Quantifying Ideological Polarization on a Network Using Generalized Euclidean Distance

Marilena Hohmann,[1,†] Karel Devriendt,[2,3,†] Michele Coscia[4,*,†]

[1]Copenhagen Center for Social Data Science, University of Copenhagen, Øster Farimagsgade 5, Copenhagen, DK
[2]Mathematical Institute, University of Oxford, Woodstock Road, Oxford, UK
[3]Alan Turing Institute, Euston Road 96, London, UK
[4]CS Department, IT University of Copenhagen, Rued Langgaards Vej 7, Copenhagen, DK

[*]To whom correspondence should be addressed; E-mail: mcos@itu.dk.
[†]Author contributed equally to the work.

Teaser. A measure for estimating ideological divergence in social networks allows to study polarization.

An intensely debated topic is whether political polarization on social media is on the rise. We can investigate this question only if we can quantify polarization, by taking into account how extreme the opinions of the people are, how much they organize into echo chambers, and how these echo chambers organize in the network. Current polarization estimates are insensitive to at least one of these factors: they cannot conclusively clarify the opening question. Here, we propose a measure of ideological polarization which can capture the factors we listed. The measure is based on the Generalized Euclidean (GE)

distance, which estimates the distance between two vectors on a network, e.g., representing people's opinion. This measure can fill the methodological gap left by the state of the art, and leads to useful insights when applied to real-world debates happening on social media and to data from the US Congress.

Introduction

Despite a multitude of studies of polarization on social media (*1–3*), it remains disputed whether political polarization in digital public spaces is on the rise. Many analyses conclude that polarization is rapidly advancing (*4–8*), while others question this interpretation (*9, 10*). The timeliness and relevance of this issue warrants a closer look at how these claims are made, and raises an important question: how can we accurately quantify the level of polarization of a social system?

The political science literature commonly distinguishes between two types of polarization: ideological and affective polarization (*11, 12*). Ideological polarization refers to increasing ideological divergence and a reduced dialogue among individuals with differing views (*13–15*). Affective polarization describes in-group favoritism and out-group hostility, and it is thus concerned with the affective attitude towards others depending on their opinions (*12, 13, 16*). Although the two types of polarization can be mutually reinforcing (*17–19*), ideological polarization and affective polarization are distinct concepts, both in terms of theory as well as empirical measurement (*11, 12*). While the measure of ideological polarization relies on data about the opinions of people, affective polarization also requires information about the valence of their relationships (*20–22*).

In this paper, we focus on ideological polarization in social networks – hereafter, whenever we do not qualify the term "polarization" we refer to ideological polarization. As outlined above, ideological polarization refers to increasing ideological divergence on the one hand, and

increasing reluctance to engage with diverging views on the other hand (*13–15, 23*). From this conceptual understanding, we derive two components of ideological polarization in social networks and an interplay between the two: a social network is more polarized than another if the opinions of its members diverge strongly (opinion component), if people with similar opinions cluster with each other in communities (structural component), and if these communities tend to organize themselves in an ideological spectrum rather than engaging with all other communities (mesoscale interplay of opinion and structure).

Since current network-based measures can only partially capture these components, we propose a new measure of ideological polarization. Our measure is based on a Generalized Euclidean (GE) distance measure (*24*), and it estimates how much effort it would take to travel from one opinion to another in the network.

The literature has advanced numerous ways to estimate ideological polarization which we briefly review here. Some methods consider exclusively opinions (*25*) or reduce the complexity of the structure (*26*), which we think does not allow to properly capture what we understand as polarization. Other approaches rely on local network measures to evaluate the structure of interactions in a network (*27–29*). The assortativity coefficient, for instance, quantifies to what extent individuals are directly linked to like-minded others (*30–32*). Similarly, there are methods which assess the average opinion of the direct neighbors of an individual (*33, 34*). However, local measures are myopic to the overall structure, and would return the same estimations even if opinions are distributed in radically different mesoscale structures such as communities (*35*).

Alternative methods explicitly divide the network into two communities to determine how well they are separated from each other (*36–41*). While these measures can account for the network structure, a two-community partition implies an expectation of polarization that might not exist. Methods that avoid the partition phase (*42*) provide node-dependent estimates, and it is unclear how to summarize them for the whole network.

Another approach builds on the opinion formation model proposed by Friedkin and Johnsen (*43*). This measure assumes that each individual has both an internal opinion, and an expressed opinion which is determined by their own internal opinion as well as the surrounding opinions in the social network (*44, 45*). However, it is questionable whether individuals have stable internal opinions on a political issue since public opinion research suggests that individual-level issue opinions are often inconsistent and volatile (*46–48*). Apart from these conceptual considerations, this approach entails a practical problem: social media data can only capture people's expressed opinions, but not their internal opinions on an issue (*49*). Our approach sidesteps this issue by not requiring to know the internal opinion of an individual.

Finally, one could use graph neural networks (*50*), but these techniques normally provide a simple classification of *whether* a structure is polarized rather than *quantifying* the polarization level.

Our approach overcomes the aforementioned issues by using data available on social media: the people's expressed opinions and their social relationships. The former is determined based on social media users' sharing behavior (*34, 51*), the latter by downloading their connections such as, e.g., follower relationships on Twitter. We estimate the Generalized Euclidean distance (*24, 52*) between two opposing opinions across all the edges of the network. By doing so, we avoid using a local approach and we do not assume a community structure by default.

In the results section, we demonstrate how our approach is the only alternative we found that is sensitive to the two components of ideological polarization outlined above, as well as their interplay. Moreover, we show that our measure allows us to make useful inferences about real-world polarized systems such as political debates on Twitter or voting patterns in the US House of Representatives.

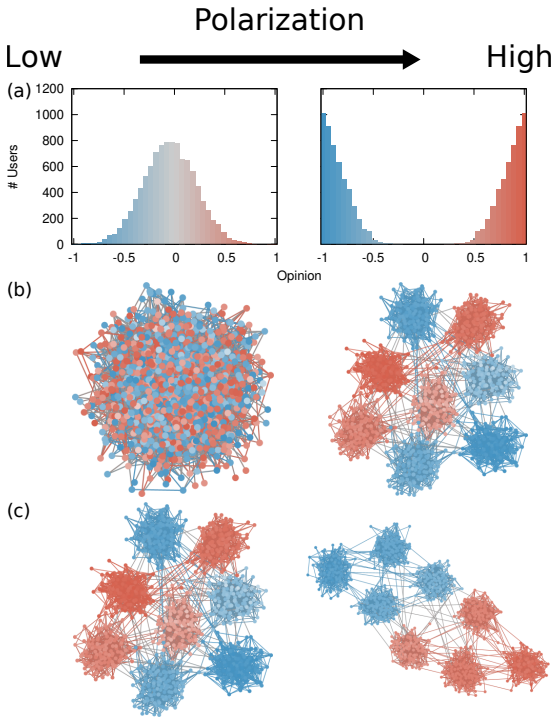

Figure 1: **The two components of polarization in a network and their interplay.** (a) Opinion component: plots show the number of people (y axis) with a given opinion value (x axis and bar color). (b) Structural component: people are nodes, connected if they are interacting with each other. Node color represents the opinion (< 0 blue, > 0 red), edge color the average opinion of the two connected nodes. One community on the left, eight communities on the right. (c) Opinion-structural interplay: same legend as (b). All communities equally interconnected on the left, each community only connected to its most similar opinion community on the right.

Methods

Definition

Based on the existing literature on ideological polarization, we define two components of polarization and their interplay (Figure 1):

- Opinion component (Figure 1(a)): Traditional political science studies of ideological polarization consider if and how people's ideological leanings diverge (*11, 14, 15*): if opinions cluster in the moderate center, polarization is low (example on the left). If they

instead disperse towards the extremes, polarization is high (example on the right).

- Structural component (Figure 1(b)): More recent approaches have emphasized the role of social connections, and especially homophily, i.e., the connections between like-minded individuals (*23, 53–56*). If there is no community structure, then there is no opinion homophily and each individual is connected and therefore exposed to many different views. In this case, polarization is low (example on the left). However, if there are clearly separated communities, individuals are only exposed to the opinions within their community, but they are not directly exposed to other opinions and polarization is therefore high (example on the right).

- Mesolevel organization of the opinion-structural interplay (Figure 1(c)): Opinion and network structure have largely been viewed as separate indicators of polarization. To integrate the two strands of the literature into a unified definition, we propose to also consider the interplay between the two components. We understand this interplay as follows: the same opinions and the same communities can give rise to different levels of polarization depending on the mesolevel organization of the system. Communities that can freely interlink regardless of their opinions indicate a lower level of polarization (example on the left) than if communities organize in progressively more extreme echo chambers (example on the right).

Since ideological polarization has previously been described both in terms of opinions and network structure, we consider these two components to be distinct, yet related aspects of polarization. This view is supported by some of the real-world examples we examine as we can show that opinion and structure are correlated in the Twitter networks we analyze (see Supplementary Materials Section 8). We therefore argue that it is important for a measure of polarization to consistently capture the two components and their interplay in a single measure.

Formulation

We refer to our polarization measure as $\delta_{G,o}$. The measure requires two inputs: the graph structure G and a vector of opinions o. $\delta_{G,o}$ takes values from 0, which implies no polarization at all, to an arbitrary positive number. The higher the value, the more polarized the network is.

The first parameter is a simple graph $G = (V, E)$, with V being the set of nodes and $E \subseteq V \times V$ the set of connections, i.e., node pairs (i, j) with $i, j \in V$. For simplicity, we assume edges in E to be unweighted and undirected ($(i, j) = (j, i)$), but our approach can consider edge weights. There are a few mandatory requirements on G. G must not contain self-loops – edges connecting a node with itself. It must also be connected, i.e., there must be at least one path between any two nodes in the graph. The polarization $\delta_{G,o}$ cannot be estimated if these conditions are not satisfied.

The second parameter is the vector of opinions o. This vector o must have length $|V|$, i.e., record a single value per node. We impose a convention on o: the opinion values must be bounded between -1 (the most extreme opinion on one side) and $+1$ (the most extreme opinion on the other side). In such a vector, 0 represents perfect neutrality between the two opinions. In real-world US politics data, -1 could be an extreme Democrat and $+1$ an extreme Republican, with 0 as perfect independents.

$\delta_{G,o}$ is based on a solution (*24*) to the node vector distance problem (*57*). In GE, one can use the pseudoinverse Laplacian to estimate the effective resistance (*58*) between two arbitrary vectors of length $|V|$ recording a variable per node of the network. We recall that the Laplacian matrix is $L = D - A$, with A being the adjacency matrix of G and D being the diagonal matrix containing the degrees of the nodes of G. Thus:

$$L_{ij} = \begin{cases} d_i \text{ if } i = j \\ -1 \text{ if } (i, j) \in E \\ 0 \text{ otherwise.} \end{cases}$$

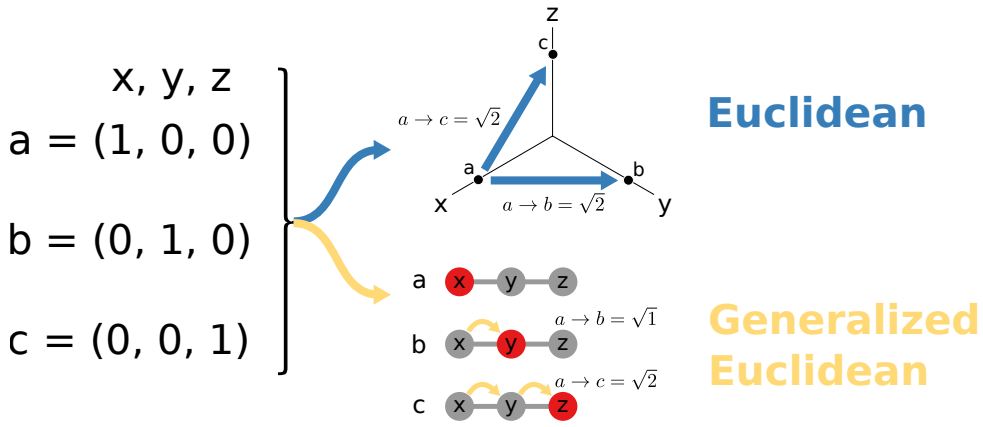

Figure 2: **Difference between Generalized and plain Euclidean.** We start from three vectors a, b, c in a 3D space x, y, z. The blue arrow points at an Euclidean space, with labeled dots showing the positions of vectors a, b, c in independent dimensions x, y, z. The yellow arrow points to an x, y, z space defined on graph G, and the node color represents the values for vectors a, b, c (red equals to 1, gray equals to zero).

To estimate the effective resistance we need to invert L, but L is singular and therefore cannot be inverted. For this reason, we take the Moore-Penrose pseudoinverse of L, symbolized as $L^{\dagger}$. Then, for two arbitrary node vectors a and b:

$$\delta_{G,a,b} = \sqrt{(a-b)^T L^{\dagger}(a-b)}.$$

Previous work shows that this formula gives a good notion of distance between vectors a and b on a network (*24, 57*). Specifically, it can recover the infection and healing parameters in a Susceptible-Infected-Susceptible (SIS) or Susceptible-Infected-Recovered (SIR) model by comparing two temporal snapshots of an epidemic – a more infectious disease with faster recovery covers more space across a social network in the same amount of time.

Figure 2 shows this intuition in the simplest possible scenario. We have three 3D vectors $a = (1, 0, 0)$, $b = (0, 1, 0)$, and $c = (0, 0, 1)$. We use x, y, and z to refer to the three spatial dimensions. In the traditional Euclidean case – following the blue arrow –, the three spatial dimensions are uncorrelated, and thus moving an equal amount in each direction contributes

equally to the distance measures. Thus a is equidistant from b and c – at a distance of $\sqrt{2}$.

However, we can use a graph G to express relationships between the dimensions as we do if we follow the yellow arrow in the figure. In that case, intuition tells us that b is closer to a than c, as the nodes with value 1 are two steps away in c and only one step away in b. In fact, $\delta_{G,a,b} = \sqrt{1}$ and $\delta_{G,a,c} = \sqrt{2}$.

To use GE for the purpose of estimating polarization, we need to split the vector o in two vectors: o^+ and o^-. o^+ contains all positive opinions and zero otherwise; o^- contains the absolute value of all negative opinions and zero otherwise. Once we do that, our $\delta_{G,o}$ measure of polarization becomes:

$$\delta_{G,o} = \sqrt{(o^+ - o^-)^T L^\dagger (o^+ - o^-)}.$$

The unit of our measure is the *step* or, to be more precise, its square root. This is the same unit as the one used by, e.g., shortest paths: if one needs to cross five edges to go from node i to node j, then i and j are five *steps* away from each other. In practice, one can interpret $\delta_{G,o}$ as the average "distance" between randomly sampled nodes in o^+ and o^-, weighted by how strongly these nodes hold their opinion (e.g., the distance between two nodes with opinions $+1$ and -1 is weighted higher than if the nodes had opinions -0.1 and $+0.1$). The units of this expected distance are "*steps*" and, as further discussed in the Analytical Approach section of the main paper and Supplementary Materials Section 4, the notion of distance in this interpretation is the so-called effective resistance.

We can see how $\delta_{G,o}$ considers all the factors we outlined in the previous section. The more well-separated the communities are – and the more they are organized at the mesolevel in an opinion spectrum – the more steps are necessary to traverse the network. The larger the opinion difference $o^+ - o^-$, the more these steps are weighted.

Results

We compare $\delta_{G,o}$ only with measures accepting the same input and providing the same output. Hence, methods working with signed networks (*20–22, 59*), or with expressed and internal opinions (*44, 45, 49*), or providing a simpler classification output in form of a yes/no value (*50*) are beyond the scope of this paper.

We specifically look at: opinion assortativity ($\rho_{G,o}$) (*32*), Random Walk Controversy (RWC_G) (*37*), density plots of opinion against average neighbor opinion (*33*); and boxplots of opinion against average opinion of the set of influenced nodes in a SIR model (*34*). The Supplementary Materials Section 1 includes details of how these measures are calculated.

Synthetic Data

We now show how $\delta_{G,o}$ is sensitive to the components of our definition of polarization, while the alternative ways of estimating polarization are insensitive to at least one of those factors. We follow the rows in Figure 1 one by one and show how the $\delta_{G,o}$ values and the alternative measures evolve over those dimensions. All numeric values reported in the figures that follow are the averages of 25 independent runs. All pairs of $\delta_{G,o}$ scores presented in the main text are statistically different, with the minimum z-score of the difference between any of the values shown being 3.4 – corresponding to a one-tailed $p < 0.001$. The density and box plots are taken from one representative run. The details on how we generate the various Gs and os, as well as the relevant parameters, are provided in the Analytical Approach section.

Supplementary Materials Sections 2 and 3 contain additional tests on the behavior of $\delta_{G,o}$ including its values for some interesting edge cases.

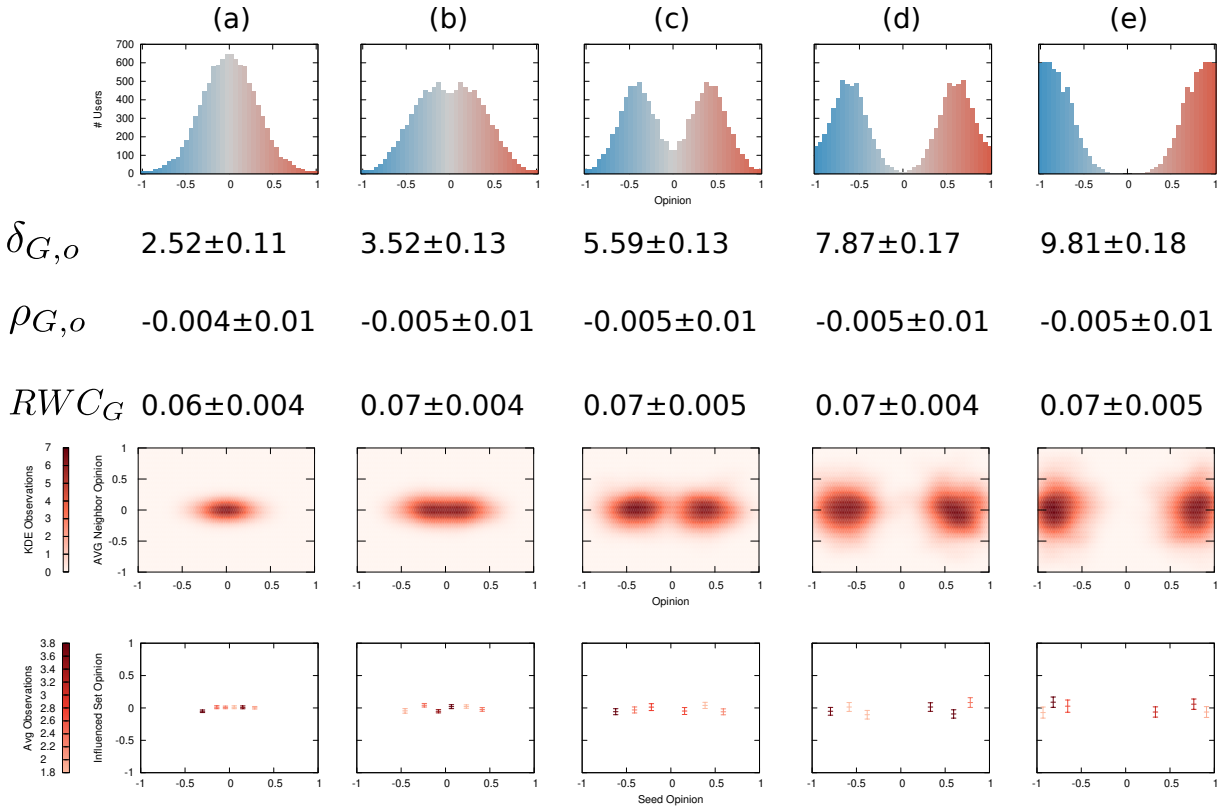

Figure 3: **The opinion component of polarization.** Each row (top to bottom): distribution of o values – number of nodes (y axis) with a given opinion (x axis and bar color) –; values of $\delta_{G,o}$, $\rho_{G,o}$, and RWC_G with their standard deviations across 25 independent runs; density maps of opinions (x axis) and average neighbor opinion (y axis); boxplots of seed opinion (x axis) and average opinion of the influenced set after a SIR propagation (y axis). The boxplots show the average for the middle tick, plus/minus its standard error for the top/bottom ticks. In the bottom two rows, color (from bright to dark) is used proportionally to the number of observations within the data point. Changes in the distribution of opinions lead to progressively increasing polarization through columns (a-e). The graph G (not depicted) has no communities.

The Opinion Component

In Figure 3, we start with a network without a community structure, in which the opinions distribute normally in the opinion spectrum (leftmost plot), and randomly over the network. This is a state of low polarization. As we move from Figure 3(a) to Figure 3(e), we create more and more polarization in the opinion vector o, keeping G as a random graph without

communities. The second row shows that the $\delta_{G,o}$ values grow by a factor of almost four, implying a significant increase in polarization. This corresponds to our intuition about the opinion component of polarization.

Neither assortativity ($\rho_{G,o}$, third row) nor RWC_G (fourth row) are able to capture this change. All their values are not statistically different from each other. This is because a random graph has zero expected assortativity (see Supplementary Materials Section 1), while RWC_G must bisect the network into two communities, regardless of how extreme the opinion difference is.

The density maps of the average neighbor opinion (fifth row) and the average influenced set opinion (sixth row) are able to capture the differences in the opinion value distributions.

The Structural Component

Next, in Figure 4, we take the most polarized opinion vector o from Figure 3 – the distribution in the top row of Figure 3(e) – and we investigate the structural component of polarization. We create eight communities in the network, each of which has a high degree of opinion homophily. As we move from Figure 4(a) to Figure 4(e), we change the connection probabilities of the nodes inside the network. We decrease p_{out}, the probability that a node will connect to a node in a different community. We then increase p_{in} – the probability of a node connecting to a node in the same community so that all networks in Figure 4 have the same expected number of edges.

The larger the difference between p_{out} and p_{in}, the higher the polarization, driven by the assortative communities in the structural component. We see that the values of $\delta_{G,o}$ (second row) follow our expectation, growing by a factor of around five. Thus, we can conclude that the measure is also sensitive to the structural component of polarization, not only to the opinion component.

Assortativity ($\rho_{G,o}$, third row) is able to distinguish between the five networks, but it is

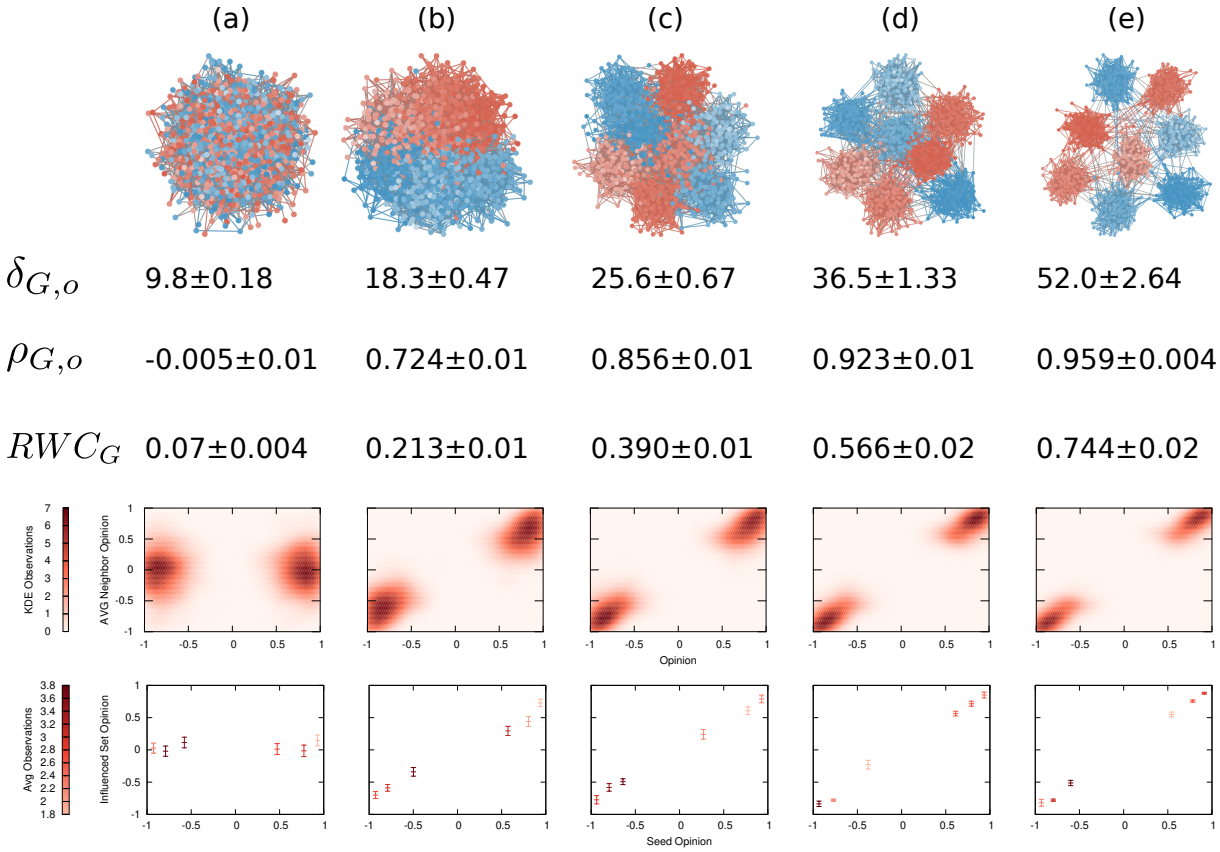

Figure 4: **The structural component of polarization.** Same legend as Figure 3. Structural changes lead to progressively increasing polarization through columns (a-e). p_{out} values: (a) 0.0085, (b) 0.0024, (c) 0.0012, (d) 0.0006, (e) 0.0003.

overly sensitive to relatively small initial changes to the random structure, downplaying the subsequent emergence of strong communities. The difference between Figure 4(a) and Figure 4(b) is more than three times as large as the difference between Figure 4(b) and Figure 4(e). This shows that, while assortativity can catch structural separation, it makes it difficult to distinguish weak communities from strong ones. The same can be said for the density maps of the average neighbor opinion (fifth row) and the average influenced set opinion (sixth row).

RWC_G (fourth row) picks up structural separation well.

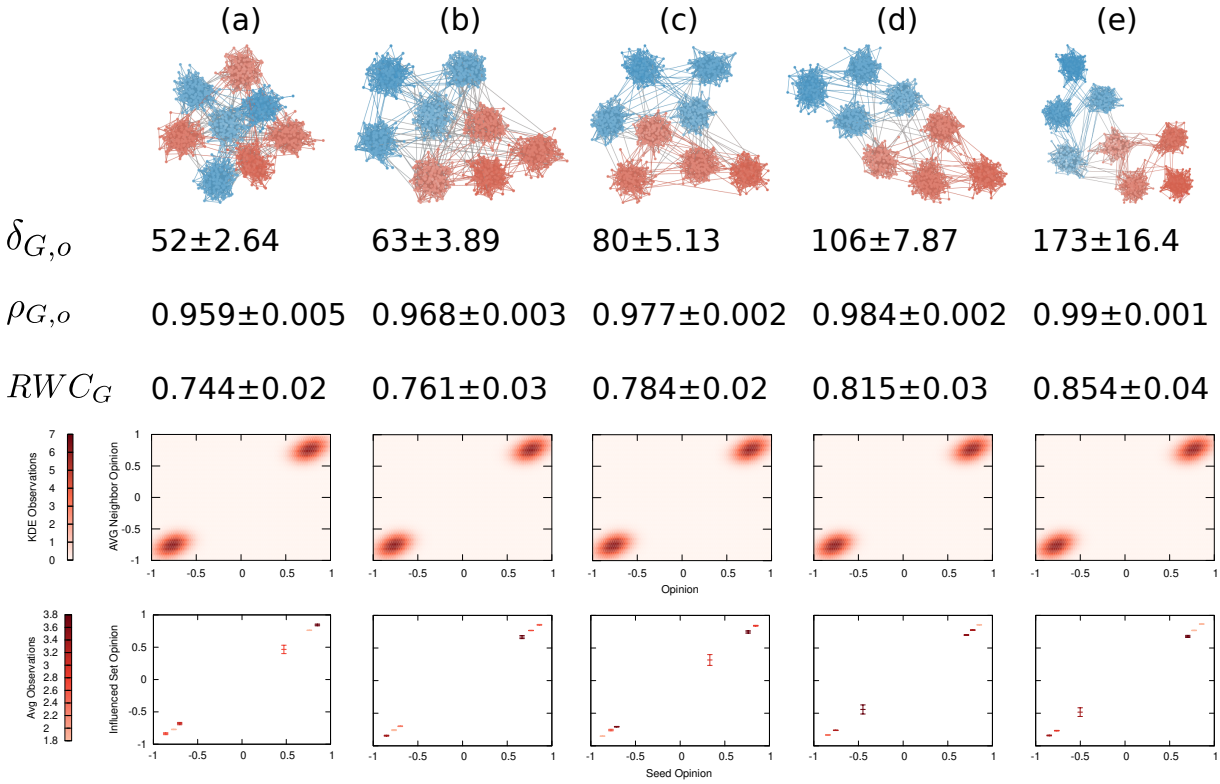

Figure 5: **The opinion-structural interplay.** Same legend as Figure 3. Changes in the opinion-structural interplay lead to progressively increasing polarization through columns (a-e). Each community in the network is connected to its (a) 7, (b) 5, (c) 4, (d) 3 and (e) 2 closest communities in terms of average opinion.

The Opinion-Structural Interplay

Finally, in Figure 5 we observe what happens when we modify the opinion-structural interplay at the mesolevel. To do so, we set some p_{out} values to zero. Specifically, each community gets progressively more and more isolated from the rest of the network as they preferentially disconnect from communities with a larger opinion difference. This mesoscale structure is something we observe empirically, as we show later when looking at data from actual debates on Twitter.

The network in Figure 5(a) is roughly equivalent to the one in Figure 4(e), where all communities connect to each other, and where the opinion distribution has values clustered around

-1 and $+1$. Starting from Figure 5(b) to Figure 5(e), we lower the number of connected neighboring communities from five to two.

Again, moving from Figure 5(a) to Figure 5(e) implies increasing levels of polarization, as it gets progressively harder for people to be exposed to differing points of view. This is reflected by a three-fold increase of the value of $\delta_{G,o}$. The large difference for each column shows that the measure is sensitive to the structural changes at the mesolevel.

Assortativity ($\rho_{G,o}$, third row) and RWC_G (fourth row) are not particularly sensitive to the mesolevel organization of the network – certainly not as much as they are to the structural component alone. Assortativity only changes at the second significant digit and always scores values near the maximum of $+1$. On the other hand, RWC_G is prone to misclassification, as the standard deviations show that there is an overlap between the higher bound of one level (for instance, Figure 5(a)) and the lower bound of the following one (in this case, Figure 5(b)). However, both measures do a reasonable job at capturing the mesolevel organization of the opinion-structural interplay.

The density maps of the average neighbor opinion are indistinguishable from each other (fifth row). This is because they exclusively look at local information, and they are blind to the mesolevel organization of the network. The average influenced set opinion (sixth row) could in principle capture the mesolevel organization as it is not bound by looking at direct neighbors, but allows the influenced set to percolate through the structure. However, the communities are too large and too well-separated for this to happen in practice, and the differences between each plot from Figure 5(a) to Figure 5(e) are minimal.

From Figures 3, 4, and 5 we can conclude that $\delta_{G,o}$ is the only measure sensitive enough to recognize each further example as a part of a continuum of increasing levels of polarization. $\delta_{G,o}$ captures the opinion and structural components, as well as their interplay happening at the mesolevel of the network. We support this statement by showing, in Supplementary Materi-

als Section 3, how $\delta_{G,o}$ varies smoothly across all the parameter values used to generate our synthetic data.

The alternative measures lack sensitivity to at least one aspect of polarization. Assortativity and RWC_G are blind to the opinion component and overemphasize the structural component over the opinion-structural interplay, while density maps of the average neighbor opinion and the average influenced set opinion in a SIR propagation do not capture the opinion-structural interplay at the mesolevel, and overemphasize the opinion component over the structural one.

Applications

We now turn to looking at real-world networks to show the insights one could gather from using $\delta_{G,o}$. We compare different political debates happening on Twitter including the 2020 US presidential election, and the evolution of US representatives over time.

Twitter Debates

Figure 6 shows examples of three debates happening on Twitter in the mid 2010s. The node color reflects ideological leaning on a US-focused liberal (blue) to conservative (red) scale. These center on three topics in the US political context which had been discussed by Twitter users between 2015 and 2016: the US Medicare reform known as Obamacare, gun control, and abortion.

$\delta_{G,o}$ shows moderate levels of polarization with values between 9 and 17. The least polarized debate is about Obamacare, while the abortion debate is the most polarized.

We call these levels of polarization "moderate" for several reasons. First, most opinions in the Obamacare network are uniformly distributed over the entire spectrum, leaving structure as the main source of polarization. The gun control network has more diverging and extreme opinions, but the distribution of ideological leanings is heavily skewed to the left, reducing

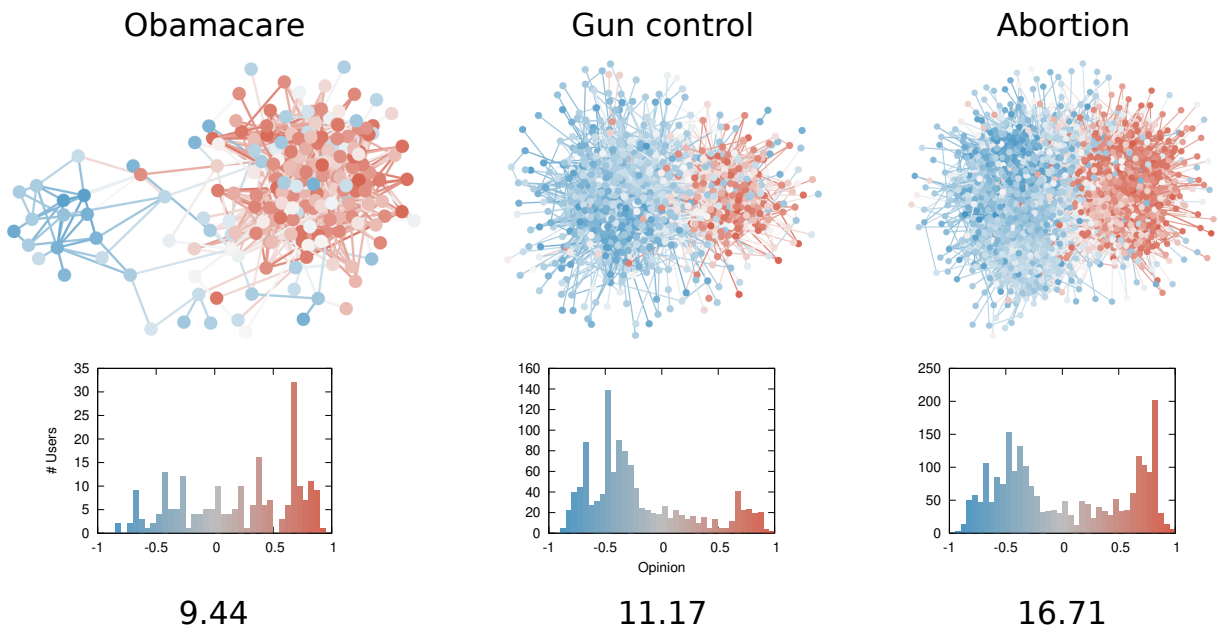

Figure 6: **The Twitter debate networks.** From top to bottom row: network topology, users as nodes, interactions as edges, opinions as colors of both nodes and edges; opinion distribution, number of users (y axis) with a given opinion (x axis and bar color); $\delta_{G,o}$ score.

polarization – polarization is low if most people agree on a position, even if it is a relatively extreme one. In this case, the vast majority of users are located left of center and, as a consequence, the $\delta_{G,o}$ score is reduced. The abortion debate is the most polarized because it has both high opinion divergence and roughly equally sized clusters. The score is still moderate because there is a high number of connections between the clusters, showing a level of communication between the faction that reduces overall polarization – 3% of all the edges of the network are between a "red" and a "blue" node, while this figure is below 2% for both the Obamacare and gun control debates.

Note that the abortion network actually has a mesoscale organization with subcommunities inside the main two opposing communities, as we detect via a stochastic blockmodel community discovery in Supplementary Materials Section 8. This provides support to our definition of polarization that includes an opinion-structural mesoscale interplay.

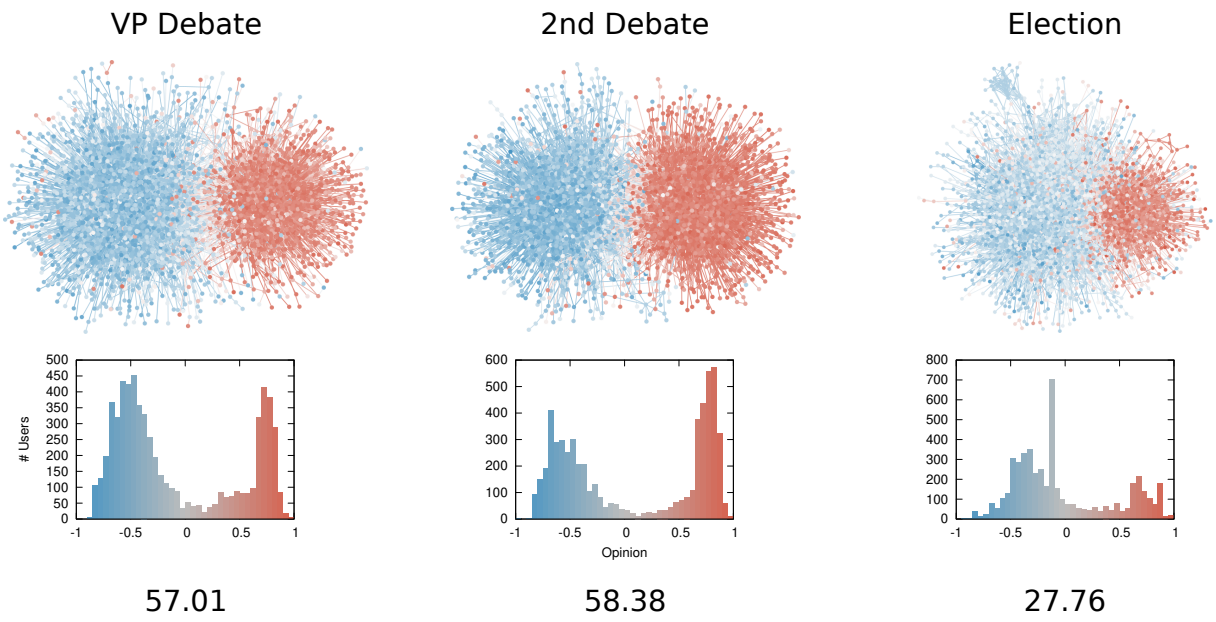

Figure 7: **The Twitter election networks.** From top to bottom row: network topology, users as nodes, interactions as edges, opinions as colors of both nodes and edges; distribution of ideological leanings, number of users (y axis) with a given opinion (x axis and bar color); $\delta_{G,o}$ score.

Note that $\delta_{G,o}$ is scale invariant as we show in Supplementary Material Section 2. It follows that differences in the polarization scores cannot be ascribed simply to the size of the network in terms of number of nodes.

Twitter Elections

Figure 7 shows the progression of the US presidential election in 2020. The networks center on the Vice-Presidential debate (October 7th, 2020), the second presidential debate (October 22nd, 2020), and election day (November 3rd, 2020).

$\delta_{G,o}$ shows high levels of polarization for the second debate and the VP debate. Both networks contain two extremely separated clusters with fewer than 1% of edges between them. Moreover, the opinion values are distributed towards the extremes. This explains why the scores are higher than the ones we show in Figure 6. During the 2020 election, users held opinions

farther from each other, and stopped interacting with disagreeing users.

Election day has significantly lower polarization due to a noticeable spike in the neutral portion of the opinion spectrum. This is caused by the necessity of sharing raw election result updates, which come from neutral and factual sources. In fact, the most shared domain during that period is from Associated Press, which has a moderate opinion value of -0.13 and is responsible for the noticeable peak in the opinion distribution. This suggests caution when estimating polarization scores in a context where people are both discussing opinions and hard facts at the same time.

US House of Representatives

We build the networks using voting records from the US House of Representatives (*60*). We connect two congressmen if they cast the same vote on the same bill a significant number of times – the Analytical Approach section provides more details. The o vector is their DW-NOMINATE score (*61*), an established way of quantifying their political leaning. We do not consider data from the Senate because senators cannot co-vote with members of the House: including them would create a disconnected network. We observe comparatively low $\delta_{G,o}$ values for two reasons. First, even though G has two opposing dense communities, the network has a small diameter and average path length of approximately $1.5 - 1.95$ (see Supplementary Table 4). This means that extreme congressmen in either community are separated by less than two steps on average, leading to low structural separation. Second, depending on the congress, 67% to 93% of the DW-NOMINATE scores are between -0.5 and 0.5, which suggests low opinion divergence as well, because the opinion values predominantly cover a smaller portion of the available $[-1, +1]$ interval.

Notwithstanding these characteristics, the US Congress has been viewed as an example of polarization escalation (*62*). Figure 8 supports this view. Up until the 98th Congress (1983-

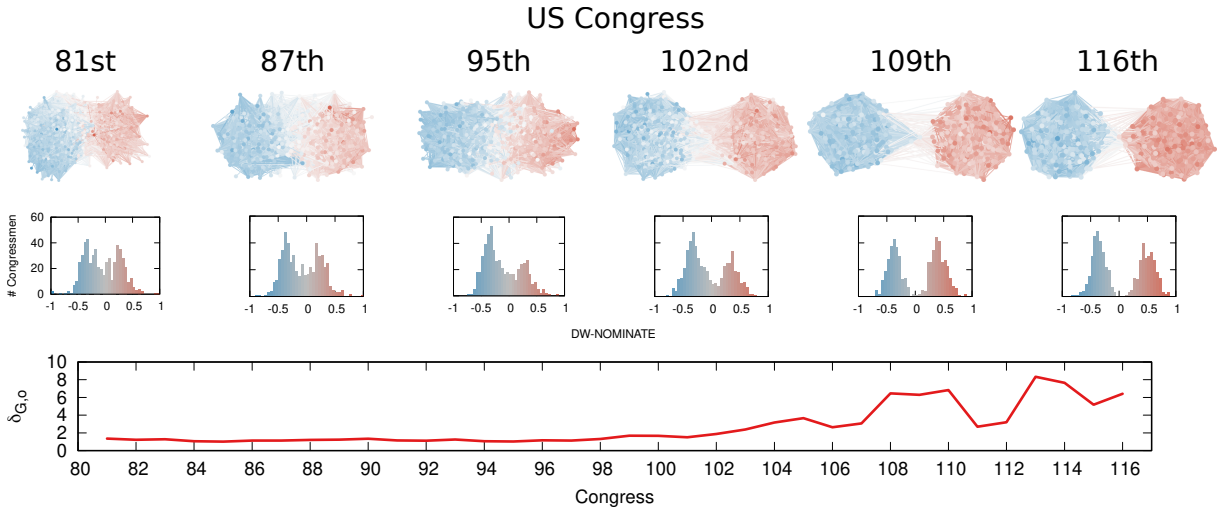

Figure 8: **Polarization in the US House of Representatives.** From top to bottom row: network topology, congressmen as nodes, co-voting relationships as edges, opinions as colors of both nodes and edges; opinion distribution, number of congressmen (y axis) with a given opinion (x axis and bar color); $\delta_{G,o}$ score. For the network and opinion rows, we show six examples of the 36 networks analyzed.

1985), polarization was almost non-existent, with $\delta_{G,o}$ scores around 1. Ever since the 98th Congress, $\delta_{G,o}$ scores have been on the rise to a maximum of more than 8.

This can be considered a high score, given the caveats we presented about how G and o are built. We should not compare these scores directly with the ones obtained from Twitter since the way of estimating o is vastly different. To contextualize the score, we can pick extreme Democrats and Republicans in the 116th Congress (2019-2021) and calculate the score we would get if they represented the entirety of their parties. If we perform this experiment using James McGovern for Democrats and Matt Gaetz for Republicans, we get a score of 14. This can be considered close to the maximum, as McGovern is part of the most left-leaning caucus of the House (the Congressional Progressive Caucus) and Gaetz is part of the most right-leaning one (the Freedom Caucus). They both have extreme DW-NOMINATE scores as well. On the other hand, the most moderate possible pair in the 116th Congress according to DW-NOMINATE is Ben McAdams and Brian Fitzpatrick who were members of the centrist caucuses Blue Dog and

Main Street Partnership. If they were composing the entirety of their parties, the polarization score would be a mere 0.2.

According to our measure, the most polarized House in the post-WWII history was the 113th (2013-2015), which coincided with the beginning of Barack Obama's second term, plus a debt-ceiling crisis following the full application of the Affordable Care Act (Obamacare), the 2014 Russo-Ukrainian conflict, strong debates about immigration reforms, and a controversial escalation of US military action in Syria and Iraq against ISIS, among other things.

Discussion

In this paper we tackled the issue of estimating the level of polarization in a social network. We ask how polarized a system is given the set of social connections and the opinions of all the individuals in the network. We decompose the polarization question in two main components and an interplay factor: how varied the opinions are (opinion component), how assortative the communities are (structural component), and how communities organize at the mesolevel of the network (opinion-structural interplay).

Intuitively, our estimate is based on the network distance between all pairs of disagreeing individuals, weighted by how strongly they hold their opinions. We show that our measurement is sensitive to all factors of polarization, a feat that is not achieved by the current state-of-the-art measures for polarization. We also show that the measure is able to unveil interesting insights in a number of real-world networks spanning from debates on Twitter to co-voting patterns in the US House of Representatives.

This is the starting point of a promising research path. However, there are a number of caveats and limitations that can be amended in future works. In general, some caution is necessary when taking the $\delta_{G,o}$ estimations at face value. If one wants to talk about ideological polarization at an entire nation's level, then they cannot rely on social media data like we do

in this paper. The social networks used here are a sample of the entire structure and, even if they considered the entirety of Twitter, it would still be a non-representative sample of the population.

If the first caveat focuses on how G is built, we also need caution when it comes to how o is estimated. $\delta_{G,o}$ scores are not compatible across networks if the ways to estimate o in different networks vary, as is the case between the Twitter and the US House of Representatives networks.

In addition, we focus mainly on measuring and summarizing the opinion and structural component and their interplay in a single, consistent measure. Another relevant question for future work may be to determine how much each component contributes to the overall level of polarization in a network. In Supplementary Materials Section 8, we show how to estimate the opinion component and the structural component on their own, as well as the strength of their correlation. A decomposition of $\delta_{G,o}$ into individual components might be interesting to understand how their importance has developed over time, and to design evidence-based strategies that help reduce polarization on social media.

Another limitation is that $\delta_{G,o}$ is only apt at describing ideological polarization, that is the extent to which opinions get farther away towards extremism and people with different opinions tend to isolate from each other. Affective polarization – which pertains to *how* people with different opinions interact with each other (*11*) – is also of great interest as it is the one truly affecting the quality of online discourse. One way we could approach affective polarization is via network co-variance (*52*) and/or correlations (*63*), since affective polarization should manifest as a correlation on the edges. Specifically, one would look whether the sentiment of a relationship is correlated with the opinion difference between the two individuals. These two approaches share commonalities with our $\delta_{G,o}$ measure – for instance, they all rely on effective resistance –, showing how future work can expect to develop a coherent framework able to

describe both ideological and affective polarization in consistent and comparable terms.

A further limitation – common in the literature – is that our measure assumes that people organize themselves in a one-dimensional opinion space with only two poles. This describes somewhat well the US political environment, and debates with a clear "for" and "against" position. However, it has two drawbacks.

First, it is grossly underpowered for a multi-pole scenario such as the multi-party political systems common in many European countries. Multiple parties does not necessarily imply that there is a corresponding ideology dimension per party, nevertheless creating a measure able to capture multiple ideological scales at the same time could be useful to avoid flattening everything on a two-pole system. There are some polarization-related studies for multi-party systems (*64, 65*), but they do not quite capture the objective of this paper: estimating a single numeric score for a given G-o pair. Instead, they return a much more complex output describing the likelihood of two nodes to connect given their characteristics – data that might be unavailable. We can explore dimensionality reduction techniques to allow $\delta_{G,o}$ to tackle a scenario with multiple different opinions at the same time, rather than just two. We outline one suggestion in Supplementary Materials Section 6.

Second, by analyzing a debate at a time we disregard the role of ideological consistency (*13, 66–69*). We can expect, e.g., a person in favor of Obamacare to also be in favor of gun control and abortion rights. There are two ways to tackle ideological consistency. The first would be to use a multilayer network, in which each layer is a debate. Then one can apply the multilayer version of $\delta_{G,o}$ (*70*) and get a polarization score that can be strengthened or weakened depending on the level of ideological consistency. Alternatively, one can calculate the network correlation between the different opinions of the individuals (*63*) to understand how consistent they are.

Our measure of polarization shares a drawback with all other data-driven approaches to

polarization: if the data estimating the opinion of the individuals is inaccurate, the measure will provide inaccurate results. However, in Supplementary Material Section 7 we show how it is possible find upper and lower bounds of a polarization estimate if one knows how uncertain the opinion measurements are.

Other limitations involve the limited scalability of our approach, which is relatively memory-hungry and thus unable to tackle networks with millions of nodes. We plan to fix this issue in future work by employing Laplacian solvers (*71, 72*). We can also work on building a better intuition for the units of our measure, and devise a way to normalize $\delta_{G,o}$ so that it takes values between, say, 0 and 1.

Analytical Approach

Interpretation of $\delta_{G,o}$

A convenient mental image to aid the interpretation of $\delta_{G,o}$ is the percolation of the opinions in a network, which can be modeled using discrete heat diffusion techniques. We can consider o as the temperature reading (opinion) of $|V|$ thermometers, each located in a node. $\delta_{G,o}$ is directly proportional to the (square root of the) time it takes for heat to diffuse across the network and bring it to equilibrium.

Figure 9 shows a graphical depiction of the diffusion process on a grid graph. The starting condition has some nodes in opposite corners at temperature -1 and $+1$. The polarization of this initial condition is $\delta_{G,o} \sim 2.95$ and we therefore expect it to take between 8 and 9 units of time for the system to converge to the average temperature (opinion), which is what we see if we run the simulation in the figure. For the simulation, we solve the discrete heat equation $\frac{do}{dt} = -Lo$ to find the solution at each time t (*73*).

The relation between $\delta^2_{G,o}$ and time to convergence is not always as direct as in this example, but in general we find that the polarization $\delta_{G,o}$ is directly proportional to the time it takes to

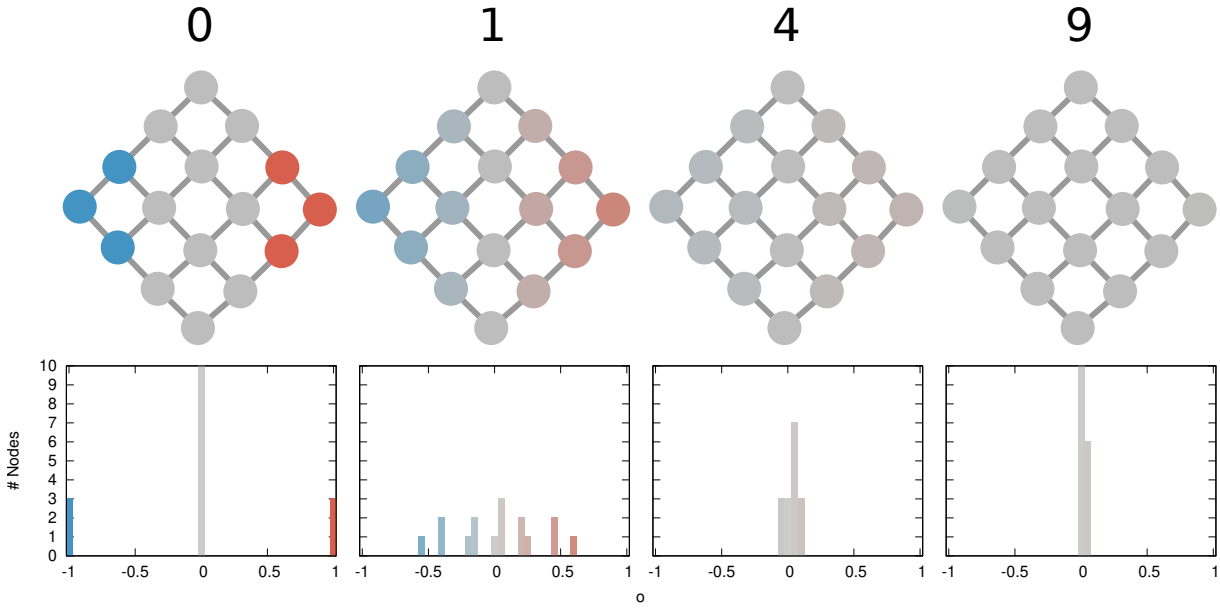

Figure 9: **Polarization as heat diffusion.** The graphs (second row) represent the status of the network at each time increment (first row). The histograms (third row) show the distribution of o values on the graph. Values in o span from -1 (blue) to $+1$ (red) passing via 0 (gray).

reach equilibrium – defined as the time t when the standard deviation of the opinion vector $o(t)$ goes below some fixed low ϵ value. We confirm this in Supplementary Materials Section 4 by repeating the diffusion experiment on many input pairs G, o. In Supplementary Materials Section 5 we further show how $\delta_{G,o}$ can be interpreted as a network version of the co-variance between the o^+ and o^- vectors.

$\delta_{G,o}$ and Effective Resistance

$\delta_{G,o}$ also has a direct relationship with "effective resistance", which is a robust way to measure distances between two nodes in a network and reflects the "effective number of steps" between two nodes (*58*). The effective resistance between two nodes i and j is denoted by ω_{ij} and defined as

$$\omega_{ij} := (e_i - e_j)^T L^\dagger (e_i - e_j),$$

where e_i is a vector with 1 at the i^{th} entry and zeros otherwise. The effective resistance is proportional to the average time it takes for a random walker to go from node i to node j, and then back again to i. In other words, this tells us how easy it is to traverse the network and move from one node to another, and back. Compared to the shortest-path distance, which measures the length of the shortest path between two nodes, the effective resistance takes into account the paths of all lengths and how they are interconnected.

To see why our measure of polarization is related to the effective resistance, let us consider a special case. Suppose that o^+ and o^- are concentrated in two nodes, i and j respectively, and zero otherwise. The polarization is then equal to

$$\delta_{G,o} = \sqrt{(o_i^+ - o_j^-)^T L^\dagger (o_i^+ - o_j^-)} = \sqrt{\omega_{ij}}.$$

In other words, the larger the effective resistance between i and j, the larger the measured polarization.

This interpretation in terms of effective resistances also holds for balanced opinion distributions. For a balanced opinion distribution, we assume that the total sum of the positive opinions equals the sum of the negative opinions, e.g. $\sum o_i^+ = \sum o_i^- = 1$ and $\sum o_i = 0$. This can be trivially achieved by normalizing o^+ and o^- with their sums. Then we make use of the fact that for any zero-sum vector x, with $\sum_i x_i = 0$, the pseudoinverse Laplacian product can be written in terms of the effective resistances as $\sum_{i,j} x_i (L^\dagger)_{ij} x_j = -\frac{1}{2} \sum_{i,j} x_i \omega_{ij} x_j$ – this follows from the definition of the effective resistance. Then:

$$
\begin{aligned}
\delta_{G,o} &= \sqrt{(o^+ - o^-)^T L^\dagger (o^+ - o^-)} \\
&= \sqrt{\sum_{i,j} (o_i^+ - o_i^-)(L^\dagger)_{ij}(o_j^+ - o_j^-)} \\
&= \sqrt{-\frac{1}{2}\sum_{i,j} (o_i^+ - o_i^-)\omega_{ij}(o_j^+ - o_j^-)} \\
&= \sqrt{\sum_{i,j} o_i^+ o_j^- \omega_{ij} - \frac{1}{2}\left(\sum_{i,j} o_i^+ o_j^+ \omega_{ij} + \sum_{i,j} o_i^- o_j^- \omega_{ij}\right)}.
\end{aligned}
$$

Since all values o_i^+ are positive and sum to one, we can interpret this value as the probability of sampling a random individual X^+ with a positive opinion. The probability of each individual is proportional to how extreme their opinion is, which is $\Pr[X^+ = i] = o_i^+$.

For instance, Alice and Bob are both in favour of gun control, so $o_A, o_B > 0$, but Alice is "twice as extreme" in her opinion as Bob and thus $o_A = 2o_B$. When we select a random individual (X^+) in favour of gun control, we will select Alice twice as likely as Bob since we get $\Pr[X^+ = A] = 2\Pr[X^+ = B]$. Similarly, we can consider a random individual X^- with a negative opinion based on the values o_i^-.

We can now formulate a probabilistic interpretation of the polarization $\delta_{G,o}$:

$$
\delta_{G,o} = \sqrt{\mathbb{E}[\omega_{X^+X^-}] - \tfrac{1}{2}\mathbb{E}[\omega_{X^+X^+} + \omega_{X^-X^-}]},
$$

where the expectation operator $\mathbb{E}$ runs over the distribution over independent random variables X^+, X^-. This formula has the following interpretation: $\delta_{G,o}$ measures the degree to which two individuals with conflicting opinions are more separated than two individuals with agreeing opinions. The *polarization is thus the difference in distance between pairs of conflicting individuals and pairs of agreeing individuals*, where individuals are selected according to the strength of their conviction. This shows that polarization is a relative measure that compares

conflicting individuals with agreeing individuals.

This expression shows a possible generalization of $\delta_{G,o}$: if we have any notion of distance d between the nodes of a graph then we can define a polarization score as:

$$\delta_{G,o} = \sqrt{\mathbb{E}[d(X^+, X^-)] - \tfrac{1}{2}\mathbb{E}[d(X^+, X^+) + d(X^-, X^-)]}.$$

This distance d could for instance be the shortest path distance between nodes in a network, the physical distance between individuals, or the travel time between locations.

Computational Complexity of $\delta_{G,o}$

If one estimates $\delta_{G,o}$'s formula naively, as we do in this paper, the most expensive part of the framework is the calculation of $L^\dagger$, the pseudoinverse of the Laplacian. This requires to solve the singular value decomposition problem for L. The cost is cubic, meaning that the algorithm can scale in the worst case as $\mathcal{O}(|V|^3)$, and hence it is inapplicable for networks larger than around 10^4 nodes. However, we do not need to explicitly calculate $L^\dagger$ to calculate $\delta_{G,o}$. We can use Laplacian solvers (*71, 72*), which can calculate the $L^\dagger(o^+ - o^-)$ portion of $\delta_{G,o}$ in near-linear time. The complexity would then be $\mathcal{O}(|V|^n)$, with $1 < n < 2$, allowing the method to scale to much larger networks.

Synthetic Data Generation

For the experiments showing the intuition and motivation of $\delta_{G,o}$, we rely on the generation of synthetic graphs G and opinion vectors o.

Each G is generated using a simple stochastic blockmodel (SBM) (*74*). To generate a SBM one needs to specify the number of nodes $|V|$, which we always set to $1,000$. The second ingredient is the assignment of nodes to communities. In our case, we create eight communities, each of the same size (125 nodes). The final two parameters are p_{in} and p_{out}, which regulate

the probability of two nodes in the same community (p_{in}) or in different communities (p_{out}) to connect to each other.

Each o is generated starting from a normal distribution of 500 values centered on 0 with a standard deviation of 0.2. Then, in Figure 3, we progressively create more and more polarization in the opinion distribution by shifting the average μ from 0 until 0.8, in 0.2 increments. We replace each value o_x higher than 1 as follows: $o_x = 1 - (o_x - 1)$. This ensures that all o values are lower than or equal to 1. Finally, we set $o = (o_0, \ldots, o_{500}, -o_0, \ldots, -o_{500})$ and sort it, making it symmetric around 0 and of length $1,000$. Each community gets a continuous portion of o, ensuring opinion homophily inside the community.

For Figure 3 we fix $p_{in} = p_{out} = 0.0085$. When $p_{in} = p_{out}$ a SBM is equivalent to a plain random $G_{n,p}$ graph (*75*). In a $G_{n,p}$, each pair of nodes has the same probability of being connected, regardless of the community affiliations of the two nodes and thus there are no communities.

For Figure 4 we progressively decrease p_{out} from 0.0085 to 0.0003, correspondingly increasing p_{in} to keep the expected number of edges constant.

For Figure 5 we set $p_{out} = 0$ between the nodes belonging to specific pairs of communities, increasing the other p_{in} and p_{out} accordingly to maintain the same expected number of edges. Specifically, we only keep connections between communities belonging to neighboring portions of the opinion spectrum o.

Data Collection

The gun control, abortion, and Obamacare networks were collected from Twitter. In all three cases, we retrieve the tweets related to each topic. To do so, we use the tweet ids provided by previous works (*34*), which follow the procedure outlined in the literature (*76*). From the tweets we obtain a list of users involved in the debate. We create the network by collecting the 5,000

most recent followers of each users, a cap that is imposed by Twitter's rate limits.

We estimate the opinion of each user by looking at the URLs they share. Each domain has an opinion score between -1 and $+1$, with the data coming from the fact-checking website `https://mediabiasfactcheck.com/`. The scores are provided directly by the website, placing each news source in a continuous -1 to $+1$ interval. The user's opinion is the average of all the URLs they have shared. This procedure is in line with the standard practice in the literature (*34*).

These networks have been used for many studies in the past (*38*), but there might be differences in their topologies due to the dynamic nature of Twitter. The original data source only provides tweet ids, not their content and no network information, as per Twitter's terms of use. As a result, we need to recollect tweets and relationships that were established when the debates took place between 2015 and 2016. In the meantime, people might delete tweets, resulting in a different estimation of o because the relative frequency of URLs shared by a user changes. Moreover, users might follow/unfollow other users, or even delete their account entirely, changing the edge and node sets of G.

We follow the literature in using only tweets which link to (at least) one of the URLs with a known opinion score. The dataset contains only users which have tweeted at least five times on either topic.

We apply the same procedure to generate the US debate Twitter datasets – by collecting the networks about the second presidential debate, the vice presidential debate, and election day, for the 2020 election. This is based on tweet ids collected by the George Washington University (*77*). Also in this case, we only use tweets which link to (at least) one of the URLs with a known opinion score. Differently from above, we only consider users which have tweeted at least three times (and not five).

For the US House of Representatives network, we collect roll call vote data from Vote-

view.com (*60*). We connect two congressmen from the US House of Representatives following the procedure obtained from the literature (*78*) – omitting votes from the Senate, as they would create a disconnected component in the network.

In practice, we connect nodes if the two members agree with one another on a vote more often than a specific Congress-dependent threshold. The threshold value is the number of agreements in a specific Congress where the pair of members is more likely to be from the same party than from opposing parties.

For most of the history of the US House of Representatives, one could find a significant number of cross-party agreements, leading to well connected communities of Democrats and Republicans. This has stopped being the case from the 98th Congress, although the two communities are still part of a single connected component (otherwise we could not apply $\delta_{G,o}$).

Note that, with this procedure, a few nodes are isolated as they did not participate in enough votes to receive a connection, and thus they are dropped from the networks.

References

1. L. A. Adamic, N. Glance, *Proceedings of the 3rd international workshop on Link discovery* (2005), pp. 36–43.

2. M. Conover, *et al.*, *Proceedings of the International AAAI Conference on Web and Social Media* (2011), vol. 5, pp. 89–96.

3. Y. Mejova, A. X. Zhang, N. Diakopoulos, C. Castillo, *arXiv preprint arXiv:1409.8152* (2014).

4. W. Quattrociocchi, A. Scala, C. R. Sunstein, *Available at SSRN 2795110* (2016).

5. M. Coscia, L. Rossi, *PloS one* **17**, e0263184 (2022).

6. E. Pariser, *The filter bubble: What the Internet is hiding from you* (Penguin UK, 2011).

7. H. F. de Arruda, *et al.*, *Information Sciences* (2021).

8. F. Cinus, M. Minici, C. Monti, F. Bonchi, *Proceedings of the International AAAI Conference on Web and Social Media* **16**, 90 (2022).

9. M. Gentzkow, *Toulouse Network for Information Technology Whitepaper* pp. 1–23 (2016).

10. L. Boxell, M. Gentzkow, J. M. Shapiro, *Proceedings of the National Academy of Sciences* **114**, 10612 (2017).

11. E. Kubin, C. von Sikorski, *Annals of the International Communication Association* **45**, 188 (2021).

12. S. Iyengar, Y. Lelkes, M. Levendusky, N. Malhotra, S. J. Westwood, *Annual Review of Political Science* **22**, 129 (2019).

13. Y. Lelkes, *Public Opinion Quarterly* **80**, 392 (2016).

14. M. P. Fiorina, S. J. Abrams, *Annu. Rev. Polit. Sci.* **11**, 563 (2008).

15. A. I. Abramowitz, K. L. Saunders, *The Journal of Politics* **70**, 542 (2008).

16. J. N. Druckman, M. S. Levendusky, *Public Opinion Quarterly* **83**, 114 (2019).

17. N. Dias, Y. Lelkes, *American Journal of Political Science* (2021).

18. Y. Lelkes, *Political Science Research and Methods* **9**, 189 (2021).

19. L. V. Orr, G. A. Huber, *American Journal of Political Science* **64**, 569 (2020).

20. F. Bonchi, E. Galimberti, A. Gionis, B. Ordozgoiti, G. Ruffo, *Proceedings of the 28th acm international conference on information and knowledge management* (2019), pp. 961–970.

21. R.-C. Tzeng, B. Ordozgoiti, A. Gionis, *Advances in Neural Information Processing Systems* **33**, 10974 (2020).

22. Z. Huang, A. Silva, A. Singh, *Proceedings of the Fifteenth ACM International Conference on Web Search and Data Mining* pp. 390–400 (2022).

23. D. Baldassarri, S. E. Page, *Proceedings of the National Academy of Sciences* **118** (2021).

24. M. Coscia, *Proceedings of the International AAAI Conference on Web and Social Media* (2020), vol. 14, pp. 119–129.

25. I. Waller, A. Anderson, *Nature* **600**, 264 (2021).

26. A. J. Morales, J. Borondo, J. C. Losada, R. M. Benito, *Chaos: An Interdisciplinary Journal of Nonlinear Science* **25**, 033114 (2015).

27. C. Musco, I. Ramesh, J. Ugander, R. T. Witter, *arXiv preprint arXiv:2110.11981* (2021).

28. M. Coletto, K. Garimella, A. Gionis, C. Lucchese, *Online Social Networks and Media* **3**, 22 (2017).

29. M. Coletto, K. Garimella, A. Gionis, C. Lucchese, *Proceedings of the International AAAI Conference on Web and Social Media* (2017), vol. 11, pp. 496–499.

30. M. E. Newman, *Physical review E* **67**, 026126 (2003).

31. M. Coscia, *arXiv preprint arXiv:2101.00863* (2021).

32. B. Mønsted, S. Lehmann, *PloS one* **17**, e0263746 (2022).

33. W. Cota, S. C. Ferreira, R. Pastor-Satorras, M. Starnini, *EPJ Data Science* **8**, 1 (2019).

34. M. Cinelli, G. D. F. Morales, A. Galeazzi, W. Quattrociocchi, M. Starnini, *Proceedings of the National Academy of Sciences* **118** (2021).

35. L. Peel, J.-C. Delvenne, R. Lambiotte, *Proceedings of the National Academy of Sciences* **115**, 4057 (2018).

36. P. Guerra, W. Meira Jr, C. Cardie, R. Kleinberg, *Proceedings of the international AAAI conference on web and social media* (2013), vol. 7, pp. 215–224.

37. K. Garimella, G. D. F. Morales, A. Gionis, M. Mathioudakis, *ACM Transactions on Social Computing* **1**, 1 (2018).

38. K. Garimella, *et al.*, Polarization on social media (2018).

39. K. Darwish, *International conference on social informatics* (Springer, 2019), pp. 188–201.

40. H. Emamgholizadeh, M. Nourizade, M. S. Tajbakhsh, M. Hashminezhad, F. N. Esfahani, *Social Network Analysis and Mining* **10**, 1 (2020).

41. A. Cossard, *et al.*, *Proceedings of the International AAAI conference on web and social media* (2020), vol. 14, pp. 130–140.

42. S. Haddadan, C. Menghini, M. Riondato, E. Upfal, *Proceedings of the 14th ACM International Conference on Web Search and Data Mining* (2021), pp. 139–147.

43. N. E. Friedkin, E. C. Johnsen, *Journal of Mathematical Sociology* **15**, 193 (1990).

44. A. Matakos, E. Terzi, P. Tsaparas, *Data Mining and Knowledge Discovery* **31**, 1480 (2017).

45. C. Musco, C. Musco, C. E. Tsourakakis, *Proceedings of the 2018 World Wide Web Conference* (2018), pp. 369–378.

46. D. Chong, J. N. Druckman, *The American Political Science Review* **104**, 663 (2010).

47. J. N. Druckman, J. Leeper, Thomas, *Dædalus* **141**, 312 (2012).

48. J. Zaller, *The nature and origins of mass opinion* (Cambridge University Press, 1992).

49. X. Chen, J. Lijffijt, T. De Bie, *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining* (2018), pp. 1197–1205.

50. S. Benslimane, J. Azé, S. Bringay, M. Servajean, C. Mollevi, *International Conference on Web Information Systems Engineering* (Springer, 2021), pp. 339–354.

51. K. Garimella, G. De Francisci Morales, A. Gionis, M. Mathioudakis, *Proceedings of the 2018 world wide web conference* (2018), pp. 913–922.

52. K. Devriendt, S. Martin-Gutierrez, R. Lambiotte, *SIAM Review* **64**, 343 (2022).

53. F. Baumann, P. Lorenz-Spreen, I. M. Sokolov, M. Starnini, *Physical Review Letters* **124** (2020).

54. C. B. Currin, S. V. Vera, A. Khaledi-Nasab, *Scientific Reports* **12** (2022).

55. M. W. Macy, M. Ma, D. R. Tabin, J. Gao, B. K. Szymanski, *Proceedings of the National Academy of Sciences* **118** (2021).

56. F. P. Santos, Y. Lelkes, S. A. Levin, *Proceedings of the National Academy of Sciences* **118** (2021).

57. M. Coscia, A. Gomez-Lievano, J. Mcnerney, F. Neffke, *ACM Computing Surveys (CSUR)* **53**, 1 (2020).

58. D. J. Klein, M. Randić, *Journal of mathematical chemistry* **12**, 81 (1993).

59. L. Akoglu, *Proceedings of the International AAAI Conference on Web and Social Media* (2014), vol. 8, pp. 2–11.

60. J. B. Lewis, *et al.*, *https://voteview. com/ (accessed 25 February 2022)* (2019).

61. K. T. Poole, H. Rosenthal, *Congress: A political-economic history of roll call voting* (Oxford University Press on Demand, 2000).

62. Z. P. Neal, *Social Networks* **60**, 103 (2020).

63. M. Coscia, *Journal of Complex Networks* **9**, cnab036 (2021).

64. M. E. Del Valle, R. B. Bravo, *International journal of communication* **12**, 21 (2018).

65. M. Esteve Del Valle, M. Broersma, A. Ponsioen, *Social science computer review* p. 0894439320987569 (2021).

66. D. Baldassarri, A. Gelman, *American Journal of Sociology* **114**, 408 (2008).

67. A. Abramowitz, K. Saunders, *The Forum* (De Gruyter, 2005), vol. 3, pp. 1–22.

68. A. Abramowitz, *The disappearing center: Engaged citizens, polarization, and American democracy* (Yale University Press, 2010).

69. C. Hare, K. T. Poole, *Polity* **46**, 411 (2014).

70. M. Coscia, *ACM Transactions on Knowledge Discovery from Data (TKDD)* (2022).

71. N. K. Vishnoi, *et al.*, *Foundations and Trends® in Theoretical Computer Science* **8**, 1 (2013).

72. K. Deweese, *Bridging the Theory-Practice Gap of Laplacian Linear Solvers* (University of California, Santa Barbara, 2018).

73. R. I. Kondor, J. Lafferty, *Proceedings of the 19th international conference on machine learning* (2002), vol. 2002, pp. 315–322.

74. P. W. Holland, K. B. Laskey, S. Leinhardt, *Social networks* **5**, 109 (1983).

75. P. Erdős, A. Rényi, *et al.*, *Publ. Math. Inst. Hung. Acad. Sci* **5**, 17 (1960).

76. H. Lu, J. Caverlee, W. Niu, *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management* (2015), pp. 213–222.

77. L. Wrubel, D. Kerchner, 2020 United States Presidential Election (2021).

78. C. Andris, *et al.*, *PloS one* **10**, e0123507 (2015).

79. G. Karypis, V. Kumar, Metis: A software package for partitioning unstructured graphs, partitioning meshes, and computing fill-reducing orderings of sparse matrices (1997).

80. M. Coscia, *Proceedings of the 2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining* (2019), pp. 1–8.

81. R. Pastor-Satorras, C. Castellano, P. Van Mieghem, A. Vespignani, *Reviews of modern physics* **87**, 925 (2015).

82. S. T. McCormick, M. R. Rao, G. Rinaldi, *Mathematical programming* **94**, 459 (2003).

83. M. E. Newman, *Proceedings of the national academy of sciences* **103**, 8577 (2006).

84. T. P. Peixoto, *Physical Review E* **89**, 012804 (2014).

Acknowledgments

MC and KD designed the analysis. MH performed the analysis. MH and MC prepared the figures. MH, KD, and MC wrote and approved the manuscript. The authors declare that they

have no competing interests. All code needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. For details on how to access the data, please see “Data Collection”. KD was supported by The Alan Turing Institute under the EPSRC grant EP/N510129/1.

List of Supplementary Materials

Supplementary Files F1-F2.

Supplementary Materials for "Quantifying Ideological Polarization on a Network Using Generalized Euclidean Distance"

Marilena Hohmann,[1,†] Karel Devriendt,[2,3,†] Michele Coscia[4,*,†]

[1]Copenhagen Center for Social Data Science, University of Copenhagen,
Øster Farimagsgade 5, Copenhagen, DK
[2]Mathematical Institute, University of Oxford, Woodstock Road, Oxford, UK
[3]Alan Turing Institute, Euston Road 96, London, UK
[4]CS Department, IT University of Copenhagen, Rued Langgaards Vej 7, Copenhagen, DK

[*]To whom correspondence should be addressed; E-mail: mcos@itu.dk.
[†]Author contributed equally to the work.

1 Alternative Measures

1.1 Assortativity

Assortativity is a way to quantify the extent to which nodes with similar values connect to each other (*31*). The specific formula we use to estimate assortativity is the one developed by Newman (*30*).

This measure is calculated by creating a matrix e. Each entry e_{xy} in the matrix contains the fraction of the edges connecting two nodes in the network with values x and y of the measure of interest. This can be considered as a probability of the edge existing, implying that $\sum_{xy} e_{xy} = 1$.

For simplicity, we also record the fraction of edges originating from a node with value x ($\sum_y e_{xy} = a_x$) and the fraction of edges ending in a node with value y ($\sum_x e_{xy} = b_y$). If a network is undirected and unipartite, then $a_x = b_y$.

The assortativity is thus the Pearson correlation coefficient of x and y, which is:

$$\rho_{G,o} = \frac{\sum_{xy} xy(e_{xy} - a_x b_y)}{\sigma_a \sigma_b},$$

where σ_a and σ_b are the standard deviations of the distributions a_x and b_y. In this formulation, one needs the topology of G and the node values from the opinion vector o to build the matrix e (and, as a consequence, a_x and b_y).

Just like the Pearson correlation coefficient, this measure ranges from -1 (perfect disassortativity) to $+1$ (perfect assortativity).

It is not hard to see that assortativity cannot capture changes in the opinion component as defined and shown in the main paper: Suppose we have a vector of opinions o. Any linear change in o will generate the exact same assortativity value, provided that the values do not go beyond the $[-1, +1]$ interval. Halving all o values will generate the same assortativity value, even though this should be recognized as a lower polarization.

Assortativity also gives diminishing returns in capturing the structural component. It is linearly related to p_{out}, the probability of nodes in a community to connect outside their community, but linear changes in p_{out} have little effect for high p_{out} and large effects for low p_{out} – as Figure S1 shows. Networks that score the medium assortativity value ($\rho_{G,o} = 0.5$, we show a network with this value in the callout in the middle) are hardly distinguishable from networks scoring the minimum value ($\rho_{G,o} = 0.0$, we show an example of a network in the callout on the left). The regime in which networks have well-defined communities occupies a small fraction of the assortativity's domain.

In summary, assortativity can capture the alignment of opinions with the network struc-

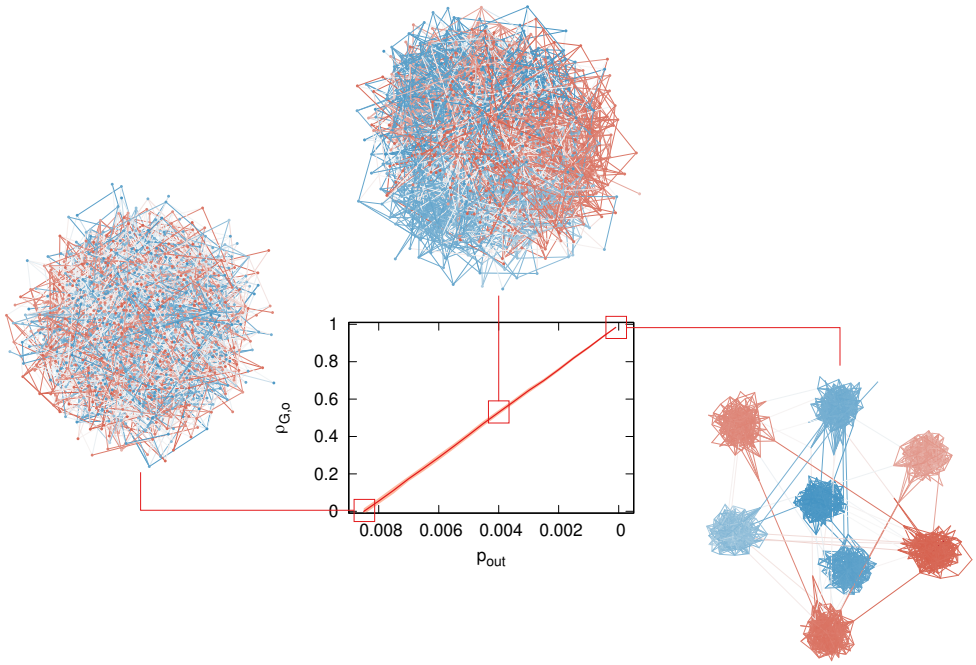

Figure S1: **The sensitivity of assortativity to the structural component**. $\rho_{G,o}$ values (y axis) for increasing levels of structural separation, i.e., decreasing p_{out} (x axis). The insets shows examples of networks at a given value of $\rho_{G,o}$, specifically (left to right): 0, 0.5, 1.

ture although it progressively loses sensitivity as we have better-defined communities, and it is wholly insensitive to the distribution of opinions.

1.2 Random Walk Controversy

Random Walk Controversy (RWC_G) is a measure that assumes the network can be partitioned into two opposing communities and then estimates the probability of a random walker to be able to transition from one community to another (*37*). The first operation is to bisect the graph into two communities. In the original paper, the authors use the METIS algorithm (*79*), although one could use any community discovery algorithm that can take the number of desired communities as input and that finds communities using a comparable definition as the one used by METIS (*80*).

Let us assume that we now have two communities C_1 and C_2, which are disjoint sets of

nodes ($C_1 \cap C_2 = \emptyset$) that cover the entire node set of G ($C_1 \cup C_2 = V$). RWC_G simulates a number of random walks – by default 10 times the number of nodes in the network – and records four probabilities: p_{C_1,C_2}, p_{C_2,C_1}, p_{C_1,C_1}, and p_{C_2,C_2}. In general, $p_{x,y}$ records the probability for a random walker to start in partition x and end in partition y. Half of the walkers start from C_1 and half from C_2. The walk terminates when it visits any node from a 10% random sample (from either side). Then:

$$RWC_G = p_{C_1,C_1}p_{C_2,C_2} - p_{C_1,C_2}p_{C_2,C_1}.$$

This measure is equal to $+1$ when all random walkers stay in their starting community, and equal to -1 when all random walkers end in the opposite community. A few things should be noted here.

First, the bisection into communities can be largely suboptimal because it is independent from the opinion vector o. If there are no clear-cut communities aligning themselves with o, then RWC_G's estimation is necessarily wrong. One could skip the community discovery and assign nodes to two communities according to whether their values in o are above or below 0. In either case, the actual distribution of o's values is irrelevant (it does not matter whether o values cluster tightly around 0 or at the extremes), and thus this measure cannot capture the opinion component by design, as we show in the main paper.

Second, the measure is not deterministic because it relies on a randomized simulation. One could fix this issue by deciding a fixed random walk length l, then take A^l – with A being the stochastic adjacency matrix of G. The result is the probability of a random walk starting from any node to reach any other node in l steps. Then one could estimate RWC_G with the formula above. In this case, RWC_G would be exact and deterministic. However, we do not perform this correction to stick with the original definition of RWC_G, noting that one could get different RWC_G values even with the same G and o.

Finally, in the original paper (*37*), the authors define multiple measures of polarization. Specifically, they also introduce versions using a random walk with restart, betweenness centrality, a 2D node embedding, Boundary Connectivity (*36*), and Dipole Moment (*26*). We do not investigate these variants as they are either highly correlated with RWC_G, or suffering from the same drawbacks about not considering the opinion distribution o.

1.3 Average Neighbor Opinion

In this approach, we estimate polarization by creating a density plot of opinions (*33*). For each node i in the network, we record its value o_i on the x axis. Then, the y axis value is derived from i's neighbor set N_i. Specifically, it is its average $\sum_{j \in N_i} o_j / |N_i|$.

The resulting density plot is then generated with a standard Kernel Density Estimation. A network is polarized if most of the density of points is concentrated among the top right and/or bottom left corners, implying that only nodes with similar values connect to each other.

As we show in the main paper, the issue with this approach is that it looks only at immediate neighbors, whereas any mesoscale organization of the network is not captured.

1.4 Influenced Set Opinion

This approach also estimates polarization graphically (*34*). Each node is considered in turn as the origin of a Susceptible-Infected-Recovered (SIR) epidemic event. In SIR, nodes start in a Susceptible state. They transition to Infected with a certain rate β when one or more of their neighbors are Infected. Finally, there is a recovery rate γ, which makes them transition to the Recovered state (*81*). Recovered nodes cannot be infected anymore. The process ends when all nodes are either in the Susceptible or Recovered state, and no more propagation can happen. In addition to epidemics, the SIR process can also model the spread of information or rumors on a network.

Say that R_i is the set of nodes in the Recovered state after the rumor propagates from node i. We can calculate their average opinion as $\sum_{j \in R_i} o_j / |R_i|$. Thus, this approach is exactly the same as the one we describe in the previous section, with the difference that we look at R_i (the set of infected nodes) rather than N_i (the direct neighbors of i).

Finally, rather than looking at the density plot, we bin all origin nodes according to their opinion value and show the distribution of the average influenced set nodes for each bin. A polarized network will show boxplots clustering on the top right and bottom left quadrants.

This can fix some of the conceptual problems of using the average opinion of the neighbours N_i, since R_i contains nodes that are not directly connected to i. However, a few more conceptual issues arise.

First, the β and γ parameters regulating the SIR event – as well as the number of bins in the plot – are chosen somewhat arbitrarily. While the authors of (*34*) show that this does not significantly impact the results as long as they are chosen within reasonable bounds, instructions on which bounds are reasonable are absent and rely on judgment calls from the analyst.

Second, just like RWC_G, the process relies on a randomized simulation, and thus can result in different estimations even with the same G and o.

Finally, while replacing N_v with R_v should help in capturing the mesoscale organization of G, this does not seem to happen in practice. Our experiments in the main paper show that different variations in of the opinion-structural interplay generate indistinguishable boxplots.

2 Properties of $\delta_{G,o}$

2.1 Lack of Maximum

$\delta_{G,o}$ lacks a maximum value, i.e., its domain is $[0, +\infty]$. That is to say, given a G and an o we can always find a G' and o' pair such that $\delta_{G,o} < \delta_{G',o'}$. We prove this statement by looking at a simple graph structure: a chain graph. In a chain graph, nodes are placed on a one dimensional

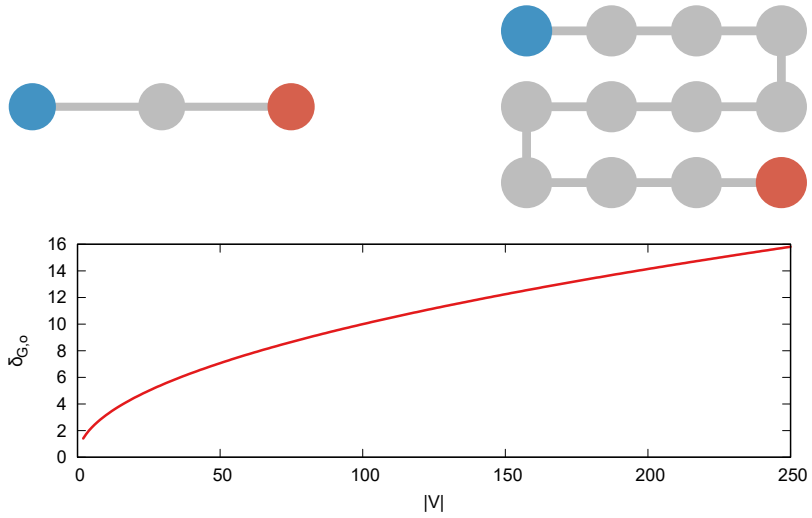

Figure S2: **The polarization values for growing chains**. (Top row) Examples of chains. The node color is proportional to the value in o (-1 blue, 0 gray, $+1$ red). (Bottom row) The value of $\delta_{G,o}$ (y axis) for a given number of nodes in the chain (x axis).

space and connected with their two closest neighbors with the exception of the two endpoints of the chain, which are only connected to their closest neighbor.

Then we generate a corresponding o vector. All entries in o are set to zero except the two entries for the endpoints of which one is set to -1 and the other to $+1$. We then grow the chain graphs by extending the chain as Figure S2 shows in the top row.

For each additional node in the chain, $\delta_{G,o}$ grows by a predictable amount. Specifically, $\delta_{G,o} = \sqrt{|E|}$ – see Figure S2 (bottom row). This is a known result in line with what was shown in the original paper defining the baseline node vector distance measure (*24*).

This is the reason why it is not possible to simply normalize $\delta_{G,o}$ to be defined between 0 and 1. There is not a well-defined maximum that can be used as a normalization factor.

One could narrow down the problem. Rather than finding the maximum of $\delta_{G,o}$ for any G, we could be looking for the maximum of $\delta_{G,o}$ for a given G, which implies finding the correct o vector. Unfortunately this is not a well-understood problem yet and thus it is not feasible to solve it here.

The possible opinion vectors must satisfy $o \in [-1,1]^{|V|}$, i.e., the only requirement is that $o_i \in [-1,1]$ for each i where $|V|$ is the number of nodes in the network. Hence, we may formalize the *maximum polarization problem* (MPP) as

$$\max_{o\in[-1,1]^{|V|}} \delta_{G,o} \tag{MPP}$$

Since the pseudoinverse Laplacian is positive semidefinite and the square root is monotone increasing on $\mathbb{R}^+$, we may simply square the objective function $\delta_{G,o} \to \delta_{G,o}^2$ and equivalently write MPP as:

$$\max_{o\in[-1,1]^{|V|}} o^T L^\dagger o$$

and then take the square root of the solution. As a first result, we can show that the opinion vector $o^\star$ that achieves the maximum polarization will necessarily have entries that are either $+1$ or -1. This implies that the domain of MPP can be changed from $[-1,1]^{|V|}$ to $\{-1,+1\}^{|V|}$, i.e., with $o_i = \pm 1$ for every node i.

Proposition 1 *The opinion vector $o^\star$ that solves MPP lies in $\{-1,+1\}^{|V|}$.*

Proof: Since both $L^\dagger$ and the domain for o are bounded, $\delta_{G,o}$ will be bounded from above by some finite number and thus at least a supremum exists. As the domain $[-1,1]^{|V|}$ is compact, this supremum is achieved by some $o^\star$, the optimal solution to MPP. There is no uniqueness, e.g., $\delta_{G,o^\star} = \delta_{G,-o^\star}$, but this is not required for the proof. Assume for contradiction that there is a node i for which $o_i^\star \in (-1,+1)$, i.e., not equal to ± 1.

Consider what happens if we increase or decrease the opinion of i, i.e., $o_i^\star \pm \epsilon$, with ϵ small enough such that this lies in $[-1,1]$. Then the node vector distance becomes:

$$(o^\star + \epsilon e_i)^T L^\dagger (o^\star + \epsilon e_i) = \delta_{G,o^\star}^2 + \epsilon^2 e_i^T L^\dagger e_i + 2\epsilon e_i^T L^\dagger o^\star,$$

where e_i is the i^{th} unit vector. Since $\epsilon^2 > 0$ and $L^\dagger$ is positive semidefinite with kernel $L^\dagger o = 0$ if and only if o is a constant vector (which e_i is not), the second term is positive. For the last term, we may choose the sign of ϵ such that the term becomes positive, i.e. $\text{sign}(\epsilon) = \text{sign}(e_i^T L^\dagger o^\star)$ then we find that

$$\delta_{G,o^\star+\epsilon e_i} > \delta_{G,o^\star}$$

But this is in contradiction with $o^\star$ being the maximum polarization vector, hence our assumption that a node i with $o_i \in (-1, 1)$ exists must be false and this concludes the proof. □

Following this proposition, we can rewrite the MPP as:

$$\max_{o\in\{-1,1\}^{|V|}} o^T L^\dagger o,$$

i.e., where the optimization domain for o is now changed. We note that instead of choosing a vector o in this domain, we may equivalently choose two subsets o^+, o^- which then determine the vector o.

In this form, MPP is equivalent to the so-called *(weighted) MAX CUT problem* (*82*) where the weight signs are determined by the off-diagonal entries of $L^\dagger$. If the off-diagonal is non-positive (≤ 0), this problem is known to be NP hard. If the entries have mixed signs, as in the case of MPP, the problem complexity transitions from hard to easy, as described in (*82*). But this transition is not fully understood and we cannot be conclusive on the complexity of MPP in general.

2.2 Scale Invariance

$\delta_{G,o}$ is scale invariant, meaning that equivalent topologies will get the same score even if they have a different number of nodes. Of course, networks with a different number of nodes also

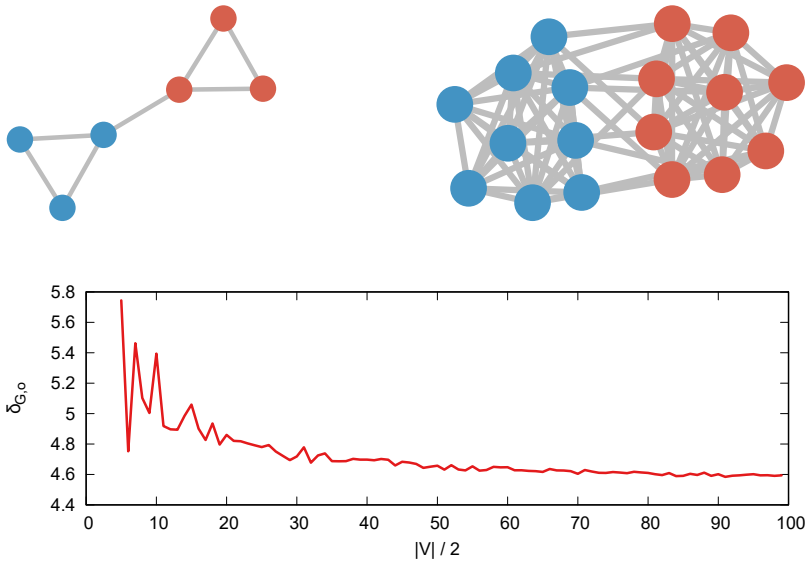

Figure S3: **The polarization values for growing connected cliques**. (Top row) Examples of connected cliques. The node color is proportional to the value in o (-1 blue, $+1$ red). (Bottom row) The value of $\delta_{G,o}$ (y axis) for a given number of nodes in the cliques (x axis).

have a slightly different topology, so the score is not exactly the same, but it will tend to the same limit.

To illustrate, let us consider a relatively simple topology. We have a network G with two cliques, connected by few edges between them. Specifically, the number of edges connecting the two cliques is 5% of the number of edges inside each clique. For instance, if the cliques contain 25 nodes, then there are 15 edges between them ($15 = 0.05 \times 25(25-1)/2$). The edges between the cliques are established randomly. The vector o assigns a value of -1 to all nodes in one of the two cliques, and $+1$ to the nodes of the other.

We can now grow the network by increasing the size of the two cliques one node at a time, while adding edges between the cliques to keep the density constant. Figure S3 shows examples of the generated graphs with their corresponding polarization score $\delta_{G,o}$. We can see that the scores tend to a finite limit. The larger the system, the closer the score gets to the limit, as the randomization of the connections between the cliques plays a smaller role in creating

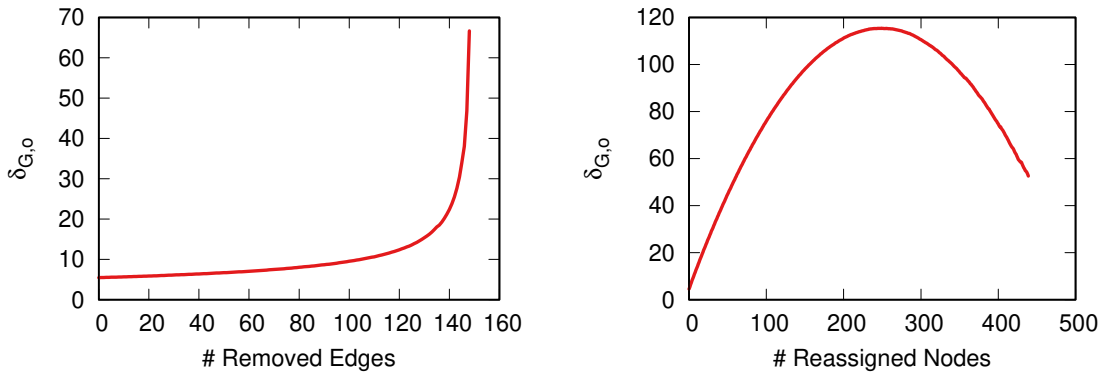

Figure S4: **The polarization values for a two-community system**. $\delta_{G,o}$ (y axis) as we modify the topology of the network. (Left) X axis records the number of inter-community edges removed. (Right) X axis records the number of nodes reassigned from the $+1$ to the -1 community.

fluctuations in the score.

On the other hand, if we modify the actual topology of the network, the score will change even if the scale – the number of nodes – remains constant. We illustrate this with two further simulations.

In the first simulation, we have two equally sized communities, each containing 250 nodes. One community has all positive random values in o while the other has negative random values. We build this with a stochastic blockmodel, ensuring that there are roughly 150 edges between the two communities. We then start removing these edges between the blocks one by one.

Intuitively, this should make the network more polarized, as it becomes harder and harder for an opinion to propagate from one community to the other. This is exactly what we see in the scores of $\delta_{G,o}$, as Figure S4 (left) shows. The first removed edges have little impact, and the impact of each removed edge grows exponentially as we get closer and closer to isolating the two communities.

In the second simulation, we also have two blocks, but we start from an unbalanced situation. One community has 495 nodes while the other only has 5. Every node in one community has

an opinion value of $+1$, while every node in the other community has an opinion value of -1. Then, we select one node at random from the large community and we move it to the small community. We remove all of its connections to its old community and we add connections to the new community at random, preserving its degree (if there are enough candidate neighbors in the community to do so). Its opinion value also flips to align with its new community.

Figure S4 (right) shows the effect. Initially, as we reassign the first nodes, $\delta_{G,o}$ grows. This is expected, since in the starting condition polarization is low as most people have the same opinion and connect to each other. Around the 250th reassignment, $\delta_{G,o}$ plateaus and then starts to decrease again. This shows how $\delta_{G,o}$ peaks when the conflicting communities are roughly the same size. As soon as one community starts getting dominant in size, this lowers the polarization of the system. The process of reassigning nodes stops at around 450 nodes because the network becomes disconnected as a result of randomly removing nodes.

2.3 Effect of Density & Fragmentation

$\delta_{G,o}$ is sensitive to density, because the more dense a network is, the easier it is for an individual to be exposed to a dissenting point of view. This sensitivity can be tested in two ways.

First, we can test it directly. We create a network of 1,680 nodes and divide it in two cliques. The probability of edges being established between the cliques is 5%, i.e., $p_{out} = 0.05$. Then we start removing edges at random from this network, rendering it progressively less dense. Figure S5 (left) shows the growth of $\delta_{G,o}$ as we remove more and more edges.

A second way for a network to become less dense is by fragmenting its community structure. A network with many smaller cliques is less dense than a network with few large ones, even when keeping the number of nodes and p_{out} constant. As before, $\delta_{G,o}$ should grow because people become more and more isolated and it is therefore harder to be exposed to both opposing but also conforming views.

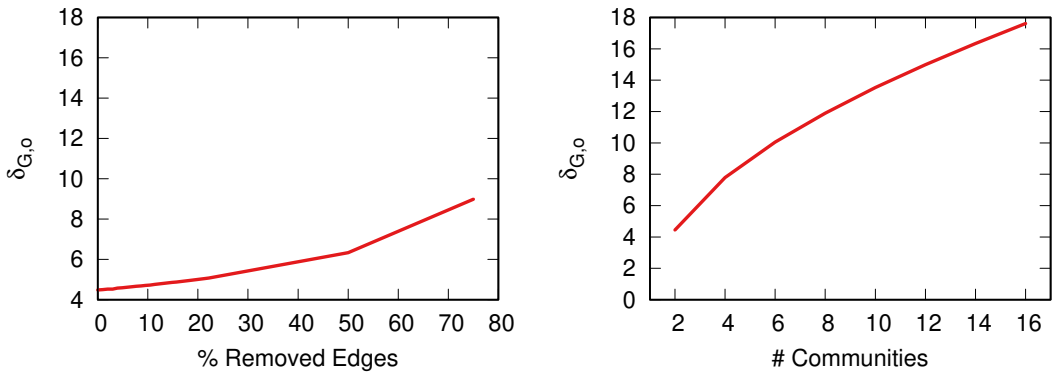

Figure S5: **The polarization values for increasingly sparse networks**. $\delta_{G,o}$ (y axis) as we: (left) remove a fraction of edges at random (x axis); or (right) fragment it into more communities (x axis).

To test this, we take our network of 1,680 nodes. We then divide it into a growing number of cliques, from 2 to 16. All nodes in each clique have an o value of either $+1$ or -1, and there is an equal number of cliques with either opinion value. We then connect cliques with $p_{out} = 0.05$. We only induce an even number of cliques to avoid having one opinion being over-represented in the leftover clique – 1,680 is the smallest number divisible by all even numbers between 2 to 16.

Figure S5 (right) shows the growth of $\delta_{G,o}$ as we increase the fragmentation in communities, which matches this intuition.

3 The Polarization Component Space

3.1 Main Results

To test our measure of polarization, we generate synthetic networks that vary along three different parameters and explore how $\delta_{G,o}$ changes along. In Figures 2 to 4 in the main text, one parameter is varied at a time with the other two fixed but here we consider the full extent of the parameter space.

We explore the three parameters as follows. The first parameter (μ) regulates the divergence

of opinions. Each side of the opinion vector o peaks at $\pm\mu$. Thus, if μ is 0.2, the left side of o peaks at -0.2 and the right side peaks at 0.2. For this reason, the higher the μ value, the more polarization there is, as Figure 2 in the main paper shows.

The second parameter (p_{out}) regulates the structural component by determining the probability of a node in a community to connect to a node in a different community – as we describe in the Materials and Methods section in the main paper. The lower p_{out}, the more polarization there is, as nodes get progressively more isolated from opinions in outside communities.

Finally, for the mesoscale interplay, we use the parameter n which indicates how many communities a community can link to. We pick possible neighboring communities among the n closest ones in the opinion spectrum o. For example, if $n = 7$, all communities can link to each other because with 8 communities there are only 7 possible neighbors. If $n = 1$, a community can only link to its most immediate neighbor. The lower the n value, the more polarization there is, because communities get progressively more isolated from potential connections with nodes in a different portion of the opinion vector o.

Table S1 shows all values of $\delta_{G,o}$ across the three parameters. It reports the average of 25 randomly initialized runs. The standard errors of these means are in the order of 1% of the mean value. This shows that these scores are stable and reliable as there are no wild fluctuations from small differences in random initialization. Table S1 shows a smooth transition across these parameters with no significant discontinuities. This supports our claim in the main paper that $\delta_{G,o}$ is sensitive to all these factors.

3.2 Necessary Conditions for Polarization

While the opinion and structural component – and their interplay – are all important to quantify polarization, there can be some degenerate cases in which one or two parameters out of three dominate the value of $\delta_{G,o}$. The most straightforward case is one in which there is no polariza-

μ	p_{out}	n						
		7	6	5	4	3	2	1
0.0	0.0085	2.56	2.75	3.10	3.54	4.33	5.80	9.89
	0.0042	3.47	3.81	4.43	5.23	6.69	9.89	19.73
	0.0024	4.55	5.01	5.81	7.00	9.13	13.93	28.91
	0.0012	6.38	7.09	8.14	9.93	13.48	20.48	43.16
	0.0006	9.20	10.41	11.82	14.45	19.16	30.73	66.46
	0.0003	13.22	14.70	17.21	20.71	28.68	44.81	100.27
0.2	0.0085	3.56	3.95	4.41	5.03	6.11	8.23	14.40
	0.0042	4.94	5.42	6.25	7.38	9.54	14.10	28.20
	0.0024	6.55	7.10	8.33	9.93	13.01	19.80	42.43
	0.0012	8.99	10.16	11.61	14.32	19.11	29.24	62.86
	0.0006	13.15	14.88	16.65	20.36	27.82	44.28	95.31
	0.0003	18.81	20.92	24.42	28.97	41.20	63.45	144.62
0.4	0.0085	5.60	6.01	6.75	7.77	9.58	13.27	23.33
	0.0042	7.88	8.47	9.64	11.51	15.14	22.79	46.10
	0.0024	10.30	11.07	12.82	15.48	20.76	31.93	67.58
	0.0012	14.46	15.66	18.15	22.32	30.24	47.19	101.40
	0.0006	20.95	23.19	25.94	32.18	43.92	70.57	153.19
	0.0003	29.88	32.43	38.52	45.72	65.51	102.47	232.88
0.6	0.0085	7.86	8.33	9.23	10.66	13.39	18.43	32.27
	0.0042	11.08	11.73	13.26	15.84	21.04	31.88	64.33
	0.0024	14.51	15.37	17.76	21.42	28.93	44.79	94.79
	0.0012	20.55	21.88	25.19	30.76	42.42	65.49	141.78
	0.0006	29.58	32.15	35.65	44.44	61.07	98.63	213.88
	0.0003	42.21	45.08	52.66	62.80	90.67	144.05	328.65
0.8	0.0085	9.79	10.27	11.24	12.96	16.37	22.61	39.81
	0.0042	13.86	14.47	16.20	19.36	25.89	39.22	79.18
	0.0024	18.20	19.08	21.76	26.11	35.44	54.83	116.96
	0.0012	25.70	26.99	30.65	37.53	51.98	80.82	174.90
	0.0006	37.00	39.70	43.70	54.28	74.59	121.25	264.56
	0.0003	52.93	55.57	64.44	76.14	110.66	176.67	405.67

Table S1: **The evolution of polarization scores across all parameters**. Each cell reports the value of $\delta_{G,o}$ for different values of the opinion component (μ, row groups), the structural component (p_{out}, rows), and the opinion-structural mesolevel interplay (n, columns). Each cell background is colored proportionally to the value of $\delta_{G,o}$, from low (bright red) to high (dark red).

tion due to the structural component. This is the case in which the network contains only one node. As soon as we have two nodes we can measure distances between individuals and, as a consequence, we have a non-zero contribution of the structural component to polarization and $\delta_{G,o}$.

The case of no polarization from the opinion component is more interesting. If everyone has the same opinion, polarization should be zero, independent of the social structure. This is the case for $\delta_{G,o}$ by definition, because if o is constant, then $o^+ - o^-$ is zero and the whole formula evaluates to zero. We can show this by creating a network with extreme structural separation – two large cliques connected by few edges – and by drawing o from a normal distribution with decreasing standard deviation.

Figure S6 shows what happens as we approach zero opinion divergence: $\delta_{G,o}$ naturally captures the decrease in diversity of opinions. Both assortativity and RWC_G are blind to those changes and still estimate a constant polarization value even for insignificant differences of opinion. The average neighbor opinion plot could capture the differences, provided the KDE is initialized with the proper parameters – here we fail to see a difference between the last two cases only because we have too large bandwidth. The average influenced set opinion can detect a lack of opinion divergence.

3.3 Secondary Polarization Components

It is possible to consider other aspects as potential components of polarization. For instance, in Figure 1 in the main paper, we move from a random $G_{n,p}$ graph with random opinion assignments to a graph with communities that have assortative opinions. In other words, we make communities whose nodes all occupy a contiguous portion of the opinion spectrum: a node with a given opinion will be embedded in a community whose nodes have a similar opinion to them.

In doing this, we are skipping over a potential polarization component: what happens if we

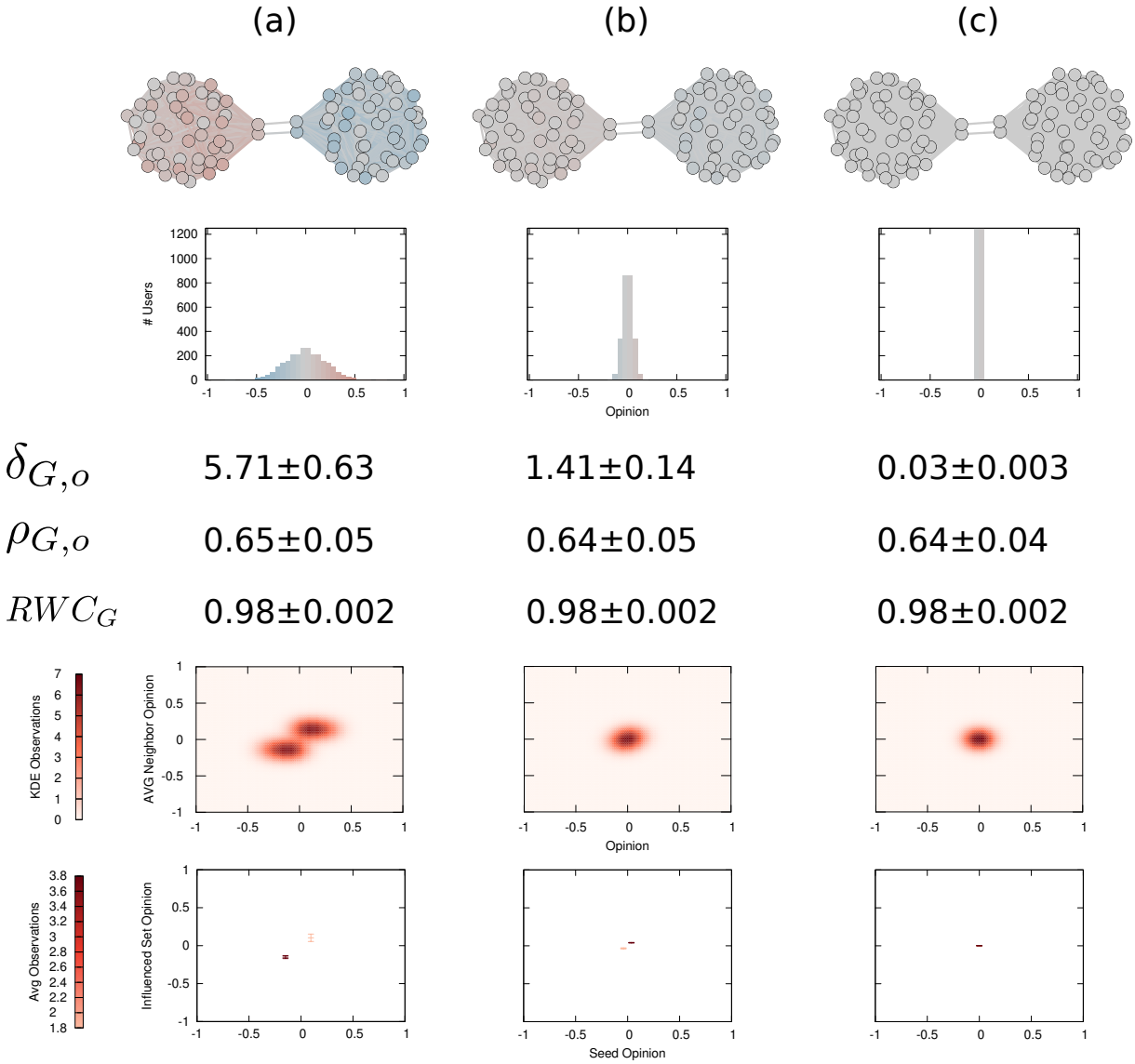

Figure S6: **Polarization for vanishing opinion divergence**. Each row shows (top to bottom): the network structure; the opinion distribution; the average values over 25 runs of $\delta_{G,o}$, $\rho_{G,o}$, and RWC_G with their standard deviations; the density maps of opinions (x axis) and average neighbor opinion (y axis); boxplots of seed opinion (x axis) and average opinion of the influenced set after a SIR propagation (y axis). The opinion distribution is a normal distribution with average zero and standard deviation of (a) 0.2, (b) 0.05, (c) 0.001.

have structural communities, but opinions are still randomly distributed within each community? In such a scenario, having a given opinion tells a node nothing about their companions in

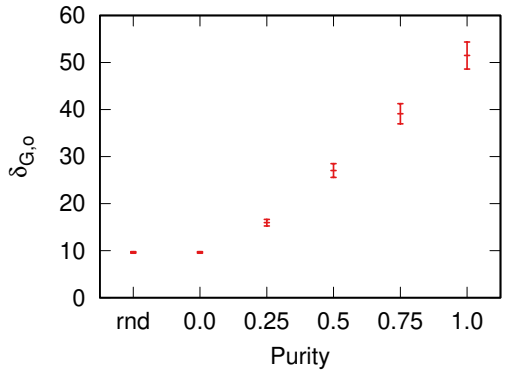

Figure S7: **Polarization and opinion purity**. The average and standard deviation value of $\delta_{G,o}$ (y axis) across 50 independent runs for a random graph and different levels of purity in a graph with strong communities.

the same community.

The reason we do not consider this as a polarization component is because polarization would not change in this extreme scenario. If opinions are distributed randomly, it does not matter whether there are communities or whether the graph is random, because each community in isolation could be considered as a sort of random graph.

To illustrate, we create a graph with 8 communities using an SBM and random o assignment with $\mu = 0.8$ (high opinion divergence) and $p_{out} = 0.0003$ (which should imply high structural separation). After 50 independent runs, we observe $\delta_{G,o} = 9.63 \pm 0.16$. For each run, we also create a randomized version of the graph by performing enough double edge swaps to render the graph effectively random. This random graph has an indistinguishable polarization score: $\delta_{G,o} = 9.64 \pm 0.17$. We do edge swaps rather than generating an alternative random graph because this way we can ensure that the graphs have the exact same number of edges.

The real factor that induces polarization in the system is a community's purity – how well-aligned communities are with o scores. If purity is 1, all nodes in all communities share similar opinion scores. If purity is 0, opinion scores are random. Figure S7 shows how $\delta_{G,o}$ varies as we increase the purity of our communities. This empirical result matches with what we would

expect theoretically: in the low purity networks, everyone is exposed to different views and the polarization score should be low.

Since one cannot have pure communities if there are no communities, we prefer using p_{out} to discuss the structural component of polarization rather than purity.

4 Interpretation of the Units

In the Materials and Methods section in the main paper, we show the relationship between $\delta_{G,o}$ and effective resistance and heat diffusion processes. Here, we show what happens if some of the assumptions we made do not hold, and we provide additional information about these relationships.

4.1 $\delta_{G,o}$ and Effective Resistance in General

If the assumptions that $\sum o_i^+ = \sum o_i^- = 1$ do not hold, and it is not reasonable to normalize the vectors such that we can make them hold by construction, we can still arrive at a similar interpretation. Let $\bar{o} := \frac{1}{|V|}\sum o_i$ be the overall mean opinion. Then we say that two individuals agree if they are on the same "side" of $\bar{o}$ (both larger, or both smaller), and disagree if they are not. Then we assign to every node the variable y equal to $y_i = \frac{(o_i - \bar{o})}{\frac{1}{2}\sum |o_i - \bar{o}|}$, and we consider the positive and negative opinions y^+ and y^- as before. In other words, the variable y_i captures how far an opinion is from the mean, and whether it is larger or smaller than the mean. We now notice that for y^+ and y^- the following holds: by construction – i.e., by removing the mean – we know that the positive sum equals the negative sum, and by normalization with $Z := \frac{1}{2}\sum |o_i - \bar{o}|$, the sums equal one. We can thus define the random variables Y^+ and Y^- with distribution given by y^+ and y^- and find that, in general, the polarization can be interpreted as:

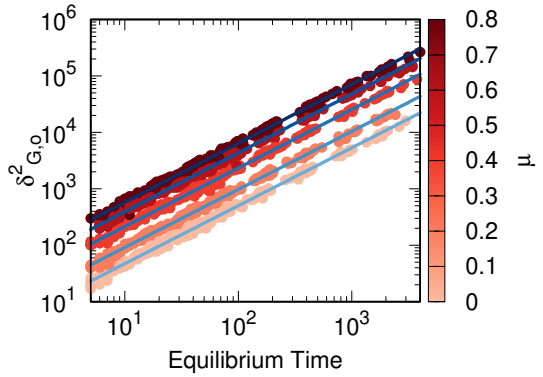

Figure S8: **The relationship between polarization score and heat diffusion equilibrium**. The squared $\delta_{G,o}$ value (y axis) against the unit of time in which heat diffusion reaches equilibrium (x axis). Color encodes the opinion divergence factor μ. The blue line shows the best fits calculated for each μ factor independently.

$$\delta_{G,o} = \left(\frac{1}{2}\sum |o_i - \bar{o}|\right)\sqrt{\mathbb{E}[\omega_{Y^+Y^-}] - \tfrac{1}{2}\mathbb{E}[\omega_{Y^+Y^+} + \omega_{Y^-Y^-}]},$$

so it is still the difference in distance between conflicting individuals and agreeing individuals, but in this case conflicting and agreeing is defined based on whether they are on the same side of the mean opinion or not; similarly the probability of an individual is proportional to their distance to the mean. Furthermore, the factor Z is included, which quantifies how far the opinions are, on average, from the mean opinion.

4.2 Heat Diffusion

In support of our interpretation of $\delta_{G,o}$ as the (square root of the) time it takes for heat to diffuse in the network, we calculate both values on the collections of SBMs used to generate Table S1. That is, we explore all possible values of opinion divergence μ, community interconnectedness p_{out}, and mesolevel organization n.

Figure S8 shows the result. We can see that the opinion divergence factor μ provides a constant multiplicative factor – if the opinion vectors are more extreme it takes a constant additional

amount of time to reach equilibrium. If we take that factor out, $\delta^2_{G,o}$ correlates almost perfectly with the amount of time required for o to have a standard deviation lower than $\epsilon = 2 \times 10^{-4}$. The choice of ϵ does not matter, provided it is a small value.

All the power laws we calculated independently for each μ value have an $R^2 \sim 0.996$, and an exponent approximately equal to 1, showing a linear relationship between $\delta^2_{G,o}$ and equilibrium time.

5 Relations to Network Covariance

As we show in the Materials and Methods section in the main paper, $\delta_{G,o}$ *de facto* measures the Generalized Euclidean distance (*24*) between two vectors o^+ and o^- on the nodes of a given graph G. We show here that this polarization measure can also be seen as a measure of covariance between the opposing opinion distributions, in the sense of (*52*). If V_+ are the nodes with positive opinion $o_i > 0$ and V_- the nodes with negative opinion, then we define the following joint distribution P between pairs of nodes

$$(P)_{ij} := \begin{cases} \frac{1}{2} o_i o_j \text{ if } i \in V_+ \text{ and } j \in V_-, \\ \frac{1}{2} o_i o_j \text{ if } i \in V_- \text{ and } j \in V_+, \\ 0 \text{ otherwise,} \end{cases}$$

which can also be written as a $|V| \times |V|$ matrix as:

$$P = \frac{1}{2} \left(o^+ o^{-T} + o^- o^{+T} \right).$$

where $o^\pm$ are the vectors with only positive (negative) opinions. This is a distribution which selects a random pair of nodes with opposing opinions, with probability of picking each node proportional to their opinion. So if $o_v = 2o_w$, then v is twice as likely to be picked as w. The covariance of this joint distribution, with respect to the effective resistance matrix $\Omega = (\omega_{ij})$, is

calculated according to (*52*) as

$$
\begin{aligned}
\mathrm{cov}_\omega(P) &= \frac{1}{2}\,\mathrm{tr}\left(\left(P\mathbf{1}\mathbf{1}^T P - P\right)\Omega\right) \qquad \text{(with } \mathbf{1} = (1,\ldots,1)\text{)} \\
&= \frac{1}{2}\mathbf{1}^T P\Omega P\mathbf{1} - \frac{1}{2}\,\mathrm{tr}\left(P\Omega\right) \\
&= \frac{1}{2}\left(\frac{1}{4}o^{+T}\Omega o^{+} + \frac{1}{4}o^{-T}\Omega o^{-} + \frac{1}{2}o^{+T}\Omega o^{-}\right) - \frac{1}{2}\left(o^{+T}\Omega o^{-}\right) \\
&= \frac{1}{2}\left(\frac{1}{4}o^{+T}\Omega o^{+} + \frac{1}{4}o^{-T}\Omega o^{-} - \frac{1}{2}o^{+T}\Omega o^{-}\right) \\
&= \frac{1}{8}\left(o^{+} - o^{-}\right)^T \Omega\left(o^{+} - o^{-}\right) \\
&= -\frac{1}{4}\left(o^{+} - o^{-}\right)^T L^{\dagger}\left(o^{+} - o^{-}\right) \qquad \text{(following the definition of effective resistance)} \\
&= -\frac{1}{4}\delta^2_{G,o}
\end{aligned}
$$

This shows that $\delta_{G,o}$ is proportional to minus the covariance of distribution P. Our polarization measure captures the covariance of a pair of nodes with opposing opinions – a high covariance signifies that they are close together in the network, while a small covariance indicates that they are far apart. This is in line with how our measure intends to quantify polarization.

6 Multidimensional Polarization

6.1 Formulation

As introduced, our polarization measure $\delta_{G,o}$ is limited to measuring the polarization that arises in a social network on a binary or unidimensional topic. Any individual can have one of two opinions $\{+,-\}$, with a given strength or conviction recorded in the vector o. While useful in the case of debates with {for, against} stances or in two-party political systems such as the United States, this approach cannot handle the case of multidimensional issues. Here, we briefly discuss how our binary measure can be extended to deal with these more complex cases.

Let us assume that any individual can support an opinion in a set of options $\{A_1, A_2, \ldots, A_k\}$ – these can for instance be k political parties – with a certain strength given by the vector o^{A_i}

that records for each individual how strongly they support opinion A_i. An overall measure of polarization can then be obtained as an average over all possible pairwise conflicts, as

$$\delta_{G,o} = \sqrt{\binom{k}{2}^{-1} \sum_{i=1}^{k} \sum_{j=i+1}^{k} (o^{A_i} - o^{A_j})^T L^{\dagger} (o^{A_i} - o^{A_j})}.$$

If all the opinion vectors are normalized, then the same derivation we use in the Methods section in the main paper can be used here, which means that this multidimensional polarization can intuitively still be interpreted as a measure of the difference in (effective resistance) distance between disagreeing and agreeing individuals. We note that other extensions to multidimensional measures might be possible, for instance assigning a different weight to the different pairwise conflicts.

6.2 Behavior

We can show that the multidimensional version of our measure behaves in the same way as the original measure across the various components and it is equally intuitive. To do so, we repeat our synthetic experiments. Rather than having two vectors o^+ and o^-, we now have eight different opinions since our synthetic networks contain eight communities. Each community i contains only nodes with nonzero values for opinion A_i and zero for all other opinions A_j. The opinion vectors are constructed with the same procedure outlined for the simple unidimensional case, with μ representing the mode value. We construct the network with the same algorithm, so also parameters p_{out} and n retain their meaning.

By running the experiments, we can create an equivalent of Table S1 for the multidimensional case. Table S2 is the result. We can see that the overall behavior of the multidimensional measure is the same as the unidimensional one. We can then conclude that the interpretation of the two measures is the same, and that we can apply our extension to the multidimensional case.

μ	p_{out}	n 7	6	5	4	3	2	1
	0.0085	1.24	1.25	1.27	1.34	1.39	1.54	2.09
	0.0042	1.60	1.60	1.69	1.79	1.91	2.40	4.07
0.0	0.0024	2.02	2.06	2.08	2.28	2.57	3.31	6.00
	0.0012	2.75	2.83	2.98	3.21	3.66	4.73	8.77
	0.0006	3.78	3.90	4.15	4.52	5.31	6.78	13.38
	0.0003	5.45	5.65	5.91	6.57	7.59	10.38	19.86
	0.0085	1.78	1.78	1.83	1.90	1.97	2.19	3.14
	0.0042	2.29	2.30	2.40	2.56	2.79	3.40	5.84
0.2	0.0024	2.88	2.97	3.00	3.29	3.81	4.71	8.62
	0.0012	3.91	4.02	4.27	4.54	5.21	6.97	13.16
	0.0006	5.63	5.61	6.08	6.59	7.66	10.12	19.64
	0.0003	7.87	8.24	8.58	9.63	11.14	15.34	29.61
	0.0085	2.80	2.81	2.86	3.00	3.15	3.57	5.25
	0.0042	3.70	3.81	3.93	4.28	4.69	5.84	10.20
0.4	0.0024	4.83	4.89	5.16	5.54	6.33	8.11	14.85
	0.0012	6.67	6.93	7.19	7.90	9.35	11.80	22.70
	0.0006	9.58	9.88	10.62	11.31	13.25	17.70	33.97
	0.0003	13.65	13.90	14.86	16.92	18.94	26.21	51.17
	0.0085	3.93	3.97	4.05	4.20	4.48	5.14	7.56
	0.0042	5.46	5.48	5.68	6.14	6.77	8.62	14.89
0.6	0.0024	7.01	7.19	7.50	8.12	9.38	11.97	22.21
	0.0012	10.01	10.22	10.58	11.55	13.49	17.56	33.84
	0.0006	14.10	14.45	15.30	16.51	19.62	25.88	50.81
	0.0003	20.14	20.86	21.93	24.53	28.28	38.41	75.07
	0.0085	4.89	4.94	5.04	5.24	5.60	6.50	9.69
	0.0042	6.85	6.88	7.21	7.73	8.66	11.10	19.19
0.8	0.0024	8.97	9.10	9.50	10.37	11.90	15.30	28.24
	0.0012	12.70	13.07	13.64	14.79	17.27	22.81	43.36
	0.0006	18.08	18.42	19.47	21.05	24.95	33.07	64.88
	0.0003	25.81	26.72	28.00	31.80	36.27	49.79	96.50

Table S2: **The evolution of multidimensional polarization scores across all parameters**. Same legend as Table S1.

The only difference is that the absolute values of the measure tend to be smaller. This might be due to the fact that the average of all pairwise conflicts decreases as we increase the number of opinions, as by construction some opinions are closer to each other. However, we leave this

investigation for future work, along with the consideration that taking the average of all pairwise conflicts might not necessarily be the best way to expand our measure.

7 Sensitivity to Measurement Errors

While we have assumed that the network G and opinion vector o are given as inputs from which we calculate the polarization, it is important to acknowledge that this data is at best an approximation of the real underlying opinions and social network connections. In particular, measurement errors and other inherent difficulties in obtaining and representing such complex data might mean that the available opinion vector o only records the opinions approximately. This can be modeled as an additive measurement error $o = \hat{o} + \epsilon$ where $\hat{o}$ is the real opinion vector and ϵ the error vector. If ϵ is large relative to the measurements, i.e., the errors are of the same order as the opinions (and uncorrelated), then there is no real hope for a measure that accurately represents the real polarization. If, on the other hand, the error vector is relatively small, we would hope that the polarization measure is a good approximation.

This is the case for our proposed polarization measure. If we assume that $\|\epsilon\|$ is small, then the polarization $\delta_{G,o}$ calculated based on the available opinion measurements o is close to the polarization $\delta_{G,\hat{o}}$ calculated based on the underlying true opinions $\hat{o}$. In the derivation below, we show that

$$|\delta_{G,o} - \delta_{G,\hat{o}}| \leq \|\epsilon\| \mu_{\min}^{-1/2}(L), \tag{1}$$

where $\mu_{\min}(L)$ is the smallest nonzero Laplacian eigenvalue, which is a fixed number for a given network. The error in our polarization measure is thus at most proportional to the measurement error $\|\epsilon\|$.

The bound (1) follows from the assumption that $\|\epsilon\|$ is small:

$$\begin{aligned}
\delta_{G,\hat{o}} - \delta_{G,o} &= \sqrt{(\hat{o}^+ - \hat{o}^-)^T L^\dagger (\hat{o}^+ - \hat{o}^-)} - \sqrt{(\hat{o}^+ - \hat{o}^- + \epsilon)^T L^\dagger (\hat{o}^+ - \hat{o}^- + \epsilon)} \\
&= \delta_{G,\hat{o}} \left(1 - \sqrt{1 + \frac{2\epsilon^T L^\dagger (\hat{o}^+ - \hat{o}^-) + \epsilon^T L^\dagger \epsilon}{\delta_{G,\hat{o}}^2}} \right) \\
&= \delta_{G,\hat{o}} \left(1 - \sqrt{1 + \frac{2\epsilon^T L^\dagger (\hat{o}^+ - \hat{o}^-)}{\delta_{G,\hat{o}}^2}} \right) \quad \text{(step 1)} \\
&= \delta_{G,\hat{o}} \left(1 - \left[1 + \frac{\epsilon^T L^\dagger (\hat{o}^+ - \hat{o}^-)}{\delta_{G,\hat{o}}^2} \right] \right) \quad \text{(step 2)} \\
\Rightarrow |\delta_{G,\hat{o}} - \delta_{G,o}| &= \left| \frac{\epsilon^T L^\dagger (\hat{o}^+ - \hat{o}^-)}{\delta_{G,\hat{o}}} \right| \\
&\leq \sqrt{\epsilon^T L^\dagger \epsilon} \quad \text{(step 3)} \\
&\leq \|\epsilon\| \mu_{\min}^{-1/2}(L) \quad \text{(step 4)}.
\end{aligned}$$

In step 1, we use the assumption that ϵ is small, such that $\epsilon^T L^\dagger \epsilon$ is negligible with respect to the other terms. In step 2, we use the Taylor expansion $\sqrt{1+x} = 1 + \frac{x}{2} + O(x^2)$ in combination with the small-error assumption. In step 3, we invoke the Cauchy-Schwarz inequality $\epsilon^T L^\dagger (\hat{o}^+ - \hat{o}^-) \leq \sqrt{\epsilon^T L^\dagger \epsilon} \sqrt{(\hat{o}^+ - \hat{o}^-)^T L^\dagger (\hat{o}^+ - \hat{o}^-)}$. Step 4, finally, follows from the fact that the largest eigenvalue of $L^\dagger$ is the inverse of the smallest nonzero eigenvalue of the pseudoinverse Laplacian, as $\mu_{\max}(L^\dagger) = \mu_{\min}^{-1}(L)$.

8 Twitter & Congress Data

8.1 Summary Statistics

Table S3 shows some summary statistics for the networks extracted from the Twitter data. The table shows, among other things, how well one could partition the network by assigning a node to a community depending on the sign of its o value. We estimate this by calculating the modularity of the partition, which compares the number of edges inside each community with

Network	$\lvert V\rvert$	$\lvert E\rvert$	ℓ	Δ	Q
Obama	204	1,377	2.762	0.183	0.027
Gun Control	1,092	22,471	2.813	0.273	0.176
Abortion	2,211	55,328	2.773	0.190	0.343
VP Debate	5,407	116,249	2.967	0.073	0.364
Second Debate	4,697	94,497	3.115	0.084	0.474
Election	3,965	40,443	3.493	0.088	0.329

Table S3: **Summary statistics of the Twitter debate networks**. For each network, we report: the number of nodes $|V|$, the number of edges $|E|$, the average shortest path length ℓ, the transitivity of the network Δ, and the modularity values Q we would get by bisecting the networks in two communities identified by having a o value of a given sign.

the one we would get with a random network with the same degree distribution as the original one but no communities (*83*). This is another rough estimate of structural separation.

Table S4 shows the same summary statistics for all the US House of Representatives networks. Consistently with what we discuss in the main paper, we can see a noticeable increase in modularity starting at the 99th Congress. We can also see that the share of congressmen with more extreme opinion scores steadily increases over time.

8.2 Polarization Components Correlation

Since our definition of polarization is made of two components – opinion and structure – and their mesoscale interplay, we are implicitly assuming that there is a correlation between them. Here, we test whether this assumption holds on the real-world data gathered on Twitter and the US House of Representatives.

We estimate the opinion component by calculating the average opinion of each side and then averaging their absolute values. For instance, in the 116th Congress, the average positive DW-NOMINATE score was 0.5, while the average negative score was -0.37. We thus infer that $\mu_{116} = (0.5 + 0.37)/2 = 0.43$.

The structural component p_{out} comes from counting the number of edges across communi-

Congress	$\|V\|$	$\|E\|$	ℓ	Δ	Q	$\|o\| > .5$
81	423	39,735	1.592	0.761	0.309	0.070
82	425	43,446	1.549	0.782	0.315	0.066
83	419	36,988	1.625	0.782	0.308	0.067
84	421	40,632	1.576	0.749	0.311	0.076
85	415	38,378	1.590	0.733	0.295	0.082
86	426	42,151	1.563	0.798	0.283	0.095
87	418	39,344	1.598	0.785	0.325	0.083
88	417	39,870	1.590	0.810	0.359	0.086
89	417	39,923	1.588	0.820	0.275	0.091
90	419	36,292	1.638	0.751	0.285	0.081
91	398	33,925	1.613	0.717	0.169	0.087
92	407	35,908	1.599	0.738	0.223	0.091
93	413	40,908	1.559	0.757	0.214	0.104
94	411	41,328	1.549	0.791	0.219	0.105
95	413	41,547	1.546	0.771	0.200	0.108
96	419	43,847	1.540	0.782	0.260	0.116
97	411	39,776	1.571	0.758	0.288	0.119
98	427	49,006	1.499	0.843	0.298	0.114
99	425	46,155	1.558	0.880	0.351	0.120
100	421	45,810	1.556	0.899	0.341	0.120
101	425	46,410	1.562	0.876	0.321	0.130
102	431	48,138	1.546	0.907	0.346	0.131
103	430	48,355	1.516	0.950	0.405	0.155
104	427	45,443	1.621	0.956	0.431	0.189
105	426	44,595	1.700	0.958	0.465	0.194
106	427	44,713	1.593	0.930	0.458	0.192
107	427	44,693	1.590	0.947	0.467	0.203
108	429	45,463	1.807	0.987	0.486	0.203
109	429	45,607	1.739	0.985	0.487	0.220
110	422	44,340	1.821	0.989	0.481	0.219
111	423	46,515	1.544	0.953	0.416	0.214
112	432	47,600	1.520	0.960	0.435	0.286
113	424	44,899	1.835	0.990	0.485	0.295
114	433	47,662	1.546	0.992	0.461	0.315
115	432	47,079	1.587	0.981	0.470	0.327
116	427	45,900	1.546	0.988	0.475	0.296

Table S4: **Summary statistics of the US House of Representatives networks**. Same legend as Table S3. The additional $|o| > .5$ columns reports the share of congressmen with an opinion score outside the $[-0.5, 0.5]$ interval.

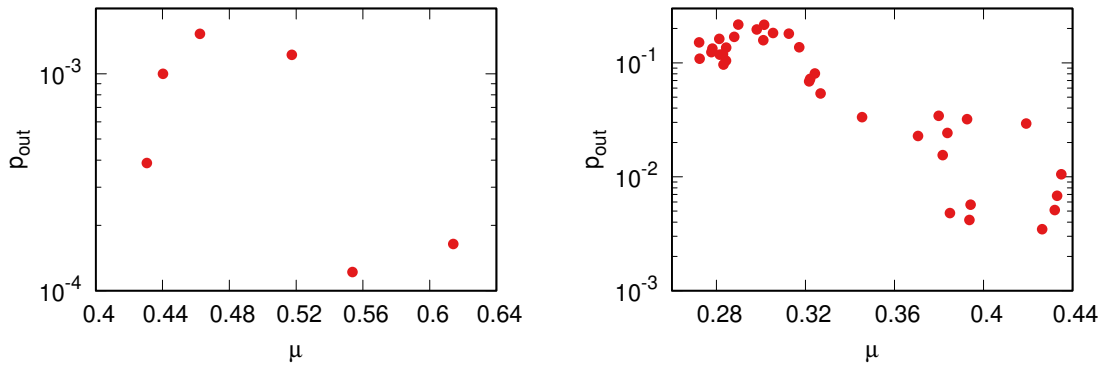

Figure S9: **The correlation between the opinion and structural components of polarization on real-world networks**. Opinion divergence on the x axis, structural separation on the y axis. (Left) Twitter, (Right) US House of Representatives.

ties over the potential number of possible edges across communities, i.e., the number of node pairs belonging to different communities. We employ a Monte Carlo Markov Chain approach for the inference of stochastic block models (*84*) (SBM).

We are unable to estimate the opinion-structural mesoscale interplay, which we indicate with n. This is because real-world data does not organize as neatly as our synthetic experiments. Even if there are multiple communities, they will always have some degree of interconnectedness. It follows that there is no easy way to calculate n, and since finding an appropriate way goes beyond the scope of this paper, we skip the estimation of this parameter.

Figure S9 shows the relationship between μ and p_{out} on all real-world networks we study. We observe the expected negative relationship. High μ correspond to high opinion divergence which we expect to cause a high structural separation by isolating the communities, causing a low value for p_{out}, which is the probability of edges appearing across communities.

The relationship is exponential (note the logarithmic scale in the y axis). In the US House of Representatives network, we obtain a -0.83 Pearson correlation coefficient and a -0.78 Spearman correlation coefficient, both significant as $p < 0.001$. We do not obtain significant coefficients for the Twitter networks, but this is most likely due to the small sample size. The

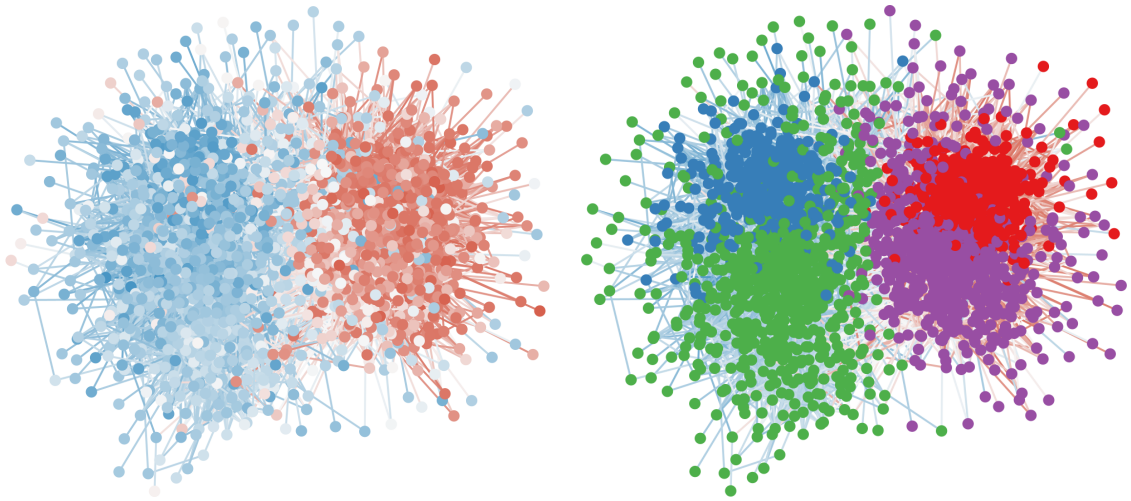

Figure S10: **The communities in the Twitter abortion network**. In both networks, nodes are Twitter users, connected by an edge if the two nodes interacted. The edge color is the average opinion value of the two connected nodes. The two networks have the exact same layout. (Left) Node color is the o value of the node (from -1 dark blue, to $+1$ dark red). (Right) Node color is the inferred community. Edge color is the same as in the left network.

coefficients for the Pearson and Spearman correlations are still negative and equal to -0.45 and -0.6, respectively.

We can conclude that indeed the structural and opinion components of our polarization definition are correlated and reinforce each other.

8.3 Communities in the Twitter Abortion Debate Network

In the main paper we argue that one of the reasons for the Twitter abortion debate network to score higher $\delta_{G,o}$ values is the fact that the network has four nested communities rather than two flat ones. Here, we provide support for this statement. To do so, we infer the community structure of the network using the same method we used in the previous section.

Figure S10 shows the result. On the left, we depict the network as it is in Figure 5 of the main paper. We can see that both the blue and the red community in the network have subcommunities, with brighter and lighter hues, showing a distinction in interaction patterns

	blue	green	purple	red
blue	0.0981	0.0078	0.0009	0.0018
green	0.0078	0.0391	0.0019	0.0003
purple	0.0009	0.0019	0.0614	0.0219
red	0.0018	0.0003	0.0219	0.1393

Table S5: **The edge connection probabilities for the communities in the Twitter abortion network**. Each cell and its color reports the probability that a node in one block connects to a node in another block.

Community	AVG o
blue	-0.52
green	-0.40
purple	0.40
red	0.56

Table S6: **The average opinion value for the Twitter abortion communities**. Each cell – and its color – reports the average o values for the nodes in each block in the Twitter abortion network.

between moderates and extremists. This is confirmed by the SBM community inference, which is shown on the right of Figure S10. Here, we can see that both the left and the right side are further split in two communities which, upon visual inspection, overlap with the brighter and darker hues on either side.

This is not just a visual artifact. Table S5 reports the probability of a block community to connect to another block community. We can see that the diagonal entries for communities paired on the same side of the network are higher than the ones with their community companion. This means that, e.g., the green community is more tightly knit with itself than with the blue community. If the network had only two communities and this hierarchical division was an artifact, there should be no difference in the edge probabilities between the green and blue block. The fact that we see such a difference shows that the network indeed has a mesoscale organization in four communities.

The communities also have distinctive opinion values. Table S6 shows the average o value

for the nodes in each community. Again, we can see noticeable differences, showing that indeed each community is distinct from the others.

9 Supplementary Material Captions

Supplementary File F1. This is a zip file containing all the code necessary to reproduce the results in the paper. Data available at the links provided in the README file. The archive is also available at `https://www.michelecoscia.com/?page_id=2105`.