UNIVERSITAT DE BARCELONA

FUNDAMENTAL PRINCIPLES OF DATA SCIENCE MASTER'S THESIS

A study of polarisaton in bimodal social networks

Author:
Adrián FERNÁNDEZ CID

Supervisors:
Dr. Emanuele COZZO
Dr. Oriol PUJOL VILA

A thesis submitted in partial fulfillment of the requirements for the degree of MSc in Fundamental Principles of Data Science

in the

Facultat de Matemàtiques i Informàtica

January 14, 2022

UNIVERSITAT DE BARCELONA

Abstract

Facultat de Matemàtiques i Informàtica

MSc

A study of polarisaton in bimodal social networks

by Adrián FERNÁNDEZ CID

Social polarisation is a central issue in the social sciences, and it has acquired mainstream interest in recent years. A prominent area of current research in computational social science studies the polarisation of social systems in terms of features of their graph representation. Such structural polarisation measures can capture well-grounded aspects of polarisation at a comparatively lower cost than content-based or distributional approaches, although some of them have been shown to depend on unrelated network properties like average degree or systematically give false positives on randomised networks. In this master's thesis, I explore a novel approach that implements an axiomatic polarisation measure with hierarchical clustering on bimodal networks, which are less studied in the literature. The clustering implements the well known Ward and centroid methods, as well as a new one, *poldist*, inspired by the polarisation measure used. In the validation use case, on the standard Southern Women dataset, results reasonably agree with the expected separation in two communities for the Ward and centroid methods, but not for poldist. On the other hand, the application use case, on data from the platform of the Conference on the Future of Europe, shows no significant dipoles neither in the topic-specific nor the global analysis, which (given the previous pipeline validation and the relatively low participation of the platform) points to a lack of polarisation in the data. However, further analysis on such data in terms of multipole partitions is underway and may yet reveal some structure. Current results show the proposed pipeline is a promising candidate for the study of polarisation in bimodal social networks and should be further explored.

Acknowledgements

I wish to thank my supervisors for their support in my wanting to find such a fulfilling project as this one to join, as well as for their understanding through the several, inconvenient postponements imposed by my health condition.

I am also deeply grateful to my family, whose love and support nourish my every aspiration.

Chapter 1

Introduction

1.1 Motivation

Social polarisation, understood as the division of individuals into coherent and strongly opposed groups based on a given attribute (such as income or opinion on a certain issue) (Fiorina and Abrams, 2008; DiMaggio, Evans, and Bryson, 1996), has long been a central topic in the social sciences (Baldassarri and Gelman, 2008; Fiorina and Abrams, 2008), recently receiving pronounced mainstream attention following the Brexit referendum in the UK and the USA presidential election of the same year, 2016.

Polarisation in social systems has been associated with undesirable features such as increased divisiveness and animosity (Mason, 2015), policy gridlock (Jones, 2001), and decreased political representation (Baldassarri and Gelman, 2008). Furthermore, it is believed to hinder resolution of such pressing issues as climate change (Zhou, 2016), immigration and race relations (Hout and Maggio, 2020), and the COVID-19 pandemic (Makridis and Rothwell, 2020).

One crucial aspect to the study of polarisation is its measurement, which has traditionally relied on distributional properties (such as bimodality or dispersion) of survey data (DiMaggio, Evans, and Bryson, 1996). More recently, the wealth and public availability of digital data on social systems has spurred the development of computational approaches (Garimella et al., 2018), among which one can distinguish two main areas: *content-based* analysis, that leverage natural language processing techniques to identify conflicting groups in a system (e.g. Belcastro et al., 2020; Demszky et al., 2019); and *structural* approaches, focusing on inferring polarisation from features of the network representation of the system (Salloum, Chen, and Kivelä, 2021). The present work focuses on the latter.

1.2 A network science approach

Structural polarisation measures are designed to identify what would be observable features of a polarised system. Their main interest lies in their ability to capture such theoretically-grounded features at a lower cost than content and survey-based approaches, which explains their widespread application in computational social science (Salloum, Chen, and Kivelä, 2021).

The typical procedure is 1) construct a graph representation of the system; 2) determine a partition of such a graph in terms of groups, or clusters, that ideally (usually) feature intra-cluster similarity as well as inter-cluster dissimilarity; and 3) compute the polarisation of the partitioned graph.

The *graph representation* of the system is determined by the definition of nodes and links: for instance, in a set of Twitter users one may take individual users as nodes and retweets as links, or take subsets of similar users as nodes, or take "follows" or "likes" as links. Furthermore, both nodes and links may be filtered by a measure of significance, or comprise more than one class (e.g. one can consider popular users as a special kind of node, yielding what is known as a bimodal graph: a graph with two different types of nodes), and links may be binary (either there is or there is not a link between two given nodes) or weighted. As one might expect, the finer the representation, the more complex (and difficult to study) the graph becomes: in particular, a greater proportion of the literature on structural polarisation measurement focuses on unimodal networks. Part of the purpose of the present study is thus to contribute to the body of research on bimodal-network polarisation.

The *clustering* step, also known as the community-detection problem in network science, is a current subject of active research. The optimal partition depends on what one defines as a community. The main difficulty of this step is that it is known to be an ill-posed problem (Fortunato and Hric, 2016), and current algorithms can find a partition even in random networks. This compromises the later computation of the polarisation, usually inflating it, in particular for sparse (i.e. low-connectivity) networks (Bagrow, 2012; Zhang and Moore, 2014; Lancichinetti, Radicchi, and Ramasco, 2010; Guimera, Sales-Pardo, and Amaral, 2004).

Finally, *polarisation measures* vary in the network features they account for. Some compare the density of in-group links to that of external links (Krackhardt and Stern, 1988; Chen et al., 2020). Others focus on the structure of groups and their interactions, e.g. by evaluating the difference in the edge betweenness centrality of external and internal links (Garimella et al., 2018). A third kind of methods rely on simulations, determining how likely a random walker is to remain in a given cluster (Garimella et al., 2018; Rabab'ah et al., 2016; Rumshisky et al., 2017; Darwish, 2019). Other approaches include boundary-based (Guerra et al., 2013) or label-propagation methods (Morales et al., 2015).

A recent review of 8 state-of-the-art structural polarisation measures (Salloum, Chen, and Kivelä, 2021) revealed the challenge with them is to avoid dependence on ostensibly unrelated network features such as average degree or degree distribution, which complicate comparison between networks. Moreover, the authors showed that all studied measures failed to give vanishing polarisation values for randomised networks. Such results call for the exploration of new measures, which is the main contribution of this master's thesis.

1.3 This master's thesis

In this project, I study *bimodal social networks*, namely the Southern Women dataset (Davis, Gardner, and Gardner, 1941) and several networks extracted from the Conference on the Future of Europe platform[1] [2] (Parliament, Council, and Commission, 2021). I will often refer to them as the SW dataset and the CFE data, respectively.

The first, intended for a validation phase of the pipeline proposed here, is a well known dataset used as a standard for community detection in bimodal networks: it consists of two classes of nodes representing women and social events, and the links encode which women attended which event. The second includes 33 previously

[1]See https://futureu.europa.eu/?locale=en.

[2]Originally, the object of study was envisaged as a bimodal network representation of a Twitter community consisting of crowd-sourced elite users (also called *influencers* in regular language) and their audiences (regular users). However, the identification of such crowd-sourced elites had already been done for a previous project, and we ultimately deemed it more appropriate to choose for the object of a master's thesis data that I would have to study from 0.

unstudied datasets obtained from the mass deliberation platform of the Conference for the purpose of this master's thesis, where the nodes represent policy proposals (by any European citizen or group) and users of the online platform, a link meaning that a given user endorsed a given proposal.

The partition of the graph is provided by the last stage of a so-called sequential, agglomerative, hierarchic, nonoverlapping (SAHN) clustering algorithm (Müllner, 2011), which gradually builds up clusters from the initial data by merging the two closest (according to a dissimilarity measure) clusters at a time. The last stage of such a process counts only two clusters. Such algorithms allow for the exploration of different dissimilarity measures, of which we consider three: the well-known Ward and centroid distances, and the novel measure we termed *poldist*, based on the measure used for evaluating polarisation (Esteban and Ray, 1994).

Finally, the polarisation measure stems from the axiomatic proposal of Esteban and Ray, 1994. The main advantage of such a measure is that its axioms guarantee *a priori* a reasonable behaviour for a polarisation measure.

The goals of this master's thesis were:

1. Understand the current polarisation measurement paradigm, namely the network approach, and leverage the data science skills acquired in the master programme in an implementation thereof.

2. Implement and study the polarisation measure proposed by Esteban and Ray, 1994 in a network-based pipeline.

3. Study a novel distance update scheme for hierarchical clustering, poldist, based on the above polarisation measure.

4. Contribute to the body of research on polarisation evaluation in bimodal social networks.

All the above objectives were met, with the following results:

1. Aside from the study of the state of the art (reflected above and in Section 2), the implementation on the use cases brought insight on the network approach, namely on the interplay between the clustering and the computation of the polarisation: the study of the CFE data stresses the importance of choosing the step of the hierarchical clustering from which to draw the partition to evaluate, as in many cases the centroid method showed relevant structure only before the final step. Among the data science skills applied, aside from all the machinery more specific to network science, the Jaccard distance and word shifts (standard tools in natural language processing) were instrumental for the evaluation of results.

2. The implementation of the measure proposed by Esteban and Ray, 1994 is validated by the reference SW dataset for the Ward and centroid methods, and it features a behaviour consistent with what is expected of a polarisation measure, as described by the authors.

3. The method poldist shows remarkable deviation from the expected results on the validation dataset. Results suggest that poldist may benefit from a factor that moderates its strong tendency to favour merging small clusters.

4. The study of the CFE data (Chapter 5) failed to find any significant polarisation, which may be due to the low participation of the platform. However, results suggest that further analysis on higher-participation extensions of the data and/or evaluating partitions at stages of the clustering other than the final one might still reveal some structure.

The rest of the thesis is organised as follows: Chapter 2 introduces the theoretical basis of the project, namely the polarisation measure, the clustering method and some analytic tools used (Sections 2.1 to 2.3); Chapter 3 presents the current proposal; Chapter 4 discusses the validation use case, on the SW dataset; Chapter 5 does so for the application use case, on CFE data; and conclusions and possibilities for further work are presented in Chapter 6.

Throughout this thesis, the words *graph* and *network* are used indistinctly, as well as the pairs *group-cluster* and *dendogram-tree*. When speaking of hierarchical clustering, I often use the term *distance* to refer to a dissimilarity index: I am aware that the latter is not a distance insofar as it might be 0 for different elements and it does not necessarily satisfy the Triangle Inequality, but such (ab)use of language seems typical in the literature (see for instance (Müllner, 2011)).

The full code of the project, extensively commented and including additional checks and plots, is available at a public Github repository[3].

[3] See https://github.com/adrifcid/polarisation.

Chapter 2

Background

2.1 An axiomatic polarisation measure

As discussed in Chapter 1, structural polarisation measures have the advantage of providing an automatic means of polarisation evaluation for any system representable as a network, but typically bear one or more shortcomings among suboptimal clustering, dependence on unrelated network features or false positives (Salloum, Chen, and Kivelä, 2021). The authors also point out that the fact that most current structural measures require a partition formed by only 2 clusters constitutes a further limitation.

Although Salloum, Chen, and Kivelä, 2021 do improve the performance of the measures they study by introducing a normalisation to extract the contribution of random features, they finish by calling to the development of novel approaches, possibly fundamentally different from the ones they study. Such is the main objective of this master's thesis.

Esteban and Ray, 1994 propose the following polarisation measure:

$$P = K \sum_{i=1}^{N_c} \sum_{j=1}^{N_c} n_i^{1+\alpha} n_j d(i,j) \tag{2.1}$$

where K is a normalisation constant, $\alpha \in (0, 1.6]$ is the *polarisation sensitivity* parameter, n_i is the size of cluster i, N_c the total number of clusters and $d(i,j)$ is the distance between i and j. The role of α is to enforce identification within a given group, while distance to other groups d accounts for inter-group alienation. If $\alpha = 0$, (2.1) becomes the Gini coefficient, a standard measure of inequality (related, but not identical, to polarisation).

The measure (2.1) is defined on a one-dimensional distribution of a given attribute (income, opinion, etc.), and it is constructed by imposing the following three intuitive axioms, illustrated in Fig. 2.1:

- *Axiom 1.* Take $p, q \gg 0$, $p > q$, $0 < x < y$. There exists $\epsilon > 0$ and $\mu > 0$ (possibly depending on p and x) such that if $\delta(x,y) < \epsilon$ and $q < \mu p$, then the joining of the two clusters at their mid-point, $(x+y)/2$, increases polarisation.

- *Axiom 2.* Take $(p, q, r) \gg 0$, $p > r$ and $x > d(x,y)$. There is $\epsilon > 0$ such that if the cluster q is moved towards r by an amount not exceeding ϵ, polarisation increases.

- *Axiom 3.* Take $p, q \gg 0$ and $x = d(x,y) \equiv d$. Any new distribution formed by shifting population mass from the central cluster q equally to the two lateral ones, each d units of distance away, increases polarisation.

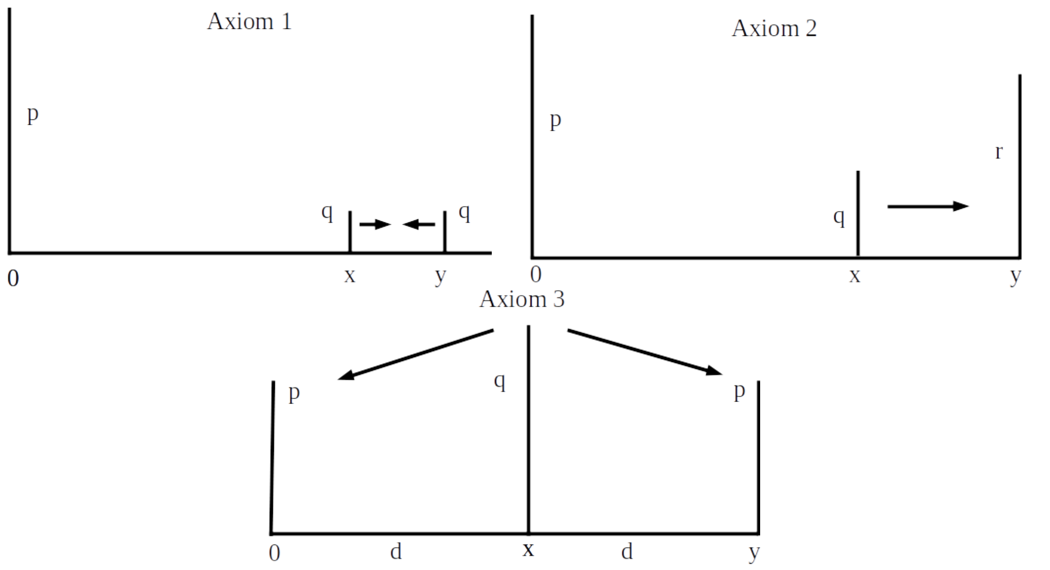

FIGURE 2.1: Illustrations of each of the three axioms imposed on the polarisation measure by Esteban and Ray, 1994. Axiom 1 imposes that the joining of clusters x and y at their centroid increase polarisation. Axiom 2 states that shifting a bit of mass from x to y increases polarisation. Axiom 3 establishes that equally shifting mass from x to the clusters at the edges also increases polarisation.

The authors also require that the sorting produced by the polarisation measure over two distributions be independent of population size $N = \sum_{i=1}^{N_c} n_i$, i.e. if we denote by $\mathbf{y}$ the vector of clusters:

- *Condition H*: If $P(\mathbf{n}, \mathbf{y}) \geq P(\mathbf{n'}, \mathbf{y'})$ for two distributions $(\mathbf{n}, \mathbf{y})$ and $(\mathbf{n'}, \mathbf{y'})$, then $P(\lambda\mathbf{n}, \mathbf{y}) \geq P(\lambda\mathbf{n'}, \mathbf{y'})\ \forall \lambda > 0$.

Theorem 1 in (Esteban and Ray, 1994) establishes that the expression in (2.1) satisfies all three of the above axioms and Condition H. The authors of (2.1) also show that the maximum polarisation is reached for a perfectly balanced 2-cluster, maximally separated configuration (i.e. half the points of a given population on either end of the distribution). Thus, if distance is normalised to 1, we have $P_{max} = 2(N_c/2)^{\alpha+2}$. This makes $K = 1/P_{max} = (2/N_c)^{\alpha+2}/2$ a reasonable choice for the normalisation constant in (2.1), and such is the value we use. We therefore have $P \in [0, 1]$.

There are a number of advantages to (2.1):

1. It is well founded: its axioms guarantee certain intuitively desirable properties of a polarisation measure.
2. Although it is defined on a distribution, it can also be implemented automatically in a network pipeline provided a notion of distance.
3. Normalisation to the maximal polarisation enables straightforward and meaningful comparisons among different systems.
4. Although here we will focus on polarisation on dipole partitions, the formula allows for an arbitrary number of clusters, contrary to most structural polarisation measures (Salloum, Chen, and Kivelä, 2021).

2.2 Graph partitioning

As I mentioned before, community detection is an ill-posed problem, and there is no all-purpose solution (Fortunato and Hric, 2016). In the present case, however, choice is facilitated by the constraints of the pipeline: namely, our polarisation measure (2.1) requires distances between clusters. A class of methods that implement cluster distances quite naturally is that of sequential, agglomerative, hierarchic, nonoverlapping (SAHN) clustering algorithms, and we use such methods here.

2.2.1 SAHN clustering algorithms

A SAHN algorithm (Müllner, 2011) gradually builds up clusters from the initial unit cluster data by merging the two closest clusters at a time (the last stage being that of only two clusters remaining). The procedure has therefore $N-1$ steps, each defining a different partition, for N initial observations.

The merging order is obtained by minimising a pairwise dissimilarity measure between the clusters that is updated according to a given scheme (e.g. any of those in Fig. 2) whenever a new cluster is formed. More concretely:

Definition. A *dissimilarity measure* on a set S is a map $d: S \times S \to [0, \infty)$ which is reflexive and symmetric, i.e. we have $d(x,x) = 0$ and $d(x,y) = d(y,x)$ for all $x, y \in S$.

In general, these algorithms take as input either the pairwise dissimilarities between the initial observations (*stored matrix approach*) or the set of observations S (*stored data approach*): in the present work, I follow the stored matrix approach.

Regarding the output, a standard is a data structure that has been called a *stepwise dendogram* by Müllner, 2011:

Definition. Given a finite set S_0 with cardinality $N = |S_0|$, a *stepwise dendrogram* is a list of $N-1$ triples (a_i, b_i, δ_i) $(i = 0, \ldots, N-2)$ such that $\delta_i \in [0, \infty)$ and $a_i, b_i \in S_i$, where S_{i+1} is recursively defined as $(S_i \setminus \{a_i, b_i\}) \cup n_i$ and $n_i \notin S \setminus \{a_i, b_i\}$ is a label for a new node.

In plain language: The set S_0 are the initial data points. At each step, n_i is the new node[1], formed by joining the nodes a_i and b_i with dissimilarity δ_i (the order of a_i and b_i within each pair is irrelevant). After step $N-1$, all N initial nodes are grouped in a single cluster.

The procedural definition of the above class of SAHN algorithms is presented in Fig. 1.

2.2.2 Input

In order to provide the input for the SAHN clustering, we need to introduce a notion of dissimilarity in the network. Given that application of the polarisation measure (2.1) in principle assumes one kind of entity, the natural approach to obtain such dissimilarities is to consider only one type of node of the bimodal network for clustering. Mukerjee, Majó-Vázquez, and González-Bailón, 2018 propose a way to contract the information of a bimodal network on one of its two kinds of nodes (that I will call the *primary* nodes) by means of the ϕ coefficient[2] (Pearson, 1900). The procedure relates the primary nodes by their overlapping ties to the secondary nodes.

[1] In this section (2.2.1) and in Appendix A I use n_i to denote the label of the cluster/node formed at step i of the hierarchical clustering, whereas in the rest of the document it refers to the size of cluster i: the two should not to be confounded.

[2] Known as the Matthews correlation coefficient in machine learning.

Figure 1 Algorithmic definition of a hierarchical clustering scheme. Taken from (Müllner, 2011).

1: **procedure** PRIMITIVE_CLUSTERING(S, d) ▷ S: node labels, d: pairwise dissimilarities
2: $N \leftarrow |S|$ ▷ Number of input nodes
3: $L \leftarrow [\,]$ ▷ Output list
4: $size[x] \leftarrow 1$ for all $x \in S$
5: **for** $i \leftarrow 0, \ldots, N-2$ **do**
6: $(a, b) \leftarrow \operatorname{argmin}_{(S\times S)\setminus\Delta} d$
7: Append $(a, b, d[a, b])$ to L.
8: $S \leftarrow S \setminus \{a, b\}$
9: Create a new node label $n \notin S$.
10: Update d with the information

$$d[n,x] = d[x,n] = \text{FORMULA}(d[a,x], d[b,x], d[a,b], size[a], size[b], size[x])$$

for all $x \in S$.
11: $size[n] \leftarrow size[a] + size[b]$
12: $S \leftarrow S \cup \{n\}$
13: **end for**
14: **return** L ▷ the stepwise dendrogram, an $((N-1) \times 3)$-matrix
15: **end procedure**

(As usual, Δ denotes the diagonal in the Cartesian product $S \times S$.)

The ϕ correlation is defined as:

$$\phi(v_1, v_2) = \frac{f_{11}f_{00} - f_{10}f_{01}}{\sqrt{f_{1*}f_{0*}f_{*1}f_{*0}}} = \frac{f_{11}F - f_{1*}f_{*1}}{\sqrt{f_{1*}f_{0*}f_{*1}f_{*0}}} \tag{2.2}$$

for any pair of observation binary vectors v_1, v_2, where F is the total number of observations, f_{11} is the number of coincidental positive (1) observations, f_{10} that of positive observations for v_1 that are also negative (0) for v_2, f_{1*} the total of positive observations in v_1 regardless of v_2, and so on.

Note that $\phi \in [-1, 1]$, the minimum being attained when $f_{11} = f_{00} = 0$ and the maximum when $f_{10} = f_{01} = 0$ (this is apparent in the first definition).

The observation vectors, in this case, correspond to the primary nodes, each vector encoding the connection value (yes: 1, no: 0) to each node of the secondary kind. F is therefore the number of secondary nodes.

As can be seen in 2.2, the ϕ coefficient is a way to have some statistical weighting of bimodal connectivity, the second definition showing that $\phi > 0$ only if the link overlap is greater than expected by chance.

From ϕ, one can compute an associated Euclidean distance as:

$$d = \sqrt{2(1-\phi)} \tag{2.3}$$

which provides the input for the SAHN algorithm. Note that $d \in [0, 2]$: in the present application, I normalise input distances to 1.

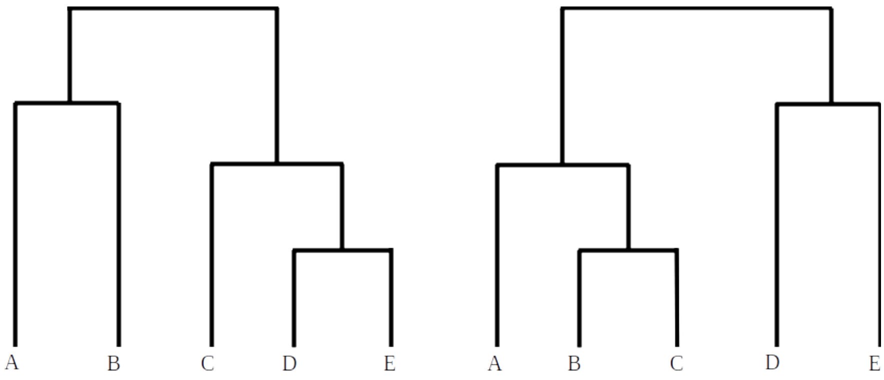

FIGURE 2.2: Example trees to illustrate the RF distance: it takes 4 steps (removing or adding one junction at a time) to go from one tree to the other.

2.3 Analysis tools

2.3.1 Robinson-Foulds distance

In order to compare the trees produced with different distance update schemes we use the (unweighted) Robinson-Foulds distance (Robinson and Foulds, 1981), a measure that captures topological differences between two trees T_1 and T_2 as the sum of the number of junctions that appear in only one of them, i.e.

$$D_{RF} = A + B, \tag{2.4}$$

where A is the sum of junctions found in T_1 but not in T_2 and B is the sum of junctions present in T_2 but not in T_1.

As is apparent from (2.4), D_{RF} is a sort of edit distance for trees, and it has the advantage of being quite intuitive: for instance, one can quickly see that it takes 4 steps (removing or adding one junction at a time) to go from the left to the right tree (or vice versa) in Fig. 2.2.

2.3.2 Jaccard distance

For the SW dataset (Chapter 4) there are two reference 2-cluster partitions (Homans, 1950; Breiger, 1974) to compare with the final partition given by the hierarchical clustering. Since the comparison involves sets, I make use of an adaptation of *Jaccard distance*.

Jaccard distance (Jaccard, 1912) is a measure of dissimilarity between sets. For sets A and B it is defined as:

$$D_J(A, B) = 1 - \frac{|A \cap B|}{|A \cup B|}$$

2.3.3 Word shifts

The CFE data (Chapter 5) consists of text, and evaluation of results relies on *word shifts* of the two clusters of the final partition.

Word shifts are a tool for pairwise comparison between texts that captures which words contribute to their difference and how. The contributions of the most relevant words are then visualized through horizontal bar charts called *word shift graphs*.

The *shifterator* Python package (Gallagher et al., 2021) implements word shifts. Shifterator's main input is the bag-of-words (BOW) representation of each of the two text to be compared, in the form of a Python dictionary whose keys are word types and whose values are the frequence of the corresponding word in the given text. The package provides several text comparison measures, which include relative frequency, Shannon entropy, Tsallis entropy, the Kullback-Leibler divergence, and the Jensen-Shannon divergence. Given that the JSD seems to be the more effective choice for stressing meaningful words (as opposed to stop words; see below) in the comparison (Gallagher et al., 2021), such is the one I implement.

A specific aspect of the BOW representation is worth noting: it may be improved by ignoring words that are deemed irrelevant for the subsequent analysis, known as *stop words* (prepositions, determiners, etc.). However, whether a given word is irrelevant depends highly on the application, so these must be handled with care.

Figure 2.3 shows an example of the word shift graphs produced by shifterator. The top of the plot shows the bars corresponding to the total JSD of each text relative to one another, while the below bar with the Σ shows the direction of their difference. By the title and Σ, we see that Cluster 1 has a JSD of about 3 times that of Cluster 2.

The subplot to the left is the cumulative contribution plot, which traces how the total JSD shift changes as we add more words according to their rank. The horizontal line shows the cutoff of the contributions of the top 50 words plotted versus those that are not, which means that in this case about 25% of the overall difference is explained by the top 50 words.

We can also see the relative size (in number of word occurrences, including repeated ones) of the texts belonging to each cluster at the right, showing that Cluster 1 is (excluding stop words) less than half the size of Cluster 2.

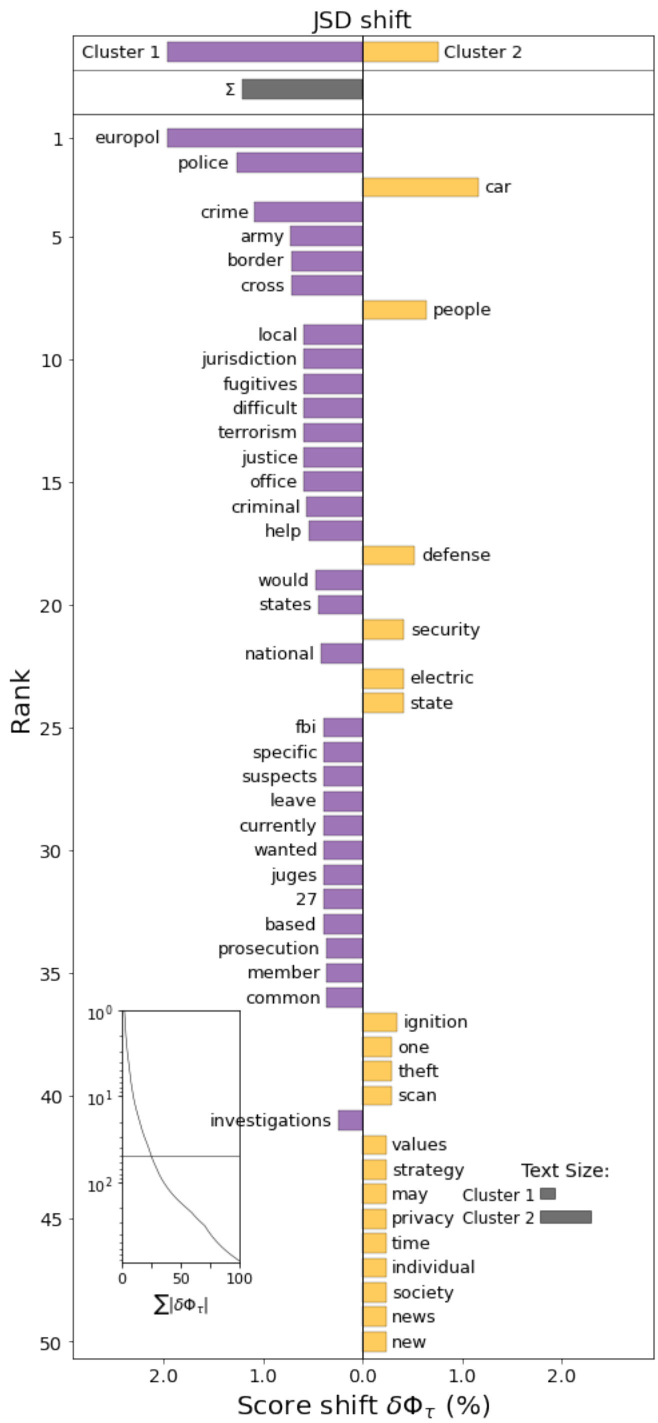

FIGURE 2.3: Example of a word shift graph.

Chapter 3

Proposal

3.1 Approach

As mentioned previously, the specific choices for each of the three steps of the network approach undertaken here are:

1. *Graph representation*. Bimodal networks, particularly of social systems.
2. *Clustering*. SAHN algorithms (Section 2.2.1).
3. *Computation of the polarisation*. The axiomatic measure (2.1).

3.2 Implementation

SAHN algorithms allow for the exploration of different dissimilarity measures as clustering criterion, of which we consider three: the well-knwon Ward and centroid distances, and the novel measure we termed *poldist*, based on the measure used for evaluating polarisation (2.1). Figure 2 shows the iterative formulas of each of these three distance update schemes, as well as their closed-form definitions.

Note that both d_W and d_p are just centroid distance weighted by a function of cluster sizes, which is ≥ 1 for d_W and ≥ 2 for d_p. For Ward, it makes clustering favour merging two unit clusters (for which $d_W = d_c$) and postpone merging two large ones. For poldist, the behaviour is qualitatively the same, although more pronounced, since we have at least (depending on α) a quadratic order on cluster sizes. Both Ward and poldist will therefore tend to make the resulting tree more "balanced" and stretched towards the root, with poldist's effect being greater.

Another feature of the poldist formula is that it coincides with global polarisation when there are only two clusters (e.g. at the last step of the clustering).

Regarding the choice of a specific algorithm, Müllner, 2011 studies different SAHN optimisations and compares their performance for several distance update schemes. In particular, he recommends the *nearest-neighbour chain clustering* algorithm (having a worst case time complexity of $O(n^2)$) for Ward clustering and the *generic clustering* algorithm (of $O(n^3)$ worst case time complexity, but usually closer to $O(n^2)$ in practice) for the centroid method. I follow such recommendations in this work.

The reason for using a less efficient algorithm with centroid distance is that the particular optimisation strategy of the nearest-neighbour clustering algorithm relies on a post-processing step that sorts the stepwise dendogram by increasing distances. This approach does not produce a valid solution for the centroid update scheme, since d_c does not necessarily produce a monotonically increasing sequence of distances (in other words, the distance to a newly merged cluster may be smaller that to any of its components).

Figure 2 Agglomerative clustering schemes. Adapted form Figure 2 in (Müllner, 2011).

Name	Distance update formula FORMULA for $d(I \cup J, K)$	Closed form of cluster distance
Ward (d_W)	$\sqrt{\frac{(n_I+n_K)d(I,K)^2+(n_J+n_K)d(J,K)^2-n_K d(I,J)^2}{n_I+n_J+n_K}}$	$\sqrt{\frac{2n_I n_J}{n_I+n_J}} \cdot \lVert\vec{c}_I-\vec{c}_J\rVert_2$
centroid (d_c)	$\sqrt{\frac{n_I d(I,K)^2+n_J d(J,K)^2}{n_I+n_J}-\frac{n_I n_J d(I,J)^2}{(n_I+n_J)^2}}$	$\lVert\vec{c}_I-\vec{c}_J\rVert_2$
poldist (d_p)	$\kappa[(n_I+n_J)^{\alpha+1}n_K+n_K^{\alpha+1}(n_I+n_J)]d_c(I\cup J,K)$	$\kappa(n_I^{\alpha+1}n_J+n_J^{\alpha+1}n_I)\cdot\lVert\vec{c}_I-\vec{c}_J\rVert_2$

Legend: Let I, J be two clusters joined into a new cluster, and let K be any other cluster. Denote by n_I, n_J and n_K the sizes (i.e. the number of elements) of clusters I, J, K, respectively.

The update formulas for the Ward and centroid methods assume Euclidean distance as dissimilarity measure (which is the one we use for all three methods). The expression $\vec{c}_X$ denotes the centroid of a cluster X. The factor κ in front of the poldist formulas is taken equal to the normalising constant used for polarisation in (2.1), and α is taken equal to the polarisation sensitivity parameter of the same expression.

All three measures defined above are dissimilarity measures in the sense that they (trivially) satisfy for all I, J the conditions of positiveness $d(I,J) \geq 0$, reflexiveness $d(I,I) = 0$ and symmetry $d(I,J) = d(J,I)$. I also refer to them as distances, although they are not proper distances insofar as the distance of different elements may be 0 (e.g. for two clusters that have the same centroid), and even if d_c does satisfy the Triangle Inequality ($d(I,J) \leq d(I,K) + d(K,J)\ \forall I,J,K$), d_p and d_W do not (e.g. take $n_I = n_J \gg 1, n_K = 1$ as a counterexample).

The generic clustering algorithm, on the other hand, can be applied to any clustering distance, which is the reason why I also apply it with the poldist method.

Both algorithms are implemented by the Python library Scipy (Virtanen et al., 2020), with the particularity of producing a slightly modified version of a stepwise dendogram the authors call a *linkage matrix*: instead of having as rows only (a_i, b_i, δ_i) triplets with the merged clusters and their distance at each step, the output matrix features tetrads that include as well the size of the newly formed cluster.

The implementations used in this work are an adaptation of Scipy's[1], with the main modification being that polarisation (2.1) is computed at every step of the clustering. Such a computation adds an $O(n^3)$ worst-case time complexity term to both the nearest neighbour and generic clustering algorithms, so their lower bounds are here more similar than in their pure versions (Müllner, 2011).

Finally, implementation of the polarisation formula (2.1) and of poldist (see Fig. 2) uses d_c (centroid distance) as distance. There are two reasons for this: i) it comes as a natural candidate; ii) other possibilities may depend on cluster size (as is the case for Ward distance), which in general will complicate the normalisation of P because the maximum distance may change at every step due to the increment in sizes.

The pseudocode of the adaptations of the generic and nearest-neighbour clustering algorithms implemented is included in Appendix A.

3.3 Evaluation

Evaluation of the results focuses on the final, dipole partition, which seems the most natural choice for studying polarisation. The procedure depends on the use case: for

[1] The adaptations are written in the Python language, both for ease of implementation and for readibility.

the SW dataset (Chapter 4), the deviation of the final partition to the 2 reference ones is computed via a modification of Jaccard distance (see Section 2.3.2); for the CFE data (Chapter 5), the final partition is evaluated through word shift graphs (Section 2.3.3).

3.3.1 Modified Jaccard distance

Jaccard distance is a measure of dissimilarity between sets, and we do not want to compare two sets but two *pairs of sets*: the final state of any clustering with either of the partitions obtained by Homans, 1950 and Breiger, 1974. Hence why I propose in this section an adaptation of Jaccard distance to the present case[2].

Since our clustering always contains all women and the two reference partitions do not, one can argue for using only the set of possible coincidental elements (instead that of all possible elements) in the denominator. Therefore, if we have the pairs of sets (R_1, R_1) and (C_1, C_2), the first being the reference, one way to adapt D_J is:

$$D_J^*((R_1, R_2), (C_1, C_2)) = 1 - S_J^*((R_1, R_2), (C_1, C_2)) \tag{3.1}$$

$$S_J^*((R_1, R_2), (C_1, C_2)) = \sigma^{-1} \left[\frac{\max(|R_1 \cap C_1| + |R_2 \cap C_2|, |R_1 \cap C_2| + |R_2 \cap C_1|)}{|R_1 \cup R_2|} - \mu \right]$$

and μ and σ are just parameters to re-center and rescale S_J^* (that we may call the *modified Jaccard similarity*) so that it varies between 0 (when there is minimal overlap between the partitions) and 1 (when the reference sets are separately contained in the obtained partition):

$$\mu = \frac{1}{2} + \frac{|R_1 \cup R_2| \; mod \; 2}{|R_1 \cup R_2|} \tag{3.2}$$

$$\sigma = 1 - \mu \tag{3.3}$$

where $x \; mod \; y$ denotes the remainder of the division of x by y.
I will call D_J^* the *modified Jaccard distance*.

3.3.2 Word shifts

Since that of the Conference on the Future of Europe is current, unstudied data, we lack a reference for comparison. However, given that such data consist of text (the titles and bodies of the proposals), one can qualitatively evaluate the final state through such text. Note that this would not be possible, or at least not directly, if one did the analysis on endorsers (see section 5.2.1).

There are a number of ways to go about such a content-based evaluation. The most obvious (and cumbersome) is to read all the proposals (more than 10^3) and apply some predefined labelling criterion that allows to evaluate the final partition.

[2] It recently came to my knowledge (yesterday, January 13, 2022) that there exist methods to compare different partitions of a set, like those based on the Rand (Rand, 1971) or the Wallace (Wallace, 1983) coefficients. I apologise to the reader already familiar with such methods for introducing a new criterion. Note however that such methods are not directly applicable here, for they are defined for two partitions of the same set, whereas the partitions reported by Homans, 1950 and Breiger, 1974 omit the women not belonging to either of the 2 main communities.

But since one would rather have at least a quantitative component and reduce subjectivity as much as possible, and given that this is not a master's thesis on sociometrics but on data science, we may leverage the power of natural language processing methods. One of such methods is *word shifts* (see Section 2.3.3).

The approach for evaluation in Chapter 5 will thus be: i) retrieve the total text of each cluster of a final partition; ii) build their respective BOW representations; iii) obtain their word shift graph; and iv) analyse the graph.

Regarding stop words, in the application in Chapter 5 I use a predefined minimal list thereof provided by the Nltk Python package (Bird and Klein, 2009), that I checked previously, to which I add some extra stop words found in the early word shift graphs. There are also some stop words candidates I found for which I was not quite sure, so I do not use them for the results reported in this thesis[3].

3.4 Pipeline

Now that we have introduced the methods used, let us review the pipeline. For a given system describable as a bimodal network:

1. Build the bimodal network representation.

2. Compute the pairwise ϕ coefficient (2.2) among the primary nodes (women in the SW use case, proposals in the CFE one). Then compute the corresponding distances $d(\phi)$ (2.3).

3. Apply hierarchical clustering with Ward, centroid and poldist distance update schemes (Fig. 2), evaluating polarisation (2.1) at every step.

Finally, evaluation follows the procedures specified in the previous section: for the SW dataset, compare the final partition of the clustering with the reference partitions (Homans, 1950; Breiger, 1974); for the CFE data, qualitatively evaluate the word shift of the texts of the two final clusters.

[3] The complete list of stop words considered is available at the Github repository of the project: https://github.com/adrifcid/polarisation.

Chapter 4

Use Case 1: The Southern Women dataset

4.1 Data

The Southern Women dataset is a table containing binary values encoding the attendance (or absence thereof) of 18 women to a series of 14 social events (a card party, a club meeting, etc.) that took place throughout a year, originally compiled (Davis, Gardner, and Gardner, 1941) with the purpose of determining the influence of social status on the forming of communities of individuals.

Although not apparent from the beginning, two groups of women became distinguishable in the reappraisals of Homans, 1950 and, in an alternative, more straigthforward way, of Breiger, 1974. Such partitions provide a reference for evaluating our results.

The SW has since become a standard dataset in computational social science, namely for testing clustering/community-detection methods in bimodal networks, and is easily available on the usual network repositories[1]. Figure B.1 in Appendix B shows the original bimodal social network.

4.2 Experimental setting

The purpose of this use case is to validate to some extent the pipeline proposed here (see Section 3.4), also used in Chapter 5.

The primary nodes for the analysis are the women. Although one could apply the same treatment event-wise, it is more interesting from a conceptual (social) perspective to do it for women, and such is also the perspective adopted by previous studies.

Evaluation of results relies on the previous work of Homans, 1950 and Breiger, 1974[2].

Homans finds that by rearranging the woman-event matrix so that women that coincided most often are grouped and events attended by most women are near the center (a tedious approach indeed, considering it had to be done manually at the time) two groups become distinct: one formed by Charlotte, Eleanor, Brenda, Theresa, Evelyn, Laura and Frances; and the other by Nora, Katherine, Helen, Sylvia

[1] See, for instance, the CASOS repository: http://casos.cs.cmu.edu/computational_tools/datasets/external/davis/index2.html.

[2] As mentioned before, Davis, Gardner, and Gardner, 1941 were not able to determine any meaningful women communities from the women-events network. The reason is that nearly all women are connected to each other by attendance to at least one common event (as pointed out by Breiger, 1974, there are 139 such links from a total of $(1/2)(18)(17) = 153$ possible ones, which makes a connectivity as high as 91%).

TABLE 4.1: Summary of results on the SW dataset for the three distance update schemes used.

Quantity	Ward	Centroid	Poldist ($\alpha = 1$)
P	0.53	0.53	0.41
$\mathbf{D}^*_{\mathbf{J,Homans}}$	0	0	0.17
$\mathbf{D}^*_{\mathbf{J,Breiger}}$	0.13	0.13	0.5

and Myrna. The rest of the women are considered not to clearly belong to either group.

In the case of Breiger, by eliminating those events connected to every other event by at least one woman (in order to leave only events that help discriminate among women) he finds two groups formed by: Charlotte, Ruth, Eleanor, Brenda, Theresa, Evelyn, Laura and Frances; and Verne, Nora, Flora, Katherine, Olivia, Helen, Sylvia and Myrna. Note that these contain Homans' groups, but add Ruth, Verne, Olivia and Flora, whom Homans judged not to belong to any group.

The previous results summarised above mean that we have two reference partitions to compare our clustering with, by means of the modified Jaccard distance D_J^* (3.1).

4.3 Results

Figure 4.1 shows the trees obtained with the Ward, centroid and poldist ($\alpha = 1$) distance update schemes, coupled with a heatmap of the input matrix distance to enhance visual interpretation (such a plot is known as a *clustermap*). It is apparent that all three methods present two main clusters, coinciding (at least qualitatively) with previous results.

Note the inversion in one branch of the left cluster in the centroid tree: indeed, as mentioned previously, the centroid update scheme allows for such a behaviour, whereas Ward's does not (Müllner, 2011). As has also been mentioned, that is precisely why the nearest-neighbour clustering algorithm does not produce a valid solution for the centroid method.

The quantitative results for Ward, centroid and poldist (with $\alpha = 1$) clustering are summarised in Table 4.1, which contains the final polarisation P and deviations D_J^* of the final partition from those of Homans, 1950 and Breiger, 1974. Only the Ward and centroid methods have $P > 0.5$, and their deviations from the reference partitions show agreement with previous work. Such is not the case for poldist, which adds to the fact that it gives a comparatively lower polarisation to a system known to have two communities: it seems its balancing tendency is too pronounced and distorts the clustering procedure.

Figure 4.2 displays the Robinson-Foulds distance (D_{RF}) of Ward, centroid and poldist-0 (with $\alpha = 0$[3]) trees to the poldist one for different α. As one might expect due to their similar definition, the distance of the Ward-poldist pair (14) is lower than that of the centroid-poldist one (18), as well as that of centroid and Ward (16, not included in the Figure). Note that the fact that the final partitions of Ward and

[3]The Gini coefficient limit of α, not strictly allowed in the polarisation formula (2.1) but included for reference.

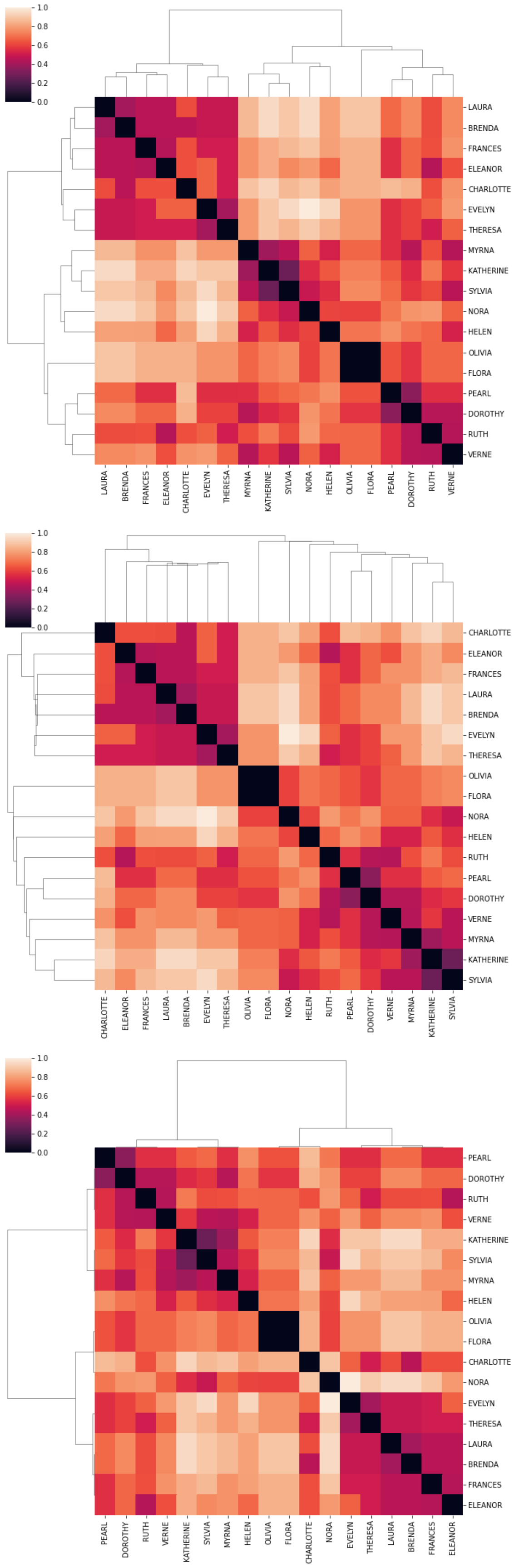

FIGURE 4.1: Southern Women clustering for the three distance update schemes considered: Ward (top), centroid (middle) and poldist ($\alpha = 1$, bottom).

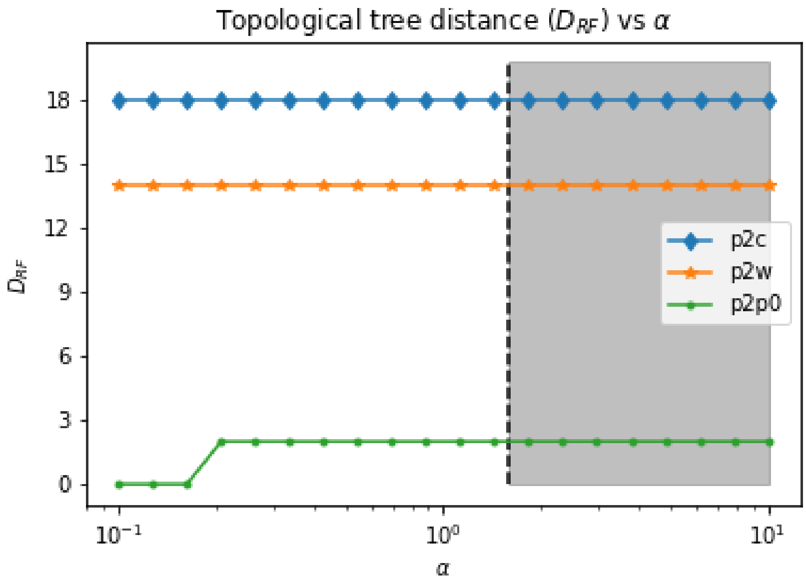

FIGURE 4.2: Robinson-Foulds distance of Ward (*p2w*), centroid (*p2c*) and poldist-0 ($\alpha = 0$; *p2p0*) trees to the poldist tree for different α, on the Southern Women dataset. Although $\alpha = 0$, the Gini coefficient limit of α, is not strictly allowed in the polarisation formula (2.1), it is included here for reference.

centroid clustering reasonably agree with previous results (see Table 4.1) does not prevent their trees from being overall rather different.

Another remarkable feature in Fig. 4.2 is the absence of evolution in the poldist tree. It seems that, aside from the change from $\alpha \approx 0.16$ to $\alpha \gtrsim 0.21$ (that yields $D_{RF} = 2$ to the tree obtained with $\alpha = 0$), there are no topological changes in the poldist tree for the different α allowed ($\alpha \in (0, 1.6]$, to the left of the vertical line), and not even up to $\alpha = 10$. The distances between branches do change (I checked), but the structure is the same[4]. This shows that one cannot hope to improve agreement with the reference partitions (Table 4.1) by tuning α.

Figure 4.3 shows the evolution of the polarisation through the clustering for the three methods considered and the allowed range of the polarisation sensitivity $\alpha \in (0, 1.6]$. Recall that with poldist the value of α is the same for the distance update scheme and the polarisation. For all three methods, we observe a convergence in the final state of the clustering for every α (the numerical values, already commented on, are shown in Table 4.1).

In addition, we see that for low α, P tends to start high and decrease with clustering step, while the behaviour is the opposite for higher values: α starts low and increases. This is consistent with the description by Esteban and Ray, 1994: the polarisation formula (2.1) reduces to the standard Gini inequality coefficient for $\alpha \approx 0$, and inequality is greater in a system with only unit clusters than when they start

[4]From the poldist formula in Fig. 2, it seems that if α is sufficiently large, the clustering is dominated by the sizes: if the sizes are equal, the two clusters to merge are picked by d_c; but otherwise the chosen pair is the one with minimal sizes, independently of d_c. The likely interpretation is thus that for every network (more specifically, for every input distance matrix d_c) there is a threshold α_t beyond which the topology of the resulting tree is invariant under changes in α. Such a threshold need not be very high, as the function on the sizes is already quadratic for $\alpha = 0$, and for the SW dataset we may have $\alpha_t \in (0.16, 0.21)$.

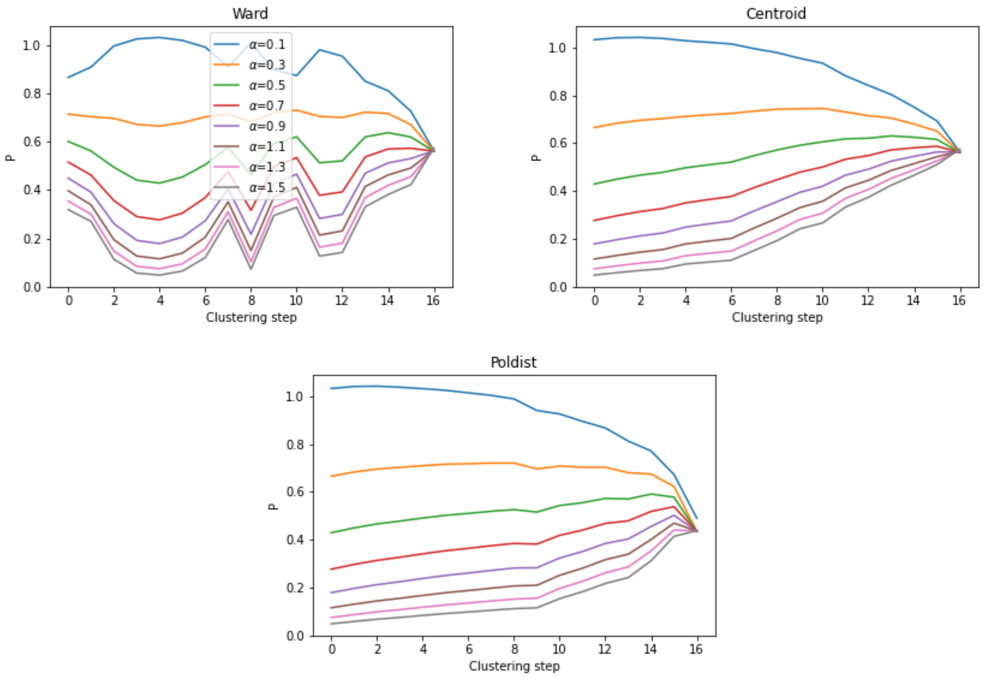

FIGURE 4.3: Polarisation throughout the clustering for different α in the Southern Women dataset, for the three distance update schemes considered: Ward, centroid and poldist.

to merge; on the other hand, the contribution of a non-vanishing α is to allow for intra-group identification to increase inter-group alienation, hence favouring greater cluster sizes. Figure 4.3 thus provides the intuition of the effect of α, and why it is called *polarisation sensitivity*.

A final result is worth mentioning. Since Breiger, 1974 modifies the original data, namely eliminating events connected to all other events, and the resulting groups of women are disconnected, it is of interest to evaluate the polarisation of such a system. One expects such a polarisation to be rather high (close to 1), and such is indeed the case: $P = 0.68$, which is higher than the previous maximum of 0.53 in Table 4.1. The details of the computation are included in Appendix C.

To conclude this chapter, following the results on the SW dataset it seems reasonable to consider the proposed pipeline validated for the Ward and centroid methods due to their agreement with previous work (displayed in Table 4.1). The poldist method, however, shows remarkable deviation from such reference values for $\alpha \in (0, 1.6]$, and is therefore not validated.

Additionally, the fact that the polarisation obtained with the validated methods is only moderate ($P = 0.53$ for both Ward and centroid) suggests polarisation in the original data is not pronounced, although it does increase when considering Breiger's subset of the data.

Chapter 5

Use Case 2: The Conference on the Future of Europe data

5.1 Data

The Conference on the Future of Europe (Parliament, Council, and Commission, 2021) is a project of the European Union led by the European Parliament, the European Council and the European Commission to engage European citizens and more widely the European civil society in democratic deliberation on EU policies. Deliberation takes place through the making and evaluation of concrete proposals and the organisation of and participation in events. By the end of the Conference (expected in the spring of 2022) the organisers commit to capture the proposals and discussions of the whole process into concrete policy recommendations.

Parcipants may be European citizens; European, national, regional or local authorities; or civil society organisations. They can engage through the multilingual platform of the Conference[1] (featuring the 24 official languages of the EU), which is an instance of the *Decidim* application for mass deliberation[2]. It allows users to propose and vote "ideas" (that one can follow or endorse), organise and attend events ("meetings"), or leave comments on any of the two, among other functionalities.

The digital platform was launched in mid-April 2021 and is foreseen to remain accessible until spring 2022, when the Conference is expected to reach conclusions. At the time the data was collected for this project, the 20th October 2021 at 03h 16min 54s (GMT+1), there were a total of approximately 9000 proposals, 10000 endorsers (users that gave their support to at least one proposal) and 45000 proposal endorsements.

The data, which include information on proposals, meetings and their respective comments, present the opportunity to apply the pipeline proposed here on a current process of mass deliberation: the bimodal network is here composed of proposals and endorsers.

5.2 Experimental setting

The results on the CFE data presented here omit those of the poldist method, as it was not validated by the study on the SW dataset (and gives ostensibly poor results here as well). The poldist results are available at the project's Github repository[3].

[1] See https://futureu.europa.eu/?locale=en.
[2] See https://decidim.org/.
[3] See https://github.com/adrifcid/polarisation.

5.2.1 Preprocessing

Before beginning the analysis on the CFE data, the following preprocessing actions were taken:

- Consider only proposals originally posted in English. This is to make sure there are no spurious distance or polarisation effects due to language[4].
- Remove amends. Proposals may be amends of previous ones, but I removed those cases to avoid spurious endorser overlap.
- Take proposals as primary nodes. Although it seems more intuitive to cluster on endorsers rather than on proposals, it is a technical constraint that the number of the former is usually much larger than that of the latter. Additionally, performing the analysis on proposals allows to make qualitative sense of the results by evaluating their content, which cannot be done with virtually anonymous endorsers, and any appreciative polarisation in the system should be present for both proposals and endorsers anyway (albeit with a relation that is not self-evident).
- Consider only proposals that include a title and have at least 1 endorsement. In the topic-specific analysis (see below), consider only topics with at least 2 proposals.

5.2.2 Analysis by topic

Proposals are categorised by topic ("Security", "Education", "Disinformation", "Coronavirus", etc.). Since it is expected that some users restrain their participation to the topics of their interest (for instance, an ecologist organisation might focus on climate change related proposals) and some topics may be more controversial than others, a topic-specific analysis imposes itself.

Table 5.1 summarises the content of the 32 pre-processed datasets used in the analysis by topic, including for every topic the number of proposals N, the number of endorsers N_e, the number of endorsements N_E, the average endorsements per proposal N_E/N and the average endorsements per endorser N_E/N_e. The bold rows correspond to the topics whose complete results are shown in Section 5.3.1. The total number of proposals considered is 1251. From the table we see that endorsement participation is rather low, the average endorsements per endorser lying between 1 and 2.

5.2.3 Global analysis

The global perspective allows to see whether endorsers participate more or less homogeneously across topics (arguably the ideal case from a participatory point of view) or, as one might expect, they hold to proposals on one or a few given topics.

Table 5.2 shows the summary of the global dataset. In this case, participation is a bit higher (average endorsements per endorser are 2.76), which means there is at least some (even if very little) participation across topics.

[4]As expected from an EU platform, the proposals may be in any language of a member state. One could in principle consider languages other than English because the platform provides an optional machine translation so that every user can vote and comment on every proposal, and the translations are included in the data. However, the complete absence of influence of language on participation cannot be guaranteed, and we prefer to avoid that factor in this first study.

TABLE 5.1: Summary of the 32 datasets used in the analysis by topic on the CFE data. The bold rows correspond to the topics whose complete results are shown in Section 5.3.1. The total number of proposals is 1251.

Topic	N	N_e	N_E	N_E/N	N_E/N_e
Using resources efficiently for a circular economy	53	335	561	10.58	1.67
Ensuring a fair and inclusive transition	46	673	791	17.2	1.18
Restoring biodiversity and cutting pollution	79	461	786	9.95	1.7
Promoting healthy lifestyles	15	75	91	6.07	1.21
Coronavirus	15	45	54	3.6	1.2
Healthcare	45	875	1162	25.82	1.33
Disinformation	5	18	18	3.6	1.0
Have your say on European Union policies	135	1817	3327	24.64	1.83
Protecting our democracies	101	579	1006	9.96	1.74
Media	13	50	59	4.54	1.18
Boosting jobs, growth and investment	**63**	**507**	**685**	**10.87**	**1.35**
A more inclusive and fairer economy	104	513	892	8.58	1.74
Coronavirus recovery	9	128	141	15.67	1.1
European rights and values	78	536	747	9.58	1.39
Security	20	134	195	9.75	1.46
Consumer rights	6	47	47	7.83	1.0
Technology for people	60	269	493	8.22	1.83
Digital economy	29	229	329	11.34	1.44
A sustainable digital society	27	149	179	6.63	1.2
Security and defence	35	504	683	19.51	1.36
Development cooperation	18	44	63	3.5	1.43
Foreign policy	60	532	907	15.12	1.7
Neighbourhood policy and enlarging the European Union	28	235	308	11.0	1.31
Trade policy	10	65	68	6.8	1.05
Humanitarian aid and civil protection	4	11	14	3.5	1.27
Legal migration and integration	25	90	138	5.52	1.53
Integrated Border Management	3	26	26	8.67	1.0
Asylum and Migration	**49**	**265**	**418**	**8.53**	**1.58**
Deepening international cooperation	7	19	21	3.0	1.11
Culture	**27**	**173**	**259**	**9.59**	**1.5**
Education	72	994	1400	19.44	1.41
Youth	10	25	26	2.6	1.04

TABLE 5.2: Summary of the global CFE dataset.

N	N_e	N_E	N_E/N	N_E/N_e
1252	5760	15896	12.7	2.76

Since in the global dataset we expect a multipole fragmentation in terms of topics, it is of most interest to evaluate not only the final partition given by the clustering, but those of the whole procedure. Such an analysis is, unfortunately, *still in progress*, current results being only for the final partition: hence why I have included them in Appendix D.

5.3 Results

5.3.1 Analysis by topic

This section presents and discusses the results of the analysis by topic. Given the quantity of plots involved and the qualitative homogeneity of results, I explicitly show here only a sample thereof: the full results are available in the project's Github repository[5].

Table 5.3 shows the number of proposals N, the final polarisation for Ward and centroid clusterings ($P_{w,c}$) and the RF distance between the corresponding trees (D_{RF}) for all 32 topics considered. The bolded rows correspond to the topics whose complete results are displayed in this section.

Overall, final polarisation is low, surpassing 0.5 only for topics with $N < 5$, which is hardly a representative size. It seems that a lower number of proposals generally tends to yield a higher final polarisation: indeed, the greater the population, the greater the number of possible configurations of the system, which means that a lesser population is in general more likely to end up with a configuration closer to the one that maximises polarisation that a greater one. Although this should be taken into account when comparing populations of different sizes, it appears to be a general fact, as opposed to one that could be attributed to the specific pipeline proposed here.

Figure 5.1 shows the clustermaps of proposals on "Culture", "Boosting jobs, growth and investment" and "Asylum and Migration" for the Ward and centroid methods[6]. In general, it seems the clustering rarely reveals any clear dipole structure around the main diagonal (which is arguably what we could identify as positive signal): the topics featured were chosen because they are among the closest to doing so, but even there the structure is not evident, with the exception of the top-right clustermap (corresponding to centroid clustering on the topic "Culture"). In any case, polarisation is not as prominent as in the SW dataset (see Chapter 4), as it stays below 0.5 (Table 5.3).

Figure 5.5 shows the evolution of polarisation for different α for Ward and centroid clustering. We observe the same overall qualitative behaviour we saw with the Southern Women dataset: a general convergence at the end of the clustering, with the Ward lines being more wobbly; and decreasing polarisation for near-vanishing α that picks up an increasing trend as the parameter augments. The centroid method, however, constitutes remarkable exception to this last feature, showing a peak (the expected increase is followed by a sharp decrease).

Indeed, centroid clustering shows overall a particular behaviour, not found in the SW dataset: it tends to form one cluster to which it then gradually adds smaller nodes. Such a transition shows clearly in the polarisation plots of Fig. 5.5 as the mentioned distinct peak, after which P decreases to vanishing values (as it should,

[5] See https://github.com/adrifcid/polarisation.

[6] The corresponding network plots of the three featured topics are included in Section B.2 of Appendix B.

TABLE 5.3: Number of proposals N, final polarisation for Ward and centroid clusterings ($P_{w,c}$) and RF distance between the corresponding trees (D_{RF}) in all 32 topics considered. The bolded rows correspond to the topics whose complete results are displayed.

Topic	N	P_W	P_c	D_{RF}
Using resources efficiently for a circular economy	53	0.21	0.04	86
Ensuring a fair and inclusive transition	46	0.26	0.05	64
Restoring biodiversity and cutting pollution	79	0.13	0.03	122
Promoting healthy lifestyles	15	0.15	0.15	16
Coronavirus	15	0.29	0.14	20
Healthcare	45	0.23	0.05	72
Disinformation	5	0.45	0.45	0
Have your say on European Union policies	135	0.18	0.02	202
Protecting our democracies	101	0.1	0.02	152
Media	13	0.31	0.17	14
Boosting jobs, growth and investment	**63**	**0.24**	**0.04**	**92**
A more inclusive and fairer economy	104	0.2	0.02	152
Coronavirus recovery	9	0.26	0.26	8
European rights and values	78	0.14	0.03	112
Security	20	0.27	0.11	22
Consumer rights	6	0.4	0.4	0
Technology for people	60	0.16	0.04	72
Digital economy	29	0.27	0.08	36
A sustainable digital society	27	0.27	0.08	34
Security and defence	35	0.18	0.07	50
Development cooperation	18	0.36	0.12	18
Foreign policy	60	0.18	0.04	92
Neighbourhood policy and enlarging the European Union	28	0.25	0.08	34
Trade policy	10	0.43	0.24	12
Humanitarian aid and civil protection	4	0.71	0.71	0
Legal migration and integration	25	0.24	0.09	34
Integrated Border Management	3	0.71	0.71	0
Asylum and Migration	**49**	**0.24**	**0.04**	**74**
Deepening international cooperation	7	0.4	0.31	8
Culture	**27**	**0.35**	**0.09**	**34**
Education	72	0.22	0.03	122
Youth	10	0.35	0.22	14

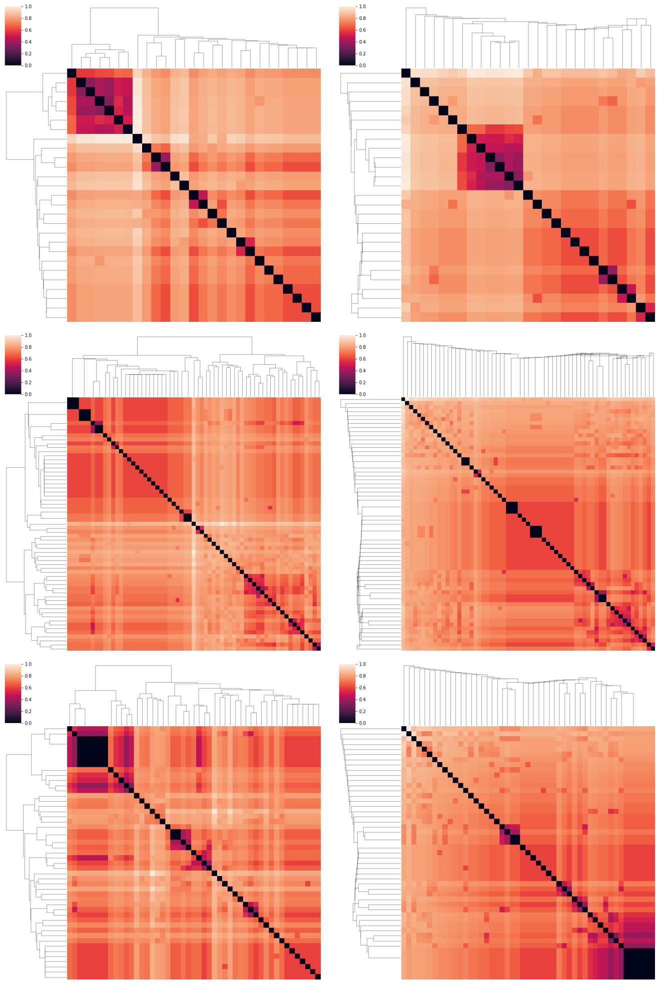

FIGURE 5.1: CFE topic clustermaps for Ward (left column) and centroid (right column) methods. From the upper to the lower row, the topics are "Culture", "Boosting jobs, growth and investment" and "Asylum and Migration".

because the algorithm is increasing the influence of a single dominant group). This behaviour is rather general across topics, as is reflected by the large RF distance of the Ward and centroid trees in Table 5.3.

The results for centroid clustering highlight the importance of evaluating P at every step to monitor the clustering, and suggest considering also configurations other than the final one in the analysis: for instance, the partition that maximises P is likely more interesting in the centroid clustering cases shown, especially that of "Culture" (top-right clustermap in Fig. 5.1).

Figures 5.2, 5.3 and 5.4 show the word shift graphs of the final 2-cluster configurations of the topics displayed. The word shift graphs are overall not very relevant due to the general lack of polarisation signal, and I have indeed been unable to find any content-based justification of the partitions by inspecting them (especially in the centroid cases, where the final partition is a particularly poor, unbalanced choice). Nevertheless, the graphs do reveal that proposals are in general ostensibly related to their topic (contrary to what would happen with spam or otherwise ill-intentioned content), which is desirable regarding the participatory intent of the platform.

Let us review the content of this Chapter before its conclusion.

The results are overall negative: the clustering failed to reveal any significant dipole structure in most of the networks considered.

The absence of positive signal could be a fault of the pipeline, but it does not seem likely: one the one hand, it has been validated (with the Ward and centroid distance update schemes) on the SW data; on the other, the low participation in the datasets used (average endorsements per endorser is between 1 and 2 in the analysis by topic, and only slightly higher –2.76– in the global case) makes it difficult for any hypothetical underlying polarising tendencies to manifest. Additionally, one may argue that users of a EU mass deliberation platform already share part of their ideology by the sheer fact of their participation, reducing the probability of controversy to arise.

Finally, following the results on the topic-specific analysis (as well as the similar, preliminary results obtained on the global dataset, presented in Appendix D), the likely interpretation seems to be that there is no polarisation signal to find in the data studied here. It will nevertheless be of interest to consider extensions of the data that feature higher participation (like an update, or including proposals in languages other than English) and to evaluate the clustering at stages other than the final one (as the behaviour of the centroid clustering suggests there may be some relevant structure before the final partition is obtained) for a more accurate assessment of the polarisation in the platform of the Conference on the Future of Europe.

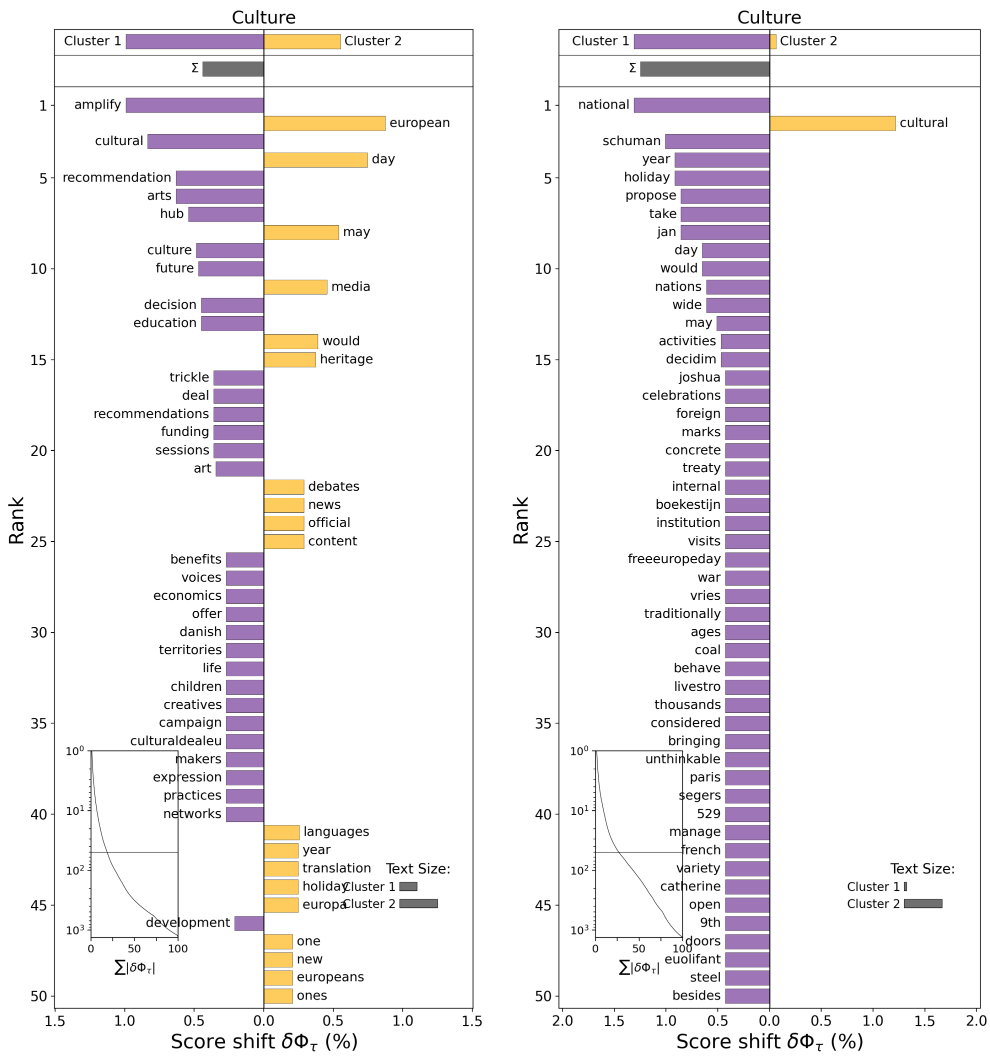

FIGURE 5.2: Word shift graphs of the final partition in the topic "Culture" of the CFE data, for the two validated distance update schemes: Ward (left) and centroid (right).

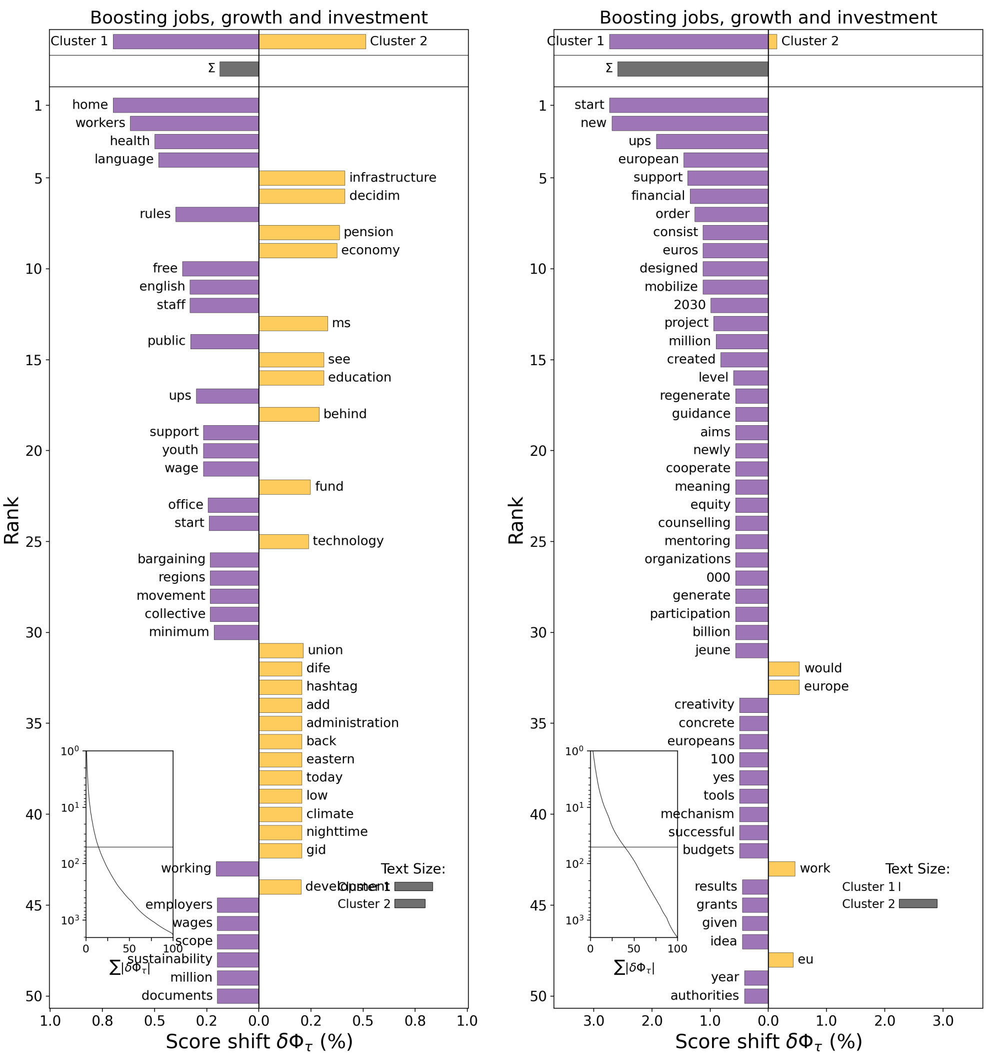

FIGURE 5.3: Word shift graphs of the final partition in the topic "Boosting jobs, growth and investment" of the CFE data, for the two validated distance update schemes: Ward (left) and centroid (right).

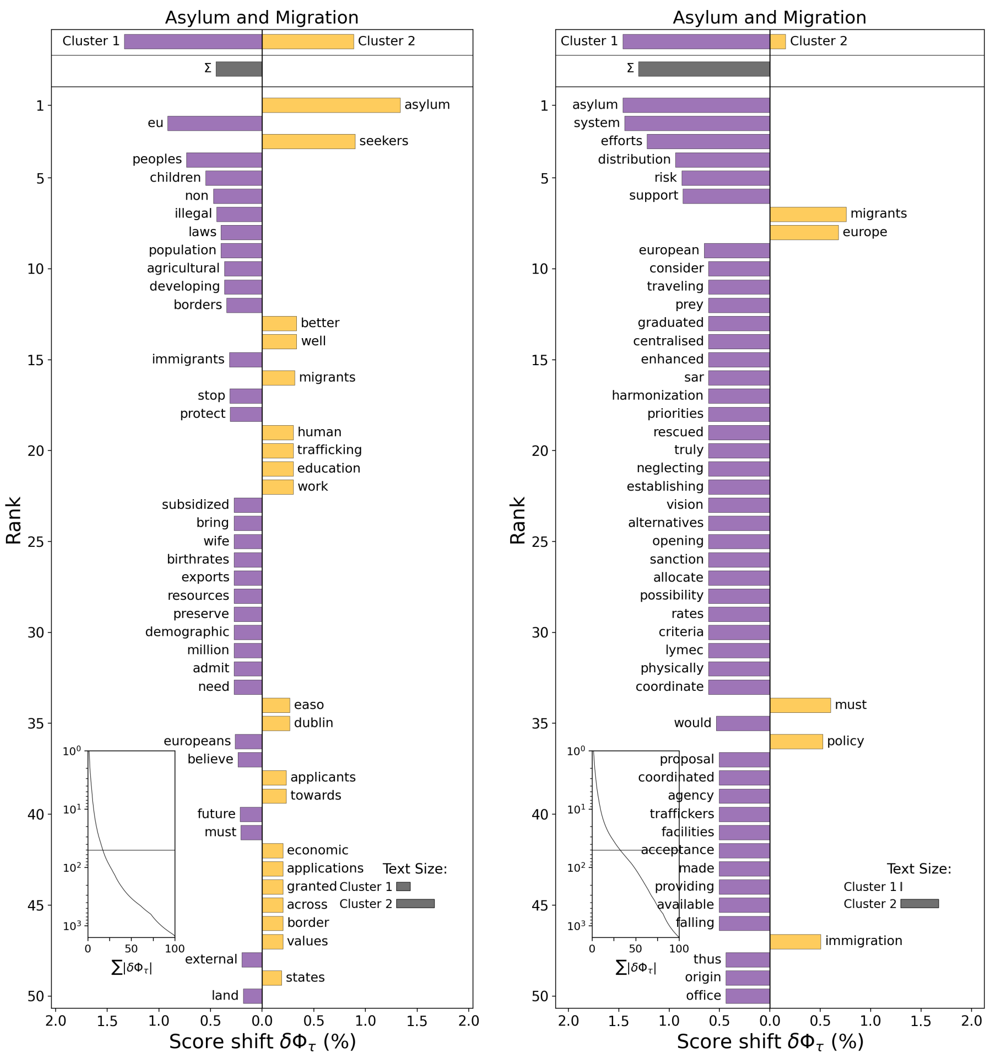

FIGURE 5.4: Word shift graphs of the final partition in the topic "Asylum and Migration" of the CFE data, for the two validated distance update schemes: Ward (left) and centroid (right).

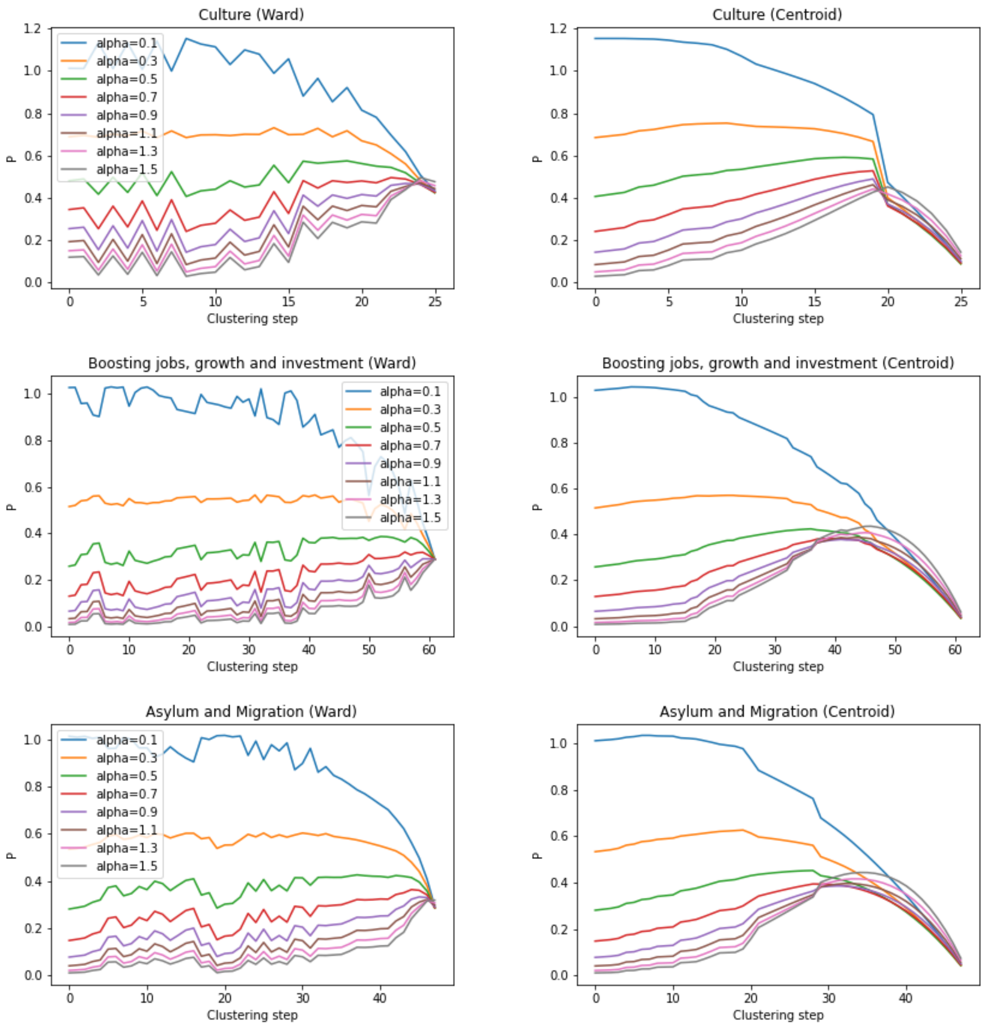

FIGURE 5.5: Polarisation throughout the clustering for different α in the topic CFE data, for Ward (left column) and centroid (right column) methods. From the upper to the lower row, the topics are "Culture", "Boosting jobs, growth and investment" and "Asylum and Migration".

Chapter 6

Conclusion

In this master's thesis, I have implemented a network-based polarisation measurement pipeline on bimodal graph representations of the Southern Women dataset (Davis, Gardner, and Gardner, 1941) and of data from the platform of the Conference on the Future of Europe (Parliament, Council, and Commission, 2021). The graphs were partitioned via hierarchical (SAHN) clustering with the Ward, centroid and the newly proposed poldist distance update schemes. The polarisation was computed on the final, dipole partition provided by the clustering by means of the axiomatic formula proposed by Esteban and Ray, 1994. Such a measure guarantees *a priori* certain intuitive properties of the notion of polarisation and thereby bears the potential to overcome the flaws of state-of-the-art structural alternatives (Salloum, Chen, and Kivelä, 2021), like dependence on unrelated network features such as average degree or degree distribution and false positives on randomised networks.

The goals of this master's thesis were:

1. Understand the current polarisation measurement paradigm, namely the network approach, and leverage the data science skills acquired in the master programme in an implementation thereof.

2. Implement and study the polarisation measure proposed by Esteban and Ray, 1994 in a network-based pipeline.

3. Study a novel distance update scheme for hierarchical clustering, poldist (see Fig. 2), based on the above polarisation measure.

4. Contribute to the body of research on polarisation evaluation in bimodal social networks.

All the above objectives were met, with the following specifications.

Objective 1 was partly met by the study of the state of the art (reflected in Chapters 1 and 2). Additionally, the implementation on the use cases brought insight on the network approach, namely on the interplay between the clustering and the computation of the polarisation: the study of the CFE data stresses the importance of choosing the step of the hierarchical clustering from which to draw the partition to evaluate, as in many cases the centroid method showed relevant structure only before the final step. Taking the final, dipole partition seemed the most natural choice for a first exploration of the polarisation pipeline, and it worked satisfactorily on the SW dataset, but results on the CFE data suggest that a more robust criterion would be to pick instead the partition that maximises polarisation. Note that such a criterion is made possible by the fact that (2.1) can account for an arbitrary number of clusters, which is a strength not shared by most alternatives (Salloum, Chen, and Kivelä, 2021).

Among the data science skills applied there is all the machinery more specific to network science, as well as the Jaccard distance and word shifts, standard tools in natural language processing that were instrumental for the evaluation of results.

Regarding *Objective 2*, the implementation of the measure proposed by Esteban and Ray, 1994 is validated by the reference SW dataset for the Ward and centroid methods, and it features a behaviour consistent with what is expected of a polarisation measure, as described by the authors. It remains to be seen in further work how the measure behaves with respect to different network features and when applied on randomised networks.

Objective 3 yielded negative resutls: the method poldist shows remarkable deviation from the reference results on the validation dataset. The poldist formula (Fig. 2) suggests this is due to the function of cluster sizes that multiplies centroid distance d_c: such a function is at least quadratic depending on polarisation sensitivity α, which means that clustering is prone to being dominated by the sizes. The poldist distance update scheme may therefore benefit from a factor that mitigates its strong tendency to favour merging small clusters, in a way similar to that of the denominator in the Ward method.

Finally, *Objective 4* is met by the contribution of the pipeline itself and the application use case, on the CFE data (Chapter 5).

The study of the CFE data failed to find any significant polarisation, which may be due to the low participation of the platform or to the absence of polarising trends altogether. However, results suggest that further analysis on extensions of the data featuring more participation (like an update, or including proposals in languages other than English) and/or evaluating partitions at stages of the clustering other than the final one may still reveal some structure. In particular, the latter would not only address the mentioned issue with the centroid method, but also allow a more accurate account of the global dataset, where we do not expect a dipole polarisation but a multipole fragmentation according to topics.

In the light the results of this master's thesis and of the above discussion, the proposed pipeline remains a promising candidate for the study of polarisation in bimodal social networks, and should be further explored. Future work on the CFE or other data and/or on the proposed pipeline may focus on any subset of the following paths:

- *A more complete evaluation of the clustering process*. I have mentioned that contrary to most structural polarisation measures, (2.1) allows for the evaluation of multipole systems, and such an analysis may yet reveal some relevant structure in the CFE data, especially in the global dataset. A particularly interesting partition to check systematically is the one that maximises polarisation.

- *Extensions of CFE data*. As discussed before, extensions featuring more participation (like an update, or including proposals in languages other than English) may reveal some currently unseen structure, and would in any case yield a more accurate representation of the system and allow to further explore the behaviour of the proposed pipeline with respect to network size.

- *Pipeline optimisation*. This point is of particular necessity if one is to consider greater network sizes, as the global analysis on the CFE data already takes $\sim 10h$ in my personal laptop. A most straightforward way to optimise the clustering in a dramatic way, now that the code is stable, is to re-write it using the Cython compiler.

- *Refinement of word shifts* (for CFE or any new text data). The text representations that provide the input for the word shifts do not necessarily have to be straight, frequency-based BOW representations: more sophisticated representations may be weighted, such as the *tf-idf* representation, or include a preprocessing step that accounts for features like synonymity in the considered text, e.g. through *topic modelling*.

- *Comparative study of diverse networks*. Salloum, Chen, and Kivelä, 2021 provide a criterion for evaluating the performance of network polarisation measurement pipelines based on studying their dependence on unrelated network features and their behaviour with randomised approximations of real networks, applied there to eight state-of-the-art candidates that reveal themselves as suboptimal. It will be important to see how our pipeline behaves in such terms.

- *Modifications of the clustering procedure*. Considering other distance update schemes for the hierarchical clustering, like a newly parameterised poldist to try and correct its present bias, or perhaps a different clustering method altogether. As I have mentioned, clustering methods are far from consensual and may influence the polarisation measurement independently of the polarisation measure used. Even if the current method does not seem flawed, it would be a useful check to compare it to others, both at the distance update and more fundamental levels.

Appendix A

Clustering algorithms implemented

Fig. 3 shows the adapted implementation of the generic clustering algorithm, while that of the nearest-neighbour clustering algorithm is featured in Figs. 4 and 5.

Figure 3 The generic clustering algorithm with polarisation computation. Adapted from (Müllner, 2011).

```
procedure GENERIC_LINKAGE(N, d)        ▷ N: input size, d: pairwise dissimilarities
    S ← (0, . . . , N − 1)
    L ← [ ]                                                ▷ Stepwise dendogram list
    P ← [ ]                                                      ▷ Polarisation list
    size[x] ← 1 for all x ∈ S
    for x in S \ {N − 1} do                 ▷ Generate the list of nearest neighbours.
        n_nghbr[x] ← argmin_{y>x} d[x, y]
        mindist[x] ← d[x, n_nghbr[x]]
    end for
    Q ← (priority queue of indices in S \ {N − 1}, keys are in mindist)
    for i ← 1, . . . , N − 1 do                                            ▷ Main loop.
        P ← [COMPUTEPOLARISATION]        ▷ With (2.1) and centroid distance.
        a ← (minimal element of Q)
        b ← n_nghbr[a]
        δ ← mindist[a]
        while δ ≠ d[a, b] do   ▷ Recalculation of nearest neighbours, if necessary.
            n_nghbr[a] ← argmin_{x>a} d[a, x]
            Update mindist and Q with (a, d[a, n_nghbr[a]])
            a ← (minimal element of Q)
            b ← n_nghbr[a]
            δ ← mindist[a]
        end while
        Remove the minimal element a from Q.
        Append (a, b, δ) to L.                      ▷ Merge the pairs of nearest nodes.
        size[b] ← size[a] + size[b]           ▷ Re-use b as the index for the new node.
        S ← S \ {a}
        for x in S \ {b} do                               ▷ Update the distance matrix.
            d[x, b] ← d[b, x] ← FORMULA(d[a, x], d[b, x], d[a, b], size[a], size[b], size[x])
        end for
        for x in S such that x < a do      ▷ Update candidates for nearest neighbours.
            if n_nghbr[x] = a then                         ▷ Deferred search; no nearest
                n_nghbr[x] ← b                           ▷ neighbours are searched here.
            end if
        end for
        for x in S such that x < b do
            if d[x, b] < mindist[x] then
                n_nghbr[x] ← b
                Update mindist and Q with (x, d[x, b])           ▷ Preserve a lower bound.
            end if
        end for
        n_nghbr[b] ← argmin_{x>b} d[b, x]
        Update mindist and Q with (b, d[b, n_nghbr[b]])
    end for
    return L, P                  ▷ The stepwise dendrogram and the polarisation list.
end procedure
```

Figure 4 The nearest-neighbour clustering algorithm with polarisation computation. Adapted from (Müllner, 2011).

```
procedure NN-CHAIN-LINKAGE(S, d)      ▷ S: node labels, d: pairwise dissimilarities
    L, P ← NN-CHAIN-CORE(N, d)
    Stably sort L and P with respect to the third column of L.
    L ← LABEL(L)                      ▷ Find node labels from cluster representatives.
    return L, P
end procedure

procedure NN-CHAIN-CORE(S, d)         ▷ S: node labels, d: pairwise dissimilarities
    S ← (0, ..., N − 1)
    chain = [ ]
    P = [ ]
    size[x] ← 1 for all x ∈ S
    while |S| > 1 do
        P ← [COMPUTEPOLARISATION]     ▷ With (2.1) and centroid distance.
        if length(chain) ≤ 3 then
            a ← (any element of S)    ▷ E.g. S[0]
            chain ← [a]
            b ← (any element of S \ {a})   ▷ E.g. S[1]
        else
            a ← chain[−4]
            b ← chain[−3]
            Remove chain[−1], chain[−2] and chain[−3]   ▷ Cut the tail (x, y, x).
        end if
        repeat
            c ← argmin_{x≠a} d[x, a] with preference for b
            a, b ← c, a
            Append a to chain
        until length(chain) ≥ 3 and a = chain[−3]   ▷ a, b are reciprocal
        Append (a, b, d[a, b]) to L                 ▷ nearest neighbours.
        Remove a, b from S
        n ← (new node label)
        size[n] ← size[a] + size[b]
        Update d with the information

            d[n, x] = d[x, n] = FORMULA(d[a, x], d[b, x], d[a, b], size[a], size[b], size[x])

        for all x ∈ S.
        S ← S ∪ {n}
    end while
    return L, P                       ▷ an unsorted dendogram and polarisation list
end procedure
```

(We use the Python index notation: *chain*$[-2]$ is the second-to-last element in the list *chain*.)

Figure 5 A union-find data structure suited for the output conversion in the nearest-neighbour clustering algorithm. Taken from (Müllner, 2011).

```
procedure LABEL(L)
    L′ ← [ ]
    N ← (number of rows in L) + 1                          ▷ Number of initial nodes.
    U ← new UNION-FIND(N)
    for (a, b, δ) in L do
        Append (U.EFFICIENT-FIND(a), U.EFFICIENT-FIND(b), δ) to L′
        U.UNION(a, b)
    end for
    return L′
end procedure

class UNION-FIND
    method CONSTRUCTOR(N)                                   ▷ N is the number of data points.
        parent ← new int[2N − 1]
        parent[0, . . . , 2N − 2] ← None
        nextlabel ← N                     ▷ SciPy convention: new labels start at N
    end method

    method UNION(m, n)
        parent[m] = nextlabel
        parent[n] = nextlabel
        nextlabel ← nextlabel + 1             ▷ SciPy convention: number new labels
consecutively
    end method

    method FIND(n)          ▷ This works but the search process is not efficient.
        while parent[n] is not None do
            n ← parent[n]
        end while
        return n
    end method

    method EFFICIENT-FIND(n)                         ▷ This speeds up repeated calls.
        p ← n
        while parent[n] is not None do
            n ← parent[n]
        end while
        while parent[p] ≠ n do
            p, parent[p] ← parent[p], n
        end while
        return n
    end method
end class
```

Appendix B

Network plots

B.1 Southern Women dataset

Figure B.1 shows the original bimodal social network.

B.2 Conference on the Future of Europe data

Figures B.2, B.3 and B.4 show the bimodal networks of the topics displayed in Chapter 5: respectively, "Culture", "Boosting jobs, growth and investment" and "Asylum and Migration". Node labels are omitted for better visibility.

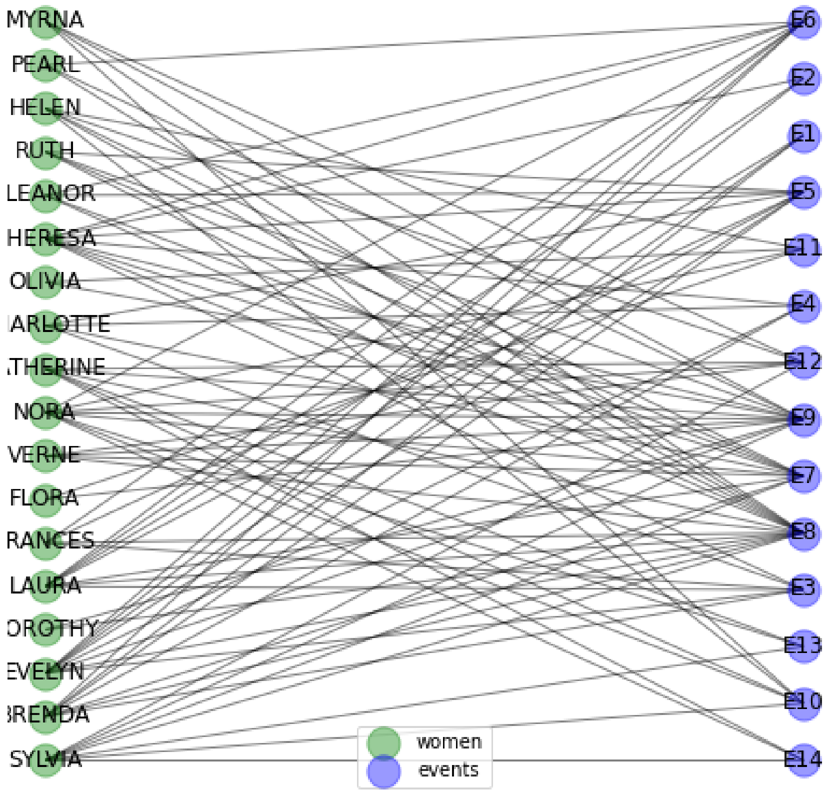

FIGURE B.1: Southern Women network.

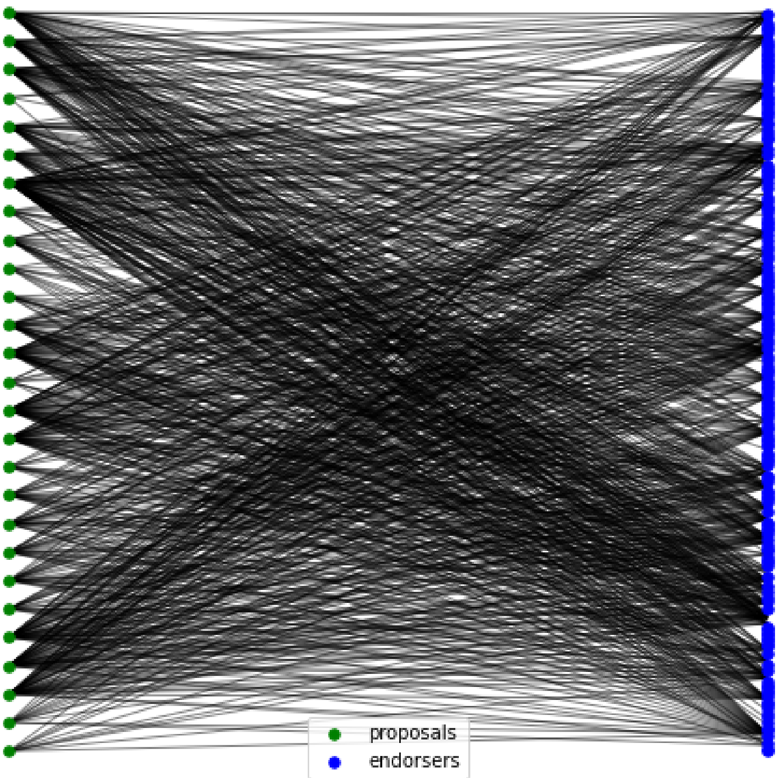

FIGURE B.2: Network of "Culture" topic.

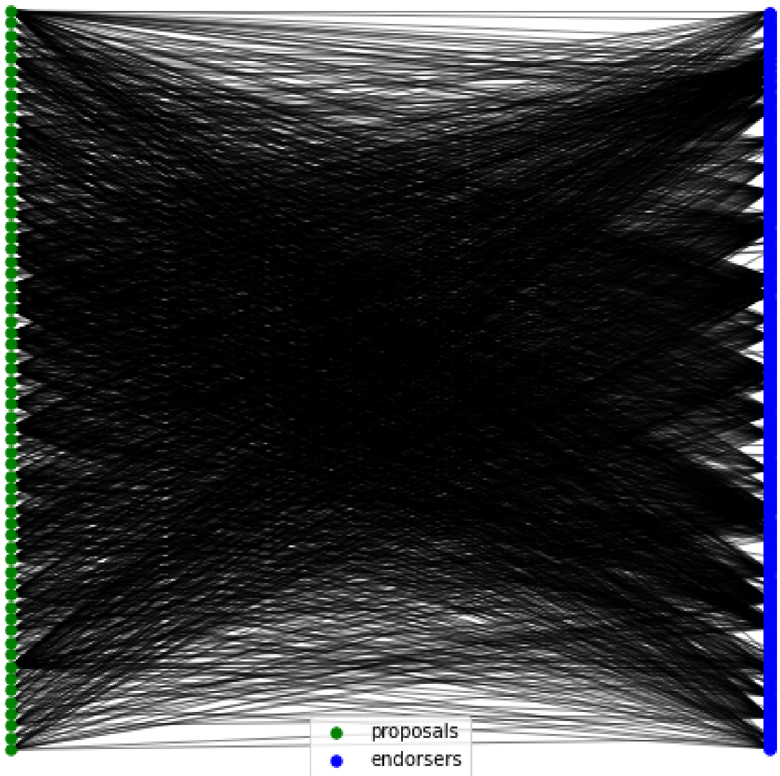

FIGURE B.3: Network of "Boosting jobs, growth and investment" topic.

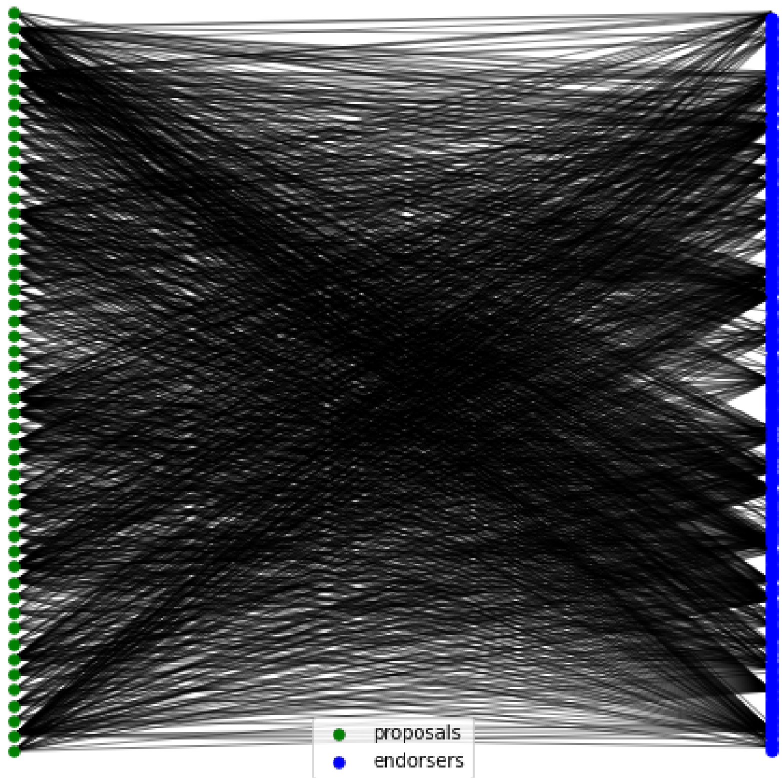

FIGURE B.4: Network of "Asylum and Migration" topic.

Appendix C

Polarisation in Breiger's partition of the SW dataset

Since Breiger, 1974 modifies the original data, namely eliminating events connected to all other events, and the resulting groups of women are disconnected, it is of interest to evaluate the polarisation of such a system for reference. One expects such a polarisation to be rather high (close to 1).

To that end, I reproduced the steps followed by Breiger[1] and applied our pipeline with centroid clustering on the resulting network, obtaining exactly Breiger's partition as the final configuration. Polarisation is indeed rather close to 1, and greater than when using the raw data: 0.68 vs the previous 0.53 in Table 4.1.

Note that, since in that case we have an even configuration $(8 + 8)$, what keeps the polarisation from its maximum is the distance between the two clusters: if the distance were 1 (the maximum) instead of 0.68 (the value obtained with the clustering), polarisation would also be 1.

The reason why the distance between Breiger's disconnected clusters is not maximal in our pipeline comes from the formula of the ϕ correlation (2.2): indeed, such a lack of connection in Breiger's case only means that the women in question do not share positive attendance values ($f_{11} = 0$), but the actual minimum in ϕ (i.e. the maximum in d) requires also that $f_{00} = 0$, i.e. that there is no overlap in positive *nor negative* values. Such is rarely the case even with Breiger's configuration, as we can see from $P2$ in (Breiger, 1974).

[1] Since we are looking for a kind of maximal reference value for the polarisation, I consider only the women in Breiger's two groups: Pearl and Dorothy, not belonging to either of them, I dismiss as noise.

Appendix D

Results for the Global CFE dataset

Given that α showed practically no influence on the final polarisation in the validation use case (Chapter 4) nor in the topic-specific analysis of the CFE data, and that the size of the data in this section makes the clustering rather slow, for the global dataset I consider only the case where $\alpha = 1$.

Figure D.1 shows the clustermaps of the global dataset obtained with the Ward and centroid methods, whose trees have a Robinson-Foulds distance of 1900. The corresponding word shift graphs are displayed in Fig. D.3, and the polarisation throughout the clustering of both validated distance update schemes is presented in Fig. D.2.

As in the analysis by topic, the clustermaps do not reveal any prominent dipole structure, as is also reflected in the low polarisation of the final partitions ($P \leq 0.20$, shown in Table D.1) and the centroid word shift graph. Remarkably though, in the Ward word shift graph one might argue for a division between ecological (in Cluster 2) and more institutional, political concerns (in Cluster 1), but the clustering and polarisation results and the low fraction of the JSD shift represented by the plotted words ($\approx 5\%$) suggest such a division is not representative of the whole dataset.

Additionally, as is apparent from Fig. D.2, the centroid method behaves as it often did in the topic-specific results, i.e. building one cluster that is then gradually augmented, thereby yielding a vanishing polarisation at the end. As mentioned in the analysis by topics, this behaviour hints towards the relevance of previous stages of the clustering, greater in this case because they would also allow to identify more clearly the expected fragmentation by topics.

TABLE D.1: Final polarisation for Ward and centroid clusterings in the global CFE dataset.

Dataset	P_W	P_c
Global	0.2	0.0

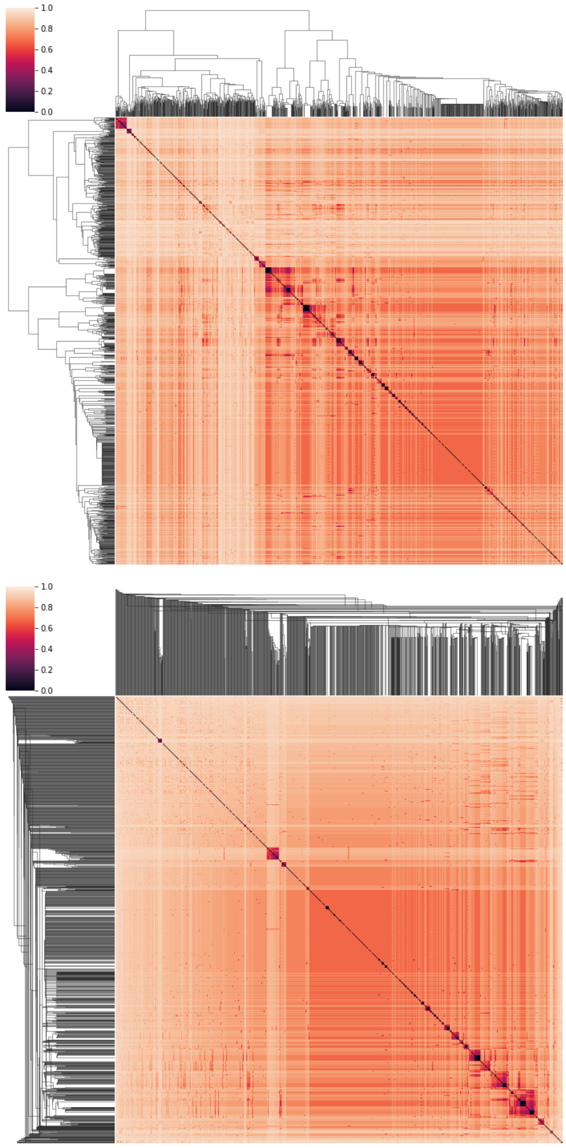

FIGURE D.1: CFE global clustermaps for Ward (top) and centroid (bottom) methods.

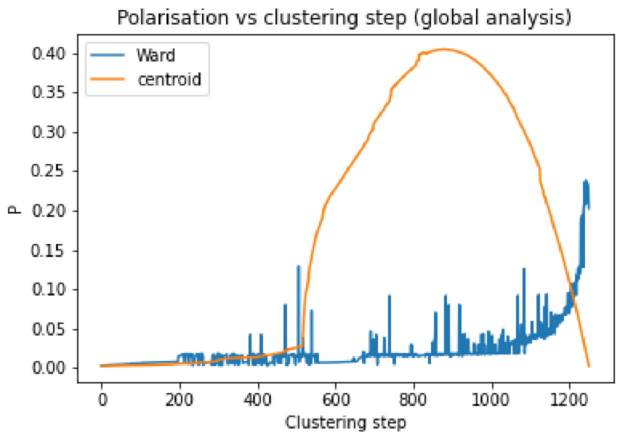

FIGURE D.2: CFE global polarisation for Ward and centroid methods.

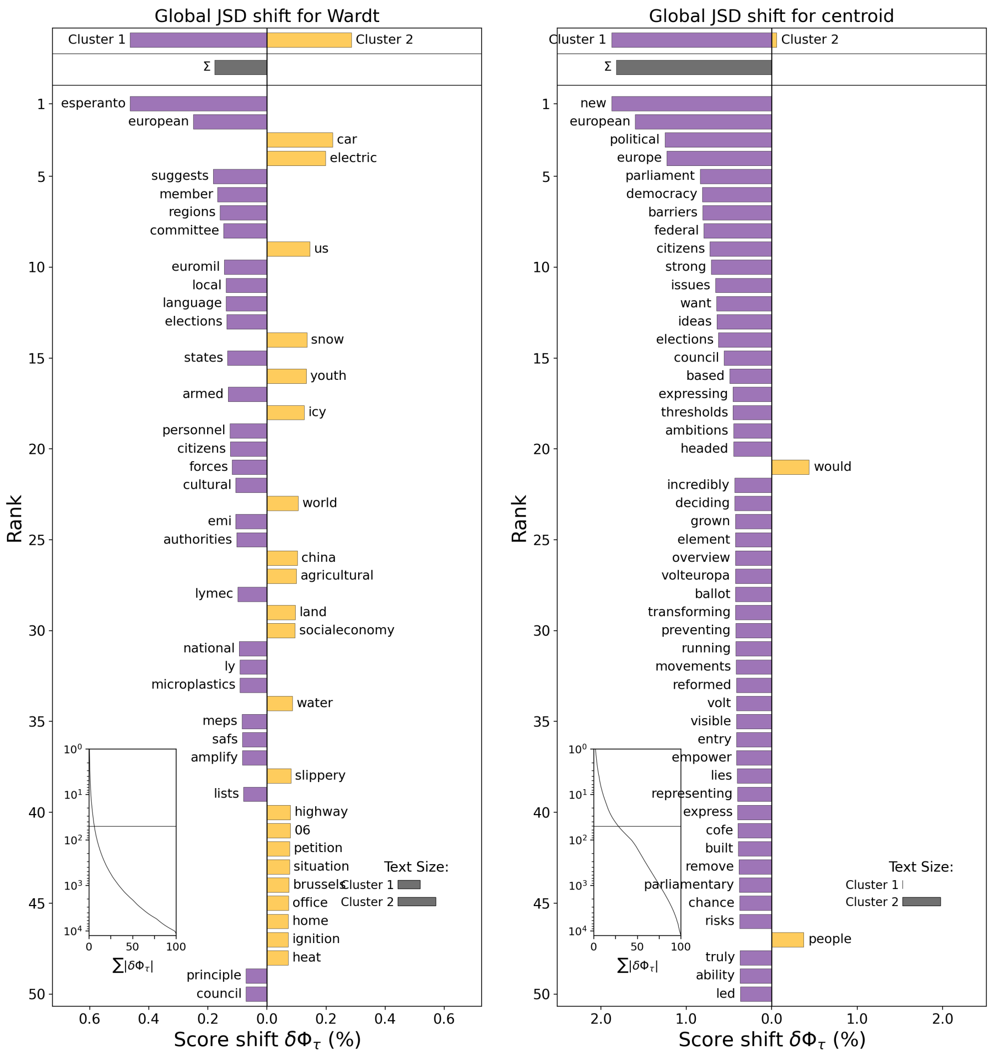

FIGURE D.3: Word shift graphs of the final partition in the global CFE dataset, for the two validated distance update schemes: Ward (left) and centroid (right).

Bibliography

Bagrow, James P (2012). "Communities and bottlenecks: Trees and treelike networks have high modularity". In: *Physical Review E* 85.6, p. 066118.

Baldassarri, Delia and Andrew Gelman (2008). "Partisans without constraint: Political polarization and trends in American public opinion". In: *American Journal of Sociology* 114.2, pp. 408–446.

Belcastro, Loris et al. (2020). "Learning political polarization on social media using neural networks". In: *IEEE Access* 8, pp. 47177–47187.

Bird Steven, Edward Loper and Ewan Klein (2009). *Natural Language Processing with Python*. O'Reilly Media Inc.

Breiger, Ronald L (1974). "The duality of persons and groups". In: *Social forces* 53.2, pp. 181–190.

Chen, Ted Hsuan Yun et al. (2020). "Polarization of Climate Politics Results from Partisan Sorting: Evidence from Finnish Twittersphere". In: *arXiv:2007.02706*.

Darwish, Kareem (2019). "Quantifying Polarization on Twitter: The Kavanaugh Nomination". In: *Social Informatics*. Cham: Springer International Publishing, pp. 188–201. ISBN: 978-3-030-34971-4.

Davis, A, BB Gardner, and MR Gardner (1941). *Deep South; a social anthropological study of caste and class.* University of Chicago Press.

Demszky, Dorottya et al. (2019). "Analyzing Polarization in Social Media: Method and Application to Tweets on 21 Mass Shootings". In: *Proceedings of the 17th Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL)*, pp. 2970–3005.

DiMaggio, Paul, John Evans, and Bethany Bryson (1996). "Have American's social attitudes become more polarized?" In: *American journal of Sociology* 102.3, pp. 690–755.

Esteban, Joan-Maria and Debraj Ray (1994). "On the measurement of polarization". In: *Econometrica: Journal of the Econometric Society*, pp. 819–851.

Fiorina, Morris P and Samuel J Abrams (2008). "Political polarization in the American public". In: *Annu. Rev. Polit. Sci.* 11, pp. 563–588.

Fortunato, Santo and Darko Hric (2016). "Community detection in networks: A user guide". In: *Physics reports* 659, pp. 1–44.

Gallagher, Ryan J et al. (2021). "Generalized word shift graphs: a method for visualizing and explaining pairwise comparisons between texts". In: *EPJ Data Science* 10.1, p. 4.

Garimella, Kiran et al. (2018). "Quantifying controversy on social media". In: *ACM Transactions on Social Computing* 1.1, pp. 1–27.

Guerra, P.H. et al. (Jan. 2013). "A measure of polarization on social media Networks-Based on community boundaries". In: *Proceedings of the 7th International Conference on Weblogs and Social Media, ICWSM 2013*, pp. 215–224.

Guimera, Roger, Marta Sales-Pardo, and Luís A Nunes Amaral (2004). "Modularity from fluctuations in random graphs and complex networks". In: *Physical Review E* 70.2, p. 025101.

Homans, George C (1950). "The human group." In:

Hout, Michael and Christopher Maggio (2020). *Immigration, Race, and Political Polarization*. DOI: 10.31235/osf.io/p7q2w. URL: osf.io/preprints/socarxiv/p7q2w.

Jaccard, Paul (1912). "The distribution of the flora in the alpine zone. 1". In: *New phytologist* 11.2, pp. 37–50.

Jones, David R (2001). "Party polarization and legislative gridlock". In: *Political Research Quarterly* 54.1, pp. 125–141.

Krackhardt, David and Robert N Stern (1988). "Informal networks and organizational crises: An experimental simulation". In: *Social psychology quarterly*, pp. 123–140.

Lancichinetti, Andrea, Filippo Radicchi, and José J Ramasco (2010). "Statistical significance of communities in networks". In: *Physical Review E* 81.4, p. 046110.

Makridis, Christos and Jonathan T Rothwell (2020). "The real cost of political polarization: evidence from the COVID-19 pandemic". In: *Available at SSRN 3638373*.

Mason, Lilliana (2015). ""I disrespectfully agree": The differential effects of partisan sorting on social and issue polarization". In: *American Journal of Political Science* 59.1, pp. 128–145.

Morales, Alfredo Jose et al. (2015). "Measuring political polarization: Twitter shows the two sides of Venezuela". In: *Chaos: An Interdisciplinary Journal of Nonlinear Science* 25.3, p. 033114.

Mukerjee, Subhayan, Sílvia Majó-Vázquez, and Sandra González-Bailón (2018). "Networks of audience overlap in the consumption of digital news". In: *Journal of Communication* 68.1, pp. 26–50.

Müllner, Daniel (2011). "Modern hierarchical, agglomerative clustering algorithms". In: *arXiv preprint arXiv:1109.2378*.

Parliament, European, European Council, and European Commission (2021). *Joint declaration on the Conference on the Future of Europe*. European Union.

Pearson, Karl (1900). "I. Mathematical contributions to the theory of evolution.—VII. On the correlation of characters not quantitatively measurable". In: *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character* 195.262-273, pp. 1–47.

Rabab'ah, Abdullateef et al. (Apr. 2016). "Measuring the Controversy Level of Arabic Trending Topics on Twitter". In: *2016 7th International Conference on Information and Communication Systems (ICICS)*. DOI: 10.1109/IACS.2016.7476097.

Rand, William M (1971). "Objective criteria for the evaluation of clustering methods". In: *Journal of the American Statistical association* 66.336, pp. 846–850.

Robinson, David F and Leslie R Foulds (1981). "Comparison of phylogenetic trees". In: *Mathematical biosciences* 53.1-2, pp. 131–147.

Rumshisky, Anna et al. (2017). "Combining Network and Language Indicators for Tracking Conflict Intensity". In: *Social Informatics*. Ed. by Giovanni Luca Ciampaglia, Afra Mashhadi, and Taha Yasseri. Cham: Springer International Publishing, pp. 391–404. ISBN: 978-3-319-67256-4.

Salloum, Ali, Ted Hsuan Yun Chen, and Mikko Kivelä (2021). "Separating Controversy from Noise: Comparison and Normalization of Structural Polarization Measures". In: *arXiv preprint arXiv:2101.07009*.

Virtanen, Pauli et al. (2020). "SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python". In: *Nature Methods* 17, pp. 261–272. DOI: 10.1038/s41592-019-0686-2.

Wallace, David L (1983). "A method for comparing two hierarchical clusterings: comment". In: *Journal of the American Statistical Association* 78.383, pp. 569–576.

Zhang, Pan and Cristopher Moore (2014). "Scalable detection of statistically significant communities and hierarchies, using message passing for modularity". In: *Proceedings of the National Academy of Sciences* 111.51, pp. 18144–18149.

Zhou, Jack (2016). "Boomerangs versus javelins: how polarization constrains communication on climate change". In: *Environmental Politics* 25.5, pp. 788–811.