

DSP: A Statistically-Principled Structural Polarization Measure

Giulia Preti
CENTAI
Turin, Italy
giulia.preti@centai.eu

Matteo Riondato
Amherst College
Amherst, MA, USA
mriondato@amherst.edu

Aristides Gionis
KTH
Stockholm, Sweden
argioni@kth.se

Gianmarco De Francisci Morales
CENTAI
Turin, Italy
gdfm@acm.org

Abstract

Social and information networks may become polarized, leading to echo chambers and political gridlock. Accurately measuring this phenomenon is a critical challenge. Existing measures often conflate genuine structural division with random topological features, yielding misleadingly high polarization scores on random networks, and failing to distinguish real-world networks from randomized null models. We introduce DSP, a Diffusion-based Structural Polarization measure designed from first principles to correct for such biases. DSP removes the arbitrary concept of “influencers” used by the popular Random Walk Controversy (RWC) score, instead treating every node as a potential origin for a random walk. To validate our approach, we introduce a set of desirable properties for polarization measures, expressed through reference topologies with known structural properties. We show that DSP satisfies these desiderata, being near-zero for non-polarized structures such as cliques and random networks, while correctly capturing the expected polarization of reference topologies such as monochromatic-splittable networks. Finally, on real-world datasets of legislative behavior in the U.S. Congress, DSP uncovers trends of increasing polarization in recent years. By integrating a null model into its core definition, DSP provides a reliable and interpretable diagnostic tool, highlighting the necessity of statistically-grounded metrics to analyze societal fragmentation.

CCS Concepts

• **Human-centered computing** → **Social network analysis**; • **Theory of computation** → *Graph algorithms analysis*; • **Applied computing** → *Sociology*.

ACM Reference Format:

Giulia Preti, Matteo Riondato, Aristides Gionis, and Gianmarco De Francisci Morales. 2026. DSP: A Statistically-Principled Structural Polarization Measure. In *Proceedings of the Nineteenth ACM International Conference on Web Search and Data Mining (WSDM '26)*, February 22–26, 2026, Boise, Idaho, USA. ACM, New York, NY, USA, 12 pages. <https://doi.org/10.1145/nnnnnnnn.nnnnnnnn>

1 Introduction

Quantifying the structural polarization of a network from its topology is crucial for identifying systemic risks—from political gridlock [25] to the proliferation of misinformation [24]—and for developing effective interventions to mitigate these harmful effects [22]. This phenomenon is characterized by the grouping of individuals into coherent communities that rarely interact with each other [7],

and it is increasingly observed across social, political, and information networks, where its effects can be seen in the emergence of ideological echo chambers [6] and adversarial dynamics [8]. Mitigating the detrimental effects of polarization on collective decision-making and overall societal stability necessitate robust methods for its quantification and analysis [4].

Existing measures of structural polarization, however, face significant limitations. Many traditional methods, such as modularity or assortativity, rely on simplistic null models that often fail to capture the complex dynamics of the real-world networks [13]. Other more recent measures, such as Betweenness Centrality Controversy [12], Boundary Polarization [13], Dipole Polarization [23], Krackhardt E/I Ratio [18], and Random Walk Controversy [12], can produce high polarization scores even for random networks, indicating an undesirable sensitivity to elementary network features such as average degree and network size [29]. These biases occur because existing controversy measures conflate topological artifacts, such as low density, small network size, or uneven group splits, with genuine structural division.

Attempts to correct these biases, such as post-hoc normalization via randomization tests [29], remain fundamentally unprincipled, as they apply arbitrary null models retroactively without rigorous justification. As such, they risk over- or under-correction, as different null assumptions (e.g., preserving degree sequences vs. erasing all structure) yield conflicting baselines [26]. Moreover, normalization is treated as an external adjustment rather than a core component of the measure’s design, perpetuating ambiguities in score interpretation.

To address the aforementioned issues, we introduce DSP, a new measure for structural polarization, built on a principled statistical model. Unlike previous proposals, DSP is designed to have an expected value of zero on $G(n, p, \ell)$ networks, a labeled extension of the Erdős-Rényi $G(n, p)$ model where edges and labels are assigned randomly, thus ensuring no structural correlation exists between the two. Our design explicitly removes the biases found in the original Random Walk Controversy (RWC) [12], by eliminating the problematic role of predefined influencers. Instead, DSP treats each vertex as a potential source for a random walk, with all other vertices as targets. This approach retains the strength of RWC in being independent of partition sizes and in leveraging random walks to effectively measure distance and information spread. Nonetheless, by considering every vertex a potential source, DSP becomes more robust to different network structures and better reflects how information spreads across the network (see Section 4).

Eliminating the idea of designating a small number of vertices as influencers and considering all vertices as potential influencers poses new challenges, as a random walk (re)starting from a given vertex v has high probability to be visiting vertex v more frequently, which introduces a new source of bias. We show how to avoid this

WSDM '26, Boise, Idaho, USA

© 2026 Copyright held by the owner/author(s).

This is the author’s version of the work. It is posted here for your personal use. Not for redistribution. The definitive Version of Record was published in *Proceedings of the Nineteenth ACM International Conference on Web Search and Data Mining (WSDM '26)*, February 22–26, 2026, Boise, Idaho, USA, <https://doi.org/10.1145/nnnnnnnn.nnnnnnnn>.

Table 1: Most common structural polarization measures.

Measure	Range	Intuition
Random Walk Controversy (RWC) [12]	$[-1, 1]$	In polarized networks, users are less exposed to cross-cutting content.
Adaptive Random Walk Controversy (ARWC) [12]	$[-1, 1]$	RWC, adjusting for community size in the number of influencers.
Betweenness Centrality Controversy (BCC) [12]	$[0, 1]$	High betweenness centrality on boundary edges indicates separation.
Boundary Polarization (BP) [13]	$[-0.5, 0.5]$	In polarized networks, authoritative users are further from the boundary.
Dipole Moment (DM) [23]	$[0, 1]$	Greater distance between positive and negative opinions indicates separation.
Krackhardt E/I Ratio (EI) [18]	$[-1, 1]$	Higher fraction of within-community edges indicates separation.
Adaptive E/I Index (AEI) [5]	$[-1, 1]$	EI, accounting for different community sizes.
Modularity (Q) [30]	$[-0.5, 1]$	More within-community edges than expected by chance indicates separation.

bias by designing a probing process tailored to our task. The resulting DSP measure can be interpreted as the average score of how likely a vertex is to receive information from a given community, where the average is taken over the distribution defined by our probing process. As a result, DSP is statistically more principled than RWC, which incorporates products of probabilities that do not have a clear statistical interpretation.

We validate our approach through extensive experiments on both synthetic and real-world networks. First, we develop a set of reference networks with prescribed values of structural polarization and confirm that DSP behaves as expected. On $G(n, p, \ell)$ networks, DSP exhibits near-zero scores as theoretically guaranteed. Second, for real-world networks, DSP distinguishes ideologically polarized communities from structurally similar but neutralized configurations, further demonstrating its robustness to group size imbalances and addressing limitations of both RWC and traditional metrics. Furthermore, we analyze the relationship between assortativity and polarization as captured by our new measure. Finally, we show that an approximate version of the measure can be computed more efficiently with good accuracy.

By integrating a statistically principled null model into its core formulation, DSP advances polarization measurement beyond ad hoc corrections. This approach answers a critical need for metrics grounded in explicit, theoretically sound baselines, which is essential for developing reliable diagnostics in an era of algorithmic fragmentation and polarized discourse. Our findings bridge key gaps in existing methodologies, enriching the toolkit for analyzing structural polarization and highlighting the value of robust measures to guide effective interventions in complex social landscapes.

2 Background and Preliminaries

Structural polarization measures aim at quantifying whether a given network represents a polarized system from its topology. Because polarization is a system-level phenomenon, these features are typically defined at the network level rather than the vertex level (such as individual behavioral mechanisms). Similar measures have been defined for economic systems, e.g., the Gini coefficient and the Theil index, as well as in other domains (e.g., spatial, occupational, and educational segregation); however, these measures deal with numerical data rather than networks. Table 1 reports the structural polarization measures most commonly used to analyse social networks, together with an intuition of the measure’s aim.

Structural polarization pipeline. Structural polarization measures are designed to be used as part of a broader pipeline [9, 29]. The typical pipeline comprises the following three stages:

Table 2: Summary of notation.

Symbol	Meaning
V	Set of vertices in the network
E	Set of edges in the network ($E \subseteq V \times V$)
A	Network adjacency matrix
n	Number of vertices ($n = V $)
R, B	Partitions of V ($R \cup B = V, R \cap B = \emptyset$)
$c(v)$	Partition of vertex v , referred to as its <i>color</i> ($c(v) \in \{R, B\}$)
ϕ_Z	Stationary distribution of a diffusion process rooted in $Z \subseteq V$ (e.g., Random Walk with Restart)

- (i) Define an appropriate directed network, e.g., a web graph, a friendship network, or an endorsement network.
- (ii) Partition this network into (typically two) separate communities, under the assumption that the network reflects polarized opinions along a single ideological dimension;
- (iii) Compute the structural polarization measure starting from the communities from the previous step.

The focus of the current work is on the third and last step. While this pipeline is commonly used, it is by no means the only way to use structural polarization measures. For instance, communities might already be defined in the data or derived from distant labels (e.g., semantically polarized hashtags).

2.1 Problem Setting and Notation

This section introduces our notation (summarized in Table 2). We consider a weakly-connected directed network $G = (V, E)$, where $n = |V|$. The set of vertices V is partitioned into two subsets, R and B , either algorithmically (e.g., by applying a graph-partitioning algorithm) or by metadata (e.g., by using vertex features). We refer to these sets as *communities*. We call *vertex color* the attribute of a vertex $v \in V$ that indicates membership to a given community, denoted as $c(v) \in \{R, B\}$. A denotes the adjacency matrix of the network G , i.e., $A_{ij} = 1$ if $(i, j) \in E$ and $A_{ij} = 0$ otherwise. If the edges of G have weights, they can be incorporated in the corresponding entries of the adjacency matrix A .

We assume a function $\phi_v : V \rightarrow \mathbb{R}^+$ that assigns a non-negative value to each vertex $w \in V$, parameterized by a vertex $v \in V$. The function ϕ_v is arbitrary, provided it does not depend on the colors of the vertices. Intuitively, it captures aspects of the network structure, so that more important vertices w tend to have higher values. Specifically, it represents a network diffusion process [21] and measures how information originating from v spreads.

As a concrete example, we consider the stationary distribution of the random walk with restart (RWR) from v . Mathematically, ϕ_v

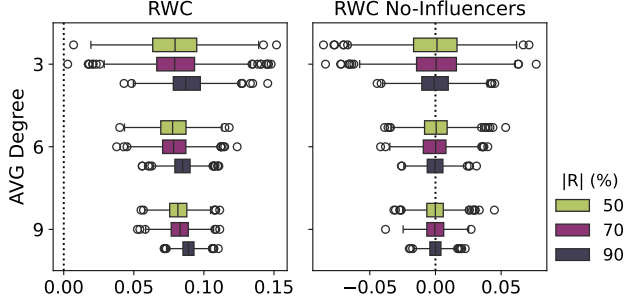


Figure 1: Polarization scores in 1000 random networks from $G(n, p, \ell)$, each with 10 000 vertices, varying average degree and partition sizes: 50% red–50% blue, 70% red–30% blue, and 90% red–10% blue. RWC (left) shows an unwarranted positive bias, caused by the overlap between the restart set and the influencers. Removing this overlap eliminates the bias, as shown for the “no-influencers” variant of RWC (right).

is the solution to the equation

$$\phi_v = \alpha A \phi_v + (1 - \alpha) \mathbb{1}_v,$$

where $\alpha \in [0, 1]$ is the follow-through probability (i.e., $1 - \alpha$ is the restart probability), and $\mathbb{1}_v$ is the indicator function for the set $\{v\}$. For $\alpha \rightarrow 1$, the RWR becomes a traditional random walk; its stationary distribution is proportional to degree centrality. For $\alpha \rightarrow 0$, the RWR remains trapped in the immediate neighborhood of v . Thus, this diffusion process can be understood as interpolating between local closeness and global degree centrality.

2.2 The RWC Measure and its Limitations

RWC [9, 12] is a widely-used structural polarization measure that informs the present work. At its core, RWC measures the relative exposure of a user to influential members of their own community versus those of an opposing community. The measure assumes that influencers are high-in-degree vertices and that information spreads via random walks.

Systematic bias. A key critique of RWC (as well as of other measures) is its systematic bias [29]. The measure reports positive polarization scores on random Erdős-Rényi networks, where edges are independent, thus no true structural division exists. Its scores on real-world networks are often preserved even after randomization with null models that preserve basic degree structure (for $d = 0$ and $d = 1$ using the dk-series terminology [20]). These results suggest RWC is sensitive to elementary network features such as size, density, and community balance, rather than capturing only genuine structural divisions. However, the proposed solution of standardizing the score by its values in a random graph null model [29] raises several crucial issues. First, it is unclear whether vertex identities (and therefore their labels) should be fixed when computing the measure in the samples from the null model, or if the labeling (i.e., the partitioning in the broader pipeline) is assumed to be part of the measure. Second, the solution, while widely applicable to all measures, is post-hoc. In addition, it requires considerable computational resources. A more principled solution would be to understand the causes of these biases in order to design a measure that is unbiased by construction.

Drawbacks of RWC. One of the main drawbacks of RWC is that it yields positive values on random network null models such as Erdős-Rényi [29]. Receiving positive scores on random networks puts the measure’s validity in doubt. The issue lies in how RWC handles its predefined set of “influencer” vertices. The measure is based on the probability of a random walk starting in one community (R or B), conditioned on it having reached an influencer in a given community (R^+ or B^+). However, a bias is introduced when the walk’s restart set overlaps with the target influencer set. This overlap artificially inflates the probability of a walk “staying” within its own community, as a walk starting from an influencer has already reached its destination.¹ As shown in Figure 1, a simple fix is to exclude influencers from the restart set, which corrects the bias for random graphs. Although this “no-influencers” variant works in this specific case, it is neither a principled nor a practical solution. This is because (i) there is no a priori way to define the set of influencers within each community, and (ii) there is no principled criterion for choosing how many influencers to include in the set. These limitations motivate our work: to design a measure that dispenses with the arbitrary notion of influencers and treats all vertices equally [3], leading to a more robust and theoretically sound assessment of structural polarization.

3 DSP: Improved Polarization Measure

This section introduces DSP, our new, principled measure for structural polarization. Similarly to previous polarization measures [9, 12], our measure employs as a core component a diffusion process over the network. In particular, we quantify how the probability mass of the diffusion process reaches the network communities when starting from different source vertices. DSP is designed to eliminate the biases of existing measures by removing the concept of “influencers” and instead considering every vertex as a potential source of diffusion. We start by defining a general diffusion process.

The diffusion process. For each vertex $v \in V$, consider a diffusion process starting from v , where the non-negative function $\phi_v : V \rightarrow \mathbb{R}^+$ (see section 2.1) measures how information originating from v spreads. The family of functions $(\phi_v)_{v \in V}$ is a parameter of our framework. An effective choice for ϕ_v , which we adopt in the remainder of this paper, is the random walk with restart (RWR) vector with restart at v . This choice connects our measure to the methodology of RWC, where RWR is used to approximate how a user’s endorsement is distributed across the network. However, other choices of ϕ_v are possible.

For a diffusion starting from a vertex $v \in V$, since we know that v is the source of the diffusion, we want to set its score equal to zero. Thus, to continue working with probabilities, for each $v \in V$, we define a probability distribution π_v over V , such that

$$\pi_v(w) \doteq \frac{\phi_v(w)}{\sum_{u \in V \setminus \{v\}} \phi_v(u)} \text{ for } w \neq v, \text{ and } \pi_v(v) \doteq 0.$$

Now, consider a thought experiment to model information flow in the network. First, we pick a “source” vertex S uniformly at random from V . Then, we pick a “target” vertex T by sampling from

¹In the original formulation, the set of influencers was negligible compared to the network (10 in networks with thousands of vertices), so the bias was not apparent. In smaller networks, or when using more influencers, this bias can be substantial.

the distribution π_S . This two-step process defines a joint probability distribution over all ordered pairs of distinct vertices $(v, w) \in V \times V$.

Let us now define a function $h_Q(v)$ as the probability that the source S of a diffusion reaching v belongs to community Q . Let, for any $Q \in \{R, B\}$ and $v \in V$,

$$h_Q(v) \doteq \Pr_{S,T}(S \in Q \mid T = v). \quad (1)$$

We refer to $h_Q(v)$ as the *exposure* of a single vertex v to a community $Q \in \{R, B\}$. Using Bayes' theorem, we can express $h_Q(v)$ as a function of the probability distributions π_v :

$$\begin{aligned} \Pr_{S,T}(S \in Q \mid T = v) &= \frac{\Pr(T = v \mid S \in Q) \Pr(S \in Q)}{\Pr(T = v)} \\ &= \frac{\left(\frac{1}{|Q|} \sum_{w \in Q} \pi_w(v) \right) \frac{|Q|}{n}}{\frac{1}{n} \sum_{w \in V} \pi_w(v)} = \frac{\sum_{w \in Q} \pi_w(v)}{\sum_{w \in V} \pi_w(v)}. \end{aligned}$$

The function h_Q depends directly on the user-specified family $(\phi_v)_{v \in V}$.

Example. In a political context, $h_R(v)$ represents the probability that a piece of content or information reaching user v originated from the community R . A comparatively higher value of $h_R(v)$ suggests that v is overly exposed to “red opinions” as they preferentially endorse that community. Conversely, if $h_R(v)$ and $h_B(v)$ are balanced, user v has a more diverse information diet.

Finally, we define a scoring function $\ell(Q, v)$ that captures whether the exposure of vertex v to community Q is a sign of polarization.

$$\ell(Q, v) \doteq \begin{cases} h_Q(v) & \text{if } v \in Q, \\ -h_Q(v) & \text{if } v \notin Q. \end{cases} \quad (2)$$

The intuition is that, when considering the exposure of a source of the same color as the target v (i.e., $v \in Q$), a high probability $h_Q(v)$ is a sign of structural homophily and contributes positively to polarization. When instead the source is of the opposite color ($v \notin Q$) a high probability $h_Q(v)$ is a sign of cross-cutting exposure and contributes negatively.

A probing process for polarization. While the $h_Q(v)$ functions describe exposure at the vertex level, a network-level polarization measure requires us to aggregate these values in a principled way. To do this, we define a *probing process* that samples three random variables: a target community Q_T , a probe target vertex Y , and a source community Q_S . Formally,

- Q_T is a color chosen uniformly at random from $\{R, B\}$.
- Y is a vertex chosen uniformly at random from community Q_T .
- Q_S is a community chosen from $\{R, B\}$ with a probability conditional on Y and defined as

$$\Pr(Q_S = R \mid Y) \doteq \frac{|B \setminus \{Y\}|}{n-1}, \text{ and } \Pr(Q_S = B \mid Y) \doteq \frac{|R \setminus \{Y\}|}{n-1}.$$

Intuition. As we mentioned before, this process is designed to test for polarization systematically without needing to define influencers. First, we select a target vertex Y without bias toward the larger community (by picking a color Q_T uniformly in $\{R, B\}$). Then, when we consider the origin of information, we choose a source community Q_S with probability proportional to the size of the *other* community. While this may seem counterintuitive, it serves to give more weight to cross-community influence in unbalanced communities, where otherwise, the smaller group could be drowned out. This design is

justified by the desirable properties of the resulting measure, such as producing a zero score for a fully connected clique regardless of the community sizes, as shown in Section 4.

The proposed polarization measure. Our new polarization measure combines the scoring function $\ell(Q, v)$ and the probing process defined above. In particular, DSP is defined as the expected value of the scoring function $\ell(Q, v)$ over the probing process:

$$\text{DSP} = \mathbb{E}_{Q_T, Y, Q_S} [\ell(Q_S, Y)]. \quad (3)$$

Intuition for DSP. In essence, DSP is the average score of how likely a vertex is to receive information from a given community. This average is over the distribution defined by our probing process.

Properties and expanded formulation. We now expand the expression for DSP to analyze its properties. By the law of total expectation, we can unroll the expectation in Equation (3).

$$\text{DSP} = \frac{1}{2} \sum_{Q \in \{R, B\}} \mathbb{E}_{Y, Q_S} [\ell(Q_S, Y) \mid Q_T = Q].$$

Applying the law of total expectation again, we get

$$\begin{aligned} \text{DSP} &= \frac{1}{2|R|} \sum_{v \in R} \mathbb{E}_{Q_S} [\ell(Q_S, v) \mid Y = v, Q_T = R] \\ &\quad + \frac{1}{2|B|} \sum_{v \in B} \mathbb{E}_{Q_S} [\ell(Q_S, v) \mid Y = v, Q_T = B]. \end{aligned}$$

Using the definition of conditional expectation, and the probability distribution for Q_S , we obtain

$$\begin{aligned} \text{DSP} &= \frac{1}{2|R|} \sum_{v \in R} \left(\frac{|B|}{n-1} h_R(v) - \frac{|R|-1}{n-1} h_B(v) \right) \\ &\quad + \frac{1}{2|B|} \sum_{v \in B} \left(\frac{|R|}{n-1} h_B(v) - \frac{|B|-1}{n-1} h_R(v) \right). \end{aligned}$$

Combining Equation (1) with the previous equation, and minimally rearranging the terms, we get

$$\text{DSP} = \quad (4)$$

$$\begin{aligned} &\frac{1}{2|R|} \left(\frac{|B|}{n-1} \sum_{v \in R} \frac{\sum_{w \in R} \pi_w(v)}{\sum_{w \in V} \pi_w(v)} - \frac{|R|-1}{n-1} \sum_{v \in R} \frac{\sum_{w \in B} \pi_w(v)}{\sum_{w \in V} \pi_w(v)} \right) \\ &+ \frac{1}{2|B|} \left(\frac{|R|}{n-1} \sum_{v \in B} \frac{\sum_{w \in B} \pi_w(v)}{\sum_{w \in V} \pi_w(v)} - \frac{|B|-1}{n-1} \sum_{v \in B} \frac{\sum_{w \in R} \pi_w(v)}{\sum_{w \in V} \pi_w(v)} \right). \end{aligned}$$

Beyond its analytical utility for studying the properties of the DSP measure, Equation (4) also provides a direct path for computation. The equation shows that calculating DSP boils down to computing four aggregate diffusion scores: within-community ($R \rightarrow R$, $B \rightarrow B$) and cross-community ($B \rightarrow R$, $R \rightarrow B$). When using random walk with restart (RWR) for the diffusion scores $\pi_v(w)$, it is required to compute the RWR vector restarting from each vertex $w \in V$. The values of these vectors at each target vertex v are then summed according to the community memberships of v and w . While computationally intensive, it is possible to use sampling-based approximations to speed up the computation.

Range. The range of DSP is $(-\frac{1}{2} \frac{n-2}{n-1}, \frac{1}{2} \frac{n}{n-1})$. The maximum is attained when diffusions from a community only reach vertices within that same community, the hallmark of extreme structural polarization. The minimum is attained when diffusions only reach

vertices of the opposite community. A key advantage of DSP is that negative values are clearly interpretable: they indicate that, on average, the network structure promotes cross-community diffusion more than within-community diffusion. This is a more intuitive interpretation than for previously proposed measures.

To see why, assume that for every $v \in R$ it is $\sum_{w \in R} \pi_w(v) = 1$, and for every $v \in B$ it is $\sum_{w \in B} \pi_w(v) = 1$. It is easy to see that these assumptions maximize the value of DSP to be

$$\frac{1}{2} \frac{n}{n-1} \rightarrow \frac{1}{2}, \text{ as } n \rightarrow +\infty.$$

The minimum value for DSP is

$$-\frac{1}{2} \frac{n-2}{n-1} \rightarrow -\frac{1}{2}, \text{ as } n \rightarrow +\infty,$$

attained when, for every $v \in R$ it is $\sum_{w \in B} \pi_w(v) = 1$, and for every $v \in B$ it is $\sum_{w \in R} \pi_w(v) = 1$.

More than two colors. It is possible to extend the DSP definition to the general case of $k \geq 2$ colors. Reusing the same notation as above, we need to define the following random variables:

- Q_T , a color chosen uniformly at random from the set of possible colors;
- Y , a vertex chosen uniformly at random from the community Q_T ;
- Q_S , a set of vertices chosen from the bi-partition $\{Q_T, V \setminus Q_T\}$, according to the following conditional probabilities

$$\Pr(Q_S = Q_T \mid Q_T) = \frac{|V \setminus Q_T|}{n-1}, \text{ and}$$

$$\Pr(Q_S = V \setminus Q_T \mid Q_T) = \frac{|Q_T| - 1}{n-1}.$$

Now, for any v and any possible community (i.e., color) Q , define the function

$$\ell(Q, v) \doteq \begin{cases} h_{c(v)}(v) & \text{if } Q = c(v), \\ -(1 - h_{c(v)}(v)) & \text{otherwise.} \end{cases}$$

Finally, we define

$$\text{DSP} \doteq \mathbb{E}_{Q_T, Y, Q_S} [\ell(Q_S, Y)]. \quad (5)$$

With $k \geq 2$ colors, the range of the DSP is roughly $(-1/k, 1/k)$. These definitions collapse to the previous ones when $k = 2$.

4 Analysis on Synthetic Data

In this section, we prove that DSP behaves as desired on a set of reference network topologies and network null models. As presented in the previous section, the DSP measure assumes a diffusion process over the network. In this section, for concreteness, we instantiate the generic ϕ_v functions using as diffusion process the random walk with restart (RWR) with restarting probability $1 - \alpha$.

In the following charts, to compare the various polarization measures while respecting their semantics, we normalized their values to a common interval. For measures that capture both the magnitude and direction of polarization, we normalize them to the interval $[-1, 1]$, preserving the neutral point at 0. For measures that quantify only the magnitude of polarization, we normalize them to the interval $[0, 1]$ to avoid introducing artificial directionality.

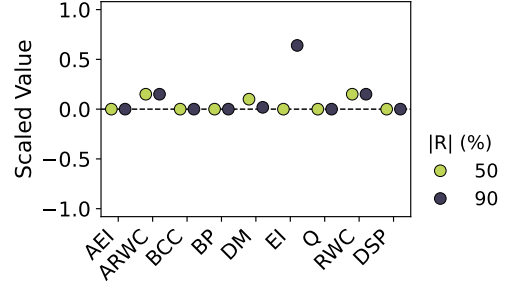


Figure 2: Polarization in a bi-colored clique with 5000 vertices and partitions of different sizes: 50% red–50% blue, and 90% red–10% blue. The dashed line denotse the desired value of 0.

4.1 Reference Network Topologies

We consider three types of networks: *cliques*, *color-alternating cycles*, and *monochromatic-splittable networks* (defined below). These specific networks have a simple enough structure that enables deriving the exact solution of the DSP polarization measure.

Clique. A clique of n vertices, arbitrarily partitioned into R and B , possibly with $|R| \neq |B|$, should exhibit no structural polarization: every vertex is connected to every other vertex, so every vertex has the same opportunities of accessing content (or interacting with vertices) from its own partition as it does from the other partition.

It holds that $\pi_v(w) = 1/(n-1)$ for every $v \in V$ and every $w \in V \setminus \{v\}$, and $\pi_v(v) = 0$ as required (see Section 3). The denominators of Equation (4) are therefore all equal to 1, while the four numerators are, in order, $(|R| - 1)/(n - 1)$, $|B|/(n - 1)$, $(|B| - 1)/(n - 1)$, and $|R|/(n - 1)$. Thus, we can express DSPas

$$\text{DSP} = \frac{1}{2|R|} \left(\frac{|B|}{n-1} |R| \frac{|R|-1}{n-1} - \frac{|R|-1}{n-1} |R| \frac{|B|}{n-1} \right) + \frac{1}{2|B|} \left(\frac{|R|}{n-1} |B| \frac{|B|-1}{n-1} - \frac{|B|-1}{n-1} |B| \frac{|R|}{n-1} \right) = 0.$$

Figure 2 shows the polarization values in a bi-colored clique with 5000 vertices and two partition ratios: one balanced (50% red, 50% blue) and one unbalanced (90% red, 10% blue). We observe that RWC and its adaptive variant ARWC exhibit a positive bias. In contrast, the other measures yield the desired polarization score of 0, except for DM, which also presents a positive bias on balanced partitions, and EI, which is sensitive to community size imbalance and yields positive scores when the partition is unbalanced.

Color-alternating Cycle. A color-alternating cycle is a cycle network with (even) n vertices, which are evenly split between B and R , and placed in an alternating fashion, i.e., each vertex v has exactly two neighbors, both of which have a color different from its own. A negative polarization is expected in this network, because each vertex is directly connected to only vertices of the other color, and thus it is more likely to access content (and interact with vertices) from the other color than its own.

The high level of symmetry in such a network implies that there is a positive $z = z(n, \alpha) \in (0, 1)$ such that, for every vertex $v \in V$, it holds that $h_{c(v)}(v) = z$ and $h_{\overline{c(v)}}(v) = 1 - z$. By plugging z in

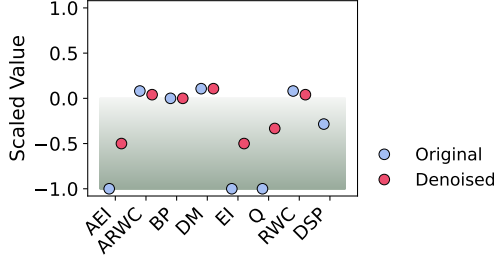


Figure 3: Polarization in a bi-colored alternating cycle with 5000 vertices, 50% red–50% blue. We show the rescaled values and the rescaled denoised values computed using the 1k-series [29]. The gradient area indicates the desired scores.

Equation (4), we obtain

$$\begin{aligned} \text{DSP} &= \frac{1}{2} \left(\frac{|B|}{n-1} z - \frac{|R|-1}{n-1} (1-z) \right) \\ &\quad + \frac{1}{2} \left(\frac{|R|}{n-1} z - \frac{|B|-1}{n-1} (1-z) \right) = z - \frac{1}{2} \frac{n-2}{n-1}. \end{aligned}$$

A lower value of α for ϕ_v (a RWR that restarts more often) leads to a lower value of z . As z is a probability, it is bounded below by zero, thus allowing us to recover the lower bound on the range of DSP as $-\frac{1}{2} \frac{n-2}{n-1}$ as α goes to zero.

Figure 3 shows the polarization values in a bi-colored alternating cycle with 5000 vertices, evenly split between red and blue. We also include the denoised [29] polarization scores. The chart displays a gradient area to indicate the desired direction of polarization. As in the clique experiment, RWC and ARWC display a positive bias. We observe that the denoising often worsens the results: measures that initially yield negative values become less negative—or even turn positive—after applying the proposed correction.

Monochromatic-splittable Networks. A monochromatic-splittable network is a network with a cut that splits the network into exactly two connected components, each monochromatic, i.e., containing only vertices of one of the two colors. As long as at least one of $|R|$ or $|B|$ is much larger than the size of the cut, these networks exhibit strongly positive polarization, because all but a few vertices of each color are connected only to vertices of the same color, and thus most interactions are between homochromatic vertices. As α decreases, then the components of Equation (4)

$$\sum_{v \in R} \frac{\sum_{w \in R} \pi_w(v)}{\sum_{w \in V} \pi_w(v)} \rightarrow |R|,$$

and similarly for B . The rationale is the same for the alternating cycle: smaller α values lead to more importance for the immediate neighbors of the starting vertex, which, in this case, are vertices with the same color as the starting vertex. Thus, the value of DSP tends to its maximum attainable value $\frac{n}{2(n-1)}$, as these are exactly the conditions that we assumed when analyzing the range of DSP (recall that $\text{DSP} \in (-1/2, 1/2)$ approximately).

When the two connected components are of the same size (i.e., $|R| = |B| = n/2$), and isomorphic, then a more precise analysis is possible. Examples of such networks are the half-split cycle network obtained by connecting both ends of two monochromatic chains (one red and one blue) of the same size, or a “barbell network” where

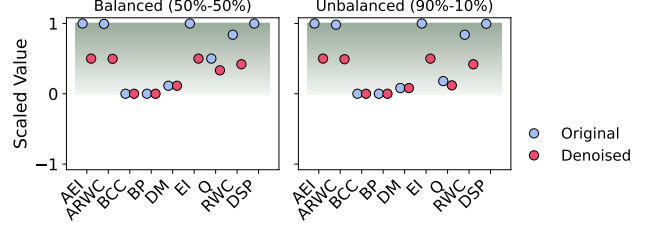


Figure 4: Polarization in a bi-colored half-split cycle with 5000 vertices and different partition sizes: 50% blue–50% red (left), and 90% red–10% blue (right). We show the rescaled values and the rescaled denoised values computed using the 1k-series [29]. The gradient area indicates the desired scores.

two monochromatic cliques of the same size are connected by a chain network that is half of one color and half of the other color. In these cases, symmetry implies that there are non-negative values $z_1 = z_1(n, \alpha)$ and $z_2 = z_2(n, \alpha)$ such that

$$\begin{aligned} \text{DSP} &= \frac{1}{2|R|} \left(\frac{|B|}{n-1} z_1 - \frac{|R|-1}{n-1} z_2 \right) + \frac{1}{2|B|} \left(\frac{|R|}{n-1} z_1 - \frac{|B|-1}{n-1} z_2 \right) \\ &= \frac{1}{n} \left(\frac{|B| + |R|}{n-1} z_1 - \frac{|R| + |B| - 1 - 1}{n-1} z_2 \right) = \frac{1}{n-1} z_1 - \frac{1}{n} \frac{n-2}{n-1} z_2. \end{aligned}$$

It holds

$$\begin{aligned} z_1 + z_2 &= \sum_{y \in R} \frac{\sum_{i \in R} \pi_i(y)}{\sum_{i \in V} \pi_i(y)} + \sum_{y \in R} \frac{\sum_{i \in B} \pi_i(y)}{\sum_{i \in V} \pi_i(y)} \\ &= \sum_{y \in R} \frac{\sum_{i \in R} \pi_i(y)}{\sum_{i \in V} \pi_i(y)} = |R| = \frac{n}{2}. \end{aligned}$$

Thus, we can express z_2 as $n/2 - z_1$. Continuing from Equation (4),

$$\text{DSP} = \frac{1}{n-1} z_1 - \frac{1}{n} \frac{n-2}{n-1} \left(\frac{n}{2} - z_1 \right) = \frac{2}{n} z_1 - \frac{1}{2} \frac{n-2}{n-1}.$$

The value z_1 , which is a function of n and α , increases as α decrease. As z_1 is upper bounded by $\frac{n}{2}$, we recover the upper bound on the range of DSP as $\frac{1}{2} \frac{n}{n-1}$.

Figure 4 presents the polarization values in a bi-colored half-split cycle with 5000 vertices, for two partitioning schemes: one with equal-sized red and blue groups, and another with a 90%–10% split. As before, the figure includes the rescaled denoised polarization scores computed using the 1k-series correction method. The gradient area highlights the desired direction of polarization. We observe that all measures yield positive scores for both partitioning scenarios; however, in several cases, the denoising approach reduces these values, hence worsening their performance. DSP consistently achieves the highest possible score, along with a few other measures.

Figure 5 reports results for a bi-colored half-split barbell network with 2000 vertices, again using both balanced and 90%–10% partitions. Rescaled denoised scores are also shown. The black dashed line represents the desired polarization value of 1. Most measures correctly reach this value in the balanced case, but performance tends to degrade—particularly after denoising—in the unbalanced setting. Overall, denoising [29] often moves the polarization scores away from their ideal values rather than correcting them. Across all scenarios, only two measures consistently align with the expected behavior: DSP and AEI.

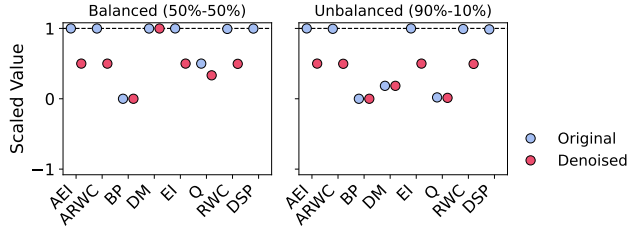


Figure 5: Polarization in a bi-colored half-split barbell network with 2000 vertices and different partition sizes: 50% blue–50% red (left), and 90% red–10% blue (right). We show the rescaled values and the rescaled denoised values computed using the 1k-series [29]. The dashed line denotes the desired value of 1.

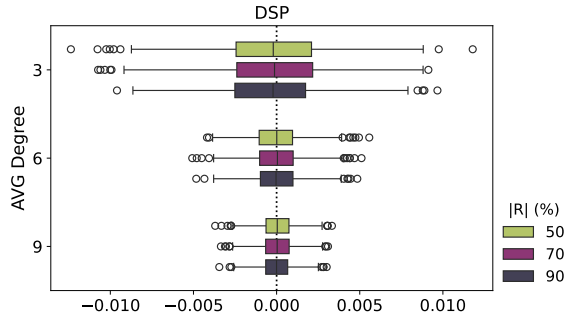


Figure 6: DSP in 1000 random networks extracted from $G(n, p, \ell)$ for different average degrees d and partition sizes $\ell(R) \in \{50\%, 70\%, 90\%\}$. We set $n = 10\,000$ and $p = d/n - 1$.

4.2 The $G(n, p, \ell)$ Null Model

The $G(n, p, \ell)$ model is an extension of the classical Erdős-Rényi random network $G(n, p)$, which generates a network with n vertices, each pair of which is connected with probability p ($0 \leq p \leq 1$). In this extension, we introduce labels for the vertices. Let $\mathcal{L} = [\ell_1, \ell_2, \dots, \ell_k]$ be the list of labels. The null model assigns the k labels uniformly at random to the vertices, subject to the constraint that exactly $\ell(i)$ vertices receive label ℓ_i .

The key property of this null model is that vertex labels are independent of the network structure, or, in other words, the edge placement process does not consider vertex labels. Thus, any observed association between labels and structural properties is purely due to chance. For this reason, any reasonable structural polarization measure should return, on expectation, zero or near-zero values.

Figure 6 shows the distribution of DSP values measured in 1000 random networks extracted from $G(n, p, \ell)$ with $n = 10\,000$, different average degrees (which determines p), and different partition skews. The average DSP score is close to 0 with minimal deviations for all average degrees and partition skews, proving the robustness of the measure to the partition size and network density.

Table 3 reports the average scaled scores for the other polarization measures. As expected, Q yields an average score of 0, since it uses an Erdős-Rényi network as its null model. Consistent with previous observations, RWC and ARWC exhibit a positive bias, as do EI, BCC, and DM. In contrast, BP shows a negative bias. Among the other measures, only AEI consistently produces the desired near-zero values across all conditions.

Table 3: Polarization in 1000 random networks from $G(n, p, \ell)$ for different average degrees d and partition sizes $\ell(R) \in \{50\%, 70\%, 90\%\}$. We set $n = 10\,000$ and $p = d/n - 1$.

d	$ R (\%)$	Metric							
		AEI	ARWC	BCC	BP	DM	EI	Q	RWC
3	50	0	0.115	0.298	-0.187	0.444	0	0	0.079
	70	0	0.115	0.299	-0.039	0.194	0.160	0	0.080
	90	-0.001	0.115	0.303	0.260	0.050	0.640	0	0.088
6	50	0	0.106	0.115	-0.416	0.264	0	0	0.078
	70	0	0.106	0.115	-0.184	0.076	0.160	0	0.079
	90	0	0.106	0.115	0.344	0.021	0.640	0	0.085
9	50	0	0.108	0.070	-0.572	0.214	0	0	0.082
	70	0	0.108	0.070	-0.342	0.052	0.160	0	0.083
	90	0	0.108	0.071	0.337	0.018	0.640	0	0.089

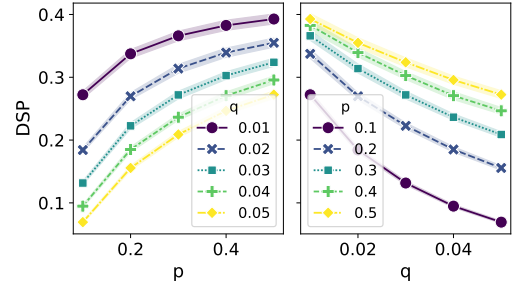


Figure 7: DSP measured in 100 networks sampled from the SBM with 1600 vertices and 2 blocks.

4.3 The Stochastic Block Model

To further test the robustness of DSP, we generate random networks using a stochastic block model (SBM) with n vertices assigned uniformly at random to 2 blocks. The degree distribution follows a Poisson distribution, with intra-block edge probability p and inter-block edge probability q . Higher values of p , especially when paired with lower values of q , correspond to higher structural polarization. In contrast, increasing q leads to greater inter-community connectivity, hence reducing structural polarization.

Figure 7 shows DSP for 100 random networks with 1600 vertices, varying p and q . DSP behaves as expected: it increases with denser intra-community connectivity and decreases as inter-community connectivity increases. For a fixed intra-community connectivity, lower inter-community connectivity gives higher scores.

5 Experiments on real-world data

In Section 4 we examined how DSP behaves on our reference networks, where the ground-truth polarization is known. In this section we consider real-world networks. First, we evaluate whether DSP can effectively distinguish between polarized and non-polarized networks and achieve classification performance comparable to that of existing polarization measures. Next, we study the performance of an approximate version of DSP that computes the measure considering only a sample of the network’s vertices. This approximation aims to reduce computational cost while maintaining accuracy. Then, we investigate the relationship between assortativity and DSP, via a null model that preserves the color assortativity of the

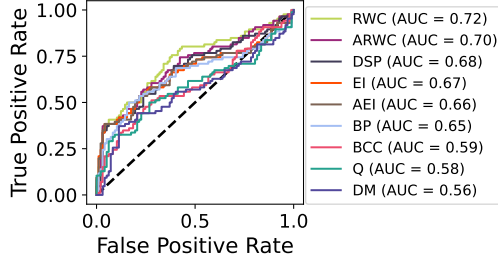


Figure 8: ROC curves and AUC values for SALLOOM.

given network [28]. Finally, we present a case study on political polarization using bill co-sponsorship data and roll-call voting records. The code is available on GitHub.²

Datasets. We consider several collections of real-world networks. SALLOOM [29] is a collection of 183 polarized and non-polarized Twitter retweet networks—150 constructed from single hashtags and 33 from multiple hashtags—collected during the 2019 Finnish Parliamentary Elections. CONGRESS-BILL-COSP [1] is a set of networks based on bill co-sponsorship data in the US Congress, covering the 93rd to the 114th Congresses. Each edge connects two legislators, with edge weights indicating the number of times they co-sponsored a bill or joint resolution [27]. CONGRESS-BILL-ROLL-CALL [19] is a set of networks constructed from roll-call voting records in the US Senate and House for the 93rd to 114th Congresses. As with the co-sponsorship dataset, we consider only bills and joint resolutions. Edges connect legislators who voted identically; edge weights denote the number of such instances.

For vertex labels, for CONGRESS-BILL-COSP and CONGRESS-BILL-ROLL-CALL, we use the legislators’ political parties (Democrat or Republican). For SALLOOM, we generate vertex labels using a graph partitioning algorithm, as commonly done in prior work [9]. We employ the Kernighan–Lin algorithm (KLIN) [17] and METIS [16].

Classification performance. We assess how well each polarization measure distinguishes polarized from non-polarized networks using the score as the output of a probabilistic classifier [29]. Each possible threshold gives a false positive rate and true positive rate, and by varying the threshold, we obtain a ROC curve that captures the performance of each score. In Figure 8 we report the unnormalized Area Under the Curve (AUC) as a summary metric of predictive accuracy. The performance of DSP is close to that of RWC, suggesting that DSP retains the discriminative power of RWC while offering a more principled foundation.

Approximate DSP. To evaluate the trade-off between accuracy and efficiency in computing the DSP score, we compare the exact score to a heuristic approximation obtained by calculating the summations in Equation (4) over a uniform random sample of the network vertices. Computing the RWR for every vertex in the network becomes computationally expensive as the network size increases. By considering only a subset of vertices, we aim to reduce computation time while incurring a small loss in accuracy. In this experiment, we consider a subset of 61 datasets from SALLOOM and, for each one, extract 50 random vertex samples with varying sample sizes. For each sample, we compute the approximate DSP score and then report the Mean Absolute Error (MAE).

²<https://anonymous.4open.science/r/diffusionBasedStructuralPolarization-1870>

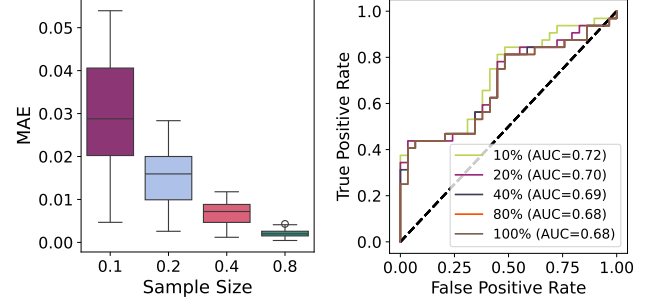


Figure 9: MAE of the approximate DSP scores computed using subsets of vertices of different size (left); and corresponding ROC curves (right).

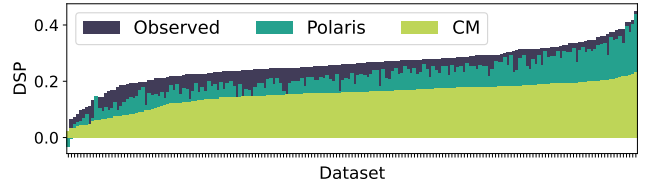


Figure 10: DSP scores and average values computed over 100 random networks from the configuration model (CM) and the Polaris model [28] on the SALLOOM dataset.

In our experiments below, we seek to estimate the empirical number of samples required to obtain a good approximation as a fraction of the network size. From a theoretical point of view, it is plausible that the required number of samples (sample complexity) is determined by a function that grows sublinearly with respect to the network size. We leave the question of theoretically deriving the sample complexity as future work. Figure 9 (top) shows that we need to retain at least 20% of the vertices to achieve a good approximation of the original score. To reduce variability in the approximation across samples, we should retain at least 40% of the vertices. Finally, the chart on the right shows that the ROC curves and the corresponding AUC values are roughly the same across sample sizes, indicating that even when using a small vertex set, the score has the same capability of distinguishing between polarized and non-polarized networks.

Color assortativity vs. polarization. We compare the DSP values measured on real datasets with those obtained from 100 samples generated using Polaris [28], an algorithm that samples from the ensemble of networks with the same degree sequence and color assortativity. Figure 10 illustrates how well DSP is preserved under the Polaris null model. The chart shows the value of DSP measured in the SALLOOM networks and the average values computed over 100 random networks from two null models: the configuration model (CM) and the Polaris null model. When the bars corresponding to the two null models are close, it indicates that color assortativity has little influence on the DSP value, since both models yield similar results despite one preserving assortativity and the other not. On the other hand, the closer the POLARIS bar is to the observed value, the more color assortativity can explain the polarization captured by DSP. A large gap between the observed value and the POLARIS baseline suggests that other structural or behavioral dynamics, beyond color assortativity, are contributing to the network’s polarization.

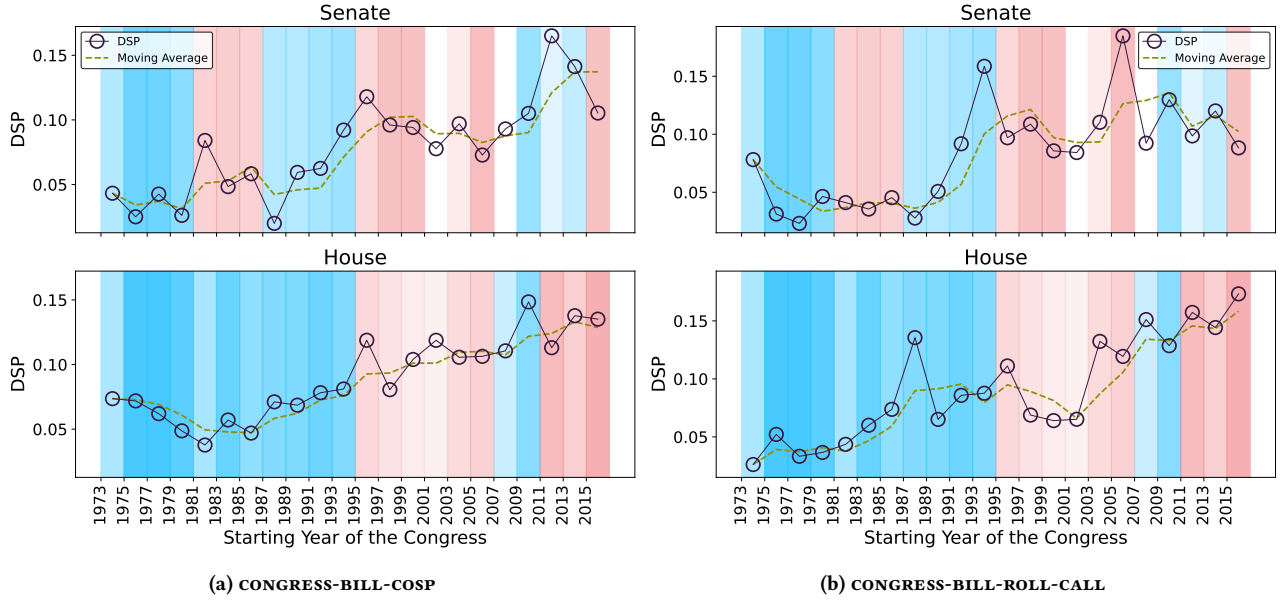


Figure 11: DSP and its 3-point moving average for both chambers across Congress sessions.

Overall, networks with higher polarization levels tend to show a stronger influence from color assortativity, as the POLARIS baseline is closer to the observed value as polarization increases. Meanwhile, the influence of the degree distribution remains relatively constant across datasets, as indicated by the consistent gap between the CM baseline and the observed values. Nonetheless, the differences between the observed and POLARIS values in less polarized networks suggest that color assortativity alone is insufficient to fully account for the observed polarization, especially when the score is low.

Polarization in the US Congress. We study how political polarization evolves and whether it correlates with the political control of the chambers of the US Congress [27]. We construct two sets of bi-colored networks: one based on bill co-sponsorship data (CONGRESS-BILL-COSP) and the other on roll-call voting data (CONGRESS-BILL-ROLL-CALL), for each chamber of Congress (Senate and House) and each session of Congress from the 93rd to the 114th. To reduce noise from widely co-sponsored legislation, we discard bills with more than 25 co-sponsors. We focus on two legislative types, i.e., Bills and Joint Resolutions, as both require passage by both chambers and the President’s signature to become law. These types are most likely to reflect strategic behavior relevant to polarization analysis.

We compute DSP on each of these networks. In this setting, the RWR vectors used in the computation of DSP incorporate edge weights, i.e., co-sponsorship/co-voting frequency influences the random walk behavior. This choice enables the polarization measure to capture stronger ties.

Figure 11 shows the DSP score for each session of Congress, separately for the Senate (top) and the House (bottom). The background color of each bar indicates which party held the majority in each chamber during that session, where darker colors correspond to stronger majorities. The figure also reports a moving average (window size 3) to smooth trends across Congress sessions (dashed line). Both co-sponsorship and roll-call voting data provide valuable yet distinct insights into the legislators’ preferences. Roll-call votes

are recorded decisions that offer a clear signal of preference. In contrast, bill co-sponsorship is a voluntary activity that indicates a positive disposition toward a bill; however, the absence of co-sponsorship has no clear interpretation. Despite these differences, previous work [2] showed that co-sponsorship data can produce estimates of ideal points that are highly correlated with those derived from roll-call votes. Consistently, we observe correlated DSP scores between the two types of congress networks, especially in the House, where both Pearson (0.73) and Spearman (0.70) are high. In the Senate, while Pearson is lower (0.59), Spearman remains high (0.74), indicating a strong agreement in trend.

Similarly, our findings on the bill-co-sponsorship networks align with previous studies [15], as we also observe low polarization levels (DSP remains below 0.16 across all sessions). This result is expected, as DSP is a structural measure, and thus, will yield high values only when the network presents a strong structural division. Nonetheless, consistent with previous research, we observe a general increase in polarization over time [10].

6 Conclusion

We introduce DSP, a statistically principled structural polarization measure, based on a probing process of information diffusion spreading as random walks from each vertex. It does not suffer from the biases exhibited by previous measures, as we show by analytically deriving DSP values on families of reference networks. The results of our experimental evaluation highlight how DSP can reliably differentiate between polarized and non-polarized networks. Interesting directions for future work include the derivation of additional properties of DSP on more reference classes, and the use of DSP to recommend edges to be added to the network to reduce polarization [11, 14].

Ethical Considerations

The proposed DSP measure is designed as a diagnostic tool to help researchers, platforms, and policymakers better understand and identify structural polarization, a phenomenon with significant societal consequences. However, like any technology that measures social dynamics, it can be used for purposes other than its intended goal. A primary concern is the potential for misuse by malicious actors. For instance, a sophisticated entity could leverage DSP to identify highly polarized and vulnerable communities to more effectively target them with tailored misinformation, deepening societal divisions for political or financial gain. Similarly, an authoritarian regime could use this measure to detect and justify the suppression of dissenting groups by labeling their cohesive, self-contained networks as a source of dangerous polarization, thereby using the metric as a pretext for censorship or control.

Beyond malicious use, harms can arise even when DSP is used as intended. A quantitative score, no matter how statistically principled, is an abstraction of complex social reality. There is a risk that a high polarization score could be used to stigmatize a community, leading to oversimplified judgments without a nuanced, qualitative understanding of that community's context. For example, a marginalized group may form a dense, insular network for mutual support, which could be misread as structural polarization. Furthermore, interventions designed to decrease a network's DSP score could have unintended negative consequences if applied naively. An automated system aimed at "depolarizing" a network might suggest interventions that disrupt vital community ties. To mitigate these risks, we stress that DSP should be used as a diagnostic aid, not a definitive verdict. Its findings must be interpreted within a broader socio-political context, and any interventions based on it should be human-centric and subject to ethical review to ensure they do not harm the communities they intend to help.

References

- [1] E Scott Adler and John D Wilkerson. 2013. *Congress and the politics of problem solving*. Cambridge University Press. 8
- [2] Eduardo Alemán, Ernesto Calvo, Mark P Jones, and Noah Kaplan. 2009. Comparing Cosponsorship and Roll-Call Ideal Points. *Legislative studies quarterly* 34, 1 (2009), 87–116. 9
- [3] Eytan Bakshy, Jake M. Hofman, Winter A. Mason, and Duncan J. Watts. 2011. Everyone's an Influencer: Quantifying Influence on Twitter. In *Proceedings of the Fourth ACM International Conference on Web Search and Data Mining (WSDM '11)*. Association for Computing Machinery, New York, NY, USA, 65–74. doi:10.1145/1935826.1935845 3
- [4] Michael J Barber and Nolan McCarty. 2015. Causes and consequences of polarization. *Solutions to political polarization in America* 15 (2015), 50. 1
- [5] Ted Hsuan Yun Chen, Ali Salloum, Antti Gronow, Tuomas Ylä-Anttila, and Mikko Kivelä. 2021. Polarization of climate politics results from partisan sorting: Evidence from Finnish Twittersphere. *Global Environmental Change* 71 (2021), 102348. 2, 11
- [6] Matteo Cinelli, Gianmarco De Francisci Morales, Alessandro Galeazzi, Walter Quattrociocchi, and Michele Starnini. 2021. The Echo Chamber Effect on Social Media. *Proceedings of the National Academy of Sciences* 118, 9 (2021), e2023301118. doi:10.1073/pnas.2023301118 1
- [7] Michael Conover, Jacob Ratkiewicz, Matthew Francisco, Bruno Gonçalves, Filippo Menczer, and Alessandro Flammini. 2011. Political polarization on twitter. In *Proceedings of the international aaai conference on web and social media*, Vol. 5. 89–96. 1, 11
- [8] Alessandro Cossard, Gianmarco De Francisci Morales, Kyriaki Kalimeri, Yelena Mejova, Daniela Paolotti, and Michele Starnini. 2020. Falling into the echo chamber: the Italian vaccination debate on Twitter. In *Proceedings of the International AAAI conference on web and social media*, Vol. 14. 130–140. 1
- [9] Kiran Garimella, Gianmarco De Francisci Morales, Aristides Gionis, and Michael Mathioudakis. 2016. Quantifying Controversy in Social Media. In *International Conference on Web Search and Data Mining (WSDM)*. ACM, 33–42. 2, 3, 8, 11
- [10] Kiran Garimella, Gianmarco De Francisci Morales, Aristides Gionis, and Michael Mathioudakis. 2017. The Effect of Collective Attention on Controversial Debates on Social Media. In *International Web Science Conference (WebSci)*. ACM, 43–52. 9
- [11] Kiran Garimella, Gianmarco De Francisci Morales, Aristides Gionis, and Michael Mathioudakis. 2017. Reducing Controversy by Connecting Opposing Views. In *International Conference on Web Search and Data Mining (WSDM)*. ACM, 81–90. 9
- [12] Kiran Garimella, Gianmarco De Francisci Morales, Aristides Gionis, and Michael Mathioudakis. 2018. Quantifying Controversy on Social Media. *ACM Transactions on Social Computing* 1, 1 (2018), 3. 1, 2, 3, 11
- [13] Pedro Guerra, Wagner Meira Jr, Claire Cardie, and Robert Kleinberg. 2013. A measure of polarization on social media networks based on community boundaries. In *Proceedings of the international AAAI conference on web and social media*, Vol. 7. 215–224. 1, 2, 11
- [14] Shahrzad Haddadan, Cristina Menghini, Matteo Riondato, and Eli Upfal. 2021. RePubBubLik: Reducing Polarized Bubble Radius with Link Insertions. In *Proceedings of the Fourteenth ACM International Conference on Web Search and Data Mining (WSDM '21)*. ACM, 139–147. 9
- [15] Marilena Hohmann, Karel Devriendt, and Michele Coscia. 2023. Quantifying ideological polarization on a network using generalized Euclidean distance. *Science Advances* 9, 9 (March 2023), eabq2044. doi:10.1126/sciadv.abq2044 Publisher: American Association for the Advancement of Science. 9
- [16] George Karypis and Vipin Kumar. 1997. METIS: A software package for partitioning unstructured graphs, partitioning meshes, and computing fill-reducing orderings of sparse matrices. (1997). 8, 11
- [17] Brian W Kernighan and Shen Lin. 1970. An efficient heuristic procedure for partitioning graphs. *The Bell system technical journal* 49, 2 (1970), 291–307. 8, 11
- [18] David Krackhardt and Robert N Stern. 1988. Informal networks and organizational crises: An experimental simulation. *Social psychology quarterly* (1988), 123–140. 1, 2, 11
- [19] Jeffrey B Lewis, Keith Poole, Howard Rosenthal, Adam Boche, Aaron Rudkin, and Luke Sonnet. 2019. Voteview: Congressional roll-call votes database. See <https://voteview.com/> (accessed 27 July 2018) (2019). 8
- [20] Priya Mahadevan, Dmitri Krioukov, Kevin Fall, and Amin Vahdat. 2006. Systematic topology analysis and generation using degree correlations. *ACM SIGCOMM Computer Communication Review* 36, 4 (Aug. 2006), 135–146. doi:10.1145/1151659.1159930 3
- [21] Naoki Masuda, Mason A. Porter, and Renaud Lambiotte. 2017. Random Walks and Diffusion on Networks. *Physics Reports* 716–717 (Nov. 2017), 1–58. doi:10.1016/j.physrep.2017.07.007 2
- [22] Antonis Matakos, Evimaria Terzi, and Panayiotis Tsaparas. 2017. Measuring and moderating opinion polarization in social networks. *Data Mining and Knowledge Discovery* 31, 5 (2017), 1480–1505. 1
- [23] Alfredo Jose Morales, Javier Borondo, Juan Carlos Losada, and Rosa M Benito. 2015. Measuring political polarization: Twitter shows the two sides of Venezuela. *Chaos: An Interdisciplinary Journal of Nonlinear Science* 25, 3 (2015). 1, 2, 11
- [24] Dimitar Nikolov, Alessandro Flammini, and Filippo Menczer. 2021. Right and Left, Partisanship Predicts (Asymmetric) Vulnerability to Misinformation. *Harvard Kennedy School Misinformation Review* (2021). doi:10.37016/mr-2020-55 1
- [25] Keith T Poole and Howard Rosenthal. 2001. D-nominate after 10 years: A comparative update to congress: A political-economic history of roll-call voting. *Legis. Stud. Q* 26 (2001), 5. 1
- [26] Giulia Preti, Gianmarco De Francisci Morales, and Matteo Riondato. 2022. ALICE and the Caterpillar: A More Descriptive Null Model for Assessing Data Mining Results. In *International Conference on Data Mining (ICDM)*. IEEE. 1
- [27] Giulia Preti, Adriano Fazzzone, Giovanni Petri, and Gianmarco De Francisci Morales. 2024. Higher-Order Null Models as a Lens for Social Systems. *Physical Review X* 14, 3 (Aug. 2024), 031032. doi:10.1103/PhysRevX.14.031032 8, 9
- [28] Giulia Preti, Matteo Riondato, Aristides Gionis, and Gianmarco De Francisci Morales. 2025. Polaris: Sampling from the Multigraph Configuration Model with Prescribed Color Assortativity. In *Proceedings of the Eighteenth ACM International Conference on Web Search and Data Mining*. 30–39. 8
- [29] Ali Salloum, Ted Hsuan Yun Chen, and Mikko Kivelä. 2022. Separating Polarization from Noise: Comparison and Normalization of Structural Polarization Measures. *Proceedings of the ACM on Human-Computer Interaction* 6, CSCW1 (March 2022), 1–33. doi:10.1145/3512962 1, 2, 3, 6, 7, 8
- [30] Andrew Scott Waugh, Liuyi Pei, James H Fowler, Peter J Mucha, and Mason A Porter. 2009. Party polarization in congress: A network science approach. *arXiv preprint arXiv:0907.3509* (2009). 2, 11

A Structural Polarization Measures

The Betweenness Centrality Controversy (BCC) [12] measure compares the edge betweenness centrality of boundary and non-boundary links, by computing the KL-divergence d_{KL} between the two corresponding distributions:

$$BCC \doteq 1 - \exp^{-d_{KL}}.$$

The intuition is that if the two communities are strongly separated, then the links on the boundary are expected to have high edge betweenness centralities.

The Boundary Polarization (BP) [13] compares the concentration of high-degree nodes within the communities (I) and their concentration in the boundary (C):

$$BP \doteq \frac{1}{|C|} \sum_{s \in C} \frac{d_I(s)}{d_C(s) + d_I(s)} - 0.5,$$

where $d_C(s)$ is the degree of s restricted to neighbors in C . The intuition is that the further away authoritative users are from the boundary, the larger the amount of polarization present in the network.

The Dipole Moment (DP) [23] measure applies label propagation for quantifying the distance between the top- k influencers of each community, which are assigned the extreme opinion scores of -1 or 1. Let gc^+ and gc^- denote the average of positive and negative opinion scores. The DP measure is a function of the distance between the means of the opposite opinion score distributions, rescaled to penalize differences in the community sizes:

$$DM \doteq \left(1 - \frac{n^+ - n^-}{n^+ + n^-}\right) \frac{|gc^+ - gc^-|}{2}.$$

The Krackhardt E/I Ratio (EI) [18] is defined as the relative density of intra-edges compared to the number of inter-edges:

$$EI \doteq \frac{EL - IL}{EL + IL},$$

where EL is the set of edges with endpoints belonging to different communities, and IL is the set of edges with endpoints belonging to the same community.

The Adaptive E/I Index (AEI) [5] extends the EI measure to account for communities with different sizes:

$$AEI \doteq \frac{\sigma_{aa} + \sigma_{bb} - 2 * \sigma_{ab}}{\sigma_{aa} + \sigma_{bb} + 2 * \sigma_{ab}},$$

where σ_{aa} is the ratio between the number of intra-edges in community a and the number of potential intra-edges, and σ_{ab} is the ratio of the number of inter-edges between community a and b and the number of potential inter-edges.

Modularity (Q) [30] compares the connectivity of the communities to that observable in random graphs extracted from the configuration model:

$$Q \doteq \frac{1}{2|E|} \sum_{i,j \in V} \left(A_{i,j} - \frac{d(i)d(j)}{2|E|} \right) \delta(i,j),$$

where A is the adjacency matrix of the graph and $\delta(i,j)$ equals one only when i and j belong to the same community.

B Additional Datasets

GARIMELLA [9] contains 10 controversial and 10 non-controversial Twitter retweet networks, constructed from Tweets collected from Feb 27 to Jun 15, 2015, using sets of related hashtags, each anchored by one manually selected hashtag. CONOVER [7] consists of two networks constructed from tweets collected between Sep 14 and Nov 1, 2010, ahead of the U.S. congressional midterm elections. Starting from the two popular political hashtags #p2 (“Progressives 2.0”) and #tcot (“Top Conservatives on Twitter”), a set of 66 related hashtags was identified and used to create a retweet and a mention network. For node labels, in CONOVER we use the node labels provided by the authors, and for GARIMELLA, we generate node labels using the Kernighan–Lin (KLIN) [17] and METIS [16] partitioning algorithms.

C Additional Experiments

More on Classification Performance. Figure 12 shows that the performance of DSP is comparable to that of RWC.

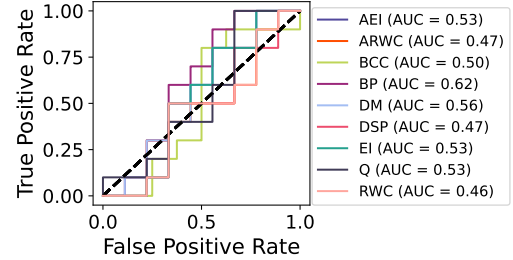


Figure 12: ROC curves and AUC values for GARIMELLA.

Similar results can be observed for the CONOVER networks (Table 4): higher values are recorded for the polarized RETWEET network, while values close to the minimum for the non-polarized MENTION network (see Figure 13).

Table 4: Polarization scores for the CONOVER networks.

Dataset	Metric								
	AEI	ARWC	BCC	BP	DM	EI	Q	RWC	DSP
MENTION	0.306	0.150	0.609	-0.030	0.643	0.315	0.101	0.101	0.057
RETWEET	0.959	0.837	0.832	0.257	0.768	0.954	0.475	0.869	0.410

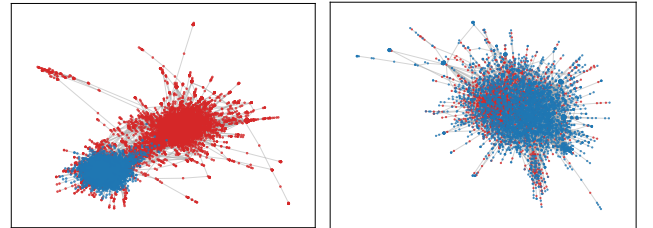


Figure 13: RETWEET (left) and MENTION (right) networks.

Impact of α on DSP. Figure 14 displays the DSP scores computed on bi-colored alternating cycles with various numbers of vertices, evenly split between red and blue, using different values of the parameter α that controls the restart probability in the PPR computation. Lower values of α correspond to more frequent restarts, making the PPR vector more influenced by the immediate neighbors of the starting node. Higher values of α emphasize more central

nodes. We observe that DSP is sensitive to the choice of α : as α decreases, the DSP score also decreases. This behavior is expected, as each node's immediate neighbors belong to the opposite community. When the diffusion process is more strongly influenced by neighbors, more probability mass flows across community boundaries, leading to a lower polarization score.

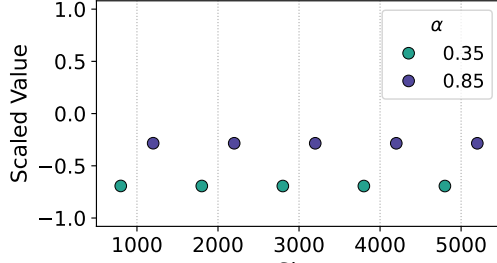


Figure 14: DSP in bi-colored alternating cycles with various numbers of vertices, 50% red and 50% blue. We show the rescaled values for different values of α .

In addition, Figure 15 shows the value of DSP for the SALLOUM networks, for two values of α . Although the polarization scores are slightly lower for larger values of α , the classification performance of DSP remains stable (0.68 using $\alpha = 0.85$ and 0.64 using $\alpha = 0.35$).

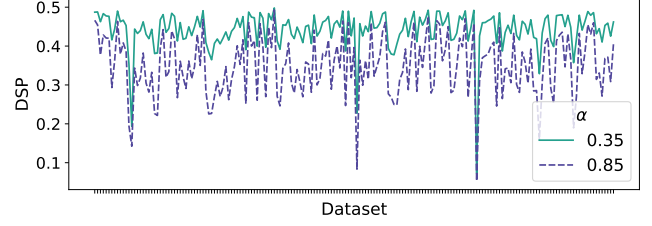


Figure 15: DSP for SALLOUM, varying the value of α .

More on Polarization and Assortativity. Figure 16 shows the distribution of the DSP values in 100 graphs sampled using Polaris, together with the color assortativity of the observed datasets. We report results for a subset of datasets for visualization purposes. This chart helps understand whether higher assortativity tends to correspond to higher polarization scores. We find that this relationship generally holds across datasets.

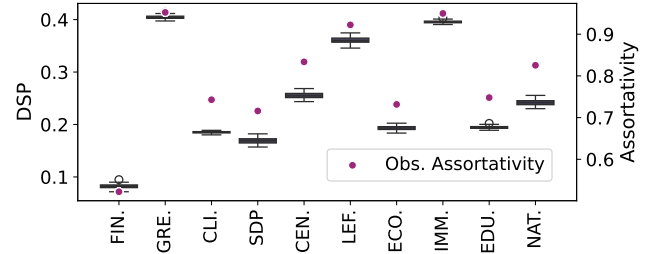


Figure 16: SALLOUM: DSP scores in the samples generated using Polaris and assortativity of the observed datasets.