Auditing the Effect of Social Network Recommendations on Polarization in Geometrical Ideological Spaces

Pedro Ramaciotti Morales, Jean-Philippe Cointet

Auditing the Effect of Social Network Recommendations on Polarization in Geometrical Ideological Spaces

Pedro Ramaciotti Morales, Jean-Philippe Cointet

Auditing the Effect of Social Network Recommendations on Polarization in Geometrical Ideological Spaces

Pedro Ramaciotti Morales
Jean-Philippe Cointet
pedro.ramaciottimorales@sciencespo.fr
jeanphilippe.cointet@sciencespo.fr
médialab, Sciences Po
Paris, France

ABSTRACT

The prevalence of algorithmic recommendations has raised public concern about undesired societal effects. A central threat is the risk of polarization, which is difficult to conceptualize and to measure, making it difficult to assess the role of Recommender Systems in this phenomenon. These difficulties have yielded two types of analyses: 1) purely topological approaches that study how recommenders isolate or connect types of nodes in a graph, and 2) spatial opinion approaches that study how recommenders change the distribution of users on a given opinion scale. The former analyses prove inadequate in settings where users are not classified into categorical types (e.g., in two-party systems with binary social divides), while the latter rely on synthetic data due to the unobservability of opinions. To overcome both difficulties we present the first analysis of friend recommendations acting on real-world sub-graphs of the Twitter network where users are embedded in multidimensional ideological spaces and in which dimensions are indicators of attitudes towards issues in the public debate. We present a polarization metric adapted to these dual topological and spatial states of social network, and use it to track both the evolution of polarization on Twitter networks where the graph evolves following well-known Recommender Systems, and opinions co-evolve following a De-Groot opinion model. We show that different recommendation principles can sometimes drive or mitigate polarization appearing in real social networks.

CCS CONCEPTS

• **Information systems → Recommender systems**; • **Computing methodologies → Simulation evaluation**.

KEYWORDS

Consequences of recommendation, network scaling, ideology, bias, polarization, polarization metrics.

1 INTRODUCTION

Social or political polarization has been an object of research and debate for several decades. Some researchers have suggested that partisanship and polarization of the political landscape in the US is on the rise [16, 22], while others have argued that it is rather the news media that is polarized [10, 15, 31]. Observers and scholars have also questioned the responsibility of social media, in particular their recommendation algorithms, whose systemic might results in radicalism or filter bubbles [25, 28, 30] which may eventually threaten democracy at large [18, 32]. However, we still lack observational data and rigorous metrics to empirically audit the role of recommendation algorithms in real-world settings [29]. This article proposes to do so operationalizing and measuring opinions in geometrical *ideological* spaces.

The study of polarization in digital social networks, and the role of Recommender Systems (RS), is two decades of works in the making and comes with mixed results: definitions largely vary. Some studies have conceptualized polarization as a property of states of social systems described in topological terms, where polarization is related to the phenomenon commonly referred to as *echo chamber* [1], in which users organize in dense graph clusters guided by affinity. Other works have conceptualized polarization as a property in which users have opinions described by spatial models, where users are placed on attitudinal scales indicating their positive or negative stances towards an issue [8] and polarization manifests as a concentration of users around different opinion positions (or attractor poles) in space. Both of these approaches (topological and geometrical) have been influenced by the political setting of the US of binary social classification of users as Democrat- or Republican-leaning [3], and user representation in a single (one-dimensional) attitudinal scale ranging from liberal to conservative views [7].

The prevalence of RS in social platforms has raised public concern about undesired societal effects [23], with polarization becoming a notion of interest. Difficulties in conceptualizing and measuring polarization have not resulted in a widely accepted metric or definition, let alone means for measuring the role of recommendation in this phenomenon. On the one hand, purely topological measures [5, 17] neglect the importance of individual attitudes in driving collective behavior [2], and rely on categorical classifications of users which are hardly applicable outside the scope of the US setting. On the other hand, attitudes of users are not directly observable, confining studies of RS in spatial opinion models to

theoretical results [11] and simulations in synthetic societies [6, 24]. Existing attitudinal estimation methods for social networks have only produced reliable results in one-dimensional spatial representations (ranging from liberal to conservative) in the US [9], and have not yet been used to study the effect of RS in opinions and in polarization.

We adopt the view that polarization is a property of states of social systems at every moment and we study the impact RS in polarization using real-world sub-graphs of the Twitter network. To overcome the aforementioned difficulties we propose the first use of multidimensional ideological embedding of social networks in the study of the evolution networks subjected to RS and the co-evolution of the opinion of users. In this model, network nodes have dynamical positions in a geometrical ideological space where dimensions stand as indicators of negative or positive attitudes towards a set of different issues. We present a polarization measure suited for this dual topological-geometrical representation of social networks, and we use it to study the evolution of polarization as the network and opinions co-evolve following well-known RS and DeGroot's opinion model [12]. It has been shown theoretically that these systems either converge in opinion to a single position in space, or diverge depending on whether users are influenced by neighbors in a biased fashion [11]. We show this empirically using real-world data for the first time, and we show that different well-known RS can disturb these regimes of convergence or divergence of such systems, affecting the evolution of polarization. In particular, we show that if the social networks are allowed to change following some RS, polarization can be limited even in settings with biased assimilation, in which increasing polarization was theoretically assured for static networks.

2 PRELIMINARIES

2.1 Directed social networks, recommendations, and the evolution of the social graph

We consider directed social networks $G = (V, E, \theta)$ embedded in an ideological space of dimensions D, where $\theta \in \mathbb{R}^{|V| \times |D|}$ is the feature matrix providing ideological position of nodes V along dimensions D. $\theta_{i,d}$ is the ideological position of user i on dimension $d \in D$. In directed social networks, edges are asymmetric channels of information (*e.g.*, *following* on Twitter) such that, if $(i, j) \in E$, information can flow from j to i and not from i to j: i is *influenced* by j as it is exposed to j's activity in its Twitter stream.

We consider changes to the social graph resulting from the adoption of friends recommended by a RS over different steps t in discrete time. Given a set of edges $E(t)$ at an instant t, we consider an operator R that (1) adds one new outward edge for each node (linking them to new friends) according to some RS, and (2) that deletes one old friendship (edges). The computation of new friends to be followed may depend on the topology of the network $E(t)$ and on the ideological position of users $\theta(t)$: $E(t + 1) = R(E(t), \theta(t))$.

We consider six different RS relying on different principles. Three are well-known RS from the state of the art: Alternating Least Squares (ALS) [19], Bayesian Personalized Ranking (BPR) [27], and

Logistic Matrix Regression (LMR) [20]. We will also consider a "random" RS (suggesting random nodes as new friends) and a nearest-neighbor (NN) RS (suggesting to each node the closest neighbor in ideological space using Euclidean distance). All RS recommend exclusively friends that are not already followed: if node i follows node j at t, a RS will not recommend j to i for $t + 1$. Finally, we will also consider a "static" RS, which does not change the social graph in time: *i.e.*, $E(t + 1) = E(t)$.

2.2 Co-evolution of ideological positions

Together with changes in the social graph, we consider changes in the ideological position of users at each step t following an asymmetric process in influence. We model this change in ideological positions due to influence of followed friends with DeGroot's well-known model of opinion formation [12]. We consider the variation proposed by Dandekar et al. [11, Eq 3] that includes a "biased assimilation" [21] parameter $b \geq 0$ with which users update their positions averaging the positions of their neighbors (followed friends), but weighting their influence differently depending on their own position. Building on [11], we formulate this update process for each user $i \in V$ along each dimension $d \in D$:

$$\theta_{i,d}(t+1) = (1-\alpha)\theta_{i,d}(t) + \alpha \left(\frac{\theta_{i,d}(t) + (\sup_i)^b s_{i,d}}{1 + (\sup_i)^b s_{i,d} + (1-\sup_i)^b \left(|\mathcal{N}(i)| - s_{i,d}(t) \right)} \right), \quad (1)$$

where $\alpha \in [0, 1]$ is a parameter that controls the importance of the previous position of user i and its neighbors in updating its position in the next step ($\alpha = 0$ gives no importance to positions of neighbors). $\mathcal{N}(i)$ is the set of neighbors of i at step t, and $s_{i,d}(t) = \sum_{j \in \mathcal{N}(i)} \theta_{j,d}(t)$ is the sum of the positions of neighbors along dimension d. $\sup_i = (\theta_{i,d} - \ell_{min,d})/(\ell_{max,d} - \ell_{min,d})$ is the degree of support of i for the most extreme position in dimension d, with $\ell_{min/max,d} = min/max_{i \in V} \theta_{i,d}(0)$. Parameter b controls the bias of the assimilation of the opinions of neighbors. $b = 0$ results in users averaging their own opinion with that of all of their friends: $\theta_{i,d}(t+1) = (1-\alpha)\theta_{i,d}(t) + \alpha \left(\theta_{i,d}(t) + s_{i,d}(t) \right)/(1 + |\mathcal{N}(i)|)$. We denote the update process from Eq. 1 as operator U: $\theta_{i,d}(t+1) = U \left(V, E(t), \theta(t)_{i,d}, \alpha, b \right)$. Dandekar et al. [11] showed that, for static networks (*i.e.*, $E(t + 1) = E(t)$), $b = 0$ results in a process where spatial opinion dispersion of users decreases monotonically on each dimension $d \in D$. This means that the positions of users converge to a single attractor pole in space.

Next, we are interested in computing the co-evolution of the social graph (resulting from the adoption of recommendations) and the ideological position of users (resulting from the evolving network), starting from a real-world dataset. Using the previously-defined friend recommendation and ideological update procedures, we specify a simple co-evolution procedure of T_{max} steps as Algorithm 1.

3 MEASURING THE POLARIZATION OF STATES OF THE SYSTEM

To address the shortcomings of the existing polarization measures, we propose to apply the Duclos-Esteban-Ray (DER) polarization measure [13] to distributions of users in ideological space. The

---

Algorithm 1: Co-evolution of social graph and ideological positions of users.

Set parameters: α, b;
Load initial network: $V, E(0), \theta(0)$;
for $t = 0, 1, \ldots, (T_{max} - 1)$ **do**
$\quad | \quad E(t + 1) = R(E(t), \theta(t))$;
$\quad | \quad \theta(t + 1) = U(V, E(t), \theta(t), \alpha, b)$;
end

DER measure is derived in an axiomatic theory of measures over 1-dimensional distribution functions. The axioms of this theory assure that the measure of polarization $P(\theta)$ of a distribution θ increases with: 1) the number of attractors in multimodal distributions, 2) the concentration of mass around attractors, and 3) the distance between attractors. Given a distribution θ defined over a 1-dimension ideological scale d, the imposition of these properties results in a parametric family of polarization measures [13, Theorem 1]:

$$P_\gamma(\theta) = \iint_{x \in d, x' \in d} \theta(x)^{1+\gamma} \theta(x')|x - x'| \mathrm{d}x \mathrm{d}x', \quad \text{for } \gamma \in [1/4, 1]. \quad (2)$$

P_γ has deep connections to diversity measures, with P_0 being the Gini coefficient [14]. We arbitrarily set $\gamma = 1$ and note P_γ simply as P (see [13, Section 3.2] for a sensibility analysis).

Before using this measure to study polarization of our online social network data, we propose two experiments using synthetic data to illustrate its use. In a first experiment, we obtain six different 1-dimensional distributions θ using a 2-component Gaussian Mixture Model with 1.000 samples, and that illustrates the ability of P for capturing the three properties underlying the axiomatic construction of the measure (see Fig 1). In the second experiment, we create a synthetic network embedded in an arbitrary 2-dimensional ideological space ($D = \{d_1, d_2\}$) and we track the changes in polarization as ideological positions evolve without changing the social graph (*i.e.*, using a "static" RS). For ease of notation we denote $\theta(t)$ as θ_t. To create the random synthetic network, first we attribute ideological positions $\theta_0 \in \mathbb{R}^{1.000 \times 2}$ to 1.000 nodes following a 2-components, 2-dimensional Gaussian Mixture Model with means $\mu_1 = (1.5, 1.5)$ and $\mu_2 = (-1.5, -1.5)$, and standard deviations $\sigma_{1,2} = (1, 1)$ (*i.e.*, 1 for each component on each dimension). Once that nodes have been positioned in space (*i.e.*, θ_0 has been set), we generate a random graph following logistic probability law $P[(i, j) \in E \mid \theta] = \mathrm{logistic}(-9.21(\|\theta_i - \theta_j\| - 0.5))$, based on the Euclidean distances between nodes in two dimensions. We then compute 5 steps of Algorithm 1 with $\alpha = 0.5$, for $b = 0$ and for $b = 20$ (see Fig. 2). We set these values arbitrarily to be used throughout this article, as tests with different alternatives did not change the substance of the conclusions. In accordance with the theoretical results of [11], $b = 0$ does not increase polarization in either dimension (even despite the imposed multimodality), while $b = 20$ does increases polarization.

4 TWITTER NETWORK IDEOLOGICAL EMBEDDING DATA

We now present real-world data on which we will evaluate polarization and the effects of different RS. We consider a Twitter sub-graph [1] in the vicinity of French parliamentarians (MPs) active on Twitter, which allows for the embedding of users in spaces where dimensions stand for indicators of attitudes. The details of this embedding procedure and its benchmarking are available in [26]. In this section we provide a brief description of the data and its quality.

We consider 831[2] (out of 925) French MPs on Twitter affiliated to 10 political parties, and we collect their followers (4.487.430 by May 2019). We follow [7] and consider only users that follow at least 3 MPs, and that have at least 25 followers (to filter bots and inactive accounts). We also removed users with repeated sets of followed MPs to obtain 368.831 followers ensuring a full-rank adjacency matrix. Following [8] we embed this bipartite network of MPs and followers in a latent 2-dimensional space, where users are positioned according to homophily (users close in space have higher probability of being connected to the same MPs). We establish referential points in space to inspect how dimensions relate to attitudinal indicators available in political surveys. For each party, we compute the spatial position as the centroid of its MPs (see Fig. 3, left subfigure). We then compared the position of parties with party positions in more than 40 attitudinal scales provided by the Chapel Hill Expert Survey (CHES) [4]. We assess the relation between position in these two different sources with a Pearson correlation, and we retain the two CHES attitudinal dimensions that are most correlated with the dimensions of our latent space. One latent dimension is related to left-right ideological and economic cleavages: we call it the "Left-Right" (LR) axis. The second latent dimension is related to attitudes towards trade protectionism and European integration: we call it the "Local-Global" (LG) axis. Next, we collect the Twitter sub-graph spanned by these users, eliminating users than had disabled permissions to have their followers collected, and users that were not connected to any other node in this sub-graph. This resulted in our empirical network composed of 230.911 users, and 67.217.556 edges (density=0.00126) where all users are positioned in our LR-LG ideological space.

5 RESULTS

5.1 Evolution of opinion and polarization in the absence of recommendations

Before investigating the effect of different RS in the evolution of the ideological positions of users, we examine the evolution of positions when the network remains static (*i.e.*, $E(t + 1) = E(t)$). We compute Algorithm 1, once for $b = 0$ (without biased assimilation), and once for $b = 20$ (with biased assimilation). For both values of b we compute 5 steps and trace the evolution of the polarization on each spatial dimension: LR and LG (see Fig. 4). As anticipated by [11], $b = 0$ produces a reduction of the spatial dispersion of
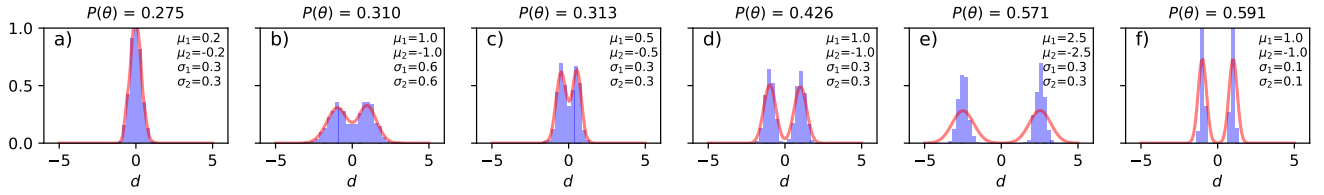

Figure 1: Illustration of the DER polarization measure P on different 1-dimensional distributions θ resulting from a 2-component Gaussian Mixture Model with 6 different sets of parameters and 1.000 random draws. The DER measures captures multimodality, distance between attractors in space, and concentration around attractors.
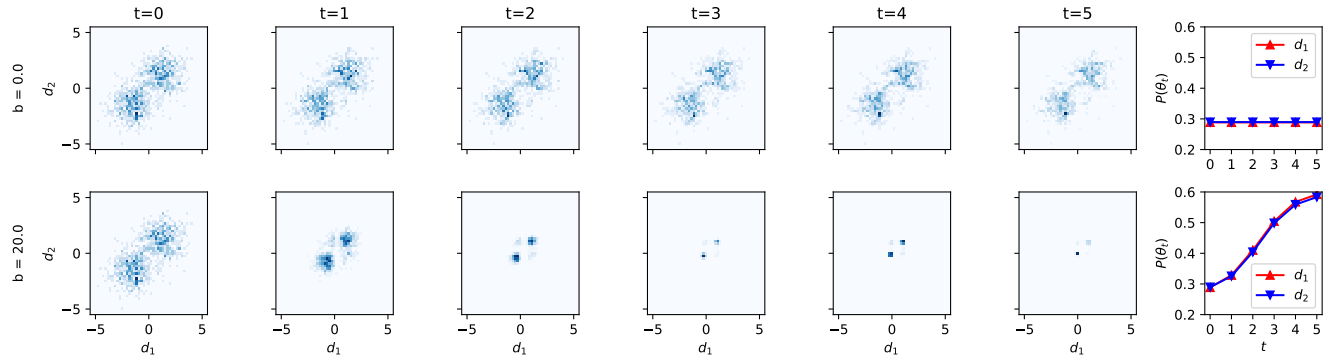

Figure 2: Evolution of the ideological positions of 1.000 users of a synthetic social network for biased assimilation $b = 0$ and $b = 20$. Starting positions follow a 2-component Gaussian Mixture Model, and edges are created with a logistic probability model based on ideological distance. Network is static (*i.e.*, no RS is applied) computing 5 steps of Algorithm 1.
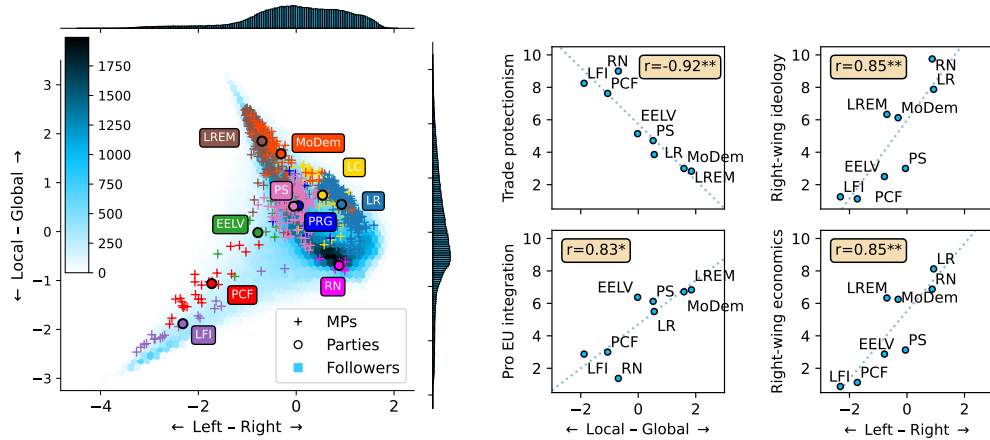

Figure 3: 2D histogram of the French Twitter network (230.911 nodes, 67.217.556 edges): MPs colored by political party with party centroids (left). Comparison of the positions of the parties in Left-Right and Local-Global axes, and the attitudinal positions towards Left-Right cleavages, trade protectionism, and EU integration given by the Chapel Hill Expert Survey (p-values: <0.01**, <0.05*). Source: Ramaciotti Morales et al. 2021 [26].

users, which in turn assures non-increasing polarization values, while $b = 20$ produces the convergence of users towards separated attractors, which is captured as increasing polarization by measure P on both dimensions. At $t = 0$ polarization on both dimensions is similar, showing small increases in steps for $b = 0$ and large increases for $b = 20$, also on both dimensions.

5.2 Evolution of polarization in recommendation regimes changing social networks

We now turn to the analysis of the effect of allowing the network to change as described in Section 2.1. We do this by computing recommendations on sub-graphs, training the RS when needed (for ALS, BPR, and LMR). We first sample sub-graphs of random 2.500 users and compute the co-evolution of the sub-graph and the ideological position of users over 5 steps. As specified in Section 2.1, at each step we compute one recommendation for each user using the sub-graph and deleting randomly a previously existing edge. For values 0 and 20 for b, and for each RS, we sample 50 sub-graphs and compute this co-evolution. Fig. 5 reports the results, showing the median polarization value, and the 10%-90% range, for the 50 computations, and comparing them to the baseline situation in which the network is kept static (i.e., $E(t + 1) = E(t)$). Even for $b = 0$, some RS can decrease polarization (random and LMR), while other can increase it (ALS and BPR), although slightly, suggesting that some RS might be leveraging spatial representations of ideologies and suggesting friends in ways that amount to biased assimilation. On the contrary, for $b = 20$, all of our tested RS resulted in polarization increments that are comparatively smaller to those observed in the absence of RS changing the network structure. This suggest that RS often expose users to dissimilar friends producing more socially diverse connections than would be otherwise achieved with high levels of bias on the part of users.

6 DISCUSSION AND CONCLUSIONS

Our work shows that it is possible to bridge the traditional gap between (data driven) social network analysis and (simulation driven) opinion dynamics studies in the analysis of the effects of recommendations in social networks by including empirical opinion spatial models. We further show how to overcome some of the shortcomings of traditional measures of polarization (based on spatial dispersion) by the introduction of an axiomatic measure capturing relevant properties of spatial distribution, lacking in previously used metrics in social network analysis. We use real-world empirical networks equipped with ideological positions, and well-known diffusion dynamics averaging opinion for which distinct polarizing and non-polarizing regimes have been proven to exist, and we show that RS can sometimes break these regimes. Well-known RS can produce behaviors that amount in their effect to biased assimilation on the part of users, while can sometimes produce the opposite effect, de-biasing connections and influence in social networks, decreasing polarization. Our geometrical analysis of RS, coupled with measures that capture important properties of polarization, offers a new path into the investigation of their effects in social systems.

REFERENCES

[1] Lada A Adamic and Natalie Glance. 2005. The political blogosphere and the 2004 US election: divided they blog. In *Proceedings of the 3rd international workshop on Link discovery*. 36–43.

[2] Icek Ajzen. 1989. Attitude structure and behavior. *Attitude structure and function* 241 (1989), 274.

[3] Christopher A Bail, Lisa P Argyle, Taylor W Brown, John P Bumpus, Haohan Chen, MB Fallin Hunzaker, Jaemin Lee, Marcus Mann, Friedolin Merhout, and Alexander Volfovsky. 2018. Exposure to opposing views on social media can increase political polarization. *Proceedings of the National Academy of Sciences* 115, 37 (2018), 9216–9221.

[4] R. Bakker, L. Hooghe, S. Jolly, G. Marks, J. Polk, J. Rovny, M. Steenbergen, and M. A. Vachudova. 2020. 2019 Chapel Hill Expert Survey. *Chapel Hill* (2020). www.chesdata.eu.

[5] Eytan Bakshy, Solomon Messing, and Lada A Adamic. 2015. Exposure to ideologically diverse news and opinion on Facebook. *Science* 348, 6239 (2015), 1130–1132.

[6] Sven Banisch and Eckehard Olbrich. 2019. Opinion polarization by learning from social feedback. *The Journal of Mathematical Sociology* 43, 2 (2019), 76–103.

[7] Pablo Barberá. 2015. Birds of the same feather tweet together: Bayesian ideal point estimation using Twitter data. *Political analysis* 23, 1 (2015), 76–91.

[8] Pablo Barberá, John T Jost, Jonathan Nagler, Joshua A Tucker, and Richard Bonneau. 2015. Tweeting from left to right: Is online political communication more than an echo chamber? *Psychological science* 26, 10 (2015), 1531–1542.

[9] Pablo Barberá and Gonzalo Rivero. 2015. Understanding the political representativeness of Twitter users. *Social Science Computer Review* 33, 6 (2015), 712–729.

[10] Yochai Benkler, Robert Faris, and Hal Roberts. 2018. *Network propaganda: Manipulation, disinformation, and radicalization in American politics*. Oxford University Press.

[11] Pranav Dandekar, Ashish Goel, and David T Lee. 2013. Biased assimilation, homophily, and the dynamics of polarization. *Proceedings of the National Academy of Sciences* 110, 15 (2013), 5791–5796.

[12] Morris H DeGroot. 1974. Reaching a consensus. *J. Amer. Statist. Assoc.* 69, 345 (1974), 118–121.

[13] Jean-Yves Duclos, Joan Esteban, and Debraj Ray. 2004. Polarization: concepts, measurement, estimation. *Econometrica* 72, 6 (2004), 1737–1772.

[14] Joan Esteban and Debraj Ray. 2012. Comparing polarization measures. *Oxford handbook of economics of peace and conflict* (2012), 127–151.

[15] Morris P Fiorina, Samuel J Abrams, and Jeremy C Pope. 2005. Culture war. *The myth of a polarized America* 3 (2005).

[16] Matthew Gentzkow, Jesse M Shapiro, and Matt Taddy. 2019. Measuring group differences in high-dimensional choices: method and application to congressional speech. *Econometrica* 87, 4 (2019), 1307–1340.

[17] Pedro Calais Guerra, Wagner Meira Jr, Claire Cardie, and Robert Kleinberg. 2013. A measure of polarization on social media networks based on community boundaries. In *Seventh international AAAI conference on weblogs and social media*.

[18] Dirk Helbing, Bruno S Frey, Gerd Gigerenzer, Ernst Hafen, Michael Hagner, Yvonne Hofstetter, Jeroen Van Den Hoven, Roberto V Zicari, and Andrej Zwitter. 2019. Will democracy survive big data and artificial intelligence? In *Towards digital enlightenment*. Springer, 73–98.

[19] Yifan Hu, Yehuda Koren, and Chris Volinsky. 2008. Collaborative filtering for implicit feedback datasets. In *2008 Eighth IEEE International Conference on Data Mining*. IEEE, 263–272.

[20] Christopher C Johnson. 2014. Logistic matrix factorization for implicit feedback data. *Advances in Neural Information Processing Systems* 27, 78 (2014), 1–9.

[21] Charles G Lord, Lee Ross, and Mark R Lepper. 1979. Biased assimilation and attitude polarization: The effects of prior theories on subsequently considered evidence. *Journal of personality and social psychology* 37, 11 (1979), 2098.

[22] Nolan McCarty, Keith T Poole, and Howard Rosenthal. 2016. *Polarized America: The dance of ideology and unequal riches*. mit Press.

[23] Silvia Milano, Mariarosaria Taddeo, and Luciano Floridi. 2020. Recommender systems and their ethical challenges. *AI & SOCIETY* 35, 4 (2020), 957–967.

[24] Cameron Musco, Christopher Musco, and Charalampos E Tsourakakis. 2018. Minimizing polarization and disagreement in social networks. In *Proceedings of the 2018 World Wide Web Conference*. 369–378.

[25] Safiya Umoja Noble. 2018. *Algorithms of oppression*. New York University Press.

[26] Pedro Ramaciotti Morales, Jean-Philippe Cointet, and Gabriel Muñoz Zolotoochin. 2021. Unfolding the dimensionality structure of social networks in ideological embeddings. (Aug 2021). https://hal.archives-ouvertes.fr/hal-03315759

[27] Steffen Rendle, Christoph Freudenthaler, Zeno Gantner, and Lars Schmidt-Thieme. 2012. BPR: Bayesian personalized ranking from implicit feedback. *arXiv preprint*
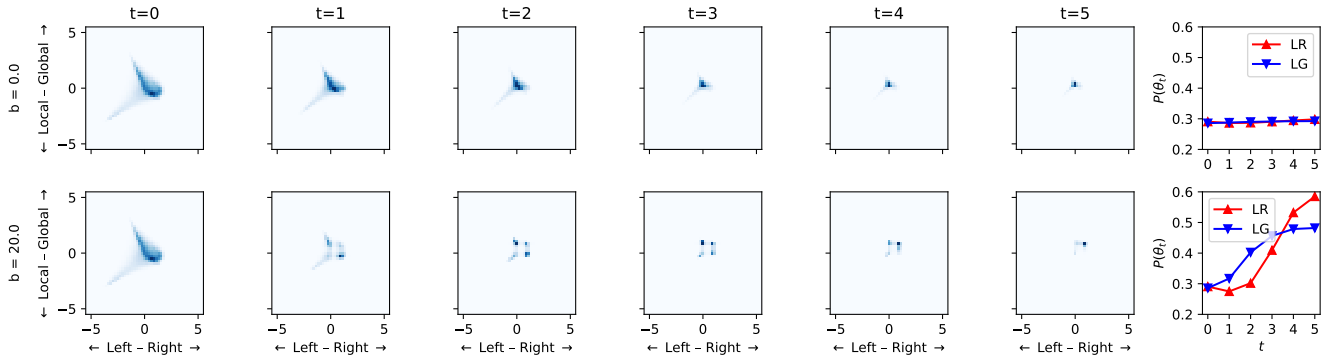
Figure 4: Evolution of ideological positions and polarization of the Twitter network on both dimensions, with a static social graph (*i.e.*, $E(t + 1) = E(t)$), and with ($b = 20$) and without ($b = 0$) biased assimilation.
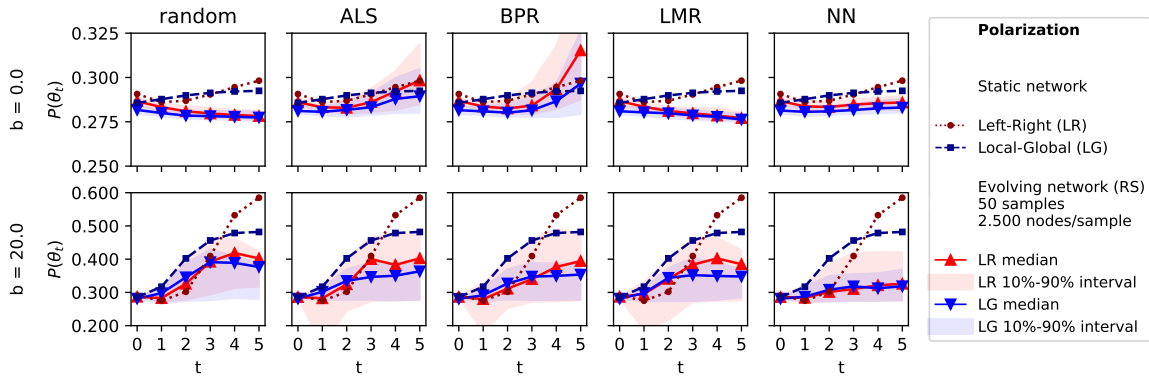

Figure 5: Evolution of polarization in the Twitter network with co-evolution of ideological positions and network structure as different RS suggest new friends, compared with polarization evolving in a static network. Recommendations are computed and proposed for a random sample of the network, reporting median and 10%-90% polarization range in comparison with values for a static network.

arXiv:1205.2618 (2012).

[28] Cass Sunstein and Cass R Sunstein. 2018. *# Republic*. Princeton university press.

[29] Andreas Tsamados, Nikita Aggarwal, Josh Cowls, Jessica Morley, Huw Roberts, Mariarosaria Taddeo, and Luciano Floridi. 2021. The ethics of algorithms: key problems and solutions. *AI & SOCIETY* (2021), 1–16.

[30] Zeynep Tufekci. 2018. YouTube, the great radicalizer. *The New York Times* 10 (2018), 2018.

[31] Alan Wolfe. 1998. One Nation After All: What Americans Really Think About God. *Country, Family, Racism* (1998).

[32] Shoshana Zuboff. 2019. *The age of surveillance capitalism: The fight for a human future at the new frontier of power*. Profile books.