Together alone

a group-based polarization measurement

Tang, Tanzhe; Ghorbani, Amineh; Squazzoni, Flaminio; Chorus, Caspar G.

DOI
10.1007/s11135-021-01271-y

Publication date
2021

Document Version
Final published version

Published in
Quality and Quantity

Citation (APA)
Tang, T., Ghorbani, A., Squazzoni, F., & Chorus, C. G. (2021). Together alone: a group-based polarization measurement. *Quality and Quantity*, *56*(5), 3587-3619. https://doi.org/10.1007/s11135-021-01271-y

Quality & Quantity
https://doi.org/10.1007/s11135-021-01271-y

Together alone: a group-based polarization measurement

Tanzhe Tang[1] · **Amineh Ghorbani**[1] · **Flaminio Squazzoni**[2] · **Caspar G. Chorus**[1]

Accepted: 14 October 2021

Abstract
The growing polarization of our societies and economies has been extensively studied in various disciplines and is subject to public controversy. Yet, measuring polarization is hampered by the discrepancy between how polarization is conceptualized and measured. For instance, the notion of group, especially groups that are identified based on similarities between individuals, is key to conceptualizing polarization but is usually neglected when measuring polarization. To address the issue, this paper presents a new polarization measurement based on a grouping method called "Equal Size Binary Grouping" (ESBG) for both uni- and multi-dimensional discrete data, which satisfies a range of desired properties. Inspired by techniques of clustering, ESBG divides the population into two groups of equal sizes based on similarities between individuals, while overcoming certain theoretical and practical problems afflicting other grouping methods, such as discontinuity and contradiction of reasoning. Our new polarization measurement and the grouping method are illustrated by applying them to a two-dimensional synthetic data set. By means of a so-called "squeezing-and-moving" framework, we show that our measurement is closely related to bipolarization and could help stimulate further empirical research.

Keywords Polarization measurement · Clustering · Bipolarization · Multi-dimensional polarization

1 Introduction

The vast and growing gap between liberals and conservatives, the prevalence of ideological extremes in public debate on social media, and the increasing income disparities within and across countries are shaking societies across the world. Researchers are required to conceptualize, define, and formally measure these various forms of polarization in a comprehensive manner. In the past few decades, polarization research has led to seminal contributions in a wide variety of sub-disciplines of social science, such as economics (e.g.,

✉ Tanzhe Tang
t.tang-1@tudelft.nl

[1] Delft University of Technology, Delft, The Netherlands

[2] University of Milan, Milan, Italy

Esteban and Ray 1994; Alichi et al. 2016), political sciences (e.g., Abramowitz and Saunders 2008; Hare and Poole 2014) and sociology (e.g., Flache and Macy 2011; Flache et al. 2017).

One insightful way to classify polarization is to distinguish whether there is a predefined group structure. For instance, the statement "the society is highly polarized in terms of political views" differs from the statement "the society is highly polarized in terms of political views *across ethnic groups*". Indeed, the latter requires a predefined group structure—in this case, based on ethnicity—whereas the former does not. This is a far-from-trivial issue because these two types of polarization reflect different social processes as well as different interpretations regarding all the possible cleavages cross-cutting our societies. A detailed explanation can be found in Sect. 2.

Like many other polarization studies (e.g., Duclos et al. 2004; Flache and Mäs 2008; Anderson 2011), here we are interested to measure the former type of polarization, which is more difficult and controversial because of the absence of a predefined group structure. We argue that even if there is no group structure that can be predetermined via theoretical hypotheses, the notion of group is still crucial to achieve a more rigorous measurement of polarization. While pre-existing theories about social cleavages could be used to hypothesize the existence of precise group structures in our complex societies, developing consistent measurements of polarization that might help scholars discover group structures while scanning empirical data is still key to conceptualizing and understanding polarization.

Indeed, scholars have traditionally used groups to conceptualize and define polarization (e.g., Esteban and Ray 2012; Danzell et al. 2019; Bauer 2019). For example, to return to the above-mentioned example of a society being highly polarized in terms of political views, referring to groups is important: intuitively, a highly polarized society would consist of a small number of groups whose political views are very similar within the group, but very different between groups. The division of these groups would solely reflect each individual's political views without any reference to other factors, such as ethnicity or religion. In other words, instead of being imposed by "exogenous" factors, these groups would emerge endogenously from the variable(s) of interest (here: political views).

Unfortunately, little attention has been paid to such "endogenously emerging" group structures in polarization measurements, with the notion of group usually omitted (e.g., Flache and Mäs 2008; Aleskerov and Oleynik 2016) or penalized by various theoretical and practical problems (e.g., Esteban and Ray 1994; Duclos et al. 2004). The usual difficulties of segmenting the social space of complex societies into group structures across various polarization dimensions would undermine the reliability of polarization measurements and so our understanding of the degree and extent of social polarization.

This paper aims to contribute to this field of research by proposing a novel way to generate groups as the basis of a generic class of polarization measurements without predefined group structures. The method, called "Equal Size Binary Grouping" (ESBG), uses clustering techniques to assign people (data points) to two groups of equal sizes according to the variable(s) of interest. On the one hand, this method can help researchers to identify "endogenously emerging" group structures starting from data. On the other hand, this permits to link the concept of groups to the variable(s) of interest without losing relevant information, which is often involved in theory-driven, ex-ante group conceptualization. The group structure generated by ESBG overcomes various problems, such as discontinuity and contradiction of reasoning, leading to polarization measurements that satisfy a range of important properties that have long been deemed desired in the field (Esteban and Schneider 2008; Gigliarano and Mosler 2009). Furthermore, ESBG-based measurements

are designed to measure both uni- and multi-dimensional polarization for discrete distributions. Although less frequently considered in literature, the latter has a great empirical value (see Sect. 4.1 for a more detailed discussion on this distinction).

The remainder of this paper is structured as follows: in Sect. 2, we provide a review of relevant literature concerning past attempts to conceptualize and measure polarization with or without predefined groups. We then propose a list of desired properties that an ideal polarization measurement without predefined groups should satisfy. In Sect. 3, we show that Equal Size Binary Grouping (ESBG) is a possible and promising approach to derive the ideal polarization measurement (of a particular form) adhering to these properties while being free from the problems mentioned above. Inspired by clustering algorithms, Sect. 4 presents the procedure for implementing ESBG and constructing corresponding polarization measurements. An illustrative example using synthetic data is given in Sect. 5, followed by a series of discussions in Sect. 6 about the relation between the proposed polarization measurements and bipolarization measurements. The relation is further explained by the so-called "squeezing-and-moving" framework. Section 7 summarizes the study and draws conclusions.

2 Background

2.1 Polarization and groups

For decades, the concept of polarization has received ample attention in various fields, yet, without a consensual definition. For instance, in the field of international relations, polarization usually refers to "the degree of which antipathetic, non-overlapping subgroups are formed" (Hart 1974), where these subgroups are defined according to the amity within each subgroup and the enmity between them. For example, the Allies and the Central Powers were two subgroups of nations during World War I. In economics, polarization is characterized as the "separation or distance across clustered groups in a distribution" (Esteban and Ray 2012). Given their particular interest in income polarization, economists consider a society to be polarized when the population can be grouped into significantly sized groups of individuals having similar incomes within each group, which differ across groups (Esteban and Ray 2012). In sociology, polarization in public opinion is conceptualized as "the degree to which the group can be separated into a small set of factions who are mutually antagonistic in opinion space and have maximal internal agreement" (Flache and Mäs 2008), which mirrors the definitions in international relations and economics.

These examples show that almost all definitions of polarization emphasize the notion of group, in the sense that members of the same group should be similar, and members of different groups should be dissimilar (in terms of the variable(s) of interest, such as income and opinion). Instead of the word "group", studies have used similar terms such as "clusters", "camps", "factions", or "subgroups". Regardless of the exact term being used, in all disciplines, groups, instead of individuals, are considered to be the crucial actor in conceptualizing polarization (Danzell et al. 2019).

In accordance with the development of polarization concepts, a growing number of polarization measurements have been formally proposed. A considerable portion of these measurements calculate polarization between groups that have been defined a priori[1] based

[1] I.e., the predefined groups as described in Sect. 1.

on an external variable (hereafter referred to as *grouping variable*), a variable that is different from the variable of interest.

To clarify this: when one says "our society is polarized in terms of X across (or between) Y" then X is the variable of interest, and Y is the grouping variable. For instance, when using Gigliarano-Mosler (GM) index to measure the income and education polarization between East and West Germany, the grouping variable is the location of each individual (East or West Germany), while the variables of interest are income and education (Gigliarano and Mosler 2009). Similar examples can be found in the measurement proposed by Zhang and Kanbur (ZK index) (2001), as well as Fusco and Silber (2014). These are sometimes called "social polarization measurements" (Fusco and Silber 2014) or "socio-economic polarization measurements" (Duclos and Taptué 2015), as groups are usually defined by social characteristics such as race and religion. For the sake of clarity and simplicity, here we call "**polarization with exogenously imposed groups**" the one between groups that are explicitly defined by grouping variable(s) instead of variable(s) of interest , because in these cases the grouping variables are exogenous to the variables of interest. Note that this type of measurements and relevant studies have focused on the congruency between opinion and demographic attributes - a crucial factor affecting team performance - and thus gaining interest in organization and management literature (Phillips 2003; Homan et al. 2007; Mäs et al. 2013).

However, in many other cases, it is more relevant to discuss polarization without exogenously imposed groups. Theoretically, polarization across particular socio-demographic strata (e.g., race, religion, ethnicity) is different from the polarization of the whole society. For instance, the opinion polarization of a society can be viewed as a result of opinion polarization across genders, races, locations, and countless other factors. Therefore, even if the degree of opinion polarization across one of these factors would be low, the society as a whole could still be highly polarized. Furthermore, there may also be practical objections in measuring polarization across exogenously imposed groups. Indeed, data of the grouping variables are not always available and in many cases, the only observation is the distribution of the variable(s) of interest. These arguments underline the importance of measuring polarization by defining groups in terms of the variable(s) of interest only.

Correspondingly, we call "**polarization with endogenously emerging groups**" the one where groups emerge based on the variable(s) of interest. Previous research has suggested two distinct lines of measurements of this type of polarization. The first line, started by Wolfson (1994), measures polarization in terms of "the decline of the middle class (i.e., the group with moderate value of the variable of interest)" (Foster and Wolfson 2010). Therefore, the polarization measurement would be large whenever the middle class is negligible. The second line, founded by Esteban and Ray (1994), has the basic idea that a system is considered polarized if (i) the degree of heterogeneity within each group is low, (ii) the degree of heterogeneity across groups is high, and (iii) there is a small number of significantly sized groups (Esteban and Schneider 2008).

Both lines are very popular, each with a large number of followers. The Wolfson's line is sometimes considered as the measurement of "bipolarization", which is conceptually different from the "polarization" measured by the Esteban and Ray's line (Deutsch et al. 2013). Furthermore, according to different sources of literature, bipolarization can be regarded as a category of polarization (Duclos and Taptué 2015) or a concept that is

Table 1 Summary of the concepts and measurements of polarization

Types of polarization		Corresponding measurements
with exogenously imposed groups		ZK index, GM index
with endogenously emerging groups	ER's line	ER index, DER index
	Wolfson's line	Wolfson's index, WT index
	Others	FM index, SSG index

distinct from polarization (Deutsch et al. 2013). We will interchangeably use the term "measurements in the Wolfson's line" and "bipolarization measurements". In this study, we primarily focus on the line originated by Esteban and Ray (1994). The relation between the two lines as well as our measurement will be further discussed in Sect. 6.

A common problem of the measurements in the Esteban-Ray's line concerns discontinuity. In the Esteban-Ray (ER) index (Esteban and Ray 1994), polarization is measured by the effective antagonism, which is a function of identification within groups and alienation between groups. Here, groups are defined in a particularly sharp form whereby members of the same group must have exactly the same value of the variable of interest. To give an extreme example, people with an income of 1000 euro and 1000.01 euro are in two distinct groups. Esteban and Ray (1994) themselves have acknowledged the risk of sharp groups, namely the "discontinuity problem": there will be a jump in the polarization measurement if two close groups merge. It is difficult to justify such a jump, making these sharp groups theoretically implausible. The DER index (Duclos et al. 2004) and the Anderson's index (Anderson 2011) can be viewed as the ER index of continuous variables and multi-variables respectively, and finding any group structure in these measurements is hardly feasible.

It is worth noting that a number of measurements are not covered by these two lines. The uncovered measurements may not involve the notion of endogenously emerging groups. For instance, in opinion dynamics literature, the FM index calculates the variance of the pairwise distance for all pairs of individuals (Flache and Mäs 2008; Flache and Macy 2011). Therefore, the notion of group is not included. A more recent example is the Schweighofer-Schweitzer-Garcia (SSG) index (Schweighofer et al. 2019), which is a function of the sum of squared pairwise difference. For multidimensional polarization (where there is more than one variable of interest), Aleskerov and Oleynik (2016) consider a multidimensional variable as a vector, and define "center of mass" as the weighted average of all vectors. Polarization is then measured by the weighted sum of the distances between each vector and the center of mass.

Table 1 provides an overview of the polarization concepts and measurements mentioned above.[2] It suggests that although the notion of group is crucial in defining and conceptualizing polarization, there has been no rigorous way to formalize it in order to measure polarization with endogenously emerging groups. We acknowledge that all measurements mentioned here were developed for particular research questions, and hence the absence of group structures would be acceptableto achieve simplest polarization measurements. However, we believe that with the intention to better understand and measure polarization, there should be an appropriate polarization measurement that clearly tells us what the group structure is, and how to measure polarization based on it.

[2] The measurement in the Wolfson's line will be introduced in Sect. 6.

2.2 Properties of polarization measurement

In order to tackle this problem, we propose a generic class of polarization measurements based on a novel method to define groups according to the variable(s) of interest. The method, called "Equal Size Binary Grouping" (ESBG), divides the population into two groups of equal sizes on the basis of similarities within each group and between different groups. We will show that a polarization measurement generated by this method, subject to certain requirements in the constructing procedure, satisfies various properties that have long been deemed desired in the field, including:

- **Continuity**: the measurement is a continuous function.
- **Dimensionality**: the measurement can be applied to both uni- or multi-dimensional discrete data.
- **Monotonicity**: the measurement decreases with within-group heterogeneity and increases with between-group heterogeneity
- **Maximum and Minimum**: the measurement is maximized when the population is equally divided into two maximally dissimilar groups, and members in the same group have the same value of the variable of interest. The measurement is minimized when everyone has the same value of the variable of interest.
- **Normalization**: The measurement should be in the range of 0 to 1.

The properties of continuity and normalization are important not only because of their omnipresence in literature (e.g., Esteban and Ray 1994; Chakravarty and Majumder 2001; Gigliarano and Mosler 2009), but also because a continuous and normalized polarization measurement is much easier to analyze than a discontinuous and non-normalized one.

The property of dimensionality echoes the growing interest in multidimensional polarization (Aleskerov and Oleynik 2016). We will further discuss this in Sect. 4.1.

The importance of the monotonicity property is widely acknowledged in polarization studies with exogenously imposed (Zhang and Kanbur 2001; Gigliarano and Mosler 2009) as well as endogenously emerging groups (Esteban and Ray 1994). Herein, while within-group heterogeneity refers to the heterogeneity or dissimilarity of members in the same group, between-group heterogeneity refers to the heterogeneity or dissimilarity between members of different groups. Different polarization measurements may use different expressions for these two variables. In many measurements of polarization with exogenously imposed groups, heterogeneity is represented by inequality (Zhang and Kanbur 2001; Gigliarano and Mosler 2009). In the ER index, given that each group only contains people with the same value of the variable of interest, the within-group heterogeneity is always zero and the between-group heterogeneity is simply the absolute difference between groups.

It is worth noting that in studies of polarization with endogenously emerging groups, while polarization level typically decreases with within-group heterogeneity (Esteban and Schneider 2008), there is no clear conclusion about the relation between polarization level and between-group heterogeneity. The only thing that has been confirmed is that the degree of between-group heterogeneity must be high in a highly polarized system (Esteban and Ray 1994; Esteban and Schneider 2008). Such a relatively vague description, which may be due to the lack of properly defined groups (see Sect. 3), breaks the symmetry and brings difficulty in polarization analysis. Ideally, we would

like to propose polarization measurements that not only decrease with *within-group* heterogeneity but also increase with *between-group* heterogeneity.

The importance of the maximum and minimum property has been highlighted in previous research (Gigliarano and Mosler 2009; Flache and Macy 2011; Fusco and Silber 2014; Bauer 2019). Particularly, there is hardly any polarization measurement that violates the maximum property regardless of how polarization is conceptualized. The minimum property indicates that a polarization measurement should be minimized at perfect equality, and originates from the so-called "normalization axiom" (Chakravarty and Majumder 2001). For instance, in studies of opinion polarization, the minimum condition refers to the state of consensus where everyone has the same opinion (Flache and Mäs 2008; Flache and Macy 2011; Schweighofer et al. 2020).

In addition, we emphasize that an ideal polarization measurement should also satisfy a number of axioms that have been used in constructing measurements (Esteban and Ray 1994), and are subject to some practical constraints, which will be further discussed in Sect. 3.4.

2.3 An alternative approach to measuring polarization

While all measurements previously discussed have tried to capture the overall picture of polarization by one single expression, there are alternative approaches that measure polarization in different aspects with respective indices, especially in sociological research. For instance, DiMaggio et al. (1996) suggest four distinct dimensions—dispersion measured by variance, bimodality measured by kurtosis, constraint (association between different dimensions of the variable of interest) measured by Cronbach's alpha, and consolidation (association between variable of interest and exogenously imposed groups) measured by "differences in groups' means over time" (McCright and Dunlap 2011). Bramson et al. (2016, 2017) decompose polarization into nine "senses", namely spread, dispersion, coverage, regionalization, community fracturing, (endogenously emerging) group distinctness, group divergence, group consensus, and group size parity. These dimensions and senses are largely overlapping and highly correlated.

As regards political polarization in the United States, Boxell et al. (2017) consider eight indices, each capturing a particular part of political polarization, such as: partisan affect polarization, ideological affect polarization, and partisan sorting. While these indices are mostly related to DiMaggio's dimensions and Bramson's senses, there are specificities that reflect the particular case of American politics. The point here is that each individual index alone is unable to reflect the whole picture, and this may lead to conflicting assessments. To fill this gap, Boxell et al. (2017) not only applied all eight indices to the data set, but also constructed an overall index of polarization based on the average of all indices. The advantage of this approach is twofold. First, as most aspects already have their own pre-existing measurements, scholars can easily apply them to their data sets, saving the effort of constructing a new measurement. Second, this approach displays more information than the single-expression approach, allowing scholars to discover trends or draw conclusions for each aspect.

However, there are drawbacks in this approach (DiMaggio et al. 1996; Bramson et al. 2016, 2017; Boxell et al. 2017). First, knowing how many aspects and which aspects are sufficient to capture polarization is hard. Therefore, choosing the optimal set of aspects can be difficult. Moreover, depending on different scenarios, certain aspects could be particularly salient while others would not. For instance, among DiMaggio's four dimensions, Baldassarri and Bearman

(2007) only use dispersion and bimodality, deciding to ignore others. Second, different aspects can be correlated and overlapping, thus making it difficult to design the overall index especially for quantitative research that aims at comparability, replication and cumulativeness.

Finally, it is worth noting that even in these alternative approaches, the concept of group, whether exogenously imposed or endogenously emerging, is still key to measuring polarization. For endogenously emerging groups, Bramson et al. (2017) define groups "directly from the histogram" of the distribution plot. Further analysis of this type of grouping methods (in the context of bimodality) and its comparison with our method (i.e., ESBG) can be found in Sect. 4.5.

3 Derivation of Equal Size Binary Grouping

When you have eliminated the impossible, whatever remains, however improbable, must be the truth. (Sherlock Holmes)

The aim of this section is to justify ESBG as an appropriate grouping method for constructing ideal polarization measurements. To achieve this aim, after clarifying the notations (Sect. 3.1), we will show that ESBG is a possible solution to the problems afflicting other grouping methods (Sect. 3.2.3): the grouping method without any constraints suffers from the discontinuity problem (Sect. 3.2.1), and the grouping method only constrained by a fixed number of groups contradicts Esteban and Ray's reasoning (Esteban and Ray 1994) (Sect. 3.2.2). Furthermore, in Sect. 3.3, we will explain how ESBG takes into account the roles of the missing variables, namely the number and size of groups, by providing some examples. Finally, in Sect. 3.4, we will test if the ESBG-based polarization measurement satisfies the axioms proposed by Esteban and Ray (1994).

3.1 Notations

We first present the following notations that will be used throughout the rest of the paper. Suppose we are interested in a discrete system (i.e. data set) $X \equiv \{\mathrm{x}_1, ..., \mathrm{x}_N\}$ consisting of N data points. A data point $\mathrm{x}_i = (x_{i,1}, ..., x_{i,D})$ $(i = 1, ..., N)$ is described by its variables $x_{i,d}$ $(d = 1, ..., D)$, where D is the dimension of the system. A grouping method $G : X \to C$ partitions the system X into K non-overlapping groups $C = \{C_1, ..., C_K\}$. The size of a group C_k is denoted by s_k, representing the number of data points in C_k. The within-group heterogeneity of a group C_k is denoted by w_k, and the between-group heterogeneity of a pair of groups C_i and C_j $(i \neq j, i, j = 1, 2, , , K)$ is denoted by $b_{i,j}$. In Sect. 4, we will further discuss how to calculate w_k and $b_{i,j}$.

At the end of Sect. 2, we have listed a range of desired properties that an ideal polarization measurement should adhere to. Assume now that we already have completed the task of partitioning the data set into groups, then the polarization measurement should be a function of at least the following two factors: within-group heterogeneity and between-group heterogeneity. Intuitively, the number of groups (K) and the size of each group ($S = \{s_1, ..., s_K\} \in \mathbb{R}^K$) may also affect the polarization level. Therefore, such a polarization measurement should have the following form:

$$P(X) = f(W, B, K, S) \tag{1}$$

where $W \in \mathrm{IR}^+$ and $B \in \mathrm{IR}^+$ are indices for within-group heterogeneity and between-group heterogeneity of the entire data set respectively. As the desired properties suggest, P should

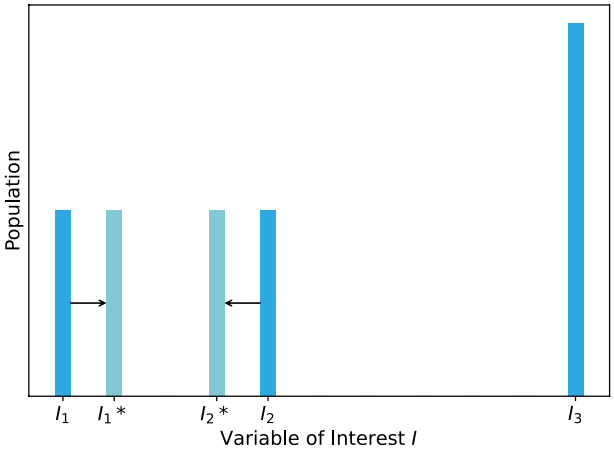


Fig. 1 Diagram to illustrate the failure of G_0

be decreasing with W and increasing with B. Following Gigliarano and Mosler (2009) where choices of W and B are mostly related to the weighted sum of each group's characteristics, we further assume that the two variables should take the following forms:

$$W = \phi\left(\sum_{k=1}^{K} \alpha_k w_k\right) \tag{2}$$

$$B = \psi\left(\sum_{i<j} \beta_{i,j} b_{i,j}\right) \tag{3}$$

where ϕ and ψ should be strictly increasing and continuous. $\alpha_k > 0$ and $\beta_{i,j} > 0$ are real number coefficients, representing the weights or importance of corresponding variables. There are good reasons for using such linear expressions ($\sum_{k=1}^{K} \alpha_k w_k$ and $\sum_{i<j} \beta_{i,j} b_{i,j}$) as inputs of ϕ and ψ. As we will see in Sects. 3.2, 3.3, and 3.4, the linearity will significantly simplify our analysis about the properties of P, by making it possible to directly obtain the changes in W and B during certain dynamical processes. These changes may be intuitive and are not of our main interest here, but not giving specific forms or using other expressions of W and B might make the formal derivation of the outcome tedious and difficult, if still possible. For example, in Fig. 1 of Sect. 3.2.1, there are three groups at I_1,I_2, and I_3 ($I_1 < I_2 < I_3$). If I_1 and I_2 move to each other for the same distance, we intuitively anticipate that $B = \psi(b_{1,2}, b_{2,3}, b_{1,3})$ should decrease, but it is not easy to prove: it is unclear if B will decrease as $b_{1,3}$ decreases but $b_{2,3}$ increases. Nonlinear forms of ψ, such as product, may require extensive efforts to confirm the result, while expression (3) can solve it easily through simple calculation (see Sect. 3.2.1). This will become clearer in the rest of the section thanks to some further examples. Given the benefit of linearity, and the lack of advantage of nonlinear expressions, we choose Eqs. (2) and (3) for the rest of the paper.

Following the linearity in W and B, in this section we further take the following assumption: $b_{i,j}$ should be the squared distance between the centers (or mean values) of C_i and C_j. Similarly, w_k should be the average squared distance between members of C_k and the center (mean value) of C_k. We do not aim to rule out other forms of $b_{i,j}$ and w_k, but this assumption will significantly simplify our analysis in the Sect. 3.2. For example, in

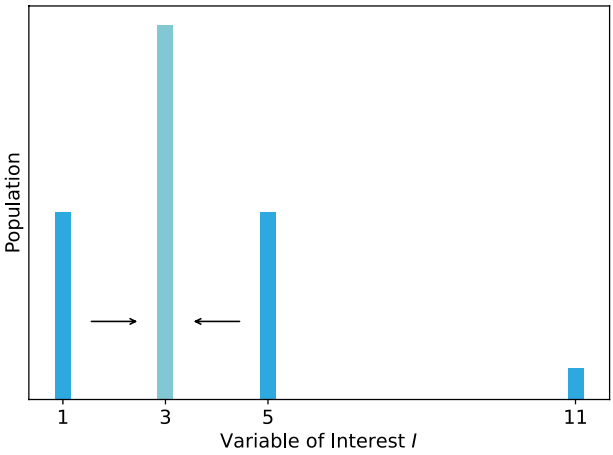


Fig. 2 Diagram to illustrate the failure of G_1

Fig. 2 of Sect. 3.2.2, there are two groups: C_1 that contains people at 1 and 5, and C_2 that contains people at 11. If people at 1 and 5 move towards each other with a same distance, we can easily show that $b_{1,2}$ stays fixed with this assumption.

3.2 Searching for grouping methods

The lack of a proper grouping method is the root for the absence of well-defined "endogenously emerging" groups in polarization measurements. In general, a grouping method is a series of steps that separate the data set into a finite number of (non-overlapping) groups, where members of the same group should be similar and members from different groups should be dissimilar according to some criteria. Additionally, multiple constraints—including the number and size of groups—can be applied to a grouping method based on prior knowledge or specific requirements. In this subsection, we consider three types of grouping methods: method without any constraint, method with a fixed number of groups, and method with both fixed number and size of groups. Conceptually, the three types represent all possible grouping methods. We will show that a particular grouping method of the last type, called "Equal Size Binary Grouping" (ESBG), which divides the data set into two equally sized groups, should be a possible solution to problems such as discontinuity and contradiction of reasoning if we want to construct an ideal polarization measurement that (i) is in the form of Eqs. (1), (2), and (3), and (ii) adheres to the desired properties.

A common requirement for endogenously emerging groups is that they should be formed on the basis of (dis)similarities between individuals (i.e. data points), so that each group is homogeneous internally but different from other groups. Let us assume that all the grouping methods discussed in this subsection satisfy this requirement. This implies that each of them is able to classify data points that are sufficiently similar into the same group and classify the data points that are sufficiently dissimilar into different groups. We will leave the question "how to perform these grouping methods to ensure that they satisfy this requirement" to Sect. 4, where more technical details will be provided.

3.2.1 Grouping method without any constraint

Suppose there is a grouping method G_0 whose *only* task is to divide the system into groups. Therefore, there is no constraint on G_0 besides the requirement mentioned above, and the number and size of the groups are determined to make the members of the same group similar and members of different groups dissimilar.

To understand why G_0 is not a proper grouping method for the polarization measurement $P(X)$, consider the uni-dimensional example given in Fig. 1, modified from Esteban and Ray (1994). In Fig. 1, initially (at $t = 0$), half of the population is equally distributed between level $I = I_1$ and $I = I_2$ (I is the variable of interest), and the other half of the population is at level $I = I_3$. Suppose $0 < I_1 < I_2 < I_3$, $I_3 - I_2 \geq I_2 - I_1$, and the three levels are sufficiently different such that G_0 will produce three non-overlapping groups C_1, C_2, and C_3, containing data points at I_1, I_2 and I_3 respectively. Therefore, $w_k = 0$ $(k = 1, 2, 3)$. Now, consider that both C_1 and C_2 move towards each other synchronously with the same speed until merging. $X(t)$ is the system at time t. During the process, there must be a moment $t = t^*$ when C_1 moves to I_1^* $(I_1 < I_1^* < I_2)$ and C_2 moves to I_2^* $(I_1^* < I_2^* < I_2)$,[3] and G_0 **starts** to recognize C_1 and C_2 as one group, denoted by C_4. The transition moment t^* fully depends on G_0 if the moving speed is given. When $t < t^*$, the between-group heterogeneity $B = \psi(\beta_{1,2}b_{1,2} + \beta_{2,3}b_{2,3} + \beta_{1,3}b_{1,3})$ is decreasing (with an intuitive condition that $\beta_{1,3} = \beta_{2,3}$).[4] Due to the fact that W, K, and S are constant, the decrease in B implies that $P(X)$ decreases with t when $t < t^*$. When $t > t^*$, there will be only two groups C_4 and C_3, and $W = \phi(\alpha_4 w_4 + \alpha_3 w_3)$ decreases with t as w_4 is decreasing, while other factors stay constant, therefore $P(X)$ is increasing. To conclude, $P(X(t))$, as a function of t, is decreasing when $t < t^*$ and is increasing when $t > t^*$.

Assume now that we have a new grouping method G_0^{II}, which is slightly different from G_0 in the sense that the transition moment for G_0^{II} is $t^{**} > t^*$. Denote the polarization measurement of X using G_0 as $P(X|G_0)$ and using G_0^{II} as $P(X|G_0^{II})$. When $t < t^*$ or $t > t^{**}$, G_0 and G_0^{II} are of no difference and hence $P(X|G_0) = P(X|G_0^{II})$. $P(X|G_0) = P(X|G_0^{II})$ when $t < t^*$ implies that $\lim_{t\uparrow t^{**}} P(X(t)|G_0^{II}) < P(X(t = t^{**})|G_0)$ if we assume both $P(X|G_0)$ and $P(X|G_0^{\mathrm{II}})$ are continous in t, as $P(X|G_0)$ is increasing during $t^* < t < t^{**}$, and $P(X|G_0^{II})$ is decreasing during the same period. Meanwhile, $P(X|G_0) = P(X|G_0^{II})$ when $t > t^{**}$ implies that $\lim_{t\downarrow t^{**}} P(X(t)|G_0^{II}) = P(X(t = t^{**})|G_0)$, therefore $\lim_{t\uparrow t^{**}} P(X(t)|G_0^{II}) < \lim_{t\downarrow t^{**}} P(X(t)|G_0^{II})$, which directly proves that the $P(X(t)|G_0^{II})$ is discontinuous at $t = t^{**}$. Given that there are countless transition moments generated by countless grouping methods without any constraint, we can conclude that $P(X(t))$ is a discontinuous function of t. Note that not only does such discontinuity exist in our example; it is likely to occur whenever two (or even more than two) groups merge.

Indeed, this problem of G_0 is the same as the discontinuity problem observed in the ER index (see Sect. 2.1). Besides being counter-intuitive, this discontinuity will cause various problems. For example, the sudden jump of the polarization level at the transition moment

[3] If $I_1^* = I_2^*$, G_0 is the same as the ER index. See Sect. 2.1.

[4] It is clear that the difference between C_1 and C_2 (i.e., $b_{1,2}$) shrinks, so we only need to prove that $\beta_{2,3}b_{2,3} + \beta_{1,3}b_{1,3}$ is not increasing, As it is intuitive to have $\beta_{1,3} = \beta_{2,3}$, the task is reduced to prove $b_{1,3} + b_{2,3}$ is not increasing. Suppose C_1 or C_2 has traveled a distance of Δ $(0 < \Delta < (I_1^* - I_1))$ at time t' $(t < t' < t^*)$ since t (C_1 and C_2 always travel the same distance), then $[b_{1,3}(t') + b_{2,3}(t')] - [b_{1,3}(t) + b_{2,3}(t)] = [(d_{1,3}(t) - \Delta)^2 + (d_{2,3}(t) + \Delta)^2] - [(d_{1,3}(t))^2 + (d_{2,3}(t))^2] = 2(d_{2,3}(t) + \Delta - d_{1,3}(t))\Delta < 0$, where $d_{i,j}(t)$ is the distance between I_i and I_j at t. Therefore $b_{1,3} + b_{2,3}$ is indeed decreasing. See Sect. 3.1 for the reason of choosing $b_{i,j} = (d_{i,j})^2$.

is hardly justifiable. If such discontinuity is accepted, one can dramatically increase or decrease the polarization level of the same data set by simply constructing a slightly different transition moment.

3.2.2 Grouping method with fixed number of groups

To solve the discontinuity problem mentioned in Sect. 3.2.1, we impose a constraint on G_0: the number of groups is fixed to $K = 2$. We denote this grouping method as G_1. The task of G_1 is to divide the systems into two groups such that the two groups are maximally different, but members in the same group are maximally similar. To show that G_1 overcomes the discontinuity problem, we also apply G_1 to $X(t = 0)$ in Fig. 1. Since $I_3 - I_2 > I_2 - I_1$, individuals at I_1 and I_2 are classified into one group, say, C_4, and individuals at I_3, as before, constitute the other group C_3. Now the dynamic process described in Sect. 3.2.1 only decreases w_4, while w_3, B, and S are not affected. Therefore $P(X(t)|G_1)$ increases continuously throughout the process, that is, it does not suffer from discontinuity.

When I_2 is closer to I_3 than to I_1, is the method still discontinuity-free? Again, consider that both individuals at I_1 and I_2 move towards each other with the same speed simultaneously. Initially, G_1 will define two groups: C_1 containing everyone at I_1 and C_5 containing everyone at I_2 or I_3. There will also be a transition moment $t = t^*$ when G_1 starts to consider individuals at I_1^* and I_2^* as one group C_4. The question is if $P(X|G_1)$ is discontinuous. Through simple analysis, we know that $P(X|G_1)$, again, decreases with t when $t < t^*$ and increases with t when $t > t^*$. However, it is intuitive to see that no matter which G_1 we choose, the transition moment t^* is always the moment when $I_3 - I_2 = I_2 - I_1$. Otherwise, the group structure will violate the basic requirement mentioned in the beginning of this section. Therefore, the method is free from discontinuity.

Although providing a solution to discontinuity, G_1 has its own problem. Consider another example modified from Esteban and Ray (1994). As shown in Fig. 2, almost all individuals are placed equally at $I = 1$ and $I = 5$, while only a sufficiently small number of individuals are at $I = 11$. G_1 will put individuals at 1 and 5 in the same group, say C_1, leaving those at 11 in another group C_2. Now, consider all individuals in C_1 merge at $I = 3$. The merge reduces w_1 to 0, while all other factors remain unchanged. Consequently, the polarization level should go up. However, according to Esteban and Ray (1994), due to the relatively small size of C_2, the initial polarization mostly comes from the dissimilarity between the individuals at 1 and 5, which is eliminated after the merge, so the polarization level should go down. This contradiction discourages using G_1 for constructing $P(X)$. Note that choosing another value for K not only lacks a strong theoretical justification but also is unable to solve the discontinuity problem in Fig. 1.

3.2.3 Grouping method with fixed number and size of groups

From Sect. 3.2.2, we know that the problem in G_1 is due to group size. To solve this problem, we impose another constraint on G_1: the size of each group must be the same. We call it **Equal Size Binary Grouping** (**ESBG**), whose task is to split the system into two equally sized groups, while maximizing between-group heterogeneity and/or minimizing within-group heterogeneity. For the sake of simplicity, we only discuss systems whose size is an even number. We will discuss how to implement this method in Sect. 4.

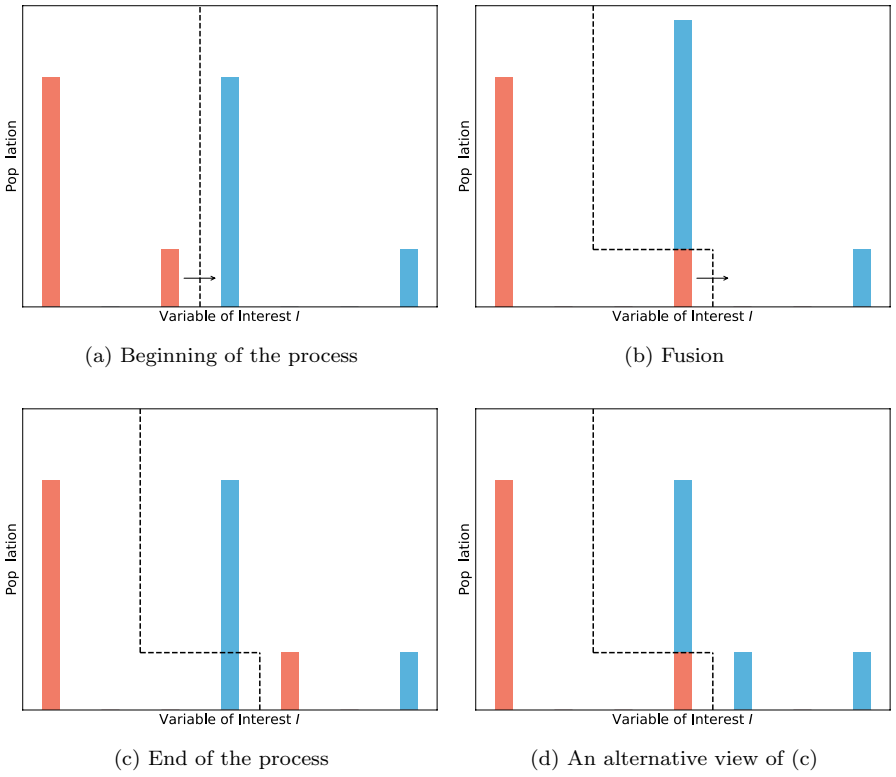
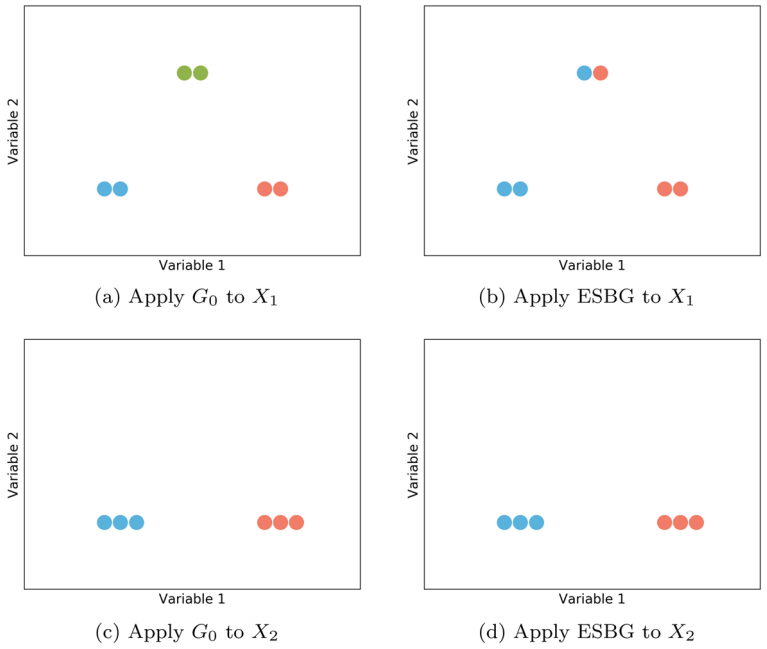


Fig. 3 Illustration of the continuity of ESBG. Colors indicate the initial membership of each individual; arrows represent the moving direction; and the dashed line is the current group boundary

First, the discontinuity problem in Fig. 1 can be completely solved by replacing G_0 with ESBG as there will be no transition moment during the process. For a more general case, consider the dynamic process described in Fig. 3, in which ESBG initially divides the uni-dimensional system into two groups: C_1 in red and C_2 in blue (Fig. 3(a)). Without loss of generality, we suppose that a portion of C_1 move towards C_2, and will stop after passing the closest member of C_2 (Fig. 3(c)). Note that the colors in the figure only indicate the **initial** group memberships.

We can see that compared to G_0, moving sufficiently close to individuals of another group can no longer trigger a transition of group membership under ESBG (Fig. 3(a)). Only when the moving red individuals pass the blue individuals near the group boundary, there will be a transition of group membership as the moving individuals, previously members of C_1, will be now identified as members of C_2 (Fig. 3(c)). However, we can take an alternative look at this situation. The identity of an individual is purely determined by its value of the variable of interest, therefore among the individuals in the middle in Fig. 3(b), ESBG cannot tell which individual just moved here from the left (red), and which individual is native (blue). Therefore, the dynamics from Fig. 3(b) to (c) can be equivalently interpreted as the dynamics from Fig. 3(b) to (d), that is, a portion of native blue individuals move to the right, and no membership transition happens during the whole process. Due to the arbitrariness of this example, we can



(a) Apply G_0 to X_1 (b) Apply ESBG to X_1

(c) Apply G_0 to X_2 (d) Apply ESBG to X_2

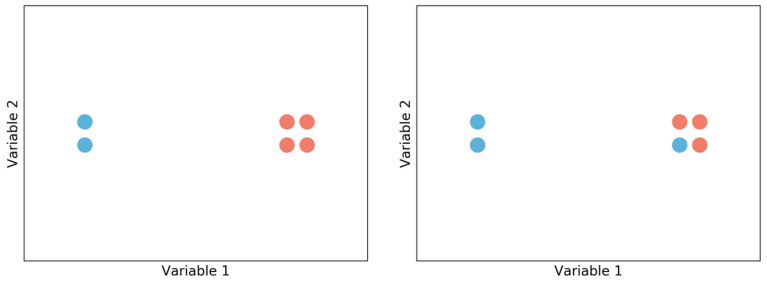
Fig. 4 Applying G_0 and ESBG to data set X_1 and X_2 respectively. In (**a**), (**c**), and (**d**), points of the same color are not only in the same group but also identical, while in (**b**), color only represents group membership. To avoid overlap and improve readability, positions of the data points have been adjusted

conclude that the discontinuity problem caused by group membership transition can be solved by ESBG.

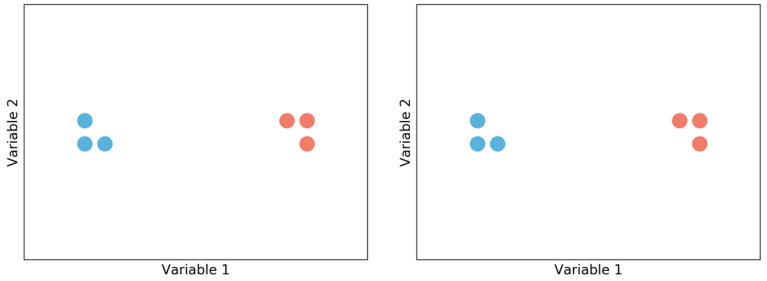
Finally, we show that ESBG has the potential to solve the problem found in G_1. In the example given in Fig. 2, ESBG will define groups differently from G_1: all individuals at 1 and a small number of individuals at 5 will form C_1, and the rest of the population will form C_2. During the merging process in Fig. 2, we can confirm that B decreases given a sufficiently small population at 11.[5] Meanwhile, w_1 decreases and w_2 increases, so it is unclear whether W increases or not. However, it gives us room to design expressions for W, B, and $f(W, B)$ in order to solve the problem, which is much better than G_1 where polarization will definitely increase (see Sect. 3.2.2). For example, Sect. 5.3 shows a particular implementation of f which should be able to solve the problem of G_1 (see Table 2).

To summarize, both G_0 and G_1 are not qualified as grouping methods for constructing the ideal polarization measurement because of discontinuity and contradiction of reasoning, while ESBG should be a possible solution.

[5] Note that this confirmation needs sufficiently small size of individuals at 11. Denote the members of C_1 at 1 as C_1^1, the members of C_1 at 5 as C_1^5, the members of C_2 at 5 as C_2^5, and the members of C_2 at 11 as C_2^{11}. During the merging process, the heterogeneity between C_1^1 and both C_2^5 and C_2^{11} decreases, while the heterogeneity between C_1^5 and C_2^5 is fixed at 0. Meanwhile, the heterogeneity between C_2^{11} and C_1^5 increases. However, if the size of C_2^{11} is small enough (the size of C_1^5 is smaller than C_2^{11}), we can conclude that B decreases.



(a) Apply G_0 to X_3 (b) Apply ESBG to X_3



(c) Apply G_0 to X_4 (d) Apply ESBG to X_4

Fig. 5 Applying G_0 and ESBG to data set X_3 and X_4 respectively. In (**a**), (**c**), and (**d**), points of the same color are not only in the same group but also identical, and the distance between points with the same color does not represent the difference between them but is introduced to improve readability. In (**b**), color only represents group membership

3.3 ESBG and the missing variables

By using ESBG, the expression $P = f(W, B, K, S)$ reduces to $P = f(W, B)$. However, removing K and S does not mean the measurement fails to include the effects of these two variables. Indeed, the role of K and S are inherited by W and B. In practice, $S = \{s_1, ..., s_K\}$ is represented by the **relative group sizes** (*RS*), which measures "how equally populated the groups are" (Gigliarano and Mosler 2009). Large *RS* implies that the group sizes are similar, and small *RS* implies unequal distribution of group sizes.

Figures 4 and 5 provide vivid examples in a two-dimensional space. In Fig. 4(a), a constraint-free grouping method (i.e. G_0) divides the data set X_1 into three groups, each containing two identical individuals. The distance (i.e., heterogeneity) between each group is assumed to be the same. Given the same data set, ESBG will divide X_1 into two groups, which means one of the three groups defined by G_0 (in Fig. 4(a), the green group) will be equally separated and taken by the remaining groups (Fig. 4(b)). Now, consider another data set X_2 where G_0 divides it into two groups (i.e., the blue and the red groups), each containing three identical individuals (Fig. 4(c)). ESBG will make the same division (Fig. 4(d)) as G_0. It is not difficult to find out that (a) and (c) have the same W, B and *RS*.[6]

[6] Both **(a)** and **(c)** have reached the maximum of *RS* (see Gigliarano and Mosler (2009) for details). A condition for **(a)** and **(c)** to have the same B is $\sum_{i<j}^{K} \beta_{i,j} = C, \forall K > 0$, where C is a constant.

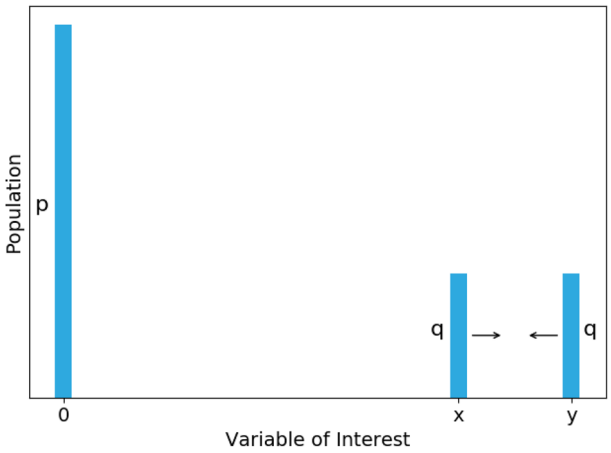


Fig. 6 Esteban-Ray's Axiom 1: Data: $p, q \gg 0$, $p > q$, $0 < x < y$. Statement: Fix $p > 0$ and $x > 0$. There exist $\epsilon > 0$ and $\mu > 0$ such that if $d(x, y) < \epsilon$ (d is the distance function) and $q < \mu p$, the joining of the two q masses at their mid-point, $(x + y)/2$, increases polarization. (Note: This statement, as well as Axiom 2 and 3, are directly taken and modified from Esteban and Ray (1994))

Therefore, the only difference between X_1 and X_2 under G_0 (i.e., (a) and (c)) is the number of groups, K. However, when using ESBG to measure polarization (i.e., (b) and (d)), both data sets have the same $K = 2$. Comparing (b) and (d), we find that the data set with a larger K (i.e., X_1) under G_0 will have a larger W but a smaller B under ESBG, indicating that X_1 is less polarized than X_2 not only under G_0 but also under ESBG (intuitively a polarization measurement is negatively related to K when $K \geq 2$). To conclude, the effect of K under G_0 is replaced by the effect of W and B under ESBG.

Figure 5 shows how ESBG converts the effect of RS to the effect of W and B. In Fig. 5(a), G_0 divides the data set X_3 into two groups, each containing two or four identical individuals (i.e., the blue and red group). Meanwhile, ESBG will divide X_3 into two groups each containing three members as shown in (b). Figure 5(c) and (d) show that both G_0 and ESBG divide another data set X_4 into two groups each containing 2 identical individuals. Assume the distances between the two groups in (a) and (c) are the same, then the only difference between X_3 and X_4 under G_0 is RS. Without any calculation, it is clear that X_3 has a smaller RS than X_4, indicating that X_4 is more polarized (because intuitively a polarization measurement is positively related to RS). Comparing (b) and (d) (where RS no longer matters), X_3 has a larger W and a smaller B, indicating that X_3 is less polarized than X_4 under ESBG, in line with the prediction by G_0. Therefore, the effect of S (via RS) under G_0 is replaced by W and B under ESBG.

3.4 Polarization axioms by Esteban and Ray

In this subsection, we test whether an ESBG-based polarization measurement, even without a particular expression, satisfies the axioms proposed by Esteban and Ray (1994). For the sake of simplicity, we reduce Eqs. (2) and (3) to $W = \phi(w_1 + w_2)$, and $B = \psi(b_{1,2})$, which will be further justified in Sect. 4.3.

Axiom 1 (Fig. 6)

To justify this statement, assume that μ is small enough such that $2q < p$ (i.e. $\mu < 1/2$). Therefore, under ESBG, the two q masses are always in the same group, say, C_2. A part of the p mass will also be in C_2, and the rest of the p mass will be in the other group C_1. Given

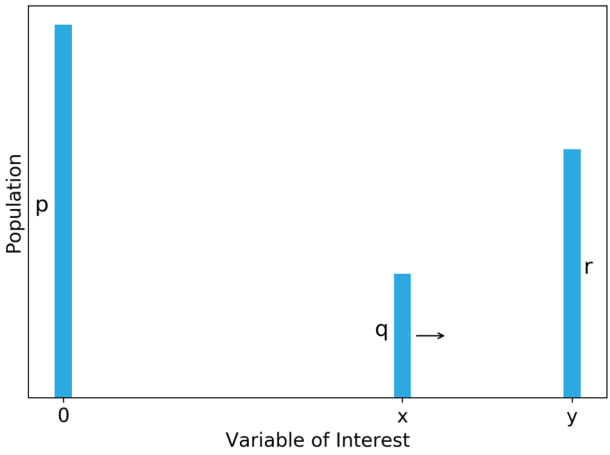


Fig. 7 Esteban-Ray's Axiom 2: Data: $(p, q, r) \gg 0$, $p > r$, $x > |y - x|$. Statement: There exists $\epsilon > 0$ such that if the population mass q is moved to the right (towards r) by an amount not exceeding ϵ, polarization goes up

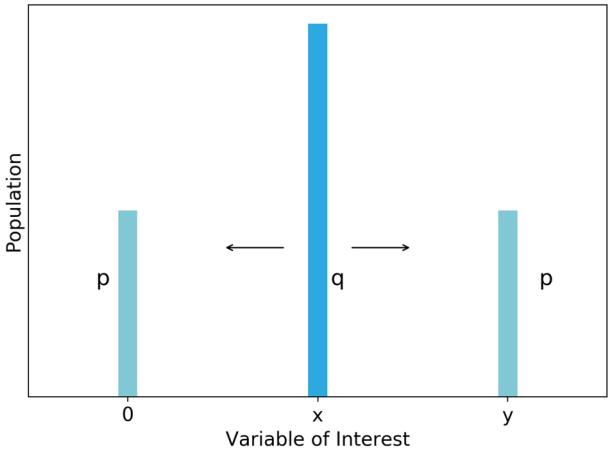


Fig. 8 Esteban-Ray's Axiom 3: Data: $(p, q) \gg 0$, $x = y - x \equiv d$. Statement: Any new distribution formed by shifting population mass from the central mass q equally to the two lateral masses p, each d units of distance away, must increase polarization

that the merge does not affect the center of C_2, B is not affected. Meanwhile w_1 is obviously not affected, but w_2 decreases, which will increase $f(W, B)$.

The condition $d(x, y) < \epsilon$, in the original paper (Esteban and Ray 1994), was proposed to ensure that the two q masses were sufficiently close. Under ESBG, this condition is no longer needed.

Axiom 2 (Fig. 7)

If $p > (p + q + r)/2$, q mass, r mass, and a part of the p mass will be in the same group C_2, while the rest of the p mass will be the other group C_1. After the move, w_1 is not affected and B goes up. If w_2 decreases, we have P increasing as the axiom requires. If w_2 increases, it does not increase as much as B does (given that B and W are on the same scale): on the one hand, the move of the q mass will decrease the heterogeneity between the q and r mass, which deteriorates w_2; on the other hand, the move increases the heterogeneity between the q mass and everyone in the p mass. However, there are more members of C_1 than members of C_2 in the p mass, implying that this move should affect B much more than w_2. Therefore, polarization should go up after the move given a properly designed measurement. Here we provide a simple intuition rather than a formal proof.

If $p = (p + q + r)/2$, the move will decrease w_2 and increase B, thus $f(W, B)$ will increase.

If $r < p < (p + q + r)/2$, more than half of the q mass will be in C_2, while the rest of the q mass will be in the other group C_1. The move of the q mass increases w_1 but reduces w_2, which should lead to a decrease in $w_1 + w_2$ because more than half of the q mass is in C_2. Due to the same reason, the move will increase B. As a result, $f(W, B)$ will go up.

Axiom 3 (Fig. 8)

Before the split, $W = B = 0$. After the split, $W = 0$ and $B > 0$, and hence the polarization measurement goes up.

4 Implementing grouping method and constructing polarization measurement

As argued above, the task of ESBG is to split the data set into two equally sized groups in such a way that members from different groups are very different, and members within each group are very similar. In Sect. 3, we did not discuss how ESBG achieves this task or how to implement ESBG, but took it for granted. In this section, using ideas of clustering techniques, we propose an implementation protocol for ESBG, especially in a multi-dimensional space, based on which the ESBG-based polarization measurement will be constructed. In addition, we compare the ESBG-based measurement with bimodality measurements as they share similar expressions.

4.1 Dimensionality

If we are only interested in polarization of a uni-dimensional data set, implementing ESBG is as simple as dividing the data set by the median value. In this subsection, we will stress on the importance of multi-dimensional polarization, justifying the necessity of implementing ESBG in multi-dimensional spaces.

Multi-dimensional polarization is not the simple aggregation of uni-dimensional polarization from different dimensions. Therefore, measuring multiple uni-dimensional polarization cannot tell how polarized the whole system is. Following the example given by Ross (1920), consider a society with half white men and half black men. Therefore, the society is ethically polarized; meanwhile, the society consists of half employees and half employers, so it is also polarized in social classes. If all white men are employed by black men, the society is polarized as a whole with half white employees and half black employers. However, if half white/black men are employers and half white/black men are employees, the society is actually split into four groups: white employer, white employee, black employer, and black employee. Therefore, the society is polarized in all dimensions, but is less polarized as a whole. This also creates problems of micro vs. macro level measurements, as suggested by research on group segregation in labour markets (e.g., Takács et al. 2018).

The majority of polarization measurements are designed to measure uni-dimensional data only. Given the necessity of measuring multi-dimensional polarization, implementing ESBG for both uni- and multi-dimensional data sets is of paramount importance.

4.2 Clustering

Clustering is one of the most important topics in data analysis and machine learning, which has been extensively studied due to its broad functionality (Jain et al. 1999; Xu and Wunsch 2009). In short, clustering is "the task of partitioning the data set into groups" (Müller and Guido 2016), and members in the same group "should display similar properties based on some criteria" (Xu and Wunsch 2009). A twin concept of clustering is called supervised classification that depends on a set of pre-classified/pre-labeled data. From the pre-classified data (also called "training data"), a supervised classification technique learns how to define groups, and then divides unlabeled data into groups (Jain et al. 1999). Unlike supervised classification, clustering deals with unlabeled data only, which means groups "are obtained solely from the (unlabeled) data" (Jain et al. 1999). This feature naturally reminds us of the "endogenously emerging groups" that are solely derived from the variable(s) of interest, indicating that clustering is fundamentally similar to the task of defining endogenously emerging groups. Thanks to the development in the field, there is a vast collection of efficient and reliable clustering algorithms (see Jain et al. (1999), Baraldi and Blonda (1999a, 1999b), Xu and Wunsch (2009) for reviews), which will pave the way for the implementation of ESBG.

A typical clustering process usually consists of the following steps (Jain and Dubes 1988; Jain et al. 1999; Xu and Wunsch 2009):

Feature selection and/or extraction: It is a necessary preprocessing step for clustering. Because not all features (in our context, dimensions) are "equally relevant" for clustering (Aggarwal 2014), for the sake of efficiency, feature selection chooses the most relevant and effective set of features for defining groups (Jain et al. 1999). In addition, feature extraction transforms original features into new forms that are more salient.

Definition of a proximity measurement: As argued above, data points are clustered into groups according to how "close" they are to each other. To implement clustering, we need to formally define a proximity measurement. The term "proximity" is the counterpart of "homogeneity" in the context of polarization. Therefore, measuring proximity in clustering echoes measuring within- and between-group homogeneity/ heterogeneity in EBSG.

Grouping/optimization: This is the main step of clustering. Given the proximity measurement, the grouping step is "an optimization problem with a specific criterion function" (Xu and Wunsch 2009), and the criterion is closely related to the proximity measurement.

Validation: This step assesses the output produced by previous steps depending on some optimal criteria (Jain et al. 1999).

4.3 Implementing ESBG

Based on the steps of a clustering process, a formal ESBG process should include the following steps:

Preprocessing: This step mirrors the feature selection and/or extraction step in clustering. A common concern about multi-dimensional polarization is the incommensurability of dimensions. For instance, when measuring the two-dimensional polarization of education and income, it is difficult to defend why an x year difference in education and a y euro difference in income are equally important (x and y are arbitrary positive numbers). Furthermore, some less relevant dimensions might harm the efficiency of ESBG. The preprocessing step should help to solve these issues by techniques such as dimension reduction and

rescaling (for details, see Aaberge and Brandolini (2015)). After the preprocessing, these dimension-related problems should no longer exist in the processed data.

Definition of a heterogeneity measurement: In this step, we need to design a heterogeneity measurement appropriate to our data. Just like proximity (Jain et al. 1999), heterogeneity can be measured by a distance function—for example, the Euclidean distance—of pairs of data points. Once the heterogeneity measurement is chosen, by denoting the two groups as C_1 and C_2, the expressions of heterogeneity within each group (w_1 and w_2) and between groups ($b_{1,2}$) can be determined. The within-group heterogeneity W and between-group heterogeneity B should be calculated according to the following equations:

$$W = w_1 + w_2 \tag{4}$$

and

$$B = b_{1,2} \tag{5}$$

Equations (4) and (5) are the reduced forms of Eqs. (2) and (3) respectively. The omission of the parameters α_1 and α_2 in Eq. (4) is due to the equity of group sizes. At the same time, we take the simplest possible expression of ϕ: $\phi(w_1 + w_2) = w_1 + w_2$. For the expression of B, since there are now only two groups, the overall between-group heterogeneity B is the same as the heterogeneity between C_1 and C_2, i.e., $b_{1,2}$.

Grouping: Given the expression of W and B, ESBG is translated to an optimization problem with the aim of maximizing B and/or minimizing W, subject to the constraint that the group number must be 2, and the sizes of the two groups must be the same.

Validation: The validation process in ESBG is almost the same as in clustering, but with an additional exam on the number and size of groups.

In practice, it is easy to choose a well developed clustering algorithm as the basis of ESBG. In Sect. 5, we will develop the implementation of ESBG based on the famous K-means clustering algorithm.

4.4 Constructing a polarization measurement

Given the endogenously emerging groups C_1 and C_2 defined by ESBG, a polarization measurement should take the following form:

$$P(X) = f(W, B) = f(w_1 + w_2, b_{1,2}) \tag{6}$$

since we have used $W = w_1 + w_2$ and $B = b_{1,2}$ (see Sect. 4.3).

When designing the expression for the measurement, it is important to ensure that all the desired properties for $P(X) = f(W, B)$ listed in Sect. 2.2 have been taken into account. These properties are formally summarized as follows:

Continuity: $P = f(W, B)$ is a continuous function of both W and B.

Dimensionality: $P : \mathbb{R}^D \to \mathbb{R}$. $D = 1$ or $D \geq 2$.

Monotonicity: $P = f(W, B)$ is strictly decreasing with W and strictly increasing with B.

Maximum: P is maximized when $W = 0$ and B is maximized.

Minimum: P is minimized when $W = B = 0$.

Normalization: For all $X \in \mathbb{R}^D$ $(D \geq 1)$, $0 \leq P(X) \leq 1$.

Combining the maximum property and the normalization property, we have $f(W = 0, B = B_{max}) = 1$, where B_{max} is the maximum between-group heterogeneity. However, in practice, determining the value of B_{max} can be troublesome. Suppose we define B_{max}

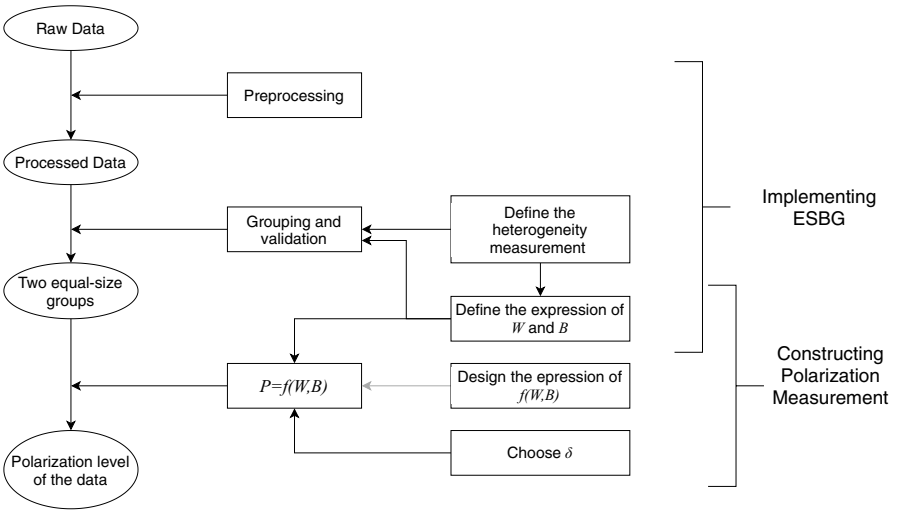


Fig. 9 Procedure of implementing ESBG and constructing polarization measurement

as the maximal pairwise distance in a data set X_1, and then we design an expression of $f(W, B)$, say f_1, such that $f_1(W(X_1) = 0, B(X_1) = B_{max}) = 1$. Then for another data set X_2 whose between-group heterogeneity $B(X_2)$ is larger than B_{max}, we will obtain $f_1(W(X_2) = 0, B(X_2)) > 1$, violating the normalization property. To solve the issue, we introduce the normalizing parameter $\delta > 0$, which should be greater than or equal to the maximum possible heterogeneity in all the data sets of interest. Formally, if we want to compare the polarization level of X_1,X_2,..., and X_M, then $\delta \geq max_m\, max_{x_i,x_j \in X_m} h(x_i, x_j)$ for all $m = 1, ..., M$, and h is the heterogeneity function. Note that in the rare case when all $h(x_i, x_j)$ are zero, δ can take an arbitrary positive value as it no longer matters. The normalizing parameter should then replace B_{max}, that is, $f(W = 0, B = \delta) = 1$. Once the value of δ is determined, it should stay constant for all data sets that are going to be compared. There is a variety of ways to determine the parameter. For example, in a recent opinion dynamics study where the data points are all in the range of -1 and $+1$, Schweighofer et al. (2020) use the "maximally possible distance" between two points in the opinion space as the normalizing parameter. This means in a D dimensional Euclidean space, their normalizing parameter is $\sqrt{4D}$.

Finally, we provide a particular form of $P = f(W, B)$ that exhibits the desired properties. It is worth noting that Eq. (7) is by no means the only possible form of $f(W, B)$.

$$f(W, B) = \frac{1}{\delta}\, g\Big(\frac{B}{W + 1}\Big) \tag{7}$$

where g is a continuous and strictly increasing function with $g(0) = 0$ and $g(1) = 1$. It is easy to prove that this form satisfies the property of continuity, dimensionality, monotonicity, maximum, and normalization. It is also obvious that when $W = B = 0$, $f(W, B)$ defined in Eq. (7) is minimized to 0. One may argue that $B = 0$ and $W \neq 0$ can also lead to $P = 0$. However, our definition of groups implies that when $B = 0$, W must also be 0, therefore $W = B = 0$ is a sufficient and necessary minimization condition.

Figure 9 summarizes the procedure of implementing ESBG and constructing polarization measurement based on ESBG. After preprocessing, the raw data are transformed to the "trouble-free" processed data. Subsequently, by defining the heterogeneity measurement,

as well as the expressions of within-group heterogeneity W and between-group heterogeneity B, we divide the processed data into two groups of equal sizes. The grouping result needs to be validated. This concludes the procedure of implementing ESBG. To construct the polarization measurement, besides W and B, we further need to design the expression of $f(W, B)$ and choose an appropriate normalizing parameter δ. By applying the measurement to the groups, we can finally obtain the polarization level of the data.

4.5 Relation with bimodality measurements

The expression given by Eq. (7) resembles a number of bimodality measurements such as Ashman's D (Ashman et al. 1994; Forchheimer et al. 2015) and the bimodal separation (Zhang et al. 2003), whose main ideas lie in the assumption that the data set X is generated or can be described by some bimodal Gaussian mixture. The density of such a mixture is (Ashman et al. 1994):

$$p(X) = \pi_1 p(X, \mu_1, \delta_1^2) + \pi_2 p(X, \mu_2, \delta_2^2) \tag{8}$$

where π_g, μ_g, and δ_g^2 (not to be confused with the normalizing parameter δ) are the fraction, mean, and variance of a Gaussian distribution g ($g = 1, 2$). Given these parameters, Ashman's D is expressed as (Forchheimer et al. 2015):

$$D \propto \frac{|\mu_1 - \mu_2|}{\sqrt{\delta_1^2 + \delta_2^2}} \tag{9}$$

and the bimodal separation is (Zhang et al. 2003):

$$BS \propto \frac{|\mu_1 - \mu_2|}{\delta_1 + \delta_2} \tag{10}$$

The polarization measurement in Eq. (7) and the above-mentioned bimodality measurements both rely on the ratio of between-group heterogeneity to within-group heterogeneity if we consider each Gaussian distribution as a group. To use D and BS, one usually needs to fit two Gaussian distributions to the data set by some technique (e.g., the KMM algorithm (Ashman et al. 1994)), which is in analogy with ESBG. With all these commonalities, it is fair to conclude that the ESBG-based polarization measurements systematically echo the bimodality measurements as they all require a bi-division of the data set and use the heterogeneity between and within the divisions.

These similarities reflect the conceptual closeness between polarization and bimodality. DiMaggio et al. (1996) regarded bimodality as one of the four key dimensions of polarization. Bramson et al. (2017) argued that bimodality takes into account at least three "senses" of polarization, including community fragmentation ("the degree to which the population can be broken into sub populations" (Bramson et al. 2016, 2017)), distinctness between groups, and distance between groups (see Sect. 2.3). Bimodality is also claimed to be an indicator (Knapp 2007) or a feature (Bramson et al. 2017) of polarization. In fact, bimodality (not necessarily D and BS introduced here) has been used as a (partial) measurement of political polarization (e.g., Baldassarri and Bearman 2007; Kim and Baek 2021) and polarization has been used as an interchangeable (yet problematic)[7] term for bimodality (e.g., Hegselmann and Krause 2002).

[7] According to Bramson et al. (2017).

Despite these similarities, these two types of measurements are fundamentally different. Polarization and bimodality are distinct concepts. A bimodal distribution is usually polarized, but a distribution with zero bimodality (e.g., a unimodal distribution) may not be of zero polarization. As pointed out by Fiorina and Abrams (2008), bimodality is a necessary but hardly sufficient condition for large degree of polarization. According to Bramson et al. (2017), bimodality does not implicitly invoke the sense of group size parity, which refers to the idea that a system is more polarized if groups are of equal sizes. This is reflected by the KMM algorithm where the size of each distribution is not relevant. Meanwhile, the ESBG-based measurement, as argued in Sect. 3.3, includes the effect of group size parity by imposing the equal size constraint.

5 An illustrative example: Equal Size Binary Grouping based on K-means clustering and corresponding polarization measurement

In this section, we provide an illustrative example of applying ESBG to a synthetic multi-dimensional data set and then constructing a polarization measurement based on the groups defined by ESBG.

5.1 K-means clustering

The implementation of ESBG in this example will utilize one of the most well-known and widely-used clustering algorithms called **k-means clustering** (Forgy 1965; MacQueen 1967; Xu and Wunsch 2009). Despite its ease of implementation, k-means clustering is an ideal choice for ESBG because the number of groups K needs to be determined a priori. To produce two groups, we simply set $K = 2$, and the only remaining problem is to ensure the sizes of the groups are equal.

In short, k-means clustering attempts to find a number (K) of centroids (sometimes called group/cluster centers), each representing a group containing the data points around the centroid (Müller and Guido 2016). Formally, the algorithm divides the system by minimizing the following distortion function (Bishop 2006):

$$J = \sum_{i=1}^{N} \sum_{k=1}^{K} r_{ik} ||x_i - \mu_k||^2 \tag{11}$$

where $r_{ik} \in \{0, 1\}$ is a binary indicator: $r_{ik} = 1$ if x_i is classified in C_k, and $r_{ik} = 0$ otherwise. The distortion function J in Eq. (11) is the sum of squared distances between each data point x_i and its centroid μ_k. The k-means algorithm chooses the optimal $\{r_{ik}\}$ and $\{\mu_k\}$ to minimize J by using an iterative procedure based on the EM algorithm: given randomly chosen initial conditions, during each iteration, first we fix μ and minimize J with respect to r_{ik} (step E); we then fix μ and minimize J with respect to r_{ik} (step M). The iteration is repeated until convergence (Bishop 2006).

From Eq. (11), we can see the proximity measurement used in k-means clustering is the (squared) Euclidean distance. Meanwhile, the distortion function J is closely related to

within-group heterogeneity (see Sect. 5.2). Therefore, from the view of grouping method, we can say that the k-means clustering algorithm defines groups by minimizing the within-group heterogeneity of the data set.

5.2 ESBG based on k-means clustering

In this subsection, we show how to implement ESBG step by step on the basis of the k-means clustering algorithm.

Preprocessing: We use a synthetic two-dimensional data set X^* containing two blobs of 100 and 200 data points respectively.[8] Assuming none of the dimension-related problems (see Sect. 4.3) exists, no preprocessing is needed for this particular case.

Definition of heterogeneity measurement: Following the k-means clustering algorithm, we use the squared Euclidean distance as the heterogeneity measurement. Within-group heterogeneity W can thus be defined as:

$$W = \frac{1}{N}\sum_{i=1}^{N}\sum_{k=1}^{2} r_{ik}||x_i - \mu_k||^2 = \frac{1}{N}\sum_{k=1}^{2}\sum_{x_i \in C_k}||x_i - \mu_k||^2 \tag{12}$$

which is the average squared Euclidean distance between each data point and its corresponding centroid, that is, $W = J/N$ with a predefined value of $K = 2$. Between-group heterogeneity B, following the same fashion, is defined as the squared distance between the centroids:

$$B = ||\mu_1 - \mu_2||^2 \tag{13}$$

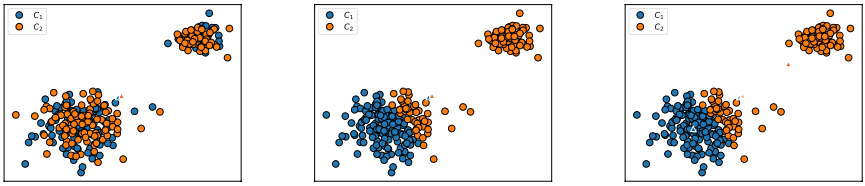
The motivation for choosing J/N instead of J as the measurement of W is to make W and B on the same scale. Otherwise, we could expect $W \gg B$ in most cases, making P extremely small. In addition, if W and B are not on the same scale, for example $W = J$, it will be difficult to defend the second axiom of Esteban and Ray (1994) (see Sect. 3.4).

Grouping: Now the task of implementing ESBG is turned into an optimization problem:

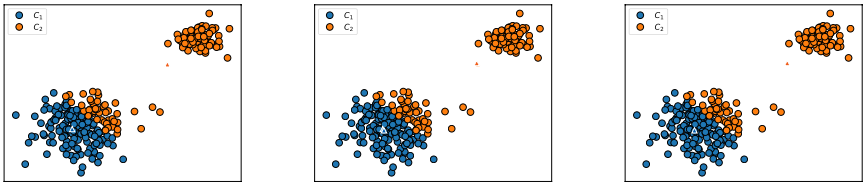
$$\begin{aligned} \min_{\{r_{ik}\},\{\mu_k\}} \quad & W = \frac{1}{N}\sum_{i=1}^{N}\sum_{k=1,2} r_{ik}||x_i - \mu_k||^2 \\ \text{s.t.} \quad & \sum_{i=1}^{N} r_{ik} = N/2, \qquad \forall k = 1, 2 \end{aligned} \tag{14}$$

Following Bishop (2006), we use the EM iteration to solve this optimization problem. The M step is the same as that in k-means, and the E step aims to minimize W with fixed μ_1 and μ_2, while constrained by the condition of equal group sizes. The basic idea is to calculate the squared distance between each data point and both centroids, respectively. For data point x_i, denote the absolute difference between its squared distances to both centroids as Δ_i. First we assign each data point to the closer centroid to generate "temporary" groups. Then, until both groups have the same size, we select a member repeatedly to move it from

[8] The data set X^* is generated using the *make_blobs* function from *sklearn* Python module: X,y=make_blobs(n_samples=[100,200], cluster_std=[.4,.8], centers=[[4,0],[0,-4]]). Then apply the MinMaxScaler from the same module and we obtain X^*.

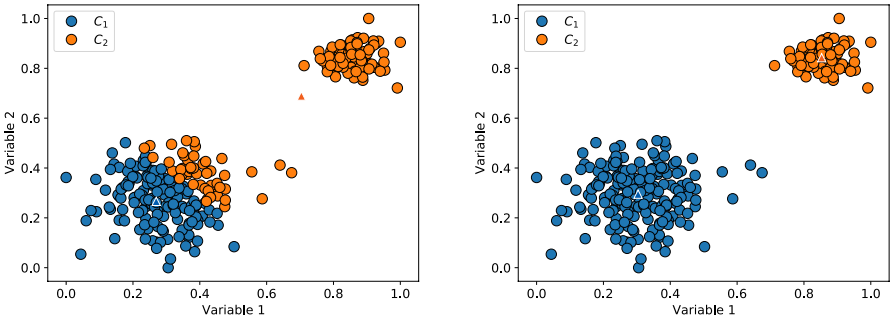


(a) Iteration 0: Initialization (b) Iteration 1: E Step (c) Iteration 1: M Step



(d) Iteration 2: E Step (e) Iteration 2: M Step (f) Iteration 3: E Step

Fig. 10 Illustration of the k-means-based ESBG using X^*. The centroids are shown by the triangles of similar colors of their corresponding groups. The triangles of lighter colors represent the centroids in the previous iteration. (**a**): Initially, the data set is randomly and equally divided into two groups C_1 and C_2, and the centroids of both groups are computed as the average of their group members. (**b**)**:** In the E step of Iteration 1, each data point is assigned to the group whose centroid is nearer, while keeping the size of each group equal. (**c**)**:** In the M step of Iteration 1, the centroid of each group is re-computed according to the new group structure updated in the last E step. (**d**)–(**f**)**:** Successive iterations. The change in the positions of centroids from (**d**) to (**e**) is relatively small and can be observed when taking a closer look. The system has reached convergence since (**f**).



(a) ESBG (b) K-means

Fig. 11 The results of applying (**a**) ESBG or (**b**) K-means clustering ($K = 2$) to X^*. Each centroid is shown by the triangle of the similar colour of its corresponding group. Note: (**a**) is the same as Fig. 10(f).

the larger group to the smaller group. The selected member should be the one with the smallest Δ_i. A similar idea can be found from the elki project.[9] The process is illustrated in Fig. 10. The outcome is two equally sized groups with members distributed around two

[9] https://elki-project.github.io/tutorial/same-size_k_means.

Table 2 Measuring polarization in the system described in Fig. 2

Pop at 1	Pop at 5	Pop at 11	W/W^m	B/B^m	f^*/f^{*m}
9	9	2	3.6000/5.1200	23.0400/2.5600	0.0501/0.0042
8	8	4	5.6000/7.6800	31.3600/10.2400	0.0475/0.0118
7	7	6	6.0000/7.6800	40.9600/23.0400	0.0585/0.0265
6	6	8	4.8000/5.1200	51.8400/40.9600	0.0894/0.0669
5	5	10	2.0000/0.0000	64.0000/64.0000	0.2133/0.6400

centroids (Fig. 11(a)). To make a comparison, we apply k-means clustering to the same data set in Fig. 11(b).

Validation: We first check if the outcome of the grouping step contains two groups, and if their sizes are the same. Then, we check whether the outcome is optimal, in other words, if W is minimized. A primary step may include checking if swapping memberships of data points can decrease W, and if each centroid is the mean of its members. In addition, a number of validation methods, criteria, and indices are available to formally justify the clustering result (Xu and Wunsch 2009).

5.3 Corresponding polarization measurement

According to Eq. (7), we choose $f^*(W,B) = \frac{1}{\delta}(\frac{B}{W+1})$ as our polarization measurement (i.e., we choose $g(x) = x$). Given this expression, we have $W = 0.05784060054$, and $B = 0.367362082$ for our synthetic data set X^*. Setting $\delta = 2$ (given that the maximum possible squared Euclidean distance in X^* is smaller than $\sqrt{2}$), the polarization level of X^* is then $f^* = 0.173637731$.

As suggested in Sect. 3.2.3, we use f^* to examine whether ESBG can solve the problem of G_1. In Table 2, we have summarized the within-group heterogeneity before (W) and after (W^m) the merge, the between-group heterogeneity before (B) and after (B^m) the merge, and the polarization level before ($f^* = f^*(W,B)$) and after ($f^{*m} = f^*(W^m,B^m)$) the merge, under different initial population distributions at 1, 5, and 11. By setting $\delta = 100$, we can see that as long as the population at 11 is no larger than 8 (recall that in the original example, it is required that the population at 11 is sufficiently small), the merge will reduce the polarization level due to the significant decrease in B and relatively small increase in W.

6 Relation with bipolarization measurements

As argued in Sect. 2, there are two notable lines of polarization measurements: the Wolfson's line (i.e. bipolarization measurement), which captures the decline of the middle class, and the Esteban & Ray's line, which focuses on how individuals are clustered in groups. It is clear that our ESBG-based measurement is in the Esteban & Ray's line as its derivation relies on the concepts, axioms, and properties proposed by Esteban and

Ray (1994). In this section, we will show that our measurement can be partly viewed as a (multidimensional) polarization measurement in the Wolfson's line.

6.1 Increased spread and increased bipolarity

The construction of a bipolarization measurement relies on two critical properties: **increased spread** and **increased bipolarity** (Wang and Tsui 2000; Chakravarty and Majumder 2001; Gigliarano and Mosler 2009). Increased spread states that given the median level fixed, polarization increases when any individual moves in the opposite direction from the median level (Wang and Tsui 2000), and increased bipolarity states that after a Pigou–Dalton transfer within the same group, polarization level should increase (Wang and Tsui 2000; Gigliarano and Mosler 2009). A Pigou-Dalton transfer is defined as a transfer from a rich individual to a poor individual, and after the transfer, the poor should not be richer than the rich before the transfer and the rich should not be poorer than the poor before the transfer (Wang and Tsui 2000).

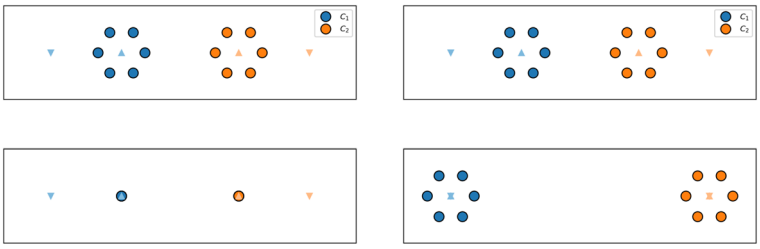
To see its relation with bipolarization measurements, we need to check if our ESBG-based measurement satisfies increased spread and increased bipolarity. From the definition of Pigou-Dalton transfer, it follows that W will be reduced after a transfer. However, estimating the effect of a Pigou-Dalton transfer on B without knowing the exact expression of B is not easy. For B defined in Eq. (13), a Pigou-Dalton transfer has no impact on it as the locations of the centroids are not affected. Therefore, at least the k-means-based polarization measurement proposed in Sect. 5.3 satisfies increased bipolarity.

Whether an ESBG-based measurement satisfies increased spread is a more complicated question. A data point's moving away from the median value (hereafter referred to as *increased-spread-move*) will definitely increase B. Meanwhile, depending on the location of the data point, the move may either increase or decrease W. Therefore, we do not know if P goes up or not. Some counter-examples can be found. For f^* given in Sect. 5.3, if we move the leftmost data point in X^* whose Variable 1 equals 0 to a more left location where Variable 1 is -10, by setting $\delta = 101$, the polarization measurement of the system drops from 0.003438371 to 0.003024571, mainly due to the significant increase in W. Although we are not sure if other expressions of $f(W, B)$ would satisfy increased spread, we could claim that this property is not generally desired by ESBG-based measurements.

6.2 Is the ESBG-based measurement a bipolarization measurement?

Even if our measurement may not satisfy increased spread, one cannot deny that it is similar to a bipolarization measurement in many aspects. First, ESBG itself is the same as the grouping method of a bipolarization measurement when $D = 1$ (see Sect. 4.1). This finding is interesting: we were looking for an appropriate grouping method for the Esteban & Ray's line, but after exploration, we end up with a grouping method similar to the one used in the Wolfson's line.

Secondly, the Wolfson's index (Wolfson 1994)—the representative of the Wolfson's line—can also be written in the form of a function of W and B. The index is originally written in the following form (Wolfson 1994; Wang and Tsui 2000):



(a) Squeezing Step (b) Moving Step

Fig. 12 Illustration of (**a**) the squeezing step and (**b**) the moving step. In each sub-figure, the configuration at the top will transfer to the configuration at the bottom after the step. The up-pointing triangles represent group centers (centroids) and the down-pointing triangles represent extremes

$$P^W = 2\frac{(2T - Gini)}{(m/\mu)} \tag{15}$$

where $T = 0.5 - L(0.5)$, and $L(0.5)$ is the share of variable of interest of the lower half of the population. *Gini* refers to the Gini index of the whole population, m is the median value, and μ is the mean value (Wolfson 1994). According to Gigliarano and Mosler (2009), the Wolfson's index can also be written as:

$$P^W = \frac{2\mu}{m}(B - W) \tag{16}$$

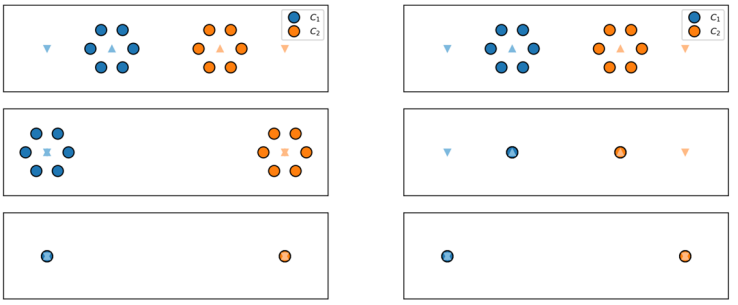
where W and B are represented by the Gini index between and within groups respectively (Gigliarano and Mosler 2009). In this sense, both the ESBG-based measurement and the Wolfson's index are in the form of $P = f(W, B)$ (note that the Wolfson's index also depends on μ and m), and both of them are increasing with B and decreasing with W.

Finally, as shown in Sect. 6.1, an ESBG-based measurement—at least a particular form of it—satisfies increased bipolarity, one of the two basic properties of bipolarization measurements.

Since the ESBG-based measurement is not expected to satisfy increased spread, it should not be considered as a bipolarization measurement. Given the similarities between them, we can roughly view the EBSG-based measurement as a (multi-dimensional) bipolarization measurement without the property of increased spread.

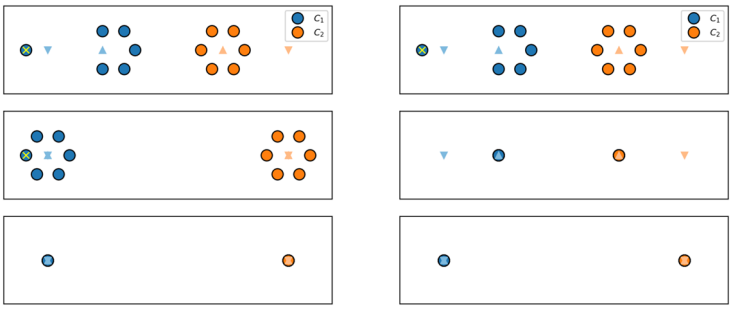
6.3 Squeezing-and-moving framework

Polarization is a slippery, context-dependent concept whenever applied to social systems. Although we could understand in principle what a maximum or a minimum polarization is, the whole range of in-between states remains poorly understood. This explains why polarization is often described as the distance to the situation of maximum polarization. For instance, Flache and Mäs (2008) stated that "polarization captures the degree to which the group can be separated into a small set of factions who are mutually antagonistic in the opinion space and have maximal internal agreement". Indeed, the general interest in polarization, whether from the public or from scholars, mainly comes from the fear of its



(a) Moving-Squeezing Procedure (b) Squeezing-Moving Procedure

Fig. 13 Illustration of (**a**) the moving-squeezing procedure and (**b**) the squeezing-moving procedure. In each sub-figure, the configuration at the top will transfer to the configuration at the bottom via the intermediate configuration in the middle. The up-pointing triangles represent group centers (centroids) and the down-pointing triangles represent extremes



(a) Moving-Squeezing Procedure (b) Squeezing-Moving Procedure

Fig. 14 Illustration of the increased-spread-move in (**a**) the moving-squeezing procedure and (**b**) the squeezing-moving procedure. The yellow saltire marks the data point that was moved here from the configuration at the top of Fig. 13 (whether (**a**) or (**b**)) by an increased-spread-move. In each sub-figure, the configuration at the top will transfer to the configuration at the bottom via the intermediate configuration in the middle. The up-pointing triangles represent group centers (centroids) and the down-pointing triangles represent extremes

destructive effect on social harmony and stability (e.g., Layman and Carsey 2002; Montalvo and Reynal-Querol 2005; Fisher and Mattson 2009). Such a fear-based interest naturally leads us to consider more carefully "how far are we from the most polarized situation?" rather than "what on earth is polarization?".

At a first glance, ESBG seems too simple to be correct. However, if we interpret polarization level as a measurement of "how far we are from the most polarized situation", it becomes clear why ESBG works. Consider that we want to transform a not-very-polarized data set into the maximum polarized situation. Therefore, the priority is to identify which data point should be relocated to which extreme. This is exactly what ESBG does.

To achieve maximum polarization, data points in each group should be later relocated to the nearer extreme. This task can be done via the following two steps: the **squeezing step** that "squeezes" the data points in the same group to the group center (Fig. 12(a)), and the **moving step** that moves[10] each group to its corresponding extreme (Fig. 12(b)). Given a group structure, within-group heterogeneity W measures how difficult the squeezing step is, and between-group heterogeneity B measures how easy the moving step is. This also explains why polarization should increase with B and decrease with W, if we consider polarization measurement as an index of the overall difficulty of achieving maximum polarization.

The concepts of squeezing and moving can help us to understand why the ESBG-based measurement and bipolarization measurement both satisfy increased bipolarity, but only the latter satisfies increased spread. A Pigou-Dalton transfer, by definition, will make the squeezing process easier (i.e. reducing W) without affecting the moving process (at least for f^* as the centroids are not affected by the transfer). Therefore, it facilitates the task and hence increases polarization. Therefore, both types of measurements satisfy increased spread.

When considering increased spread, the picture is different. If the moving step is executed before the squeezing step (i.e., the moving-squeezing procedure, see Fig. 13(a)), an increased-spread-move makes the moving process easier (i.e., increasing B) without affecting the squeezing process (see Fig. 14(a)). It will therefore increase polarization. However, if the moving step is executed after the squeezing step (i.e., squeezing-moving procedure, see Fig. 13(b)), an increased-spread-move makes the squeezing process more difficult (i.e., increasing W), while (maybe slightly) facilitating the moving process (i.e., increasing B) because the relevant centroid will be closer to its extreme due to the move (see Fig. 14(b)). This implies that we cannot determine if the move will decrease polarization or not without knowing the exact expression of $f(W, B)$. From a result-oriented point of view, we can then conceptualize the bipolarization measurement as a realization of the moving-squeezing procedure, and the ESBG-based measurement as a realization of the squeezing-moving procedure, explaining why our measurement satisfies increased bipolarity but not increased spread.

7 Conclusion

In the vast literature on polarization, the notion of group, especially groups based on similarities between individuals, is the elephant in the room: everyone considers groups when defining or conceptualizing polarization, but it is difficult to understand what exactly such groups are. The only recurrent argument is that members of the same group should be similar, whereas members from different groups should be dissimilar. This is neither sufficient to capture the nuances of the various group structures, which are caused by various social cleavages that characterize our complex societies, nor it contributes to a consistent measurement of polarization. The mismatch between how we understand and how we measure polarization undermines the reliability of measurements, thus hampering our understanding of society in its complex and multifaceted aspects.

[10] Strictly speaking, the "move" is a space translation that moves every point in the same group by the same distance and in the same direction.

In this study, we have proposed a grouping method for constructing polarization measurements called "Equal Size Binary Grouping" (ESBG) that divides a data set into two groups of equal sizes according to similarities between data points. We showed that ESBG can be a suitable solution to certain theoretical and practical problems that trouble other grouping methods, such as discontinuity and contradiction of reasoning. While alternative approaches exist that over-impose pre-existing group structures or explore various dimensions of polarization, we believe that significant advances in polarization studies in complex societies can be made if measurements are consistent and possibly capable of discovering endogenous structures from data that are coherent with the variable(s) of interest.

Following clustering algorithms, we presented a procedure containing four steps to implement ESBG. Based on ESBG, a novel class of polarization measurements can be constructed to measure both uni- and multi-dimensional polarization. The measurements increase with between-group heterogeneity and decrease with within-group heterogeneity, and are not affected by other variables such as the number or size of groups. We also showed that the measurements satisfy a range of properties that have long been deemed desired in the field, such as continuity, normalization, maximization and minimization. Subsequently an illustrative example of applying ESBG and the related measurement to a synthetic data set was demonstrated.

As a final remark, we investigated the relation between the ESBG-based measurement and bipolarization measurement. The ESBG-based measurement can be roughly viewed as a multidimensional bipolarization measurement without the property of increased spread. This is because both types of measurements use the same grouping method when $D = 1$, and satisfy the same property of increased bipolarity. Furthermore, we developed a so-called "squeezing-and-moving" framework to help explain the relation between them.

With all due caveats due to our general approach and the lack of appropriate data on which to test these measurements, we believe that future research will help to improve the design of the measurement, while contributing to the debate on the key role of group definition in current measurements. Although useful to explore group structures within data starting from the variable(s) of interest, our method drastically simplifies the possible variety of groups co-existing in the same society, due to the varying cleavages that characterize the complex fabric of our social systems. However, we hope that our measurement could also stimulate new empirical research on polarization that improves comparability, replicability, and cumulativeness. As an avenue for future research, we suggest comparing the ESBG-based measurement with existing polarization measurements in the context of various attribute distributions such as distributions with two or more peaks.

Funding This study has received funding from the European Research Council: Consolidator GrantBEHAVE (Grant Agreement No. 724431).

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

Ethical statement This study does not involve Human Participants and/or Animals. All authors have agreed with the manuscript and approved the submission.

References

Aaberge, R., Brandolini, A.: Multidimensional poverty and inequality. In: Atkinson, A.B., Bourguignon, F. (eds.) Handbook of Income Distribution, vol. 2, pp. 141–216. Elsevier, New York (2015)
Abramowitz, A.I., Saunders, K.L.: Is polarization a myth? J. Politics **70**(2), 542–555 (2008)
Aggarwal, C.C.: An introduction to cluster analysis. In: Aggarwal, C.C., Reddy, C.K. (eds.) Data Clustering: Algorithms and Applications, pp. 1–28. CRC Press, London (2014)
Aleskerov, F., Oleynik, V.: Multidimensional polarization index and its application to an analysis of the Russian state Duma. arXiv:1608.01351 (2016)
Alichi, A., Kantenga, M. K., Sole, M. J.: Income Polarization in the United States. International Monetary Fund (2016)
Anderson, G.: Polarization measurement and inference in many dimensions when subgroups can not be identified. Economics: The Open-Access, Open-Assessment E-Journal 5 (2011)
Ashman, K.M., Bird, C.M., Zepf, S.E.: Detecting bimodality in astronomical datasets. Astron. J. **108**, 2348–2361 (1994)
Baldassarri, D., Bearman, P.: Dynamics of political polarization. Am. Sociol. Rev. **72**(5), 784–811 (2007)
Baraldi, A., Blonda, P.: A survey of fuzzy clustering algorithms for pattern recognition. I. IEEE Trans. Syst. Man Cybern. Part B (Cybern.) **29**(6), 778–785 (1999a)
Baraldi, A., Blonda, P.: A survey of fuzzy clustering algorithms for pattern recognition. II. IEEE Trans. Syst. Man Cybern. Part B (Cybern.) **29**(6), 786–801 (1999b)
Bauer, P.C.: Conceptualizing and Measuring Polarization: A Review. Working paper (2019)
Bishop, C.M.: Pattern Recognition and Machine Learning. Springer, Berlin (2006)
Boxell, L., Gentzkow, M., Shapiro, J.M.: Greater Internet use is not associated with faster growth in political polarization among US demographic groups. Proc. Natl. Acad. Sci. **114**(40), 10612–10617 (2017)
Bramson, A., Grim, P., Singer, D.J., Fisher, S., Berger, W., Sack, G., Flocken, C.: Disambiguation of social polarization concepts and measures. J. Math. Sociol. **40**(2), 80–111 (2016)
Bramson, A., Grim, P., Singer, D.J., Berger, W.J., Sack, G., Fisher, S., Sack, G., Holman, B.: Understanding polarization: meanings, measures, and model evaluation. Philos. Sci. **84**(1), 115–159 (2017)
Chakravarty, S.R., Majumder, A.: Inequality, polarisation and welfare: theory and applications. Aust. Econ. Pap. **40**(1), 1–13 (2001)
Danzell, O.E., Yeh, Y.Y., Pfannenstiel, M.: Determinants of domestic terrorism: an examination of ethnic polarization and economic development. Terror. Polit. Violence **31**(3), 536–558 (2019)
Deutsch, J., Fusco, A., Silber, J.: The BIP trilogy (bipolarization, inequality and polarization): one saga but three different stories. Econ. Open Access Open Assess. E-J. **7**(2013–22), 1–33 (2013)
DiMaggio, P., Evans, J., Bryson, B.: Have Americans social attitudes become more polarized? Am. J. Sociol. **102**(3), 690–755 (1996)
Duclos, J.Y., Taptué, A.M.: Polarization. In: Atkinson, A.B., Bourguignon, F. (eds.) Handbook of Income Distribution, vol. 2, pp. 301–358. Elsevier, London (2015)
Duclos, J.Y., Esteban, J., Ray, D.: Polarization: concepts, measurement, estimation. Econom. J. Econ. Soc. **72**(6), 1737–1772 (2004)
Esteban, J.M., Ray, D.: On the measurement of polarization. Econom. J. Econ. Soc. **62**(4), 819–851 (1994)
Esteban, J., Ray, D.: Comparing polarization measures. In: Garfinkel, M.R., Skaperdas, S. (eds.) Oxford Handbook of Economics of Peace and Conflict, pp. 127–51. Oxford University Press, Oxford (2012)
Esteban, J., Schneider, G.: Polarization and conflict: theoretical and empirical issues. J. Peace Res. **45**(2), 131–141 (2008)
Fiorina, M.P., Abrams, S.J.: Political polarization in the American public. Annu. Rev. Polit. Sci. **11**, 563–588 (2008)
Fisher, C.S., Mattson, G.: Is America fragmenting? Ann. Rev. Sociol. **35**, 435–455 (2009)
Flache, A., Macy, M.W.: Small worlds and cultural polarization. J. Math. Sociol. **35**(1–3), 146–176 (2011)
Flache, A., Mäs, M.: How to get the timing right. A computational model of the effects of the timing of contacts on team cohesion in demographically diverse teams. Comput. Math. Organ. Theory **14**(1), 23–51 (2008)

Flache, A., Mäs, M., Feliciani, T., Chattoe-Brown, E., Deffuant, G., Huet, S., Lorenz, J.: Models of social influence: towards the next frontiers. J. Artif. Soc. Soc. Simul. **20**(4), 2 (2017)
Forchheimer, D., Forchheimer, R., Haviland, D.B.: Improving image contrast and material discrimination with nonlinear response in bimodal atomic force microscopy. Nat. Commun. **6**(1), 1–5 (2015)
Forgy, E.: Cluster analysis of multivariate data: efficiency vs. interpretability of classifications. Biometrics **21**, 768–780 (1965)
Foster, J.E., Wolfson, M.C.: Polarization and the Decline of the Middle Class: Canada and the US. J. Econ. Inequal. **8**(2), 247–273 (2010)
Fusco, A., Silber, J.: On social polarization and ordinal variables: the case of self-assessed health. Eur. J. Health Econ. **15**(8), 841–851 (2014)
Gigliarano, C., Mosler, K.: Constructing indices of multivariate polarization. J. Econ. Inequal. **7**(4), 435–460 (2009)
Hare, C., Poole, K.T.: The polarization of contemporary American politics. Polity **46**(3), 411–429 (2014)
Hart, J.: Symmetry and polarization in the European international system, 1870–1879: a methodological study. J. Peace Res. **11**(3), 229–244 (1974)
Hegselmann, R., Krause, U.: Opinion dynamics and bounded confidence models, analysis, and simulation. J. Artif. Soc. Soc. Simul. **5**, 3 (2002)
Homan, A.C., Van Knippenberg, D., Van Kleef, G.A., De Dreu, C.K.: Interacting dimensions of diversity: cross-categorization and the functioning of diverse work groups. Group Dyn. Theory Res. Pract. **11**(2), 79 (2007)
Jain, A.K., Dubes, R.C.: Algorithms for Clustering Data. Prentice-Hall Inc, London (1988)
Jain, A.K., Murty, M.N., Flynn, P.J.: Data clustering: a review. ACM Comput. Surv. (CSUR) **31**(3), 264–323 (1999)
Kim, J., & Baek, S. K.: Democracy and political polarization in the National Assembly of Republic of Korea. arXiv:2101.03490 (2021)
Knapp, T.R.: Bimodality revisited. J. Mod. Appl. Stat. Methods **6**(1), 3 (2007)
Layman, G.C., Carsey, T.M.: Party Polarization and "Conflict Extension" in the American Electorate. American Journal of Political Science **46**(4), 786–802 (2002)
MacQueen, J.: Some methods for classification and analysis of multivariate observations. In Proceedings of the Fifth Berkeley Symposium **1**, 281–297 (1967)
Mäs, M., Flache, A., Takács, K., Jehn, K.A.: In the short term we divide, in the long term we unite: demographic crisscrossing and the effects of faultlines on subgroup polarization. Organ. Sci. **24**(3), 716–736 (2013)
McCright, A.M., Dunlap, R.E.: The politicization of climate change and polarization in the American publics views of global warming, 2001–2010. Sociol. Q. **52**(2), 155–194 (2011)
Montalvo, J.G., Reynal-Querol, M.: Ethnic polarization, potential conflict, and civil wars. Am. Econ. Rev. **95**(3), 796–816 (2005)
Müller, A.C., Guido, S.: Introduction to Machine Learning with Python: A Guide for Data Scientists. O Reilly Media, Newton (2016)
Phillips, K.W.: The effects of categorically based expectations on minority influence: the importance of congruence. Pers. Soc. Psychol. Bull. **29**(1), 3–13 (2003)
Ross, E.A.: The Principles of Sociology. Century Company, New York (1920)
Schweighofer, S., Schweitzer, F., Garcia, D.: A weighted balance model of opinion hyperpolarization. J. Artif. Soc. Soc. Simul. **23**(3), 5 (2019)
Takács, K., Bravo, G., Squazzoni, F.: Referrals and information flow in networks increase discrimination: a laboratory experiment. Soc. Netw. **54**, 254–265 (2018)
Wang, Y.Q., Tsui, K.Y.: Polarization orderings and new classes of polarization indices. J. Pub. Econ. Theory **2**(3), 349–363 (2000)
Wolfson, M.C.: When inequalities diverge. Am. Econ. Rev. **84**(2), 353–358 (1994)
Xu, R., Wunsch, D.: Clustering. Computational Intelligence Series, Wiley-IEEE Press, New York (2009)
Zhang, X., Kanbur, R.: What difference do polarisation measures make? An application to China. J. Dev. Stud. **37**(3), 85–98 (2001)
Zhang, C., Mapes, B.E., Soden, B.J.: Bimodality in tropical water vapour. Q. J. R. Meteorol. Soc. **129**(594), 2847–2866 (2003)