Multidimensional political polarization in online social networks

Antonio F. Peralta [1,2,*], Pedro Ramaciotti [3], János Kertész [1,4] and Gerardo Iñiguez [1,5,6,7,†]

[1] *Department of Network and Data Science, Central European University, A-1100 Vienna, Austria*
[2] *Helmholtz Institute for Functional Marine Biodiversity (HIFMB), 26129 Oldenburg, Germany*
[3] *CNRS, Complex Systems Institute of Paris Ile-de-France (ISC-PIF), Sciences Po médialab & LPI, Université Paris Cité, France*
[4] *Complexity Science Hub, A-1080 Vienna, Austria*
[5] *Faculty of Information Technology and Communication Sciences, Tampere University, FI-33720 Tampere, Finland*
[6] *Department of Computer Science, Aalto University School of Science, FI-00076 Aalto, Finland*
[7] *Centro de Ciencias de la Complejidad, Universidad Nacional Autonóma de México, 04510 Ciudad de México, Mexico*

(Received 21 June 2023; accepted 21 December 2023; published 16 February 2024)

Political polarization in online social platforms is a rapidly growing phenomenon worldwide. Despite their relevance to modern-day politics, the structure and dynamics of polarized states in digital spaces are still poorly understood. We analyze the community structure of a two-layer, interconnected network of French Twitter users, where one layer contains members of Parliament and the other one regular users. We obtain an optimal representation of the network in a four-dimensional political opinion space by combining network embedding methods and political survey data. We find structurally cohesive groups sharing common political attitudes and relate them to the political party landscape in France. The distribution of opinions of professional politicians is narrower than that of regular users, indicating the presence of more extreme attitudes in the general population. We find that politically extreme communities interact less with other groups as compared to more centrist groups. We apply an empirically tested social influence model to the two-layer network to pinpoint interaction mechanisms that can describe the political polarization seen in data, particularly for centrist groups. Our results shed light on the social behaviors that drive digital platforms towards polarization and uncover an informative multidimensional space to assess political attitudes online.

DOI: 10.1103/PhysRevResearch.6.013170

I. INTRODUCTION

Understanding how people share information and influence each other in their political attitudes, potentially leading to ideological partisanship and political polarization [1], is a relevant yet challenging issue that has been tackled for decades using theories and methods from fields as diverse as sociology [2], political science [3,4], economics [5] and, more recently, complexity and computational social science [6–9]. Mathematical modeling is a frequently applied method to elucidate the mechanisms behind social influence and ideological polarization [10–15]. Often times, however, models that are otherwise conceptually robust and even inspired by empirical data, are investigated only theoretically through analytical derivations and numerical simulations on idealized synthetic populations [16,17]. An example are models of continuous political opinions that position individuals in, e.g., liberal-conservative scales, where the dimension capturing political ideology is defined a priori and not as the result of data analysis in the social context of interest. Indeed, one of the most challenging aspects of bridging opinion dynamics models and empirical observations of political attitudes in social networks is the number of dimensions determining social influence [18].

Ideal point estimation models [19] have been used to position large numbers of social media users in a liberal-conservative scale in several platforms [20,21], amounting to a single-dimensional opinion analysis where users are classified from the most liberal to the most conservative. And yet, social scientists acknowledge that political systems in Europe [22] and also increasingly in the US [23] are structured by several dimensions of opinion. Recent advancements in multidimensional political opinion estimation methods allow to embed structural data, such as communication networks coming from social media, into political spaces with multiple political dimensions [24]. In these spaces, dimensions act as continuous indicators of positive or negative attitudes towards identifiable issues of political debate. As an example, the DeGroot model of opinion dynamics [26] has recently been used to estimate multidimensional political attitudes of users in online platforms [27].

Here we propose a methodology to uncover and understand patterns of online political polarization in social media by combining embedding methods with empirically grounded opinion models in a multidimensional ideological space.

*peraltaaf@ceu.edu
†iniguezg@ceu.edu

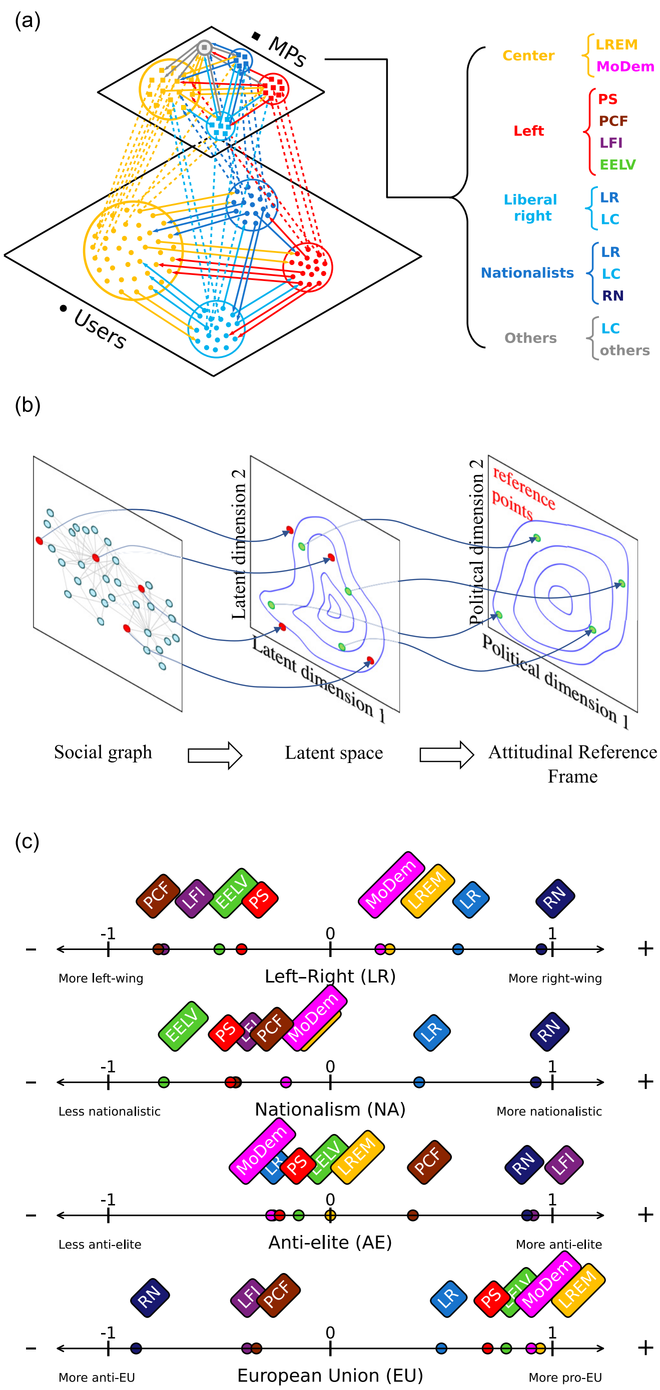

FIG. 1. Uncovering political opinions via online social network data. (a) Schematic diagram of network structure in French Twitter, with Members of Parliament (MPs) on top and regular users in the bottom. Colored circles highlight communities found by the planted partition model [28], which divides the network by assortativity for both MPs and users. Colors follow overall political leaning. MPs belong to five communities: *Center*, *Left*, *Liberal right*, *Nationalist,* and *Others* (see Ref. [25], Sec. S2). Parties in each community are shown by acronyms (for details on parties see Results). Color of links corresponds to the community of the source nodes; User → User and MP → MP links are represented by solid arrows, and User → MP links by dashed arrows (we disregard MP → User links). (b) Scheme of method to obtain politically relevant ideological positions of users and MPs. Twitter data are embedded in a multidimensional latent space preserving homophily: users who are close in this space have higher probability of following the same set of MPs. We compute positions of political parties in latent space as the mean position of MPs of the same party. Using these points and the corresponding party positions in political survey data, we map the network onto the opinion dimensions of the survey, forming a political Attitudinal Reference Frame (ARF, see MM and Ref. [25], Sec. S1.2.2). (c) Position of French political parties used as reference points to map the position of users on latent space (resulting from homophiliy embedding) onto the ARF. Party positions are taken from the 2019 Chapel Hill Expert Survey (CHES) data [29], built by a panel of experts in European politics (for details see MM and Ref. [25], Sec. S1.2.2).

Using follower networks in the online micro-blogging platform Twitter, with data from both professional politicians and regular users in France, we estimate the ideological positions of individuals along an optimal number of four politically relevant dimensions: left-right stance and attitudes towards nationalism, elites, and the European Union. By means of a network community detection method based on stochastic block-modelling, we first classify politicians and regular users into groups, according to assortative patterns in the Twitter interaction structure. We then embed the social graph in a latent space preserving homophily, where dimensions are interpreted as ideological indicators using a survey of political experts. Our four political dimensions are optimal in the sense that they capture main differences between competing parties in France. Relying on the community partition of the network and on the inferred positions along the detected four political dimensions, we propose formal measures that uncover polarized states and diverging ideologies. Finally, we introduce an opinion dynamics model capable of reproducing the large-scale behavior of empirical data, providing a plausible explanation for the influence mechanisms underlying the structure and dynamics of political polarization in multidimensional ideological spaces.

II. RESULTS

We gather network data from the follower → followed relations between the Twitter accounts of $M = 813$ Members of Parliament (MPs) in France and $N = 230\,254$ of their followers (here denoted regular users), who follow at least 3 MPs and follow another user that follows at least 3 MPs [Fig. 1(a)]. Based on this directed online social network, we infer politically relevant ideological positions of MPs and users via a two-step embedding method [Fig. 1(b)]. First, we embed nodes of the network onto a homophily preserving latent space (users close in space follow similar sets of MPs). Then, positions in this latent space are correlated to the attitude dimensions of a standard political survey, the 2019 Chapel Hill Expert Survey (CHES) [29]. The resulting space is comprised by four real-valued variables, or political dimensions, that represent attitudes towards: (i) the political left or right (LR), intended to measure the overall ideological stance of an individual (i.e., without specifying particular political issues to survey respondents); (ii) nationalism (NA); (iii) the European Union (EU); and (iv) the establishment and elites, intended to measure anti-elite sentiment (AE). Beyond the standard left-right dimension of political cleavage, our embedding process is able to identify additional dimensions capturing relevant differences between party positions in France (Fig. 1(c); for more details see Materials and Methods [MM], Ref. [24] for an implementation of the algorithm, and Ref. [25], Sec. 1).

A. Structurally cohesive groups share political attitudes

Since MPs arguably carry the political agenda by highlighting topics of interest to their parties and the general public, we choose to focus first on the part of the network involving MPs only, i.e., the MP → MP links [Fig. 1(a)]. We run a standard community detection algorithm on the MP layer to find the best partition into assortative groups (groups more connected to themselves than to others), by minimizing the description length of the network from an information-theoretical perspective [28,30].

We find four assortative communities, named *Center*, *Left*, *Liberal right*, and *Nationalists* by following traditional distinctions in French politics [31], plus a nonassortative group denoted *Others* (Fig. 2(a); see also MM and Ref. [25], Sec. S2). These names are determined by the positions of the corresponding MPs along the identified political dimensions [Figs. 1(c), 2(b), and 2(c)]. The *Center* is composed mainly of members of the French parties LREM (Macron's party Republic on the Move), and MoDem (Moderate Democrats), displaying centrist positions in all dimensions except EU (where it is the most pro-Europe community). The *Left* has a markedly left-leaning distribution, assembling most MPs from known left-wing parties (LFI, PCF, PS, and EELV, standing respectively for Indomitable France, the Communist, Socialist, and Ecologists parties). The rightmost communities in the LR dimension are named *Liberal right* (including some MPs from MoDem, but mostly from the LC and LR parties, standing for the Centrist and Republican parties) and *Nationalists* (including MPs from LC and LR, and notably all MPs from Le Pen's National Rally party, RN) to account for their differences along the NA and AE dimensions.

The *Others* group is composed of several parties across the whole political spectrum, and the structural patterns of its MPs do not fit any of the other groups. We observe that the dimensions of the latent space properly capture the attitudes of politicians expected by their party allegiance, and MPs of the same group are clustered together in opinion space, except for *Others*. We also identify interesting features in the opinion overlaps between groups: the *Liberal right* and *Nationalists* exhibit some overlap in their LR attitudes, but occupy different regions in the NA and AE dimensions [Fig. 2(c)].

These results further cement the need and real-world relevance of the political attitude dimensions comprising our multidimensional latent space. Beyond the traditional left-right cleavage, we find a dimension of attitudes towards institutions and elites (previously identified as relevant in French politics and political Twitter in general [32]), an ideological position towards nationalism that differentiates between two right-wing tendencies (*Liberal right* and *Nationalists*), and a variable encapsulating opinions with respect to the EU, also deemed significant in French politics [33] (see MM for a detailed discussion on the selection of these dimensions).

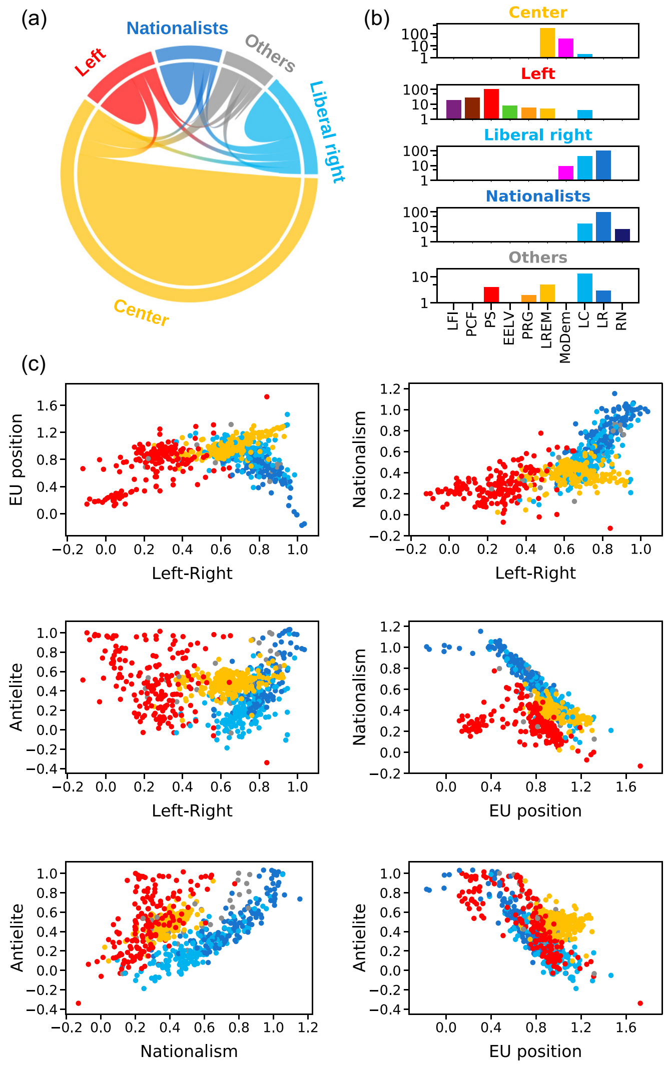

FIG. 2. Communities and ideological positions of professional politicians. Using the best partition (minimum description length) of the planted partition model [28], we find 5 MP communities: *Center*, *Left*, *Liberal right*, *Nationalists*, and *Others*. (a) Chord diagram indicating the connectivity (number of links) between and inside communities. The angular size of each community in the diagram is proportional to the number links that depart from it. (b) Party composition of each community. The list of parties (horizontal axis) is ordered according to their positions in the LR dimension. Bars indicate the number of MPs that belong to each party in the specified community (rows). We choose convenient colors for communities and parties in order to better visualize their political attitudes (see Ref. [25], Sec. S2). (c) Political positions of MPs (coloured according to their communities) in various two-dimensional projections of the four-dimensional political space, leading to six possible pairs of opinion variables: LR-EU, LR-NA, LR-AE, EU-NA, NA-AE, and EU-AE. The positions of MPs motivate the naming of each community.

B. Professional politicians have less extreme attitudes than regular users

We analyze the positions of regular users in latent space and compare them to the ideological positions of the MPs they follow (Fig. 3). Political positions of MPs lie

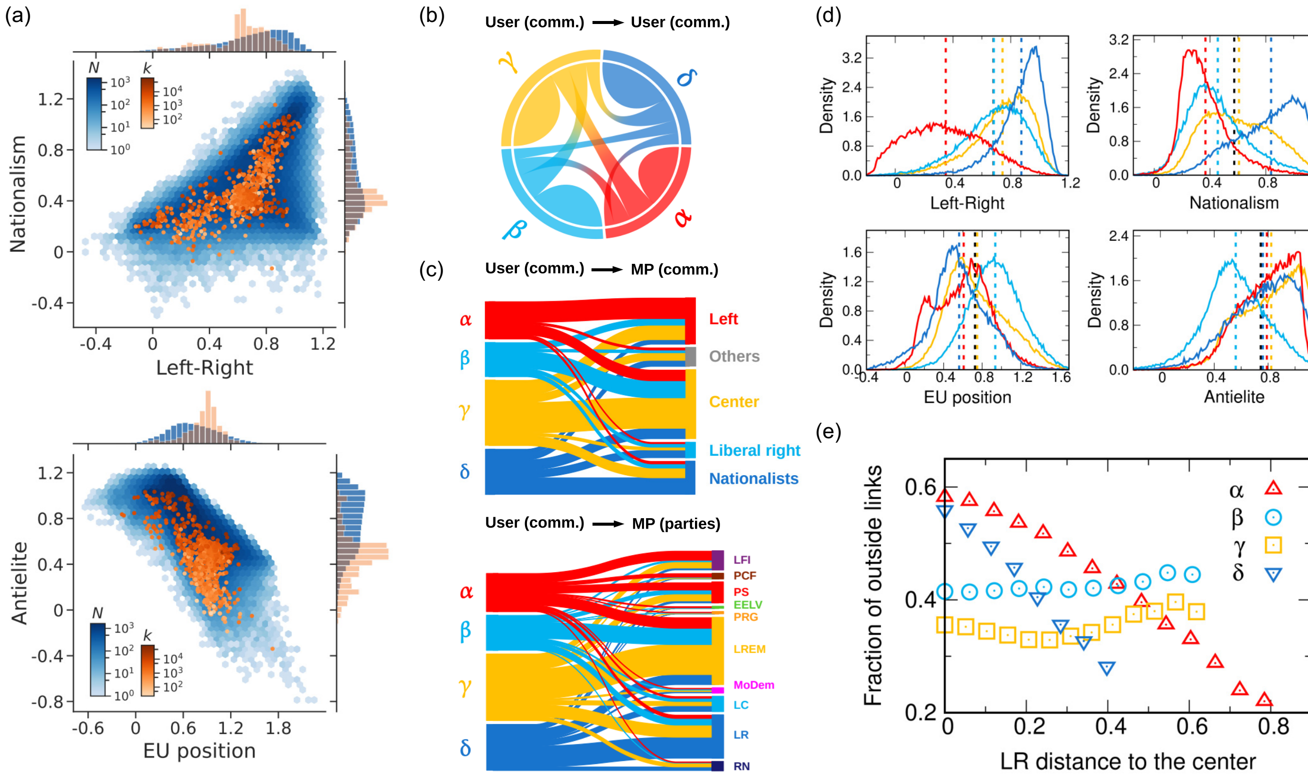

FIG. 3. Communities and ideological positions of regular users and their relation to professional politicians. (a) Number distribution of ideological positions of users (MPs), represented by blue (orange) dots in a two-dimensional opinion space for pairs of opinion variables: LR-NA (top) and EU-AE (bottom) (other pairs in Ref. [25], Sec. S1.3). Color shading for MPs is proportional to the number of followers k (users) of each MP in logarithmic scale, i.e., in-degree in the User → MP network, $k \equiv k_m^{\text{in/um}}$ (see Ref. [25], Sec. S1.1.1). Corresponding marginal probability densities of users (blue) and MPs (orange) are plotted in linear scale. (b) Communities in the user layer correspond to the best partition (minimum description length) of the planted partition model [28] with a fixed number of communities equal to four. The color of each community is chosen according to its characteristic political attitude in relation to the MP layer. The chord diagram indicates the connectivity (number of links) between and inside communities of users. (c) Sankey diagrams indicating the connectivity between user groups and both MP communities (top) and their parties (bottom). Size of flows is proportional to the number of links in the User → MP network, whose source nodes belong to a particular community of users (indicated by colors). Link colors are chosen according to user communities. (d) Probability densities of opinion variables (LR, NA, EU, and AE) of users in each community (as indicated by line color). Colored dashed lines represent the average opinion of each community, and the black dashed line is the global average. (e) Fraction of links of users (in the user layer) pointing outside of their community as a function of the distance of their opinion from the average opinion of all users. Plot corresponds to the LR dimension (for others see Ref. [25], Sec. S3.3.4). While members of β and γ connect freely to other communities despite of political differences, this function rapidly decreases with ideological distance for members of α and δ.

exclusively within the limits of the distribution of values for users. This means that, in our sample, the most extreme attitudes in French Twitter are held by the regular audience of the platform [Fig. 3(a)]. The difference in ideological extremism between MPs and users is most salient for the positions along the AE dimension, with MPs having a noticeable less anti-elite leaning than users (see Ref. [25], Sec. S1.3). From a political science strategic standpoint, this is to be expected, as politicians seek to position themselves as appealing to the largest possible number of users [34] (see Ref. [25], Sec. S1.3). Users also tend to align with the most popular MPs based on their number of followers, which ultimately produces a high concentration of opinions around popular MPs (see Ref. [25], Sec. S1.3). This is a first indication of the social influence mechanisms potentially driving the dynamics of political Twitter, such as opinion imitation or assimilation [17,35], which we explore further below.

We compare the structural patterns of connectivity of users and MPs by running the same community detection algorithm in the user layer, but now under a constraint of four groups, the same number of assortative communities found for MPs [Fig. 3(b)]. In other words, we focus on the particular level of the hierarchy of community structure among the general audience of Twitter that corresponds to the group divisions imposed by professional politicians (for details see Ref. [25], Sec. S2). We denote the resulting four communities by α, β, γ, and δ. At this level of granularity, the assortative communities of users also exhibit opinion coherence, both in terms of the distribution of political attitudes and their relation to the communities and parties of MPs [Fig. 3(c)].

The predominant political positions of user groups are strikingly informative [Fig. 3(d)]. The α community is the only one having a marked left-wing stance over the LR dimension. The δ group, the rightmost in the LR dimension, is the only one skewed towards the nationalist side of the NA spectrum. The more centrist communities β and γ lie in between α and δ in the LR dimension and show a low nationalist stance. Yet they differ in their attitudes towards elites and the EU, with the γ community showing a marked antielite sentiment. Three of these communities (α, β, and δ, respectively) somewhat correspond to identifiable political ideologies in France and their associated MP groups: the traditional left and two types of center to right-wing stances, the liberal and the nationalist right [33]. On the other extreme, community γ does not correspond to a single group in the MP layer; its users mainly follow *Center* politicians [see Fig. 3(c)] and show a strong antielite sentiment.

C. Groups with extreme political attitudes are more segregated

Assortative communities of users in political Twitter also differ in the way they connect to each other despite their ideological disagreement, as captured by our attitudinal latent space [Fig. 3(e)]. For each user community, we partition one of the dimensions of latent space (say, LR) into chunks, and compute the fraction of links (of users in that community and opinion interval) that lead to one of the other groups. This is a measure of community segregation, or political polarization, as a function of attitudinal positions in LR space (for details and other dimensions see Ref. [25], Sec. S3.3.4). The β and γ communities, roughly corresponding to the *Liberal right* and *Center*, show a flat trend, meaning that individuals identifying with these political ideologies interact with other groups despite their differences. Notably, the more right-wing β community is slightly better connected to others than the more centrist γ group (i.e., the fraction of outside links is larger on average). On the other hand, the politically extreme α and δ communities, mostly associated to the *Left* and *Nationalists*, have a decreasing trend in their connectivity to users with diverging ideologies, highlighting their segregation both in terms of structural connectivity and attitudinal positions in latent space. The α and δ groups are in this sense more heterogeneous yet assortative, since the political stances of their users are a strong indication of their degree of homophily with peers, echoing recent findings of increasing political polarization in US Twitter [8,9].

D. Modeling multidimensional political polarization online

The positioning of professional politicians and regular users of French Twitter in a multidimensional attitudinal space indicates that people form structurally cohesive groups that become more segregated as their political ideologies diverge. Regular users also tend to be more extreme towards topics of political debate, while concentrating their attention on popular MPs. In order to identify potential idealized mechanisms that might explain this behavior, we explore an opinion dynamics model with social influence processes based on the results of controlled psychological experiments [36,37].

In the model, each user i holds a vector opinion $\vec{v}_i(t) = (x_i(t), y_i(t), z_i(t), w_i(t))$ at time t that determines its position in attitudinal latent space (the LR, NA, EU, and AE dimensions, respectively). Each MP m has a static vector opinion $\vec{V}_m = (X_m, Y_m, Z_m, W_m)$ that we extract from data. This assumption can be understood as a separation of timescales: regular users of Twitter change their minds and who they follow faster than MPs, who pursue the political agenda of their parties at a slower pace. Explicitly, the fast timescale, the opinion dynamics of users, is coupled to the slow one, the opinion dynamics of MPs, following its changes closely; which simplifies the analysis within the context of the slower dynamics.

In reality, both user and MP opinions evolve over time and influence each other. However, due to the visibility of the political position of MPs in the public sphere, we can argue that their opinion is in some sense more influential than that of regular users. Factors supporting this premise include the presence of MPs in the news across various media beyond Twitter, such as newspapers, television, radio, other social platforms, and websites. Another factor that bolsters this assumption is that MPs, due to the organization of political parties and the nature of the political agenda, are less prone to idiosyncratic changes. From a political theoretical standpoint, this follows a strategic logic in which politicians position themselves by displaying ideological and issue positions as the supply, which the public (the demand side) can chose via voting, or, in the case of our study, via following [34]. This distinction in the nature of opinions can be summarized as a public of users holding ideologies and attitudes that determine their choices, while MPs choose to display a set of ideologies and opinions not subject to change due to the positions of other MPs. In summary, we assume that users are influenced by other users and MPs, but not the other way around [meaning we ignore links from MPs to users; see Fig. 1(a)].

The dynamics of the model is as follows [Fig. 4(a)]. In a time step $\Delta t = 1/N$, a randomly selected user i interacts with one of its neighbors, either another user j or an MP m, who influences the opinion of i according to

$$\vec{v}_i(t + \Delta t) = \vec{v}_i(t) + I_{ij}[\vec{v}_j(t) - \vec{v}_i(t)] \tag{1}$$

or

$$\vec{v}_i(t + \Delta t) = \vec{v}_i(t) + I_{im}[\vec{V}_m - \vec{v}_i(t)], \tag{2}$$

where the influence factors I_{ij} and I_{im} are drawn at each time step from a predefined probability density function $f(I)$ [Fig. 4(b)] (note that in our model, opinion dimensions do not significantly interact; for related modeling approaches see Refs. [15,38]). We follow the dynamics until the system is stationary, that is, until the distribution of opinion values in all dimensions is stable. The distribution $f(I)$ captures a spectrum between prototypical influence processes, here denoted *Keep* ($I = 0$) and *Adopt* ($I = 1$), i.e., not being influenced by a neighbor or fully imitating its behavior [for details on the choice of $f(I)$ see MM].

In Ref. [37], the authors quantify the change in the opinion of subjects under the influence of others, Eqs. (1) and (2), and obtain the probability kernel $f(I)$. This exhibits two pronounced peaks at $I = 0$ and 1, with some dispersion for intermediate values. Inspired by these experimental results, we propose a parametrization of the kernel consisting of two Gaussians [Fig. 4(b)]. The free parameters of the kernel are the probability p of having an influence factor around $I = 0$ ($1 - p$ around $I = 1$), and the standard deviations σ_K and σ_A of the Gaussians peaked around $I = 0$ and $I = 1$, respectively.

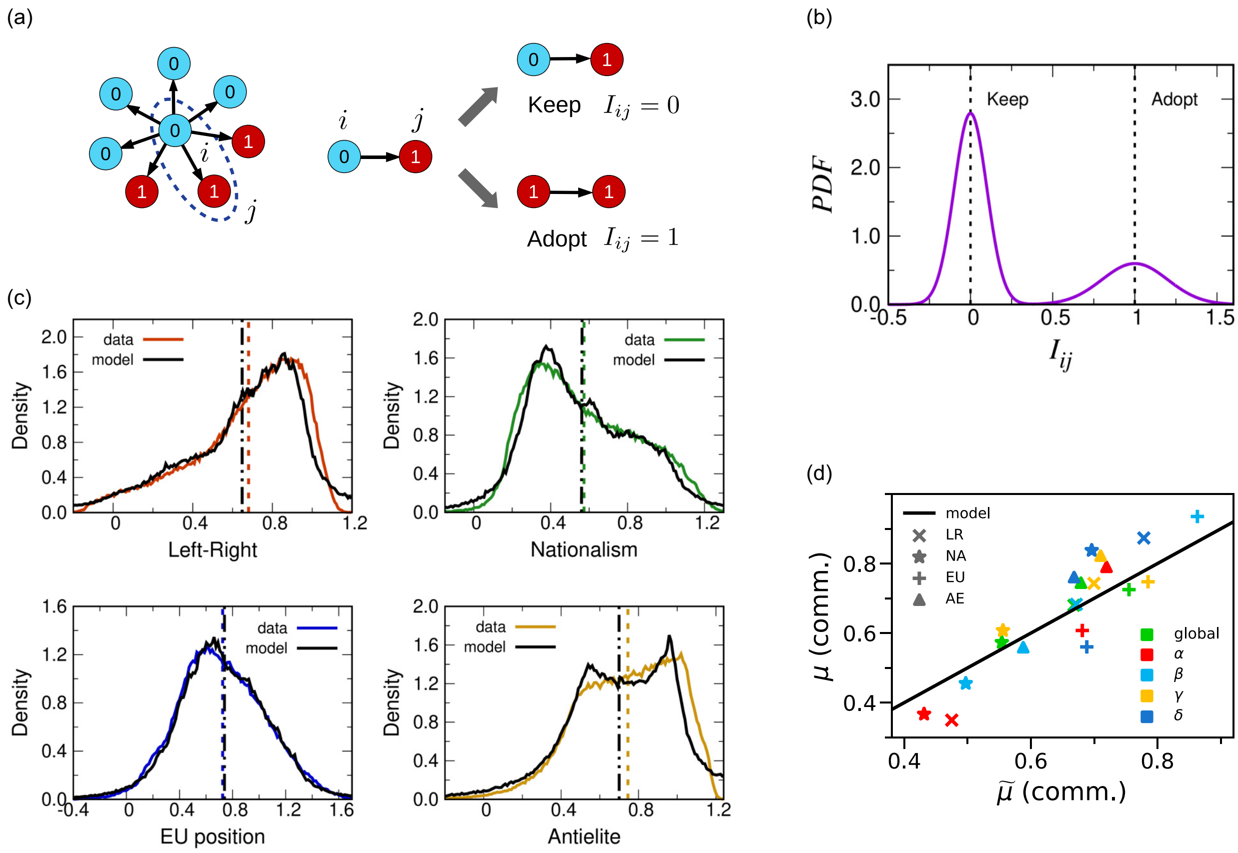

FIG. 4. Modelling multidimensional political polarization online. (a) In our minimal social influence model, user i interacts with its neighbor, another user j or MP m, and decides to either keep its own opinion or incrementally adopt the neighbor's position according to influence factor I_{ij} [see Eqs. (1) and (2)]. (b) We model the influence factor I_{ij} as a sum of two Gaussians peaked around $I_{ij} = 0$ and 1. The height and width of the peaks are parametrized by (p, σ_K, σ_A), which we determine by fitting (see MM). (c) Probability density function of attitudinal positions of users in latent space (LR, NA, EU, and AE) in both empirical data (colored solid lines) and best fit of stationary state of the model (black solid lines). Colored (black) dashed lines represent the average opinion of data (model) in a given political dimension. (d) Average opinion μ of users as a function of the weighted average opinion $\widetilde{\mu}$ of the MPs they follow [see Eq. (3)]. Weights are the in-degrees of the MPs coming from the considered set of users. Each point corresponds to either the whole network (global) or certain community of users (indicated by colors; see Fig. 3) and for a given opinion variable (LR, NA, EU, and AE). The straight line is the degree-based mean field approximation $\mu = \widetilde{\mu}$ (see Ref. [25], Sec. S3.2).

We also assume that the two influence factors I_{ij} and I_{im} are drawn from the same distribution. Crucially, we introduce a parameter λ controlling the ratio of rates of interactions with either users [Eq. (1)] or MPs [Eq. (2)] (see Ref. [25], Sec. S3.1).

We fit the model by estimating $f(I)$ (i.e., p, σ_K, and σ_A) and λ such that the difference in the marginal distributions of all opinion components between data and model are minimized (for details see Ref. [25], Sec. S3.3.1). The fitted parameters take reasonable values, and the shape of $f(I)$ is comparable with that of experiments [37]. Notably, we obtain a high value of p, which indicates that keeping your opinion after an interaction is more probable than adopting another one. We also obtain a high value of λ, showing that MPs influence users more often than other users do (further details on fitting, its accuracy, and a table of parameter values in Ref. [25], Sec. S3.3.1). Despite its simplicity, numerical simulations of the stationary state of the fitted model recover the attitudinal positions of most users across the entire latent space [Fig. 4(c)], with some deviations at the extremes of the political spectrum, particularly in the LR, NA, and AE dimensions. Our results imply that the collective decisions of users to either keep their own opinions or incrementally get influenced by others are compatible with the amount of political polarization seen in data.

The levels of political polarization across communities in this latent space are further clarified by a degree-based mean field analysis [39] of our model [Fig. 4(d)]. Since the attitudes of MPs are static, the average opinion μ of users along a given dimension is approximately equal to the degree-weighted average opinion $\widetilde{\mu}$ of MPs they follow. In terms of, say, the stationary opinions $x_i^{(\mathrm{st})}$ in the LR dimension, we have

$$\frac{1}{N}\sum_{i=1}^{N} x_i^{(\mathrm{st})} = \mu \approx \widetilde{\mu} = \frac{\sum_{m=1}^{M} k_m^{\mathrm{in/um}} X_m}{\sum_{m=1}^{M} k_m^{\mathrm{in/um}}}, \qquad (3)$$

where $k_m^{\mathrm{in/um}}$ is the number of users following MP m (see Ref. [25], Sec. S3.2). Attitudinal positions in French Twitter roughly follow the mean-field trend $\mu \approx \widetilde{\mu}$ in all dimensions of the latent space, both at the global level and when separating users by their political communities (see Fig. 3). In data, however, this approximately linear relation has a slope higher than 1, implying that users have even more radical attitudes than the MPs they follow, especially at the extremes of the multidimensional political spectrum [see Fig. 3(a)].

III. DISCUSSION

Our results show that political polarization in online social networks cannot be reduced to a single dimension. This contrasts with a stream of recent research leveraging ideological scaling in social media data, which focuses in uni-dimensional left-right models. Using embedding methods based on large-scale Twitter and political survey data, we uncover at least four dimensions that capture relevant attitudinal differences across political groups in France. We observe that both professional politicians and regular users of Twitter create cohesive communities of similarly minded people, but users are more extreme in their attitudes and may distance themselves from groups with dissimilar political leanings, further polarizing the online platform. Indeed, there is a clear but nuanced relationship between the group segregation in this multidimensional latent space and the political party structure in France, highlighting how real-world political cleavages are reflected in online activity.

Political polarization is intrinsically multidimensional and thus depends on particular topics of public debate. In France, the political left-right and the nationalism issue segregate online communities the most, while attitudes towards the European Union and against socio-economic elites are less polarizing. We observe a strong relationship between the intra- and interconnectivity of communities and the political opinions of their members. The centrist communities β and γ interact quite uniformly with other groups, while the more extreme communities α and δ (in the left and right of the political spectrum) connect less with other groups as the political disagreement between them increases.

Identifying and understanding the characteristics of individuals in distinct regions of multidimensional political spaces is of importance to several lines of research, with broad implications for policy making, political campaigning, grassroots movements, and collective social phenomena in democratic spaces. By virtue of their engagement with professional politicians, the inferred attitudinal positions of a sample set of citizens could be harnessed in, e.g., the run up to elections. From a political space competition perspective [40], eligible candidates might take positions appealing to voters in certain regions of a previously identified latent space. Identifying the users and political spatial regions under-served by existing candidates is a potential benefit of our methodology, which, together with text analysis of opinions in social media, may have relevant implications for political strategies. Other applications include the study of online social movements [41], discovery of political biases in algorithms [42–44], and polarization in online news media consumption [8,45,46].

To complement our statistical analysis, we have explored a model that replicates the positions of professional politicians and Twitter users in opinion space and pinpoints the basic social mechanisms, such as imitation, that might drive the levels of political polarization seen online. The fitted parameter λ (a ratio of the frequency of interactions with politicians) indicates, in accordance with empirical observations, that MPs lead the dynamics. The relation between the opinions of users and MPs predicted by the model shows a good fit with data. Notably, the global opinion average of users is independent of model parameters and might be thought of as a fundamental property of the proposed imitation mechanism. The model recovers this fundamental property at the global level, but there are some deviations for individual communities. The discrepancies are mostly at the extremes: the average opinion of communities with extremist individuals is more extreme than predicted by the model. This indicates that, in addition to imitation, further mechanisms are potentially at play in the dynamics of polarization in French Twitter.

Taking into account influence mechanisms based on similarity or other socio-cognitive biases is a further step to investigate in the future. Indications that the introduction of biases would enhance the accuracy of the model are suggested by our results [see Fig. 4(d)]. The linear trend that results from the mean field approach matches the data, but the slope seems to be higher than predicted by the model. The introduction of biases in the interaction mechanisms, like bias assimilation as proposed in Ref. [35], may increase this slope. Additionally, the deviations observed between model and data for extremist users might be corrected by introducing biases.

The emergence of political cleavages as indicated by interactions in online social media is an inherently temporal and cultural phenomenon. As the political agenda evolves and the topics of national debate transition from one administration to the next, the dimensions of our ideological space relevant to political polarization will also change. The results of the embedding process might also depend on the selected online platform and the country for which data is gathered. How does this opinion space vary across countries and time? And, perhaps more crucially, what characteristic dimensions of political polarization are common around the world, despite cultural differences? Our results offer a flexible framework to further explore these tantalizing questions.

IV. MATERIALS AND METHODS

A. Twitter network data

The network is obtained via the Twitter accounts of Members of Parliament (MPs) in France for deputies [47] and Ref. [48] for senators. We have data on 813 MPs (out of 925), including 348 senators and 577 deputies, each one belonging to one of ten political parties: LREM (La République en Marche), LR (Les Républicains), PS (Le Parti Socialiste), LFI (La France Insoumise), LC (Les Centristes), RN (Rassemblement National), PCF (Parti Communiste Français), MoDem (Mouvement démocrate), PRG (Parti Radical de Gauche), and EELV (Europe Écologie–Les Verts). Followers of the MPs were collected in May 2019, from which we keep only the 230 254 users with sufficiently high number of political interactions on Twitter (see Ref. [25], Sec. S1.1 for details on how we filter data).

Considering these two types of nodes, MPs and users, we categorize their links (follower → followed interactions collected for the period August–December 2020) as User → User (63 625 921), User → MP (3 351 359), MP → MP (113 596) and MP → User (515 882). The average number of followers of the MPs (4122) is higher than that of users (276) (see Ref. [25], Sec. S1.1 for additional statistics on data collection).

B. Latent space embedding

For the political positions of MPs and users we rely on the computation of Ref. [24] (see Acknowledgments for details). From the described data, political positions of individuals in a four-dimensional space are computed in two steps as follows [see Fig. 1(b)]. In the first step, we consider the bipartite network of MPs and users (User → MP links) and create an embedding in a multidimensional latent space preserving homophily: Users closer in space have higher chances of following the same MPs, and MPs closer in space have higher chances of being followed by the same users. To produce this embedding, a generative homophilic process is considered for the bipartite network of MPs and their follower users [20,32]:

$$P(\text{User}_i \rightarrow \text{MP}_j) = \text{logit}^{-1}(\alpha_i + \beta_j - \gamma \|\phi_i - \phi_j\|^2), \quad (4)$$

where $P(\text{User}_i \rightarrow \text{MP}_j)$ is the probability of observing User_i following MP_j, α_i is the level of activity of User_i (number of followed friends), β_j is the popularity of MP_j (number of followers), γ is a sensitivity parameter, and ϕ_i and ϕ_j are unobservable positions of User_i and MP_j in latent space.

The first step takes the bipartite graph of MPs and users as observations to compute Bayesian inference of ϕ values for them [see Fig. 1(b)]. This is done by performing a correspondence analysis [49] of the adjacency matrix of the bipartite graph as an approximation of the unobservable positions of MPs and users in Eq. (4) [50]. Correspondence analysis, being a factor analysis method, preserves global properties such as distance, i.e., up to affine transformations. This latent space embedding for the bipartite graph assures that the relation of relative distances are preserved, in contrast to other network embedding methods [51] (see Ref. [52] for an evaluation of this approximation, and Ref. [25], Sec. S1.2.1 for the first step leading to the latent space embedding).

C. Political survey data

The second step of the embedding process uses political survey data to map latent space positions onto a second space where dimensions do have explicit meaning, as they stand for attitudes towards identifiable issues of political debate [see Fig. 1(b)]. The 2019 Chapel Hill Expert Survey (CHES) data [29] contains positions of political parties in France (and across Europe) in 51 policy and ideological dimensions. We call this space the Attitudinal Reference Frame (ARF; Ref. [25], Sec. S1.2.2). To map positions from the latent space onto this ARF, we use positions of political parties to compute an affine transformation. For each party, we compute the position in latent space as the centroid or mean of the positions of MPs that belong to that party. Knowing party positions on both the latent space and the ARF, we compute an affine transformation mapping positions of the former onto the latter by choosing the number of latent dimensions that fully determine the parameters of the affine transformation (see Ref. [25], Sec. S1.2.2 for more details on this transformation).

The positions of French political parties, as captured by the 51 CHES dimensions, can be described almost completely with only four dimensions, as shown by principal component analysis of the CHES dimensions (see Sec. IV in Ref. [53]). The four dimensions deemed relevant for our analysis are left-right (LR, variable lrgen in CHES), antielite salience (AE, variable antielite_salience in CHES), attitudes towards European integration (EU, variable eu_position in CHES), and nationalism (NA, variable nationalism in CHES). The ARF is built with explicit spatial reference points, e.g., the question that experts answer to position parties on the left-right scale is "Where do you position the party in terms of its overall ideological stance, 0 being extreme left, 5 being centrist, and 10 being extreme right?" (for the questions defining all four dimensions in CHES data see Ref. [25], Sec. S1.2.2). We further normalize the scales so that bounds of each dimension of the survey match the [0, 1] interval, making them comparable.

D. Validation of embedding and robustness

To test the validity of positions in ARF and their robustness, we use text written by users on their Twitter profiles. We select subsets of users by keywords that reveal their political leaning in their bio profiles, and check that they are correctly positioned in, e.g., the left-right scale (see Ref. [54] for a detailed presentation of this text-based validation approach). We focus on a limited set of keywords that must be correctly positioned: "left" ("gauche") and "right" ("droite") on the LR dimension; "Europe" and "European" ("européen") on the EU dimension; "people" ("peuple") and "elites" on the AE dimension; and "patriot" ("patriote") on the NA dimension. When computing metrics for a correct positioning of users that use these keywords along our four political dimensions, we further filter out users that have written a bio profile with negative sentiment (as computed via a sentiment analysis model), to minimize the probability of a user uttering criticism rather than support for an issue.

We evaluate two qualities for the density of users that express support for these issues. First, we evaluate positioning; users that express support should be concentrated in the corresponding region of ideological space. For example, most users with the keyword "right" with positive sentiment are positioned to the right of value 0.5 on the LR dimension. Second, we evaluate the monotonicity of the density of users that use a keyword over the various dimensions. For example, the proportion of users with the keyword "right" on their profile with positive sentiment grows with higher values along the LR dimension (see Ref. [55] for a bootstrap robustness analysis of the positions of users, and Ref. [25], Sec. S1.2.2 for more details on this dataset and the quality metrics for positioning individuals along the four dimensions).

E. Network community detection

The community analysis of the network data is performed with the PYTHON library GRAPH-TOOL [56]. We use the minimum description length of communities as a measure of goodness of fit, which is optimal in the sense that it avoids under- and overfitting and minimizes the occurrence of spurious communities. For the community detection model, we choose a version of the stochastic block model known as planted partition model [28], which constrains the community search to find structural patterns based on assortativity properties. Assortative groups are characterized by nodes that are connected mostly to other nodes of the same group (see Ref. [25], Sec. S2). First we apply the method to find communities in the MP layer and find five groups, four of which are assortative and one nonassortative. Then we identify the modular structure in the user layer by constraining the search to four communities (for more details see Ref. [25], Sec. S2).

F. Opinion dynamics model

The agent-based model defined by Eqs. (1) and (2) establishes the dynamics of opinions and evolves in discrete time steps. It considers two possible outcomes for each user at every time step: one represents the adoption or imitation of the opinion of a neighbor ("Adopt"), and the other the preservation of its current opinion ("Keep"). We consider a bimodal distribution for the influence factor I, determining whether a user keeps its opinion or adopts a new one [see Fig. 4(b)]. We parametrize this bimodal distribution as a mixture model: $f(I) = p\mathcal{N}(I; 0, \sigma_K^2) + (1-p)\mathcal{N}(I; 1, \sigma_A^2)$, where $\mathcal{N}(x; \mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} \exp[-(x-\mu)^2/2\sigma^2]$ is a normal distribution with mean μ and variance σ^2. We use the same distribution for both I_{ij} and I_{im} in Eqs. (1) and (2).

The parameter input of the model is as follows: (i) (p, σ_K, σ_A) for the influence distribution $f(I)$; (ii) λ as a measure of the frequency at which users interact with MPs compared to other users [both (i) and (ii) are fitting parameters]; and (iii) the network of interactions and the opinions of MPs, $\{X_m\}_{m=1,\dots,M}$, which we assume to be constant and extract from the data. The dynamics and stationarity of the model can be obtained by means of numerical (Monte Carlo) simulations, for which we can optionally apply boundary conditions in opinion space (see Ref. [25], Sec. S3.3.2). At the degree-based, mean-field level (Ref. [25], Sec. S3.2), the average opinion of users in Eq. (3) depends only on input (iii) through the degree-weighted average of MP opinions. The variance of user opinions depends also on (i) and (ii).

ACKNOWLEDGMENTS

This work has been funded by the "European Polarisation Observatory" (EPO) of the CIVICA Consortium. P.R. acknowledges support by the Data Intelligence Institute of Paris (diiP) through the French National Agency for Research (ANR) Grant No. ANR-18-IDEX-0001 "IdEx Université de Paris" and SoMe4Dem (Grant No. 101094752) Horizon Europe project. G.I. and J.K. acknowledge support from AFOSR (Grant No. FA8655-20-1-7020), project EU H2020 Humane AI-net (Grant No. 952026), and CHIST-ERA project SAI (Grant No. FWF I 5205-N). J.K. acknowledges support from European Union's Horizon 2020 research and innovation programme under Grant Agreement No. ERC No 810115 - DYNASNET. Data declared in 19 March 2020 and 15 July 2021 at the registry of data processing of Fondation Nationale de Sciences Politiques (Sciences Po) in accordance with General Data Protection Regulation 2016/679 (GDPR) and Twitter policy.

A.F.P., P.R., J.K., and G.I. conceived and designed the study. P.R. collected Twitter data and performed the embedding process. A.F.P. analyzed the data and explored the opinion dynamics model. J.K. and G.I. contributed to the network analysis. A.F.P., P.R., J.K., and G.I. interpreted the results and wrote the paper.

[1] N. McCarty, *Polarization: What Everyone Needs to Know* (Oxford University Press, New York, 2019)

[2] D. Baldassarri and P. Bearman, Dynamics of political polarization, Am. Sociol. Rev. **72**, 784 (2007).

[3] M. P. Fiorina and S. J. Abrams, Political polarization in the american public, Annu. Rev. Polit. Sci. **11**, 563 (2008).

[4] M. Prior, Media and political polarization, Annu. Rev. Polit. Sci. **16**, 101 (2013).

[5] A. K. Dixit and J. W. Weibull, Political polarization, Proc. Natl. Acad. Sci. USA **104**, 7351 (2007).

[6] M. Conover, J. Ratkiewicz, M. Francisco, B. Gonçalves, F. Menczer, and A. Flammini, Political polarization on Twitter, in *Proceedings of the International AAAI Conference on Web and Social Media*, Vol. 5 (AAAI, 2011), pp. 89–96.

[7] C. A. Bail, L. P. Argyle, T. W. Brown, J. P. Bumpus, H. Chen, M. F. Hunzaker, J. Lee, M. Mann, F. Merhout, and A. Volfovsky, Exposure to opposing views on social media can increase political polarization, Proc. Natl. Acad. Sci. USA **115**, 9216 (2018).

[8] J. Flamino, A. Galeazzi, S. Feldman, M. W. Macy, B. Cross, Z. Zhou, M. Serafino, A. Bovet, H. A. Makse, and B. K. Szymanski, Political polarization of news media and influencers on Twitter in the 2016 and 2020 US presidential elections, Nat. Human Behav. **7**, 904 (2023).

[9] M. Hohmann, K. Devriendt, and M. Coscia, Quantifying ideological polarization on a network using generalized Euclidean distance, Sci. Adv. **9**, eabq2044 (2023).

[10] T. C. Schelling, Dynamic models of segregation, J. Math. Sociol. **1**, 143 (1971).

[11] M. Granovetter, Threshold models of collective behavior, Am. J. Sociol. **83**, 1420 (1978).

[12] R. Axelrod, The dissemination of culture: A model with local convergence and global polarization, J. Conf. Resolut. **41**, 203 (1997).

[13] C. Castellano, S. Fortunato, and V. Loreto, Statistical physics of social dynamics, Rev. Mod. Phys. **81**, 591 (2009).

[14] P. Holme and F. Liljeros, Mechanistic models in computational social science, Front. Phys. **3**, 78 (2015).

[15] F. Baumann, P. Lorenz-Spreen, I. M. Sokolov, and M. Starnini, Emergence of polarized ideological opinions in multidimensional topic spaces, Phys. Rev. X **11**, 011012 (2021).

[16] P. Sobkowicz, Modelling opinion formation with physics tools: Call for closer link with reality, JASSS **12**, 11 (2009).

[17] A. F. Peralta, J. Kertész, and G. Iñiguez, Opinion dynamics in social networks: From models to data, arXiv:2201.01322.

[18] K. Benoit and M. Laver, The dimensionality of political space: Epistemological and methodological considerations, Eur. Union Polit. **13**, 194 (2012).

[19] K. Imai, J. Lo, and J. Olmsted, Fast estimation of ideal points with massive data, Am. Polit. Sci. Rev. **110**, 631 (2016).

[20] P. Barberá, Birds of the same feather tweet together: Bayesian ideal point estimation using twitter data, Polit. Anal. **23**, 76 (2015).

[21] R. Bond and S. Messing, Quantifying social media's political space: Estimating ideology from publicly revealed preferences on Facebook, Am. Polit. Sci. Rev. **109**, 62 (2015).

[22] R. Bakker, S. Jolly, and J. Polk, Complexity in the european party space: Exploring dimensionality with experts, Eur. Union Polit. **13**, 219 (2012).

[23] 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, Am. J. Polit. Sci. **65**, 877 (2021).

[24] P. Ramaciotti Morales, J.-P. Cointet, G. Muñoz Zolotoochin, A. Fernández Peralta, G. Iñiguez, and A. Pournaki, Inferring attitudinal spaces in social networks, Social Network Analysis and Mining **13**, 14 (2023).

[25] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevResearch.6.013170 for additional details about the data, model, methodology used, and extra results obtained. It is structured in three main sections: (i) Data details, (ii) Communities, and (iii) Model. Each of these sections is further divided into subsections that address specific aspects. The file includes 14 figures and 2 tables.

[26] M. H. DeGroot, Reaching a consensus, J. Am. Stat. Assoc. **69**, 118 (1974).

[27] S. Martin-Gutierrez, J. C. Losada, and R. M. Benito, Multipolar social systems: Measuring polarization beyond dichotomous contexts, Chaos, Solitons & Fractals **169**, 113244 (2023).

[28] L. Zhang and T. P. Peixoto, Statistical inference of assortative community structures, Phys. Rev. Res. **2**, 043271 (2020).

[29] S. Jolly, R. Bakker, L. Hooghe, G. Marks, J. Polk, J. Rovny, M. Steenbergen, and M. A. Vachudova, Chapel Hill Expert Survey trend file, 1999–2019, Electoral Studies **75**, 102420 (2022), www.chesdata.eu.

[30] P. D. Grunwald, *The Minimum Description Length Principle* (The MIT Press, Cambridge, MA, 2007).

[31] R. Rémond, *Les droites en France* (Aubier, 2014).

[32] 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), pp. 333–338.

[33] L. Hooghe and G. Marks, Cleavage theory meets europe's crises: Lipset, rokkan, and the transnational cleavage, J. Eur. Public Policy **25**, 109 (2018).

[34] W. H. Riker and P. C. Ordeshook, A theory of the calculus of voting, Am. Pol. Sci. Rev. **62**, 25 (1968).

[35] P. Dandekar, A. Goel, and D. T. Lee, Biased assimilation, homophily, and the dynamics of polarization, Proc. Natl. Acad. Sci. USA **110**, 5791 (2013).

[36] M. Moussaïd, J. E. Kämmer, P. P. Analytis, and H. Neth, Social influence and the collective dynamics of opinion formation, PLoS ONE **8**, e78433 (2013).

[37] A. Chacoma and D. H. Zanette, Opinion formation by social influence: From experiments to modeling, PLOS ONE **10**, e0140406 (2015).

[38] X. Wang, A. D. Sirianni, S. Tang, Z. Zheng, and F. Fu, Public discourse and social network echo chambers driven by socio-cognitive biases, Phys. Rev. X **10**, 041042 (2020).

[39] M. A. Porter and J. P. Gleeson, Dynamical systems on networks, Front. Appl. Dyn. Syst. Rev. Tutor. **4** (2016).

[40] A. Downs, An economic theory of political action in a democracy, J. Political Econ. **65**, 135 (1957).

[41] J.-P. Cointet, P. R. Morales, D. Cardon, C. Froio, A. Mogoutov, B. Ooghe, and G. Plique, What colours are the yellow vests? An ideological scaling of Facebook groups, Statistique et Société **9**, 79 (2021).

[42] E. Bozdag, Bias in algorithmic filtering and personalization, Ethics Inf. Technol **15**, 209 (2013).

[43] A. F. Peralta, M. Neri, J. Kertész, and G. Iñiguez, Effect of algorithmic bias and network structure on coexistence, consensus, and polarization of opinions, Phys. Rev. E **104**, 044312 (2021).

[44] F. Huszár, S. I. Ktena, C. O'Brien, L. Belli, A. Schlaikjer, and M. Hardt, Algorithmic amplification of politics on Twitter, Proc. Natl. Acad. Sci. USA **119**, e2025334119 (2022).

[45] D. J. Watts, D. M. Rothschild, and M. Mobius, Measuring the news and its impact on democracy, Proc. Natl. Acad. Sci. USA **118**, e1912443118 (2021).

[46] M. Falkenberg, A. Galeazzi, M. Torricelli, N. D. Marco, F. Larosa, M. Sas, A. Mekacher, W. Pearce, F. Zollo, W. Quattrociocchi, and A. Baronchelli, Growing polarization around climate change on social media, Nature Climate Change **12**, 1114 (2022).

[47] Obtained from http://www2.assemblee-nationale.fr/deputes/liste/reseaux-sociaux for deputies and http://www.senat.fr/espace_presse/actualites/201402/les senateurs_sur_twitter.html for senators.

[48] http://www.senat.fr/espace_presse/actualites/201402/les.

[49] M. Greenacre, *Correspondence Analysis in Practice* (Chapman and Hall, 2017).

[50] W. Lowe, Understanding wordscores, Polit. Anal. **16**, 356 (2008).

[51] T. Chari and L. Pachter, The specious art of single-cell genomics, PLoS Comput. Biol. **19**, e1011288 (2023).

[52] 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? Psychol. Sci. **26**, 1531 (2015).

[53] P. Ramaciotti Morales and Z. Vagena, Embedding social graphs from multiple national settings in common empirical opinion spaces, in *Proceedings of the 2022 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining* (IEEE, 2022).

[54] P. Ramaciotti Morales and G. Muñoz Zolotoochin, Measuring the accuracy of social network ideological embeddings using language models, in *Information Technology and Systems: Proceedings of ICITS 2022* (Springer, 2022), pp. 267–276.

[55] P. Ramaciotti Morales, M. Berriche, and J.-P. Cointet, The geometry of misinformation: embedding twitter networks of users who spread fake news in geometrical opinion spaces, in *Proceedings of the International AAAI Conference on Web and Social Media*, Vol. 17 (AAAI, 2023), pp. 730–741.

[56] T. P. Peixoto, The GRAPH-TOOL PYTHON library, Figshare (2014), doi:10.6084/m9.figshare.1164194.