CORE
Metadata, citation and similar papers at core.a
Provided by Elektronische Publikationen der Wirtschaftsuniversität Wien

Department of Strategy and Innovation
Working Paper No. 13/2020

Polarization in Networks: Identication-alienation Framework

Kenan Huremović
Ali Ozkes

September 2020

WIRTSCHAFTS UNIVERSITÄT WIEN VIENNA UNIVERSITY OF ECONOMICS AND BUSINESS

EFMD EQUIS ACCREDITED
AACSB ACCREDITED
ASSOCIATION OF MBAS ACCREDITED

Polarization in Networks: Identification-alienation Framework*

Kenan Huremović† Ali Ozkes‡

September 16, 2020

Abstract

We introduce a model of polarization in networks as a unifying setting for the measurement of polarization that covers a wide range of applications. We consider a substantially general setup for this purpose: node- and edge-weighted, undirected, and connected networks. We generalize the axiomatic characterization of Esteban and Ray (1994) and show that only a particular instance within this class can be used justifiably to measure polarization in networks.

JEL codes: D63, D70, P16

Keywords: measurement, networks, polarization

*We thank Joan Esteban, Jan Fałkowski, and Fernando Vega-Redondo for their comments and to Debraj Ray for our discussion at the beginning of the project. K. Huremović acknowledges financial support from ANR (project ANR 18-CE26-0020-01).

†IMT School for Advanced Studies Lucca, Piazza S. Francesco, 19, 55100 Lucca, Italy. E-mail: kenan.huremovic@imtlucca.it.

‡Wirtschaftsuniversität Wien, Institute for Markets and Strategy, Welthandelsplatz 1, 1020, Vienna, Austria. E-mail: ali.ozkes@wu.ac.at

1 Introduction

Polarization in a population denotes an intensified disconnect among its groups. The analysis of the sources and the consequences of polarization depends highly on what is measured and how, which, in turn, is strictly contingent on the particular context. For instance, while in the context of American politics polarization is perceived as the division of masses into the cultural camps of liberals and conservatives, in the context of European multi-party parliaments, it is seen as the existence of ideologically cohesive and distinct party blocks.[1] So even the term "political polarization" is not indicative of what is being measured and how. Existing literature reflects this complexity, and there is an abundance of measures without a unified formalism that applies to comparable contexts.

Although there are substantial differences among existing measures across fields, one ubiquitous feature can be identified. Namely, most of the current measures are proposed in settings with a uni-dimensional scalar attribute on which the polarization is assumed to occur. However, conflicts in societies are in general related to an irreducibly complex set of attributes and most of the empirical work rely on categorical data on various characteristics.[2] Dimensionality reduction approaches are called for in many instances, because the existing polarization measures allow for only a uni-dimensional, or at most a bi-dimensional domain (Hill and Tausanovitch, 2015). However, reduced dimensions can be questionable for their capacity to represent the actual phenomenon of interest (Kam et al., 2017).

In this paper, we propose the formalism of network theory to study the measurement of polarization which delivers the desired generality and spans a large variety of contexts. We fully characterize a polarization measure following the axiomatic setting introduced by Esteban and Ray (1994) (henceforth ER) for distributions on the real line. Same as ER, we restrict ourselves to distributions with finite support.

Our setup is built on undirected networks in which both nodes and links are weighted. A node in the network represents a certain attribute or grouping of individuals in the population. The weight of a node corresponds to the number of individuals in the population that are characterized by the attribute or members of the group (*e.g.,* a political party). The weighted links describe (direct) bilateral relationships between nodes. This setup is quite general and can represent a wide range of settings in which measuring polarization is an issue of first-order importance. We describe a number of important examples in the

[1] See Fiorina et al. (2005) and Maoz and Somer-Topcu (2010) for the two different contexts.

[2] Examples include ethnolinguistics (Montalvo and Reynal-Querol, 2008), ethnic power relations (Wimmer et al., 2009), and political retweets (Conover et al., 2011).

next section, with a particular focus on the political domain, not only because it is a central point of discussion, but also because it comprises of a variety of aspects that can be captured distinctly within network formalism. For instance, we show how elite polarization can be modeled within our framework through networks of politicians, parties, or policy space. Mass polarization, on the other hand, can be modeled through a network of opinions or preferences. Going beyond the political domain, we furthermore discuss how our approach can be used to study polarization in any setting with multidimensional distributions with finite support.

The axiomatic approach developed by ER for distributions with finite support on the real line led to the development of measures in several other domains, such as measures for continuous distributions as in Duclos et al. (2004) and measures for binary classifications as in Montalvo and Reynal-Querol (2008) (henceforth MRQ). Most of the applications employing measures within this line of work lie in the fields of income inequality and social conflicts. In their seminal contribution in this context, ER conceptualize polarization as the aggregate antagonism in a population. The effective antagonism an individual feels against another individual depends on how alienated she feels from the other's group and how identified she feels with her own group. According to ER, a population in which individuals are identified within groups is polarized if there is a high level of intra-group homogeneity, a high level of inter-group heterogeneity, and a small number of large-enough groups. They deliver a characterization of a class of polarization measures, based on an axiomatization built around distributional properties and not confined to incomes or wealth, although the main motivations of ER were about income and wealth distributions.

Following ER, we provide an axiomatic characterization for measures of network polarization. We argue that networks represent a powerful tool to capture any distribution with a finite support and a notion of distance. Thus, the strength of our contribution lies in the fact that we deliver an axiomatic foundation for a family of measures that are applicable in a significantly larger set of domains. Furthermore, as any distribution considered in ER or MRQ can be represented as a network, our work can be seen as a unifying generalization, with ER and MRQ as special cases.

The class of measures characterized by ER is identified by the range of values that parameter α, which captures the importance of identification in the effective antagonism, can take. Our first result shows, quite surprisingly, that this class is thinned down by a unique value, *i.e.,* $\alpha = 1$ (Theorem 1).[3] Note that adaptations of these axioms are neither trivial nor straightforward, as networks allow for a much larger generality in representing

[3]MRQ also identify $\alpha = 1$ in their setup, which is a special case of ours. Furthermore, their axiomatization is different than ours and ER.

discrete distributions than the real line. Recent literature on the measurement of polarization carried along the restricting assumption that the attributes can be captured by the values of a scalar variable. We take off where ER leave, and deliver an analysis that does not "sweep a serious dimensionality issue under the rug" (ER, p. 823). Our approach accommodates a significantly larger variety of settings that are not confined to scalar attributes, and naturally include the case of the Euclidean distance on the real line as a special case. This entails a solution to an unresolved issue in this line of research as a by-product, in that our results point to the choice of an exact value within the interval $(0, \alpha^* \simeq 1.6]$.[4]

It is desirable that polarization measures attain their maximum at the symmetric bipolar distribution. Contrary to the real intervals, in networks there can be any finite number of nodes with maximal distance between them. Still, we show that any measure within the family we characterize is maximized at the symmetric bipolar distribution — when the population is symmetrically distributed among the *two* most distant nodes in the network (Proposition 1).

Finally, we show that if we restrict our attention to particular classes of networks emerging in certain domains such as language trees (class of tree networks) or income distributions (class of line networks), one of the axioms, Axiom 3, can be weakened in a systematic way to allow for a wider class of measures that can be used consistently (Theorem 2). For instance, in the special case of line networks that can be used to represent income distributions, our set of axioms and the class of measures reduce to the ones in ER.

Related literature

It presents a challenge to pay a fair tribute to the ever-growing literature on the measurement of polarization. Here, we refer to a set of papers in different domains and discuss a few closely related ones. We mention several other works in Section 5.

Polarization is studied in social sciences (particularly in economics and political science) in relation to economic inequality (Esteban et al., 2007, Esteban and Ray, 2012, Zhang and Kanbur, 2001), social conflict (Desmet et al., 2017, Montalvo and Reynal-Querol, 2008, Østby, 2008), political economy (Aghion et al., 2004, Desmet et al., 2012, Lindqvist and Östling, 2010), international relations (Maoz, 2006b), political ideologies (Abramowitz and Saunders, 2008, Fiorina and Abrams, 2008, Lelkes, 2016, Martin and Yurukoglu, 2017, Ozdemir and Ozkes, 2014), political sentiments (Boxell et al., 2017, Garcia et al., 2015), and social attitudes (DiMaggio et al., 1996, Lee et al., 2014, McCright and Dunlap, 2011), among others.

[4]ER proposes further restrictions in that regard by imposing an additional axiom (Axiom 4) that brings about a lower bound, *i.e.,* $\alpha \geq 1$.

We want to emphasize that we are not the first to consider an ER-type approach to the measurement of polarization in networks. For instance, both Esteban and Ray (1999) and Esteban and Ray (2011) explore this issue. However, to the best of our knowledge, this is the first paper to provide an axiomatic characterization for measures of polarization in networks.[5] Fowler (2006a,b) and Maoz (2006b) are among the leading examples where network formalism is proposed for the measurement of polarization, without an axiomatic treatment.[6] Finally, Permanyer and D'Ambrosio (2015) characterize a distinct family of measures for categorical attributes by using identification-alienation framework and a number of additional axioms.

The rest of the paper is organized as follows. In Section 2 we describe the environment we study and illustrate the wide applicability of our approach. In Section 3 we define polarization, state the axioms, and deliver our major results. In Section 4 we discuss the importance of network structure in terms of polarization and formally illustrate the connection between our work and previous literature. We conclude in Section 5.

2 Networks and polarization

We consider a population in which individuals belong to $n > 0$ mutually exclusive groups of potentially different sizes. A group may be, for instance, a political party, ethnic group, or a set of individuals that share the same attributes. For each group i, $\pi_i \geq 0$ denotes the number of individuals in group i. When $\pi_i = 0$ we say that group i is *empty*. Vector $\boldsymbol{\pi} \in \mathbb{R}^n_{\geq 0}$ describes the distribution of a population among n groups.

Bilateral relationships between groups are described with an undirected weighted graph (UWG) $\mathbf{g}$, with the set of nodes, $N(\mathbf{g})$, equal to the set of groups, and the set of undirected links (or edges) $E(\mathbf{g}) = \{\{i,j\} : \{i,j\} \in N^2 \text{ and } i \neq j\}$. As usual, we denote the edge between nodes i and j in graph $\mathbf{g}$ with ij and the weight of that edge with $g_{ij} \geq 0$. While we treat weights quite generally, it is useful to think of g_{ij} as the direct distance between two connected nodes i and j – a higher g_{ij} implies a *weaker* connection between i and j.[7] For the remaining part of the paper we write $ij \in \mathbf{g}$ instead of $\{i,j\} \in E(\mathbf{g})$ to indicate that there is an edge between nodes i and j in $\mathbf{g}$. When nodes i and j are not directly connected,

[5] Esteban and Ray (1999) arrive at $\alpha = 1$ in their attempt to connect the intensity of conflict to polarization, without an axiomatic discussion, while Esteban and Ray (2011) supplement the four axioms in Duclos et al. (2004) with a fifth axiom that delivers $\alpha = 1$.

[6] The measure Maoz (2006b) uses is developed in the unpublished working paper by Maoz (2006a), and while inspired by Duclos et al. (2004), it is only shown to satisfy an extended and qualitatively different set of properties.

[7] The particular interpretation of weights $(g_{ij})_{i,j \in N}$ depends on the application, as we demonstrate in Section 2.1.

we write $ij \notin \mathbf{g}$. Moreover, since groups are represented as nodes, we use words *group* and *node* interchangeably.

We restrict our attention to connected graphs, *i.e.*, graphs in which there is a path connecting any two nodes.[8] The distance between nodes i and j in $\mathbf{g}$, denoted with $d_{\mathbf{g}}(i,j)$, is measured using the notion of the shortest path. That is, while there may be different routes one can take to reach node j starting from node i and moving along the links in $\mathbf{g}$, the distance between i and j is the length of the shortest path. This notion of distance, also known as the *geodesic distance*, is the standard in graph theory and the theory of networks (Newman, 2003, Jackson, 2008).

Let $\mathscr{G}_n$ denote the set of all UWGs with n nodes, and let $\{\mathscr{G}_n\}_{n\in\mathbb{N}}$ denote the family of all UWGs with any finite number of nodes. The main object of our analysis is the ordered pair $(\mathbf{g},\boldsymbol{\pi}) \in \mathscr{G}_n \times \mathbb{R}^n_{\geq 0}$, which represents a weighted (node-weighted and link-weighted) network. We use $\mathscr{N}$ to denote the set of all networks with finite number of nodes.

In the special case when $\boldsymbol{\pi} = \mathbf{1}$, $(\mathbf{g},\boldsymbol{\pi})$ coincides with the standard notion of a (link-) weighted network.[9] If, additionally, $g_{ij} = 1$ whenever $ij \in \mathbf{g}$, then $(\mathbf{g},\boldsymbol{\pi})$ is a binary network. Thus $(\mathbf{g},\boldsymbol{\pi})$ is a fairly general object that can be used to represent any undirected network we observe, allowing for weights on nodes and edges. In Section 4 we show that any distribution studied in ER or any classification covered by MRQ can be represented as a network.

A polarization measure is a mapping $\mathscr{P} : \mathscr{N} \to \mathbb{R}_{\geq 0}$ that assigns to each network $(\mathbf{g},\boldsymbol{\pi}) \in \mathscr{N}$ a non-negative real number.

Before turning to the axiomatic analysis, we discuss a number of examples in which data can be represented as a network and measuring polarization is of interest.

2.1 Examples

2.1.1 Polarization in political networks

We consider several networks that arise in politics, each of which encodes a different aspect of the prevailing political climate. In particular, we consider situations in which collection of individuals express their preferences over alternatives, natural examples of which include a parliament voting on bills and an electorate choosing among candidates. We discuss how these two can be modeled as networks in order to measure elite and mass polarization.[10]

We start with the case of a parliament with possibly more than two parties. Let there

[8] We consider only connected graphs in this paper. Our insights can be extended, in a somewhat ad-hoc manner, to cases when $\mathbf{g}$ is unconnected, for instance, by defining the distance between nodes from different components of $\mathbf{g}$ to be equal to the longest path between any two connected nodes in $\mathbf{g}$.

[9] Alternatively, one can think of $(\mathbf{g},\boldsymbol{\pi})$ as a distribution $\boldsymbol{\pi}$ on graph $\mathbf{g}$.

[10] See Kearney (2019) for a review focusing on networks in the political domain from a general perspective.

be $N \in \mathbb{N}$ representatives denoted by $\mathscr{R} = \{1, ..., N\}$ and $T \in \mathbb{N}$ parties denoted by $\mathscr{T} = \{t_1, \ldots, t_T\}$. Suppose there are $k \in \mathbb{N}$ bills that are sponsored by representatives, either individually or in groups, which are thereafter voted for approval in the parliament. Let $v_{ij} \in \{0, 1\}$ denote the vote of i for the bill $j \in \{1, 2, ..., k\}$ and $\mathscr{V} = \{0, 1\}^k$ denote the set of possible vote combinations.

Network of representatives, $(\mathbf{g}', \boldsymbol{\pi}')$, link-weighted.
The set of nodes in graph $\mathbf{g}'$ is $\mathscr{R} = \{1, \ldots, N\}$. For any two representatives i and j, let $g'_{ij} \geq 0$ denote the share of bills on which they do not vote in the same way.[11] Thus, g'_{ij} stands for the (inverse of the) strength of their connection, where $g'_{ij} = 0$ indicates that i and j always vote the same way.[12] When they never vote the same way on any bill, they are not directly connected, hence $ij \notin \mathbf{g}'$. The size of every node $i \in \mathscr{R}$ is $\pi'_i = 1$, thus $\boldsymbol{\pi}' = \mathbf{1}$, as each node represents a unique representative. An example of networks as such can be found in Andris et al. (2015).

Network of co-sponsorships, $(\hat{\mathbf{g}}', \hat{\boldsymbol{\pi}}')$, unweighted.
The set of nodes in $\hat{\mathbf{g}}'$ is $\mathscr{R} = \{1, \ldots, N\}$. $\hat{g}'_{ij} = 1$ if i and j co-sponsored al least one bill together, and $ij \notin \hat{\mathbf{g}}'$ otherwise.[13] The size of each node $i \in \mathscr{R}$ is $\hat{\pi}'_i = 1$, thus $\hat{\boldsymbol{\pi}}' = \mathbf{1}$, since each node represents a unique representative. Fowler (2006a) studies this type of networks.

Network of votes, $(\tilde{\mathbf{g}}', \tilde{\boldsymbol{\pi}}')$, node-weighted.
The set of nodes in graph $\tilde{\mathbf{g}}'$ is $\mathscr{V} = \{v_1, \ldots, v_{2^k}\}$. Two nodes (vote combinations) v_i and v_j are connected, *i.e.*, $ij \in \tilde{\mathbf{g}}'$, whenever v_i and v_j differ only in a single coordinate (bill). Each link in $\tilde{g}'$ has a weight 1. $\tilde{\pi}'_i$ denotes the number of individuals with voting profile v_i, and $\tilde{\boldsymbol{\pi}}'$ is the corresponding distribution. Brams et al. (2007) and Moody and Mucha (2013), among others, study this type of networks.

Network of parties, $(\bar{\mathbf{g}}', \bar{\boldsymbol{\pi}}')$, node- and link-weighted.
The set of nodes in $\bar{\mathbf{g}}'$ is $\mathscr{T} = \{t_1, \ldots, t_T\}$. $\bar{g}'_{ij}$ denotes the share of bills on which a majority of representatives in both parties vote the same way.[14] Thus, $ij \notin \bar{\mathbf{g}}'$ indicates that there is no bill that is supported (or opposed) by a majority of representatives in both parties. The size of a node $t_i \in \mathscr{T}$, $\bar{\pi}'_i$, denotes the number of seats of the party i in the parliament.[15] See

[11] Alternatively, one can model that two representatives are connected (with weight 1) if they vote together for more than 50% of the bills and not connected otherwise, in which case we would have an unweighted network.

[12] The fact that $g'_{ij} = 0$ does not indicate that link between i and j does not exist, but that the distance between i and j is 0.

[13] Alternatively, $\hat{g}'_{ij}$ may reflect how many bills i and j co-sponsored together, in which case, we would have a link-weighted network.

[14] $\bar{g}'_{ij}$ captures the ideological distance *i.e.,* the extent the policies of two parties overlap, which can be measured in different ways. Maoz and Somer-Topcu (2010) take, for instance, the similarities in party manifestos.

[15] Alternatively, $\bar{\boldsymbol{\pi}}'$ can be taken as $\mathbf{1}$, disregarding party sizes and focusing on closeness among parties, in

Maoz and Somer-Topcu (2010) for an analysis on party networks.

Each network we describe above focuses on a different aspect of the political activities in the parliament. Accordingly, the corresponding measures of polarization provide different, yet complementary, insights into congressional polarization. For instance, $\mathscr{P}(\mathbf{g}', \boldsymbol{\pi}')$ tells us how polarized the policy positions of representatives based on their vote histories are, regardless of their party affiliations, whereas $\mathscr{P}(\bar{\mathbf{g}}', \bar{\boldsymbol{\pi}}')$ measures the party-level polarization.[16] Also, $\mathscr{P}(\tilde{\mathbf{g}}', \tilde{\boldsymbol{\pi}}')$ is informative about the polarization with respect to policy space, while $\mathscr{P}(\hat{\mathbf{g}}', \hat{\boldsymbol{\pi}}')$ captures the polarization among representatives with respect to policy cooperation.

For illustration, let us more closely compare networks $(\mathbf{g}', \boldsymbol{\pi}')$ and $(\tilde{\mathbf{g}}', \tilde{\boldsymbol{\pi}}')$, which are based on exactly the same data, *i.e.,* votes on bills. Consider the following example with 3 bills and 8 representatives, where "+" in (1) represents approval for a bill and "−" represents disapproval.

	R_{1-3}	R_4	R_{5-6}	R_7	R_8
I	+	−	−	+	+
II	−	+	+	−	+
III	−	−	+	+	+

(1)

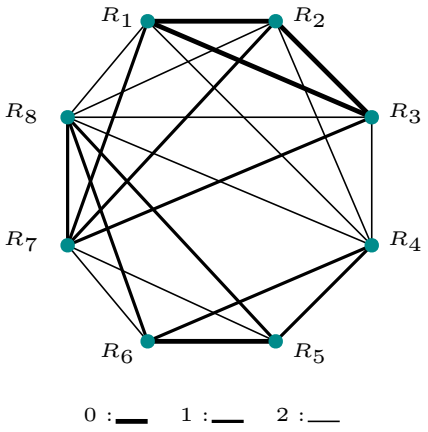
Panel (a) of Figure 1 below shows the corresponding network of representatives, whereas the panel (b) shows the corresponding network of votes.[17] Since the two networks describe two different sets of relations in the legislation, we may expect that the measured level of polarization differs between them. Nevertheless, any polarization measure in our framework is applicable to both cases. To obtain a deeper insight, for instance, one can also compare polarization of networks representing different types of relationships with a suitable normalization *e.g.,* by dividing the polarization index with the maximal value it can attain.

We next turn to the case of mass polarization. Our example is concerned with an electorate choosing among candidates for an office (or individuals expressing preferences over policy alternatives such as *remain*, *soft-Brexit*, and *hard-Brexit*). Let there be a set of alternatives $X = \{x_1, \ldots, x_m\}$ and each individual $i \in \{1, \ldots, n\}$ be endowed with a preference

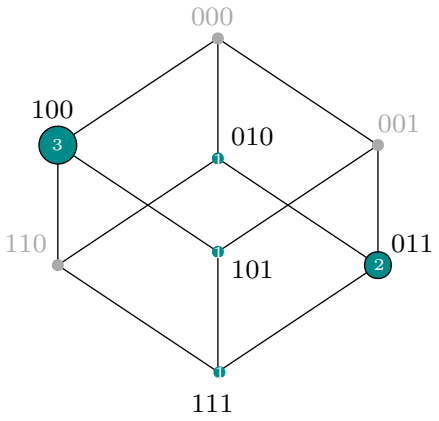
which case we would have a link-weighted network.

[16] We write $\mathscr{P}(\mathbf{g}, \boldsymbol{\pi})$ in place of $\mathscr{P}((\mathbf{g}, \boldsymbol{\pi}))$ with a slight abuse of notation.

[17] Both networks $(\mathbf{g}', \boldsymbol{\pi}')$ and $(\tilde{\mathbf{g}}', \tilde{\boldsymbol{\pi}}')$ have a level of "structural regularity." Graph $\mathbf{g}'$ leads to a complete network structure in the sense that each node is connected to any other node, even though there is a substantial heterogeneity across weights of the links. Graph $\tilde{\mathbf{g}}'$ has a lattice structure. This is by no means necessary for our approach, which is applicable to connected networks with arbitrary structure. For instance, as in Andris et al. (2015), two representatives can be connected if they vote the same way sufficiently many times, then the $\mathbf{g}'$ will not have the complete graph structure. Co-sponsorship networks such as $(\hat{\mathbf{g}}', \hat{\boldsymbol{\pi}}')$ have, in general, quite irregular structures, as in Fowler (2006b).



(a) Network of representatives. Nodes denote representatives and two nodes are not connected if they do not agree on any issue. The thickness of edges indicate weights.



(b) Network of votes. The nodes represent all possible vote combinations *e.g.*, 100 represents the approval of only first bill.

Figure 1: Two possible network representations of the same profile of votes of representatives.

$P_i \subseteq X \times X$ that is a linear order, *i.e.*, a complete, antisymmetric, and transitive binary relation on X. Let $\mathscr{L}$ denote the set of all preferences over X and $\mathscr{L}^n$ be the set of profile of preferences.

Network of preferences, $(\mathbf{g}'', \boldsymbol{\pi}'')$. The set of nodes is $\mathscr{L} = (p_1, \ldots, p_{m!})$. Two nodes p_i and p_j are connected with $g''_{ij} = 1$, whenever p_i can be obtained from p_j by switching only one binary preference, *i.e.,* the Kemeny distance between p_i and p_j is 1 (Kemeny, 1959).[18] We denote with π''_i the number of individuals with preference p_i, and with $\boldsymbol{\pi}''$ the corresponding distribution. See Cervone et al. (2012) for a study on preference networks.[19]

For an illustration, let $\{a, b, c\}$ be the set of alternatives and consider the preference profile with 11 individuals represented by (2).

$$\begin{array}{cccc} 2 & 3 & 2 & 4 \\ \hline a & b & c & c \\ b & a & a & b \\ c & c & b & a \end{array} \tag{2}$$

[18] A network of preferences can be represented as a special network of votes, in which each bill represents a pairwise comparison of alternatives and transitivity is imposed.

[19] Often without explicitly using the language of networks, graph theoretical representations of preferences are studied in the social choice literature widely. There is also a growing interest in measuring polarization in preference profiles, as in Can et al. (2015, 2017). Note that network $(\mathbf{g}'', \boldsymbol{\pi}'')$ could alternatively be defined using a weighted metric as in Can (2014).

This profile of preferences can be represented with a network as depicted as in Figure 2.

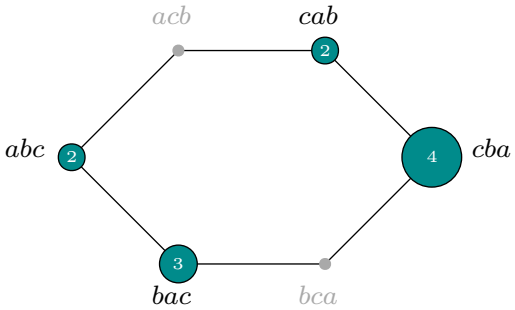


Figure 2: A distribution over a preference network with 3 alternatives and 11 individuals.

2.1.2 Beyond the political domain

While we paid a close attention to examples of networks from the political domain, our approach can naturally be applied in a much wider range of applications, not necessarily confined to those that are commonly studied using networks. For instance, our setting can be adopted to study multidimensional polarization in any distribution with a discrete support. To see how, take the example of polarization in a society with respect to income and education (both measured on some discrete, increasing scale). The set of all pairs of income (ι) and education (ϵ) levels defines the set of nodes in the network. Two nodes $\boldsymbol{x} = (x_\iota, x_\epsilon)$ and $\boldsymbol{y} = (y_\iota, y_\epsilon)$ are connected, with link $\boldsymbol{xy}$ of weight s ($g_{\boldsymbol{xy}} = s$), if, for instance, $|x_\iota - y_\iota| + |x_\epsilon - y_\epsilon| = s$, that is if the Manhattan distance between $\boldsymbol{x}$ and $\boldsymbol{y}$ is equal to s.

Other potential applications include conflicts between groups (Esteban and Ray, 1999, 2011), private provision of public goods (Bramoullé and Kranton, 2007), research output and citation networks (Leskovec et al., 2005), friendship networks (Calvó-Armengol et al., 2009), and trust networks (Richardson et al., 2003).

3 Identification-alienation framework and axiomatization

To recall, our objective in this paper is two-fold. First, we propose network theory as a unifying formalism to study polarization without any constraint on dimensionality. Second, we present a theoretical foundation for a family of polarization measures in this setting. For the latter, we closely follow the axiomatic approach in ER, who envisage polarization as the aggregate antagonism in a population, based on the identification and alienation among individuals.

First, as in ER, we require polarization measures to satisfy the following property that ensures invariance of the measure with respect to the size of the population $\sum_{i\in N(\mathbf{g})}\pi_i$. Thus, in fact, $\boldsymbol{\pi}$ may represent also a probability mass function.

Assumption 1 (Homotheticity)
$\mathscr{P}(\mathbf{g},\boldsymbol{\pi}) \geq \mathscr{P}(\mathbf{g}',\boldsymbol{\pi}') \implies \mathscr{P}(\mathbf{g},\lambda\boldsymbol{\pi}) \geq \mathscr{P}(\mathbf{g}',\lambda\boldsymbol{\pi}')$ *for all* $(\mathbf{g},\boldsymbol{\pi}),(\mathbf{g}',\boldsymbol{\pi}') \in \mathscr{N}$ *and* $\lambda > 0$.

The antagonism between individuals depend on how they identify themselves and how alienated they feel from others. In the network setup we propose, individuals in a population are identified only with their definitive attributes, which are represented as nodes in the network. As emphasized before, these attributes are by no means restricted to singletons or a uni-dimensional space.

The effect of the feeling of *identification* of each individual on her antagonism towards another is measured in relation to the presence of others that share the same attributes, hence are in the same node. This effect is the basis of the intra-group homogeneity, and we denote it with $I(\pi_i)$. Thus, when the nodes represent individuals, each individual feels the same level of identification, whereas when nodes represent groups of individuals, the identification an individual feels is a function of the size its node ($I(\pi_i)$).[20] The only assumption we make on the identification function $I:\mathbb{R}_{\geq 0}\to\mathbb{R}_{\geq 0}$ is that $I(\pi_i)>0$ whenever $\pi_i>0$.

The distance an individual perceives between herself and any other individual is a natural component of the antagonism between individuals as it forms the basis of the inter-group heterogeneity. We measure this *alienation* component as a function of the distance between individuals $a(d(i,j))$. We assume that the alienation function $a:\mathbb{R}_{\geq 0}\to\mathbb{R}_{\geq 0}$ is a continuous and nondecreasing function with $a(0)=0$.

Finally, the *effective antagonism* of group i towards group j is measured by continuous and strictly increasing function $T(I_i,a_{ij})$ of the identification of group i, $I_i=I(\pi_i)$, and the alienation between groups i and j, $a_{ij}=a(d(i,j))$, satisfying $T(I_i,0)=0$. As in ER, we consider polarization measures $\mathscr{P}:\mathscr{N}\to\mathbb{R}_{\geq 0}$ defined as the sum of effective antagonisms:

$$\mathscr{P}(\mathbf{g},\boldsymbol{\pi}) = \sum_{i=1}^{n}\sum_{j=1}^{n}\pi_i\pi_j T\Big(I(\pi_i), a\big(d_{\mathbf{g}}(i,j)\big)\Big). \tag{3}$$

As we shall see, our axioms will pin down specific functional form for $T\Big(I(\pi_i), a\big(d_{\mathbf{g}}(i,j)\big)\Big)$.

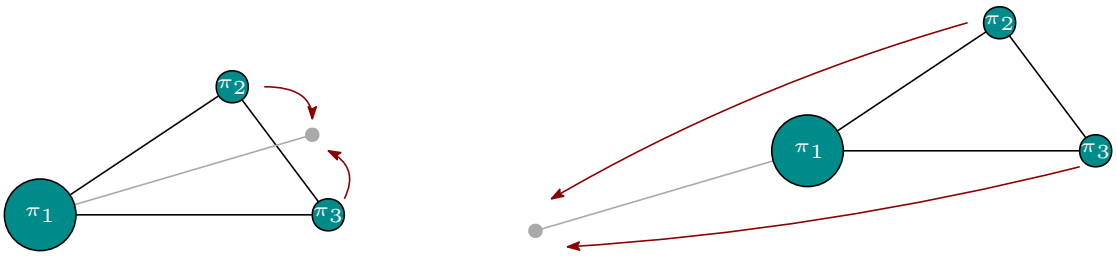
Our goal is to follow the axiomatization in ER as closely as possible, and modify it only when the network setting requires. As it turns out, the first two axioms can be restated only with slight changes in the nomenclature. Axiom 3 needs an important adjustment.

[20] This implies that two groups (nodes) of the same size exhibit the same level of identification. While potentially restrictive, this is standard in the identification-alienation framework (Esteban and Ray, 1994, 2012).

Axiom 1

Data: Network $(\mathbf{g}, \boldsymbol{\pi})$ *with* $n \geq 3$ *nodes such that* $\pi_x > \pi_y = \pi_z > 0$ *and* $\pi_i = 0\ \forall i \in N(\mathbf{g}) \setminus \{x, y, z\}$. *Furthermore,* $d_{\mathbf{g}}(x, y) \leq d_{\mathbf{g}}(x, z)$.
Statement: Fix π_x *and* $d_{\mathbf{g}}(x, y)$. *There exists* $\epsilon > 0$ *and* $\mu = \mu(\pi_x, d_{\mathbf{g}}(x, y)) > 0$ *such that* $d_{\mathbf{g}}(y, z) < \epsilon$ *and* $\pi_y < \mu\pi_x$ *imply that for any* $(\mathbf{g}', \boldsymbol{\pi}') \in \mathscr{N}$ *with* $n \geq 2$ *nodes such that* $\pi'_{x'} = \pi_x$, $\pi'_{w'} = \pi_y + \pi_z$, $d_{\mathbf{g}'}(x'w') = \frac{1}{2}\left(d_{\mathbf{g}}(x, y) + d_{\mathbf{g}}(x, z)\right)$ *and* $\pi'_{i'} = 0$, $i' \in N(\mathbf{g}') \setminus \{x', w'\}$, *we have* $\mathscr{P}(\mathbf{g}', \boldsymbol{\pi}') > \mathscr{P}(\mathbf{g}, \boldsymbol{\pi})$.

The Axiom 1 captures the situations where two small groups join while keeping the (average) distance the same.



(a) The move shown by arrows increase polarization.

(b) Axiom 1 applies when the new node is further away from the two small nodes as well.

Figure 3: Axiom 1.

Suppose in $(\mathbf{g}, \boldsymbol{\pi})$ there is a node with large group and there are two other smaller and equal-sized groups that are close to each other but further away from the larger group. Then network $(\mathbf{g}', \boldsymbol{\pi}')$, in which smaller groups are joined at a node which is located in $\mathbf{g}'$ at a distance equal to their average distance (in $\mathbf{g}$) to the large group, is more polarized. Figure 3 illustrates such moves.[21] Note that the distance of the fourth node to smaller nodes is not restricted in the axiom, allowing for moves such as the one depicted in panel (b) of Figure 3.

Axiom 2

Data: Network $(\mathbf{g}, \boldsymbol{\pi})$ *with* $n \geq 3$ *nodes such that* $\pi_x > \pi_z > 0$, $\pi_y > 0$, *and* $\pi_i = 0, \forall i \in N(\mathbf{g}) \setminus \{x, y, z\}$. *Furthermore,* $d_{\mathbf{g}}(x, z) > d_{\mathbf{g}}(x, y) > d_{\mathbf{g}}(y, z)$.
Statement: There exists $\epsilon > 0$ *such that for any network* $(\mathbf{g}', \boldsymbol{\pi}')$ *with* $(\pi'_{x'}, \pi'_{y'}, \pi'_{z'}) = (\pi_x, \pi_y, \pi_z)$, *and* $\pi'_{i'} = 0$, $i' \in N(\mathbf{g}') \setminus \{x', y', z'\}$ *such that* $d_{\mathbf{g}}(x, z) = d_{\mathbf{g}'}(x', z')$, $0 < d_{\mathbf{g}'}(x', y') - d_{\mathbf{g}}(x, y) = d_{\mathbf{g}}(y, z) - d_{\mathbf{g}'}(y', z') < \epsilon$ *we have* $\mathscr{P}(\mathbf{g}', \boldsymbol{\pi}') > \mathscr{P}(\mathbf{g}, \boldsymbol{\pi})$.

Axiom 2 applies when the group at one extreme is larger than the one at the other extreme and a third group is closer to the smaller of these two. When the group in-between moves

[21] Note that $\mathbf{g}$ and $\mathbf{g}'$ do not have to be different and in our depictions we present axioms on the same graphs.

slightly closer to the smaller group and away from the larger group, polarization increases.[22] Note that the relative size of the group in the middle is not restricted. Figure 4 illustrates such moves.

(a) The move shown by the arrow increases polarization.

(b) Axiom 2 applies in such a move as well, which is not possible on the real line.

Figure 4: Axiom 2.

Note that the described move makes the middle group closer to the smaller group, but its new location does not have to be close to its original position, as seen in panel (b) of Figure 4. This kind of a move is not possible on the real line.

Axiom 3

Data: Network $(\mathbf{g}, \boldsymbol{\pi})$ *with* $n \geq 3$ *nodes such that* $\pi_x > 0$, $\pi_y = \pi_z > 0$ *and* $\pi_i = 0$ $\forall i \in N(\mathbf{g}) \setminus \{x, y, z\}$. *Furthermore,* $d_{\mathbf{g}}(x, y) = d_{\mathbf{g}}(x, z) = d > 0$.
Statement: For any $\Delta \in (0, \frac{\pi_x}{2}]$ *and any network* $(\mathbf{g}', \boldsymbol{\pi}')$ *with* $(\pi'_{x'}, \pi'_{y'}, \pi'_{z'}) = (\pi_x - 2\Delta, \pi_y + \Delta, \pi_z + \Delta)$, *and* $\pi'_{i'} = 0$, $i' \in N(\mathbf{g}') \setminus \{x', y', z'\}$ *such that* $d_{\mathbf{g}'}(x', y') = d_{\mathbf{g}'}(x', z') = d$ *and* $d_{\mathbf{g}}(y, z) = d_{\mathbf{g}'}(x', z')$, *we have* $\mathscr{P}(\mathbf{g}', \boldsymbol{\pi}') > \mathscr{P}(\mathbf{g}, \boldsymbol{\pi})$ *whenever* $d_{\mathbf{g}}(y, z) = cd$, *for any* $c > 1$.

Axiom 3 states that as long as the distance between two lateral groups is greater than the distance between the "middle group" and a lateral group, a network in which individuals from the group in the middle are reallocated to extreme points will exhibit higher polarization. Note that the relative size of the group in node x is not restricted. Furthermore, in a network, $d_{\mathbf{g}}(x, y) = d_{\mathbf{g}}(x, z) = d$ implies only that $d_{\mathbf{g}}(y, z) \leq 2d$, whereas on the real line $y \neq z$ and $|x - y| = |z - x| = d$ imply that $|z - y| = 2d$. We will come back to this crucial point in Section 4.

We are now ready to state our central result, which identifies the measures of polarization in networks that satisfy Axioms 1–3.

Theorem 1

A polarization measure $\mathscr{P}$ *of the family defined in* (3) *satisfies Axioms 1–3 and homotheticity*

[22] Axiom 2 is rather weak as it applies to only those (small) moves such that an increase in distance from one extreme is equal to a decrease in the distance to the other extreme.

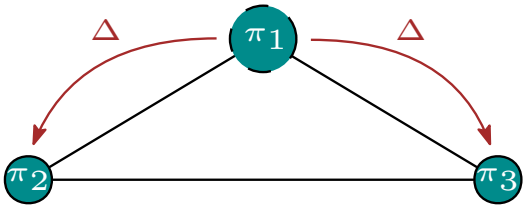


Figure 5: Axiom 3 dictates that the dissolution of the middle group into two extreme nodes increases polarization.

if and only if

$$\mathscr{P}(\mathbf{g}, \boldsymbol{\pi}) = K \sum_{i \in N(\mathbf{g})} \sum_{j \in N(\mathbf{g})} \pi_i^2 \pi_j d_{\mathbf{g}}(i, j), \tag{4}$$

for some constant $K > 0$.

Proof.

Sufficiency. Without loss of generality set $K = 1$. We prove that Axiom 1 and Axiom 2 are satisfied for

$$\mathscr{P}_\alpha(\mathbf{g}, \boldsymbol{\pi}) = K \sum_{i \in N(\mathbf{g})} \sum_{j \in N(\mathbf{g})} \pi_i^{1+\alpha} \pi_j d_{\mathbf{g}}(i, j), \tag{5}$$

whenever $\alpha > 0$. Clearly, (5) becomes (4) when $\alpha = 1$. Establishing this claim is important also for the proof of Theorem 2.

Axiom 1. Let $\pi_x = p$ and $\pi_y = \pi_z = q$. Using $d_{\mathbf{g}'}(x', w') = \frac{d_{\mathbf{g}}(x,y) + d_{\mathbf{g}}(x,z)}{2}$ we get that

$$\mathscr{P}_\alpha(\mathbf{g}, \boldsymbol{\pi}) = p^{1+\alpha} q d_{\mathbf{g}}(x, y) + p^{1+\alpha} q d_{\mathbf{g}}(x, z) + 2q^{1+\alpha} q d_{\mathbf{g}}(y, z) + q^{1+\alpha} p d_{\mathbf{g}}(x, y) + q^{1+\alpha} p d_{\mathbf{g}}(x, z),$$

while

$$\mathscr{P}_\alpha(\mathbf{g}', \boldsymbol{\pi}') = p^{1+\alpha}(2q)\frac{d_{\mathbf{g}}(x, y) + d_{\mathbf{g}}(x, z)}{2} + (2q)^{1+\alpha} p \frac{d_{\mathbf{g}}(x, y) + d_{\mathbf{g}}(x, z)}{2}.$$

After simplification we get:

$$\mathscr{P}_\alpha(\mathbf{g}, \boldsymbol{\pi}) = (d_{\mathbf{g}}(x, y) + d_{\mathbf{g}}(x, z))(p^{1+\alpha} q + q^{1+\alpha} p) + 2q^{2+\alpha} d_{\mathbf{g}}(y, z), \text{ and}$$
$$\mathscr{P}_\alpha(\mathbf{g}', \boldsymbol{\pi}') = (d_{\mathbf{g}}(x, y) + d_{\mathbf{g}}(x, z))(p^{1+\alpha} q + q^{1+\alpha} p) + (2^\alpha - 1)(d_{\mathbf{g}}(x, y) + d_{\mathbf{g}}(x, z)) q^{1+\alpha} p,$$

which implies $\mathscr{P}_\alpha(\mathbf{g}, \boldsymbol{\pi}) > \mathscr{P}_\alpha(\mathbf{g}', \boldsymbol{\pi}')$ whenever $(2^\alpha - 1)(d_{\mathbf{g}}(x, y) + d_{\mathbf{g}}(x, z))p > 2q d_{\mathbf{g}}(y, z)$. When $d(y, z)$ is small enough ($d(y, z) < \epsilon$) this inequality will hold for any $\alpha > 0$ and q small enough relative to p ($q < \mu p$), as required by Axiom 1.

Axiom 2. Let $\pi_x = p$, $\pi_y = q$, and $\pi_z = r$. Subtracting we get:

$$\begin{aligned}\mathscr{P}_\alpha(\mathbf{g},\boldsymbol{\pi}) - \mathscr{P}_\alpha(\mathbf{g}',\boldsymbol{\pi}') = q^{1+\alpha}[p(d_{\mathbf{g}'}(x',y') - d_{\mathbf{g}}(x,y)) + r(d_{\mathbf{g}'}(z',y') - d_{\mathbf{g}}(y,z))] \\ + q[p^{1+\alpha}(d_{\mathbf{g}'}(x',y') - d_{\mathbf{g}}(x,y)) + r^{1+\alpha}(d_{\mathbf{g}'}(z',y') - d_{\mathbf{g}}(y,z))],\end{aligned}$$

which is positive for any $\alpha > 0$ whenever $r < p$, since $d_{\mathbf{g}'}(x',y') - d_{\mathbf{g}}(x,y) = d_{\mathbf{g}'}(z',y') - d_{\mathbf{g}}(y,z)$, and therefore $\mathscr{P}_\alpha$ satisfies Axiom 2.

Axiom 3. We now show that $\mathscr{P}$ satisfies Axiom 3. To this end let $d_{\mathbf{g}}(x,y) = d_{\mathbf{g}}(x,z) = d$, and let $d_{\mathbf{g}}(y,z) = cd$ with $c > 1$. Furthermore, let $\pi_x = p + 2\Delta$ and $\pi_y = \pi_z = q - \Delta$. We can write:

$$\mathscr{P}_\alpha((\mathbf{g},\boldsymbol{\pi});\Delta) = 2cd\big((q-\Delta)^{2+\alpha}\big) + 2d\Big((p+2\Delta)(q-\Delta)\big((p+2\Delta)^\alpha(q-\Delta)^\alpha\big)\Big). \quad (6)$$

To prove that $\mathscr{P}$ satisfies Axiom 3 it is sufficient to show that $\frac{\partial \mathscr{P}_\alpha((\mathbf{g},\boldsymbol{\pi}),\Delta)}{\partial \Delta}\Big|_{\Delta=0,\alpha=1} < 0$ for every $(p,q) \gg 0$, except for at most one ratio p/q. Differentiating (6) at $\Delta = 0$ and dividing by $2d$ (≥ 0) we get:

$$\frac{\partial \mathscr{P}_\alpha}{\partial \Delta}\Big|_{\Delta=0} < 0 \iff -p^\alpha\big(p - 2(1+\alpha)q\big) + q^\alpha\big(-(1+\alpha)p + 2q - (2+\alpha)cq\big) < 0.$$

Dividing by $p^{1+\alpha} > 0$ and using $z = q/p$ we get:

$$\frac{\partial \mathscr{P}_\alpha}{\partial \Delta}\Big|_{\Delta=0} < 0 \iff f(z,\alpha,c) < 0,$$

where $f : \mathbb{R}^2_{\geq 0} \times [1,2] \to \mathbb{R}$ is defined with:

$$f(z,\alpha,c) = (1+\alpha)\left[z - \frac{z^\alpha}{2} + \frac{z^{1+\alpha}}{2}\frac{(2 - c(2+\alpha))}{1+\alpha}\right] - \frac{1}{2}. \quad (7)$$

One can easily verify $f(z,1,c) < 0$ ($f(z,1,c)$ is a quadratic function in z) for any $c \in [1,2]$, except for $z = 1$ when $c = 1$, therefore $\mathscr{P}$ satisfies Axiom 3 as well.

Necessity. The proof is analogous to the proof of Theorem 1 in ER. We describe it briefly, and refer the reader to ER for detailed derivation. Axioms 1–2 imply that function T is linear in its second argument, thus $\theta(\pi,\delta) \equiv T(I(\pi), a(d_{\mathbf{g}}(i,j)))$ can be written as $\theta(\pi,\delta) = \phi(\pi)\delta$. Furthermore, Axiom 1 implies that $\phi(\cdot)$ is an increasing function.[23] Homotheticity implies that $\phi(\pi) = K\pi^\alpha$ for some constants $(K,\pi) \gg 0$. Finally, Axiom 3 implies $f(z,\alpha,1) \leq 0$ for all z. As seen in the first part of the proof this holds for $\alpha = 1$. To see that it does not hold for any other value $\alpha > 0$, first note that $f(1,\alpha,1) = 0$. Furthermore, function $f(z,\alpha,1)$ is increasing in z at $z = 1$ when $\alpha < 1$ and decreasing in z at $z = 1$ when $\alpha > 1$. By continuity, there exist $\epsilon_1 > 0$ and $\epsilon_2 > 0$ such that $f(1+\epsilon_1,\alpha,1) > 0$ when $\alpha < 1$, and $f(1-\epsilon_2,\alpha,1) > 0$

[23] See Kawada et al. (2018) for a solution to a technical problem arising from the original formulation of Axiom 1 in ER.

when $\alpha > 1$. □

A few comments are in order. First, recall that ER characterize measures of polarization on the real line as

$$P^{ER}(\boldsymbol{\pi}) = K \sum_{i=1}^{n} \sum_{j=1}^{n} \pi_i^{1+\alpha} \pi_j |i-j|, \tag{8}$$

with $K > 0$ and $\alpha \in (0, \alpha^*]$, with $\alpha^* \simeq 1.6$. The main difference between (4) and (8) is that the index in (4) implies $\alpha = 1$. The reason for this difference lies in the nature of the distances, discussed in relation with Axiom 3. It requires that a move from a middle mass (π_x) to the lateral points (π_y and π_z) equidistant from the middle increases polarization whenever they are individually further away from each other than they are to the midpoint. Contrary to the real line, in $(\mathbf{g}, \boldsymbol{\pi})$, $d_{\mathbf{g}}(y,z)$ is not determined by $d_{\mathbf{g}}(x,y) = d_{\mathbf{g}}(x,z)$, and in fact it can very well happen that $d_{\mathbf{g}}(y,z) < d_{\mathbf{g}}(x,y)$ even when $d_{\mathbf{g}}(x,y) = d_{\mathbf{g}}(x,z)$. We revisit this important matter in Section 4.3 below. Note that Axioms 1 and 2 also require adaptation for the network setup, but these adaptations are minor and do not have important implications on the form of the characterized family of measures.

Intuitively, a society is polarized if it can be grouped in a small number of homogeneous groups of similar sizes that are very different from each-other and polarization is often conceptualized to capture the level of bipolarity (or bimodality).[24] Thus, it is desirable that a polarization measure is maximized at a bipolar distribution. A bipolar network is one where the population is split equally into two extreme (most distant) nodes. The maximal distance between two nodes in graph $\mathbf{g}$ is called the *diameter* of $\mathbf{g}$ and is denoted by $d(\mathbf{g})$.[25] For any graph $\mathbf{g}$ let $\boldsymbol{\pi}^B(\mathbf{g})$ denote the distribution in which the population is split equally across two nodes at distance $d(\mathbf{g})$. Our next result shows that $(\mathbf{g}, \pi^B(\mathbf{g}))$ is more polarized than any other network $(\mathbf{g}, \boldsymbol{\pi})$ under any measure within our characterization.

Proposition 1
$\mathscr{P}(\mathbf{g}, \boldsymbol{\pi}^B(\mathbf{g})) > \mathscr{P}(\mathbf{g}, \boldsymbol{\pi})$ for any $(\mathbf{g}, \boldsymbol{\pi})$ with $\boldsymbol{\pi} \neq \boldsymbol{\pi}^B(\mathbf{g})$ and any measure $\mathscr{P}$ defined in (4).

Proof.
We first prove that for any network $(\mathbf{g}, \boldsymbol{\pi})$ such that $\boldsymbol{\pi}$ has at lest 4 nonzero mass points, there exists a 3 node network $(\mathbf{g}^*, \boldsymbol{\pi}^*)$ with $g^*_{ij} = d(\mathbf{g})$ for $i, j \in N(\mathbf{g}^*)$, and $\sum_{i=1}^{3} \pi_i^* = \sum_{i \in N\mathbf{g}} \pi_i$ such that $\mathscr{P}(\mathbf{g}, \boldsymbol{\pi}) < \mathscr{P}(\mathbf{g}^*, \boldsymbol{\pi}^*)$.

The proof is constructive. Assume, without loss of generality, that in $(\mathbf{g}, \boldsymbol{\pi})$, we have $\pi_1 \geq \pi_2 \geq \cdots \geq \pi_n$ with $\pi_k > 0$ and $\pi_{k+1} = 0$ for some $k \geq 4$. Fixing $K = 1$ in (4) (without

[24] See Foster and Wolfson (2010) for a discussion on bipolarity of income distributions and DiMaggio et al. (1996) for a more general discussion on bimodality, among others.

[25] More formally, $d(\mathbf{g}) = \max_{i,j \in N(\mathbf{g})} d_{\mathbf{g}}(i,j)$. See Vega-Redondo (2007) or Jackson (2008).

loss of generality) we get:

$$\begin{aligned}\mathscr{P}(\mathbf{g},\boldsymbol{\pi}) \leq & d(\mathbf{g})\sum_{i=1}^{k}\sum_{j=1}^{k}\pi_i^2\pi_j d_{\mathbf{g}}(i,j)\\ = & d(\mathbf{g})\left[\sum_{i=1}^{k-2}\sum_{\substack{j=1\\ j\neq i}}^{k-2}\pi_i^2\pi_j+\pi_{k-1}^2\sum_{\substack{j=1\\ j\neq k-1}}^{k}\pi_j+\pi_k^2\sum_{\substack{j=1\\ j\neq k}}^{k}\pi_j+\pi_{k-1}\sum_{j=1}^{k-2}\pi_j^2+\pi_k\sum_{j=1}^{k-2}\pi_j^2\right].\end{aligned} \tag{9}$$

Denote the right hand side expression in (9) with $\mathscr{P}(\mathbf{g}',\boldsymbol{\pi}')$, where $g'_{ij}=d(\mathbf{g})$ for all $i,j\in N(\mathbf{g})$ and $\pi'_i=\pi_i$ for all $i\in N(\mathbf{g})$. Consider now a change in $(\mathbf{g}',\boldsymbol{\pi}')$ such that masses in nodes k and $k-1$ are merged at one of these nodes to obtain $(\mathbf{g}'',\boldsymbol{\pi}'')$. Simple algebra gives:

$$\mathscr{P}(\mathbf{g}'',\boldsymbol{\pi}'')=d(\mathbf{g})\left[\sum_{i=1}^{k-2}\sum_{\substack{j=1\\ j\neq i}}^{k-2}\pi_i^2\pi_j+(\pi_{k-1}+\pi_k)^2\sum_{j=1}^{k-2}\pi_j+(\pi_{k-1}+\pi_k)\sum_{j=1}^{k-2}\pi_j^2\right].$$

Subtracting $\mathscr{P}(\mathbf{g}',\boldsymbol{\pi}')$ we get:

$$\mathscr{P}(\mathbf{g}'',\boldsymbol{\pi}'')-\mathscr{P}(\mathbf{g}',\boldsymbol{\pi}')=d(\mathbf{g})\pi_{k-1}\pi_k\left[2\sum_{j=1}^{k-2}\pi_j-(\pi_{k-1}+\pi_k)\right]>0, \tag{10}$$

where the inequality follows from the choice of k and $k-1$ and the fact that $k\geq 4$. Thus, for any network $(\mathbf{g},\boldsymbol{\pi})$ with $|N(\mathbf{g})|\geq 4$ have $\mathscr{P}(\mathbf{g},\boldsymbol{\pi})<\mathscr{P}(\mathbf{g}'',\boldsymbol{\pi}'')$.

If $k=4$, $(\mathbf{g}^*,\boldsymbol{\pi}^*)=(\mathbf{g}'',\boldsymbol{\pi}'')$. If $k>4$, the above described procedure of joining the masses in nodes $k-1$ and $k-2$ can be iteratively applied.

To conclude the proof of the proposition, consider 3 different cases for $(\mathbf{g},\boldsymbol{\pi})$:

(i) $|\{i\in N(\mathbf{g}):\pi_i>0\}|=2$. Clearly $\mathscr{P}(\mathbf{g},\boldsymbol{\pi}^B(\mathbf{g}))>d(\mathbf{g})2\left(\frac{\pi_i+\pi_j}{2}\right)^3>d_{\mathbf{g}}(i,j)\big(\pi_i\pi_j(\pi_i+\pi_j)\big)=\mathscr{P}(\mathbf{g},\boldsymbol{\pi})$ for any $(\mathbf{g},\boldsymbol{\pi})\neq(\mathbf{g},\boldsymbol{\pi}^B)$.

(ii) $|\{i\in N(\mathbf{g}):\pi_i>0\}|=3$. Consider $(\mathbf{g}',\boldsymbol{\pi}')$ obtained from $(\mathbf{g},\boldsymbol{\pi})$ such that $\boldsymbol{\pi}=\boldsymbol{\pi}'$, $N(\mathbf{g})=N(\mathbf{g}')$, $E(\mathbf{g})=E(\mathbf{g}')$ and $g'_{ij}=d(\mathbf{g})$ for any $i,j\in N(\mathbf{g}')$. Clearly $\mathscr{P}(\mathbf{g},\boldsymbol{\pi})\leq\mathscr{P}(\mathbf{g}',\boldsymbol{\pi}')$. Construct $\boldsymbol{\pi}''$ from $\boldsymbol{\pi}'$ by reallocating mass from any point to the other two points as in Axiom 3. We have:

$$\mathscr{P}(\mathbf{g},\boldsymbol{\pi}^B(\mathbf{g}))=\mathscr{P}(\mathbf{g}',\boldsymbol{\pi}^B(\mathbf{g}))\geq\mathscr{P}(\mathbf{g}',\boldsymbol{\pi}'')>\mathscr{P}(\mathbf{g}',\boldsymbol{\pi}'),$$

where the first inequality is implied by (i), while the second is a consequence of Axiom 3.

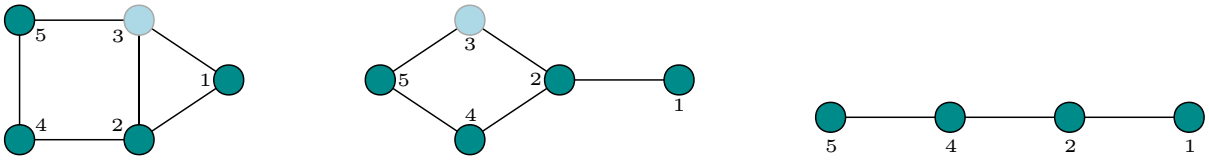
(iii) $|\{i\in N(\mathbf{g}):\pi_i>0\}|\geq 4$. The claim follows from the first part of the proof and (ii).□

4 Discussion

In this section, we first discuss some important properties of the measures we characterize in relation to the structure of networks. Then we show how our work is related to previous papers in the literature. We conclude this section with a discussion on how the weakening of the Axiom 3 can relate our characterization to the one in ER, by exactly describing the relationship between the importance of identification (α) and the network structure.

4.1 Network structure and polarization

We first want to emphasize that the structure of a graph $\mathbf{g}$ determines the distance between any two nodes in $N(\mathbf{g})$. A change in the structure of a graph $\mathbf{g}$, *e.g.,* deleting a link, may affect the measured levels of polarization, even if $\boldsymbol{\pi}$ stays the same. Although empty (zero-weight) nodes do not directly contribute to the level of polarization, they may be important "indirectly" if, for instance, they are located on the shortest path between some non-empty nodes. Figure 6 illustrates this point.



(a) $(\mathbf{g}, \boldsymbol{\pi})$. (b) $(\mathbf{g}', \boldsymbol{\pi})$. (c) $(\mathbf{g}'', \boldsymbol{\pi}'')$.

Figure 6: Three networks where each link has weight 1 and each node except the node 3 has weight 1 ($\pi_3 = 0$). $(\mathbf{g}', \boldsymbol{\pi})$ is obtained from $(\mathbf{g}, \boldsymbol{\pi})$ by deleting the link g_{13}. $(\mathbf{g}'', \boldsymbol{\pi}'')$ is obtained from $(\mathbf{g}, \boldsymbol{\pi})$ by deleting the node 3. Note that $\pi_i = \pi_i''$ for all $i \in N(\mathbf{g}'')$. We have $\mathscr{P}(\mathbf{g}, \boldsymbol{\pi}) < \mathscr{P}(\mathbf{g}', \boldsymbol{\pi}) = \mathscr{P}(\mathbf{g}'', \boldsymbol{\pi}'')$.

Next, we want to note that given Proposition 1, we have that $d(\mathbf{g}) > d(\mathbf{g}')$ implies $\mathscr{P}(\mathbf{g}, \boldsymbol{\pi}^B(\mathbf{g})) > \mathscr{P}(\mathbf{g}', \boldsymbol{\pi}^B(\mathbf{g}'))$. That is, comparing two bipolar networks, the larger the diameter, the higher the polarization.

Finally, in the special case when $\boldsymbol{\pi} = \mathbf{1}$, $\mathscr{P}(\mathbf{g}, \boldsymbol{\pi})$ is proportional to the *average shortest path* in the graph $\mathbf{g}$.[26] Thus, the closer the individuals are, on average, the less polarized the network is.

4.2 Relation to previous results

We argue that the settings considered in ER and MRQ are special cases of our setting, and hence our results can be seen as generalizations of theirs. To start with, recall that ER

[26]The average shortest path in a network is closely related to the "closeness" measure (Vega-Redondo, 2007, Jackson, 2008).

consider distributions on the real line with a finite support (p. 830). It is straightforward to note that any distribution as such can be described as a network. To see this, let $\boldsymbol{\pi}$ be a distribution with a set of N mass points. Consider graph $\mathbf{g}$ with N nodes such that $g_{ij} = |i-j|$ for any two adjacent mass points i and j on the real line, and $ij \notin \mathbf{g}$ otherwise.[27] Indeed, we can represent any distribution on an $m-$dimensional space with finite number of mass point as a network by simply setting $g_{\boldsymbol{ij}} = \|\boldsymbol{i} - \boldsymbol{j}\|$, where $\|\cdot\|$ can be any norm.

In the setting considered in MRQ the distance between any two different groups equals to 1. It is immediate to note that this setting can be described by the network $(\mathbf{g}, \boldsymbol{\pi})$ where $\mathbf{g}$ is the complete graph such that $g_{ij} = 1$ for any pair of nodes $i, j \in N(\mathbf{g})$. MRQ propose a different set of axioms.[28]

4.3 Axiom 3 and its role in the network setting

Axiom 3 requires that the described change in $(\mathbf{g}, \boldsymbol{\pi})$ leads to an increase in polarization only when the distance between lateral nodes is at least as large as the distance between the center node and lateral nodes. We now discuss less demanding versions of Axiom 3, labeled systematically as Axiom 3(c), in which we require that the scenario in Axiom 3 leads to an increase in polarization only if the lateral nodes are "far enough" (quantified by the scalar c) from each other. This is of interest also because some settings imply a specific network structure in which there is a clear lower bound for the distance between two lateral nodes contemplated in Axiom 3. For instance, as we saw before, any discrete distribution on a real line can be represented with a line network. On any line network, the distance between lateral nodes is the double of the distance between the middle node and a lateral node, as it is on the real line.

Axiom 3(c)
Data: Network $(\mathbf{g}, \boldsymbol{\pi})$ *with* $n \geq 3$ *nodes,* $\pi_x > \pi_y = \pi_z > 0$ *and* $\pi_i = 0$ *for all* $i \in N(\mathbf{g}) \setminus \{x, y, z\}$. *Furthermore,* $d_{\mathbf{g}}(x, y) = d_{\mathbf{g}}(x, z) = d > 0$.
Statement: Fix $c \in (1, 2]$. *For any* $\epsilon \in (0, \pi_x]$ *and any network* $(\mathbf{g}', \boldsymbol{\pi}')$ *with* $(\pi'_{x'}, \pi'_{y'}, \pi'_{z'}) = (\pi_x - \epsilon, \pi_y + \frac{\epsilon}{2}, \pi_z + \frac{\epsilon}{2})$, *and* $\pi'_{i'} = 0$, $i' \in N(\mathbf{g}') \setminus \{x', y', z'\}$ *such that* $d_{\mathbf{g}'}(x', y') = d_{\mathbf{g}'}(x', z') = d$ *and* $d_{\mathbf{g}}(y, z) = d_{\mathbf{g}'}(y', z')$, *we have* $\mathscr{P}(\mathbf{g}', \boldsymbol{\pi}') > \mathscr{P}(\mathbf{g}, \boldsymbol{\pi})$ *whenever* $d_{\mathbf{g}}(y, z) \geq dc$.

[27] This is not the unique way to represent a discrete distribution with n mass points as a network. However, any consistent representation that relies on the same metric will lead to a network with the same polarization.

[28] The logical dependence between our axioms and the ones in MRQ is an interesting question that is left for future research.

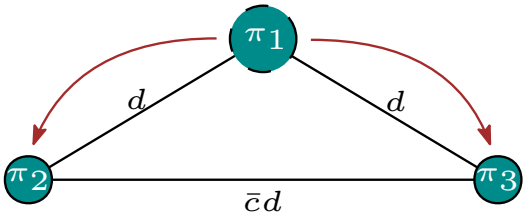


Figure 7: Axiom 3(c) requires that the move shown by the arrows should increase polarization if $\bar{c} \geq c$.

When $c = 1$, we have the same statement as in Axiom 3, while for $c = 2$ we have essentially the Axiom 3 in ER. The particular value of c has important implications on the resulting measure of polarization, as stated in Theorem 2.

Theorem 2

Fix $c \in (1, 2]$. There exists an interval $[\underline{\alpha}(c), \bar{\alpha}(c)] \subseteq (0, \alpha^*]$ with $\alpha^* \simeq 1.6$ such that the polarization measure $\mathscr{P}$ of the family defined in (3) *satisfies Axioms 1, 2, and 3(c) and homotheticity if and only if*

$$\mathscr{P}_\alpha(\mathbf{g}, \boldsymbol{\pi}) = K \sum_{i \in N(\mathbf{g})} \sum_{j \in N(\mathbf{g})} \pi_i^{1+\alpha} \pi_j d_{\mathbf{g}}(i, j) \tag{11}$$

for some constant $K > 0$ whenever $\alpha \in [\underline{\alpha}(c), \bar{\alpha}(c)]$. Furthermore, $c_2 > c_1 \implies [\underline{\alpha}(c_1), \bar{\alpha}(c_1)] \subset [\underline{\alpha}(c_2), \bar{\alpha}(c_2)]$.

Proof.

See the proof of Theorem 1 for the proofs of claims regarding Axiom 1 and Axiom 2 (the *Sufficency* and the *Necessity* part). Similarly, Axiom 3(c) holds iff α is such that $f(z, \alpha, c) < 0$ except for at most one point z, where f is defined in (7). To conclude the proof, two observations about $v(\alpha, c) = \max_{z \geq 0} f(z, \alpha, c)$ are important. First, v is increasing in $\alpha \in (1, 2]$ for any fixed $c \in (1, 2]$ and changes the sign on the considered interval. Thus, there exists $\bar{\alpha}(c)$ such that $v(\alpha, c) \leq 0$ for $\alpha \in (1, \bar{\alpha}(c)]$. Since $v(\alpha, c)$ is decreasing in c, $\bar{\alpha}(c)$ is increasing in c. Second, for $\alpha < 1$ and fixed c, v decreases in α whenever $v(\alpha, c) \geq 0$ eventually becoming negative as $v(1, c) < 0$ for $c > 1$. This implies the existence of $\underline{\alpha}(c) \in [0, 1]$. Since v decreases in c we have that $\underline{\alpha}(c)$ increases in c. From these two observations[29] we conclude $2 \geq c_2 > c_1 > 1 \implies [\underline{\alpha}(c_1), \bar{\alpha}(c_1)] \subset [\underline{\alpha}(c_2), \bar{\alpha}(c_2)]$. □

Theorem 2 shows that as we make Axiom 3 less demanding, the range of values of parameter α for which our axioms is satisfied expands monotonically. In particular, if we restrict ourselves to line networks, then the network structure implies that any move described in

[29] See Lemma 1 and 2 in Appendix A for the formal statements and proofs of these two observations.

Axiom 3 is consistent with Axiom 3(c) for $c = 2$, and Axioms 1, 2 and 3(c) can be seen as restatements of the Axioms 1–3 in ER.

Finally, it should be noted that the claim in Proposition 1 holds only for measures characterized in Theorem 1, and not for any other measure as in (5) with $\alpha \neq 1$. To see this, take any graph $\mathbf{g}$ such that $N(\mathbf{g}) = \{x, y, z\}$ with $0 < g_{xy} = g_{xz} \leq g_{yz}$. Then for any $\alpha \in \mathbb{R}_{\geq 0} \setminus \{1\}$, there exists a distribution $\boldsymbol{\pi} \neq \boldsymbol{\pi}^B(\mathbf{g})$ and $\epsilon > 0$ such that $\mathscr{P}(\mathbf{g}, \boldsymbol{\pi}) > \mathscr{P}_\alpha(\mathbf{g}, \boldsymbol{\pi}^B(\mathbf{g}))$ whenever $g_{yz} = g_{xz} + \epsilon$. This is a direct consequence of the fact that for $\alpha \neq 1$, $\mathscr{P}_\alpha$ does not satisfy Axiom 3 when c is arbitrary close to 1.

5 Conclusion

We have introduced a model of polarization in networks. This model can be used to study the levels and trends of polarization in a wide range of applications. In Section 2, we discussed several examples from political processes in parliaments and public preferences. The potential of our proposal is by no means restricted to these examples as pointed to before. To name a few areas beyond the domain of polity, for which a recent survey is provided by Battaglini and Patacchini (2019), Bail (2016) constructs weighted networks between advocacy organizations based on the frequency of words in the shared vocabulary of their posts. Stewart et al. (2018) construct retweet networks to study the impact of suspicious troll activity on the levels of polarization on Twitter (see, also, Conover et al., 2011). Farrell (2016) constructs a network of organizations based on the activities of affiliates to study polarization on climate change issues among organizations. O'Connor and Weatherall (2018) propose the network formalism to study polarization in scientific communities around beliefs based on scientific knowledge. DiFonzo et al. (2013) employ a network-based approach on capturing polarization of rumor beliefs in the context of social impact theory.

Reconstructing the axiomatic analysis of ER, we characterized a family of measures within our model. Importing the axiomatic approach needs a careful attention due to the distinct nature of the geodesic distance on networks compared to the Euclidean distance on the real line. Our characterization result shows that the class of measures characterized by ER carries almost intact to the networks. The only bite is in the value of the parameter for the effect of identification on effective antagonism. We find that $\alpha = 1$ is a necessary and sufficient condition for the measures of polarization in the form of aggregate antagonisms to satisfy the aforementioned axioms, together with hometheticity. We demonstrate that polarization is maximized when the population is allocated on the two most distant nodes in the network. Finally, we discuss how restricting to specific class of network structures may expand the class of polarization measures.

Our model can be further developed along different dimensions. One promising avenue for future research pertains to extending the measures so as to capture the intra-group heterogeneity, which could also be described as a network. In that case, the identification function should additionally depend on the within-group structure. Another direction for future research concerns the existence of interesting characterizations outside the identification-alienation framework but with the same axioms, as these two are independent.

References

Abramowitz, A. I. and Saunders, K. L. (2008). Is polarization a myth? *The Journal of Politics*, 70(2):542–555.

Aghion, P., Alesina, A., and Trebbi, F. (2004). Endogenous political institutions. *The Quarterly Journal of Economics*, 119(2):565–611.

Andris, C., Lee, D., Hamilton, M. J., Martino, M., Gunning, C. E., and Selden, J. A. (2015). The rise of partisanship and super-cooperators in the us house of representatives. *PloS one*, 10(4):e0123507.

Bail, C. A. (2016). Combining natural language processing and network analysis to examine how advocacy organizations stimulate conversation on social media. *Proceedings of the National Academy of Sciences*, 113(42):11823–11828.

Battaglini, M. and Patacchini, E. (2019). Social networks in policy making. *Annual Review of Economics*.

Boxell, L., Gentzkow, M., and Shapiro, J. M. (2017). Greater internet use is not associated with faster growth in political polarization among us demographic groups. *Proceedings of the National Academy of Sciences*, 114(40):10612–10617.

Bramoullé, Y. and Kranton, R. (2007). Public goods in networks. *Journal of Economic Theory*, 135(1):478–494.

Brams, S. J., Kilgour, D. M., and Sanver, M. R. (2007). A minimax procedure for electing committees. *Public Choice*, 132(3-4):401–420.

Calvó-Armengol, A., Patacchini, E., and Zenou, Y. (2009). Peer effects and social networks in education. *The Review of Economic Studies*, 76(4):1239–1267.

Can, B. (2014). Weighted distances between preferences. *Journal of Mathematical Economics*, 51:109–115.

Can, B., Ozkes, A., and Storcken, T. (2015). Measuring polarization in preferences. *Mathematical Social Sciences*, 78:76–79.

Can, B., Ozkes, A., and Storcken, T. (2017). Generalized measures of polarization in preferences. Technical report, Aix-Marseille School of Economics, France.

Cervone, D. P., Dai, R., Gnoutcheff, D., Lanterman, G., Mackenzie, A., Morse, A., Srivastava, N., and Zwicker, W. S. (2012). Voting with rubber bands, weights, and strings. *Mathematical Social Sciences*, 64(1):11–27.

Conover, M. D., Ratkiewicz, J., Francisco, M., Gonçalves, B., Menczer, F., and Flammini, A. (2011). Political polarization on twitter. In *Fifth international AAAI conference on weblogs and social media*.

Desmet, K., Ortuño-Ortín, I., and Wacziarg, R. (2012). The political economy of linguistic cleavages. *Journal of development Economics*, 97(2):322–338.

Desmet, K., Ortuño-Ortín, I., and Wacziarg, R. (2017). Culture, ethnicity, and diversity. *American Economic Review*, 107(9):2479–2513.

DiFonzo, N., Bourgeois, M. J., Suls, J., Homan, C., Stupak, N., Brooks, B. P., Ross, D. S., and Bordia, P. (2013). Rumor clustering, consensus, and polarization: Dynamic social impact and self-organization of hearsay. *Journal of Experimental Social Psychology*, 49(3):378–399.

DiMaggio, P., Evans, J., and Bryson, B. (1996). Have american's social attitudes become more polarized? *American journal of Sociology*, 102(3):690–755.

Duclos, J.-Y., Esteban, J., and Ray, D. (2004). Polarization: concepts, measurement, estimation. *Econometrica*, 72(6):1737–1772.

Esteban, J., Gradín, C., and Ray, D. (2007). An extension of a measure of polarization, with an application to the income distribution of five oecd countries. *The Journal of Economic Inequality*, 5(1):1–19.

Esteban, J. and Ray, D. (1994). On the measurement of polarization. *Econometrica: Journal of the Econometric Society*, pages 819–851.

Esteban, J. and Ray, D. (1999). Conflict and distribution. *Journal of Economic Theory*, 87(2):379–415.

Esteban, J. and Ray, D. (2011). Linking conflict to inequality and polarization. *American Economic Review*, 101(4):1345–74.

Esteban, J. and Ray, D. (2012). Comparing polarization measures. *Oxford Handbook of Economics of Peace and Conflict*, pages 127–151.

Farrell, J. (2016). Corporate funding and ideological polarization about climate change. *Proceedings of the National Academy of Sciences*, 113(1):92–97.

Fiorina, M. P. and Abrams, S. J. (2008). Political polarization in the american public. *Annu. Rev. Polit. Sci.*, 11:563–588.

Fiorina, M. P., Abrams, S. J., and Pope, J. C. (2005). Culture war. *The myth of a polarized America*, 3.

Foster, J. E. and Wolfson, M. C. (2010). Polarization and the decline of the middle class: Canada and the us. *The Journal of Economic Inequality*, 8(2):247–273.

Fowler, J. H. (2006a). Connecting the congress: A study of cosponsorship networks. *Political Analysis*, 14(4):456–487.

Fowler, J. H. (2006b). Legislative cosponsorship networks in the us house and senate. *Social Networks*, 28(4):454–465.

Garcia, D., Abisheva, A., Schweighofer, S., Serdült, U., and Schweitzer, F. (2015). Ideological and temporal components of network polarization in online political participatory media. *Policy & internet*, 7(1):46–79.

Hill, S. J. and Tausanovitch, C. (2015). A disconnect in representation? comparison of trends in congressional and public polarization. *The Journal of Politics*, 77(4):1058–1075.

Jackson, M. O. (2008). *Social and economic networks*. Princeton university press Princeton.

Kam, C., Indridason, I., Bianco, W., et al. (2017). Polarization in multiparty systems.

Kawada, Y., Nakamura, Y., and Sunada, K. (2018). A characterization of the esteban–ray polarization measures. *Economics Letters*, 169:35–37.

Kearney, M. W. (2019). Analyzing change in network polarization. *new media & society*, 21(6):1380–1402.

Kemeny, J. G. (1959). Mathematics without numbers. *Daedalus*, 88(4):577–591.

Lee, J. K., Choi, J., Kim, C., and Kim, Y. (2014). Social media, network heterogeneity, and opinion polarization. *Journal of communication*, 64(4):702–722.

Lelkes, Y. (2016). Mass polarization: Manifestations and measurements. *Public Opinion Quarterly*, 80(S1):392–410.

Leskovec, J., Kleinberg, J., and Faloutsos, C. (2005). Graphs over time: densification laws, shrinking diameters and possible explanations. In *Proceedings of the eleventh ACM SIGKDD international conference on Knowledge discovery in data mining*, pages 177–187. ACM.

Lindqvist, E. and Östling, R. (2010). Political polarization and the size of government. *American Political Science Review*, 104(3):543–565.

Maoz, Z. (2006a). Network polarization. *mimeo*.

Maoz, Z. (2006b). Network polarization, network interdependence, and international conflict, 1816–2002. *Journal of Peace Research*, 43(4):391–411.

Maoz, Z. and Somer-Topcu, Z. (2010). Political polarization and cabinet stability in multiparty systems: A social networks analysis of european parliaments, 1945–98. *British Journal of Political Science*, 40(4):805–833.

Martin, G. J. and Yurukoglu, A. (2017). Bias in cable news: Persuasion and polarization. *American Economic Review*, 107(9):2565–99.

McCright, A. M. and Dunlap, R. E. (2011). The politicization of climate change and polarization in the american public's views of global warming, 2001–2010. *The Sociological Quarterly*, 52(2):155–194.

Montalvo, J. G. and Reynal-Querol, M. (2008). Discrete polarisation with an application to the determinants of genocides. *The Economic Journal*, 118(533):1835–1865.

Moody, J. and Mucha, P. J. (2013). Portrait of political party polarization. *Network Science*, 1(1):119–121.

Newman, M. E. (2003). The structure and function of complex networks. *SIAM review*, 45(2):167–256.

O'Connor, C. and Weatherall, J. O. (2018). Scientific polarization. *European Journal for Philosophy of Science*, 8(3):855–875.

Østby, G. (2008). Polarization, horizontal inequalities and violent civil conflict. *Journal of Peace Research*, 45(2):143–162.

Ozdemir, U. and Ozkes, A. (2014). Measuring public preferential polarization. *Ecole Polytechnique Departement d'Economie, Cahier de recherche*, (2014-06).

Permanyer, I. and D'Ambrosio, C. (2015). Measuring social polarization with ordinal and categorical data. *Journal of Public Economic Theory*, 17(3):311–327.

Richardson, M., Agrawal, R., and Domingos, P. (2003). Trust management for the semantic web. In *International semantic Web conference*, pages 351–368. Springer.

Stewart, L. G., Arif, A., and Starbird, K. (2018). Examining trolls and polarization with a retweet network. In *Proc. ACM WSDM, workshop on misinformation and misbehavior mining on the web*.

Vega-Redondo, F. (2007). *Complex social networks*. Number 44. Cambridge University Press.

Wimmer, A., Cederman, L.-E., and Min, B. (2009). Ethnic politics and armed conflict: A configurational analysis of a new global data set. *American Sociological Review*, 74(2):316–337.

Zhang, X. and Kanbur, R. (2001). What difference do polarisation measures make? an application to china. *Journal of development studies*, 37(3):85–98.

A Appendix: Proofs

In what follows, we denote the maximal value of parameter α in ER with α^* (so that $\alpha^* \simeq 1.6$).

Lemma 1
Let $1 < \alpha \leq \alpha^*$ *and* $c \in (1, 2]$. *There exists* $\bar{\alpha} = \bar{\alpha}(c) \in (1, \alpha^*]$ *such that* $\max_{z \geq 0} f(z, \alpha, c) \leq 0$ *whenever* $\alpha \leq \bar{\alpha}$. *Furthermore,* $\bar{\alpha}$ *is increasing in* c.

Proof of Lemma 1. For $\alpha \geq 1$, f is concave in z. Thus, the maximum of f is given by the first order condition:

$$\frac{1}{2}(1+\alpha)\left(2 - \alpha z^{\alpha-1} + (2-(2+\alpha)c)z^{\alpha}\right) = 0. \tag{12}$$

Taking derivative of the value function $v(\alpha, c) = \max_{z \geq 0} f(z, \alpha, c)$ with respect to c, and applying the envelope theorem, we get:

$$\frac{\partial v}{\partial c} = \frac{\partial f}{\partial z}\frac{\partial z}{\partial c} + \frac{\partial f}{\partial c} = \frac{\partial f}{\partial c} = -\frac{1}{2}(\alpha+2)z^{\alpha+1} < 0,$$

so the value function is (strictly) decreasing in c. This means that $v(\alpha, c) > v(\alpha, 2)$, for any $c \in [1, 2)$.

We know from the observations on the function f in (7) discussed in the proof of Theorem 1 that, for any $c > 1$, $v(1, c) < 0$. Since $f(z, 2, 2) > 0$, as pointed out in ER (p. 833) and $v(\alpha, c)$ is decreasing in c, it must be that $v(2, c) > 0$ for any $c \in (1, 2]$. Therefore $v(1, c) < 0$ and $v(2, c) > 0$ for any $c \in (1, 2]$. To show that there exist $\bar{\alpha}(c)$ from the claim of the Lemma, we show that $v(\alpha, c)$ is increasing in α. Indeed:

$$\frac{\partial v}{\partial \alpha} = \frac{\partial f}{\partial z}\frac{\partial z}{\partial \alpha} + \frac{\partial f}{\partial \alpha} = \frac{\partial f}{\partial \alpha} = \frac{1}{2}(2z - z^{\alpha}(1 + cz + (1 + \alpha + (-2 + (2+\alpha)c)z)\ln z)).$$

To see that the above derivative is positive, first note that the first order condition (12) implies that at the maximum of f:

$$z^{\alpha-1} = \frac{2}{\alpha - z\big(2 - (2+\alpha)c\big)}. \tag{13}$$

Equation (13) together with the fact that $\alpha - 1 > 0$ and $c > 1$ implies that $z < 1$. Indeed, if $z > 1$ then the RHS of (13) would be greater than 1, while the LHS of (13) would be smaller than 1, since the denominator $\alpha - z\big(2 - (2+\alpha)c\big)$ would be greater than 2 since $(2+\alpha)c > 3$

and $z > 1$. Plugging (13) into the expression for $\frac{\partial v}{\partial \alpha}$ from above. we get:

$$\begin{aligned}\frac{\partial v}{\partial \alpha} &= \frac{1}{2}(2z - z^{\alpha}(1 + cz + (1 + \alpha + (-2 + (2 + \alpha)c)z)\ln z)) \\ &= \frac{1}{2}(2z - \frac{2}{\alpha - z\big(2 - (2 + \alpha)c\big)}(1 + cz + (1 + \alpha + (-2 + (2 + \alpha)c)z)\ln z)) \\ &= -z\frac{(1 - \alpha) + (2 - 3c - 2\alpha c)z + [1 + 2\alpha + (-2 + 2c + \alpha c)z]\log z}{\alpha - z\big(2 - (2 + \alpha)c\big)},\end{aligned}$$

which is clearly positive for $z < 1$, $\alpha \geq 1$ and $c > 1$.

Therefore, the intermediate value theorem implies that, for any $c \in [1, 2]$ there exist $\bar{\alpha}(c) \in [1, 2]$ such that $\max_{z \geq 0} f(z, \alpha, c) \leq 0$ whenever $\alpha \leq \bar{\alpha}(c)$ (with equality only when $\alpha = \bar{\alpha}$). Finally, $\frac{\partial v}{\partial c} < 0$ and $\frac{\partial v}{\partial \alpha} > 0$ imply that $\bar{\alpha}(c)$ increases with c for $c \in (1, 2]$. □

Lemma 2

Let $0 \leq \alpha < 1$ and $c \in (1, 2]$. There exists $\underline{\alpha} = \underline{\alpha}(c) \in [0, 1]$ such that $\max_{z \geq 0} f(\alpha, z, c) \geq 0$ whenever $\alpha \leq \underline{\alpha}(c)$. Furthermore, $\underline{\alpha}$ is decreasing in c.

Proof of Lemma 2. Let $\alpha < 1$. We first prove that $f(z, \alpha, c) \geq 0$ only if $z \geq 1$. Then we show that for $z \geq 1$ f is decreasing in α. Therefore $v(\alpha, c) = \max_{z \geq 0} f(z, \alpha, c)$ is decreasing in α.

To show that $f(z, \alpha, c) \geq 0 \Rightarrow z \geq 1$ we first note that $f(z, \alpha, c) \geq 0 \Rightarrow f(z, \alpha, 1) \geq 0$ since f is decreasing in c. Hence, to show $f(z, \alpha, c) \geq 0 \Rightarrow z \geq 1$ it is sufficient to show that $f(z, \alpha, 1) \geq 0 \Rightarrow z \geq 1$. To prove this implication, we show that $z < 1 \Rightarrow f(z, \alpha, 1) < 0$. We have

$$\begin{aligned}f(z, \alpha, 1) =& \frac{1}{2}\left(-1 + 2(1 + \alpha)z - (1 + \alpha)z^{\alpha} - \alpha z^{1+\alpha}\right) \\ <& -1 + (1 + \alpha)(2z - z) - \alpha z^{1+\alpha} \\ <& -1 + (1 + \alpha)z - \alpha z^2 = (1 - \alpha z)(z - 1) < 0,\end{aligned}$$

where the inequalities follow the fact that $\alpha, z < 1$.

Next we prove that if $z \geq 1$, then f is decreasing in α. We have that:

$$\frac{\partial f}{\partial \alpha} = \frac{1}{2}\left[2z - z^{\alpha} - cz^{\alpha+1} - z^{\alpha}\left(1 + \alpha + 2(c - 1)z + \alpha c z\right)\ln z\right]$$

We need to show that $2z - z^{\alpha} - cz^{\alpha+1} - z^{\alpha}\left(1 + \alpha + 2(c - 1)z + \alpha c z\right)\ln z \leq 0$. Dividing by z, we obtain the inequality

$$2z^{1-\alpha} \leq 1 + cz + (1 + \alpha + 2(c - 1)z + \alpha c z)\ln z,$$

which holds as the LHS is not greater than 2, because $z \geq 1$ and $\alpha - 1 < 0$ and the RHS is not smaller than 2, because $c \geq 1$ and $z \geq 1$.

Therefore, whenever $f(z, \alpha, c) \geq 0$, f is decreasing in α. This implies that $v(\alpha, c) =$

$\max_{z\geq 0} f(z,\alpha,c)$ is decreasing in α whenever $v(\alpha,c)\geq 0$. We choose $\underline{\alpha}(c)$ to be equal to a zero of function $v(\alpha,c)$, whenever this zero exists on $[0,1)$, and equal to 0 otherwise. Since v is decreasing in c, and decreasing in α whenever $v(\alpha,c)\geq 0$ we have that $\underline{\alpha}(c)$ decreases when c increases. $\square$