From Lil Wayne to Mozart, Network dynamics in Spotify

Tobin South

The University of Adelaide,
ARC Centre of Excellence for Mathematical & Statistical Frontiers (ACEMS)
Supervisors: Lewis Mitchell & Matthew Roughan

November 28, 2019

What is Spotify and why should I care?

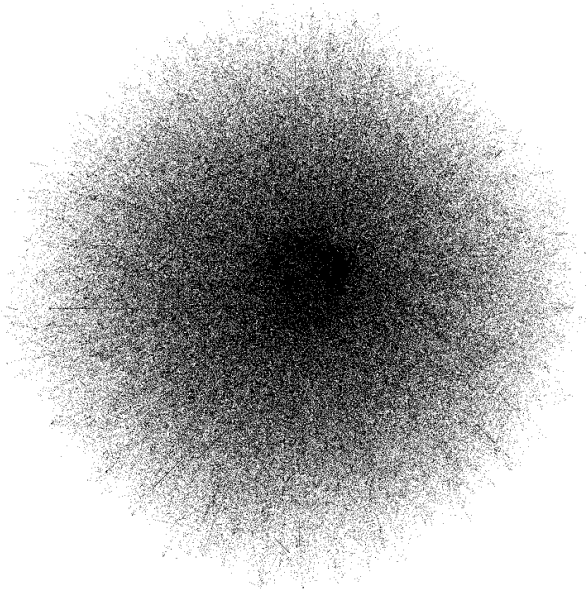
The Spotify Artist Collaboration Graph

- Nodes
  - Unique Artists on the Spotify Platform.
- Edges
  - 'Collaborations' between artists.
  - Any two artists that appear on the same song or album.
  - This is specified by artists during their upload.
  - This can include sampling or cover songs.
  - Unweighted, undirected.

Collection Method

- Snowball Sampling.
- Started from Kayne West (chosen for high collaboration rate and popularity).
- 1,250,065 artists collected.
- 3,766,631 edges.
- Limited by API limits.
- Collected metadata about 918,504 artists (random sample).

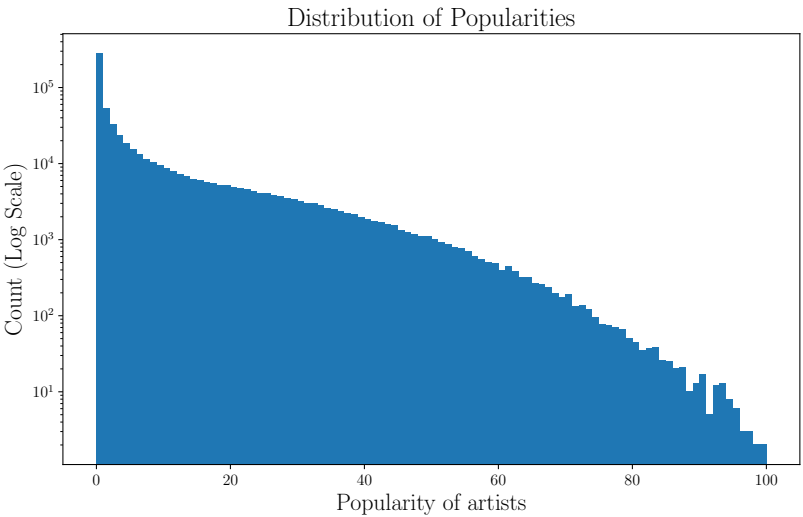
The graph



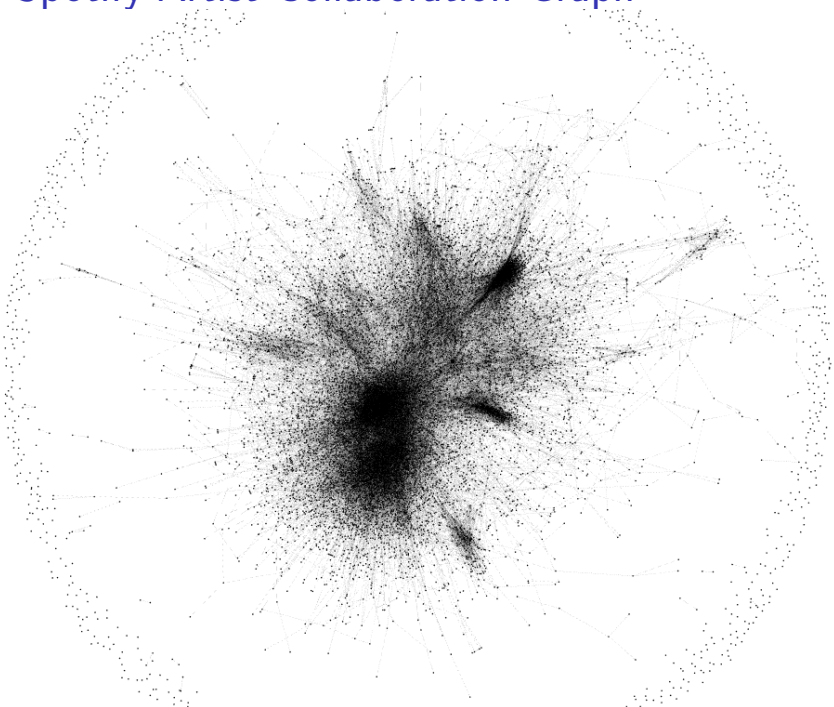
What do we know about our data?

- Who these artists have collaborated with.
- Popularity.
- Number of Followers.
- What genres are these artists described by.
- Some other stuff we don't care about.

Popularity


Distribution of Popularities
Count (Log Scale)
10^5
10^4
10^3
10^2
10^1
0
20
40
60
80
100
Popularity of artists

The Spotify Artist Collaboration Graph



What do we want to know

- We can do better than popularity.
- We seek a measure of importance within the graph.
- There are lots of metrics of centrality in a graph.

It pays to have friends: Eigenvector Centrality

- Gives a measure of the influence of a node in a graph.
- Connections to high-scoring nodes contribute more to the score of the node than connections to low-scoring nodes.
- Solving it is seeking a solution to the Eigenvector problem $Ax = \lambda x$ where A is the adjacency matrix.
- Usually seen as robust [**?**, **?**].

Who's cool?

Figure: Network filtered by nodes with highest eigenvector centrality. Colour is the relative centrality, with red being the most central. Size is proportional to popularity.

But: Sampling by high popularity

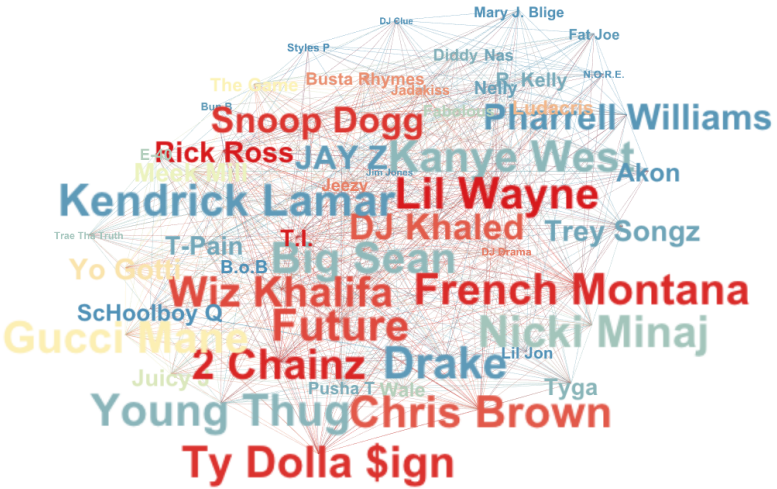


Figure: Network filtered by popularity of nodes. Nodes with highest new eigen centrality after popularity filtering are shown.

The change

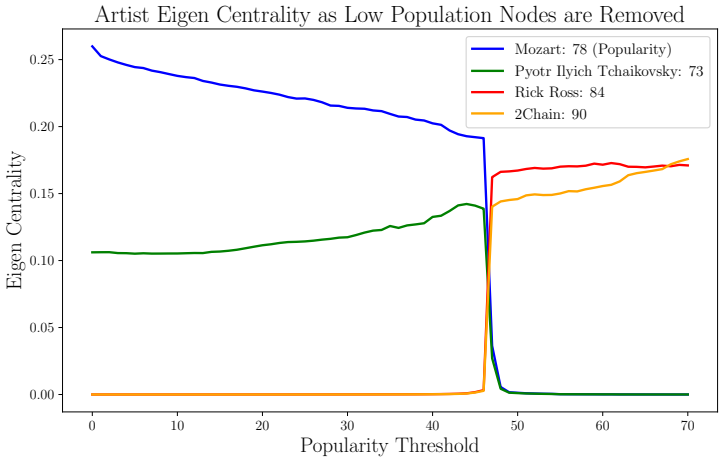
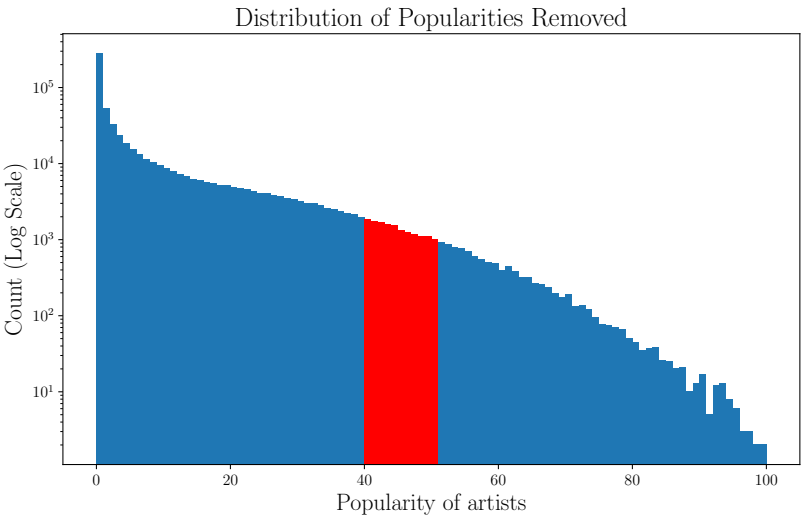


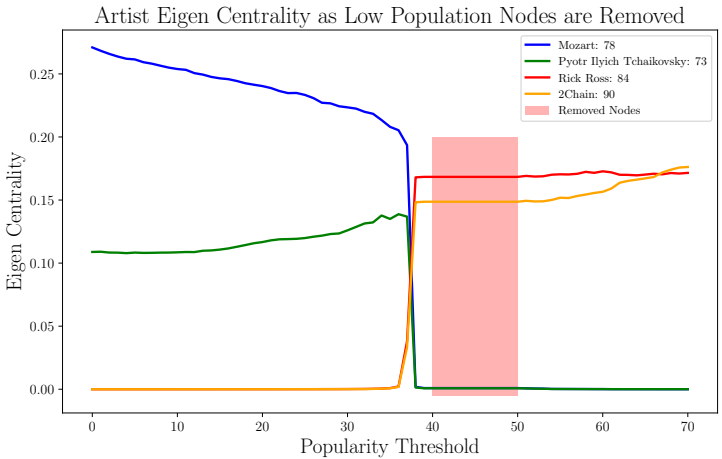
Figure: 70 subgraphs are constructed using only nodes with popularity greater than the enumerated threshold. Eigen centrality is calculated on each subgraph. Centrality coefficients are given for a selection of artists in each of the subgraphs.

Removing the critical nodes



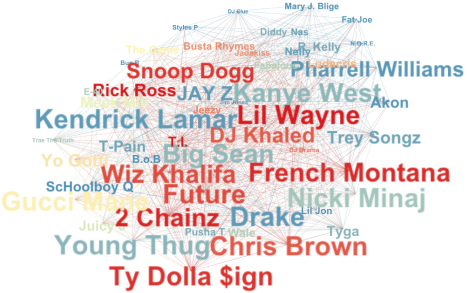
Similar behaviour with critical nodes removed

Nodes with, $40 \leq$ popularity ≤ 50, were removed from the network entirely, and analysis was rerun.




TO BE
CONTINUED...

What we do have



Conclusion

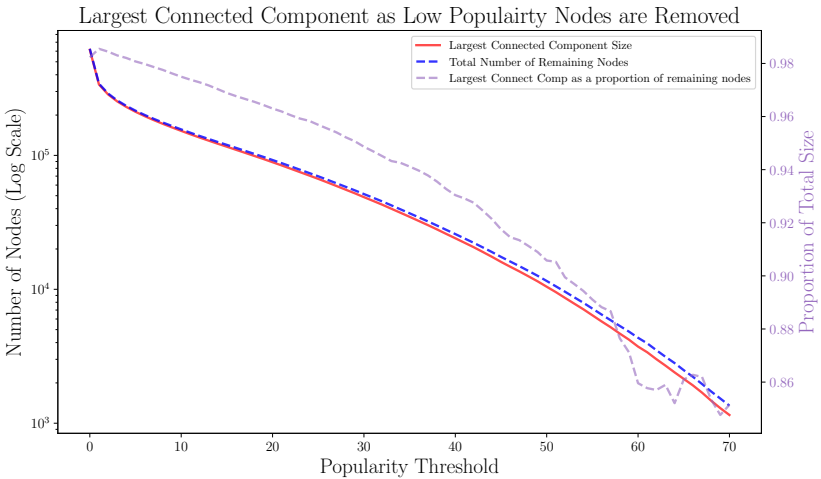
- *Viva la Classics*.
- Modern streaming services provides fantastic data sources of complex systems.
- Biased examinations of complex system can lead to false conclusions.

Thanks for listening!

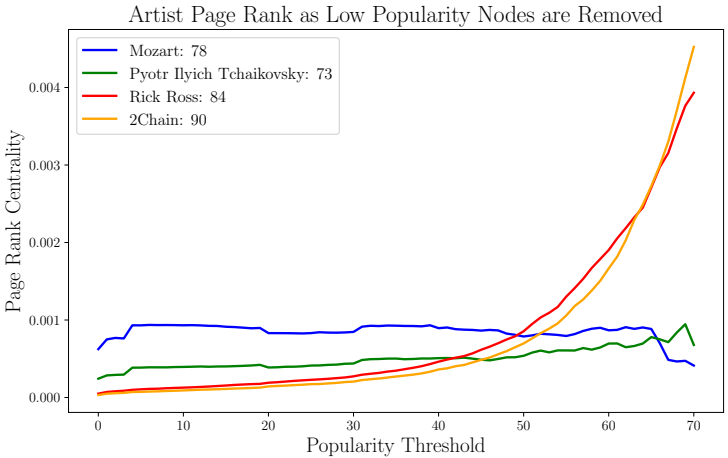
Thanks to:

and my fantastic supervisors,
Matt Roughan and Lewis Mitchell

Checking for network fissures



An alternative centrality: Page Rank



References

S. P. Borgatti, K. M. Carley, and D. Krackhardt.

On the robustness of centrality measures under conditions of imperfect data.

Social networks, 28(2):124–136, 2006.

E. Costenbader and T. W. Valente.

The stability of centrality measures when networks are sampled.

Social networks, 25(4):283–307, 2003.