

The network of music collaborations: Exploring how collaborations influence music genres and artist popularity

Ty Irving¹, Wasif Ud Dowlah², and Haoyang Shi³

¹CPSC 572, 30105319

²CPSC 572, 30130706

³CPSC 572, 30105296

April 4th, 2024

1. Project Summary

Historically and currently collaborations between artists play a pivotal role in shaping how music is created and shared to individuals. Our project looks into the network of Spotify collaborations and aims to try and discern the impact of collaborations on artists' popularity. The network that is intended to display the impact of collaborations is an undirected network between artists which intends on displaying the collaboration of artists when they collaborate on songs together this demonstrates the ability of collaboration on artist performance when studied. The data in which this network is derived is gathered from the [Spotify Web API](#) which contains the artists songs and collaborations which will be gone through to create edges for the undirected network. The resulting network demonstrated that the artists mainly collaborated with a few main artists in the community and generally the popularity of the artists who collaborated with each other were similar. Due to the results that were gathered it was obvious that the impact of collaboration tends to be related with individuals who had collaborated a ton generally had a high following or popularity though the amount of data that had been used could be improved. Within the future the ability to gather and analyze more data can greatly increase the understanding of how collaboration impacts individuals though with the small sample size it is evident that there is a correlation between collaborations and popularity.

2. Research Questions

We are looking to address the following research questions:

1. Which artists are the most influential in terms of collaboration?
2. Does increased collaboration correspond to an increase in popularity?

To be able to properly be able to address the questions above we plan on checking statistics, for the case of which artists are most influential we will be looking at a subset of North American artists which will improve our ability to know how these artists are influential and gather valuable data. Comparing the collaboration to popularity will use the node's popularity which each artist has and it will be compared against the amount of collaboration that the artist has which will demonstrate whether increased collaboration corresponds to an increase in popularity.

3. Introduction

Investigating the inner networks of collaboration within the music industry has long been a topic for many researchers. The importance of these types of studies are in the impact that collaboration has and whether it is important to continue to collaborate or not. Scholars have increasingly delved into this topic of whether collaboration impacts the popularity of artists and songs for a long time and this is important because it helps artists understand the importance of collaboration.

Studies such as the study done by Abhishek Deshmane and Victor Matinez de Albeniz (2023) examined the collaboration within songs in the music industry and concluded that on average collaboration garners more than twice the number of plays per week. Our study aims to follow the artists as a whole viewing their entire career of collaborations rather than a song. By broadening our scope of viewing artists as a whole rather than looking at songs individually we plan on deepening our understanding of collaborations' influence on artist success metrics on Spotify and being able to view how collaboration dynamics shape the music industry.

Other similar insights such as research done by Wichaya Peechapat and Natapon Puttanapong (2024) studies the social network analysis of Thailand's music industry and this provides valuable insight on interactions between artists and the industry. This draws very similar data as what we aim to study which provides valuable insight on the complex relationships between artists and the collaborations and how it impacts an artist and their popularity. Not only does the paper provide valuable insight on artists relationships it also helps us navigate the complex struc-

ture of networks that we are dealing with.

Upon looking at the literature that analyzes the complex relationship between collaborations and success metrics, our project aims to investigate the impact of Spotify collaborations on the artist popularity. By being able to draw upon these complex network analysis techniques show within existing literature we aim to be able to show the impact on artists careers based on these results that we gather from the network. We strive to find similar results to other literature within the music industry commenting on the collaboration of artists and the metrics involved.

4. Dataset Description

The data that we have used is from the [Spotify Web API](#) and the way that we were properly able to gather our data was through three calls to this API using a python script to go through numerous artists.

1. Get Artist

```
1  {
2    "external_urls": {
3      "spotify": "string"
4    },
5    "followers": {
6      "href": "string",
7      "total": 0
8    },
9    "genres": [
10     "Prog rock",
11     "Grunge"
12   ],
13   "href": "string",
14   "id": "string",
15   "images": [
16     {
17       "url": "https://i.scdn.co/image/ab67616d00001e02ff9ca10b55ce82
18       ae553c8228",
19       "height": 300,
20       "width": 300
21     },
22   ],
```

```

22     "name": "string",
23     "popularity": 0,
24     "type": "artist",
25     "uri": "string"
26 }
27

```

Returns the following however the only items that we stored were the following: name, followers, popularity, and genres.

2. Get Artists Albums

```

1  {
2    "href": "https://api.spotify.com/v1/me/shows?offset=0&limit=20",
3    "limit": 20,
4    "next": "https://api.spotify.com/v1/me/shows?offset=1&limit=1",
5    "offset": 0,
6    "previous": "https://api.spotify.com/v1/me/shows?offset=1&limit=
1",
7    "total": 4,
8    "items": [
9      {
10         "album_type": "compilation",
11         "total_tracks": 9,
12         "available_markets": [
13           "CA",
14           "BR",
15           "IT"
16         ],
17         "external_urls": {
18           "spotify": "string"
19         },
20         "href": "string",
21         "id": "2up30PMp9Tb4dAKM2erWXQ",
22         "images": [
23           {
24             "url": "https://i.scdn.co/image/ab67616d00001e02ff9ca10b
55ce82ae553c8228",
25             "height": 300,
26             "width": 300
27           }

```

```

28     ],
29     "name": "string",
30     "release_date": "1981-12",
31     "release_date_precision": "year",
32     "restrictions": {
33         "reason": "market"
34     },
35     "type": "album",
36     "uri": "spotify:album:2up30PMp9Tb4dAKM2erWXQ",
37     "artists": [
38         {
39             "external_urls": {
40                 "spotify": "string"
41             },
42             "href": "string",
43             "id": "string",
44             "name": "string",
45             "type": "artist",
46             "uri": "string"
47         }
48     ],
49     "album_group": "compilation"
50 }
51 ]
52 }
53

```

Returns the artists albums this was used to get all of the artists albums the only things that we ended up storing was the artists album ids so we could write a script to go through each album to see the collaborations

3. Get Album

```

1     {
2         "album_type": "compilation",
3         "total_tracks": 9,
4         "available_markets": [
5             "CA",
6             "BR",
7             "IT"
8         ],

```

```

9      "external_urls": {
10          "spotify": "string"
11      },
12      "href": "string",
13      "id": "2up30PMp9Tb4dAKM2erWXQ",
14      "images": [
15          {
16              "url": "https://i.scdn.co/image/ab67616d00001e02ff9ca10b55
ce82ae553c8228",
17              "height": 300,
18              "width": 300
19          }
20      ],
21      "name": "string",
22      "release_date": "1981-12",
23      "release_date_precision": "year",
24      "restrictions": {
25          "reason": "market"
26      },
27      "type": "album",
28      "uri": "spotify:album:2up30PMp9Tb4dAKM2erWXQ",
29      "artists": [
30          {
31              "external_urls": {
32                  "spotify": "string"
33              },
34              "href": "string",
35              "id": "string",
36              "name": "string",
37              "type": "artist",
38              "uri": "string"
39          }
40      ],
41      "tracks": {
42          "href": "https://api.spotify.com/v1/me/shows?offset=0&limit=
20",
43          "limit": 20,
44          "next": "https://api.spotify.com/v1/me/shows?offset=1&limit=
1",
45          "offset": 0,

```

```

46     "previous": "https://api.spotify.com/v1/me/shows?offset=1&
limit=1",
47     "total": 4,
48     "items": [
49         {
50             "artists": [
51                 {
52                     "external_urls": {
53                         "spotify": "string"
54                     },
55                     "href": "string",
56                     "id": "string",
57                     "name": "string",
58                     "type": "artist",
59                     "uri": "string"
60                 }
61             ],
62             "available_markets": [
63                 "string"
64             ],
65             "disc_number": 0,
66             "duration_ms": 0,
67             "explicit": false,
68             "external_urls": {
69                 "spotify": "string"
70             },
71             "href": "string",
72             "id": "string",
73             "is_playable": false,
74             "linked_from": {
75                 "external_urls": {
76                     "spotify": "string"
77                 },
78                 "href": "string",
79                 "id": "string",
80                 "type": "string",
81                 "uri": "string"
82             },
83             "restrictions": {
84                 "reason": "string"

```

```

85         },
86         "name": "string",
87         "preview_url": "string",
88         "track_number": 0,
89         "type": "string",
90         "uri": "string",
91         "is_local": false
92     }
93 ]
94 },
95 "copyrights": [
96     {
97         "text": "string",
98         "type": "string"
99     }
100 ],
101 "external_ids": {
102     "isrc": "string",
103     "ean": "string",
104     "upc": "string"
105 },
106 "genres": [
107     "Egg punk",
108     "Noise rock"
109 ],
110 "label": "string",
111 "popularity": 0
112 }
113

```

Using this we were able to create the edges with our artists IDs mapping each artist to the collaborations that they had. Before using the nodes and edges list made from the spotify API the data cleaning was an important part of the process. Since the original data has more than 20,000 nodes, there are nearly 40,000 more edges, and many nodes and edges do not correspond. Some points (that is, artists) mentioned in the side files do not appear in the point files. Some points also don't appear in the side files. Secondly, there is a lot of missing information in the click file, such as the actual popularity, the number of people being tracked, and a lot of missing data. To supplement this information, you need to retrieve it from different websites, which will lead to data inconsistency, so the best way is to delete it entirely. Because we have a lot of data we are not worried that deletion will affect research. Secondly,

since the database contains musicians from different countries around the world, the spelling of their names affects the data display. In order to make the research more targeted, we decided to screen artists whose names are in normal English, so characters from any other country will not be included. show. Finally, in order to ensure the rationality of the research, we again use symmetrical edge files and point files to ensure that the points mentioned in the edge files are all existing points.

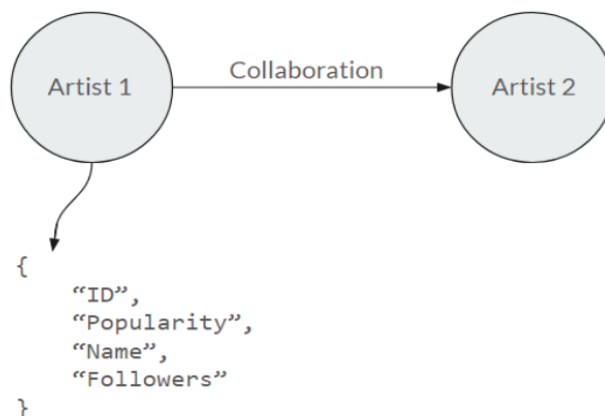


Figure 1: Node & Edge Visual

All nodes within this network follow the setup of the diagram above with each artist node containing the following: id, popularity, name, and followers this is important for our undirected network because of the ability to have ids that link each node to another this is how our edges are created they are 1:1 edges from id to id if the individual has any song/album that they collaborated with another artist.

5. Basic Statistics

Below includes the following statistics for our network:

Measures	
Number of nodes	3040
Number of edges	3426
Number of connected components	3040
Clustering coefficient	0.204
Average path length	4.782
Graph Density	0.001

Table 1: Basic statistics

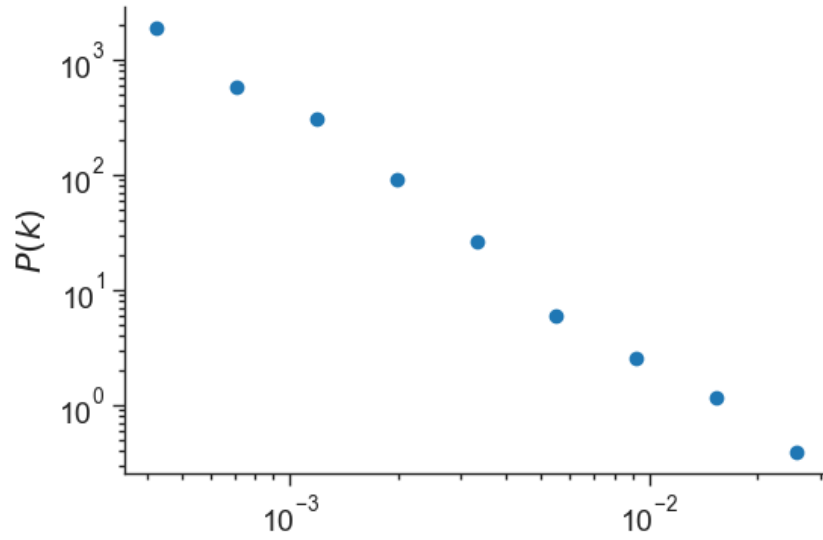


Figure 2: Degree Distribution

6. Network Visualization

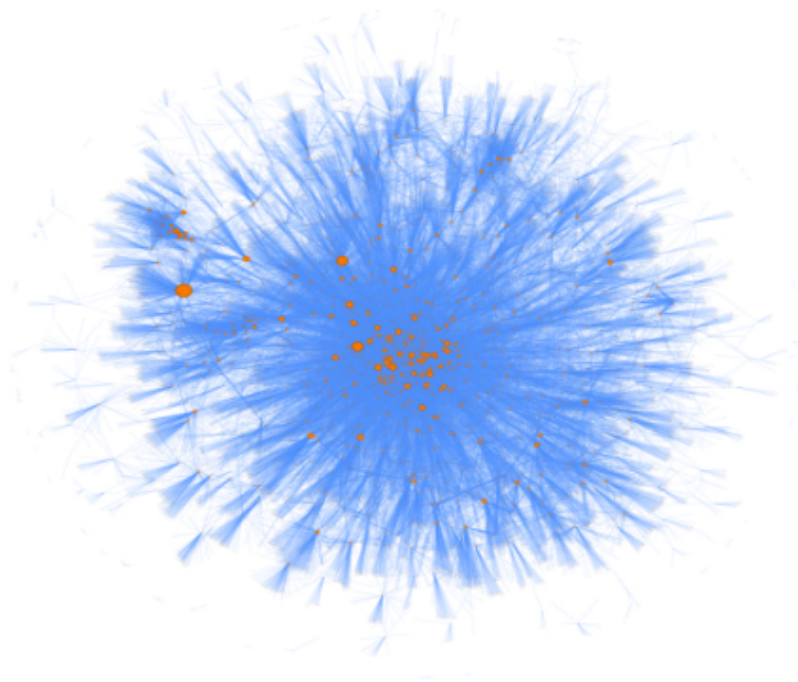


Figure 3: Entire Network - 1



Figure 4: Entire Network - 2

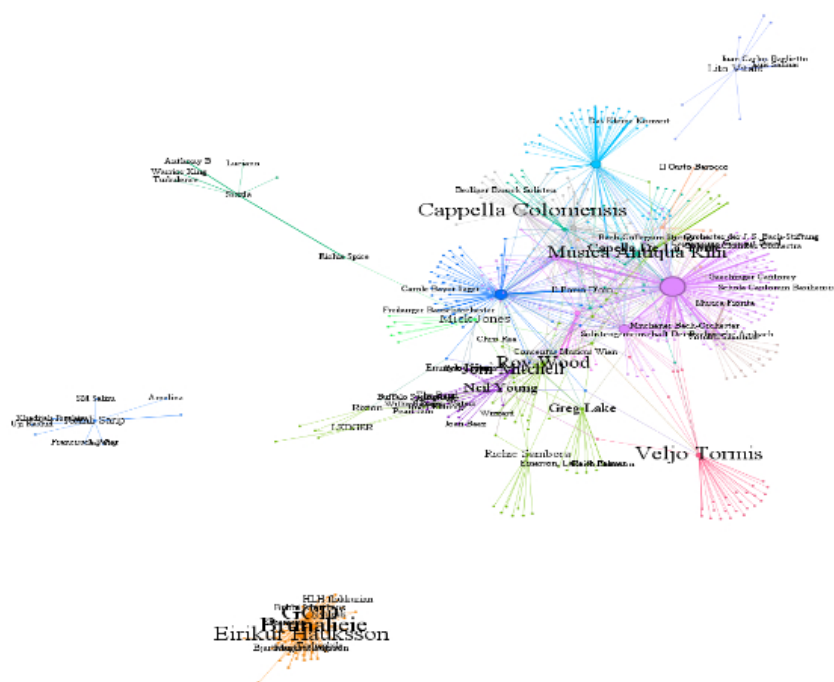
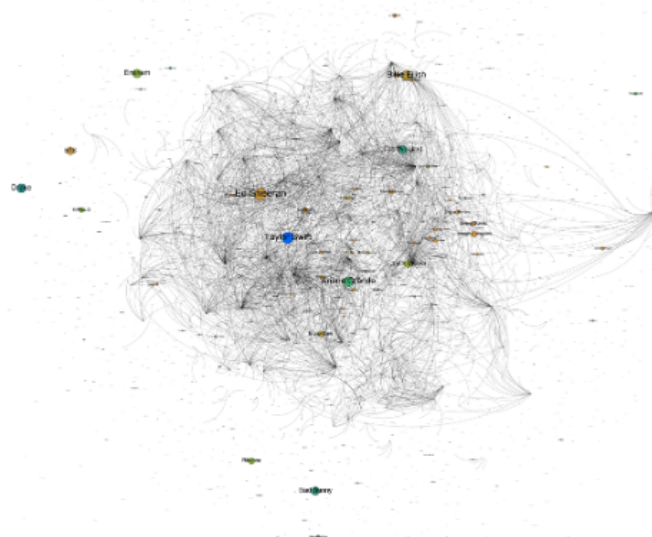


Figure 5: Subset of 500 Artists



7. Results

1. Who are the top artists when it comes to working together with others

In Figure 6, the size of a node indicates the number of connections, and the node label's size is proportional to the node size. The larger the node, the clearer the label becomes. The node's color, changing from light orange to orange to red, illustrates the node's popularity change. The darker the color, the higher the popularity.

The predominance of orange hues and larger nodes suggests that individuals of medium popularity are more inclined to collaborate with each other. Thus, artists like Rage Against The Machine, Mazzy Star, and Boston are more likely to collaborate with other artists as they have the largest nodes in the graph. These artists tend to have a mid-level popularity, indicated by the node's color being somewhere between orange.

In Figure 7, the node size represents the number of fans, and the node label's size is proportional to the node size. The larger the node, the larger and clearer the label. The color gradient, from light purple through light blue, orange, green, to bright blue, signifies the change in the number of fans per node. As the color transitions, the larger the node, the greater the number of fans.

From this graph, it is evident that well-known artists like Taylor Swift and Ed Sheeran are less likely to collaborate with others. It logically follows that major artists might avoid collaborating with those who have fewer followers.

The Python script used to generate the top 10 in terms of degree centrality supports our assumption. With degree centrality and betweenness centrality values listed for artists like Mazzy Star, Rage Against The Machine, and Rixton, a higher degree centrality is indicative of more collaborations. Artists such as Rixton, Boston, and The Ronettes, with higher degrees of centrality, appear to be more connected than others on the list. Furthermore, a higher betweenness centrality suggests a stronger influence on the network's flow. For instance, Rixton, having a notably high betweenness centrality, likely plays a crucial role in connecting disparate groups or artists within the network.

2. Do artists who work with others a lot also get more famous?

This question is challenging to address due to the limitations of our dataset in capturing changes in fame over time. To effectively analyze how artists "get more famous," we would require a more comprehensive dataset that tracks changes in

popularity metrics. However, from the data we have, which includes metrics like the number of nodes (3040), number of edges (3426), average path length (4.782), average clustering coefficient (0.204), and graph density (0.001), we can draw some conclusions about the structure of the music industry network.

The average path length being less than 5 suggests a "small-world" effect within the network, meaning any two artists are, on average, separated by just a few steps despite the network's large size. This indicates a high degree of interconnectedness in the music industry, which could facilitate the rapid spread of trends, styles, or collaborative opportunities.

However, the graph density of 0.001 is very low, pointing to a sparse network. This sparsity could be due to the dataset's quality or reflect the reality that many potential connections between artists are not realized. In such a sparse network, achieving fame through a few collaborations might be challenging. It suggests that for artists to significantly increase their fame, collaborating with as many different artists as possible might be a more effective strategy. This approach could increase their visibility and connectivity within the network, potentially leading to more opportunities for exposure and recognition.

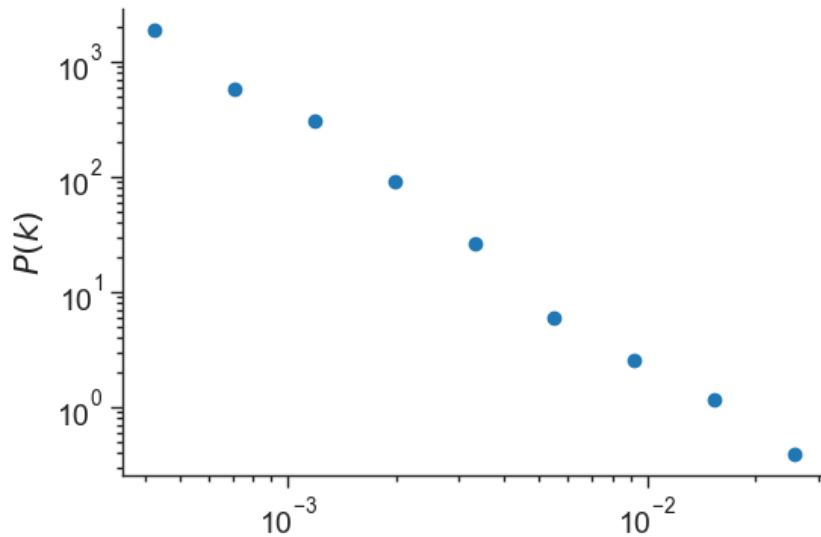


Figure 8: Degree Distribution of Figure 6 and 7

The overall downward trend suggests that nodes with a higher degree are less common, which is typical in most networks - as the degree increases, fewer nodes have that many connections. For artists, this could mean that the path to increasing their popularity and influence may involve not just producing quality music but also strategically increasing their visibility and connections within the network.

8. Comparison to Suitable Null Model

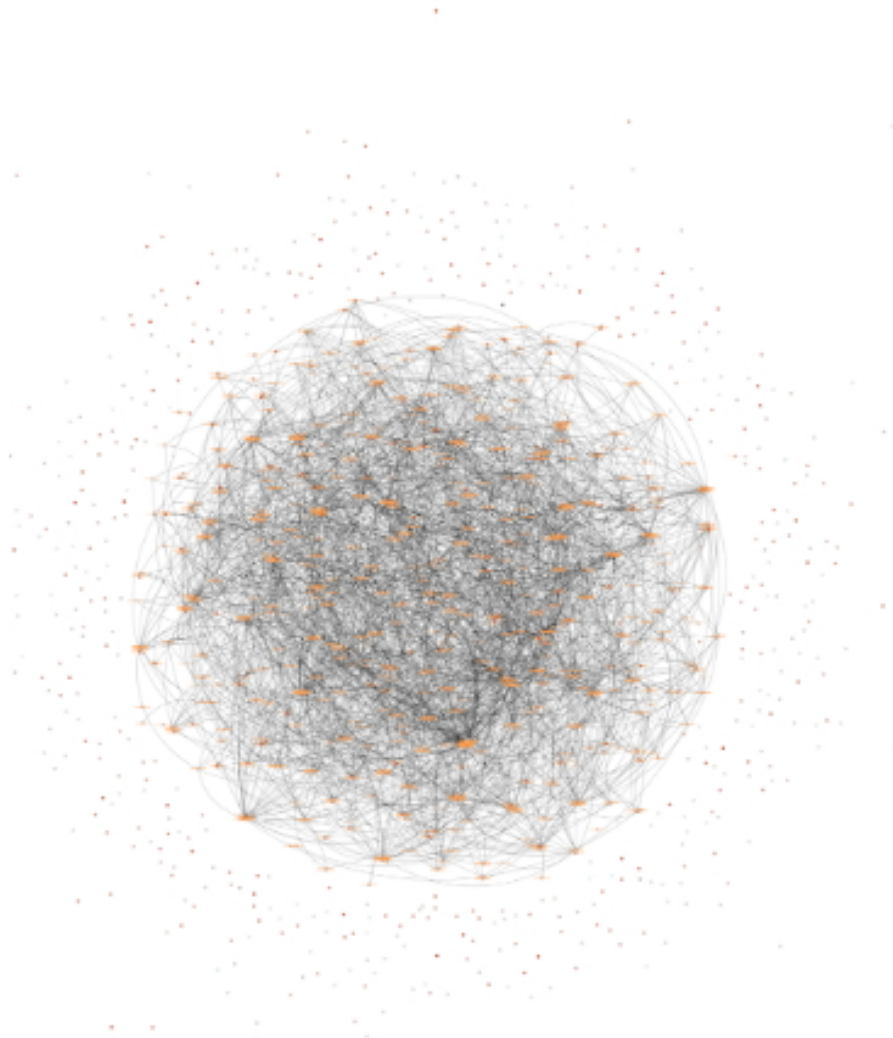


Figure 9: Null Model

Explanation of this Null Model

1. The size of the node indicates the number of connections, and the node label is proportional to the node size. The larger the node, the clearer the label;
2. The color of the node, the gradient from light orange to orange to red, illustrates the change in the popularity of the node. The darker the color, the higher the popularity. There are more oranges in the figure and the nodes are larger, indicating that people with medium popularity are more likely to cooperate with each other.
3. The node label color is the same as the node color and has no special meaning, which is convenient for displaying changes in popularity. The thickness of the edge indicates the frequency of cooperation;

Measures	
Number of nodes	3040
Number of edges	4557 (+1517)
Number of connected components	3040
Clustering coefficient	0.001 (-0.203)
Average path length	6.11 (+1.328)
Graph Density	0.001

Table 2: Null Model Statistics

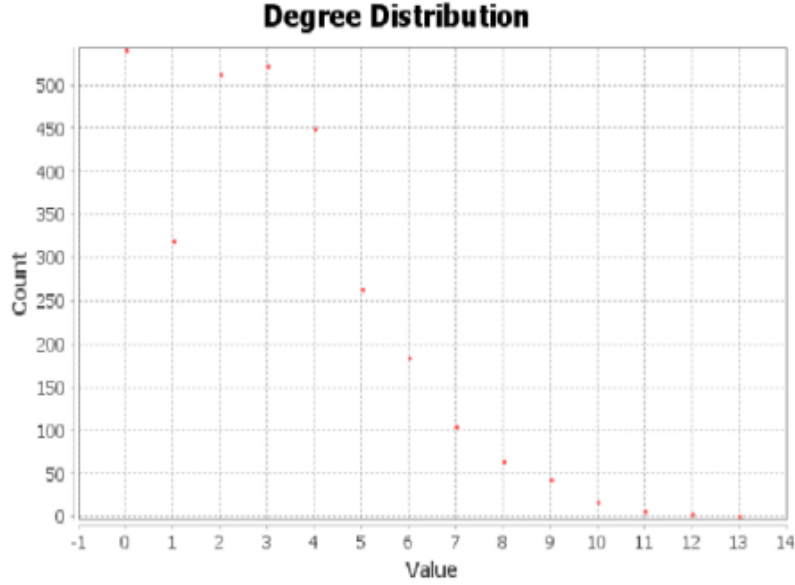


Figure 10: Degree Distribution Null Model

The regular model's higher average clustering coefficient (0.204 vs. 0.001 in the null model) and shorter average path length indicate a compact, "small-world" network. This suggests artists are more interconnected, forming tight-knit communities where collaboration or influence can spread quickly and easily, a characteristic common in social networks like those in the music industry.

9. Discussion

In our investigation into the collaborative networks of artists, we have adopted a refined dataset to gain insights into the intricate web of relationships that shape musical innovation and popularity. Comparing our findings with established literature, we see parallels with studies that identify collaboration as a key driver for artistic creativity and success (Cattani and Ferriani, 2008; Uzzi and Spiro, 2005). Our network's low-density points to a vast potential for new collaborative ventures, a characteristic underscored by Uzzi and Spiro in their exploration of Broadway

creative networks.

Upon analysis, we have discovered that there are surprisingly short paths of collaboration on average, suggesting a 'small world' phenomenon within the artist network. However, the lower-than-expected clustering coefficient leads us to speculate that while some clusters of tight-knit collaboration exist, they are not widespread.

Our research questions aimed to uncover the most influential genres and artists in terms of collaborations and whether increased collaboration is linked to an increase in popularity. The lack of data on genre-specific collaborations presents a limitation, which means the question regarding genre influence remains partially unanswered. Despite this, our network metrics have allowed us to identify highly connected artists, which could serve as a proxy for influence, albeit incomplete without the genre context. Artists like Jonathan Sebastian Bach and Mozart have the highest degree centrality, which means they have the most collaborations and therefore the most influence in terms of collaborations.

For future work, constructing a multi-layered network incorporating genres as nodes connected by edges representing collaborations would enhance our understanding of genre-specific influences. Furthermore, integrating comprehensive popularity indices, such as streaming counts or social media metrics, would refine our analysis of the relationship between collaborations and artist popularity. Cross-referencing our findings with economic models could also shed light on the commercial implications of artist collaborations. As we strive for a holistic understanding, expanding our dataset to include global collaborations is an avenue for further exploration, potentially revealing the universality or specificity of our current findings.

10. Methods

To determine the most influential artists in terms of collaborations in our network, we focused on network centrality measures. These measures can help identify the most important or influential nodes (in our case, artists) within the network. With our network size, (48,651 nodes and 87,063 edges), here are the key centrality measures we calculated using NetworkX:

Degree Centrality: This is the simplest measure of node influence. It is based on the number of connections (collaborations) an artist has. In our collaboration network, a high degree of centrality means the artist has collaborated with many

other artists.

Betweenness Centrality: This measure indicates how often a node appears on the shortest paths between other nodes. Artists with high betweenness centrality can be seen as influential in connecting different groups or clusters within the network.

Closeness Centrality: This reflects how close a node is to all other nodes in the network. It indicates how closely connected an artist is to the network as a whole, potentially influencing a broader range of artists.

Eigenvector Centrality: This measure considers not just the number of connections an artist has, but also the quality of these connections. It means that being connected to artists who themselves are highly connected is beneficial for an artist's centrality score.

An artist scoring high across multiple centrality measures can be considered particularly influential. For instance, an artist with high degree centrality but lower betweenness might be influential within a certain genre or community but not as much across different music communities. The code block below outlines the usage of Python library, Networkx, in calculating different centrality measures. Parallel processing was implemented to optimize the calculation of betweenness centrality.

```
1 def calculate_betweenness centrality(G, k=None):
2     return nx.betweenness centrality(G, k=k, normalized=True)
3
4 def main():
5     # Read data efficiently
6     nodes = pd.read_csv('nodes_Copy.csv')
7     edges = pd.read_csv('edges.csv')
8
9
10    G = nx.from_pandas_edgelist(edges, 'Source', 'Target')
11
12    #add node attributes
13    nodes = nodes.drop_duplicates(subset='id')
14    attr_dict = nodes.set_index('id').to_dict('index')
15    nx.set_node_attributes(G, attr_dict)
16
17    print(f"Number of nodes: {G.number_of_nodes()}")
18    print(f"Number of edges: {G.number_of_edges()}")
19
20    # Calculate degree centrality
21    degree centrality = nx.degree centrality(G)
```

```

22
23 # Parallel processing for betweenness centrality to reduce runtime
24 pool = multiprocessing.Pool(processes=multiprocessing.cpu_count())
25 betweenness Centrality = pool.apply(calculate_betweenness_Centrality, args=(G,
26 1000))
27 pool.close()
28 pool.join()
29
30 # other centralities
31 closeness_Centrality = nx.closeness_Centrality(G)
32 eigenvector_Centrality = nx.eigenvector_Centrality(G, max_iter=1000)
33
34 # Sort and get top 10 for each centrality measure
35 top_Centralities = {
36     'degree': sorted(degree_Centrality.items(), key=lambda x: x[1], reverse=True)
37    )[:10],
38     'betweenness': sorted(betweenness_Centrality.items(), key=lambda x: x[1],
39     reverse=True)[:10],
40     'closeness': sorted(closeness_Centrality.items(), key=lambda x: x[1],
41     reverse=True)[:10],
42     'eigenvector': sorted(eigenvector_Centrality.items(), key=lambda x: x[1],
43     reverse=True)[:10]
44 }

```

Listing 1: Python example

11. Code

Our code in which we mainly used for creating our data and making our visualizations are available on our [Github](#) the following table demonstrates what each file was used for.

File Name	File Description
NetworkAnalysis.py	Creates the degree distribution graph with the library networkx
fetchalbums.py	Fetches all of the artists albums and creates an edge list.
fetchdata.py	Creates a node list containing all the artists information.
handle_edges.py	Cleans edges file and creates a new edge list.
handle_nodes.py	Cleans node file and creates a new node list.
networkStatTest.py	Calculates metrics about our network and creates files based on it.

Table 3: Code file descriptions

References

- [1] Peechapat, W., & Puttanapong, N. (2024). Collaboration and Competition: A Social Network Analysis of Thailand's Music Industry. *Economies*, 12(2), 45. <https://doi.org/10.3390/economies12020045>
- [2] Deshmane, A., & Martínez-de-Albéniz, V. (2023). Come Together, Right Now? An Empirical Study of Collaborations in the Music Industry. *Management Science*. Advance online publication. <https://doi.org/10.1287/mnsc.2023.4743>
- [3] Cattani, G., & Ferriani, S. (2008). A Core/Periphery Perspective on Individual Creative Performance: Social Networks and Cinematic Achievements in the Hollywood Film Industry. *Organization Science*, 19(6), 824-844.
- [4] Uzzi, B., & Spiro, J. (2005). Collaboration and Creativity: The Small World Problem. *American Journal of Sociology*, 111(2), 090090. Retrieved from VSJ.