Measuring Cultural Dynamics Through the Eurovision Song Contest

David García and Dorian Tanase
Chair of Systems Design, ETH Zurich, Weinbergstrasse 56/58
8092 Zurich, Switzerland
dgarcia@ethz.ch, dtanase@alumni.ethz.ch

January 14, 2013

Abstract

Measuring culture and its dynamics through surveys has important limitations, but the emerging field of computational social science allows us to overcome them by analyzing large-scale datasets. In this article, we study cultural dynamics through the votes in the Eurovision song contest, which are decided by a crowd-based scheme in which viewers vote through mobile phone messages. Taking into account asymmetries and imperfect perception of culture, we measure cultural relations among European countries in terms of cultural affinity. We propose the Friend-or-Foe coefficient, a metric to measure voting biases among participants of a Eurovision contest. To validate how this metric represent cultural affinity, we designed a model of a random, biased Eurovision contest. Simulations of this model show how our metrics can detect negative affinities and serve as an estimator for positive affinities. We apply this estimator to the historical set of Eurovision contests from 1975 to 2012, finding patterns of asymmetry and clustering in the resulting networks. Furthermore, we define a measure of vote polarization that, when applied to empirical data, shows a sharp increase within countries of the EU during 2010 and 2011. As a result, we measure how the recent political decisions of EU states influence the way their citizens relate to the culture of other EU members, leading to stronger cultural biases in the way they vote in the Eurovision song contest.

Keywords: *Cultural dynamics; international networks; social simulation*

1 Introduction

How do cultures evolve? How do they influence each other? These questions are not only central to human sciences, like anthropology or ethnology, but play a major role in politics, economics, and international relations. Among the scientific tools to study human cultures, agent-based modeling provides quantitative insights to culture and opinion dynamics [5]. The original works of Axelrod [2, 3] motivate how computational modeling can be used to understand the evolution of human cultures. These agent-based models, while being essential simplifications of a more complicated phenomenon, allow us to draw the conditions for the emergence of macroscopic social behavior [4, 5], such as polarization [14] or clustering [30], as well as to predict future social phenomena [36, 39]. Furthermore, modeling and simulation can open new questions that drive future research, allowing –in an ideal case– a deeper understanding of human societies through multidisciplinary research [15].

As well as sociological theories need to be empirically testable, computational models of social behavior need to be formulated over assumptions that can be verified against empirical data. When such behavior is objectively measurable, e.g. economic decisions [6] or voting [8, 26], datasets can be produced in a way such that we can directly measure the state of a human. On the other hand, when a model includes subjective elements, such as emotions or beliefs, measuring the internal states and dynamics of a human becomes a cumbersome task. As an example, Axelrod's model introduces the internal state of an agent as a vector of cultural dimensions, or opinions, which change according to certain rules [2]. To validate this kind of dynamics, we need to be able to measure the subjective internal state of a human, and how it changes when interacting with others. While survey data can shed light on opinions and culture [23], there is a subconscious component of cultural behavior that cannot be encoded in words [42]. Nevertheless, this component can be indirectly measured, for example through physiological responses [18, 25], or through online traces such as expression biases [17], and behavioral patterns in computer mediated interaction [16].

Quantitative models need to be validated on the dynamics of individual states, but are often aimed to reproduce macroscopically observable collective behavior. When addressing cultures or societies as a whole, issues of data availability become critical. It can be expensive to query large amounts of individuals, limiting the application of subjective reports and surveys. In addition, approaching these questions through experimental studies suffers additional problems. For example, experiments cannot reproduce *natural exposure* in the context of culture and popularity [31], limiting the representativeness of any experimental study. The emerging field of computational social science [20, 28] aims at overcoming these limitations, studying human behavior through the statistical analysis of large-scale datasets. Such datasets, when available, offer the opportunity to validate the macroscopic behavior explained by computational models of social interaction. Following the example of Axelrod's model, its validation requires to measure how whole cultures change in time, as well as the distances between different cultures.

In this article, we aim at providing a way to measure the relations between cultures through their voting patterns in a set of song contests, in particular looking for biases in the way they evaluate each other. This way, we are measuring the dynamics of culture i) at a large-scale level usually unreachable for independent research, and ii) measuring subjective biases that are not explicitly expressed by the studied individuals. It is of special relevance to measure these kind of relations in a timely manner, in order to address possible changes in the relation between pairs of countries. The political decisions of a country, the results of sport events, or the current state of the economy might impact the evolution of the public opinion of one society towards another. For the case of Europe, the policies of the European Union regarding the debt crisis might have an impact on the *"state of the union"*, or how countries within the EU perceive each other. Studying data with a time component, we measure how these events play a role in the manifestation of cultural relations, with the aim of providing a macroscope tha measures the state of the union of Europe at large.

2 The Eurovision Song Contest

In this article, we present our study of the relations between European countries through the set of results of the Eurovision Song Contest, an annual competition held among the country members of the European Broadcasting Union. Every year, each participating country chooses a representative artist to compete by performing a song, which is included in a live event broadcasted simultaneously in the whole Europe. After the performance of each participant, voting countries gather televotes and jury votes [34], creating a local ranking of songs from other contestants. Afterwards, each voting country publicly announces which other countries receive points from 1 to 8, 10, and 12, according to their local rankings. The winner of the contest is the country with the song that accumulated the highest amount of points. Extensive and detailed descriptions about the contest, its rules, and its history can be found elsewhere [19, 40].

While the contest rules and participating countries have changed over the years, this contest offers a timely source of cultural evaluations across most European countries. Eurovision has been subject of substantial research, up to the point of the usage of the term *"eurovisiopsephology"* [19], defined as the study of the results of the votes casted in the Eurovision song contest. Initial research focused on the possible existence of voting clusters or alliances [40, 41], generally due to geographical locations, diaspora effects, language, and religious similarities [38]. Further studies combined network analysis with simulations of maximally random contests, revealing how Eurovision results have high clustering [12], which results in voting blocks [19, 34], and higher chances to win for countries depending on their position in the voting network [9, 37].

Since 2004, all the countries participating in the contest choose their votes according to televoting, a method that uses phone calls and mobile phone messages of viewers to decide how a country votes. Since 2009, these televotes were combined with some expert judges, turning Eurovision in an experimental ground to compare popular and expert choices. Recent studies show the statistical changes due to televoting [34], while older works measure how expert judges chose their votes according to song quality rather than cultural biases [22]. Either way, the results of this contest highlight the stable cultural relations between countries [38], where voting trades or game theoretical decisions do not seem to play a role [21].

2.1 Controversies and applications

Recently, Google set up a Eurovision predictor based on search queries, leading to the correct prediction in 2009 and 2010[1] . This was discontinued in 2011, after a contested prediction result between Lena, the previous German winner who was competing again, and the Irish participants called Jedward[2]. The outcome of this prediction failed, as both countries were defeated by Azerbaijan by more than 100 points. In addition, seems that users were exploiting the search engine to try to push their country higher in the prediction[3] as if searching for your representative would increase its chance to win. This reaction to a prediction mechanism shows how social systems, as complex adaptive systems, can change their behavior due research results, leading to the invalidation of prediction tools or even to self-fulfilling predictions.

Our approach does not aim to predict contest outcomes or to reveal voting alliances, but to use Eurovision as a social macroscope for the relations across European countries. Initial results show how Eurovision outcomes can predict international trade [11, 27], which motivates the measuring of the cohesion of European countries and the EU through Eurovision [34]. Popular culture and mass media criticize the contest organization, claiming that some countries are treated as European only in Eurovision[4], as a limitation for a "Europeaniziation process"[24]. In addition, the contest rules and results are periodically claimed to be unfair, biased[5] , or even farcical[6], portraying the contest as a European popularity survey rather than an artistic competition. In this article, we precisely aim to measure these biases as relevant quantities, focusing on the political, social, and cultural component of the contest rather than on its artistic one.

2.2 Exploring Eurovision data

We gathered the whole historical set of Eurovision results from `Wikipedia`[7], which contains a page for each edition of the contest, and from the official website of the contest[8].

For each year, we count with a matrix with the values $p_{v,c}$, where each entry corresponds to the amount of points given by a country c_v to the competing song of another country c_c. As explained before, $p_{v,c}$ is contained in the set $\{0, 1, 2, 3, 4, 5, 6, 7, 8, 10, 12\}$, and chosen according

[1] calmyourbeans.wordpress.com/2012/05/22/no-google-eurovision-predictor-this-year/

[2] wiwibloggs.com/2011/05/07/google-prediction-jedwards-lead-grows-denmark-and-estonia-climbing-update-2/10942/

[3] thedailyedge.thejournal.ie/google-trends-predict-eurovision-near-miss-for-jedward-130899-May2011/

[4] "I'm sick of being European just on Eurosong" https://www.youtube.com/watch?v=IK8fVHNk0oM

[5] http://news.bbc.co.uk/2/hi/uk_news/wales/south_east/3719157.stm

[6] http://news.bbc.co.uk/2/hi/entertainment/6654719.stm

[7] For an example of a contest result page, see: http://en.wikipedia.org/wiki/Eurovision_Song_Contest_2012

[8] http://www.eurovision.tv/page/history/year

to the ranking of televotes and jury votes. Our dataset comprises the whole set of results of Eurovision editions from 1957 to 2012, including 11775 voting relations between 51 country members of the European Broadcasting Union.

The straightforward approach to understand this data is to look into the voting network formed every year [12], where nodes are participating countries. A directed edge $c_v \rightarrow c_c$ connects two nodes if c_v assigned that year a nonzero amount of points to c_c. Edge weights are assigned to be the amount of points given by the vote, $p_{v,c}$. The left panel of Figure 1 shows this network for the edition of 2008, with edge darkness according to weight, and node darkness proportional to the final score $s_c = \sum_{c_v} p_{v,c}$ of each country c_c in the contest. The topological properties of these networks have been widely explored, finding symmetrical relations, triadic clustering, and highly connected blocks that map to geographically close, and culturally related countries [19, 37, 40].

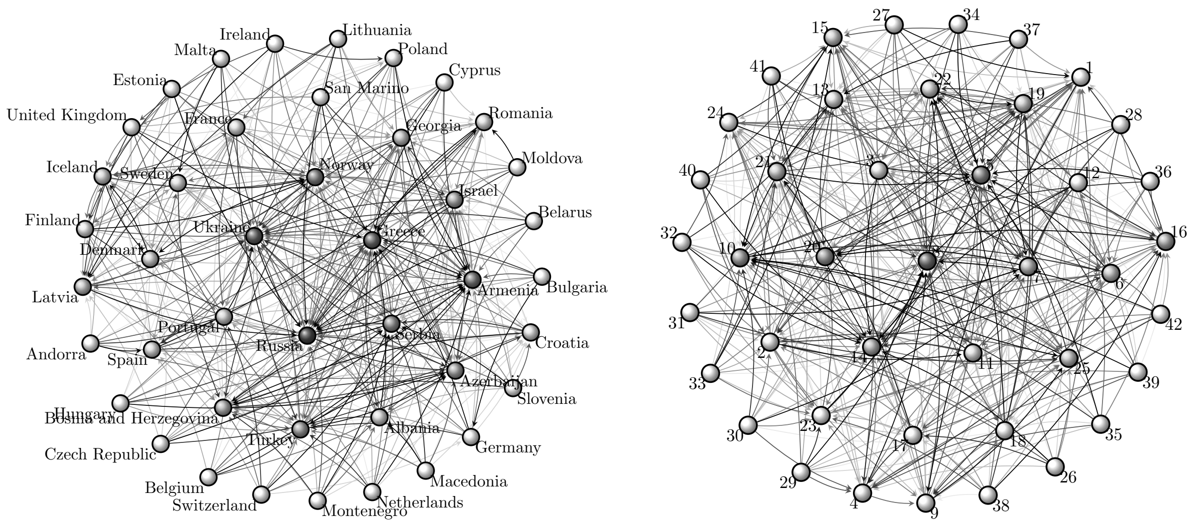

Figure 1: Left: network of votes for the 2008 edition of Eurovision. Nodes are participating countries, with color darkness proportional to the final score of the country. Directed edges represent the votes given by one country to another, with darker color according to the amount of points given by the vote. Right: simulation results with the same size as the 2008 contest, visualized in the same manner as the left side.

Visual inspection of this network, as shown in the left panel of Figure 1, reveals a significant heterogeneity in node darkness. This corresponds to the large deviation of final scores usually present in this contest. Initial editions of the contest had multiple draws, so the voting scheme was changed to the current one in 1975, in order to encourage the selection of a single winner. The resulting heterogeneity is relevant to test the existence of winner-takes-all effects as in cultural markets [35], and product reviews [29]. To do so, we calculated the relative score $s'_c = \frac{s_c}{T}$, where

$T = \sum_{c_c} s_c$ is the total amount of points given in an edition of the contest, which depends on the amount of countries participating in a given year. This way we can aggregate all participant scores since 1975, as shown in the histogram of Figure 2. Similarly to the cultural markets mentioned above, the distribution of s' shows a large variance, and positive skewness. On the other hand, the log-log histogram shown in the inset of Figure 2 allows us to notice that there are no scaling relations, probably due to the finite size of the contest. We can say that the contest has a large variance of final scores, yet these do not allow arbitrarily large values, as opposed to previous experience in popularity analysis. We will use these final scores to compare individual votes with final results, as explained below.

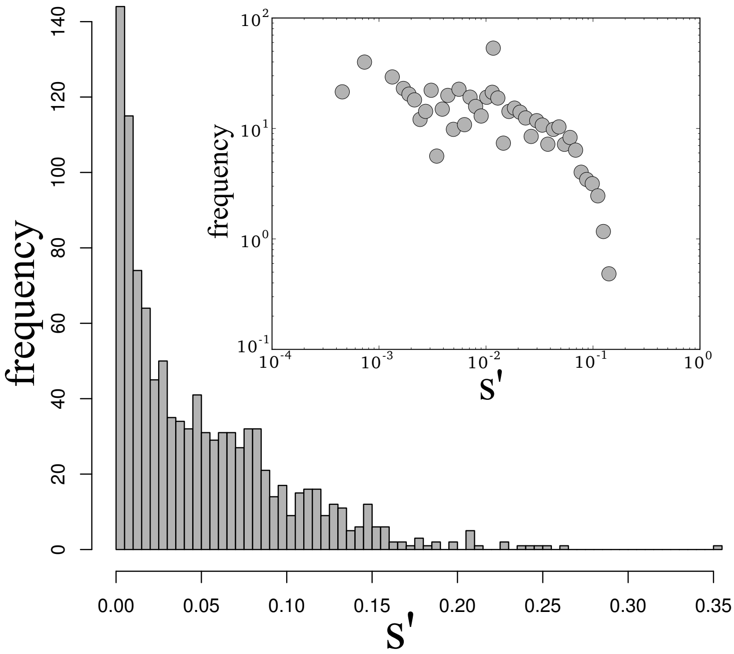

Figure 2: Relative score s' distribution for contest results from 1975 to 2012. Inset: log-log version of the distribution. We use this empirical distribution as input for our simulations.

3 Measuring cultural relations through Eurovision

3.1 Perception of culture

Most agent-based models of culture dynamics include agent interactions based on their internal states, usually depending on the distance between the values of their cultural features. While valid to reproduce the emergence of opinion groups and cultures [32], there is still ample room to validate and empirically test the existence of this kind of dynamics. The existence of cultural dimensions was first introduced by Hofstede [23], in a study of surveys across different countries. These dimensions were detected by means of dimensionality reduction on the survey responses, and have been applied in numerous studies about culture [13], including a study on Eurovision [21]. On the other hand, measuring culture through surveys has clear limitations [1], in particular in the interpretation of the meaning of the results of dimensionality reduction.

Apart from dimensional structures, a key component in models of culture dynamics is the set of rules that determine which agents interact and how. While these rules can represent spontaneous events of influence between cultures, in other scenarios work as a mechanism in which agents perceive the state of others. In a realistic setup, the perception of cultural differences might be constrained by imperfect communication, and path dependencies like historical events or stereotypes. Such phenomena can shape the way culture is perceived across a society, leading to new structures to take into account in future models. For example, with the minimal assumption that humans can only perceive a set of dimensions from another culture, the perceived distance between cultures could have asymmetric properties. In the schema of Figure 3, we sketch two cultures with binary feature vectors of five dimensions. If the left one can only perceive the three first features of the other, its perceived Hamming distance would be 1, as they just differ in the third feature. At the same time, if the right one can only perceive the last three features of the left one, the perceived distance would be 3, leading to asymmetric perception of cultural differences.

Another possibility is that each society might have a reference point, i.e. an expected or "acceptable" maximum cultural distance towards another. This would lead to the existence of negative cultural relations, which would be a plausible explanation for multiple international conflicts present in History. This possibility is commonly ignored when taking into account cultural distances in discrete spaces, and might very well be a property of realistic cultural dynamics.

Given possible asymmetries and signed values, we will define *cultural affinity* of a society towards another as *"the differences perceived by the members of a society in relation to the culture of another society, evaluated as a comparison with a reference point"*. Cultural affinity takes maximum values for very close cultures, and negative values towards very different cultures. By analyzing Eurovison results, we want to explore the two properties that differentiate cultural distance from cultural affinity, i.e. i) the presence of asymmetry in the perceived differences between cultures, and ii) the existence of negative and positive cultural evaluations among European cultures.

3.2 The Friend-or-Foe coefficient

To measure cultural affinity, we need to define a way to estimate it from the raw Eurovision scores of our dataset. For this, we define the Friend-or-Foe (FoF) coefficient of country c_v towards

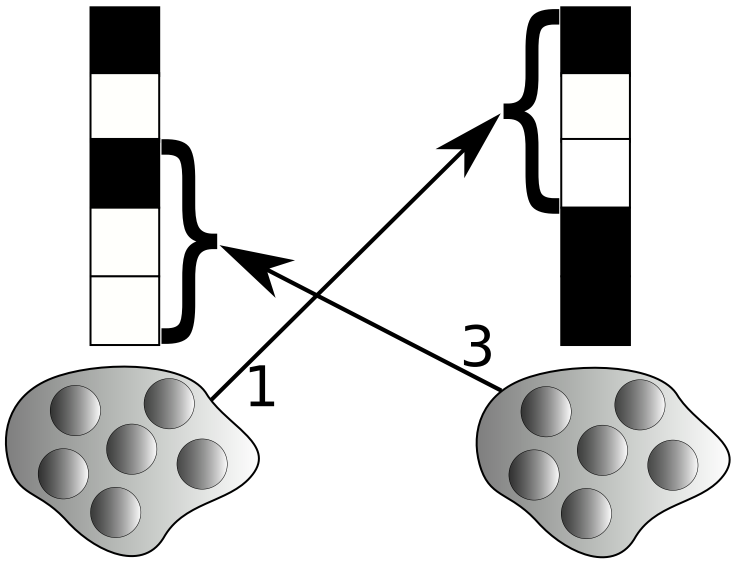

Figure 3: Schema with a possible scenario of feature-based cultures under imperfect perception. Each culture is composed of a set of agents with the same values of the cultural vector, shown over them. Their asymmetric relation is a result of partial perception of the cultural features of the other.

country c_c, as estimated from a particular edition of the contest:

$$Fof(c_v, c_c) = \frac{p_{v,c}}{12} - \frac{s_c - p_{v,c}}{12(N-2)} \quad (1)$$

where $p_{v,c}$ are the points assigned to c_c by c_v, s_c is the final score of c_c, and N is the total amount of countries voting in the studied edition of Eurovision. The first term of the right hand side of Equation 1 represents a normalized value of the score given by c_v to c_c, ranging from 0 for no points given, to 1 when 12 points were assigned to c_c. The second term corrects for the final score of c_c in the whole contest, calculating the total amount of points given by *other* countries different than c_v. The maximum value of this score is $12(N-2)$, as one country cannot vote itself and we have already subtracted c_v from the calculation.

We designed the Friend-or-Foe coefficient to measure the overvoting or undervoting bias from a country to another, correcting for "song quality" as estimated by the final contest result [21]. This way, we aim at removing the effects of the artistic component of the contest, highlighting the political or cultural biases that are commonly claimed to exist in Eurovision. If a country c_v assigns 12 points to c_c, while all the others assign 0, then $FoF(c_v, c_c) = 1$, which would be the maximum value of an overvoting bias. If c_v assigns 0 points to c_c but all the other countries assign 12, then $FoF(c_v, c_c) = -1$, representing the maximally negative Friend-or-Foe coefficient given the contest rules.

After this definition, we need to assess if the FoF is a valid measure to estimate the real cultural affinity of one coutry towards another, as described above. In the following, we perform a case study of the values of the FoF for different years between pairs of countries with known cultural similarities, as well as countries with explicit conflicts. We continue by defining and simulating a model for Eurovision contests, including heterogeneous song quality, and an underlying network of cultural affinities. By numerical analysis of such model we want to measure the quality of this approximation, as the contest rules might distort and limit the quality of the FoF as an estimator for cultural affinity.

3.3 Dyadic relations over time

For any pair of countries c_1 and c_2 we can calculate the FoF coefficients between them in each contest in which they participated together. The values of $FoF(c_1, c_2)$ and $FoF(c_2, c_1)$ might depend on effects that influence Eurovision votes, such as language and geographical proximity. In Figure 4 we present the FoFs for some years between pairs of countries with known similarities and conflicts. In the following, we make a case study on how we can use the FoF as an approximation for measuring the relations between European cultures, unifying the already known biases in previous literature.

- **Cultural proximity.** The standard example for the expression of cultural similarity in Eurovision is Cyprus and Greece [12, 19, 34]. Figure 4A shows the FoF between these two countries from 2002 until 2012. Both values are positive in each edition of the contest, never dropping below 0.3. This way, the FoF is consistent with previous research [38], where cultural proximity and language are major components in the voting trends of Eurovision. Note that, while we have a value of $FoF(Cyprus, Greece)$ for every year, some values of $FoF(Greece, Cyprus)$ are missing. This is due to the fact that, since 2002, Greece has always competed in the final, but Cyprus did not qualify for the final round every year.

- **Asymmetric effects.** One of the relations we want to explore is the possible asymmety of cultural affinities between the inhabitants of two countries. The FoF pair of Turkey and Armenia, shown in Figure 4D is a clear example of an asymmetric relation between countries. $FoF(Turkey, Armenia)$ keeps a significantly positive value, representing Armenian diaspora living in Turkey. This same effect of 'patriotic voting' was suggested for Turkish migrants across Europe [38], and our FoF coefficient reflects it in this case. On the other hand, $FoF(Armenia, Turkey)$ is significantly low and mostly below 0. This negative relation is a possible expression of negative relations due to historical conflicts between both countries.

Another example of asymmetric relations is the Greece and Germany pair, shown in Figure 4C. While $FoF(Germany, Greece)$ is significantly positive, the $FoF(Greece, Germany)$ is significantly negative. A plausible explanation might be the economic and political relations between both countries in the last years, but this asymmetry seems to stretch to past times before the creation of the Euro. As an alternative, another possible explanation is the existence of perceived cultural asymmetry, in a way such that German inhabitants feel a closer to Greek culture, or a diaspora effect caused by Greek immigrants in Germany.

- **Negative relations.** The second type of cultural relation we want to explore is the possibility of negative relations between pairs of countries. Couples of countries with explicit territorial conflicts show this way symmetric negative FoFs, as shown in Figure 4B and 4F. Turkey and Cyprus have diplomatic conflicts regarding the status of Northern Cyprus[9], and Armenia and Azerbaijan are still officially in war since the Nagorno-Karabakh conflict[10]. This negativity is evident in their FoF coefficients, as these countries consistently avoid voting each other. But political conflicts might not be the only reason for negative FoFs, as illustrated in the example of Greece and Norway on Figure 4E. We do not have any plausible assumption for this negative pair rather than the mere cultural distance between one of the northmost and one of the southmost European countries. This way, the FoF coefficient would be a valid estimator for possible negative cultural relations, in the case that two countries consider each other too far in terms of cultural distance.

3.4 A model for biased Eurovision contests

The above case study shows that the Friend-or-Foe coefficient consistently represents previous findings and types of symmetric, asymmetric, as well as positive and negative cultural relations. Yet this qualitative study does not allow us to determine the actual relation between the FoF and any underlying, quantifiable, cultural affinity. In the following, we propose a model to simulate Eurovision contests under heterogeneous song quality and the influence of cultural affinity. Our aim is to study the relation between the FoF as manifested through contest simulations, given a network of affinities between countries.

Our model receives as an input a network in which nodes represent countries, connected by edges with weights that represent cultural affinities, i.e. a measure that takes high positive values for culturally similar countries, and negative values for highly dissimilar cultures. Thus, an edge $e_{v,c}$ connecting the node of country c_v to the node of country c_c has a weight $w_{v,c}$ that measures the cultural affinity of c_v towards c_c. This network is composed of two subnetworks, with participant

[9] `http://en.wikipedia.org/wiki/Cyprus_dispute`
[10] `http://en.wikipedia.org/wiki/Nagorno-Karabakh_War`

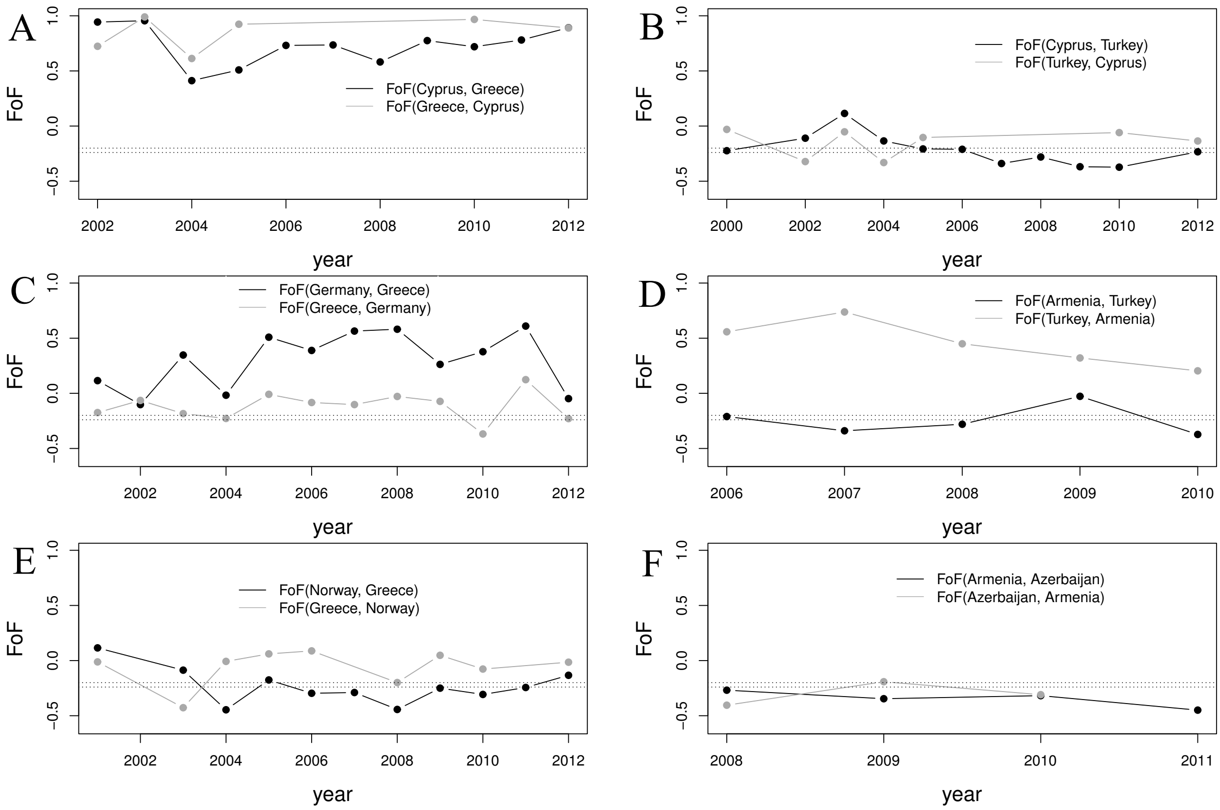

Figure 4: Time evolution of Friend or Foe coefficients between selected pairs of countries. The example of Greece and Cyprus show a symmetric positive relation, possibly due to cultural similarity. On the other hand, the example of Greece and Norway shows a symmetric negative relation, possibly due to cultural differences. The case of Greece and Germany shows an asymmetric relation, where Germany shows significantly larger FoF towards Greece than vice versa. The right column shos possible political conditions of the FoF, where country pairs with regional conflicts have symmetric negative FoF (Cyprus-Turkey, Armenia-Azerbaijan), or diaspora effects create asymmetries like the Turkey-Armenia pair. The horizontal dashed lines represent the baselines calculated in Section 3.5.

and only voting countries. The subnetwork between participants is fully connected, directed, weighted network, and the only voting countries are connected to all the participant ones by unidirectional weighted links.

At the beginning of a simulation, we assign a quality value q to each participant, sampled uniformly at random from the distribution of rescaled scores s'_c shown in Figure 2. This value is an approximation of the artistic quality of a song [21], which is supposed to determine the final outcome 7of the contest. A simulation of the model is composed of two steps:

1. Each country c_i constructs the ranking of the other participant countries, by computing a value r_j that is a function of the weight $w_{i,j}$ of the edge $e_{i,j}$, and the quality of the song, q_j. Our initial assumption is that the form of this function is

$$r_j = f(w_{i,j}, q_j) = \alpha q_j + (1 - \alpha) w_{i,j} \tag{2}$$

   This way, $f(w, q)$ is a linear combination of w and q, with a proportionality factor α. This function represents the combination of jury votes and televotes, as empirical studies show that the jury is more influenced by the artistic quality of a song than the televotes [22], which seem to be driven by geographical and cultural biases. The current rules of the contest give the same weight to both votes, so we will choose $\alpha = 0.5$ for our simulations.

2. Given the rankings of each node, the agents cast their votes in order, assigning them according to the voting scheme of Eurovision, awarding a set of points from 1 to 8, then 10, and finally 12 to the neighbor with the top value in the ranking.

As an initial assumption, we take edge weights $w_{i,j}$ sampled from a uniform distribution with $w_{min} = -1$ and $w_{max} = 1$. This way, simulations of our model can equally span positive and negative values of cultural relations, with possible asymmetries.

The output of the model is an artificial voting result that can be compared with the real world data. The right panel of Figure 1 shows the outcome of a simulation, displayed with the same method as the empirical data of the left panel. Both networks show similar properties of degree distribution and density, but the purpose of our simulations was not to reproduce topological properties of the contest results. In the following section, we present an analysis of the relation between simulated FoFs and edge weights, exploring how to estimate cultural affinities through Eurovision results.

3.5 Numerical analysis of the Friend-or-Foe coefficient

The countries participating in Eurovision cannot distribute the points they give in an arbitrary way, they should follow rules that constrain these values as explained above. Under this restriction, it follows to ask about the possible distortion that these rules create on the estimation of cultural affinities, as calculated by the Friend-or-Foe coefficient. After a simulation of our model for biased Eurovision contests, we can analyze the relation between the weight of an edge $w_{i,j}$, and the FoF calculated on the simulation result. Figure 5 shows a scatter and barplot of the FoF in a simulation of our model as a function of edge weight. There is a linear relation for weights above zero, while the FoF seems to keep a constant value for negative weights. To quantify this, we calculated a linear regression of the form

$$FoF(e_{i,j}) = b_1 w_{i,j} + b_0 \tag{3}$$

for values of $w_{i,j}$ above 0. The dashed diagonal line of Figure 5 shows the regression results, where $b_1 = 0.93593$, and $b_0 = -0.39460$, with $R^2 = 0.75$ and p-values below 10^{-15}. The horizontal dashed line shows the mean FoF of negative weights, -0.31

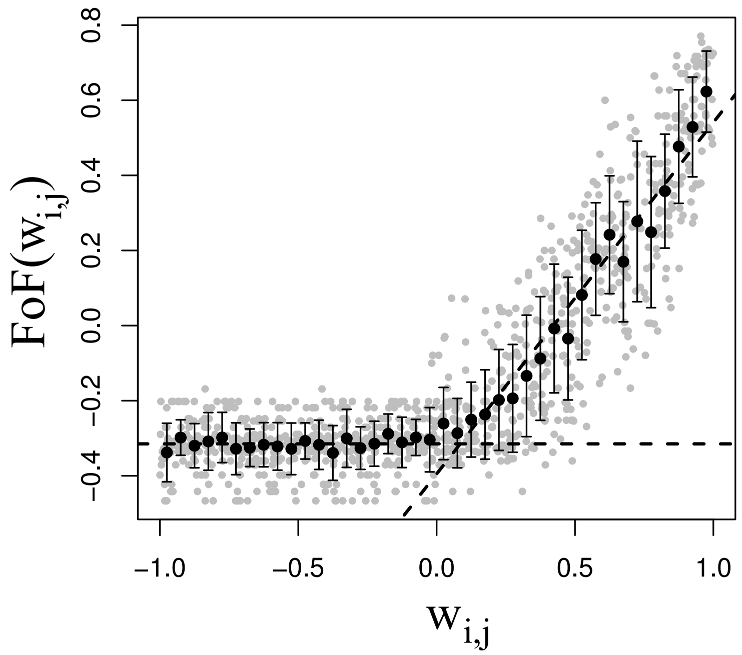

Figure 5: Scatter plot of cultural affinities and FoF for a simulation of the model. The barplot shows mean and standard deviation of FoF in bins of width 0.05. The dashed line shows piecewise regression results, showing the proportional effect of positive weights on the FoF, versus the negative constant baseline of negative weights.

The relation shown in Figure 5 suggests that the Friend-or-Foe coefficient is a good estimator for positive cultural affinities, if it is above certain value. Below a threshold, the FoF can be used as a way to detect negative affinities, but it does not allow us to estimate their magnitude. To apply these results to empirical data from Eurovision we need to find the value Θ_{neg} below which we can conclude that a FoF represents a negative relation, and the value Θ_{pos} over which we can estimate a positive positive relation through the FoF. For this decision problem, we define the following classification rule:

$$\begin{array}{llll} & if \quad FoF(e_{i,j}) > \Theta_{pos} & then & w_{i,j} > 0 \\ else & if \quad FoF(e_{i,j}) < \Theta_{neg} & then & w_{i,j} < 0 \\ & & else & w_{i,j} = 0 \end{array} \tag{4}$$

Each possible set of values for Θ_{pos} and Θ_{neg} defines two discriminants that map the FoF coefficients to the sign of the affinity between pairs of countries $w_{i,j}$. Given this prediction, we can classify each pair given its FoF, having certain amounts of correctly and incorrectly classified relations. This way, for a given value of Θ_{pos}, we compute the values of precision and recall [33] over our simulation as

$$Precision = \frac{tp}{tp + fp} \qquad Recall = \frac{tp}{tp + fn} \tag{5}$$

where tp and fp are the amounts of true and false positives, i.e. pairs with $FoF(e_{i,j}) > \Theta_{pos}$ that had $w_{i,j} > 0$ for tp, and $w_{i,j} < 0$ for fp. The amount of false negatives fn is the count of pairs with $FoF(e_{i,j}) < \Theta_{pos}$ that had $w_{i,j} > 0$. In an equivalent manner, we compute precision and recall for Θ_{neg}, inverting the rules of Equation 5. To independently assess the quality of the discriminant given a value for each threshold, we compute the F_1 scores combining precision and recall in a unidimensional metric [33]

$$F_{pos} = 2 \cdot \frac{Precision_{pos} \cdot Recall_{pos}}{Precision_{pos} + Recall_{pos}} \qquad F_{neg} = 2 \cdot \frac{Precision_{neg} \cdot Recall_{neg}}{Precision_{neg} + Recall_{neg}} \tag{6}$$

Figure 5 shows the value of F_{pos} and F_{neg} as a function of the value of the thresholds Θ_{pos} and Θ_{neg}. Both classification rules have maximum F_1 values for $\Theta_{pos} = -0.2$ and $\Theta_{neg} = -0.24$. These F_1 are above 0.8, and $\Theta_{pos} > \Theta_{neg}$, providing us with a valid and consistent method to determine the sign of a cultural relation given its FoF in a contest. The dashed horizontal lines of Figure 4 are placed at -0.2 and -0.24, to guide us in deciding negative affinities between couples of countries. Using the inverse form of Equation 3, we can formulate the following method to estimate the affinity of one country for another from the FoF in a given contest:

$$\begin{array}{llll} & if \quad FoF(e_{i,j}) > \Theta_{pos} & then & w_{i,j} = \dfrac{FoF(e_{i,j}) - b_1}{b_0} \\ else & if \quad FoF(e_{i,j}) < \Theta_{neg} & then & w_{i,j} < 0 \\ & & else & w_{i,j} = 0 \end{array} \tag{7}$$

where the values of b_0, b_1, Θ_{pos}, and Θ_{neg} have been estimated from computer simulations as explained above.

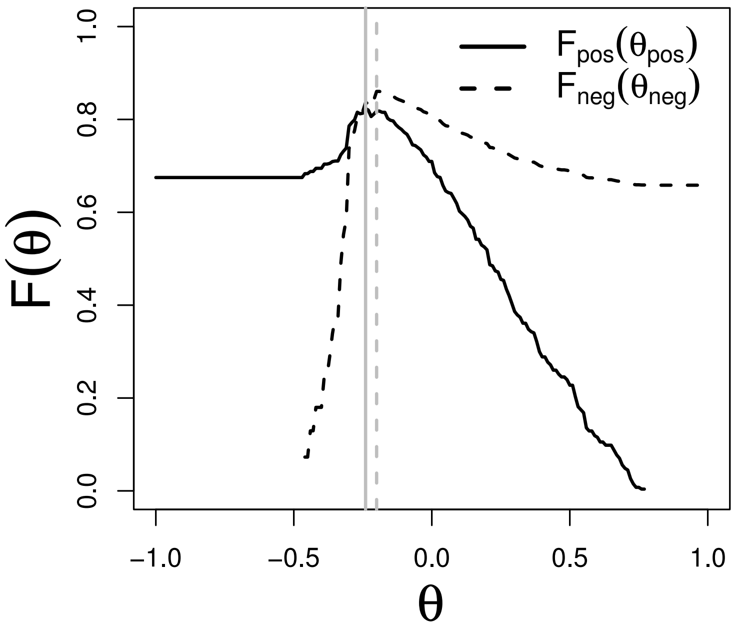

Figure 6: F values for negative and positive weight classifications as a function of the chosen threshold. Vertical lines indicate the maximum points for both, at $\Theta_{pos} = -0.2$ and $\Theta_{neg} = -0.24$

4 Analysis of Cultural Affinity in Eurovision

4.1 The network of affinity among European countries

The analysis of simulated biased contests allowed us to define the estimator of Equation 7, which lets us detect negative affinities between countries, and estimate values of positive affinities. In this section, we will apply this rule to the FoF coefficients across several contests, analyzing the properties of the network of cultural affinity between European countries.

For each pair of countries c_c and c_v, we have the FoF coefficients in the contests where they appeared $FoF^t(e_{c,v})$, where t denotes a year where both countries participated. We apply Equation 7 to convert this time series of FoFs into a time series of affinity estimations $w^t_{c,v}$. First, we summarize the history of cultural affinities of c_c towards c_v with two metrics, the average positive affinity

$$P_{c,v} = \frac{1}{T} \sum_t \Theta[w^t_{c,v}] \cdot w^t_{c,v} \tag{8}$$

and the ratio of negative affinities

$$N_{c,v} = \frac{1}{T} \sum_{t} (1 - \Theta[w_{c,v}^{t}]) \qquad (9)$$

where T is the length of the studied time period, and $\Theta[w]$ is the Heaviside step function, with value 1 if $w > 0$ and 0 otherwise. These two metrics allow us to aggregate the historical relations between European countries, as estimated from the Eurovision results in a certain time period. Each metric defines a network between countries, one with average positive affinities as weights, and another one with frequencies of negative affinites as weights. We show these two networks for the period 2007-2012 in Figure 7, with edge width and darkness according to edge weight ($P_{c,v}$ and $N_{c,v}$ respectively), and node darkness according to node strength, i.e. the sum of the weights of incident edges.

The left panel of Figure 7 shows the network of positive relations displayed with the weighted ARF algorithm of the Cuttlefish Network Workbench [11], which highlights node clusters. This confirms previously found voting coalitions [12, 19, 34], including a Balkan cluster of former Yugoslavian republics, a Scandinavian cluster, and the pairs Cyprus-Greece and Moldova-Romania. In addition, Turkey has a significantly high strength, which confirms previous conjectures of the effect of Turkish diaspora [38]. Among all the countries studied in this period, the strongest one is Belgium, which to our knowledge was never found to play a special role in this contest. Its possible mediator role in the European Union, combined with immigration from all across Europe might be the reason that drives Belgium to this privileged position.

The right panel of Figure 7 shows the network with negative affinity ratios as edge weights, with node darkness proportional to strength calculated as the sum of weights of incident edges. Using the same layout algorithm as for its positive counterpart, there is no clear community structure in this negative network. A possible explanation for this lack of clustering is the balanced nature of signed networks [10], which naturally avoids negative triangles. The strongest nodes in this network are Greece, Azerbaijan, Turkey, and Armenia, all involved in a series of territorial and political conflicts with each other. While we quantitatively test for the reasons for the structure of these networks, the FoF coefficient allows us to quantify the cultural affinities between European countries.

4.2 Time evolution of cultural affinities

While the networks of Figure 7 show the time aggregates of cultural relations among European countries, we want to measure how these networks evolve in time. To do so, we compute the time

[11] `www.cuttlefish.sourceforge.net`

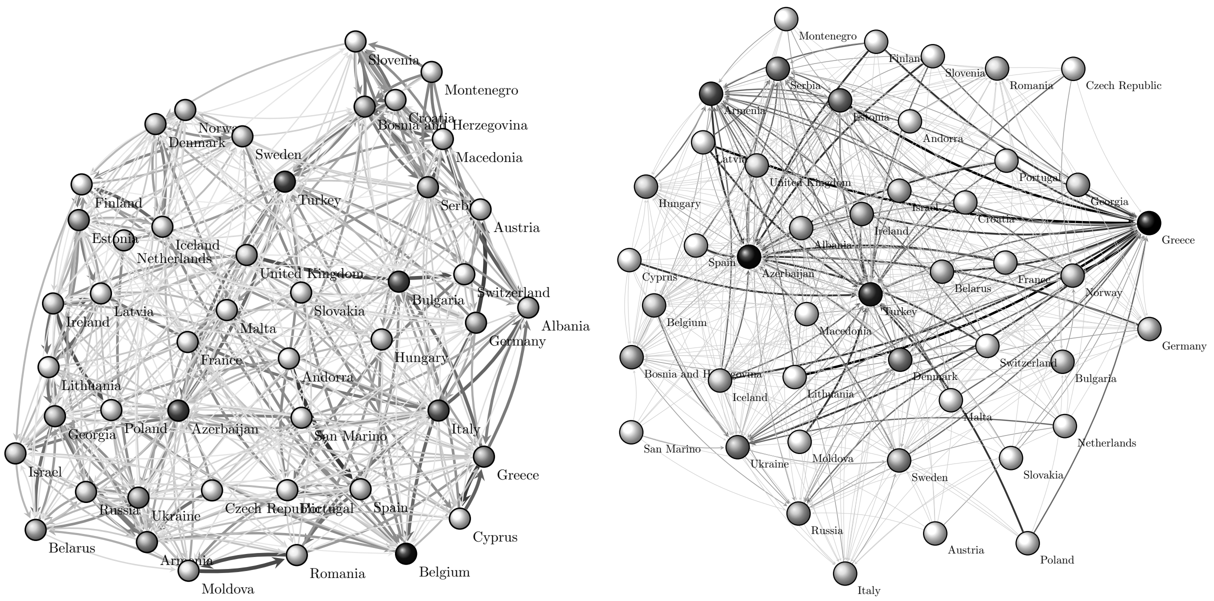

Figure 7: Left: network of positive FoF averaged between the years 2007 to 2012. Edge color, weight, and width are proportional to the average FoF from one country to another. Node darkness corresponds to the total sum of incoming edge weights. Right: Network of counts of negative FoF for the same time period as in the left. Edge width, weight, and color is proportional to the amount of negative instances of FoF from one country to another, and node darkness is proportional to the sum of incoming edge weights.

series of network asymmetry, defined as

$$asymmetry(t) = \frac{1}{E} \sum_{c,v} |w_{c,v}^t - w_{v,c}^t| \tag{10}$$

where E is the total amount of edges, i.e. pairs of voter-competing countries in the network. The rationale behind this metric is to estimate how strong are the typical asymmetries illustrated in Section 3.3, by computing the average distance between reciprocal pairs of estimated affinities.

In addition, we also calculate the time series of network polarization as

$$polarization(t) = \sqrt{\frac{1}{E} \sum_{c,v} (w_{c,v}^t - \langle w^t \rangle)^2} \tag{11}$$

which is essentially the standard deviation of estimated affinities across all the edges in the network. This polarization metric takes higher values when cultural and political biases are strong, in comparison to the artistic component of the contest. If all countries agreed on the best songs in the same manner, the polarization would have a value close to zero, as the resulting FoFs would lead to moderate values of w as calculated in Equation 7.

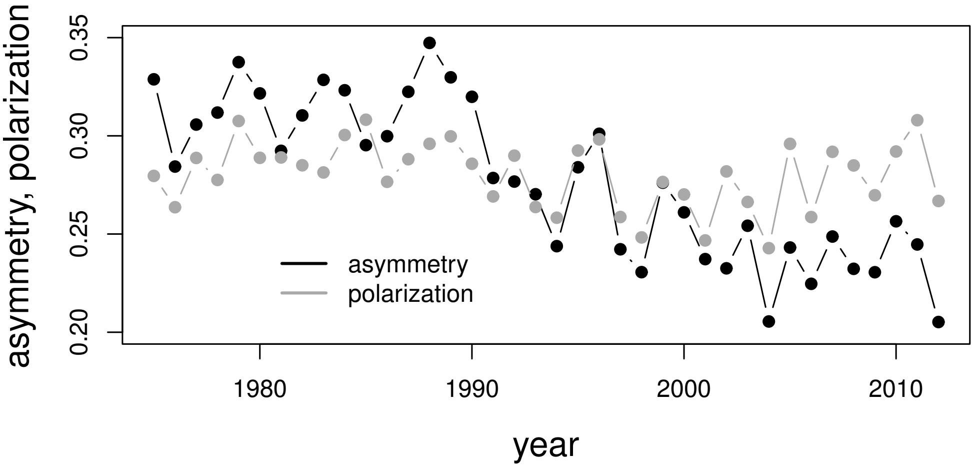

Figure 8: Time series of asymmetry and polarization in the Eurovision editions since 1975. Asymmetry shows a negative trend in the last two decades, but keeping always above 0.2.

As mentioned in Section 2.1, the organization of the Eurovision song contest has been accused of having an increasing level of unfairness and lack of artistic content. If that is the case, voting clusters should decrease asymmetry, and political and popularity effects should create a heterogeneity in the votes that increases polarization. To test those hypotheses, we created the time series of asymmetry and polarization for each contest edition, shown in Figure 8. These were calculated on the affinity estimations over the each year from 1975 to 2012. From 1990, there is a negative trend in asymmetry that leads to more symmetrical contest outcomes towards 2012, while polarization does not seem to follow this pattern. This decreasing asymmetry indicates that countries tend to vote each other as the other one voted them before. This increasing reciprocity can be built up along many interactions, where the already known cultural clusters form and strengthen, due to voting dependencies as well as the changes on the political landscape of Europe. On average, the overall asymmetry of the contest has a value of 0.27, giving us a quantization of the cultural reciprocity between the countries of Europe.

Figure 8 also shows a lack of a significant trend in polarization, meaning that, at this high level of aggregation, we cannot detect any pattern of political influence in the contest outcome. This does not rule out the possibility of increasing polarization for groups of countries, or around individual participants, which we discuss in the following section.

5 Measuring International Relations in the EU

One of the main motivations for studying Eurovision data is its representativeness of the whole Europe, giving the possibility to study the relations among countries. It is particularly relevant to study if the events of the financial crisis across members of the European Union can be traced through our statistical measures, looking for a macroscope of the state of the union in the EU. In this section, we focus on two subsets of countries of the EU: the 12 founding members of the Eurozone, and the EU-15, which is the set of members of the European Union since 1995. For each subset we compute the value

$$P_G(t) = \frac{1}{|G|} \sum_{(c,v) \in G} \Theta[w^t_{c,v}] \cdot w^t_{c,v} \tag{12}$$

which is the average positive affinities between the studied countries G.

Figure 9 shows the time series of P for the countries in the Eurozone, in the EU-15, and in the whole European Broadcasting Union. In 2010 and 2011, P shows a significantly higher value for the Eurozone and EU-15 countries than in previous years, where it was not easily distinguished from the rest of the contest. To explore this comparison between a subset of countries and the whole Europe, we computed the Z-score of P as $z_G(t) = \frac{P_G(t) - P(t)}{\sigma_{pos}(t)}$, where $P(t)$ is the mean and $\sigma_{pos}(t)$ is the standard deviation of positive weights for all Europe in year t. Figure 10 shows the time evolution of $z(t)$, for both the Eurozone and the EU-15. We can appreciate the fluctuation above zero in 2001 and 2011, as well as a fluctuation below zero in 2007.

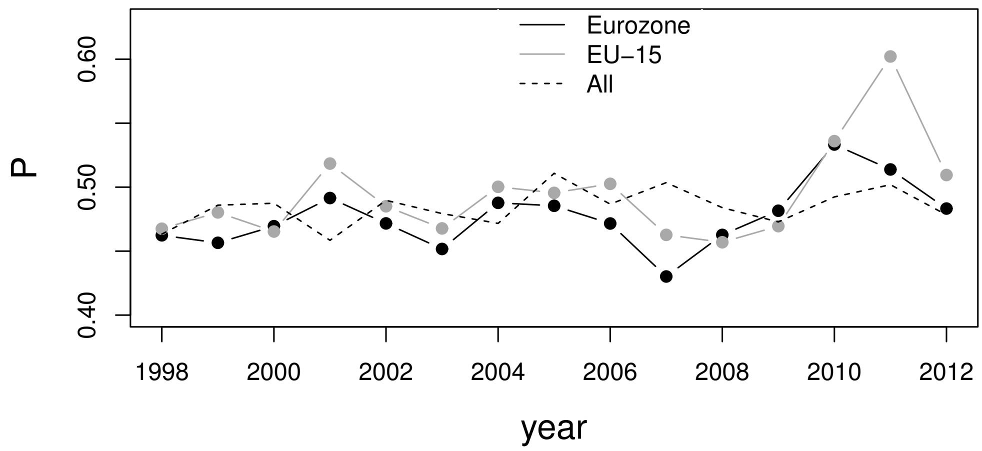

Figure 9: Time series of P from 1998 to 2012, for countries in the Eurozone, in EU-15, and all countries.

These Z-scores keep below 0.3, showing an anecdotal fluctuation rather than a consistent result. Nevertheless, they indicate the presence of a pattern that could be measured with topological

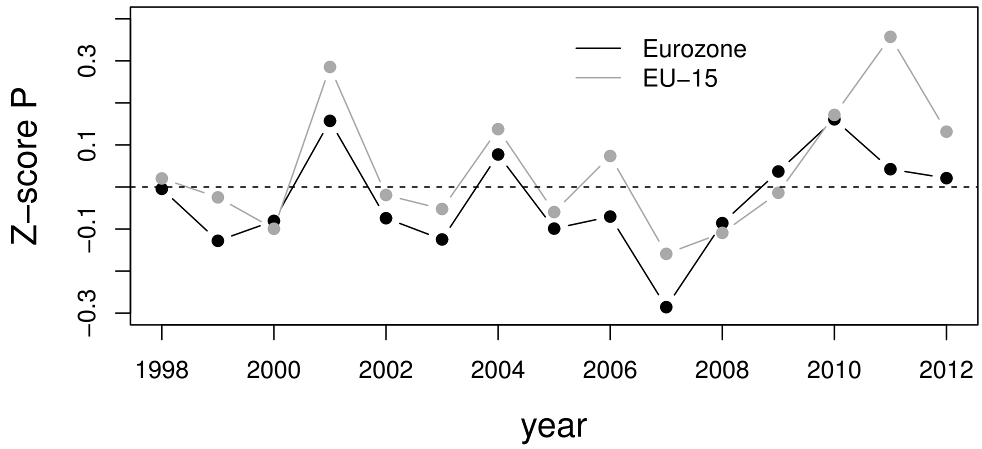

Figure 10: Time series of the Z-score of P per year since 1998, comparing the distribution of P in the whole contest versus the value of P for countries in the Eurozone and in EU-15.

metrics that also include negative weights, such as the asymmetry and polarization of Equation 10 and Equation 11 respectively. Focusing on the EU-15, we calculated $asymmetry(t)$ and $polarization(t)$ for the subnetwork formed by the 15 countries of EU-15, which are shown from 1998 to 2012 in Figure 11. These two metrics show a strong fluctuation in 2010 and 2011, the first two years of the debt and austerity measures across the studied countries. The increase in asymmetry reflects the lower cohesion among countries in the EU-15, which voted outside their usual geographical clusters. This effect is also present when analyzing the Eurozone, clearly influencing countries related to the common currency. The increase in polarization shows that their votes where less homogeneous and their biases led to strong FoFs, reflecting an influence exogenous to the artistic component of the contest.

Figure 12 shows the EU-15 subnetworks for the contest editions between 2009 and 2012. Blue edges represent positive affinities with a width and darkness proportional to the estimated affinity of a country towards the other. Red edges represent negative affinities of a constant width as we cannot estimate their magnitude. These visualizations allow us to conjecture about the origin of the changes in asymmetry and polarization after the crisis started, in particular in 2011. The density of the network is much higher in 2011, with a large amount of negative affinities, but also with stronger positive ones. The ARF layout algorithm highlights a network division which includes countries in debt on one side (Spain, Portugal, Italy, Greece), and countries not yet in crises on the other (Netherlands, Germany, Sweden, Finland), with the exception of Ireland in this second cluster, and the UK as a bridge between both.

While we cannot precisely quantify the dependence of the debt crisis and the metrics of asymmetry and polarization, our observations indicate a significant change in 2010 and 2011, coinciding

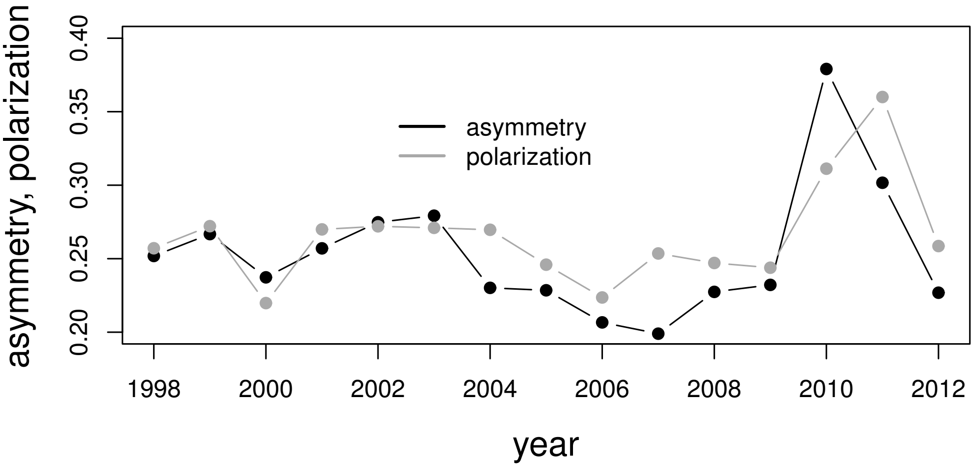

Figure 11: Time series of asymmetry and polarization among the countries in EU-15, between 1998 and 2012. A strong peak is present in 2010 and 2011, coming back to normal values in 2012.

with the loans and austerity measures in Portugal, Ireland, Italy, Greece, and Spain. In addition, these two metrics where quite stable since 1998, the year when televoting was introduced in Eurovision, only having such change in 2010 and 2011.

6 Conclusions

We have studied the cultural relations among European countries through the behavioral biases present in the Eurovision song contest. To do so, we gather a dataset of the historical contest outcomes, which aggregate the votes of large amounts of viewers who simultaneously vote by sending mobile phone messages. Our approach is centered around the statistical analysis this large-scale dataset, producing metrics that compose a macroscope of the cultural cohesion in Europe at large.

The first metric defined here is the Friend-or-Foe coefficient, a metric that reveals asymmetric positive and negative relations between European countries. We validated how this metric estimates underlying cultural affinities against a model of biased Eurovision contests. This allowed us to find a statistical rule to detect negative affinities, and a way to estimate the value of positive affinities given the outcome of a contest. Applying such rule, we could extract the networks of positive and negative relations across countries, which reveal known clusters and dyadic relations. Over this network data, we designed metrics of asymmetry and polarization, which detect changes in the voting patterns among participating countries. Applying these metrics to the votes between countries in the EU-15, we find a significant change in 2010 and 2011, the most

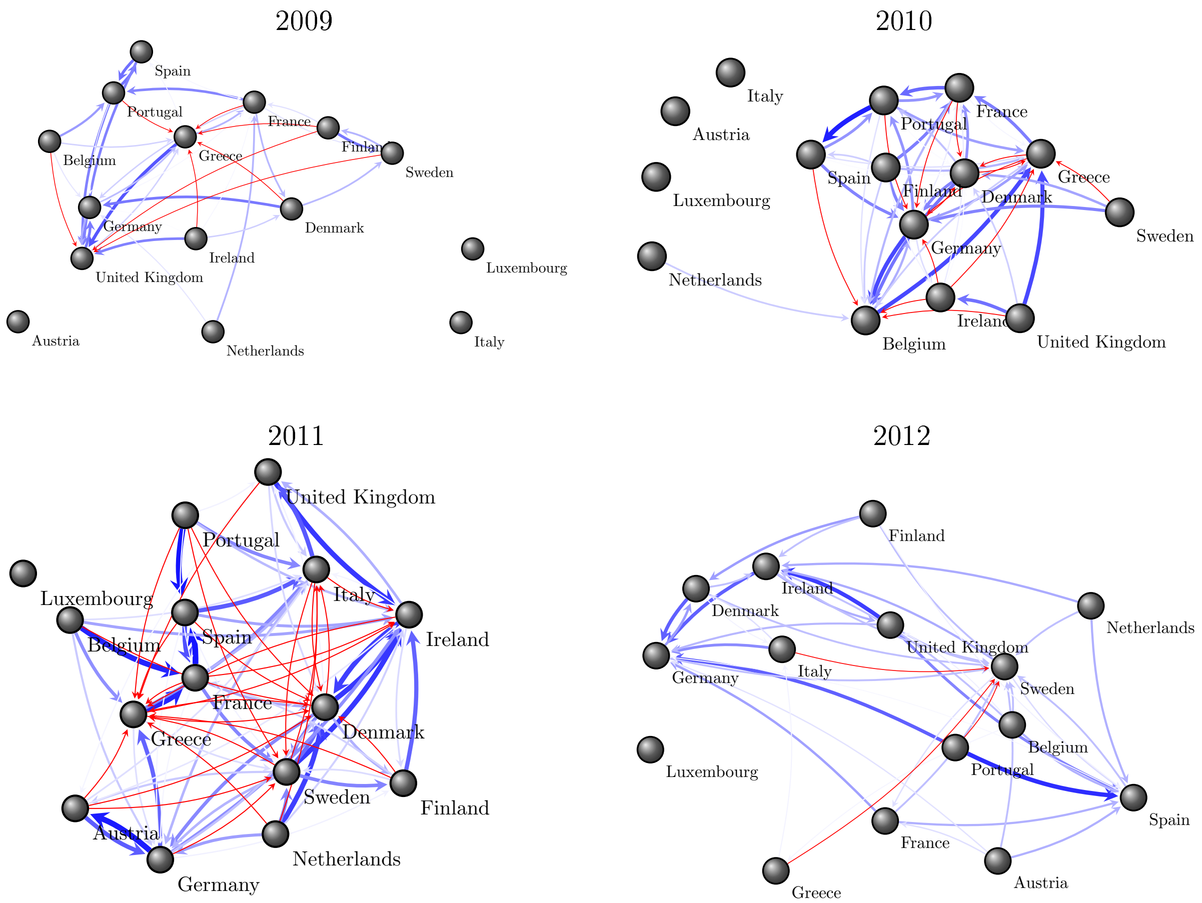

Figure 12: Affinity networks between the EU-15 countries for the 2009, 2010, 2011, and 2012 editions of Eurovision. Each node represents a country and each directed edge connects two countries if their affinity was found different than 0. Blue edges represent positive affinities with width and darkness proportional to the estimated affinity. Red edges denote instances of negative affinities. All the networks where plotted with the weighted ARF layout algorithm, as explained above.

turbulent years with respect to debt and austerity measures in the EU. These suggest, in turn, that the political climate on the EU can influence the perception of culture across countries, and economic decisions of the involved states can change the way societies relate to each other in the EU.

Additional datasets are available to extend our results. `Wikipedia` also contains the results of the semifinals of the contest, which can be used to refine the Friend-or-Foe coefficient as an aggregation of all the available data. Furthermore, our structural analysis can be combined with statistics of online behavior, such as measures of search queries, website visits, or amounts of views and comments for the `Youtube` videos of participants. These extended datasets, in turn, could be used to create testable predictors for the outcome of Eurovision contests.

A clear limitation of our analysis is the country level restriction given the data provided by the Eurovision song contest. Cultures do not need to map to countries, as ethnic minorities or pan-state cultures are neglected in this analysis. This limitation is present in the current state-of-the-art studies [1], waiting for sources of cultural data at different levels of aggregation. Our estimation of cultural affinity through the Friend-or-Foe coefficient depends on the way these two relate to each other, which we explored through model simulations. In these simulations, we have chosen a uniform distribution of the affinities between -1 and 1, on an unbiased random network. For the purpose of completeness, future work should consider other distributions (for example, normal distribution centered around 0 with some deviation) and other mechanisms of link formation (e.g. reciprocity and clustering). The nature of this network, as well as the rules to decide votes according to song quality, are empirically testable, and can be validated if more precise data becomes available.

The modeling and analysis approach presented here can be applied to other contests, for example in artist popularity competitions [7], or beauty contests. The Friend-or-Foe coefficient can be adapted to other contest schemes, looking for alternative support on the way cultural affinity creates voting biases. Finally, the metrics of positivity, asymmetry, and polarization compose a "European mood" metric, which we have shown to be related to the economical and political decisions of the European Union, and its member states. Applying this metric to future contests, we can investigate how political decisions influence the state of cultural cohesion between the inhabitants of European countries.

Acknowledgments

We would like to thank Uwe Serdült, Andreas Flache, and Frank Schweitzer for providing useful comments and discussions.

References

[1] G. Ailon. Mirror, Mirror on the Wall: Culture's Consequences in A Value Test of its Own Design. *Academy of Management Review*, 33(4):885–904, October 2008.

[2] R. Axelrod. The Dissemination of Culture: A Model with Local Convergence and Global Polarization. *Journal of Conflict Resolution*, 41(2):203–226, April 1997.

[3] Robert Axelrod and Leigh S. Tesfatsion. *A Guide for Newcomers to Agent-Based Modeling in the Social Sciences.* 2006.

[4] Philip Ball. Social science goes virtual. *Nature*, 448(7154):647–648, August 2007.

[5] Claudio Castellano, Santo Fortunato, and Vittorio Loreto. Statistical physics of social dynamics. *Reviews of Modern Physics*, 81(2), 2009.

[6] Simla Ceyhan, Xiaolin Shi, and Jure Leskovec. Dynamics of bidding in a P2P lending service. In *Proceedings of the 20th international conference on World wide web - WWW '11*, page 547, New York, New York, USA, March 2011. ACM Press.

[7] Fabio Ciulla, Delia Mocanu, Andrea Baronchelli, Bruno Gonçalves, Nicola Perra, and Alessandro Vespignani. Beating the news using social media: the case study of American Idol. *EPJ Data Science*, 1(1):8, July 2012.

[8] R.N. Costa Filho, M.P. Almeida, J.E. Moreira, and J.S. Andrade. Brazilian elections: voting for a scaling democracy. *Physica A: Statistical Mechanics and its Applications*, 322(null):698–700, May 2003.

[9] A. Dekker. The Eurovision Song Contest as a âĂŸFriendship âĂŹ Network. *Connections*, 27(3):53–58, August 2007.

[10] Giuseppe Facchetti, Giovanni Iacono, and Claudio Altafini. Computing global structural balance in large-scale signed social networks. *Proceedings of the National Academy of Sciences of the United States of America*, 108(52):20953–8, December 2011.

[11] Gabriel J. Felbermayr and Farid Toubal. Cultural proximity and trade. *European Economic Review*, 54(2):279–293, February 2010.

[12] Daniel Fenn, Omer Suleman, Janet Efstathiou, and Neil F. Johnson. How does Europe Make Its Mind Up? Connections, cliques, and compatibility between countries in the Eurovision Song Contest. *Physica A: Statistical Mechanics and its Applications*, 360(2):576–598, February 2006.

[13] Denise Rotondo Fernandez, Dawn S. Carlson, Lee P. Stepina, and Joel D. Nicholson. Hofstede's Country Classification 25 Years Later. *The Journal of Social Psychology*, 137(1):43–54, February 1997.

[14] Andreas Flache and Michael W. Macy. Small Worlds and Cultural Polarization. *The Journal of Mathematical Sociology*, 35(1):146–176, January 2011.

[15] Serge Galam. Sociophysics: a personal testimony. *Physica A: Statistical and Theoretical Physics*, 336(1-2):49–55, May 2004.

[16] Antonios Garas, David Garcia, Marcin Skowron, and Frank Schweitzer. Emotional persistence in online chatting communities. *Scientific Reports*, 2:402, May 2012.

[17] David Garcia, Antonios Garas, and Frank Schweitzer. Positive words carry less information than negative words. *EPJ Data Science*, 1(1):3, October 2012.

[18] David Garcia, Arvid Kappas, Dennis Kuester, Mathias Theunis, Elena Tsankova, Antonios Garas, Peter Kuppens, and Frank Schweitzer. Measuring the Dynamics of Individual Emotions Under Online Interaction Through Subjective and Physiological Responses. In *SPR 52nd annual meeting*, New Orleans, Lousiana, June 2012.

[19] Derek Gatherer. Comparison of Eurovision Song Contest Simulation with Actual Results Reveals Shifting Patterns of Collusive Voting Alliances. *Journal of Artificial Societies and Social Simulation*, 9(2):1, March 2006.

[20] Jim Giles. Computational social science: Making the links. *Nature*, 488(7412):448–450, August 2012.

[21] Victor Ginsburgh and Abdul G. Noury. The Eurovision Song Contest. Is voting political or cultural? *European Journal of Political Economy*, 24(1):41–52, March 2008.

[22] Marco A. Haan, S. Gerhard Dijkstra, and Peter T. Dijkstra. Expert Judgment Versus Public Opinion ? Evidence from the Eurovision Song Contest. *Journal of Cultural Economics*, 29(1):59–78, February 2005.

[23] Geert Hofstede. *Culture's Consequences: International Differences in Work-Related Values*, volume 1980. SAGE, 1980.

[24] S. Jones and J. Subotic. Fantasies of power: Performing Europeanization on the European periphery. *European Journal of Cultural Studies*, 14(5):542–557, October 2011.

[25] Arvid Kappas, E. Tsankova, M. Theunis, and D. Kuester. CyberEmotions: Subjective and physiological responses elicited by contributing to online discussion forums. In *Poster presented at the 51st Annual Meeting of the Society for Psychophysiological Research, Boston, Massachusetts*, 2011.

[26] Peter Klimek, Yuri Yegorov, Rudolf Hanel, and Stefan Thurner. It's not the voting that's democracy, it's the counting: Statistical detection of systematic election irregularities. January 2012.

[27] Ari Kokko and Patrik Gustavsson Tingvall. The Eurovision Song Contest, Preferences and European Trade. Technical report, January 2012.

[28] David Lazer, Alex Pentland, Lada A. Adamic, Sinan Aral, Albert-Laszlo Barabasi, Devon Brewer, Nicholas A. Christakis, Noshir Contractor, James H. Fowler, Myron Gutmann, Tony Jebara, Gary King, Michael W. Macy, Deb Roy, and Marshall Van Alstyne. Social science. Computational social science. *Science*, 323(5915):721–723, February 2009.

[29] Jure Leskovec, Lada A. Adamic, and Bernardo A. Huberman. The dynamics of viral marketing. *ACM Transactions on the Web*, 1(1):39, May 2007.

[30] Michael Mas, Andreas Flache, and Dirk Helbing. Individualization as driving force of clustering phenomena in humans. *PLoS computational biology*, 6(10):1–8, January 2010.

[31] William N. McPhee. *Formal theories of mass behavior.* The Free Press of Glencoe, Collier-Macmillian, London, 1963.

[32] Q. Michard and Jean-Philippe Bouchaud. Theory of collective opinion shifts: from smooth trends to abrupt swings. *The European Physical Journal B*, 47(1):151–159, September 2005.

[33] David L. Olson and Dursun Delen. Advanced Data Mining Techniques. March 2008.

[34] Gema Bello Orgaz, Raul Cajias, and David Camacho. A study on the impact of crowd-based voting schemes in the 'Eurovision' European contest. In *Proceedings of the International Conference on Web Intelligence, Mining and Semantics - WIMS '11*, May 2011.

[35] Sherwin Rosen. The Economics of Superstars. *The American Economic Review*, 71(5):845–858, 1981.

[36] Diego Rybski, Sergey V. Buldyrev, Shlomo Havlin, Fredrik Liljeros, and Hernan A. Makse. Scaling laws of human interaction activity. *Proceedings of the National Academy of Sciences*, 106(31):12640, August 2009.

[37] Serguei Saavedra, Janet Efstathiou, and Felix Reed-Tsochas. Identifying the underlying structure and dynamic interactions in a voting network. *Physica A: Statistical Mechanics and its Applications*, 377(2):672–688, April 2007.

[38] L. Spierdijk and M.H. Vellekoop. Geography, culture, and religion: Explaining the bias in Eurovision song contest voting, February 2006.

[39] Chunyan Wang and Bernardo a. Huberman. How Random are Online Social Interactions? *Scientific Reports*, 2, September 2012.

[40] G. Yair and D. Maman. The Persistent Structure of Hegemony in the Eurovision Song Contest. *Acta Sociologica*, 39(3):309–325, January 1996.

[41] Gad Yair. âĂŸUnite Unite EuropeâĂŹ The political and cultural structures of Europe as reflected in the Eurovision Song Contest. *Social Networks*, 17(2):147–161, April 1995.

[42] Slavoj Zizek. *The Sublime Object of Ideology*. Verso, 1989.