WEIGHING THE HEAVIEST BAND OF ALL TIME

Kazi Minhazur Rahman
Graduate Student
Master of Information Program
University of Maryland
minhaz@gmail.com
9 Nov 2012

Data Set

Led Zeppelin has been described by the Rolling Stones magazine as “the heaviest band of all time”. Zeppelin is still popular after the breakup in 1978. In 2007, on their reunion tour, they set the world record for “Highest Demand for Tickets for One Music Concert”. Recently, there have been rumors about a new “Led Zeppelin V” album – which the fans are eagerly waiting for. In this light, we take a look at their influence on other music genre and artists with similar music aspiration.

The data was collected using last.fm API (www.last.fm). Last.fm is a music-based social network site that collects users’ music consumption data and publishes the aggregate results. The following steps were taken for data collection:

1. Top 40 similar artist names for Led Zeppelin were collected. (Selection A)
2. For each member of Selection A, top 10 similar artist names were collected. (Selection B)
3. For each member of Selection B, top 5 similar artist names were collected. (Selection C)
4. Last.fm provides a similarity index for any pair of artist (ranged 0 to 1). This index, named Match, was collected for the relationships among the above groups.
5. For each artist, listeners (number of unique listeners) and playcount (number of unique track plays) were collected.
6. Dataset was cleaned for missing and abnormal data resulting in 489 vertices and 1271 edges.

Javascript, Chrome browser and Excel were used for data collection and cleanup.

Findings

Zeppelin’s influence spanned across multiple genres

Workflow

1. Through Automate option, duplicate edges were counted and merged and also clusters were created.
2. After laying out the groups in boxes using the HKFM (Harel-Koren Fast Multiscale) layout, all groups except the first 13 were set to ‘collapsed’ as they seemed visually less significant.
3. Vertex size was autofilled based on listeners and Vertex opacity was autofilled based on Betweenness Centrality. For Betweenness Centrality, outliers were excluded since Led Zeppelin was the extreme center due to the data collection method. Edge opacity was autofilled using Match.
4. Using Dynamic Filters, Vertices were filtered for ‘degree>3’ and ‘listeners > 600,000’.
5. All edges with ‘high opacity’ were manually organized for better visibility. Lower ranking groups were set to skip by setting visibility = 0.

Details

7 clusters can be picked out from the graphic that are distinguishable in terms of broad music genre.

1. Purple (top left) : Heavy metal
2. Cyan (bottom left) : Classic rock
3. Dark green (top mid) : Hard rock
4. Red (bottom mid) : Progressive rock
5. Orange (bottom mid) : Country rock
6. Green (top right) : Alternative
7. Light green (center mid) : Blues rock

The clusters were identified based on background research of the artists as well as from personal experience. A lot of these artists started after Zeppelin’s breakup but yet the similarity level of music is significant.
Everything goes with Cream of Clapton

Workflow

1. From the last graphic, all groups’ visibility was set to “2(hide)” except the second group.
2. The vertices of the second group were selected and were arranged in a circular layout.
3. Vertices were manually arranged for better visibility.

Details

Within the Classic Rock group, Zeppelin’s similarity with other major artists is significant. However, Cream seems to not only topple that but also has better connections with other in the network as its Betweenness Centrality is quite low compared to neighbor vertices. However, it should be noted that Cream from a close-knit group with Eric Clapton, Derek and the Dominos and The Yardbirds – a sort of semi cluster. The explanation is that Eric Clapton is a common member in all of them. Similarly, Jimi Hendrix’s influence is also visible through the visible edges. On the other hand, Beatles and Rolling Stones has weak similarity with almost all the major artists but yet their high Betweenness Centrality suggest that they might be connecting this group to other major groups.
Rock is not dead

**Workflow**

1. From the last graphic, all the teams were set to visible and none were left collapsed.
2. Vertices’ $x$ and $y$ were mapped to *listeners* and *playcount* respectively.
3. *Vertex opacity* autofill was removed and set to 0.7 for better visibility. *Edge opacity* was set to 1 for all.
4. *Vertex size* was set to *Betweenness Centrality*.
5. Low values in *playcount* and *listeners* were filtered out.

**Details**

Several artists from the 7 groups have more than 100 million *playcount* where as a number of the artists have 1 million+ *listeners*. It can be assumed Beatles is an outlier(); Metallica and Pink Floyd are known for their fan base as well as fan loyalty. As Zeppelin’s numbers trail them, it can be said easily that rock is not dead. What’s interesting here is that Nirvana and Foo Fighters – both alternative rock back, have a high listener count. It is also interesting that Dave Grohl is a member of both. It might indicate a tendency to become more mainstream.
Critique of NodeXL

Good sides

+ The ability to use the tool through Excel is a highly innovative and useful idea. This will aid a lot of users without specialized knowledge to use this.
+ Even within Excel, the spreadsheet layout, planning and automation is well thought out and amply geared to make the user feel ‘at ease’. Using regular Excel features to augment the tool has worked out really well (example: tooltip in comment).
+ Though it has been advised otherwise, using Excel functions with direct cell references to construct new columns and values worked out flawlessly.
+ Though the tool crashed several times, it never needed to be restarted and no work was lost. They errors are managed well.

Areas for improvement

– The dropdown menus on the cells worked only on the first row.
– For own columns that are added manually, the dynamic filter step’s is set to 1. Thus if the value range is low (e.g. 0 to 1), the user needs to manually multiply all the values by a certain number.
– Clicks on vertices in graph does not get recognized nearly 30-40% of times.
– Refresh filters, Refresh graph button and Lay Out Again button have a sync issue. If dynamic filters are applied and then the graph is refreshed, the filters retain their value but are not applied to the graph. One need to click on the dynamic filter button to apply the filter values.
– The buttons on NodeXL ribbon do not have user intuitive keyboard accelerators. This can be a significant issue for Excel power users.
– XY graph image export did not export the axes.

Suggestions

☒ A feature is absolutely necessary to save the visual layout. There should also be multilevel ‘undo’ capability.
☒ The color management system needs to be upgraded significantly. The 64 color palette is not sufficient neither is it flexible enough. Integrating Excel 2007/higher color management would be a good option. It is also possible to show preview of the color as the cell background.
☒ Ability to use custom svg files as shapes would open a lot of possibilities for unique visualizations.
☒ The Save Image button should be a big one on the ribbon! Additionally it can be “Save and share” which will enable the image to be uploaded and shared on social networks.