# Quantifying Polarization in Models of Opinion Dynamics

Christopher Musco
cmusco@nyu.edu
New York University

Indu Ramesh
ir914@nyu.edu
New York University

Johan Ugander
jugander@stanford.edu
Stanford University

R. Teal Witter
rtealwitter@nyu.edu
New York University

## ABSTRACT

It is widely believed that society is becoming increasingly polarized around important issues, a dynamic that does not align with common mathematical models of opinion formation in social networks. In particular, measures of polarization based on opinion variance always *decrease* over time in models like the popular DeGroot model. Complementing recent work that seeks to resolve this inconsistency by modifying opinion models, we instead resolve the inconsistency by proposing changes to *how polarization is quantified*.

We present a natural class of group-based polarization measures that capture the extent to which opinions are clustered into distinct groups. Using theoretical and empirical arguments, we show that these group-based measures display interesting, non-monotonic dynamics, even in the simple DeGroot model. In particular, for natural social networks, group-based metrics can increase over time, and thereby correctly capture perceptions of increasing polarization.

Our results build on work by DeMarzo et al., who introduced a group-based polarization metric based on ideological alignment. We show that a central tool from that work, a limit analysis of individual opinions under the DeGroot model, can be extended to the dynamics of other group-based polarization measures, including established statistical measures like bimodality.

We also consider local measures of polarization that operationalize how polarization is perceived in a network setting. In conjunction with evidence from prior work that group-based measures better align with real-world perceptions of polarization, our work provides formal support for the use of these measures in place of variance-based polarization in future studies of opinion dynamics.

## 1 INTRODUCTION

Polarization of individual opinions and beliefs has become a topic of intense interest in recent years, especially in relation to politics [9], and politically sensitive issues like climate change [44] and public health [28, 34]. Polarization is often believed to threaten social stability; for example, it has been blamed for legislative deadlock [6, 7], decreased trust and engagement in the democratic process [37, 39], and hindered responses to crises like the COVID-19 pandemic [28]. In response to its impact, there is growing interest in using mathematical models of *opinion dynamics* to formally study how polarization arises and evolves. Such models provide simple rules for how an individual's opinion on a topic changes in response to influence from that individual's social connections. Mathematical models of opinion dynamics offer a useful abstraction for studying important real-world phenomena [5]. For example, they have been used to study the impact of biased assimilation [12, 32] and the effect of outside actors[1] on polarization [8, 24, 31, 46, 50].

To continue effectively leveraging such models, we first need to address a basic and important question:

*How should the broad and imprecise concept of polarization be quantified in mathematical models of opinion dynamics?*

Surprisingly, this question has received little attention. Most prior work defaults to quantifying polarization based on the overall *variance* of societal opinions (opinions are typically encoded as real valued numbers) [8, 17, 24, 43]. While mathematically convenient, any variance-based approach faces a basic challenge: standard models of opinion formation in social networks, like the ubiquitous DeGroot learning model [15], predict gradual convergence of opinion variance towards zero over time. This inevitable decrease stands in contradiction to the fact that, qualitatively, polarization is considered to exhibit far more interesting dynamics. For example, it is widely believed that polarization is currently *increasing* across the globe on a variety of issues [9, 44], and that its dynamics have been impacted by forces such as the rise of social media [48].

### 1.1 Our Approach and Main Results

Given the shortcomings of opinion variance as a measure of polarization, we address the central question of how to best quantify polarization by evaluating metrics through a *dynamic lens*. In particular, our goal is to identify natural metrics whose dynamics under simple models of opinion formation, like the DeGroot model, agree with observed dynamics of polarization in society.

Towards this end, we introduce a class of *group-based* metrics for polarization. We use "group-based" to reference the idea of a measure that is high when there are well-separated groups of individuals with different opinions. We formalize this notion in Section 2 by assuming shift- and scale-invariance, hich are properties that naturally align with axiomatic treatments of clustering [38]. For now, we leave "group-based" as an intuitive definition and illustrate with an example. Consider the following opinion vectors on a six node social network (each entry is one individual's opinion):

$$\mathbf{a} = [-1, -.6, -.2, .2, .6, 1] \qquad \mathbf{b} = [.5, .5, .5, -.5, -.5, -.5]$$

While $\mathbf{a}$ has larger variance than $\mathbf{b}$ (2.8 vs. 1.5), $\mathbf{b}$ would have higher polarization under a group-based measure, since opinions are more clearly clustered into two groups. This cluster structure could be quantified, for example, by any statistical measure of bimodality, like Sarle's bimodality coefficient (see Def. 3 for more details). Sarle's coefficient is equal to $(\gamma^2 + 1)/\kappa$, where $\gamma$ is skewness and $\kappa$ is kurtosis, and evaluates to 0.58 for $\mathbf{a}$, but a higher value of 1 for $\mathbf{b}$.

In this work we explore two main types of group-based polarization measures:

**Statistical Measures (Section 4)** This class includes functions that, like Sarle's bimodality coefficient, measure group structure

[1] Actors like news agencies, social media companies, advertisers, and governments can influence opinions in a social network by swaying the strength of social connections, possibly by promoting or hiding social media posts, creating fake user accounts and content, or running advertisements. By modeling these actions mathematically within an opinion dynamics framework, researchers can better understand how susceptible networks are to adversarial attacks [3, 26] and how "filter bubbles" emerge [11, 48].

in an opinion distribution by looking at moments beyond the second (i.e., beyond variance). For example, the bimodality coefficient incorporates third and fourth moment information.

**Local Measures (Section 5)** This class includes metrics that take into account local social connections on perceptions of group-structure. For example, we study *local agreement*, defined as the average percentage of an individuals social connections who agree on a particular topic (i.e., have an opinion on the same side of the mean). Networks with high local agreement may appear more polarized to individuals, who feel isolated in opinion bubbles.

We show that these group-based measures behave very differently than variance-based measures, exhibiting interesting, non-monotonic dynamics even in the simple DeGroot model. In particular, we prove that any group-based measure converges to a value that *depends on the structure of the underlying social network* governing the opinion dynamics. So, instead of always converging to zero like variance-based measures, group-based measures can *increase* over time for certain networks. Our work builds on a result of DeMarzo, Vayanos, and Zwiebel [16], who study a group-based metric that we call "ideological alignment". Their work is based on an analysis of the limiting behavior of each individual's divergence from the mean opinion under the DeGroot opinion dynamics model. We show that this analysis extends to other measures.

Moreover, we demonstrate empirically that increases in group-based polarization are not only possible, but actually common in natural synthetic and real-world social networks. A sample result for average local agreement measure (discussed in Section 5) appears in Figure 1. We conclude that group-based measures not only have the capacity to model interesting dynamics, but also better align with perceptions of increasing polarization in reality.

For specific group-based measures, we provide additional theoretical support for increasing polarization over time. For example, in Section 4, we give a heuristic analysis for the limiting Sarle's bimodality of opinions in stochastic block-model graphs. We show that the equilibrium value of this measure under the DeGroot dynamics is large for social networks with a small number of communities, a reasonable assumption of real-world networks. In Section 5, we also show that average local agreement in a social network converges to a value that depends on the second eigenvalue of the normalized adjacency matrix $\mathbf{D}^{-1}\mathbf{A}$. Polarization increases to a larger value when this eigenvalue is close to 1, which is empirically the case in a variety of real-world social network graphs.

## 1.2 Conclusions and Recommendations

Our findings provide formal support for using group-based measures to quantify polarization in mathematical models of opinion dynamics. The unrealistic monotonic dynamics of variance-based measures have led past studies to abandon simple opinion models like the DeGroot dynamics, and to adopt alternative, more complicated models to mathematically recover interesting polarization dynamics. For example, the Friedkin-Johnson dynamics [23], bounded confidence model [42], and geometric models have all seen recent attention [25, 31]. A central conclusion of our work is that, alternatively, it may be the *definition of polarization*, not the model, that lacks richness for understanding societal polarization. By turning from variance-based to natural group-based polarization measures, we see interesting dynamics even in the simplest models.

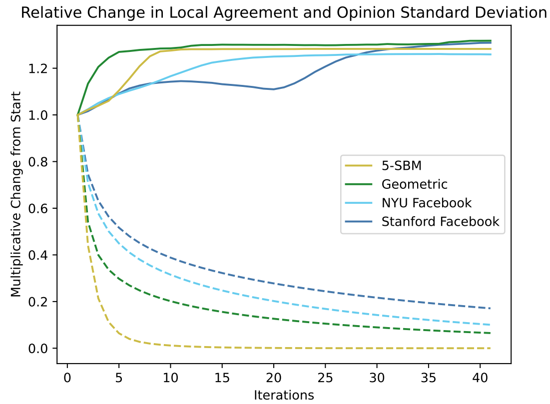


**Figure 1: Number of iterations of DeGroot's model vs. two measures of polarization in synthetic and real-world social networks.[2] Dotted lines plot opinion standard deviation, a variance-based measure, while solid lines plot average local agreement, a natural group-based measure, discussed in Section 5. In contrast to standard deviation, for all networks the group-based measure increases over time, which is consistent with real-world perceptions of how polarization can evolve. This finding provides evidence that the group-based measure may be a more appropriate method for quantifying polarization than the variance-based one.**

Beyond our work, the recommendation to use group-based measures is also supported by empirical evidence that these measures are more aligned with how individuals perceive polarization in society than variance-based metrics [20]. In particular, research suggests that perceived polarization does not correlate with significant *absolute* differences in opinion (which drive overall opinion variance) [10, 41]. Instead, it has been argued that perceptions of polarization stem from perceptions of group-structure [41]. In fact, even the origin of the term "polarization" in the physical sciences suggests a group-based interpretation [1]. While sociological and psychological arguments for how to best quantify polarization are beyond the scope of this paper, the initial alignment between prior work and our findings is promising.

## 1.3 Relation to Prior Work

Prior results have largely defaulted to variance-based measures with the exception of Guerra et al. [29] who introduce a community boundary measure of polarization. As mentioned, our work is most closely related to that of DeMarzo, Vayanos, and Zwiebel [16], whose limit analysis we adopt (providing a new proof in Section 3). The main novelty of our work over [16] is two fold. First, we show that the limit analysis has implications for a wider class of group-based polarization measures beyond "ideological alignment". In particular, it implies convergence of any group-based polarization metric to a graph dependent value, which *can* be large. Second, for different group-based measures, we provide experimental evidence and novel theoretical arguments to show these measures *will* converge to large values for natural social network graphs.

[2] NYU and Stanford are graphs from the Facebook100 data set [51]. 5-SBM is a five community Stochastic Block Model graph on 1000 nodes with intra- and inter-community edge probabilities equal to $p = .1$ and $q = .01$, respectively. Geometric is a proximity graph with 1000 nodes on the unit square with $r = .1$, generated using NetworkX [30].

## 2 PRELIMINARIES

**Graph Notation.** The DeGroot opinion dynamics model studied in this work is based on representing social connections via a weighted, undirected social graph, which we denote $G = (V, E)$. $G$ has $|V| = n$ nodes and $|E| = m$ edges, possibly including self-loops. Let $\mathbf{A}$ be the adjacency matrix of $G$, with $A_{ij} = A_{ji} > 0$ if there is an edge between $i$ and $j$, and $A_{ij} = A_{ji} = 0$ otherwise. Let $\mathcal{N}(i) \subseteq \{1, \ldots, n\}$ denote the set of neighbors of node $i$, which includes all $j$ for which $A_{ij} \neq 0$. If $G$ contains a self-loop at node $i$ then $A_{ii} > 0$ and $i \in \mathcal{N}(i)$. Let $d_i = \sum_{j \in \mathcal{N}(i)} A_{ij}$ denote the degree of node $i$ and let $\mathbf{D}$ be a diagonal matrix containing $d_1, \ldots, d_n$ on its diagonal.

**Vector Sign Normalization.** For a non-zero vector $\mathbf{x}$, let $[\mathbf{x}]^{\pm}$ denote the vector $\text{sign}(x_i) \cdot \mathbf{x}$, where $x_i$ is the first non-zero entry in $\mathbf{x}$. That is, $[\mathbf{x}]^{\pm}$ is equal to either $+\mathbf{x}$ or $-\mathbf{x}$, with the sign chosen to ensure that the first non-zero entry is positive.

### 2.1 DeGroot Opinion Dynamics

Mathematical models of opinion dynamics have been studied for decades in economics [35] , applied math [47], computer science [13], and a variety of other fields [27]. We refer the reader to the survey in [5]. Such models typically view society as a graph, where nodes represent individuals and edges represent social connections of various strength. Simple rules and procedures then define how an individual's opinion on an issue (represented as a single discrete or continuous value, or as a vector) evolves over time.

We focus on one of the earliest and most elegant models of opinion formation: the DeGroot opinion dynamics [15, 22]. This model is based on the idea that opinions on a topic, encoded as continuous values, propagate through the social graph via simple averaging. Nodes incorporate the beliefs of their neighbors into their own opinion over time. We formally describe the model below.

Definition 1 (DeGroot Opinion Dynamics). *Let $G = (V, E)$ be a weighted, undirected graph with $n$ nodes, $m$ edges, adjacency matrix $\mathbf{A}$, and degree matrix $\mathbf{D}$. For time steps $t = 0, 1, \ldots, T$, we associate the nodes of $G$ with an opinion vector $\mathbf{z}^{(t)} \in \mathbb{R}^n$ containing numerical values that represent each individual's current view on an issue. Starting with a fixed vector of initial opinions $\mathbf{z}^{(0)}$, opinions under the DeGroot model evolve via the update:*

$$z_i^{(t+1)} = \frac{1}{D_{ii}} \sum_{j \in \mathcal{N}(i)} A_{ij} z_i^{(t)}, \textit{ or equivalently, } \mathbf{z}^{(t+1)} = \mathbf{D}^{-1}\mathbf{A}\mathbf{z}^{(t)}.$$

The DeGroot model generalizes to directed graphs, but we consider the undirected case for simplicity.

**Convergence to Consensus**. Like many other models of opinion dynamics, it is well known that the DeGroot dynamics converges to *consensus* in the limit. Formally, we have:

Fact 1. *If $G$ is a connected, undirected, non-bipartite graph then,*

$$\mathbf{z}^* = \lim_{t \to \infty} \mathbf{z}^{(t)} = c \cdot \vec{\mathbf{1}} \qquad \textit{where} \qquad c = \sum_{i=1}^{n} \frac{d_i}{\sum_{j=1}^{n} d_j} z_i^{(0)}.$$

*Note that $c$ is equal to the degree-weighted average opinion at time 0.*

As discussed, a common approach to measuring polarization on a single issue at time $t$ is to consider the overall opinion variance:

$$\text{Var}[\mathbf{z}^{(t)}] = \|\mathbf{z}^{(t)} - \text{mean}(\mathbf{z}^{(t)}) \cdot \vec{\mathbf{1}}\|_2^2$$

Of course, if all opinions converge to the same constant value $c$, as guaranteed by Fact 1, opinion variance eventually converges to zero. While this asymptotic observation only speaks to the model's behavior after a very long time, in the short term, variance also tends to decrease *monotonically* with $t$, a fact that can be proven rigorously for regular graphs [12].

### 2.2 Group-based Polarization

In this work, we study group-based polarization metrics, which we broadly define to include any function with three properties: invariant to a shift in mean opinion, invariant to sign flips, and invariant to scaling. Formally:

Definition 2 (Group-based Polarization). *Let $f(G, \mathbf{z})$ be a function that maps an $n$-node graph $G$ and vector of opinions $\mathbf{z} \in \mathbb{R}^n$ to a measure of polarization. Then $f(G, \mathbf{z})$ is "group-based" if:*

*(1) $f(G, \mathbf{z}) = f(-\mathbf{z})$,*
*(2) $f(G, \mathbf{z}) = f(\mathbf{z} + c\vec{\mathbf{1}})$ for any $\mathbf{z}$ and scalar $c$, and*
*(3) $f(G, \mathbf{z}) = f(c\mathbf{z})$ for any non-zero scalar $c$.*

While variance many other measures depend only on $\mathbf{z}$, the local measures studied in Section 5 do depend on the underlying social network which is why we include $G$ as a parameter to $f$. Variance-based measures of polarization satisfy properties (1) and (2) of Defn. 2, but not (3). The last property reflects the fact that group structure should depend on *relative* differences in opinions instead of absolute differences. I.e. an opinion vector would be considered polarized if we have two groups whose mean opinions are further apart than opinions within each group, regardless of absolute opinion difference. The resulting requirement of scale-invariance has also appeared in axiomatic treatments of clustering objectives [38], which are closely related to group-based measures.

## 3 LIMIT ANALYSIS

While Fact 1 implies that any variance-based measure of polarization converges to zero, this is not true for the group-based measures. Since they are both shift and scale invariant, they are insensitive to both the mean opinion (this is also true of variance-based measures) and to constant rescaling of the opinions. As such, to analyze these measures, we prove a separate convergence result for the mean-centered, normalized opinion vector, which was also observed in [16]. In particular, we study the vector:

$$\frac{\mathbf{z}^{(t)} - \text{mean}(\mathbf{z}^{(t)}) \cdot \vec{\mathbf{1}}}{\|\mathbf{z}^{(t)} - \text{mean}(\mathbf{z}^{(t)}) \cdot \vec{\mathbf{1}}\|_2}.$$

We show that, under mild conditions, in the DeGroot model this vector converges to a fixed function of the second eigenvector of the normalized social network adjacency matrix, $\mathbf{D}^{-1}\mathbf{A}$. We give a full proof below, which uses simpler arguments than [16].

Theorem 1. *Let $G$ be a connected graph with adjacency and degree matrices $\mathbf{A}$ and $\mathbf{D}$. Let $\mathbf{v}_1, \ldots, \mathbf{v}_n$ and $\lambda_1, \ldots, \lambda_n$ be the eigenvectors and eigenvalues of $\mathbf{D}^{-1}\mathbf{A}$, in order of magnitude. I.e., $|\lambda_1| \geq \ldots \geq |\lambda_n|$ . Let $\mathbf{z}^{(0)}, \ldots, \mathbf{z}^{(t)}$ be a sequence of opinion vectors updated via the DeGroot opinion dynamics as in Definition 1. Let $\bar{\mathbf{z}}^{(t)} = \mathbf{z}^{(t)} -$ $\text{mean}(\mathbf{z}^{(t)}) \cdot \vec{\mathbf{1}}$ be the mean-centered opinion vector at time $t$, and let*

$\bar{\mathbf{v}}_2 = \mathbf{v}_2 - \text{mean}(\mathbf{v}_2) \cdot \vec{\mathbf{1}}$. *If* $|\lambda_2| \neq |\lambda_3|$ *and* $\langle \mathbf{D}^{1/2}\mathbf{v}_2, z^{(0)} \rangle \neq 0$ *then:*

$$\bar{\mathbf{s}}^* \overset{\text{def}}{=} \lim_{t\to\infty} \frac{[\bar{\mathbf{z}}^{(t)}]^{\pm}}{\|\bar{\mathbf{z}}^{(t)}\|_2} = \frac{[\bar{\mathbf{v}}_2]^{\pm}}{\|\bar{\mathbf{v}}_2\|_2}.$$

*Recall that for a non-zero vector* $\mathbf{x}$, $[x]^{\pm}$ *denotes* $[x]^{\pm} = \text{sign}(x_i) \cdot x$, *where* $x_i$ *is the first non-zero entry in* $\mathbf{x}$.

Theorem 1 holds under two mild assumptions. First, we require that $\mathbf{z}^{(0)}$ has non-zero inner product with $\mathbf{D}^{1/2}\mathbf{v}_2$. This holds with probability 1 whenever $\mathbf{z}^{(0)}$ involves any isotropic random component. Second, we require that $|\lambda_2| \neq |\lambda_3|$, which will hold for any natural social network, as it can be guaranteed by assuming some randomness in the edges of the network[3]. Under these conditions, Theorem 1 shows that the normalized opinions converge to a vector $\bar{\mathbf{s}}^*$ that depends on the social graph $G$ (through its second eigenvector) but *does not depend* on the initial opinion vector $\mathbf{z}^{(0)}$.

Proof. From the linear algebraic form of the DeGroot update rule, we have that:

$$\mathbf{z}^{(t)} = (\mathbf{D}^{-1}\mathbf{A})^t \mathbf{z}^{(0)} = \mathbf{D}^{-1/2} \left(\mathbf{D}^{-1/2}\mathbf{A}\mathbf{D}^{-1/2}\right)^t \mathbf{D}^{1/2}\mathbf{z}^{(0)}. \quad (1)$$

Let $\mathbf{D}^{-1/2}\mathbf{A}\mathbf{D}^{-1/2} = \mathbf{V}\Sigma\mathbf{V}^T$ denote the eigendecomposition of the symmetric normalized adjacency matrix $\mathbf{D}^{-1/2}\mathbf{A}\mathbf{D}^{-1/2}$. $\Sigma$ is a diagonal matrix that contains real-valued eigenvalues identical to those of $\mathbf{D}^{-1}\mathbf{A}$. $\mathbf{V}$ is an orthogonal matrix whose columns contain eigenvectors $\mathbf{v}'_1, \ldots, \mathbf{v}'_n$ where $\mathbf{v}'_i = \mathbf{D}^{1/2}\mathbf{v}_i / \|\mathbf{D}^{1/2}\mathbf{v}_i\|_2$. The eigenvalues of the normalized adjacency matrix of an undirected graph always lie in $[-1, 1]$ and, since $\mathbf{A}$ is connected, there is exactly one eigenvector with eigenvalue $\lambda_1 = 1$.[4] It can be verified that the corresponding eigenvector of $\mathbf{D}^{-1/2}\mathbf{A}\mathbf{D}^{-1/2}$ is equal to $\mathbf{v}'_1 = \mathbf{D}^{1/2}\vec{\mathbf{1}} / \|\mathbf{D}^{1/2}\vec{\mathbf{1}}\|_2$.

We expand (1), using that $(\mathbf{D}^{-1/2}\mathbf{A}\mathbf{D}^{-1/2})^t = \mathbf{V}\Sigma^t\mathbf{V}^T$ since $\mathbf{V}$ is orthogonal. For $i = 1, \ldots, n$, let $c_i = \langle \mathbf{v}_i, \mathbf{D}^{1/2}\mathbf{z}^{(0)} \rangle$. We have that:

$$\mathbf{z}^{(t)} = \mathbf{D}^{-1/2} \cdot \left(c_1\lambda_1^t\mathbf{v}_1 + c_2\lambda_2^t\mathbf{v}_2 + \ldots + c_n\lambda_n^t\mathbf{v}_n\right),$$

and thus $\bar{\mathbf{z}}^{(t)} = \mathbf{z}^{(\mathbf{t})} - \text{mean}\left(\mathbf{z}^{(t)}\right)$ equals:

$$\bar{\mathbf{z}}^{(t)} = c_1\lambda_1^t\mathbf{D}^{-1/2}\mathbf{v}'_1 - \text{mean}\left(c_1\lambda_1^t\mathbf{D}^{-1/2}\mathbf{v}'_1\right) \cdot \vec{\mathbf{1}}$$
$$+ \ldots + c_n\lambda_n^t\mathbf{D}^{-1/2}\mathbf{v}'_n - \text{mean}\left(c_n\lambda_n^t\mathbf{D}^{-1/2}\mathbf{v}'_n\right) \cdot \vec{\mathbf{1}}.$$

Note that $\mathbf{D}^{-1/2}\mathbf{v}'_1$ is a scaling of the all ones vectors, so the first term in the sum above is zero. Letting $\bar{\mathbf{v}}_i = \mathbf{D}^{-1/2}\mathbf{v}'_i - \text{mean}(\mathbf{D}^{-1/2}\mathbf{v}'_i) \cdot \vec{\mathbf{1}}$, we are left with:

$$\frac{\bar{\mathbf{z}}^{(t)}}{\|\bar{\mathbf{z}}^{(t)}\|_2} = \frac{c_2\lambda_2^t\bar{\mathbf{v}}_2 + c_3\lambda_3^t\bar{\mathbf{v}}_3 + \ldots + c_n\lambda_n^t\bar{\mathbf{v}}_n}{\|c_2\lambda_2^t\bar{\mathbf{v}}_2 + c_3\lambda_3^t\bar{\mathbf{v}}_3 + \ldots + c_n\lambda_n^t\bar{\mathbf{v}}_n\|_2}.$$

We first note that $\|\bar{\mathbf{v}}_i\|_2 > 0$ for all $i = 2, \ldots, n$. To see why this is the case, observe that to have $\|\bar{\mathbf{v}}_i\|_2 = 0$, it must be that $\mathbf{v}'_i = c\mathbf{D}^{1/2}\vec{\mathbf{1}}$ for some constant $c$. However, this cannot be the case because $\mathbf{v}'_i$ is orthogonal to $\mathbf{v}'_1 = c\mathbf{D}^{1/2}\vec{\mathbf{1}}$. Combined with our assumption that $c_2 = \langle \mathbf{D}^{1/2}\mathbf{v}_2, z^{(0)} \rangle \neq 0$, it follows that $\|c_2\bar{\mathbf{v}}_2\|_2 \geq 0$. Then, by our assumption that $|\lambda_2| \neq |\lambda_3$, we have $|\lambda_2| > |\lambda_i|$ for all $i = 3, \ldots, n$.

[3]Informally, suppose we are given a fixed adversarial example network with $|\lambda_2| = |\lambda_3|$. A small random perturbation of the edges in the network will ensure that the second and third eigenvalue are no long *exactly* equal with high probability.

[4]We follow the convention that an eigenvalue equal to $-1$ would be denoted as $\lambda_2$.

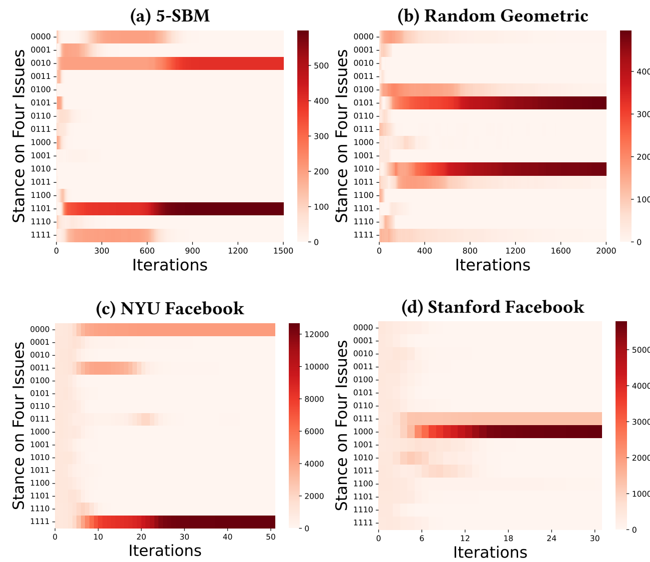


**Figure 2: Heat map of "binary opinion profiles" across $m = 4$ different issues by iteration of the DeGroot model. Each row corresponds to one of $2^m$ difference opinion profiles, and the color at time $t$ indicates the number of nodes whose current opinions match that profile. As predicted by Corollary 2, all individuals eventually sort into just two profiles, leading to perfect ideological alignment.**

So, for any $\epsilon > 0$, there is some $t$ such that $\|c_3\lambda_3^t\bar{\mathbf{v}}_3 + \ldots + c_n\lambda_n^t\bar{\mathbf{v}}_n\|_2 \leq \epsilon\|c_2\bar{\mathbf{v}}_2\|_2$. We conclude that:

$$\lim_{t\to\infty} \frac{[\bar{\mathbf{z}}^{(t)}]^{\pm}}{\|\bar{\mathbf{z}}^{(t)}\|_2} = \lim_{t\to\infty} \frac{[c_2\lambda_2^t\bar{\mathbf{v}}_2 + \ldots + c_n\lambda_n^t\bar{\mathbf{v}}_n]^{\pm}}{\|c_2\lambda_2^t\bar{\mathbf{v}}_2 + \ldots + c_n\lambda_n^t\bar{\mathbf{v}}_n\|_2} = \frac{[c_2\lambda_2^t\bar{\mathbf{v}}_2]^{\pm}}{\|c_2\lambda_2^t\bar{\mathbf{v}}_2\|_2}.$$

This proves the theorem since $\frac{[c_2\lambda_2^t\bar{\mathbf{v}}_2]^{\pm}}{\|c_2\lambda_2^t\bar{\mathbf{v}}_2\|_2} = \frac{[\bar{\mathbf{v}}_2]^{\pm}}{\|\bar{\mathbf{v}}_2\|_2}$ for any $c_2, \lambda_2$. □

## 3.1 Implications for Ideological Alignment

In the work of DeMarzo, Vayanos, and Zwiebel [16], Theorem 1 is used to explain a phenomenon involving *multiple opinion vectors*, each defined for a different issue. They call the phenomenon "unidimensional opinions", but we prefer the terminology *ideological alignment.* Ideological alignment occurs when large groups of individuals simultaneously differ in opinion on many issues [4, 19]. Also refereed to as "party sorting" [20], this phenomenon is well-documented in the real-world, and there is strong survey-based evidence that it has increased in recent years [40, 49]. Since it accentuates differences between groups, ideological alignment has likely contributed to increased perception of polarization [4].

Theorem 1 provides striking mathematical support for the emergence of ideological alignment. In particular, an immediate corollary of the result is that, in the limit, individuals will perfectly sort into exactly two groups that simultaneously disagree on *all issues* –i.e., for each issue, the members of one group will all have opinions on the opposite side of the mean as the other group. We formalize their observation from [16] in Corollary 2.

Corollary 2. *Consider a social graph G and m different initial opinion vectors* $z_1^{(0)}, \cdots, z_m^{(0)}$ *satisfying the assumptions of Theorem 1.*

*Apply the DeGroot opinion dynamics to each vector for $t$ steps to obtain opinions $\mathbf{z}_1^{(t)}, \cdots, \mathbf{z}_m^{(t)}$, and let $\mathbf{s}_i^{(t)} = \text{sign}(\mathbf{z}_i^{(t)} - \text{mean}(\mathbf{z}_i^{(t)}) \cdot \vec{\mathbf{1}})$. Consider the matrix $\mathbf{S}^{(t)} = [\mathbf{s}_1^{(t)}, \ldots, \mathbf{s}_m^{(t)}]$. In the limit as $t \to \infty$, $\mathbf{S}^{(t)}$ will only contain two unique rows.*

Each row of $\mathbf{S}^{(t)}$ corresponds to a single node (individual) in $G$. It contains $\{+1, -1\}$ entries indicating if that individual has opinion below or above the mean for each of the $m$ topics at time $t$. The row can thus be viewed an individual's "binary opinion profile". The takeaway from Corollary 2 is that, while there are $2^m$ possible opinion profiles, for large enough $t$, just two will dominate, becoming adopted by every individual. We visualized this alignment for four social networks in Figure 2. Opinions were initialized randomly, so the rows of $\mathbf{S}^{(0)}$ are distributed evenly between all $2^m$ possible binary opinion profiles. However, as $t$ increases, we eventually see convergence to a state where $\mathbf{S}^{(t)}$ has just two unique binary rows. The number of iterations until convergence varies by network.

While an interesting phenomenon, one limitation of ideological alignment as a polarization measure is that, like variance-based measures, it converges to the same extreme state for all social networks – albeit to a state that is fully polarized instead of a fully in consensus. In contrast, the other group-based measures of polarization discussed in this paper converge to *network dependent quantities*, so their dynamics naturally differ within different social structures and can be impacted by outside influences that effect that structure, like social media or propaganda.

## 3.2 Implications for Group-Based Polarization

The foundation of our work is the insight that Theorem 1 actually has implications on the limiting behavior of *any* group-based measure of polarization. Formally:

Corollary 3. *Let $f(G, \mathbf{z})$ be a group-based polarization metric according to Definition 2 that is continuous with respect to the argument $\mathbf{z} \in \mathbb{R}^n$. If the conditions of Theorem 1 hold, then*

$$\lim_{t \to \infty} f(G, \mathbf{z}^{(t)}) = f(G, \mathbf{v}_2)$$

*where $\mathbf{z}^{(t)}$ and $\mathbf{v}_2$ are as defined as in Theorem 1.*

Corollary 3 implies that, unlike variance-based measures which always converge to zero, under the mild assumptions of Theorem 1, any group-based measure of polarization converges to a value that depends on the social graph $G$. At the same time, the value does not depend on the starting opinions $\mathbf{z}^{(0)}$. With Corollary 3 in place, we next analyze several different group-based measures.

# 4 STATISTICAL MEASURES

We start with statistical measures that, like variance, consider only the numerical values in an opinion vector $\mathbf{z}$, without taking into account the ordering of entries or their structure with respect to $G$. For example, the following common statistical measure of bimodality incorporates 3rd and 4th moment information from $\mathbf{z}$:

Definition 3 (Sarle's Bimodality Coefficient). *Consider an opinion vector $\mathbf{z}$ and let $\bar{\mathbf{z}}$ denote $\bar{\mathbf{z}} = \mathbf{z} - \text{mean}(\mathbf{z})$. Then the bimodality $\beta(\mathbf{z})$ is written in terms of the skewness $\gamma$ and kurtosis $\kappa$ as follows:*

$$\beta(\mathbf{z}) = \frac{\gamma^2 + 1}{\kappa} \quad \text{where } \gamma = \frac{\text{mean}(\bar{\mathbf{z}}^3)}{\text{mean}(\bar{\mathbf{z}}^2)^{3/2}} \text{ and } \kappa = \frac{\text{mean}(\bar{\mathbf{z}}^4)}{\text{mean}(\bar{\mathbf{z}}^2)^2}.$$

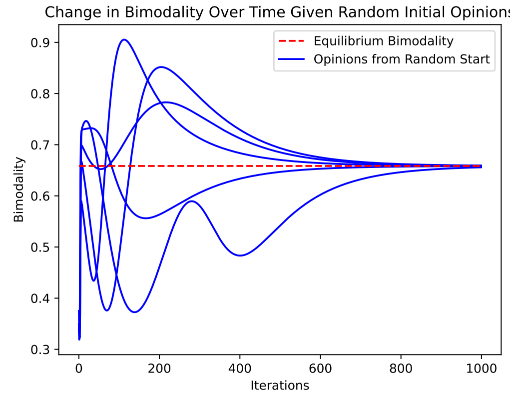


**Figure 3: Opinion bimodality $\beta(\mathbf{z})$ plotted by iteration of DeGroot's opinion dynamics model run on the same 5 block SBM graph, initialized with five randomly generated starting opinion vectors. As predicted by Corollary 3, in all cases opinion bimodality converges to a fixed non-zero equilibrium bimodality that depends on the graph.**

The bimodality coefficient of Definition 3 has been used as a measure of opinion polarization, e.g. in [18], where it was compared against variance-based measures. The measure lies between 0 and 1, with 1 indicating maximum polarization. However, even a random isotropic vector $\mathbf{r}$ (e.g., a vector with i.i.d. random Gaussian entries) will have bimodality $\beta(\mathbf{r}) \approx 1/3$, since the skewness of a normal random variable is 0 and the kurtosis is 3. Accordingly, we consider a vector of opinions "polarized" if the bimodality is larger than 1/3.

We demonstrate Corollary 3 in Figure 3. We generate a Stochastic Block Model (SBM) network [2, 33] on $n = 1000$ nodes with five communities (blocks). The probability of an edge within a block is $p = 1/10$ and the probability of an edge between blocks is $q = 1/100$. We then initialize five random starting opinion vectors, each with i.i.d. standard normal entries. We plot the bimodality of opinions as they evolve via the DeGroot dynamics. By 1000 iterations, there is clear convergence to the bimodality of the second eigenvector of the SBM, which, at .658, is much larger than the bimodalities of the starting opinions around 1/3. So, while bimodality evolves in a highly non-monotonic way, it ultimately increases over time.

| 1st Quartile | Median | 3rd Quartile |
|---|---|---|
| .805 | .917 | .952 |

**Table 1: Statistics of equilibrium Sarle's bimodality coefficient for 100 college social networks from the Facebook100 data set [51]. Notably values tend close to the maximum coefficient of 1.**

Increases in bimodality are even more pronounced in real-world social networks. We ran a similar experiment for 100 college social networks from the Facebook100 data set [51] and observed that for all but five networks, bimodality *increases* under the DeGroot dynamics with random starting opinions. The median and quartiles of the equilibrium bimodality (computed directly from the second eigenvector of each network) are included in Table 1. We conclude that the simple bimodality coefficient offers a clear contrast with variance-based measures of polarization that decrease over time.

An informal analysis of SBM graphs offers theoretical support for increases in bimodality in natural social networks with a small number of well connected communities. Specifically, we argue that any SBM with a small number of blocks typically has equilibrium bimodality greater than 1/3. We thus expect increasing bimodality under the DeGroot model if opinions are randomly initialized.

Observation 1. *For a $k$-block SBM graph, the equilibrium bimodality is approximated by the sample bimodality of a normal random variable when $k$ samples are taken, which has expected value greater than* $1/3$ *for small* $k$.

We sketch a proof of Observation 1: While the true bimodality of the normal distribution is 1/3, the empirical bimodality computed from a finite number of samples tends to be an over-estimate. While it is difficult to obtain an exact expression for the expected sample bimodality, the sample kurtosis has expectation $3\frac{k-1}{k+1}$ [36]. It is thus an underestimate for small $k$, explaining the overestimate of bimodality, which depends on the inverse kurtosis. Now consider the expected normalized adjacency matrix $\bar{\mathbf{D}}^{-1/2}\bar{\mathbf{A}}\bar{\mathbf{D}}^{-1/2}$ of an SBM graph, where $\bar{\mathbf{D}} = \mathbb{E}[\mathbf{D}]$ and $\bar{\mathbf{A}} = \mathbb{E}[\mathbf{A}]$. The top $k$ eigenvectors of $\bar{\mathbf{D}}^{-1/2}\bar{\mathbf{A}}\bar{\mathbf{D}}^{-1/2}$ can be spanned by $\vec{1}$ as well as $k$ block indicator vectors, each which is 1 for the nodes in a single community, and 0 for all other nodes. Since the actual normalized adjacency matrix $\mathbf{D}^{-1/2}\mathbf{A}\mathbf{D}^{-1/2}$ can be viewed as a perturbed version of $\bar{\mathbf{D}}^{-1/2}\bar{\mathbf{A}}\bar{\mathbf{D}}^{-1/2}$, we roughly expect its first $k$ eigenvectors to also be spanned by $\vec{1}$ and the $k$ block vectors – a formal statement could be made by appealing to the Davis-Kahan perturbation theorem [14]. Moreover, the $2^{\text{nd}}$ through the $(k-1)^{\text{st}}$ eigenvalues of $\bar{\mathbf{D}}^{-1/2}\bar{\mathbf{A}}\bar{\mathbf{D}}^{-1/2}$ are all the same, so we roughly expect the second eigenvector of $\mathbf{D}^{-1/2}\mathbf{A}\mathbf{D}^{-1/2}$ to be a *random* linear combination of the $k$ block indicator vector, plus some scaling of $\vec{1}$ (which has no impact on bimodality). If the random linear combination is isotropic, the second eigenvector will look exactly like $k$ samples from a random Gaussian distribution, each repeated $n/k$ times. This vector has the same bimodality as $k$ random Gaussian samples.

Observation 1 is visualized in Figure 4, which was generated by computing the equilibrium opinion bimodality for 100 random $k$-SBM graphs with 1000 and 2000 nodes. While it approaches 1/3 as $k$ increases, equilibrium bimodality is much larger for small $k$. We also plot the sample bimodality of $k$ i.i.d Gaussian samples (also computed using 100 trials), which as predicted by Observation 1, correlates well with the observed bimodality of the $k$-SBM.

## 5 LOCAL MEASURES

Another interesting class of group-based polarization measures are those that take into account local structure of the social graph $G$. Such measures are motivated by the fact that individuals are most heavily exposed to the opinions of their social connections – i.e., their neighborhood in $G$. Individuals likely also have a sense of the overall mean opinion in $G$ (e.g., from the news), but do not simultaneously sense all opinions in a social network.

In this section we introduce and study one such measure, which we call *average local agreement* that takes these considerations into account. In particular, we define the local agreement of a vertex $i$ to be the ratio of $i$'s neighbors whose opinion falls on the same side (above or below) the mean opinion mean($\mathbf{z}$) as $i$. We posit that

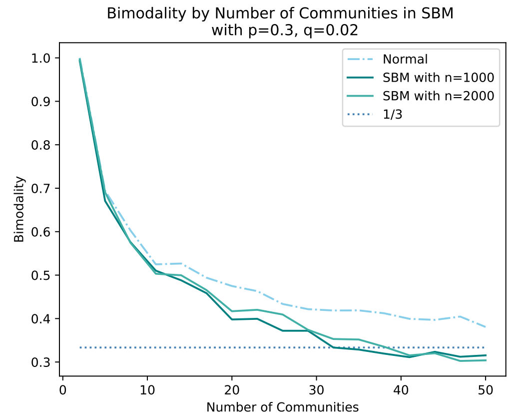


**Figure 4: Average equilibrium bimodality of $k$-SBM graphs with intra- and inter-community edge probability .3 and .02. Bimodality converges to that of a random normal variable for large $k$, which is 1/3, but as predicted in Observation 1, can be larger for small $k$.**

*high local agreement* correlates with *high perceptions of polarization*, as individuals who feel more isolated in a group, away from those differing opinion, tends to experience feelings of polarization [41].

We formally define average local agreement below. We use sign($\mathbf{x}$) to denote the operation that rounds every entry of a vector $\mathbf{x}$ to +1 or -1, taking the convention that if $x_i = 0$, $[\text{sign}(\mathbf{x})]_i = +1$.

Definition 4 (Average Local Agreement). *Let $G$ be a social network on $n$ nodes and let $\mathbf{z} \in \mathbb{R}^n$ be an opinion vector. Let $\mathbf{s} = \text{sign}(\mathbf{z} - \text{mean}(\mathbf{z}) \cdot \vec{1})$. The average local agreement $\mathcal{L}(G, \mathbf{z})$ equals:*

$$\mathcal{L}(G,\mathbf{z}) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{d_i}\sum_{j\in\mathcal{N}(i)}\mathbb{1}[s_i = s_j]$$

*where $\mathbf{s} = \text{sign}(\mathbf{z} - \text{mean}(\mathbf{z}) \cdot \vec{1})$, $\mathcal{N}(i)$ denotes the neighborhood of node i, and $\mathbb{1}[\cdot]$ evaluates to 1 if the expression in brackets is true, and to 0 otherwise.*

Like bimodality, average local agreement is a group-based measure, so by Corollary 3, we have that in the DeGroot dynamics, under the assumptions of Theorem 1, $\lim_{t\to\infty}\mathcal{L}(G,\mathbf{z}^{(t)}) = \mathcal{L}(G,\mathbf{v}_2)$, where $z^{(t)}$ and $\mathbf{v}_2$ are as defined as in the theorem.

| $1^{\text{st}}$ Quartile | Median | $3^{\text{rd}}$ Quartile |
|---|---|---|
| .904 | .947 | .960 |

**Table 2: Statistics of the equilibrium average local agreement, $\mathcal{L}(G, \mathbf{v}_2)$ for the Facebook100 data set [51].**

Average local agreement is bounded between [0, 1] and we expect a value of 1/2 for randomly initialized opinions. So, any value above 1/2 is considered "polarized". As shown in Table 2, we observe very high average local agreement in the limit for real-world social networks. For all but two of the 100 networks in the Facebook100 data set, this measure of polarization converged to a value above .6, and was typically well above .9. In Figure 5 we also visualize local agreement over time for a random 5-SBM graph and a random geometric graph, as well as the Swarthmore Facebook graph (chosen for its small size). In all cases, "bubbles" of high local agreement visibly emerge, with average local agreement increasing to .785, .954, and .941 for the three graphs, respectively.

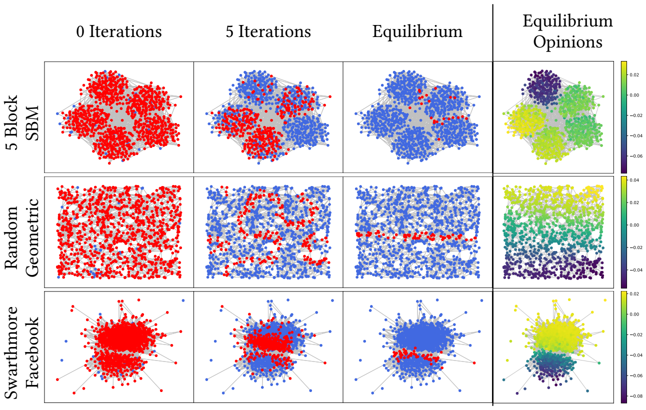


**Figure 5: A visualization of local agreement by number of DeGroot iterations for three social networks. Nodes are colored blue for individuals with $\geq 2/3$ of their neighbors on the same side of the mean opinion, and red otherwise. The last column shows the normalized differences from the mean opinion at equilibrium ($\bar{\mathbf{s}}^*$ from Theorem 1). In all cases, strong clusters of high local agreement emerge, which may lead to increased perceptions of opinion polarization.**

To better understand the steep increase in this group-based polarization metric theoretically, we show that for an unweighted, regular graph $G$, average local agreement has a simple linear algebraic form. Ultimately, the following claim will help us relate the measure to spectral properties of the underlying social graph $G$.

Claim 1. *Let $G$ be an unweighted $d$-regular graph with no self-loops. Let $\mathbf{z}$ be a vector of opinions and let $\mathbf{s} = \text{sign}(\mathbf{z} - \text{mean}(\mathbf{z}) \cdot \vec{\mathbf{1}})$. Then, the average local agreement $\mathcal{L}(G, \mathbf{z})$ equals:*

$$\mathcal{L}(G, \mathbf{z}) = \frac{\mathbf{s}^T \mathbf{A} \mathbf{s}}{2nd} + \frac{1}{2} \qquad \textit{where} \quad \mathbf{s} = \text{sign}(\mathbf{z} - \text{mean}(\mathbf{z}) \cdot \vec{\mathbf{1}}).$$

Proof. For a node $i$, let $p_i = \sum_{j \in \mathcal{N}(i)} \mathbb{1}[s_j = +1]$ denote the number of nodes in $\mathcal{N}(i)$ that are on the positive side of the mean and let $q_i = \sum_{j \in \mathcal{N}(i)} \mathbb{1}[s_j = -1]$ denote the number of nodes on the negative side of the mean. Let $a_i$ denote the number of nodes in $\mathcal{N}(i)$ that agree with node $i$ (i.e., are on the same side of the mean) and let $b_i$ denote the number of nodes that disagree. We can write:

$$a_i = \begin{cases} p_i & \text{if } s_i = +1 \\ q_i & \text{if } s_i = -1 \end{cases} \qquad\qquad b_i = \begin{cases} p_i & \text{if } s_i = -1 \\ q_i & \text{if } s_i = +1 \end{cases}$$

Observe that the $i^{\text{th}}$ entry of $\mathbf{A}\mathbf{s}$ equals $p_i - q_i$, and thus:

$$\mathbf{s}^T \mathbf{A} \mathbf{s} = \sum_{i=1}^{n} s_i (p_i - q_i) = \sum_{i=1}^{n} a_i - b_i.$$

Next note that $a_i + b_i = d$ and thus $nd = \sum_{i=1}^{n} a_i + b_i$. So we have $\mathbf{s}^T \mathbf{A} \mathbf{s} + nd = \sum_{i=1}^{n} 2a_i$. Dividing by $2nd$ gives the result because $\mathcal{L}(G, \mathbf{z}) = \frac{1}{n} \sum_{i=1}^{n} \frac{a_i}{d}$. $\square$

With Claim 1 in place, we make the following observation:

Observation 2. *For an unweighted graph $G$, we can approximate the equilibrium average local agreement $\lim_{t \to \infty} \mathcal{L}(G, \mathbf{z}^{(t)})$ by*

$$\lim_{t \to \infty} \mathcal{L}(G, \mathbf{z}^{(t)}) \approx \frac{\lambda_2}{2} + \frac{1}{2},$$

*where $\lambda_2$ is the second eigenvalue of $G$'s normalized adjacency matrix.*

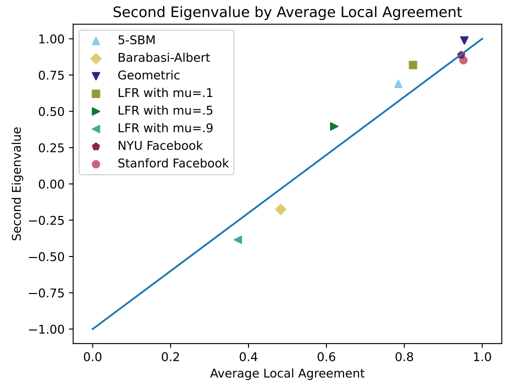


**Figure 6: Average equilibrium local agreement plotted against the second normalized adjacency matrix eigenvalue for several random graphs generated with NetworkX [30], and for the NYU and Stanford Facebook graphs. The values closely align with the linear relationship predicted by Observation 2 (plotted as a solid line).**

According to this claim, we expect high equilibrium local agreement – i.e., *increasing* polarization – in graphs with second eigenvalue close to 1, which includes any social network with strong community structure. For example, the Facebook100 networks had an average second eigenvalue of .871. Observation 2 predicts that this would lead to a mean equilibrium average local agreement of $\approx$ .936, which is extremely close to the observed mean of .948.

To establish Observation 2, again assume that $G$ is $d$-regular with no self-loops. Note that for a regular graph, the second eigenvalue of $\mathbf{D}^{-1}\mathbf{A}$ is equal to that of $\mathbf{D}^{-1/2}\mathbf{A}\mathbf{D}^{-1/2}$. By Corollary 3, we have that $\lim_{t \to \infty} \mathcal{L}(G, \mathbf{z}^{(t)}) = \mathcal{L}(G, \mathbf{v}_2)$. And by Claim 1:

$$\begin{aligned}\mathcal{L}(G, \mathbf{v}_2) &= \frac{\text{sign}(\mathbf{v}_2^T)\mathbf{A}\,\text{sign}(\mathbf{v}_2)}{2nd} + \frac{1}{2} \\ &= \frac{\text{sign}(\mathbf{v}_2^T)\mathbf{D}^{-1/2}\mathbf{A}\mathbf{D}^{-1/2}\,\text{sign}(\mathbf{v}_2)}{2n} + \frac{1}{2}\end{aligned}$$

Observation 2 then immediately follows by noticing that

$$\text{sign}(\mathbf{v}_2^T)\mathbf{D}^{-1/2}\mathbf{A}\mathbf{D}^{-1/2}\,\text{sign}(\mathbf{v}_2) \approx n\mathbf{v}_2^T\mathbf{D}^{-1/2}\mathbf{A}\mathbf{D}^{-1/2}\mathbf{v}_2 = n\lambda_2.$$

The approximation is exact if all entries in the unit vector $\mathbf{v}_2$ have magnitude $1/\sqrt{n}$. It tends to hold a close approximation in other networks, which can be formalized via well-known connections between balanced cut problems and the second eigenvector [45].

In Figure 6, we confirm the relationship described in Observation 2 by examining a variety of random graphs with widely varying second eigenvalue. The correlation between average local agreement and second eigenvalue in the Facebook100 data set is statistically significant ($p = 5e^{-5}$) with a Pearson correlation of $r = .392$.

Finally, we comment on the rate at which average local agreement converges to its equilibrium value. Since this rate depends on how quickly the normalized difference vector converges to $\bar{\mathbf{s}}^*$ under the DeGroot model, we expect it to scale linearly with the inverse of the *second* eigenvalue gap, $\frac{|\lambda_2| - |\lambda_3|}{|\lambda_2|}$. We confirm this relationship on the Facebook100 data set, where we see a statistically significant ($p = 7e^{-6}$) correlation between inverse second eigengap and average number of iterations until convergence to the final average local agreement when starting with a random opinion vector. The Pearson correlation coefficient of the relationship is $r = .451$.

In contrast, the rate at which the opinion vector converges to $\mathbf{z}^*$ depends inversely on the *first* eigengap $\frac{|\lambda_1|-|\lambda_2|}{|\lambda_1|}$. As such, when the second eigengap is large compared to the first, we expect local agreement to increase more quickly than opinion variance decreases, which might contribute to perceptions of growing polarization.

## 6 FUTURE DIRECTIONS

In this paper, we established that natural group-based polarization measures display interesting dynamics under the standard DeGroot opinion formation model. Unlike heavily studied variance-based measures, we showed empirically and theoretically that group-based measures can increase over time, and often do increase quite significantly in natural social networks. We leave a number of questions for future research. As discussed, recent work on mathematical models of opinion dynamics has sought to understand the impact of outside actors (who can modify the graph $G$ is some way) on individual opinions and polarization [3, 25]. There is little work on how such modifications impact group-based polarization, and if they can accelerate its emergence. Another challenging empirical question it to determine the "right" group-based measure of polarization for use in opinion dynamics studies – i.e., to better understand what measures best align with perceived polarization in the real-world. There is some evidence for the value of ideological alignment as a meaningful polarization metric [21, 40], but statistical measures of bimodality and "local" metrics have received less attention.

## REFERENCES


[1] [n.d.]. "polarization, n.". In *Oxford English Dictionary*. Oxford University Press. https://www.oed.com/view/Entry/146757.
[2] Emmanuel Abbe. 2017. Community detection and stochastic block models: recent developments. *The Journal of Machine Learning Research* 18, 1 (2017), 6446–6531.
[3] Rediet Abebe, T.-H. Chan, Jon Kleinberg, Zhibin Liang, David Parkes, Mauro Sozio, and Charalampos E. Tsourakakis. 2021. Opinion Dynamics Optimization by Varying Susceptibility to Persuasion via Non-Convex Local Search. *ACM Trans. Knowl. Discov. Data* 16, 2 (2021).
[4] Alan I. Abramowitz and Kyle L. Saunders. 2008. Is Polarization a Myth? *The Journal of Politics* 70, 2 (2008), 542–555.
[5] Daron Acemoglu and Asuman Ozdaglar. 2011. Opinion Dynamics and Learning in Social Networks. *Dynamic Games and Applications* 1, 1 (2011), 3–49.
[6] W. Bianco and R. Smyth. 2020. The Bicameral Roots of Congressional Deadlock: Analyzing Divided Government Through the Lens of Majority Rule. *Social Science Quarterly* 101 (2020), 1712–1727.
[7] Sarah Binder. 2014. Polarized we govern? https://www.brookings.edu/research/polarized-we-govern/
[8] Heather Brooks and Mason Porter. 2020. A model for the influence of media on the ideology of content in online social networks. *Phys. Rev. Research* 2, 2 (2020).
[9] Thomas Carothers and Andrew O'Donohue (Eds.). 2019. *Democracies Divided: The Global Challenge of Political Polarization.* Brookings Institution Press.
[10] John R. Chambers, Robert S. Baron, and Mary L. Inman. 2006. Misperceptions in Intergroup Conflict: Disagreeing About What We Disagree About. *Psychological Science* 17, 1 (2006), 38–45.
[11] Uthsav Chitra and Christopher Musco. 2020. Analyzing the Impact of Filter Bubbles on Social Network Polarization. In *Proceedings of the 13th International Conference on Web Search and Data Mining*. 115–123.
[12] Pranav Dandekar, Ashish Goel, and David T Lee. 2013. Biased assimilation, homophily, and the dynamics of polarization. *Proceedings of the National Academy of Sciences* 110, 15 (2013), 5791–5796.
[13] Abhimanyu Das, Sreenivas Gollapudi, and Kamesh Munagala. 2014. Modeling opinion dynamics in social networks. In *Proceedings of the 7th International Conference on Web Search and Data Mining (WSDM)*. ACM, 403–412.
[14] C. Davis and W. Kahan. 1970. The Rotation of Eigenvectors by a Perturbation. III. *SIAM J. Numer. Anal.* 7, 1 (1970), 1–46.
[15] Morris H. Degroot. 1974. Reaching a Consensus. *J. Amer. Statist. Assoc.* 69, 345 (1974), 118–121.
[16] Peter M. DeMarzo, Dimitri Vayanos, and Jeffrey Zwiebel. 2003. Persuasion Bias, Social Influence, and Unidimensional Opinions. *The Quarterly Journal of Economics* 118, 3 (2003), 909–968.
[17] Dominic Difranzo and Kristine Gloria-Garcia. 2017. Filter bubbles and fake news. *XRDS: Crossroads, The ACM Magazine for Students* 23 (04 2017), 32–35.
[18] Paul DiMaggio, John Evans, and Bethany Bryson. 1996. Have American's Social Attitudes Become More Polarized? *Amer. J. Sociology* 102, 3 (1996), 690–755.
[19] John H. Evans. 2003. Have Americans' Attitudes Become More Polarized?—An Update. *Social Science Quarterly* 84, 1 (2003), 71–90.
[20] Morris P. Fiorina and Samuel J. Abrams. 2008. Political Polarization in the American Public. *Annual Review of Political Science* 11, 1 (2008), 563–588.
[21] Morris P. Fiorina, Samuel J. Abrams, and Jeremy C. Pope. 2005. *Culture war: The myth of a polarized America.* Pearson-Longman.
[22] John R. P. French Jr. 1956. A formal theory of social power. *Psychological Review* 63, 3 (1956), 181–194.
[23] Noah E. Friedkin and Eugene C. Johnsen. 1990. Social influence and opinions. *The Journal of Mathematical Sociology* 15, 3-4 (1990), 193–206.
[24] Jason Gaitonde, Jon Kleinberg, and Eva Tardos. 2020. Adversarial perturbations of opinion dynamics in networks. In *ACM Conference on Economics and Computation.*
[25] Jason Gaitonde, Jon Kleinberg, and Éva Tardos. 2021. Polarization in Geometric Opinion Dynamics. *ACM Conference on Economics and Computation* (2021).
[26] Daniel Geschke, Jan Lorenz, and Peter Holtz. 2019. The triple-filter bubble: Using agent-based modelling to test a meta-theoretical framework for the emergence of filter bubbles and echo chambers. *Brit J Soc Psychol* 58, 1 (2019), 129–149.
[27] Alvin Goldman. 2002. Knowledge in a social world. *Philosophy and Phenomenological Research* 64, 1 (2002).
[28] Jon Green, Jared Edgerton, Daniel Naftel, Kelsey Shoub, and Skyler J. Cranmer. 2020. Elusive consensus: Polarization in elite communication on the COVID-19 pandemic. *Science Advances* 6, 28 (2020).
[29] Pedro Guerra, Wagner Meira Jr, Claire Cardie, and Robert Kleinberg. 2013. A measure of polarization on social media networks based on community boundaries. In *AAAI Conference on Web and Social Media*, Vol. 7. 215–224.
[30] Aric A. Hagberg, Daniel A. Schult, and Pieter J. Swart. 2008. Exploring Network Structure, Dynamics, and Function using NetworkX. In *Proceedings of the 7th Python in Science Conference.* 11 – 15.
[31] Jan Hązła, Yan Jin, Elchanan Mossel, and Govind Ramnarayan. 2019. A Geometric Model of Opinion Polarization. *arXiv:910.05274* (2019).
[32] Rainer Hegselmann and Ulrich Krause. 2002. Opinion dynamics and bounded confidence: Models, analysis and simulation. *Journal of Artificial Societies and Social Simulation* 5 (2002), 1–24.
[33] Paul W. Holland, Kathryn Blackmond Laskey, and Samuel Leinhardt. 1983. Stochastic blockmodels: First steps. *Social Networks* 5, 2 (1983), 109–137.
[34] Harald Holone. 2016. The filter bubble and its effect on online personal health information. *Croatian medical journal* 57, 3 (2016), 298.
[35] Matthew O. Jackson. 2008. *Social and economic networks.* Princeton Univ. Press.
[36] D. N. Joanes and C. A. Gill. 1998. Comparing Measures of Sample Skewness and Kurtosis. *Journal of the Royal Statistical Society. Series D* 47, 1 (1998), 183–189.
[37] David R. Jones. 2015. Declining Trust in Congress: Effects of Polarization and Consequences for Democracy. *The Forum* 13, 3 (2015), 375–394.
[38] Jon Kleinberg. 2003. An Impossibility Theorem for Clustering. In *Advances in Neural Information Processing Systems 16 (NeurIPS)*, Vol. 15.
[39] Geoffrey C. Layman, Thomas M. Carsey, and Juliana Menasce Horowitz. 2006. Party Polarization in American Politics: Characteristics, Causes, and Consequences. *Annual Review of Political Science* 9, 1 (2006), 83–110.
[40] Matthew Levendusky. 2009. *The partisan sort.* University of Chicago Press.
[41] Matthew Levendusky and Neil Malhotra. 2015. (Mis)perceptions of Partisan Polarization in the American Public. *Public Opinion Quarterly* 80, S1 (10 2015).
[42] Jan Lorenz. 2007. Continuous Opinion Dynamics Under Bounded Confidence: A Survey. *International Journal of Modern Physics C* 18, 12 (2007), 1819–1838.
[43] Antonis Matakos, Cigdem Aslay, Esther Galbrun, and Aristides Gionis. 2020. Maximizing the Diversity of Exposure in a Social Network. *IEEE Transactions on Knowledge and Data Engineering* PP (11 2020), 1–1.
[44] Aaron M McCright and Riley E Dunlap. 2011. The politicization of climate change and polarization in the American public's views of global warming, 2001–2010. *The Sociological Quarterly* 52, 2 (2011), 155–194.
[45] Frank McSherry. 2001. Spectral Partitioning of Random Graphs. In *Proceedings of the 42nd Annual IEEE Symposium on Foundations of Computer Science (FOCS).*
[46] Cameron Musco, Christopher Musco, and Charalampos E. Tsourakakis. 2018. Minimizing Polarization and Disagreement in Social Networks. In *Proceedings of the 27th International World Wide Web Conference (WWW).* 369–378.
[47] Reza Olfati-Saber, J Alex Fax, and Richard M Murray. 2007. Consensus and cooperation in networked multi-agent systems. *Proc. IEEE* 95, 1 (2007), 215–233.
[48] Eli Pariser. 2011. *The filter bubble: what the internet is hiding from you.* Penguin.
[49] Pew Research Center. 2014. Political Polarization in the American Public. (2014).
[50] Hafizh Prasetya and Tsuyoshi Murata. 2020. A model of opinion and propagation structure polarization in social media. *Comp. Social Networks* 7, 1 (2020), 2.
[51] Amanda L Traud, Peter J Mucha, and Mason A Porter. 2012. Social structure of Facebook networks. *Physica A* 391, 16 (2012), 4165–4180.