NETWORK SCIENCE
PROJECT

THE NETWORK OF SINGERS IN SPOTIFY

Pignatelli Nicolò
13831A

A.A. 2023-2024

THE AIM

How are singers connected according to users' listenings?

Every node is an artist

A pair of nodes is connected if the two singers have both been listened to on Spotify by at least 2 users

More than 1 because maybe it happened by chance

THE AIM

What can we discover about the set of artists we listen to?

- Do people divide between who listens to popular singers and who does not?
- Is it possible to find patterns of artists listened to together?

THE DATA SET

Publicly available

- > 12mln rows so random sampling (125k)
- 4 columns, the two of interest: “user_id” and “artistname”
- Removed every duplicate “user_id”-”artistname”

We do not take into account how many times a user listens to a singer

TOP 5 ARTISTS

1. Coldplay: 298 users
2. Daft Punk: 288 users
3. JAY Z: 268 users
4. Kanye West: 252 users
5. Radiohead: 237 users

Will we find the same artists during our analysis?

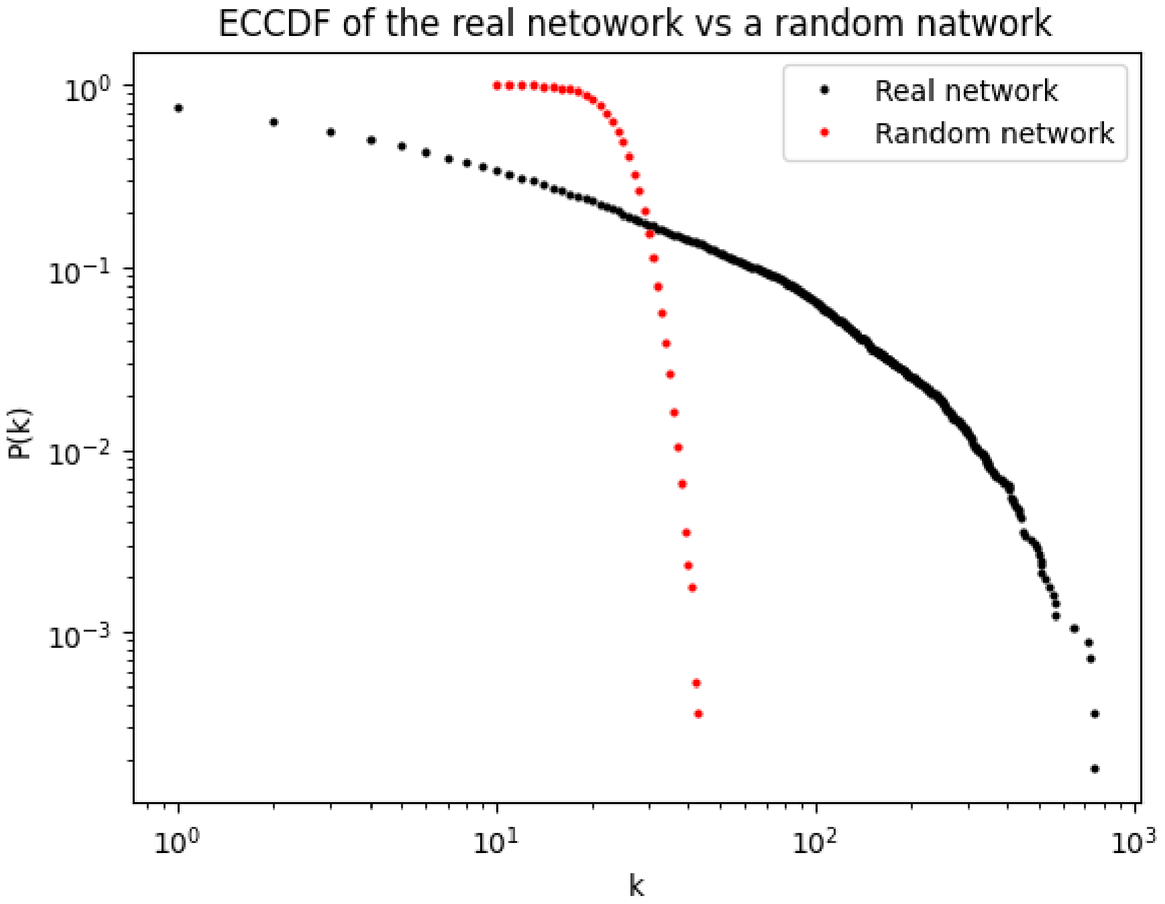
THE GRAPH

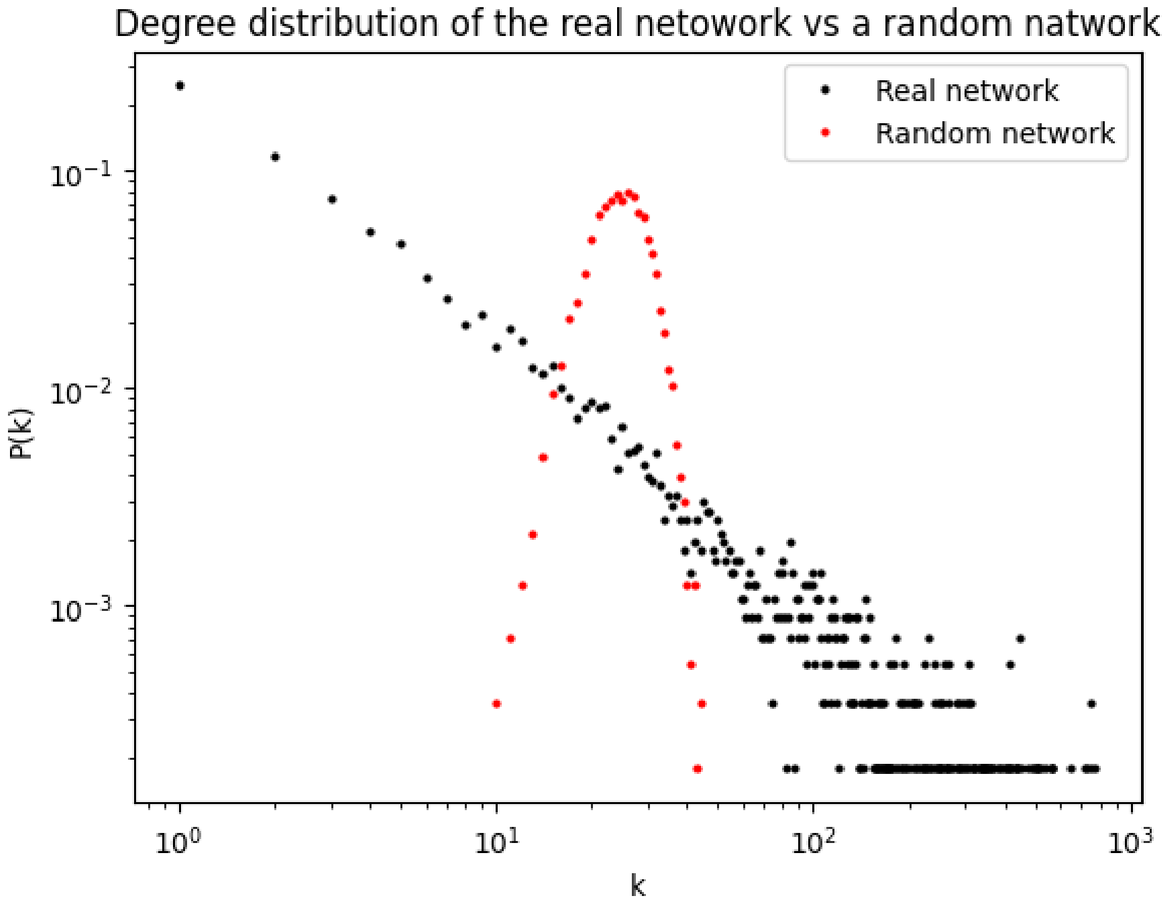
- Undirected and weighted
- 5586 nodes
- 70430 edges
- Weights: how many users have listened to the pair of artists. By construction, the minimum is 2.

NODE DEGREES AND DENSITY

- Average degree: 25 (number of other listened artists)
- Minimum: 1
- Maximum: 771 (beyond genres)
- Standard deviation: 62.37
- Median: 5
- Density: 0.0045, sparse as expected for a real network

DEGREE DISTRIBUTION





CENTRALITY

- Degree: David Bowie
- Betweenness: David Bowie
- Closeness: Coldplay
- Eigenvector: Radiohead

CENTRALITY

David Bowie is the artist that is the most listened to with other singers

Maybe others have more listeners, but they listen to fewer other artists with respect to David Bowie’s users

He is also the one who is easier to come across when listening to new singers

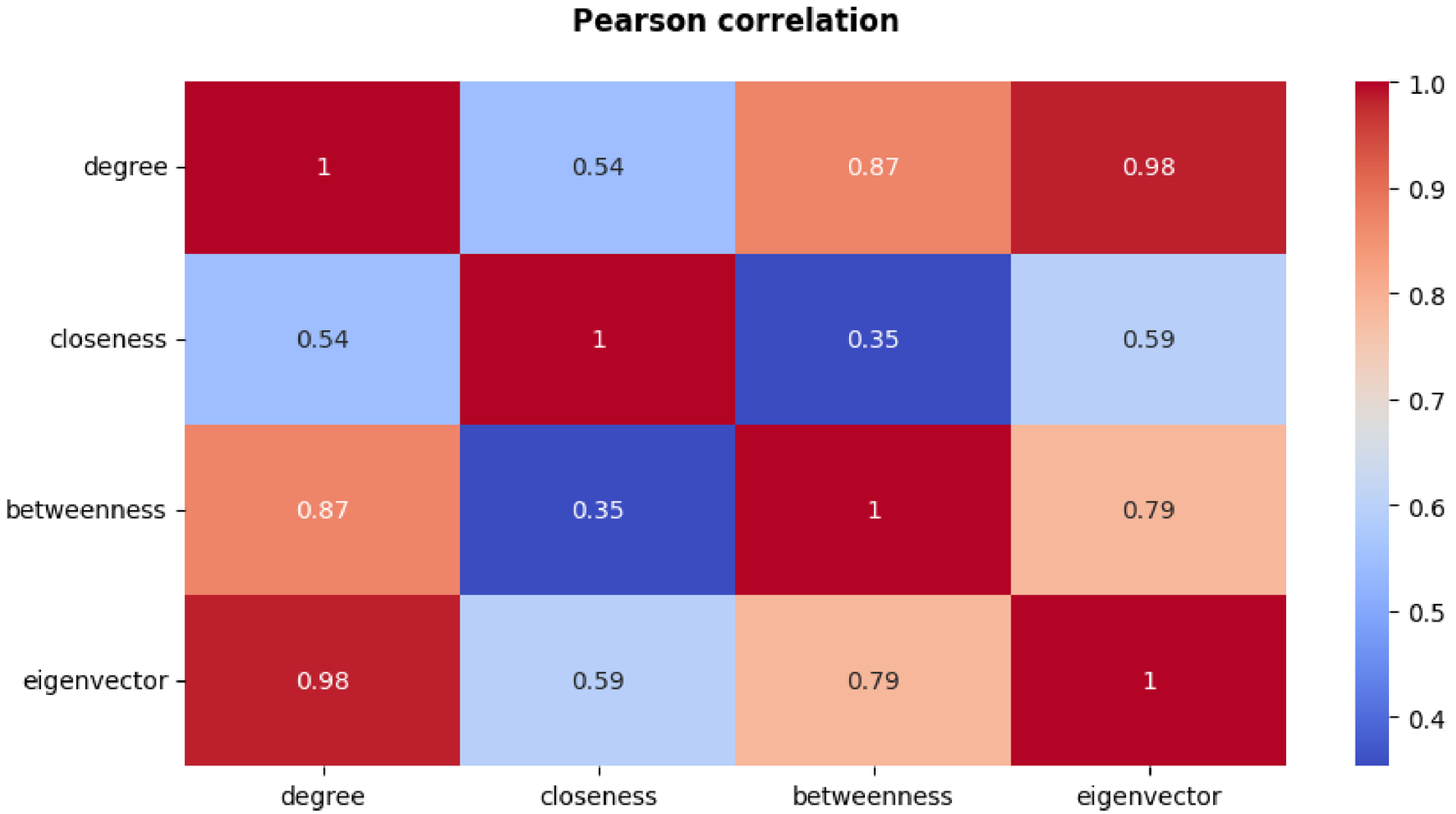
CENTRALITY

Radiohead are the most listened artists with other popular singers

Coldplay are the closest to other artists, so they really seem beyond genres

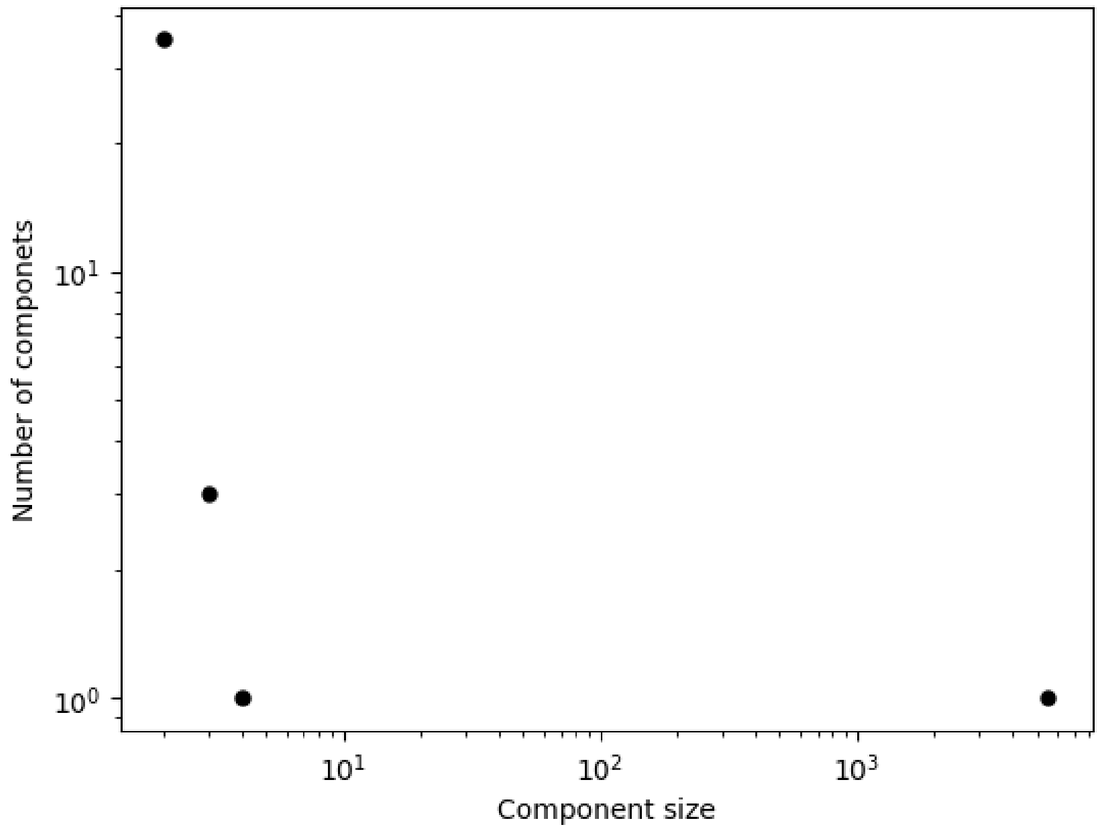
Top 5 of centralities are similar, so they detect the same hubs

CENTRALITY


Pearson correlation
degree
closeness
betweenness
eigenvector
1
0.54
0.87
0.98
0.54
1
0.35
0.59
0.87
0.35
1
0.79
0.98
0.59
0.79
1
degree
closeness
betweenness
eigenvector
1.0
0.9
0.8
0.7
0.6
0.5
0.4

CONNECTIVITY

- Graph is disconnected
- 40 connected components
- 39 with < 5 nodes
- Largest: 98.5% of all nodes
- 1.95% of all edges are bridges



CONNECTIVITY

Considering the largest connected component:

- 5503 nodes and 70386 edges
- Average degree and density slightly higher
- average shortest path length: 3.1
- diameter: 8

TRANSITIVITY

- Global Clustering Coefficient: 0.19
- Average Local Clustering Coefficient: 0.36

Why this difference?
Local clusters of artists, maybe related to genres, not so conncted to each other

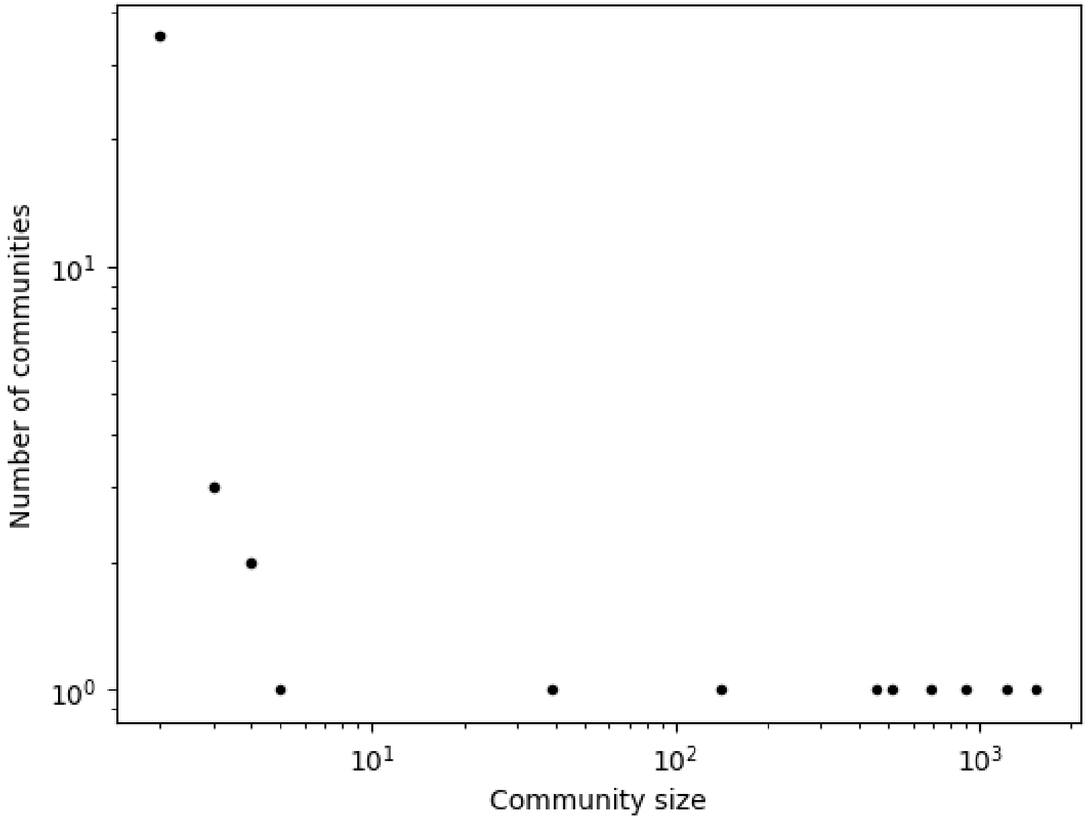
ASSORTATIVITY

Degree assortativity: -0.13

The graph is dissortative, so high degree nodes tend to be connected to low degree ones
Again, it seems that some artists are present in many users' listenings with many ones present only occasionally and only with that specific artist.

COMMUNITIES

- 49 with Louvain Algorithm
- Average size: 114 nodes
- Maximum: 1527 nodes
- Top 5 central are in 3 different ones



FINDINGS

- The network seems to have the expected properties
- Most listened singers are not always also the most connected ones (listened to with many other artists); still overlapping
- Artists tend to be listened to by users together with often the same singers (in-genre conformism)
- Some popular artists are listened to together with many not so popular singers, which on the other hand are listened to together with that specific singer (hierarchy)

RECOMMENDATION SYSTEM

- Graph Machine Learning
- Thanks to Node2Vec, we are able to create a recommendation system
- Given an artist, the top 5 most similar artists are retrieved

RECOMMENDATION SYSTEM

- Both DFS and BFS
- Similar results in terms of similarities coefficients and artists retrieved
- It seems to follow degree assortativity: for popular artists, it retrieves not popular and viceversa
- Completely different from Spotify’s “Fans also like”

Thanks for your attention!