Eurovision Song Contest: Who votes Whom?

Viet-An Nguyen (vietan@cs.umd.edu), Udayan Khurana (udayankhurana@gmail.com)

Abstract
Despite being one of the longest and most successful television program in the world, Eurovision Song Contest has been criticized about the political bias in its voting results. The objective of this project is to analyze the voting data from the Eurovision contest [1] [2] by means of visualizing the voting patterns of different countries. Using NodeXL [5], an open source tool for network data visualization and ManyEyes [4] by IBM we found evidences in favor of the above criticism. More specifically, by having the right visualization we can observe that countries in the same region tends to vote highly for each other.

Eurovision Song Contest and Data
Eurovision Song Contest

Eurovision Song Contest is one of the largest song contests in Europe and one of the world’s longest-running television program [3]. It has been held annually since 1956 among active member countries of the European Broadcasting Union. Each year, a country is chosen to be the host which is usually the winner of the previous year in recent years. Each participating country will perform a song and others will vote. Since 1998, the “televoting” format has been implemented in which a country awards a set of points from 1 to 8, 10 and 12 to other songs in the competition (with 12 to the most favorable song). More details about the contest along with its history and rules can be found in [1].

Despite its long history and successes, the contest has been accused of political bias where "the voting outcomes don't simply reflect performance quality but are influenced by factors such as regional politics, expatriate populations, alliances" [3].

Data

Our dataset is a subset of the data provided by Kaggle [3] for the past competition on forecasting the votes in Eurovision Song Contest 2010. The dataset consists of historical data for Eurovision Finals from 1998 to 2009. The attributes contained in this dataset are shown in Table 1.

Table 1: Descriptions of attributes in our dataset

<table>
<thead>
<tr>
<th>Column</th>
<th>Attribute description</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Year of competition, from 1998 - 2009</td>
</tr>
<tr>
<td>B</td>
<td>Competitor country</td>
</tr>
<tr>
<td>C</td>
<td>Region of competitor country</td>
</tr>
<tr>
<td>D</td>
<td>Language of song entered by competitor country</td>
</tr>
<tr>
<td>E</td>
<td>Name of performing artist</td>
</tr>
<tr>
<td>F</td>
<td>Name of song</td>
</tr>
</tbody>
</table>
One aspect of the data which is particularly important to our analysis is the region classifications:

- **Scandinavia** - Denmark, Finland, Iceland, Norway, Sweden
- **Western Europe** - Andorra, Austria, Belgium, France, Germany, Greece, Italy, Luxembourg, Malta, Monaco, Netherlands, Portugal, San Marino, Spain, Switzerland
- **Independent** - Cyprus, Ireland, Israel, Turkey, United Kingdom
- **Former Socialist Bloc** - Albania, Armenia, Azerbaijan, Belarus, Bulgaria, Czech Republic, Estonia, Georgia, Hungary, Latvia, Lithuania, Moldova, Poland, Romania, Russia, Slovakia, Ukraine
- **Former Yugoslavia** - Bosnia and Herzegovina, Croatia, Kosovo, Macedonia, Montenegro / Serbia and Montenegro / Serbia, Slovenia

### Visualizations and Discoveries

**Countries in the same region tends to vote highly for each other**

We represent the voting data as a graph where nodes are countries and directed edges from a node to another node represents a vote from one country to another one. The weight of an edge is the aggregated point a country has given to the other country from 1998 to 2009. In fact, the votes are made for a candidate artist but we prefer to aggregate them by countries the contestants belong to due to the nature of our questions. At first we found a graph that was quite dense (as shown in Figure 1) and it was hard to make any conclusions given the labyrinth except for

- Slovakia and Georgia are barely connected to other countries. This is due to the fact that the two countries did not participate frequently. Since 1998, Georgia participated twice in 2007 and 2008; whereas Slovakia participated once in 1998.
The highest betweenness centrality scores correspond to different regions which can be considered the "regional representatives".

Figure 1: Voting graph without any edge filtering. The node sizes are proportional to their betweenness centrality. Countries are color-coded with respect to their region.

In order to concentrate on the stronger relationships, we pruned away weaker connections in the graph using the 'Filtering' feature provided by NodeXL. Figure 2 and 3 shows the graph layouts after edges having strength below 15 and 60 are pruned respectively.

Upon filtering further we saw small clusters with exceptions that can be explained by historic or cultural associations between these countries.

- Different clusters (color coded) of countries from Scandinavian, former Soviet bloc, Western European and Mediterranean can be observed in both figures.
- Figure 3 shows the strongest voting directed pairs in which we can easily notice that the "representative" countries Russian, Sweden and Bosnia & Herzegovina have been highly voted for by countries in their respective regions.
• Another interesting observation is that Turkey has been highly voted for by various countries including Germany, Belgium, the Netherlands, France and Macedonia. We conjecture that this is due to the large Turkish population living in these countries.

Figure 2: Voting graph when edges having weight (aggregated points) less than 15 are filtered. Node sizes are proportional to their betweenness centrality. Edge widths and opacity are proportional to their weights. Countries are color-coded with respect to their region.
Figure 3: Voting graph when edges having weight (aggregated points) less than 60 are filtered. Node sizes are proportional to their betweenness centrality. Edge widths and opacity are proportional to their weights. Countries are color-coded with respect to their region.

We also used NodeXL’s Clustering/Group feature to automatically identify communities in the graph. We found three communities and on plotting them on a map of Europe using ManyEyes, reveals the following interesting East-West Divide shown in Figure 4.

Figure 4: Automatically identified groups by NodeXL plotted using ManyEyes showing the East-West Divide of Europe.
Countries in the same region tends to vote similarly

Next, we would like to examine the co-voting pattern of different countries in Eurovision. We represent the co-voting graph using an undirected graph where nodes are countries as before; edges between two nodes represent how similarly the two countries vote over time. In order to capture the voting similarity between two countries, we use Spearman’s rank correlation coefficient [6].

Thus, the edge weight between two countries in the co-voting graph is defined as the average of Spearman’s rank correlation coefficient of the two countries’ votes over the years.

In order to capture the strongest correlations, we filtered out edges having small weight. Figures 6 and 7 show the co-voting graph when edges having weight less than 0.5 and 0.75 are filtered away respectively. In Figure 6, we can observe that regional groups of countries are clustered together even though there are still a lot of edges going across clusters. Moving to Figure 7, we can observe very strong correlations among countries in Former Socialist Bloc such as Belarus and Russia, Ukraine and Poland, Lithuania and Latvia.

![Figure 5: Co-voting network when edges having weight less than 0.5 are filtered. Edge widths and opacity are proportional to their weights. Countries are color-coded with respect to their region.](image-url)
Figure 6: Co-voting graph when edges having weight less than 0.75 are filtered. Edge widths and opacity are proportional to their weights. Countries are color-coded with respect to their region.

Critiques and Suggestions

We used NodeXL effectively to complete the above visualizations with a lot of ease. While it is a great tool (there were features we wish we could have used, but didn’t find time), we wish the following issues to be addressed:

- It does not provide a means to compare a network over time or a feature with time axis.
- The arrows heads of directed edges are quite cluttered up near the nodes. If these arrows heads could be moved slightly away from the nodes, clarity can possibly be increased.
- We wished to compare the reciprocal nature of relationships, e.g. which countries have voted for some other country but never received a vote back or vice-verse. apparently, there was no functionality that could help us analyze pair wise relationships in NodeXL.

Manyeyes was helpful in quickly plotting the cluster information over the world map, but its list of countries needs to be updated - Moldova was one of the countries not listed.

References


[6] Spearman’s Rank Correlation Coefficient  
http://en.wikipedia.org/wiki/Spearman%27s_rank_correlation_coefficient