

Ne gre le za melodijo: kako Evropa glasuje za svoje najljubše skladbe

Anej Svete, Jakob Hostnik, Lovro Šubelj

Univerza v Ljubljani, Fakulteta za računalništvo in informatiko, Večna pot 113, SI-1000 Ljubljana

as5108@student.uni-lj.si, jh9934@student.uni-lj.si, lovro.subelj@fri.uni-lj.si

Izveček

Tekmovanje za pesem Evrovizije je priljubljeno mednarodno pevsko tekmovanje, ki ga organizira Evropska radiodifuzna zveza. Zmagovalca določi glasovanje publike in strokovne žirije vsake sodelujoče države, kar pomeni, da analiza glasovalnega omrežja tekmovanja ponuja dragocen vpogled v faktorje, ki, poleg kvalitete nastopov, vplivajo na odločitve za glasovanje.

V članku predstavimo rezultate analize glasovalnega omrežja in opišemo rezultate napovednega sistema rezultatov, ki temelji na dosedanjih uvrstitvah. Opišemo metodologijo in predstavimo podatke, na katerih je bila izvedena analiza. Rezultati med drugim vsebujejo nekaj splošnih lastnosti omrežja, najdene skupine držav, ki si med seboj izmenjajo znatno več točk, kot bi pričakovali, in razpravo o tem, kaj so vzroki tega trenda. Samo na podlagi algoritmične obdelave izpostavimo znana razmerja med državami, kot je favoriziran odnos med Ciprom in Grčijo, med skandinavskimi državami in med državami nekdanje Jugoslavije. Na splošno se izkaže, da sosednje države pogosto sodelujejo, sploh če so kulturno ali geografsko ločene od ostalih. Očitno je tudi, da gradnja takih odnosov izboljša uspeh v tekmovanju in da se trend vključevanja v take skupine povečuje. Hkrati pa je opazen tudi negativen vpliv vpletenosti v odnose zanemarjanja drugih držav na uvrstitev v tekmovanju. Preizkusimo tudi model, ki bi na podlagi zgodovinskega glasovanja in preferenc posameznih držav napovedal prihodnje rezultate, vendar se model izkaže za slabšega v primerjavi z napovedmi, ki jih omogočajo stavnice tekmovanja.

Ključne besede: Tekmovanje za pesem Evrovizije, Pristranskost glasovanja, Struktura skupnosti, Predvidevanje izidov glasovanja

Abstract

The Eurovision Song Contest is a popular annual international song competition organized by the European Broadcasting Union. The winner is decided by the audience and expert juries from each participating nation, which is why the analysis of its voting network offers a great insight into what factors, beside the quality of the performances, influence the voting decisions.

In this paper, we present the findings of the analysis of the voting network together with the results of a predictive model based on the collected data. We touch upon the methodology used and describe the dataset that we are analyzing. The results include a number of general features of the voting networks, the exposed communities of countries that award significantly more points to each other than would be expected and predictions on what the biggest factors that lead to this phenomenon are. Multiple known favourable relationships are exposed, such as the one between Greece and Cyprus, between Scandinavian countries and between the countries of the former Yugoslavia. A general trend of neighbouring countries forming alliances is observed, which is especially apparent if they are culturally or geographically separated from others. Furthermore, it is also observed how membership in communities of common point exchange helps achieve better results in the competition and that this trend is on the rise. At the same time, countries involved in relationships of neglect often achieve poorer results. We also try out a model in order to predict the votes based on the network structure of both previous votes and song preferences of nations, which was found not to offer significant improvement of predictions compared to betting tables alone.

Keywords: Eurovision Song Contest, voting bias, community structure, voting inference.

1 INTRODUCTION

The Eurovision Song Contest (ESC) has been held every year since 1956. Its initial purpose was to unite the European nations after the Second World War and has since evolved into an annual entertainment spectacle followed by millions of people. Every rendition of the contest except for the first featured one song entry by each participating country. The countries involved are mostly European, with the recent addition of some nations from outside the continent, e.g. Israel and Australia.

Although some rules have changed throughout the years, the main principles of the competition remain the same. Each participating nation awards some number of points to the performances chosen by the public and jury of that country. Countries cannot give points to themselves. The song with the highest number of points wins. The current system is in place since 2004 and consists of two semi-finals and a final. Each country gets the same number of points to distribute, and they are equally split between the jury and televoting votes. Both are converted on a scale of points ranging from 1 to 12, with the exception of 9 and 11 points, which are not awarded. This means that every country awards points to 10 performances (EBU, 2019), (EBU, 2019).

The nature of voting offers an intriguing opportunity to explore what European countries base their voting decisions on. Some of the commonly attributed factors include geographical proximity, language similarity (Dekker, 2007), ethnic structure (Spierdijk & Vellekoop, 2006), common history, political preference and cultural similarity (Ginsburgh & Noury, 2008). We try to extract as much information on the deciding factors as possible from the network structure and the nation's attributes to pinpoint the most important influences and to leverage that information to infer future voting choices. We have not found any previous work that tried to use the network information for future predictions. It is a challenging task, partly because of the challenge of obtaining enough quality data and partly because of the changing format of the competition. We take these historical differences into account during the analysis. However, since one of our main goals is making future predictions, the most recent results are the most important, and these are enough to expose some main trends.

Besides making predictions, we are also interested in how different similarities between nations correlate

to their voting patterns. Just by paying some attention to points distribution, it can be seen that there is some correlation between the points awarded and geographic proximity. We try to leverage the network features (e.g. the strength of bias shown throughout the years and the community structure this implies) to extract useful information more precisely and then reason about the biggest deciding factors on the voting. To achieve that, we perform community detection on the network of shown bias to infer the influences.

There have also been some suggestions that building these friendships allows the participants to achieve better scores and rank higher in the competition. One part of the paper thus also focuses on finding out if this is really the case by finding a correlation between the community structure of specific nations and their success in the competition.

To prevent too much biased voting, some measures have already been taken by the ESC committee. For example, they try to minimize the number of neighboring countries competing in the same semi-final and since only the countries performing that night can vote, neighboring countries have less of a chance to help each other get into the final (EBU, 2019), (EBU, 2019). This measure can of course not be taken in the final.

Another thing that we investigate is the notion of neglect between countries. By this we mean the behavior when countries which are somehow linked to one another seldom award each other a significant number of points. In other words, neighbors that do not exchange points could be regarded as neglecting each other.

When the major influences on the voting behavior are exposed, we turn our attention to the actual predictor of future voting behavior and use the information about the communities in the second part of the project together with some additional data that we learn through song and artist features to try to infer the number of points countries will award in future competitions. The model is used together with betting predictions since they are considered to be the best existing way to predict the outcome of the competition. As described in Data collection and presentation, we gathered data from various sources in hope of making some confident predictions.

The rest of the paper is structured as follows. We review some of the previous work on the topic, ranging from specific analysis of the ESC voting net-

work to the more general methods of examining the data. Then we present the dataset we worked with and how it was obtained. We also present some of its main characteristics. Then we describe in more detail the methodology used. Since some of the needed data turned out to be very difficult to get hold of, some compromises had to be made and these are discussed as well. In the last part we present the results of the project and we end by brainstorming some future work ideas and concluding the paper in a summary.

2 RELATED WORK

Since we mainly deal with analysis of the ESC voting network, we discuss some previous papers covering the topic in terms of applying network techniques on the problem, introducing some ideas and techniques that will prove useful for our project. Although they all deal with the competition as a network problem to some extent, none of them use more advanced network analysis tools such as community detection and link prediction on the graphs, which we implement. Both these methods can then be used for future projections, which is also not dealt with in any of the papers.

(Mantzaris, Rein, & Hopkins, Preference and neglect amongst countries in the Eurovision Song Contest, 2018) investigate different possible explanations for the voting patterns which deviate significantly from a uniform distribution, specifically focusing on the notion that nations try to build reciprocal voting connections that lead to them receiving more points from their “partners”, and thus ranking higher. Therefore, they try to find correlation between the number of collusive edges a nation has and their success in the competition. They build on previous work in (Mantzaris, Rein, & Hopkins, 2017), (Gatherer, 2006), analyzing the voting behavior by simulation voting, since analytical identification of statistically significant trends in the competition would be mathematically too complex because of its changing nature. Capturing the different voting systems in place throughout the years mathematically is untraceable, therefore simulation provides a good compromise.

The authors extend the algorithm presented in (Gatherer, 2006) for finding significant exchange of points awarded between participants. The original paper focused on a limited interval of competitions when the voting rules were mostly homogeneous, therefore *Mantzaris et al.* provide a more general sampling technique. To be able to do that, they iden-

tify the three principles of voting used by ECS since its start in 1956. These can be grouped as *allocated*, *sequential* and *rated*. The algorithm samples the uniform distribution of points throughout a time period, based on the rules in place at the time and then extracts the highest-weighted edges. Network is formed based on those colluding edges between countries, showing patterns of biased voting. They then perform community detection on obtained structures and base their results on those. They consider both one-way and two-way relationships and thus lay groundwork for thorough network inspection in terms of both motif and community detection.

They find significant patterns of both preference and neglect spanning throughout the participating nations, showing that voting is geographically influenced, linking it to mutual history, similar ethnic features and the feeling of “brotherhood” of neighboring countries. They also conclude that the participants with higher number of colluding edges achieve better success in the competition, showing it does pay off to build partnerships. This, together with the changing nature of the network that is more and more concentrated around the colluding edges, implies that nations are actively trying to build these relationships.

(Dekker, 2007) provides a different take on the analysis of the voting network. The techniques the authors demonstrate have a more general applicability, spanning away from the ESC, and can also be used for analyzing other types of friendship networks. They focus on the votes from the 2005 rendition of the contest and come up with ways to adjust votes for song quality. With that, they produce a friendship network with valued links (the value of the link being the strength of the friendship). They find that friendships are often not returned, which reveals their asymmetric nature, especially visible in countries with a large number of immigrants.

They run a more statistical analysis by removing the influence of song quality or popularity and it shows that friendship between countries is determined in a big part by geographical proximity. Another factor they find are large immigrant groups voting for their home country. Other factors, such as population size, language similarity and economy were found to be insignificant. They expose a visible five-bloc structure, the blocs being the Eastern (former USSR countries, together with Romania, Hungary and Poland), Nordic (Norway, Sweden, Denmark, Finland

and Iceland), Balkan (former Yugoslavia and Albania), Eastern Mediterranean (Greece, Cyprus, Malta, Bulgaria and Turkey) and Western (Portugal, Spain, Ireland, Andorra, Israel, the UK, France, Monaco, Germany, Belgium and the Netherlands). Preferences among the different blocs are also analyzed, finding that some blocs are more connected than others. Grouping countries computationally by exposing the strongly connected components, they find three different blocs. Using taxonomic trees proved to be ineffective and only finding one bloc.

(Ginsburgh & Noury, 2008) analyze 29 years of the Eurovision Song Contest, specifically the competitions held between 1975 and 2003. Its main goal is to find any correlations between the points awarded and country similarity, performance type, etc. The authors find some meaningful properties impacting the scores and extract some clusters that exchange votes regularly. They propose what could lead to this behavior, stating that there exist cliques of countries that award points among themselves and even trading with votes. But these blocs are found not to base on politics, but rather on language and cultural similarities. To measure the language impact, they rely on the Morris-Swadesh method for analyzing linguistic differences.

To infer the influence of each factor, *Ginsburgh et al.* formulate a weighted expression, for which weights are assigned based on the voting behavior. The major takeaway of it is that the biggest factor influencing the voting decision is still the music quality. As with the previously discussed work, they also notice an important role of immigrants that vote for their country of origin. These observations are, however, not algorithmic but rather the results of looking at the formed communities and discussing the prevailing similarities in them.

3 METHODS

3.1 Bias detection

In the first major goal of the paper is determining the community structure of the voting networks. The most important step is the formation of edges that reflect a consistent bias between nations (both in terms of positive and negative relationships) and we approach that in two ways, described in this section. Both methods are used to detect bias over all the selected time periods. Altogether, this gives us more than 4400 different networks which are later used to present some statistical facts about the distribution of points.

```
function Gatherer (start_year, end_year)
  conf_high = bias threshold // 1 % in our case
  conf_low = neglect threshold // 90 % in our case
  avg_simulation = []
  // simulate voting enough times to obtain
  // a reliable expectation (100000 times in our case)
  for selected number of iterations:
    simulation = []
    for year in start_year..end_year:
      score = expected number of votes
              received by a contestant
              // depends on the voting scheme
      append(simulation, score)
    avg_sim = mean(simulation)
    append(avg_simulation, avg_sim)
  sort_ascending(avg_simulation)
  positive_bias = percentile(avg_simulation, conf_high)
  // more than bias number of points reflect biased voting
  negative_bias = percentile(avg_simulation, conf_low)
  // more than bias number of points reflect neglect
```

Figure 3.1 The Gatherer algorithm

Firstly, we follow the methodology described in (Mantzaris, Rein, & Hopkins, 2017), (Mantzaris, Rein, & Hopkins, 2018), (Gatherer, 2006) and use the Gatherer algorithm. This turned out to be the most effective method and very important for our analysis, thus, we describe the pseudo code in Figure 3.1 and

Figure 3.2. Its main idea is to estimate the number of points that a participant is expected to receive in a certain time period, based on the rules in place at that time. It then uses these estimates to find bias, i.e. behavior where nations exchange more than the expected number of points in a time period.

```

function Determine_Bias_Gatherer (start_year, end_year)
  period_length = start_year - end_year + 1
  participants = nations that took part in the ESC
  in the period
  for c1 in participants:
    for c2 in participants:
      if participated_together > period_length / 5
        // only take into account the participants that took
        // part in 20 % of all competitions in that period
        points_awarded_1, points_awarded_2 =
          number of points awarded by c1 to c2
          (and by c2 to c2) in the period
        threshold_high = the threshold number of points
        for that period showing bias
        calculated by the Gatherer algorithm
        threshold_low = the threshold number of points
        for that period showing neglect
        calculated by the Gatherer algorithm
        if points_awarded_1 > threshold_high and
        threshold_high > points_awarded_2:
          // one-way bias
          add edge (c1, c2) with weight
            (points_awarded_1 - points_awarded_2)
            to the directed bias network
        if points_awarded_1 > threshold_high and
        points_awarded_2 > threshold_high:
          // two-way bias
          add edge {c1, c2} with weight
            ((points_awarded_1 + points_awarded_2) / 2 -
            threshold_high)
            to the undirected bias network
        if c1 less than 3 hops away from c2:
          // only consider countries that are close
          // geographically to avoid noise
          if points_awarded_1 < threshold_low
          and points_awarded_2 < threshold_low:
            // two-way neglect
            add edge {c1, c2} with weight
              (threshold_low - (points_awarded_1
              + points_awarded_2) / 2)
              to the undirected neglect network

```

Figure 3.2: **Method of forming a bias voting network based on statistics calculated by the Gatherer algorithm**

The other method of obtaining the structure was developed by us and it accounts for the number of points a country has received each year in the selected period. We create a directed edge from country 1 to country 2 if the first one awarded the second one

more than the average number of points received by the second one in more than 75 % of the competitions in that period. If the bias is shown both ways, we add the edge to the undirected network. The pseudo code is described in Figure 3.3.

```

function Determine_Bias_Average (start_year, end_year)
overshot = defaultdict(int)
period_length = start_year - end_year + 1
threshold = 0.75
// threshold of how many times more than the average
// number of points need to be awarded for bias to occur
participants = nations that competed in the period
for year in start_year..end_year:
    determine the average number of points
    for each participant in the time period
    for c1 in participants:
        for c2 in participants:
            points_awarded_1, points_awarded_2 =
                number of points awarded by c1 to c2
                (and by c2 to c1) in the period
            determine how many times each country has awarded any
            other more than the average number of points
            received by the second in the time period
            appearances_1, appearances_2 =
                number of times c1 (and c2)
                participated in the ESC in the period
            overshoot_1, overshoot_2 =
                number of times c1 gave more than the average
                number of points received by c2 to c2 in the period
            if overshoot_1 > threshold * appearances_2
                and
                overshoot_2 < threshold * appearances_1:
                    // one-way bias
                    add edge (c1, c2) with weight
                        (overshoot_1 - overshoot_2)
                    to the directed bias network
            if points_awarded_1 > threshold * appearances_2 and
                points_awarded_2 > threshold * appearances_1:
                    // two-way bias
                    add edge {c1, c2} with weight
                        ((overshoot_1 + overshoot_2 -
                        threshold * appearances_1 -
                        threshold * appearances_2) / 2)
                    to the undirected bias network

```

Figure 3.3: Method of forming a bias voting network based on the average number of points received by the countries in the time period

We have generally found that the simulated voting implemented by the Gatherer algorithm gives clearer and less noisy results. It proves much more useful for detecting neglect, since the average points method generates too much noise. Even the graphs generated by the Gatherer algorithm were tricky to work with, which is why we added an additional criterion to detect a neglect between 2 countries. Since geographic proximity proved to be very important, we also demand that two neglecting countries lie no more than 3 hops (borders) away on the map thus reducing the amount of random edges.

Since the Gatherer algorithm is also used in (Mantzaris, Rein, & Hopkins, 2017), (Mantzaris, Rein, & Hopkins, 2018), (Gatherer, 2006), it is well tested and reliable. Therefore, we focus mainly on its results for graph formation from here on. Although we were able to extract some valuable information with the second method and it performed very similarly to the Gatherer algorithm for the longer periods, it behaves inconsistently on the shorter time spans, picking up too much randomness, while the Gatherer algorithm performs consistently no matter the period length, which led us to this decision.

We form two types of graphs: undirected, showing mutual affinity between contestants (i.e. an undirected edge between two nodes is added if both show bias towards each other), and directed, showing only one-way bias. Here, we are more interested in actual one-way relationships - an edge was therefore added only if one country shows bias towards the other, but the other does not show any bias for the first. Both graphs use weighted edges, the weight denoting the difference between the actual and expected (simulated / average) number of votes.

Despite the consistent performance by the Gatherer algorithm, it still needs some tuning. Besides the noise picked up in neglect detection, it also struggles on the directed networks, adding insignificant edges. This is why we also post-process the directed graphs and remove the edges whose weights were below the average in the network, giving us much more readable results.

3.2 Basic graph features and communities

In order to get the general oversight over the voting networks, we calculate some basic statistics about the voting and bias networks, such as the average number of nodes in a certain time period, the aver-

age number of edges and the average degree. All the methods are already implemented in the *networkx* library (Hagberg, Swart, & Chult, 2008).

After extracting the biased voting trends, we extract the communities using the Louvain (Blondel, Guillaume, Lambiotte, & Lefebvre, 2008) community detection algorithm in the undirected and the Newman's leading eigenvector method (Newman, 2006) in the directed ones. Both are implemented in the CDLib library (Rossetti, Milli, & Cazabet, 2019). The extracted communities in the undirected network depict blocs of countries "collaborating" in the competition. The directed networks are analyzed somewhat differently, since the actual communities do not play such a vital role here, as there is no mutual point exchange. However, they still expose some interesting behavior that would be missed if we only focused on the undirected networks. The results of both types are presented in Results.

The number of extracted graphs also allows us to find the most commonly co-occurring nations in communities. Those were extracted with the *apriori* algorithm, implemented in MLxtend library (Raschka, 2018).

The plots throughout the paper were generated with the Gephi visualization tool (Bastian, Heymann, & Jacomy, 2009).

3.3 Correlation between the community structure and success in the competition

One of the main goals of bias edge construction and community detection was determining the affect the bias behavior has on the final score of participants. In other words, we wanted to find out whether being in a large community or having many friends in the network pays off. Thus, based on the communities a node (country) belongs to, we gather some voting data (points, points from community, percentage of points from community and final place) and aggregate it for each community type and period length.

We find the most interesting aggregations to be points per degree in community, portion of points received from communities and total number of points received from communities. Therefore, we decided to interpret those more carefully in Results.

3.4 Preference detection and future projections

One of the hypotheses we set was that users and jury vote based on three main factors: the song popularity

and features, the country of the performance and the artist features. We split those categories to subcategories and obtain as much data as possible about them. We build a knowledge graph which connects all possible properties that can be considered together into relations of different types. With this graph we perform similarity scoring and link prediction where we try to predict the “voting relations” based on other connections. We use different link prediction algorithms which have to consider the rich structure of the formed graph. In addition to network analysis techniques, the dataset was also examined with the Orange package (Demšar, et al., 2013).

3.5 Prediction performance evaluation

In the second part of the analysis, we focus on the predictor of the success in the competition. The performance is measured against the performance of the betting tables. We use two different scores to calculate the success of the model. The first score is the mean absolute error (MAE) of the ranks inferred by the predictor based on the actual results and the second score the recall at n (Recall@ n) score for $n = 3, 5$ and 10 . MAE, too, is measured at distinct intervals: for the whole set of performing nations and just for the top 10 performances each year.

4 DATA COLLECTION AND PRESENTATION

4.1 Collection

The data set used was obtained by scraping various web pages. The voting data was collected from (European Broadcasting Union) and the information about specific countries, songs and performers was downloaded from (Flecht, 2019) (Wikimedia Foundation, Inc, 2019). The available voting data includes all points awarded by every country to every other participant, both for the final and the semi-finals (when both were held), with the exception of the first ever competition in 1956, since the data is not available. For the period between the years 2016 and 2019 we even got separated votes from jury and audience, since this is when the EBU started sharing these figures.

For most countries we obtained their names, Wikipedia category entries, languages, the currency, calling code, ethnic groups, religions, neighborhood. For the participants we have their country of origin, how old they were when they represented their nation, name, Wikipedia categories, music genres, in-

struments and occupations. Data for songs was scrapped from Wikipedia. We have among other things the genres, categories, languages and released date. To analyze songs even better we scrapped lyrics, chords and scores from (Fandom, Inc., 2019), (Musixmatch, 2019), (Musescore BVBA, 2019), (Naidev, 2019). The biggest challenge presented the data about songs, performers and performances themselves. We have tried to obtain as much as we could from Wikipedia, at least for the latest entries, which were better represented. We therefore focus mainly on those. We also try to extract some other important properties (the tones, harmony, metrum, melody...) about song quality from the chords and scores and the prevailing themes and motifs with text mining. Those features will be useful to pinpoint the preferences of specific countries and the factors that contribute most to success.

We have also obtained the betting tables for each competition between the years 2004 and 2019 (Eurovision World Betting Odds). These allow us to combine our models with the expected outcomes based on the betting odds. The data is stored in structured JSOG format (extended JSON format which can work with references and is therefore better for graphs). The scraping was done in Java, but data analysis is done in Python because of its numerous robust libraries for data management.

4.2 The inferred networks

Based on the collected voting information, we are able to form a large number of graphs, showing the voting behavior throughout the history of the competition.

Firstly, we just create the voting network for each contest separately and for all of them together (the all-time voting network). The networks for each competition are directed and any edge between two countries depicts the number of points awarded by one to another in that year. The in-strength of any node therefore shows its total score that year. Similarly, the all-time directed network depicts the total number of points awarded in the competition history.

These networks are then used to form the bias networks as described in Methods. To observe the changes throughout the competition history, we opt to form networks that represent biases and neglect in certain periods - those were chosen to be 1, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 60, 63 years. For each period,

both directed and undirected networks are created. This gives us more than 4400 networks altogether, but we do not need to analyze all thoroughly. The main focus are the networks that show the all-time preferences (period length 63 years), the ones that depict different 10 year periods, since this can show any changing nature of the voting, and the ones that depict the last 20, 25 and 30 year periods, showing long-term but still recent trends.

The all-time voting network has 52 nodes, one for each country that has ever competed. The 10 most successful (the nations with the highest number of points collected throughout the history) were Sweden, Norway, the UK, Germany, France, Spain, Denmark, Greece, the Netherlands, and Ireland. The countries with the lowest number of points obtained so far have been Monaco, Bulgaria, Australia, San Marino, Montenegro, Czech Republic, Slovakia, Andorra and Morocco. However, these scores should not be too surprising and taken too seriously, since the most successful nations are also the ones that have participated in the competition the longest and many of the least successful ones have only taken part a few times. On the other hand, Australia has only participated 5 times so far and has achieved great success each time, which cannot be captured with this kind of analysis.

4.3 Betting tables accuracy

The baseline for measuring the performance of our prediction model was using the betting tables as the only means for predicting the outcome of the competition. Thus, this baseline needed to be determined. The results were obtained using the performance metrics described in Preference detection and future projections, averaged over the whole period for which we have obtained the betting tables. To limit how much was known about the outcome of the competition when the tables were updated, we only included data from 20-35 days before the final rounds of the competition. They are presented in Table 4.1. We can see the MAE of the tables improves for the higher part of the table, while the recall does not seem to be affected much by the range. These figures are the

baseline for our model, which provided results described in Prediction performance evaluation.

Tabela 4.1: Performance evaluation of betting tables as predictors.

Performance measure used	Results
Mean averaged error over the whole set	4,3391
Mean averaged error over the top 10	4,0421
Recall@3	0.4386
Recall@5	0,54737
Recall@10	0,56842

5 RESULTS

The results present the communities of positive bias, countries showing neglect, the correlation between the community structure of a country and its success, what we think causes this behavior, the inferred preferences of specific countries and the prediction results.

5.1 Communities

The number of generated networks makes it possible to reason about the different trends and influences on the voting. Although we could have focused on any period in the competition history, we chose to further inspect the most recent results and mostly summarize the older.

Figure 5.1 shows the communities formed if we consider the results from the start of the competition in 1956. There are 10 communities in total and they are strongly geographically influenced, forming the following blocs: Northern (Sweden, Denmark, Norway, Iceland), Western (Ireland, United Kingdom, Germany, Luxembourg), Southern (Italy, Malta, Spain, Portugal), Central (Netherlands, Hungary, Belgium, Austria), Baltic (Lithuania, Estonia, Latvia, Finland), Eastern (Poland, Ukraine, Russia, Belarus), the Balkan (Greece, Cyprus, Albania), South-Western (Moldova, Romania, Turkey), Yugoslavian (Croatia, Slovenia) and Cross-Continental (Israel, France). Especially prominent are the connections between Cyprus and Greece, Greece and Albania, Romania and Moldova, Italy and Malta and the former USSR countries.

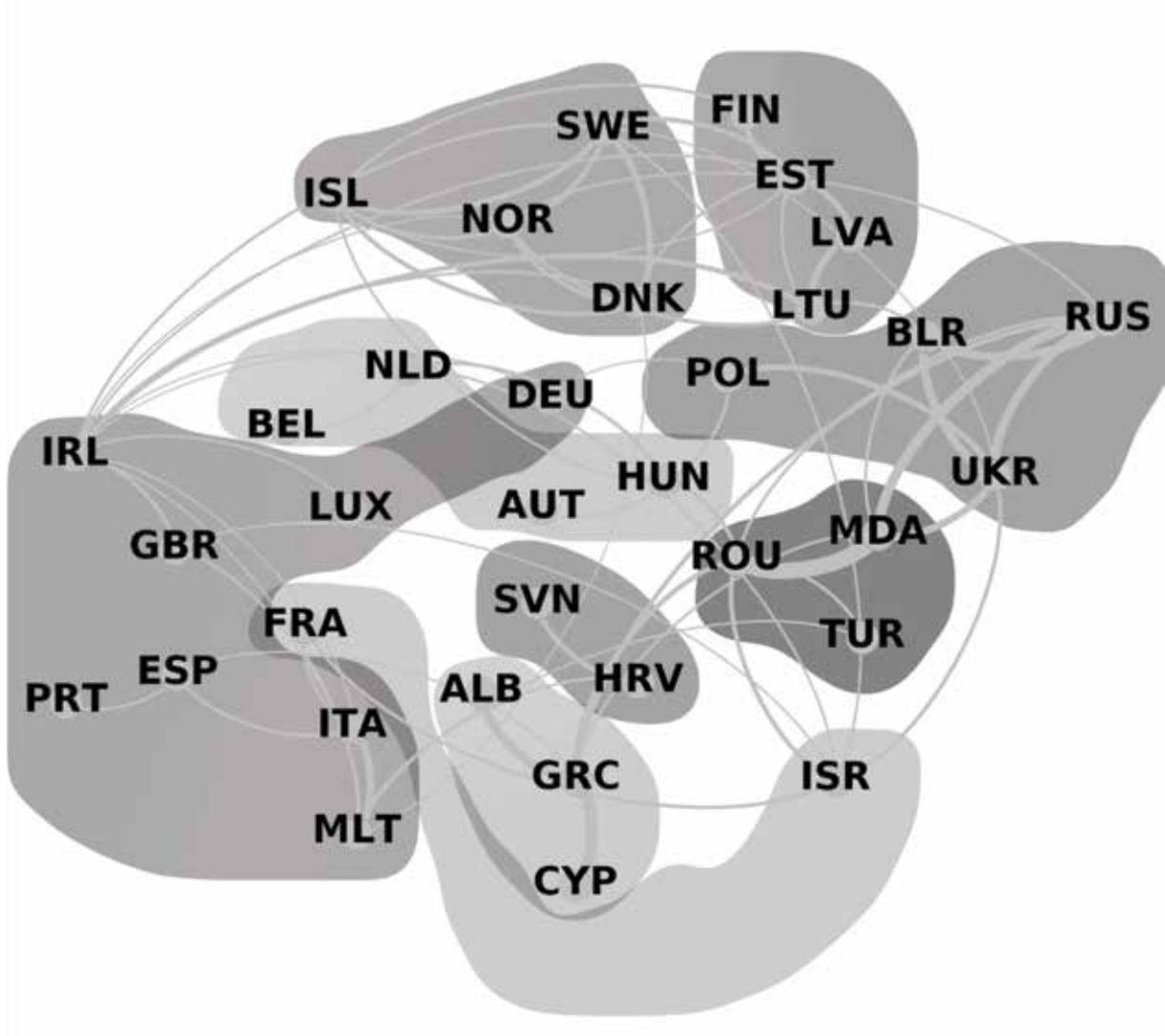


Figure 5.1: **Bidirectional bias from the start of the competition.**

Although this is the network that includes the most data and is thus seemingly the most important, we only mention it here for the sake of completeness. We are more interested in the networks depicted in

Figure 5.2, Figure 5.3 and Figure 5.4, and for the later parts of the analysis as they speak of the more recent trends.

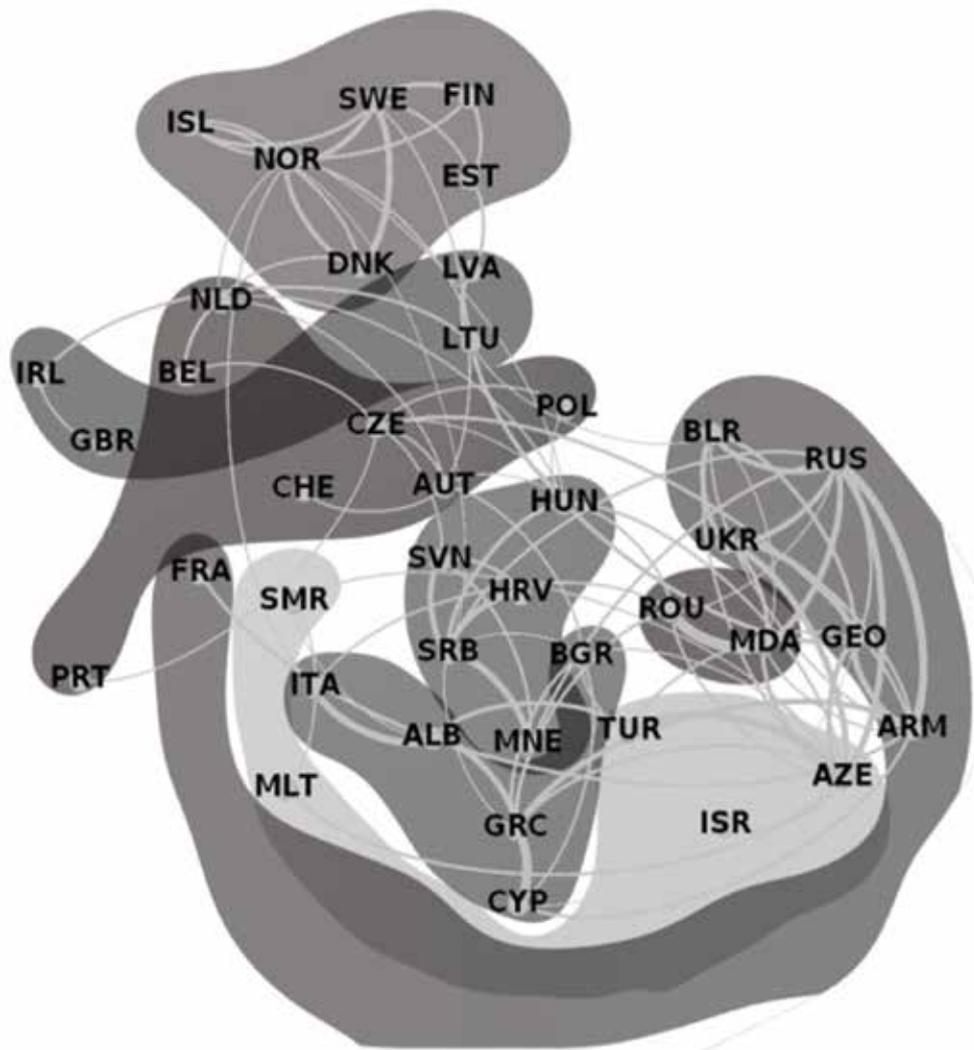


Figure 5.2: Bidirectional bias from the last 20 years (1999-2019).

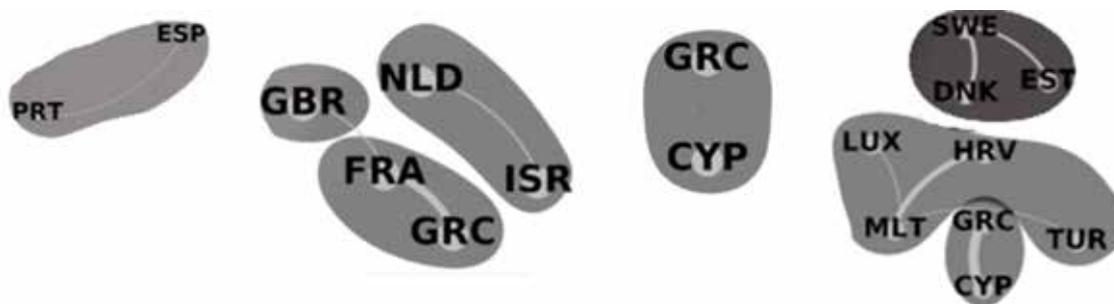


Figure 5.3: Bidirectional bias in 10-year periods (1959-1969, 1969-1979, 1979-1989, 1989-1999).

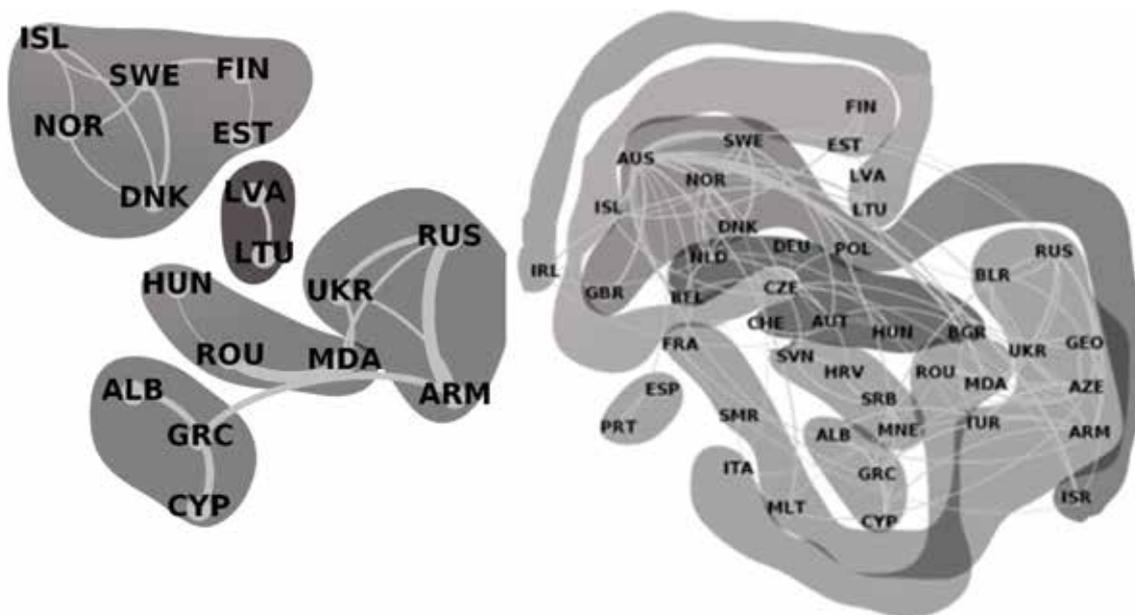


Figure 5.4: Bidirectional bias in 10-year periods (1999-2009, 2009-2019).

Figure 5.3 and Figure 5.4 show how the bias networks have evolved and grown, although the main communities remain the same. Clearly, there is more and more biased voting, but it remains concentrated in the same blocs in all periods. It is interesting to see how Australia got mixed into the Northern bloc in the last 10 years. This may be one of the reasons for their reasonable success so far. During the first five time they have taken part in the competition, they showed a very focused voting behavior and at the same time managed to collect many points from until then a very closed bloc.

We think the most interesting and current network is the one in Figure 5.2, since it shows the recent trends, while still taking into account a longer time period. The communities are very similar to the ones implied by the all-time bias network, showing the persistence of these relationships.

Also interesting is the network shown in Figure 5.5, showing the all-time network of one-way relationships. The edges depict relationships where only one country awards more than average number of points and the other do not. Communities are not that prominent in this network, but still visible. One reason why the community structure is limited in the fact that historically very successful countries such as Sweden receive a high number of points from others

very often and they cannot “return” the votes to all of them. Therefore, they have a very high in-degree and this does not infer any preference, just the fact that they were successful. However, some relationships in the all-time network are still quite interesting, like the strong edge from Croatia to Bosnia and Herzegovina and the edges from the former USSR nations to Russia.

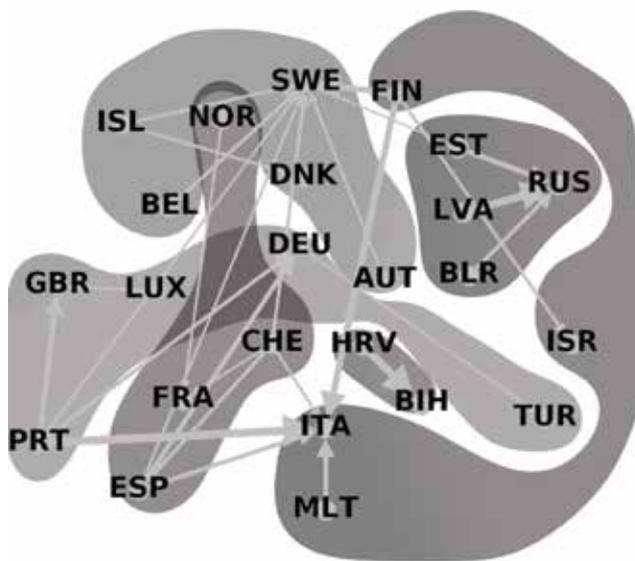


Figure 5.5: Unidirectional bias from the start of the competition.

The data also allows us to find sets of countries that end up in the same communities most often. The results are presented in Table 5.1. As the table shows, the countries that co-occur in a community most often are Cyprus and Greece, which are a part of more than 10 % of the formed communities. They are followed by some Scandinavian countries and the most regular participants in the competitions, such as the UK, Ireland, and Switzerland. The most common set of size three contains Denmark, Sweden, and Norway. We also notice a strong relationship between Portugal and Spain, Romania and Moldova, Slovenia and Croatia and, interestingly, France and Israel. All the relationships are also visible in the figures below.

The graphs in Figure 5.3 and Figure 5.4 provide a different view as to how the bias has evolved throughout the history and it is clear that there are more and more biased connections. This can also be seen if we look at the average degree of the bias undirected network throughout the history, depicted in Figure 5.6. The degree has been rising consistently, which means that the countries are actively forming more and more friendship communities and concentrating their votes among specific “partners”.

Tabela 5.1: Countries that ended up in the same community most often and the relative number of times.

Rank	Countries	Relative support
1	Cyprus, Greece	0,109
2	Denmark, Sweden	0,081
3	Sweden, Noreway	0,069
4	Switzerland, United Kingdom	0,066
5	Denmark, Norway	0,062
6	United Kingdom, Ireland	0,062
7	Denmark, Sweden, Norway	0,055
8	Spain, Portugal	0,053
9	Sweden, Iceland	0,043
10	Germany, United Kingdom	0,042
11	Denmark, Iceland	0,041
12	Romania, Moldova	0,037
13	Belgium, Netherlands	0,036
14	Slovenia, Croatia	0,035
15	Israel, France	0,034
16	Denmar, Sweden, Iceland	0,033
17	Norway, Iceland	0,033
18	Germany, Iceland	0,033
19	Finland, Sweden	0,029
20	Estonia, Litvia	0,028

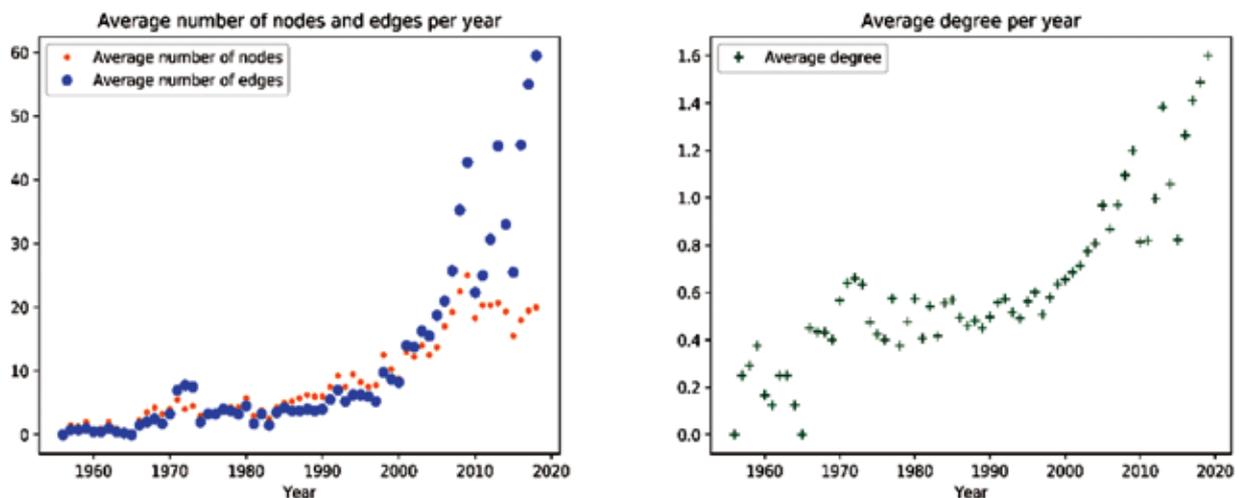


Figure 5.6: Average number of nodes and edges, average degree and clustering coefficient in the bias networks throughout the years.

5.2 Correlation between the community structure and success

Table 5.2 shows what percentage of points countries got from their communities. For example, averaged over 25 years, members of neglect communities only received 6.1 % of their points from that community, while members of communities in directed graphs get on average 17.6 % of their points from that cluster.

Table 5.3, Table 5.4 and Figure 5.7 show a recurring relationship between the success of a nation and its community structure. Being in a positive bias community pays off because countries get on average higher scores and achieve higher places, a trend clearly visible in the plots. We can also see that it is better to avoid neglect clusters, since membership in those usually means lower ranking.

Table 5.2: Average percentage of points received from communities..

Period	bidirectional bias		bidirectional nelgct	
	Average	STD. deviation	Average	STD. deviation
1	192.80	121.30	10.00	0.00
5	532.97	272.12	179.53	122.62
10	812.57	402.08	334.40	216.89
20	1160.40	466.33	656.59	485.73
63	2877.68	732.65	2151.29	871.49

Table 5.2: Average percentage of points received from communities..

Period	bidirectional bias		bidirectional nelgct	
	Average	STD. deviation	Average	STD. deviation
1	0.22	0.10	0.03	0.05
5	0.18	0.09	0.01	0.06
10	0.18	0.09	0.03	0.04
20	0.17	0.09	0.05	0.06
63	0.15	0.06	0.09	0.06

Table 5.3: Average place.

Period	bidirectional bias		bidirectional nelgct	
	Average	STD. deviation	Average	STD. deviation
1	4.02	8.85	1.00	0.00
5	11.28	12.19	11.24	7.41
10	14.76	11.74	19.94	10.59
20	15.29	10.93	23.03	11.22
63	18.47	10.45	28.42	11.56

Table 5.4: Average number of points.

Period	bidirectional bias		bidirectional nelgct	
	Average	STD. deviation	Average	STD. deviation
1	192.80	121.30	10.00	0.00
5	532.97	272.12	179.53	133.62
10	812.57	402.08	334.40	216.89
20	1160.40	466.33	656.59	458.73
63	2877.68	732.65	2151.29	871.49

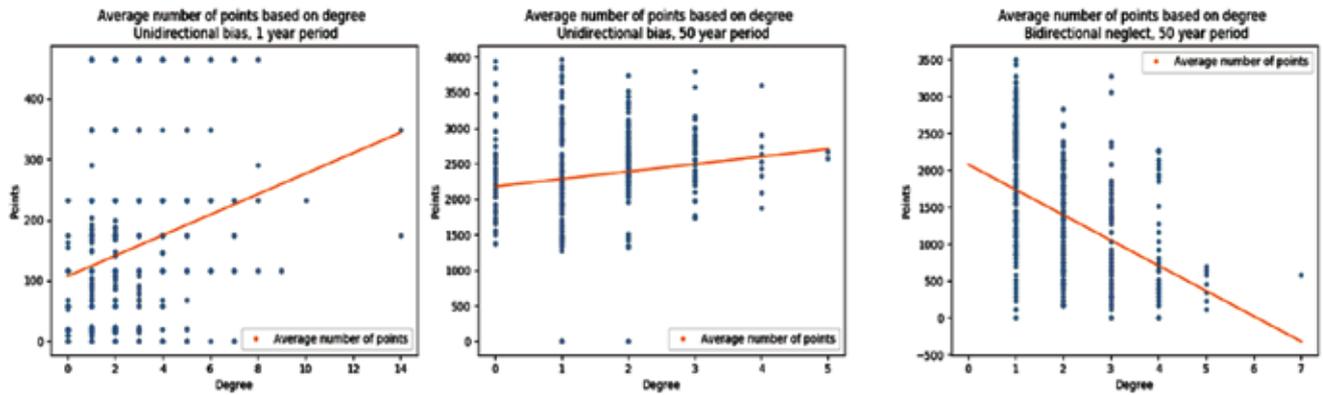


Figure 5.7: Relationship between the degree in the bidirectional bias (left and center) and neglect (right) and the number of points received in the last 50 years.

5.3 Neglect

We form the neglect networks in a similar manner to the positive bias ones and the results are shown in Figure 5.8, Figure 5.9 and Figure 5.10. As expected, some distinct neglect relationships are visible between nations, most notable between Macedonia and

both Greece and Cyprus. Similar holds for the pair Cyprus and Turkey and more recently, for Azerbaijan and Armenia. Interestingly, there is also a strong evidence of neglect between Germany and the pair Belarus and Ukraine. As seen in the more recent networks, the trends persist.

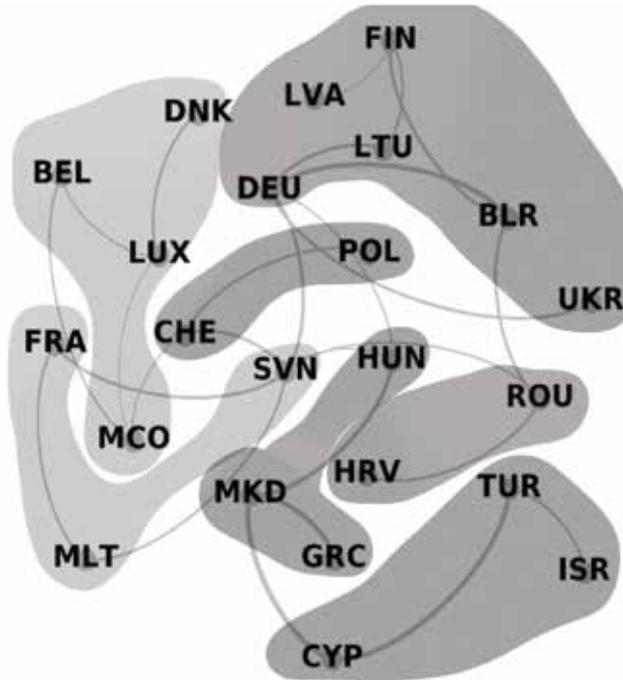


Figure 5.8: Bidirectional neglect from the start of the competition.

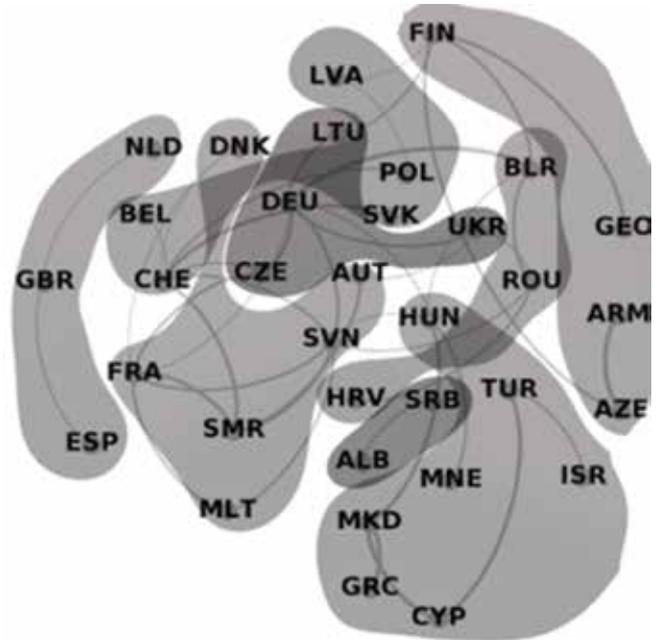


Figure 5.9: Bidirectional neglect from the last 30 years (1989-2019).

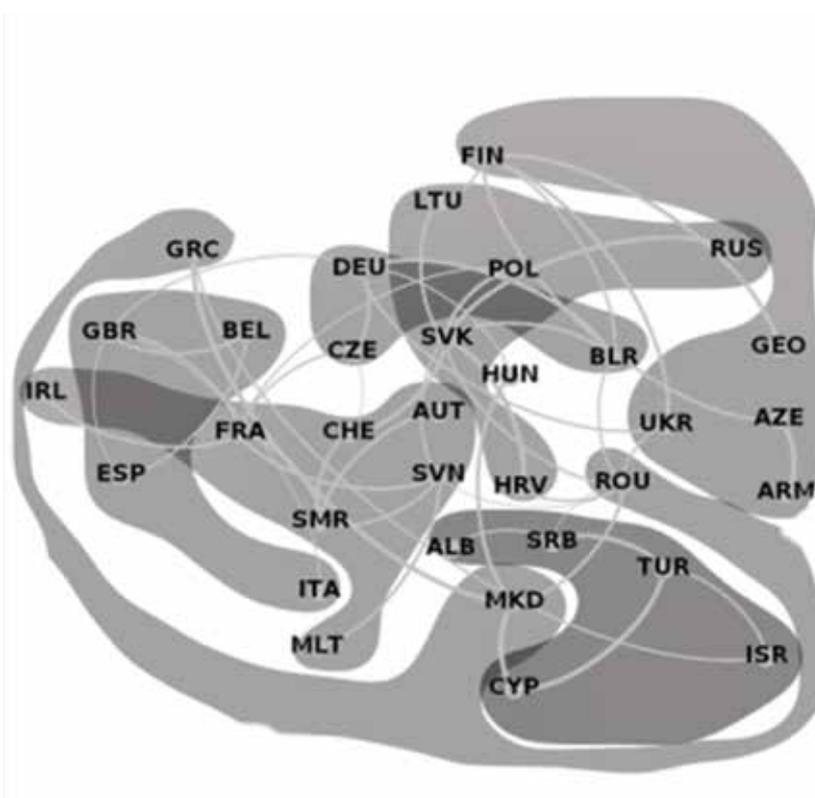


Figure 5.10: Bidirectional neglect from the last 10 years (2009-2019)

5.4 Possible influences and motivations

As found in the discussed literature, geographical proximity seems to influence the voting behavior most, as can be seen through the geographically local communities that form. Moreover, affinity between nations such as the UK and Malta stress that language similarity also plays a role. Common historical background could be attributed to the affinity between the former Yugoslavian and USSR nations, since the communities rarely extend beyond the bounds of the former unions.

The one-way relationships are trickier and less obvious. However, they can be explained to some extent by the number of immigrants (e.g. votes from Croatia to Bosnia and Herzegovina, Germany and France to Turkey and Switzerland to Serbia) and historical significance of one country to the other (e.g. the votes from the former USSR countries to Russia). Other reasoning is hard to ground since the highest in-degrees can be explained purely on the success in the competition.

It is worth noticing that the positive bias behavior is most strongly represented by pairs of nations that are more isolated, either geographically (e.g. Spain and Portugal, the UK and Ireland, Romania and Moldova, the Scandinavian countries), or culturally (e.g.

Cyprus and Greece, Greece and Albania, the Baltic countries and, again Romania and Moldova).

The countries showing most neglect have notable reasons as well. Especially the historical relationship between Macedonia and Greece and more recently between Albania and Serbia can be explained by their non-friendly neighborhood relations.

5.5 Nation's music preferences

The constructed knowledge graph and Orange visualization tools offer a glimpse into what genres, music styles and other performance features caught the voters' attention. Unfortunately, due to the lack of data, we are only able to extract the crudest of relationships, thus, we do not discuss them here thoroughly. Some trends we observe, though, are the fondness of Slovenia towards Croatian songs (both in the form of the language and the origin), Australia towards songs in English and we again confirm the strong relationship between Greece and Cyprus.

5.6 Prediction performance evaluation

The performance measures indicate that the built model did not increase the accuracy of the betting tables. Much of this can be attributed to the fact that the data was often very sparse and not structured

very well. Even after preprocessing and filtering the whole dataset, we were still left with too many unreliable and altogether not very useful entries.

In Table 5.5 we report the performance measures when we also consider the predictor data together with the data from the betting tables in variable amo-

unts. The hyperparameter β indicates how much the predictions made by our model are taken into account ($\beta = 0$ means only the betting tables are used and $\beta = 1$ means we rely only on our predictor). We can see, the predictor does not improve the betting tables performance.

Table 5.5: Performance evaluation of our prediction model.

Performance measure used	Results	Results	Results
Mean averaged error over the whole set	6.00	6.52	6.76
Mean averaged error over the top 10	6.48	7.46	7.80
Recall@3	0.14	0.09	0.07
Recall@5	0.23	0.20	0.12
Recall@10	0.42	0.35	0.31

6 DISCUSSION

This section presents some of the problems faced during the analysis and concludes the paper with our closing remarks.

6.1 Problems and compromises

The incompleteness of the data has turned out to be a problem very early on, as we were initially unable to construct graphs based on some similarities, namely the ethnic groups, immigrant numbers and economic exchange. We thus resorted to manual inspection of the probable causes of some trends. The inferred relationships are thus based only on our domain knowledge and presumptions.

As expected, the availability of the data about the performances, songs and authors is also limited, but we have managed to obtain a reasonably diverse and complete dataset and hoped we could make use of it, especially in the second part of the paper.

Another problem we encountered was the noise in the less robust networks such as the directed ones and the ones dealing with neglect. They needed a lot of tuning and some post-processing to present any usable information, but the final outcome is still quite non-deterministic and open to numerous interpretations.

Motif counting and detection was also found to be not as effective as we had hoped. The process of extracting the motif structure itself was not very straight-forward since the functionality is not as widely implemented as some other tools and at the same time the results were not as informative and

interpretable as the community structure itself. For example, the notion of the reciprocal point exchange is summed up in the undirected positive bias networks. Thus, we think that a thorough inspection of the motif structure would not provide better enough understanding of the network. We therefore abandoned this idea and focused on other analysis tools.

If we were able to manage the dataset deficits in the first part of the paper, they really came forward in the second part, since the shortcomings disabled us to build a valid and useful model for prediction. We leave this feat for future work.

6.2 Future work

During the analysis, we came across a few possible applications to other fields. Firstly, the ideas and method discussed here do not necessarily apply only on the Eurovision voting network. Such analysis can be applied to any voting system, especially ones with a smaller number of voting entities, such as the participating countries discussed in this paper. We would find analysis of the voting behavior in sports where points are awarded by judges from different countries very interesting. Similarly, taking a closer look at the voting for awards would presumably reveal interesting trends. One of such awards is the Ballon d'Or prize in soccer, where journalists and players from around the world vote for the best footballer each year. Each nation is represented by its journalists and players, which is similar to the voting structure of the ESC.

A different field we would also be interested in is the voting a political environment such as the European Parliament. As representatives from the whole EU vote for propositions which come from different backgrounds, one might find some trends in the way the representatives from specific countries vote. Lastly, we consider our own implementations and dataset. Some methods we implemented did not take into account all the specifics in the ESC dataset (e.g. the change of Macedonia to North Macedonia was handled manually) and could be extended to further increase the result reliability. One of the main objectives for future work would also be the aforementioned expansion of the dataset that could allow a better model of the behavior.

6.3 Summary and conclusions

In this paper, we analyzed the trends in the ESC voting network. The results show strong and recurring patterns of mutual point exchange between neighboring countries. We observed the most commonly recurring friendships and one-way relationships together with some persistent behavior of neglect. As discussed in the previous work, they can be explained by geographical proximity and language similarity, as well as ethnic structure and historical bonds. Having a large number of biased relationships positively correlates to the success in the competition and we observed more and more relationships the more recent years. Isolation of sets of countries seems to make bonds among the members of the set stronger.

We also described the methodology used in more detail and explained how the data was structured to obtain the information. The obtained data was then used to build predictor for future contests. To the extent possible, we leveraged the distinct music preferences of individual nations to extract which genres and music styles achieve the greatest success in different countries. This involves both the points given by the nation to other countries for their performances and also their representative artists. This data was combined with betting tables, since they are widely considered to be the best predictors about the success of participants. The resulting model did not outperform the betting tables alone with its main weakness being the lack of reliable data.

We look forward to future extensions of our work on similar fields or the same project with a more promising prediction model.

REFERENCES

- [1] Bastian, M., Heymann, S., & Jacomy, M. (2009, 3). Gephi: An Open Source Software for Exploring and Manipulating Networks. doi:10.13140/2.1.1341.1520
- [2] Blondel, V., Guillaume, J.-L., Lambiotte, R., & Lefebvre, E. (2008, 4). Fast Unfolding of Communities in Large Networks. *Journal of Statistical Mechanics Theory and Experiment*, 2008. doi:10.1088/1742-5468/2008/10/P10008
- [3] Dekker, A. (2007, 1). The Eurovision Song Contest as a 'Friendship' Network 1. *Connections*, 27.
- [4] Demšar, J., Curk, T., Erjavec, A., Gorup, Č., Hočevar, T., Milutinovič, M., . . . Zupan, B. (2013). Orange: Data Mining Toolbox in Python. *Journal of Machine Learning Research*, 14, 2349-2353. Retrieved from <http://jmlr.org/papers/v14/demsar13a.html>
- [5] EBU. (2019, October 20). *European Broadcasting Union*. Retrieved from European Broadcasting Union: <https://www.ebu.ch/home>
- [6] EBU. (2019, October 20). *Eurovision Song Contest*. Retrieved from Eurovision Song Contest: <https://eurovision.tv/>
- [7] Eurovisionworld. (2019, October 20). *Eurovision World Betting Odds*. Retrieved from Eurovision : <https://eurovisionworld.com/odds/eurovision>
- [8] Fandom, Inc. (2019, October 20). *LyricWiki*. Retrieved from Lyrics Fandom: <https://lyrics.fandom.com/wiki/LyricWiki>
- [9] Licht, M. (2019, October 20). *ESCHome*. Retrieved from ESCHome: <https://eschome.net>
- [9] Gatherer, D. (2006, 3). Comparison of Eurovision Song Contest Simulation with Actual Results Reveals Shifting Patterns of Collusive Voting Alliances. *Journal of Artificial Societies and Social Simulation*, 9.
- [10] GeoDataSource. (2019, October 16). *Country Borders*. Retrieved from GitHub repository: <https://github.com/geodatasource/country-borders>
- [11] Ginsburgh, V., & Noury, A. G. (2008). The Eurovision Song Contest. Is voting political or cultural? *European Journal of Political Economy*, 24, 41-52. doi:<https://doi.org/10.1016/j.ejpoleco.2007.05.004>
- [12] Hagberg, A., Swart, P., & Chult, D. (2008, 1). Exploring Network Structure, Dynamics, and Function Using NetworkX.
- [13] Lao, N., Mitchell, T., & Cohen, W. (2011, 1). Random Walk Inference and Learning in A Large Scale Knowledge Base., (pp. 529-539).
- [14] Mantzaris, A. V., Rein, S. R., & Hopkins, A. D. (2017). Examining collusion and voting biases between countries during the Eurovision song contest since 1957.
- [15] Mantzaris, A. V., Rein, S. R., & Hopkins, A. D. (2018, 9 01). Preference and neglect amongst countries in the Eurovision Song Contest. *Journal of Computational Social Science*, 1, 377-390. doi:10.1007/s42001-018-0020-2
- [16] Musescore BVBA. (2019, October 20). *Musescore*. Retrieved from Musescore: <https://musescore.com>
- [17] Musixmatch . (2019, October 20). *Musixmatch*. Retrieved from Musixmatch : <https://www.musixmatch.com/>
- [18] Naidenov, E. (2019, October 20). *Ultimate Guitar*. Retrieved from Ultimate Guitar: <https://www.ultimate-guitar.com/>
- [19] Newman, M. (2006, 10). Finding Community Structure in Networks Using the Eigenvectors of Matrices. *Physical review. E, Statistical, nonlinear, and soft matter physics*, 74, 036104. doi:10.1103/PhysRevE.74.036104
- [20] Raschka, S. (2018, 4). MLxtend: Providing machine learning and data science utilities and extensions to Python's scientific computing stack. *The Journal of Open Source Software*, 3, 638. doi:10.21105/joss.00638

- [21] Rossetti, G., Milli, L., & Cazabet, R. (2019, 12). CDLIB: a python library to extract, compare and evaluate communities from complex networks. *Applied Network Science*, 4. doi:10.1007/s41109-019-0165-9
- [22] Spierdijk, L., & Vellekoop, M. H. (2006, 2). *Geography, culture, and religion: Explaining the bias in Eurovision song contest voting*. University of Twente, Department of Applied Mathematics.
- [23] Wikimedia Foundation, Inc. (2019, October). *Wikipedia*. Retrieved from Wikipedia: <https://en.wikipedia.org/>

■

Anej Svete končuje dodiplomski študij računalništva in matematike na Fakulteti za računalništvo in informatiko Univerze v Ljubljani. Zanima ga splošno področje podatkovnih ved, še posebej obdelava naravnega jezika, uporaba umetne inteligence v robotiki in spodbujevano učenje, kar je tema njegovega diplomskega dela. Med študijem izkušnje nabira tudi pri razvojnem podjetju XLAB d.o.o. Magistrski študij nadaljuje na področju podatkovnih znanosti na univerzi ETH.

■

Jakob Hostnik je študent prve stopnje univerzitetnega študija računalništva in matematike na Univerzi v Ljubljani, je ustanovitelj podjetja Etilk d.o.o. in v podjetju HPE d.o.o. devops in tehnični vodja ekipe. Profesionalno in zasebno se najbolj ukvarja managementom, arhitekturo programske opreme, devops delom in računalništvom v oblaku.

■

Lovro Šubelj je docent za računalništvo na Fakulteti za računalništvo in informatiko Univerze v Ljubljani. Diplomiral je leta 2008 na Fakulteti za matematiko in fiziko in Fakulteti za računalništvo in informatiko ter doktoriral leta 2013 na temo analize velikih omrežij. Je avtor ali soavtor več kot šestdeset znanstvenih prispevkov in patentov ter urednik prestižnih mednarodnih znanstvenih revij. Njegovo preteklo delo je bilo izbrano kot izjemen znanstveni dosežek v Sloveniji ter predstavljeno na uglednih mednarodnih univerzah kot sta Stanford in UCSD. Sodeloval je že pri številnih uspešno zaključenih raziskovalnih in razvojnih projektih v sodelovanju s podjetji Petrol, Celtra, Optilab, Iskratel in drugimi.