathbb{R}^N$. AE, on the contrary, leverages reference points and always aligns labeled users to the selected dimensions, having dimensions that have explicit meaning and bounds and reference points in space (*e.g.*, *0* and *10* being the most extreme positions for parties on each dimension).

VI. Conclusions and Discussion

In this article we have shown that social graphs data alone (to the exclusion of text data) can be used to embed large numbers of users from different national settings into common attitudinal reference frames (ARFs) taken from traditional instruments such as polls and surveys. These referential spaces have explicit meaning for dimensions, removing the need to inductively interpret what dimensions stand for in methods such as multi-dimensional scaling. In addition, this referential space has additional advantages, such as being endowed with explicit referential points in space, providing explicit reference for positions deemed, *e.g.*, as being centrist, extreme left or right. Most importantly, doing so 1) in a space common for several national settings, 2) on dimensions relevant for different disciplines, and 3) in a way that is language-independent, opens a path for large-scale comparative studies in social media, and that use positions of users, such as algorithm audit [7], social psychology [6], media consumption [8], or the study of social movements [11], [33]. Accuracy analysis of positions using independent data such as text profile descriptions shows that this method is comparable with ideological scaling while displaying the aforementioned advantages.

Acknowledgments

This work has been funded by the "European Polarisation Observatory" (EPO) of the CIVICA Consortium and by the French National Agency for Research (ANR) through grants ANR-19-CE38-0006 "Geometry of Public Issues" (GOPI) and ANR-18-IDEX-0001 "IdEx Université de Paris". Data declared the 19 March 2020 and 15 July 2021 at *Fondation Nationale de Sciences Politiques* (Sciences Po) in accordance with General Data Protection Regulation 2016/679 (GDPR) and Twitter policy. For further details and the respective legal notice, please visit https://medialab.sciencespo.fr/en/activities/epo/.

References

[1] K. Imai, J. Lo, J. Olmsted *et al.*, "Fast estimation of ideal points with massive data," *American Political Science Review*, vol. 110, no. 4, pp. 631–656, 2016.

[2] E. Bakshy, S. Messing, and L. A. Adamic, "Exposure to ideologically diverse news and opinion on facebook," *Science*, vol. 348, no. 6239, pp. 1130–1132, 2015.

TABLE III

ACCURACY OF A LOGISTIC REGRESSION CLASSIFIER USING THE PAIRS OF LABELED USERS ON IDEOLOGICAL AND ATTITUDINAL DIMENSIONS.

			France			Germany			Italy			Spain		
Partition	Method	Dimension	Prec.	Rec.	F1	Prec.	Rec.	F1	Prec.	Rec.	F1	Prec.	Rec.	F1
Left/Right	IS	δ_1	0.68	0.60	0.64	0.62	0.90	0.73	0.63	0.70	0.66	0.61	0.80	0.70
Left/Right	IS	δ_2	0.76	0.75	0.76	0.70	0.60	0.65	0.76	0.79	0.78	0.37	0.26	0.30
Left/Right	AE	$d_{\text{Left-Right}}$	0.87	0.76	0.81	0.64	0.67	0.66	0.80	0.81	0.81	0.64	0.94	0.76
Anti-elitism	IS	δ_1	0.96	0.81	0.87	1.0	0.43	0.60	0.36	0.37	0.37	0.74	0.68	0.71
Anti-elitism	IS	δ_2	0.82	0.54	0.65	0.97	0.45	0.61	0.66	0.67	0.67	0.93	0.61	0.73
Anti-elitism	AE	$d_{\text{Anti-elite}}$	0.97	0.78	0.86	0.99	0.51	0.68	0.63	0.58	0.60	0.73	0.67	0.70

[3] R. Bond and S. Messing, "Quantifying social media's political space: Estimating ideology from publicly revealed preferences on facebook," *American Political Science Review*, vol. 109, no. 1, pp. 62–78, 2015.

[4] P. Barberá, "Birds of the same feather tweet together: Bayesian ideal point estimation using twitter data," *Political analysis*, vol. 23, no. 1, pp. 76–91, 2015.

[5] P. Barberá, J. T. Jost, J. Nagler, J. A. Tucker, and R. Bonneau, "Tweeting from left to right: Is online political communication more than an echo chamber?" *Psychological science*, vol. 26, no. 10, pp. 1531–1542, 2015.

[6] P. Ramaciotti Morales, J.-P. Cointet, and J. Laborde, "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)*, 2020, pp. 217–221.

[7] P. Ramaciotti Morales and J.-P. Cointet, "Auditing the effect of social network recommendations on polarization in geometrical ideological spaces," in *15th ACM Conference on Recommender Systems, RecSys' 21*, 2021.

[8] J.-P. Cointet, P. Ramaciotti Morales, D. Cardon, C. Froio, A. Mogoutov, B. Ooghe-Tabanou, and G. Plique, "What colours are the yellow vests? an ideological scaling of facebook groups," *Statistique et Société*, 2021.

[9] E. Grossman and N. Sauger, "Economic internationalization and the decline of the left–right dimension," *Party Politics*, vol. 25, no. 1, pp. 36–49, 2019.

[10] P. Ramaciotti Morales, M. Berriche, and J.-P. Cointet, "Fake news diffusion in geometrical attitudinal spaces," in *8th International Conference on Computational Social Sciences*, Chicago, United States, Jul. 2022. [Online]. Available: https://hal.archives-ouvertes.fr/hal-03653473

[11] P. Ramaciotti Morales, J.-P. Cointet, B. Benbouzid, D. Cardon, C. Froio, O. F. Metin, B. Ooghe, and G. Plique, "Atlas multi-plateformes d'un mouvement social: Le cas des gilets jaunes," *Statistique et Société*, 2021.

[12] J.-P. Cointet, D. Cardon, A. Mogoutov, B. Ooghe-Tabanou, G. Plique, and P. Ramaciotti Morales, "Uncovering the structure of the french media ecosystem," *arXiv preprint arXiv:2107.12073*, 2021.

[13] P. Ramaciotti Morales, J.-P. Cointet, and G. Muñoz Zolotoochin, "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*, 2021.

[14] P. Ramaciotti Morales, J.-P. Cointet, G. Muñoz Zolotoochin, A. Fernández Peralta, G. Iñiguez, and A. Pournaki, "Inferring Attitudinal Spaces in Social Networks," 2022.

[15] K. T. Poole and H. Rosenthal, "A spatial model for legislative roll call analysis," *American Journal of Political Science*, pp. 357–384, 1985.

[16] M. McPherson, L. Smith-Lovin, and J. M. Cook, "Birds of a feather: Homophily in social networks," *Annual review of sociology*, vol. 27, no. 1, pp. 415–444, 2001.

[17] P. F. Lazarsfeld, R. K. Merton *et al.*, "Friendship as a social process: A substantive and methodological analysis," *Freedom and control in modern society*, vol. 18, no. 1, pp. 18–66, 1954.

[18] P. Barberá and G. Rivero, "Understanding the political representativeness of twitter users," *Social Science Computer Review*, vol. 33, no. 6, pp. 712–729, 2015.

[19] P. Ramaciotti Morales and G. Muñoz Zolotoochin, "Measuring the accuracy of social network ideological embeddings using language models," in *2022 International Conference on Information Technology & Systems (ICITS22)*, 2022.

[20] J. E. Uscinski, A. M. Enders, M. I. Seelig, C. A. Klofstad, J. R. Funchion, C. Everett, S. Wuchty, K. Premaratne, and M. N. Murthi, "American politics in two dimensions: Partisan and ideological identities versus anti-establishment orientations," *American Journal of Political Science*, vol. 65, no. 4, pp. 877–895, 2021.

[21] P. Ramaciotti Morales, J.-P. Cointet, and A. Pournaki, "Text-free multi-issue attitude mining from social network graph data," in *8th International Conference on Computational Social Sciences*, 2022.

[22] R. Bakker, L. Hooghe, S. Jolly, G. Marks, J. Polk, J. Rovny, M. Steenbergen, and M. A. Vachudova, "2019 chapel hill expert survey," *Chapel Hill*, 2020, www.chesdata.eu.

[23] R. C. Luskin, "Explaining political sophistication," *Political behavior*, vol. 12, no. 4, pp. 331–361, 1990.

[24] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," *arXiv preprint arXiv:1810.04805*, 2018.

[25] M. Greenacre, *Correspondence analysis in practice*, 2017.

[26] W. Lowe, "Understanding wordscores," *Political Analysis*, pp. 356–371, 2008.

[27] J. D. Carroll, E. Kumbasar, and A. K. Romney, "An equivalence relation between correspondence analysis and classical metric multidimensional scaling for the recovery of euclidean distances," *British Journal of Mathematical and Statistical Psychology*, vol. 50, no. 1, pp. 81–92, 1997.

[28] R. Penrose, "On best approximate solutions of linear matrix equations," in *Mathematical Proceedings of the Cambridge Philosophical Society*, vol. 52, no. 1. Cambridge University Press, 1956, pp. 17–19.

[29] I. Dokmanić and R. Gribonval, "Beyond moore-penrose part i: generalized inverses that minimize matrix norms," *arXiv preprint arXiv:1706.08349*, 2017.

[30] C. Mudde, "The populist zeitgeist," *Government and opposition*, vol. 39, no. 4, pp. 541–563, 2004.

[31] K. A. Hawkins and C. Rovira Kaltwasser, "Measuring populist discourse in the united states and beyond," *Nature human behaviour*, vol. 2, no. 4, pp. 241–242, 2018.

[32] I. Mani and I. Zhang, "knn approach to unbalanced data distributions: a case study involving information extraction," in *Proceedings of workshop on learning from imbalanced datasets*, vol. 126. ICML, 2003, pp. 1–7.

[33] P. Ramaciotti Morales, J.-P. Cointet, and C. Froio, "Posters and protesters," *Journal of Computational Social Science*, pp. 1–29, 2022.