Digital Object Identifier

Eurovision Diplomacy: Investigating Neighbourhood Voting Through Data Analytics

N. PORTELLI
Department of Artificial Intelligence, University of Malta. (e-mail: nathan.portelli.19@um.edu.mt)

ABSTRACT
This research investigates voting patterns in the European Song Contest through statistical, network, correlation, and clustering analyses on voting data spanning from 1957 to 2023, and country characteristics. The study aims to determine whether neighbourhood voting, particularly between countries with geographical, linguistic, and ethnic commonalities, is prevalent within the contest. The results reveal that the existing voting patterns are non-random, with voting blocs in geographic proximity, as well as language and ethnic similarities to a lesser extent, having a significantly larger chance of voting for one another.

INDEX TERMS Data Analytics, Machine Learning, Network Analysis, Voting, Statistics

I. INTRODUCTION

THE European Song Contest (ESC) is an annual musical competition where countries select one original song to vie against others in a televised show. First held in 1956, it has become one of the world's most-watched non-sporting events, celebrating cultural diversity and European unity.

The voting process of the contest involves each participating country awarding points to other countries' songs based on a combination of jury and public votes, the results from which are tallied to determine the winner.

Yet, amidst the spectacle, there have always been persistent rumours and allegations of collusion and strategic voting. Countries are often accused of voting for their neighbours or allies rather than solely based on the merit of the performances [1]. Moreover, the contest serves as one of the few international platforms where nations can express their sentiments towards others, free from political or economic constraints [2].

Analyses of the voting patterns strongly indicate that linguistic similarity, ethnic connections, and, predominantly, geographic proximity have a profound influence on voting behaviours. Statistical tests have also shown that the voting tendencies are non-random. Moreover, network and clustering analyses have emphasised the existence of cohesive voting blocs and regional alliances, thereby highlighting the significant role of geographical proximity in shaping voting dynamics.

A. AIMS & OBJECTIVES

The purpose of this study is to investigate the presence and extent of neighbourly voting in the ESC, using simulations and statistical analysis. Building on existing research and incorporating recent data up to 2023, this study aims to determine whether certain countries consistently vote for each other more than would be expected at random, thus shedding light on the influence of geopolitical factors on the outcome of ESC.

The approach will involve:

- Simulating random vote allocation annually to establish a baseline for comparison.
- Using statistical analysis on ESC voting data up to 2023 and country profiles to identify neighbourhood voting patterns.
- Investigating correlations between voting patterns and factors such as geographic proximity, linguistic connections, and ethnic affiliations to determine whether consistent voting tendencies exist.
- Identifying voting clusters to analyse countries more likely to vote for each other, examining their characteristics.

II. DATA COLLECTION AND CLEANSING

To achieve the objectives outlined above, a comprehensive data collection process was undertaken. This involved gathering two main types of data, being information on the participating countries, and the ESC voting history.

A. PARTICIPATING COUNTRIES AND THEIR BORDERS

Countries close to each other often share historical, cultural and political ties, which may influence their voting behaviours. To assess the impact of geographic proximity on voting, countries were compared to one another based on whether they share a border.

The first step was to compile a list of all participating countries in the ESC throughout the years and identify their respective bordering countries.

This was done by collecting data from two publicly available datasets. The first source included a list of countries with land borders [3]. While this data was manually verified for accuracy, and filtered to contain only countries which have at some point participated in the ESC, it only listed countries sharing a land border. This limited the true representation of Europe, as many countries share strong cultural ties despite being separated by sea.

To address this limitation, the definition of a "neighbour" was expanded to include countries with significant maritime borders. This included the following additions:

- Cyprus and Turkey
- Denmark with Norway and Sweden
- France and the United Kingdom
- Iceland with the United Kingdom and Norway
- Malta and Italy

The second source contained a historical list of every contestant who had ever participated in the contest between 1956 and 2023, their song, the number of votes, their country of origin, and their final position on the scoreboard [4]. This list was primarily used to determine which countries participated in the contest per year, due to the ever-changing list of participants, and the total number of votes provided by other countries.

Various cleaning operations were performed on the data to fix inconsistencies in country names. For example, 'Russian Federation' was replaced with 'Russia' and 'Czechia' was replaced with 'Czech Republic'.

The political history of some countries also caused some issues. Particularly, although Yugoslavia participated as a unified nation for 27 contests before its dissolution in 1992 [5], attribution of its votes to any successor states proved impractical due to its diverse ethnic makeup. Therefore, the following approach was taken:

- 1961 - 1991: Treated as a single entity within the dataset to maintain historical consistency.
- 1992, 2004 & 2005: Votes transferred to Serbia's history, including its brief participation as the 'State Union of Serbia and Montenegro.' This reflects Serbia's dominant role following Yugoslavia's dissolution.

Additionally, the introduction of a single vote for the 'Rest of the World' *(wld)* in 2023 [6] posed methodological considerations. These votes were excluded from this study due to their lack of specificity regarding the origin of each vote, which deviated from the traditional voting structure of participating countries.

B. LANGUAGE AND ETHNICITIES

While 'neighbours' can be simply defined as countries sharing boundaries [7], the European continent is a complex region with the connections between countries being impacted by various factors beyond mere geographical proximity.

1) Linguistic Connections

Countries sharing a common language or linguistic heritage may have stronger voting ties due to linguistic affinity and cultural similarities [8, 9, 10]. To explore the influence of linguistic connections on voting, countries with shared languages were identified and examined voting patterns among these groups.

A dataset containing information on the languages spoken in every country was used [11], which was filtered to extract and categorise the language names, eliminating the unnecessary information from the dataset. This was done by first filtering for participant countries, then extracting only words that started with an uppercase, which from inspection of the dataset only consisted of languages. The languages were then split up into different records per country, and a merge operation was used to combine the languages and the relevant country with the country code for ease of data usage.

2) Ethnic Affiliations

Ethnic affiliations could also shape voting behaviour in the ESC [8, 9, 12], particularly in cases where significant ethnic minorities reside in foreign countries. To investigate the potential impact, data was extracted from the CIA's The World Factbook [13], which contains a summary of significant ethnic minorities in each country. This dataset also underwent filtering to remove unnecessary information, following a process similar to that of the linguistic dataset.

C. VOTING HISTORY

A comprehensive record of all voting results throughout the ESC was also acquired from [4]. This dataset encompasses the total points given by each country to every other country from 1957 until 2023, with televoting and jury points individually listed when appropriate. A description of voting procedures over the years can be found in Section IV-B.

This voting dataset also underwent a cleansing process, with each type of analysis for country information, being geographic, linguistic, and ethnic data, addressed separately. In general, two columns containing country codes of the source and recipient countries were removed, as this information was already present in other columns. Subsequently, empty values, notably in columns representing televoting and jury votes separately, were replaced with *'0'* for ease of formatting. These empty values corresponded to years and countries where televoting or jury voting was not utilised. Furthermore, as mentioned earlier, records including the *'wld'* votes were also eliminated.

III. ANALYSIS OF THE VOTING PATTERNS

The extensive history and wide array of participants in the ESC provide a rich dataset for analysing influence in European diplomacy. Examining voting patterns, particularly considering factors like geographical closeness, linguistic similarities, and ethnic ties, can shed light on the underlying forces that shape these patterns.

For instance, Figure 1 illustrates the countries that voted for Denmark, serving as a prime example of this phenomenon. The map demonstrates that Denmark receives the majority of its votes from neighbouring states, particularly the Nordic countries with which it shares geographic, linguistic, and ethnic backgrounds.

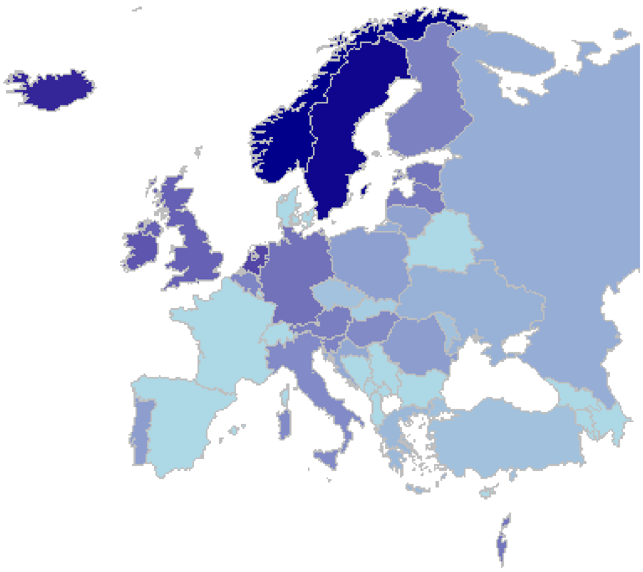

FIGURE 1. Votes given to Denmark by ESC participants. A darker blue indicates that a higher percentage of the total points were received from that country.

Therefore, to rigorously evaluate the presence of neighbourhood voting in the ESC, the following hypothesis is posed:

- *Null Hypothesis (H0):* Countries do not exhibit neighbourhood voting in the ESC; voting patterns are random and not influenced by geographic proximity, linguistic connections, or ethnic affiliations.
- *Alternative Hypothesis (H1):* Countries exhibit neighbourhood voting in the ESC; voting patterns are influenced by geographic proximity, linguistic connections, and ethnic affiliations.

A. A COMPARISON WITH RANDOMISED VOTING

To create a baseline for comparison, a null model is developed that simulates random voting patterns. This model assumes that each country assigns votes independently of geographic, linguistic, or ethnic factors, only recognising the evolving rules of the contest throughout the years, as shown in Section IV-B, as well as the average amount of voting counts given.

These complexities and the dynamic nature of the voting procedures in the ESC made it difficult to apply simulations such as the Monte Carlo method to accurately represent the full range of possible voting patterns. Therefore, the process was subdivided between the years where the voting patterns remained consistent, with a random number generator, probability-weighted to account for the votes already outputted by a country, and the possible votes each country could provide.

Let C_i denote a specific country in the dataset, and S_{ij} represent the score assigned by country C_i to country C_j. With random voting, the probability of country C_i assigning a score to country C_j can be expressed as:

$$P(S_{ij}) = \frac{N-1}{1} \tag{1}$$

where N is the total number of participating countries. This formula accounts for the fact that a country cannot vote for itself. This ensures that each country has the same chance of providing a score to every other participating country.

To easily visualise the voting dynamics across the entire matrix of possible random and actual voting scores a heatmap was created for both datasets that shows the score groups provided based on the colour gradient, with darker shades indicating higher scores and lighter shades indicating lower scores. Furthermore, diagonally darker shades within the heatmap signify a high chance of mutual voting between those specific countries, symbolised by an abbreviation established by the ISO 3166-1 alpha-2 standard [14], which can be found in Section IV-B.

By visually comparing the voting patterns between the random and actual voting patterns, as illustrated by Figures 2 and 3 respectively, one can already see significant differences such as a broader distribution of scores and a decrease in the maximum total votes allotted.

Later analysis in Sections III-D1 and III-E will compare this data to the actual voting patterns in order to further assess the null hypothesis and determine the extent to which neighbourhood voting influences voting patterns of the contest.

B. NETWORK ANALYSIS

To better analyse the actual voting patterns in the ESC, a network analysis approach is used to visualise and quantify the voting relationships and patterns between participating countries over the years.

As depicted in Figure 3, certain reciprocal high-voting tendencies between countries become immediately apparent, such as Greece and Cyprus, Moldova and Romania, Sweden and Norway, the Netherlands, and Belgium, among others. However, this figure does not immediately show whether this trend is the result of neighbourhood voting.

We therefore attempt to represent these exchanged votes through a directed network, where each competing country is represented by a vertex in the network. The directed edges within this network signify preferential voting behaviours, with their directionality indicating the origin and recipient of votes. The magnitude of each edge corresponds to the frequency or strength of voting connections between respective countries across successive events. Bidirectional vertices signify instances of mutual reciprocation in voting outcomes.

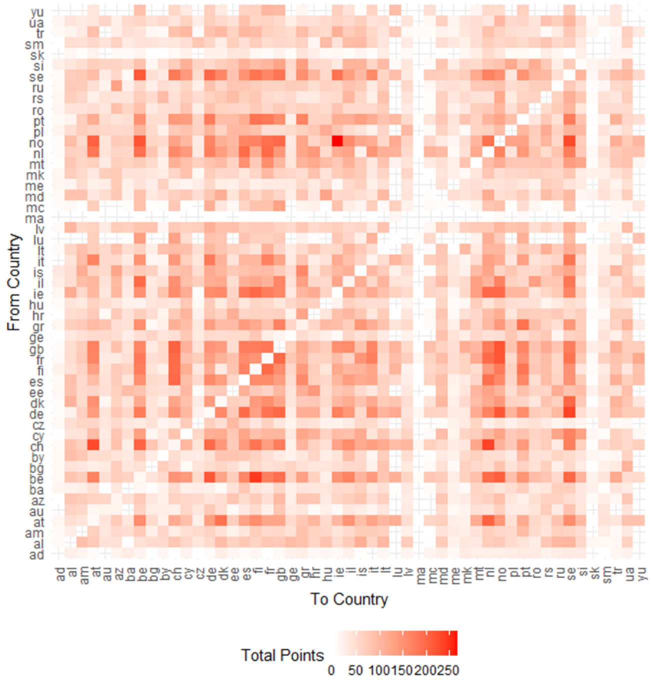

FIGURE 2. Heatmap of a Random Voting Pattern in the ESC

FIGURE 3. Heatmap of Actual Voting Patterns in the ESC

In Figure 4, each node corresponds to a participating country in the contest, with directed edges indicating the top recipient of votes from one country to another over the years. In order to differentiate between regions, each vertex is colour-coded based on the regions specified by [15].

The edges are also colour-coded: green signifies connections between countries that are geographical neighbours, while red indicates connections between non-neighbouring countries. Analysing this aspect reveals that approximately 48.08% of countries vote for countries that are bordering one another. This finding underscores the significance of geographic factors in shaping voting behaviours and highlights the potential role of cultural and regional alliances in the contest's voting dynamics.

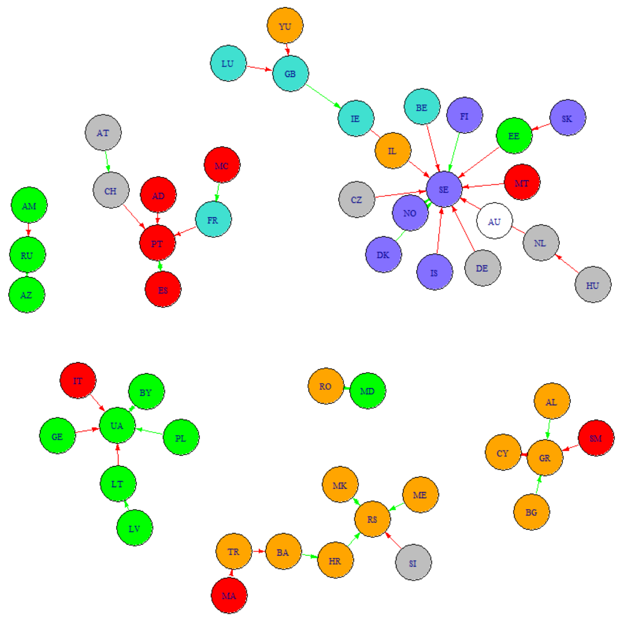

FIGURE 4. Network Graph of Bordering Countries. The vertices are colour-coded based on the following regions [15]: South-West (red), North-West (turquoise), North (blue), Central (grey), South-East (orange), East (green) and Other (white).

As a simple test, the data was filtered to the three highest voters for each country, and their shared language, ethnicity and proximity were analysed. It was found that 47.06% of these countries share a border, approximately 23.72% have a common language, and 35.87% have the same ethnic groups. This suggests that voting patterns in the contest are notably influenced by regional interconnectedness, with a considerable portion of votes coming from neighbouring countries. Additionally, the presence of shared language and ethnic ties also plays a discernible role, albeit to a lesser extent.

For a better understanding of these connections, network centrality analysis is utilised, particularly degree centrality, where the overall influence of a country within the voting network is quantified through the number of direct connection, and closeness centrality, where the ease with which a country can reach all other countries in the network is assessed by measuring the average length of the shortest paths to all other nodes.

These measures show that countries with higher degree centrality values, such as the countries listed in Table 1, are more influential in the voting network as they receive votes from a larger number of countries. This suggests that these countries have a significant impact on the overall voting dynamics of the ESC.

Closeness centrality continues to ascertain this measure, with countries having a greater proximity and interaction with other countries showing that they have a more central role within the network, playing a crucial role in facilitating voting interactions and influencing voting outcomes.

Country	Degree Centrality	Closeness Centrality
Sweden	3531	0.019230769
Norway	3483	0.019230769
United Kingdom	3378	0.019230769
Germany	3351	0.019230769
Spain	3311	0.019230769
France	3305	0.019230769
Belgium	3245	0.019230769
Netherlands	3232	0.018867925
Switzerland	3158	0.019230769
Finland	3120	0.019230769

TABLE 1. Countries with the Highest Centrality Measures in the ESC

C. CORRELATION ANALYSIS

The strength and direction of relationships between the countries and their characteristics was then measured using correlation analysis, in order to better understand the voting patterns.

Using the collected voting data, alongside information regarding countries sharing a common border, language, and ethnic groups, correlation coefficients between each pair of countries are calculated to quantify the degree of association in their voting behaviour. Specifically, the Pearson Correlation Coefficient is used, which measures the linear correlation between two values.

In this case, these variables are the points given by one country to another and whether the countries share a border, are linguistic similarity, or share major ethnic groups. Three distinct analyses are conducted to assess their separate influences on the voting patterns:

1) *Geographic Analysis*: Correlation coefficients are calculated between pairs of countries based on their geographic proximity, as determined by sharing a common border. The output reveals a strong positive correlation coefficient of *0.8393*, indicating that border-sharing countries are highly likely to reciprocate votes in the ESC.
2) *Linguistic Analysis*: Similarly, correlation coefficients are computed between countries based on their shared linguistic backgrounds, analysing whether common languages exhibit a stronger tendency to vote for each other. This yields a correlation coefficient of *0.1020263*, indicating a significantly weaker but positive correlation between linguistic similarity and voting patterns.
3) *Ethnic Analysis*: Finally, correlation coefficients on pairs of countries sharing common ethnic groups are analysed to uncover potential cultural ties and historical connections. The output results in a correlation coefficient of *0.1009814*, similarly suggesting a weak positive correlation between ethnic affinity and voting behaviour.

Similar to the conclusions from the network analysis, these results suggest that while linguistic similarity and ethnic affinity factors might play a role, countries with a shared border have a far more substantial impact on voting patterns in the ESC.

D. CLUSTERING ANALYSIS

1) Clustering Coefficients

The clustering coefficient algorithm provides valuable insights into the voting behaviours within the network, identifying clusters or communities of countries that exhibit a higher tendency to vote for each other based on shared voting patterns [2]. This analysis goes beyond simple correlations, revealing the underlying structure of the voting network and shedding light on the presence of voting blocs and regional alliances influencing voting behaviours.

As demonstrated in Table 2, the analysis unveils a highly clustered network, characterised by a mean clustering coefficient of *0.957*, a median of *0.972*, and a Standard Deviation of *0.033*. These statistics underscore the pronounced tendency for countries to cast their votes within tightly-knit groups, highlighting the cohesive voting behaviours prevalent within these clusters. The uniformity in clustering across the network, as indicated by the relatively low SD, implies that most countries participate in these voting blocs.

Figure 5 provides more detailed information on the clustering coefficient of each country, with higher clustering coefficient meaning that they have a high tendency for their neighbours to be interconnected, indicating the presence of more highly clustered regions.

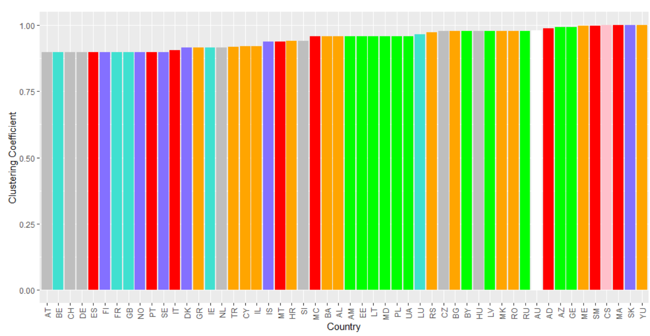

FIGURE 5. Clustering Coefficients by Country and Region. The vertices are colour-coded based on the following regions [15]: South-West (red), North-West (turquoise), North (blue), Central (grey), South-East (orange), East (green) and Other (white).

This discrepancy between actual and random results shown in Table 2 indicates that the observed clustering in the ESC voting network is not merely a random occurrence, but rather a characteristic feature of the voting behaviours exhibited by participating countries. The consistently higher clustering coefficients in actual votes compared to random votes continue to suggest that countries tend to form cohesive voting blocs.

Furthermore, the relatively small SD for both actual and random data indicate a degree of consistency in clustering across different years, with the actual clustering coefficients consistently higher than those of random data. This suggests that the observed clustering patterns in the ESC voting network are not only prevalent but also stable over time, further highlighting the significance of voting blocs and regional alliances in shaping voting behaviours.

Year	Actual Votes			Random Votes		
	Mean	Median	SD	Mean	Median	SD
2013	0.6508	0.6909	0.1421	0.5768	0.5556	0.0885
2014	0.6299	0.6000	0.1314	0.5316	0.5211	0.0825
2015	0.7538	0.7937	0.1927	0.5385	0.5327	0.0737
2016	0.7485	0.7338	0.1319	0.7553	0.7218	0.0955
2017	0.7944	0.8396	0.1583	0.7457	0.7066	0.0928
2018	0.7712	0.7536	0.1384	0.7676	0.7291	0.1006
2019	0.7807	0.8301	0.1487	0.7567	0.7218	0.0944
2021	0.7820	0.7908	0.1446	0.7517	0.7241	0.0796
2022	0.7899	0.7947	0.1451	0.7292	0.6858	0.0899
2023	0.7630	0.7428	0.1014	0.7507	0.7367	0.0732
Total	0.9568	0.9718	0.0332	0.9526	0.9663	0.0346

TABLE 2. Summary Statistics of Clustering Coefficients for Actual and Random Data (2013-2023, excluding 2020).

2) Community Detection

In addition to analysing clustering coefficients, community detection methods further enhances understanding of voting patterns and group dynamics within the ESC by uncovering hidden patterns of cooperation. The Louvain method, a popular algorithm for community detection in complex networks, is applied to identify cohesive communities or clusters of countries based on their voting behaviours.

This method works by optimising a modularity function to detect communities with high internal connectivity and low external connectivity. By iteratively optimising the modularity score, the algorithm identifies communities that exhibit strong intra-community ties while minimising inter-community connections. This approach allows for the identification of distinct voting blocs or communities within the ESC voting network, shedding light on the underlying structures and relationships among participating countries.

Due to the ever-changing diplomatic structure of the European region, it was determined that taking a subset of the dataset, particularly votes between 2003 and 2023, would provide a more accurate representation of the current voting blocs, as well as build on studies that used older data [10, 16, 17, 18].

Based on this information, four distinct clusters are identified, which are described in Table 3. As can be better seen in Figure 6, these groups are primarily based on geographical proximity, with the clusters being made up of Western Europe, North-Central Europe, Mediterranean-Eastern Europe, and the Alpine-Balkan region.

Cluster	Countries	Colour in 5
1	Andorra, Spain, France, Monaco, Portugal	Blue
2	Albania, Austria, Bosnia and Herzegovina, Bulgaria, Switzerland, Serbia and Montenegro, Croatia, Montenegro, North Macedonia, Serbia, Slovenia, Turkey	Red
3	Armenia, Azerbaijan, Belarus, Cyprus, Georgia, Greece, Israel, Italy, Lithuania, Latvia, Moldova, Malta, Romania, Russia, San Marino, Ukraine	Orange
4	Australia, Belgium, Czech Republic, Germany, Denmark, Estonia, Finland, United Kingdom, Hungary, Ireland, Iceland, Netherlands, Norway, Poland, Sweden, Slovakia	Green

TABLE 3. Clusters identified through the voting data between 2003 and 2023

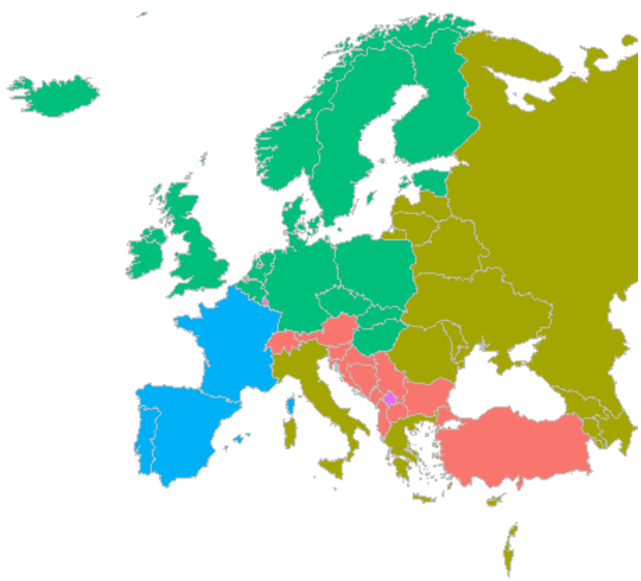

FIGURE 6. Map of clusters identified through the voting data between 2003 and 2023. The countries are colour-coded according to Table3.

E. STATISTICAL TESTS

Using the simulated random vote distribution as a reference, the actual voting data is subjected to various statistical tests.

1) Chi-Square Test

The Chi-Square test is conducted to determine if there is a significant association between countries in terms of the points they give to each other. It is calculated as follows;

$$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i}$$

where O_i is the observed frequency for category i, made up of the actual voting data, and E_i is the expected frequency for category i. The latter is the theoretical frequency that is expected if the null hypothesis is true, usually derived from a probability distribution or contingency table.

The results reveal a significant association between countries in terms of the points they award each other, with a *p-value < 2.2e-16*, suggesting that the observed distribution is significantly different from the expected distribution under the null hypothesis. Therefore, bordering countries tend to vote for each other in a non-random manner. This supports the hypothesis that geographical proximity has a significant influence on voting behaviour.

2) T-Test

The Welch Two Sample t-test is then used to compare the average points given by bordering countries to each other against those given by non-bordering countries. This was calculated as follows;

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

where

- $\bar{X}_1$ and $\bar{X}_2$: The sample means of the two groups being compared.
- s_1^2 and s_2^2: The sample variances of the two groups.
- n_1 and n_2: The sample sizes of the two groups.

The analysis revealed a significant difference between the two groups ($t = -37.248$, $df = 102640$, $p < 2.2 \times 10^{-16}$), indicating that the true difference in means is not equal to zero. This suggests that proximity plays a crucial role in voting preferences.

Furthermore, the 95 percent confidence interval for the difference in means ranged from *-1.0836* to *-0.9752*. The sample estimates indicate that the mean points given by bordering countries ($\bar{X}_1 = 3.1221$) are notably lower than those given by non-bordering countries ($\bar{X}_2 = 4.1515$).

These tests collectively underscore the influence of commonalities between countries, with neighbouring countries displaying distinct preferences compared to their non-neighbouring counterparts, further revealing the non-random nature of the voting behaviours.

Given the consistent evidence from multiple analytical approaches and statistical tests, the null hypothesis (*H0*) that countries do not exhibit neighbourhood voting in the ESC is rejected. Instead, the data robustly supports the alternative hypothesis (*H1*) that voting patterns in the ESC are significantly influenced by geographic proximity, as well as linguistic and ethnic affiliations to a lesser extent. These findings highlight the complex interplay of European diplomacy, alliances and cultural ties in shaping the voting dynamics of the ESC.

IV. CONCLUSION

In conclusion, this study robustly demonstrates that voting patterns in the European Song Contest (ESC) are significantly influenced by geographic proximity, as well as ethnic and linguistic commonalities to a lesser extent, with neighbouring countries exhibiting a strong tendency to vote for each other. Statistical analyses, including Chi-Square and T-Tests, consistently revealed non-random voting behaviours, network analysis further highlighted reciprocal high-voting tendencies between geographically linked countries, while correlation analysis showed strong positive correlations between shared borders and voting patterns. Clustering analysis revealed a highly clustered network with significant voting blocs and regional alliances, reinforcing the impact of geographical proximity on voting behaviours.

The findings collectively highlight the complex interplay of European diplomacy, alliances, and cultural ties, rejecting the null hypothesis of random voting and supporting the alternative hypothesis of significant neighbourly voting influence.

A. ASSUMPTIONS AND LIMITATIONS

Some key assumptions were made when analysing the retrieved data:

- The song's language did not factor into the analysis. Instead, the primary languages spoken in each country are used instead to identify commonalities. Given the focus on neighbourhood voting, emphasising the countries themselves over the songs would more accurately capture if such as relationship exists.
- Similarly, no analysis is included on the quality of a song, as, apart from the reasoning provided by [2], this is extremely subjective and varies greatly between individuals and throughout the years.
- No analysis of news sources or the changing public relations of countries over the years is performed, as this is beyond the scope of the investigation.

B. FUTURE WORK

Further research could investigate the impact of changes in voting procedures on the prevalence of neighbourly voting over time. Another valuable area of research could involve the integration and sentiment analysis of news sites and social networks to gain deeper insights into public voting behaviours and perceptions. Additionally, incorporating the quality of songs and their performances, perhaps through crowd-sourced ratings or expert reviews, could provide a more nuanced understanding of the factors influencing voting patterns. Finally, expanding the scope of analysis to include geopolitical events and their temporal relationship with voting trends could offer a broader perspective on the dynamics of international relations as reflected in the ESC.

REFERENCES

[1] Eurovision: Jury results of six countries removed after 'voting irregularities' identified. *Euronews*, May 2022.

[2] 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, 2006.
[3] GeoDataSource. https://www.geodatasource.com/addon/country-borders, Mar 2019. Online; accessed 15-April-2024.
[4] Janne Spijkervet. The Eurovision Dataset, March 2020.
[5] Yugoslavia - eurovision song contest. https://eurovision.tv/country/yugoslavia. Online; accessed 15-April-2024.
[6] Voting changes (2023) faq | eurovision song contest. https://eurovision.tv/voting-changes-2023-faq. Online; accessed 15-April-2024.
[7] Marta Blangiardo and Gianluca Baio. Evidence of bias in the eurovision song contest: modelling the votes using bayesian hierarchical models. *Journal of Applied Statistics*, 41(10):2312–2322, 2014.
[8] 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.
[9] Nicholas Charron. Impartiality, friendship-networks and voting behavior: Evidence from voting patterns in the Eurovision Song Contest. *Social Networks*, 35(3):484–497, July 2013. ASSIGNMENT.
[10] Alexander V. Mantzaris, Samuel R. Rein, and Alexander D. Hopkins. Examining collusion and voting biases between countries during the Eurovision song contest since 1957. *Journal of Artificial Societies and Social Simulation*, 21(1):1, 2018. arXiv:1705.06721 [stat].
[11] Languages spoken across various nations. https://www.kaggle.com/datasets/shubhamptrivedi/languages-spoken-across-various-nations. Online; accessed 15-April-2024.
[12] Zrinka Borić and Ana Radović Kapor. The European Song Contest as a tool of cultural diplomacy. *Zbornik sveučilišta Libertas*, 1-2(1-2):225–240, December 2017. Publisher: Libertas meunarodno sveučilište.
[13] The world factbook - ethnic groups. https://www.cia.gov/the-world-factbook/field/ethnic-groups/. Online; accessed 15-April-2024.
[14] Codes for the representation of names of countries and their subdivisions – Part 1: Country codes. Standard, International Organization for Standardization, Geneva, CH, 2020.
[15] Alexander V. Mantzaris, Samuel R. Rein, and Alexander D. Hopkins. Preference and neglect amongst countries in the eurovision song contest. *Journal of Computational Social Science*, 1(2):377–390, Sep 2018.
[16] Gad Yair. 'unite unite europe' the political and cultural structures of europe as reflected in the eurovision song contest. *Social Networks*, 17(2):147–161, 1995.
[17] Derek Gatherer. Birth of a meme: the origin and evolution of collusive voting patterns in the eurovision song contest. *Journal of Memetics - Evolutionary Models of Information Transmission*, 8(2), Mar 2004.
[18] 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, 2006.

APPENDIX

COUNTRY NAME ABBREVIATIONS

The abbreviations used within the figures to refer to each participant nation were derived from the ISO 3166-1 alpha-2 standard [14], and are defined as the following:

AB (Albania), AD (Andorra), AM (Armenia), AU (Australia), AT (Austria), AZ (Azerbaijan), BY (Belarus), BE (Belgium), BA (Bosnia and Herzegovina), BG (Bulgaria), Croatia (HR), Cyprus (CY), Czechia (CZ), DK (Denmark), EE (Estonia), FI (Finland), FR (France), GE (Georgia), DE (Germany), GR (Greece), HU (Hungary), IS (Iceland), IE (Ireland), IL (Israel), IT (Italy), LV (Latvia), LT (Lithuania), LU (Luxembourg), MT (Malta), MD (Moldova), MC (Monaco), ME (Montenegro), MA (Morocco), NL (Netherlands), MK (North Macedonia), NO (Norway), PL (Poland), PT (Portugal), RO (Romania), RU (Russia), SM (San Marino), RS (Serbia), SK (Slovakia), SI (Slovenia), ES (Spain), SE (Sweden), CH (Switzerland), TR (Turkey), UA (Ukraine), GB (United Kingdom), YU (Yugoslavia).

ABBREVIATIONS

ESC	European Song Contest
SD	Standard Deviation

VOTING PROCEDURES

The competition's voting methods have undergone significant changes since its inception in 1956, making straightforward homogeneous sampling difficult. Notably, data from 1956, the only year where countries could submit multiple entries, have never been made public. Therefore, this year has been excluded from the analysis.

From 1957 until 1961, and again between 1967 and 1970, and later again in 1974, each participating country c awarded 10 points, denoted as c_i, to distribute among other participants, where the number of participants eligible for votes from a particular country is *c-1*, as self-voting is prohibited.

The voting scheme changed for 1962 and 1963, with the maximum amount of available points $c_i max$ set to 3 and 5, respectively.

A distinctive system emerged from 1964 to 1966, with countries assigned c_i points from a set of $\{1, 3, 5, 6, 9\}$ points, with a possible $c_i max$ of 9, although a single maximum vote of 9 was never used.

The period between 1971 and 1973 saw a shift to a simpler rating system, where each country rated entries on a c_i scale of $\{1, 2, 3, 4, 5\}$.

From 1975 to 2015, a structured voting system was established, allowing countries to assign c_i to {1, 2, 3, 4, 5, 6, 7, 8, 10, 12}. improving granularity in scoring. Apart from jury voting, televoting was slowly introduced in different countries, either sharing the voting power by 50/50 or taking over the jury role entirely. As the voting power remained the same for countries regardless of televoting, this aspect was not considered.

Subsequently, from 2016 to 2022, while maintaining the structure introduced in 1975, televoting became universal, providing both jury panels $c_i j$ and televoters $c_i t$ with separate but identical voting points of {1, 2, 3, 4, 5, 6, 7, 8, 10, 12}.

In 2023, the voting scheme once again shifted. While maintaining the previous voting power and point allocation, the jury votes were eliminated from the semi-final competitions. Additionally, a single vote was introduced for the rest of the non-competing world, denoted as *wld*. This vote functions as an additional participating country **c** without a jury and the inability to vote. Consequently, the number of participants eligible for votes changed to *c-wld-1*.

R SCRIPT DESCRIPTIONS

The following section contains a brief description of the content of the scripts used to analyse the notion of neighbourhood voting in the ESC. The full scripts and datasets, which including additional details on their contents as well as additional statistics and metrics which were not mentioned above, can be found on GitHub.

1) borders_data_preparation.R

This script was used to analyse geographic proximity and potential neighbourhood voting patterns among ESC participants. Data from the *geodatasource-country-borders.csv* was used [3], which provided information on countries and their land borders.

However, this dataset alone was insufficient to capture the full scope of neighbouring relationships, as many European nations share strong neighbourhood relationships despite being separated by the sea. To address this limitation, additional bordering relationships were manually incorporated for countries such as Malta and Italy, Denmark, Norway and Sweden, Iceland, the United Kingdom and Norway, France and the United Kingdom, and Cyprus and Turkey. Furthermore, Israel and Australia, which lack any close borders with other participants, were included and assigned an empty value for their bordering countries to ensure their retention in the dataset.

Data cleaning was then performed, standardising country names and handling special cases such as the dissolution of Yugoslavia and the brief participation of the State Union of Serbia and Montenegro. This data was then filtered to include only those countries that have participated in the contest, making use of a list of unique participating countries extracted from *contestants.csv*. The resulting filtered dataset was exported to the *filtered_countryborders_data.csv* file for further analysis.

2) borders_voting.R

The processed information from *borders_data_preparation.R* is then used by this to analyse the voting patterns of countries that physically border one another, in order to see whether they have a higher chance of voting for each other.

The main datasets used are *filtered_countryborders_data.csv*, extracted by the *borders_data_preparation.R* script and *votes.csv*, which contains the voting history of each country from 1957 until 2023.

A general heatmap of the voting patterns in the ESC is created with the x-axis containing the country to which the votes were given, and the y-axis showing which country gave the vote, as shown in Figure 3. If the colour intensity is diagonally symmetrical, it means that the voting trend is reciprocated between the two countries.

Network graphs were also used to visualise the voting patterns of countries with a shared border. The goal was to visually explore if neighbouring countries exhibit higher instances of mutual voting.

Degree and Closeness Centrality were the two primary network metrics analysed. Pearson's correlation coefficient was also used used to quantify the relationship between the proportion of neighbouring points relative to the total points. Two key statistical tests, the Chi-Square Test and the T-Test, are utilised to check whether the patterns observed are statistically significant or due to random chance.

3) clustering_analysis.R

This script implements a comprehensive analysis of clustering coefficients and community detection based on voting patterns among the participant countries. Initially, the script preprocesses the actual and random voting data from *votes.csv* and *random_votes.csv* respectively, the latter of which was created by the *random_voting_patterns.R* script. The clustering coefficients of the actual and random votes are subsequently calculated and compared, for both the total combined years, as well as individual years between 2013 and 2023. Summary statistics such as mean, median, and standard deviation of clustering coefficients are computed and stored in *clustering_stats_votes.csv* and *clustering_stats_random_votes.csv*. The output of these statistics can be found in Table 2.

Moreover, the script employs the Louvain method for community detection within the voting network, identifying clusters of countries with similar voting behaviours. These clusters are visualised on a map of Europe using different colours, which can be seen in Figure 3.

4) ethnic_data_preparation.R

Processing ethnic group data for ESC participants, the script makes use of *cia-factbook-ethnic-groups.csv*, which provides detailed information on ethnic groups in various countries [13], and *unique_countries.csv*, which lists countries that have participated in the ESC from 1957 to 2023.

The ethnic group names data was cleaned by removing unnecessary characters, spaces, and specific terms like 'EU' and

'European'. Manual reviews followed to eliminate duplicates and refine ethnicity names further.

The data was then filtered to focus exclusively on ESC participant countries, and the ethnic groups were split into individual records. Then each ethnic group was sent to its respective country.

Country names were then matched with their respective country codes, which was then merged and saved as *separated_ethnicities_coded.csv*.

5) ethnic_voting.R

This script performs an analysis of voting patterns in ESC based on shared ethnicities between participating countries, investigating whether countries with common ethnic groups tend to vote for each other more frequently. Two primary datasets are used in this analysis are *votes.csv*, containing voting history data for the ESC, and *separated_ethnicities_coded.csv*, which was extracted from *ethnic_data_preparation.R*.

Countries with the same ethnicities are compared by creating all possible combinations of countries for comparison, excluding self-comparisons. A custom function determines whether two countries share any common ethnic groups, and the results are recorded in *ethnicity_comparison.csv*.

The connections between countries with shared ethnicities are visualised using network graphs. For further exploration, a subset of the data focusing on countries with a Russian minority is also visualised.

The correlation between the total points awarded and the presence of shared ethnicities between countries is calculated by merging the voting and ethnicity data, the relationship between these variables. This provides insights into voting behaviour based on ethnic affiliations.

Network metrics are then computed to assess the importance and influence of countries with shared ethnicities within the voting network. Degree centrality, closeness centrality, and betweenness centrality measures are calculated, the results of which are saved in *ethnic_centrality_measures.csv*.

Correlation analysis, including degree, closeness, and betweenness centrality calculations, as well as statistical tests, including the chi-square test and the t-test, are calculated in a similar fashion as was done in *border_voting.csv* and *language_voting.csv*, with the results being saved to *ethnic_centrality_measures.csv*.

6) language_data_preparation.R

The script processes language data for countries participating in the ESC. It utilises two primary datasets, being *countries-languages.csv*, which details the languages spoken across various nations, and *unique_countries.csv*, which lists ESC participants from 1957 until 2023.

To ensure data accuracy, country and language names were standardised, cleaned and filtered to include only ESC participants. The languages were then split for each country into individual records, which wre subsequently merged with country code information. The final output is saved as *separated_languages_coded.csv*.

7) language_voting.R

This script analyses Eurovision voting patterns to investigate whether countries sharing languages vote for each other more frequently. It makes use of the *votes.csv* and *separated_languages_coded.csv*, which was extracted by *language_data_participation.R*.

The total points for each country pair are aggregated and the languages spoken in each pair are compared to identify shared languages. This results in a dataset indicating whether two countries share any common language, which was then saved in *language_comparison.csv*.

Network graphs are plotted to visualise relationships between countries with shared languages. A general network graph shows all countries with shared languages. An additional graph was created as an example highlighting countries connected to the Russian language.

Correlation analysis, including degree, closeness, and betweenness centrality calculations, as well as statistical tests, including the chi-square test and the t-test, are calculated in a similar fashion as was done in *border_voting.csv* and *ethnic_voting.csv*, with the results being saved to *lang_centrality_measures.csv*.

8) random_voting_patterns.R

This script was developed to generate simulated voting patterns for the ESC, serving as a baseline for comparison against the observed voting data. The datasets used for this *unique_countries.csv*, containing a list of countries that have participated in the contest, and *votesċsv*, containing a record of all voting results [4]. An empty votes dataset was then created by removing actual vote counts while retaining the necessary structure. To ensure a higher degree of randomness, a function was implemented to shuffle the recipient countries *(to_country)* within each group of voting countries *(from_country)*. The script then proceeded to simulate random voting patterns based on the evolving rules and conditions of the contest throughout the years, as described in Section IV-B, along with additional restrictions to maintain fairness. Furthermore, the script accounted for the introduction of televoting and the separation of jury and public votes from 2016 onwards. Finally, the resulting simulated random voting patterns were exported to the *random_votes.csv* file for further analysis and comparison with the observed data.

9) random_voting_analysis.R

The random votes generated by *random_voting_patterns.R*, which were stored in *random_votes.csv*, are compared with actual ESC voting data in *votes.csv*. The objective is to identify patterns and biases in the actual voting process.

Both datasets include the same details, being the country that voted, the recipient of the votes, the total points awarded, and distinctions between jury and televoting points. For consistency, *wld* entries were removed and "NA" values replaced with "0".

The aggregated data was used to calculate the total points given by each country and a heatmap was created for the random voting dataset, shown in Figure 2 to visually compare the distribution of votes with the actual voting results in Figure 3.

Statistical analysis included a correlation analysis to understand the relationships between countries in both datasets, using correlation matrices. Additionally, a Chi-Square Test was performed to compare the distribution of votes between the random and actual voting data. The test results indicated significant differences between the two datasets. Finally, a T-Test was conducted to compare the means of total points in the random and actual voting data, further highlighting discrepancies and potential biases in the actual voting process.