Not Your Toy? - Is The Eurovision Just A Song Contest?

Ori Heldman, Koren Gershoni, Efrat Ravid Ben Gurion University of the Negev
SISE Dept.
Be'er Sheva, Israel
heldman@post.bgu.ac.il; korenger@post.bgu.ac.il; efratrav@post.bgu.ac.il

1 Introduction

The Eurovision Song Contest (ESC) is one of the most successful and longest-running international music competitions and television shows worldwide. Though the organizer of the contest has long marketed it as a nonpolitical and unifying force, and although the ESC is a musical entertainment event, there have always been social, cultural, and political themes and tensions that simmer at the edges of the contest (Jackson 2017).

A common public question that is raised every year around the ESC is how the voting behavior is affected by other factors than performance and song quality [1]. For example concerns were raised that votes were impacted by social factors such as stance against the Israel-Palestinian conflict, LGBT support, the Brexit and more. In addition, previous studies have empirically examined the affect of cultural alliances, cliques and the extent of political bloc (Yair 2019; Fenn et al. 2006). (Charron 2013) examined voting blocs that are based on mutual culture and historic friend partners like the Balkan or ex-Soviet countries. In this project we wish to spread more light on the hidden motives in the ECS votes.

First, we wish to understand if and how much social subjects are taking into account in voting consideration. As social conversations have moved online to social media platforms where citizens discuss everything from song reviews to politics and shape their opinion; we wonder if the social conversation regarding the ESC revolves less around the song quality and performance and more on different social topics. As the social conversing in such platforms plays a key factor in shaping the public opinion, it may also play a key role in shaping the voting behaviors in the ESC.

Hence, we wish to evaluate if social subjects are taking a significant role in the discussion of ESC songs. We collected Youtube comments on Eurovision songs between the years 2017-2019 and 2021. Using the comments we trained a BERT based model ((Devlin et al. 2018)) to classify if a comment is discussing the song quality/performance or not. The analysis quantifies the differences between different countries in different years, reflecting how much of the discourse was related to the act or on other topics.

Next, empirically evaluate specific political and social factors as indicators of voting behaviours. Our hypothesis is that these indicators can reflect on the voting behavior in the ESC. Specifically, we focused on economic and tourism relationships between countries, as both reflect social and political relationships and have temporal representations like the ESC votes.

In addtion to the annual ESC voting results, we collected annual trade and tourism statistics between ESC countries and evaluated their correlations with the jury and public telephone voting behavior.

Formally, we ask the following research questions:

RQ1. Are social issues becoming a vital factor in the ESC voting patterns?

RQ2. Are economic relationships of countries reflected in their ESC voting behavior?

RQ3. How are tourism relationships of countries reflected in their ESC voting behavior?

Relevant background on the ECS can be found in section 2. Section 3 contains details on the datasets we collected for this project, their processing and the methods we used for answering the above research questions. Section 4 contains the results of the methods carried out. Finally, we discuss the results and future work in section 5.

[1] https://www.washingtonpost.com/opinions/2019/05/13/eurovision-is-political-this-year-it-is-every-year/

2 Background

2.1 Eurovision Song Contest

Active country members of the European Broadcasting Union (EBU), as well as invited associate members, are eligible to compete, and as of 2021, 52 countries have participated at least once. Each participating broadcaster sends one original song. Each country awards two sets of 1–8, 10 and 12 points to their favourite songs. The first set is the jury votes, based on the views of an assembled group of music professionals chosen by each country. The second set is the public televoting set. The song receiving the most points is declared the winner.

Since 2008 each contest is typically formed of three live television shows held over one week: two semi-finals are held on the Tuesday and Thursday, followed by a grand final on the Saturday. All participating countries compete in one of the two semi-finals, except for the host country of that year's contest and the contest's biggest financial contributors known as the "Big 5" — France, Germany, Italy, Spain and the United Kingdom. The remaining countries are split between the two semi-finals, and the 10 highest-scoring entries in each qualify for the final to produce 26 countries competing in the grand final.

3 Methods

3.1 Datasets

For answering our research questions, we combined several sources of data.

ECS Votes The 'Eurovision Votes' dataset contains the ESC vote records from 1975-2019[2]. The vote records in the dataset contain both the semi-final and the final from both the jury and the public televotes. Recall that since 2016 each country awards two sets of points, professional jury, and general public through telephone. For our research, we updated the dataset with the latest Eurovision 2021 vote records from a Eurovision fan community website[3].

Europe Trade Dataset The 'Eurovision Trade' dataset contains the trade statistics between Eurovision participating countries between the years 2013-2019. Using the 'World Integrated Trade Solution'(WITS) [4] database, we collected the exported trade statistics for each Eurovision participating country for each year mentioned above. Specifically, the 'Eurovision Trade' dataset contains the export goods in USD to each Eurovision country and the entire world. The WITS does not contain trade statistics on San Marino, which is one of the participating countries in the ESC.

Europe Tourism Dataset The 'Eurovision Tourism' dataset contains the tourism statistics between Eurovision participating countries between the years 2013-2019. Using the 'Eurostats' database[5] of the number of trips by country / world region of destination, we collected the tourism statistics for each Eurovision participating country for each year mentioned above.

For both the Europe trade dataset and the Europe tourism dataset we do not collect the data after 2019 for 2 reasons; first, due to COVID-19 the ECS was not held in 2020. In addition, there is no data available yet for 2021.

Eurovision YouTube Channel Each year before the contest dates, the official Eurovision Song Contest YouTube channel[6] uploads official videos of the songs participating in the contest. The official videos for each song are uploaded approximately 2 months before the contest and serves as a mean for the audience to get to know the competitors from each country before the competition is broadcasted. Commenting on the videos is more available than voting in the competition because it is not bounded by time. In addition, during the semi final, only votes from countries that participated in the specific semi final are considered and in the finals you can only vote for countries who were qualified for the finals. These limitations do not hold for commenting on the official YouTube videos.

We used the Google Client API to download the comments on the official videos from the competition in 2017-2019 and in 2021. In addition, we used the Google Cloud Translation API to translate all comments to English. In total, this dataset contains over 1M comments from 166 videos.

3.2 Performance Related Discourse

Every year the European Broadcasting Union publish the songs of the countries in their YouTube channel. YouTube users can leave a comment on any song that they want and reply to each other as well. Part of the comments are song related and the other part are related to other topics (e.g. greeting, politics, etc.) We may expect that songs with a high pro-

[2]https://www.kaggle.com/datagraver/eurovision-song-contest-scores-19752019

[3]https://eurovisionworld.com/eurovision/2021

[4]https://wits.worldbank.org/countrystats.aspx?lang=en

[5]https://ec.europa.eu/eurostat/web/main/data/database

[6]https://www.youtube.com/c/EurovisionSongContest/featured

Song	Year	Youtube Comment	Label
Toy	2018	Why it is so hard for people to understand that the chicken voices refers to the man, who is a chicken, coward because he treats a woman like a toy...	0
I Won't Break	2018	Julia, why are these lips, these cheekbones? Who is Nicki Minaj? The makeup is ugly for Julia.	0
Arcade	2019	Very good song! If I didn't know it was a part of Eurovision song contest I'd say it was made by top world hitmakers! 12 points from Croatia! ;)	1
I Don't Feel Hate	2021	The worst song Ive ever heard, like it was written by a child with developmental disabilities ...	1

Table 1: Examples of labeled YouTube comments.

portion of song related comments will get a good score from the jury/televoting because unrelated comments tends to appear when the song has a poor performance or other reasons such as political issues.

We manually labeled over 600 comments with a binary label that reflects if the comments is related to the song/performance (1) or not (0). Table 1 contains examples of labeled comments. As can be seen in the table, the first two comments received a label of 0 as they do not talk about the song quality or performance. For example, the second comment topic is about the 'looks' of the singer, giving criticism about makeup. On the other hand, the third and forth items in the table are labeled 1 as they are talking about the song quality either positive or negative. For example, the forth comment gave criticism to the lyrics of the song.

Based on these labels, we fine-tune a BERT model to classify whether or not a comment is related to the song or to the performance. The model is then used to generate a weak label for each of the comments in the dataset in order to evaluate to what extent other topics are reflected in the discourse on the songs in the ECS.

Alongside to looking at the proportions of song related comments to non song related comments, we split the comments into 2 groups according to the labels obtained by our classifier. We use word cloud generators on the groups to visualize the most frequent words and phrases in the comments. We created word clouds using all the comments of the country's videos, one using song quality comments and another using non-song quality. Next we created pairs of word clouds for every year to understand the change over time.

3.3 Measuring Social Effects

Our research question aims to find hidden motives to the voting patterns in the ECS. We hypothesise that the trade relations between countries and the tourism patterns between the countries affect votes in the contest.

Correlation Analysis If trading relations are considered when a country casts its vote in the Eurovision, we would expect it to vote for countries to which it exports to the most to maintain the good relationship. Therefore, we can induce a ranking for each voting country based on the velocity of their trade with the other countries. For every Eurovision contest we compare the statistics from the previous year to the votes of that year. For example, if Italy exported the most to Germany in 2017 compared to how much they exported to other countries competing, we would expect Italy to give 12 points to Germany in the contest of 2018. In the same way, if most of the outgoing tourism from Italy in 2017 was to Germany with respect to the other countries in the finals, we would expect Italy to give 12 points to Germany in the contest of 2018.

Inspired by the analysis in (Kumpulainen et al. 2020), we measure the correlation between the voting matrix produced be these rankings and the true voting matrix from the Eurovision finals between the years 2015-2018 for the jury votes and between 2016-2018 for the televoting. The televoting results start from 2016 because this was the year televoting was introduced. This analysis is performed for the trading data and the tourism data separately. In addition, separating between the Televoting and the Jury voting lets us differ between the different sources of correlation.

Predicting Points In addition to measuring the correlation between the rankings, we wondered if the trade and tourism information can be used to predict the ECS votes for each year. To do so, for each year and each relation type we create a graph representing the relation type. For example, the graph for the trade data from 2016 contains all eurovision participating countries in 2016 as nodes. An edge is added between node A to B with a weight representing how much A exports to B. The out edges weights were normalized by the total export for each country. On this graph, we learn

Country	Related Comments Proportion
Azerbaijan	0.638
Russia	0.731
Serbia	0.752
San Marino	0.768
Moldova	0.781
Czech Republic	0.798
Cyprus	0.805
Israel	0.814
Ukraine	0.818
Netherlands	0.818

Table 2: The countries with the lowest related comments proportion

Country	Related Comments Proportion
Portugal	0.87
Albania	0.87
Slovenia	0.872
Ireland	0.875
Croatia	0.88
United Kingdom	0.88
Denmark	0.88
Hungary	0.895
Belgium	0.897
Austria	0.904

Table 3: The countries with the highest related comments proportion

embeddings for each country using node2vec (Grover and Leskovec 2016). Then, we train a network which receives as input 2 countries, represented by their learnt embeddings, and predicts how much points the first country will assign to the second country. If the trade / tourism patterns are indicative to the voting patterns, we expect the network to be able to learn to some extent to predict the votes between the countries. We experiment with both a regression setting with a MSE loss and a classification setting (11 classes for each of the points options) and a cross entropy loss.

4 Evaluation

4.1 Performance Related Discourse

Labels Analysis The BERT model was fine-tuned with the following parameters: 100 epochs, train batch size of 16, val batch size of 20, 500 warm-up steps for the learning rate scheduler and a weight decay of 0.01. We took the weights

Figure 1: The world clouds generated of the Israeli comments labeled as song related (group 1) vs non (group 2)

Figure 2: The world clouds generated of the Israeli comments from 2021.

of the iteration with the lowest loss on the validation set which was 20% of the data. The trained model achieves an accuracy of around 90% on the validation set.

For each country, we calculated the proportion of song related comments on all the songs collected (according to the weak labels). Table 2 contains the 10 countries with the lowest proportions and table 3 presents the 10 countries with the highest proportions. It is clear to see that most of the discourse regarding the ECS, quantifying to over 80% for the vast majority of the countries, is still related to the song or the performance in the contest. It is interesting to see that in the lowest ranking countries, most of the countries aren't a part of the European Union, containing a lot of slavic countries. The proportion of the song related comments of Israel were placed 33rd out of 40 countries. These results could have been affected by the translation of the comments, achieving lower quality translation for the languages of the slavic countries.

In addition, there was no correlation between the jury/televoting/total voting and the proportion of the song related comments.

Word Clouds We picked Israel and the UK as interesting countries to evaluate the discourse on their songs as they both got controversial diplomatic attention in the past 5 years.

Figure 1 presents the world clouds generated of the Israeli comments labeled as song related (group 1) vs non

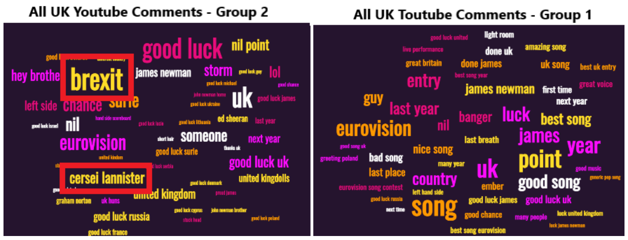

Figure 3: The world clouds generated of the UK comments labeled as song related (group 1) vs non song related (group 2).

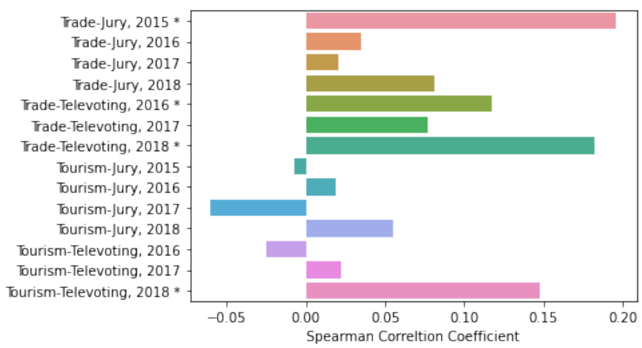

Figure 4: The Spearman correlation coefficient values between the rankings induced by the trade and tourism patterns to the true rankings in the Eurovision. Rows marked with * are statistically significant.

song related (group 2) on the videos from all ESC contests. As can be seen, the word cloud of group 1 contains only song related comments (with alot of love from Azerbaijan). When focusing on group 2, we can see that the Israeli-Palestinian conflict is a hot topic in the discussion. For example, ”Palestine”, ”Hamas shooting” and more. Next we created word clouds per year using the same method. The topic of the Israeli-Palestinian conflict has escalated over the years, dominating in 2021, 2. This is probably due to the ”Guardian of the Walls” operation that accured between May 10th-21st 2021, right before the ESC where the semi-finals where held on May 18th and 20th and the final on May 22nd.

Figure 3 presents the world clouds generated of the UK comments labeled as song related (group 1) vs non (group 2). As can be seen, the word cloud of group 1 contains only song related comments. When focusing on group 2, we can see that the Brexit is a hot topic in the discussion.

4.2 Social Considerations

Correlation Analysis Figure 4 presents the Spearman correlation coefficient values between the voting matrix induced by the trade and tourism patterns to the actual voting matrix in the Eurovision. The correlation is calculated for each year, relation type (trade or tourism) and voting type (televoting or jury). Results marked with an asterisk are statistically significant after applying a Bonforroni correction for multiple hypotheses ($p < 0.05$).

The graph presents that there is a relatively high correlation between the televoting and the trade relations in all years evaluated. We suspect this relationship is attributed to the ”Big 5” countries - France, Germany, Italy, Spain and the United Kingdom. These countries are countries with elevated trading relationships with all Eurovision countries, meaning in the voting matrix induced by the trade relations they will receive many points from all countries. These countries are also the countries that financially enable the ECS. We believe the audience votes for the ”Big 5” countries as a token of gratitude and an act of appreciation for funding the contest. To back this intuition, we present figure 5 which contains heatmaps of the voting matrices from the jury and the televoting separately for the ECS of 2018. When looking at the columns containing votes for Italy and Germany you can see that the televoters casted more votes for these countries than the Jury.

With that been said, it is important to state that correlation is not causation. Though the correlation between the audience votes and the trade patterns could be an initial sign for hidden motives behind the voting patterns of the ECS, it could be a coincidence.

The only significant correlation with the tourism patterns is with respect to the Televoting in 2018. In 2018 the ECS was held in Lisbon Portugal. Israel’s Netta Barzilai won with the song ”Toy” and Cyprus came in second with ”Fuego” and the third place was occupied by Austria. None of these countries are highly toured by Europeans. Due to the lack of explanation and the fact that in all other years both from the televoting and the jury voting there is not significant correlation, we conclude that there is no evidence to say tourism patterns are a motive when voting in the ECS.

Predicting Points For both the regression setting and the classification setting, though trying several different architectures and parameters, the networks were unsuccessful in learning the voting patterns based on the tourism embeddings or the trade embeddings. In the regression setting, the networks converged to predicting the mean of the voting options and in the classification setting the networks converged

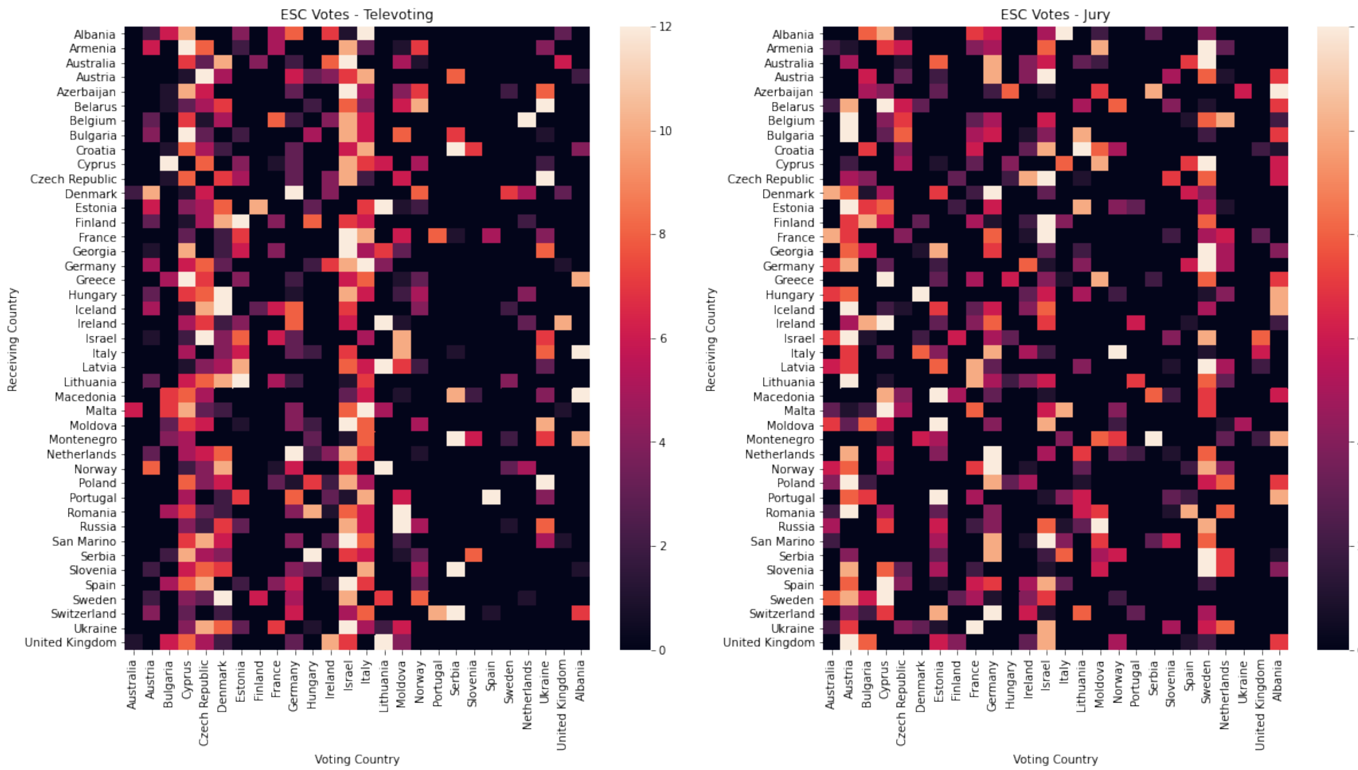

Figure 5: Heatmaps of the voting matrices in the ECS of 2018, seperated to the jury votes and the televoting. The audience voted more than the jury for countries from the "Big 5" like Italy and Germany.

to predicting the majority class (0 points). Attempts to handle the imbalance in the datasets, like weighting and undersampling, were not successful in improving the network's performance. If the networks were able to learn something based on any of the relation types, it would have been an indication that the relation type contains information regarding the voting patterns in the Eirovision. Unfortunatly, the fact that we weren't able to train successfull networks does not reject the hypotheses.

To summarize, from the results presented in the above evaluation we conclude that there is evidence supporting the hypothesis that trade relations are a factor in the eurovision voting system, but tourism patterns aren't taken into account.

5 Discussion

In this project, we analyze affects of different aspects on the voting behaviour in the Eurovision Song Contest. Our results present that though there is a difference between countries, most of the discourse around the songs in the ESC is still related to the song or the performance. In addition, when deep diving to two specific aspects - economic and tourism relations between the countries, we see correlation between the trade relations and the televoting. Other than that, no clear correlations where observed, which is aligned with most of the discourse being related to the songs.

One limitation of our work is that when analyzing the performance related discourse, we train a model based on a relatively small subset of the data. In addition, when labeling the comments, there are ambiguous comments for which it is not clear what the writer meant, making the labels ambiguous as well. Another limitation of our work is that it measures correlation to the voting patterns. Future work can examine different causes for the votes in the ECS.

References

Charron, N. 2013. Impartiality, friendship-networks and voting behavior: Evidence from voting patterns in the Eurovision Song Contest. *Social Networks* 35(3): 484–497.

Devlin, J.; Chang, M.-W.; Lee, K.; and Toutanova, K. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805* .

Fenn, D.; Suleman, O.; Efstathiou, J.; and Johnson, N. F. 2006. 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.

Grover, A.; and Leskovec, J. 2016. node2vec: Scalable feature learning for networks. In *Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining*, 855–864.

Jackson, P. 2017. The Politics of Belonging at the Eurovision Song Contest. *EuropeNow* .

Kumpulainen, I.; Praks, E.; Korhonen, T.; Ni, A.; Rissanen, V.; and Vankka, J. 2020. Predicting Eurovision Song Contest Results Using Sentiment Analysis. In *Conference on Artificial Intelligence and Natural Language*, 87–108. Springer.

Yair, G. 2019. Douze point: Eurovisions and Euro-Divisions in the Eurovision Song Contest–Review of two decades of research. *European Journal of Cultural Studies* 22(5-6): 1013–1029.