Music Collaboration Networks

Erik Malmqvist Jakobsen
erja@itu.dk
Student

Maria Astefanoaei
msia@itu.dk
Supervisor

Course code: BIBAPRO1PE *15/05–2025*

Contents

Preface

This bachelor project was made as part of the Software Development program at the IT University of Copenhagen during spring 2025. The project was written by Erik Malmqvist Jakobsen and supervised of Maria Astefanoaei.

All code and other related resources are available in the GitHub repository:

`https://github.com/emjakobsen1/music-collaboration-networks-ba`

The dataset was made public at:
`https://www.kaggle.com/datasets/minormajor/spotify-songs`

The visualizations were created with a mixture of tools available at `https://cytoscape.org`, `app.diagrams.net` and `https://matplotlib.org/`

Abstract

This project explores the structural and temporal dynamics of artist collaborations in music through the lens of network science. Using Spotify data, artist relationships are modeled as a temporal graph with yearly snapshots from 1986 to 2016. The evolution of the network and its giant component are analyzed to track the growth, structure, and connectivity over time. Betweenness centralities of the artists, are analyzed to track collaboration patterns within different music genres, and to identify central artists in the collaboration network, who potentially bridge different groups and communities. Entropy is used as a metric to measure diversity and identify cross-genre collaborators. The results show an increasingly connected collaboration network, with rising activity over time, yet one that becomes more centralized around a smaller set of highly connected artists. The analysis concludes by identifying the top collaborating artists, who have had impact on the collaboration landscape across the decades.

1 Introduction

Music is a phenomenon that researchers and scientists still do not fully understand the origin of. One popular theory suggests that music emerged in prehistoric times among early human-like species and evolved alongside the development of a holistic proto-language, serving as a foundational element in early human expression and interaction.[1]. Music is surely, both as a concept and it's role in culture and society, strongly connected to social domains.

Collaborations in music and have historically played a significant role in pushing cultural change. As an example of this, during the 1940s, jazz underwent a strongly innovative period, shaped by the African American community, leading to the emergence of bebop, which laid the foundation for modern jazz [2]. The importance of collaborations within music and art, serves as motivation to reason about these as social networks to study its structures and dynamics.

Network theory provides a powerful framework to model relationships and interactions between individuals. With the increasing availability of large-scale, structured datasets from platforms like Spotify, it has become possible to apply network analysis to domains such as music. Viewing artist collaborations through the lens of network science allows us to apply mathematical and computational techniques to analyze which structural properties and patterns they exhibit, and better understand the underlying factors driving their evolution.

Network theory is also a concept with a wide range of use cases, providing a broad abstraction model capable of representing everything from cellular processes and road systems to matchmaking algorithms and social networks. That also means that a problem which has a known solution in one field, surprisingly can lead to answers or insights to questions we have with regards to music collaboration networks that might not initially seem connected. This flexibility gives the opportunity to e.g. explore the evolution of music genre collaborations as a temporal graph or to identify central artists who act as structural bridges across genres or communities.

This study aims to explore the evolution of music collaboration networks over the period 1986 to 2016. Specifically, it seeks to address two main questions: (1) How has the volume and structure of collaboration changed over time? and (2) Which high-collaboration artists can be identified through quantitative network measures? By answering these questions, the project contributes to a better understanding of patterns in artistic collaboration and the roles of individual artists within the network. The following section reviews related work in this area and highlights various approaches to situate the context of this project.

2 Related Works

In recent years, numerous studies have investigated music collaboration networks from a variety of perspectives. Depending on the research focus, these networks can be approached through different views and with different goals. Coscia, for example, constructs a network of Italian composers with particular attention to regional background and how geographical closeness influences collaboration [3]. Schich et al. take a more historical angle by analyzing the network of British classical composers, focusing on how shared educational backgrounds contribute to community formation and how network position correlates with success [4]. Gienapp et al. emphasize genre differences, constructing a case of Jazz and Hip Hop to demonstrate how these different genres lead to distinct network structures and patterns of evolution [5]. Meanwhile, Oliveira et al. analyzes how inter-genre collaborations shape both regional and global music markets [6].

A methodological challenge is how to model the network itself. In most studies, a music collaboration network is represented as a graph structure, where each node corresponds to an artist, and an edge is added between two artists if they collaborated together. The data itself is embedded as node and edge attributes, such as artist name, genre, and the songs shared between collaborators. When building the graph, one approach, as done by Levy et al. is to begin with a mainstream artist, in this case *Drake*, and then iteratively expand the network through *Drake's* collaborators and their collaborators using a Breadth-First Search (BFS) [7]. The advantage of this approach is that it quickly leads to an exploration of a dense and influential part of the network. However, this is also its main limitation: it excludes artists who do not collaborate with the chosen starting node or its close neighbors, or artists who do not collaborate at all. Since the starting point is manually selected, this method also potentially introduces sampling bias and may not provide a representative picture of the broader collaboration landscape. To avoid this, this project instead constructs a bipartite graph linking artists and songs directly, and then projects it onto an artist–artist collaboration network where edges represent a collaboration, inspired by the work of Park et al. [4].

A structural modeling decision to be made, is which fundamental characteristics the graph should have. Let's say, that you wanted to emphasize certain roles of a collaboration, such as an artist contributing to another artists release e.g. as instrumentalist. To capture this asymmetry, you could make an argument for a graph with directed edges as done by Gienapp et al. [5]. However since collaborations are typically treated as a symmetric relationship, undirected edges are the most common choice.

Another consideration is the selection of appropriate node and edge attributes for the graph. In music collaboration networks, nodes often carry metadata such as genre, popularity, nationality, or periods of activity [3]. Edges, on the other hand, are typically weighted to reflect the intensity of a collaboration, e.g. measured by the number of co-released tracks between two artists. While edge weights can be good to capture collaboration strength, they also constrain the available tools. For example, Park et al. [4] uses clustering coefficients to examine the network of western classical composers, which is useful for identifying community structures. However, standard clustering coefficients do not account for edge weights and may not accurately reflect the strength of collaboration in weighted networks. While weighted variants exist [8], they are less commonly used.

When working with these kinds of networks, a commonly used concept is the giant component. The giant component refers to the largest connected subgraph in which any two nodes are connected by a path, and serves as a measure to understand the overall connectivity of a network. It has been used to study the evolution of collaboration networks over time, showing that the size of the giant component grows in proportion to the total number of nodes [9], which suggests that the network becomes more connected as more collaborators join.

Another concept in network analysis is centrality — in particular, betweenness centrality, which quantifies how often a node appears on the shortest paths between all pairs of nodes in the network. As a metric, betweenness centrality can highlight artists who act as bridges across genre or community boundaries, facilitating connections between otherwise separate groups of artists. In the context of research citation networks, Leydesdorff demonstrates that betweenness centrality can serve as an indicator of interdisciplinarity, as it reflects an journals role in linking multiple fields [10]. In a music setting, this translates to artists who bridge otherwise separate communities or genre clusters. From a computational perspective, the calculation of betweenness centrality is resource-heavy, particularly in large-scale graphs. One of the well known solutions, is Brandes' algorithm that provides an efficient approach for exact computation [11], but later sampling-based methods also exists, offering more scalable approximations for use in larger or dynamic networks [12].

Beyond structural metrics, recent work has explored diversity in networks using entropy. Entropy is a concept and metric widely used across fields, from physics to information theory, to have a quantitative measure of the balance between diversity and concentration within a set of entities. Oliveira et al. uses entropy to quantify the interdisciplinarity of scientific publications [13], which in a music context would reflect how genre-diverse an artist's collaboration neighborhood is, offering a way to quantify stylistic breadth. The most common way to measure this, is with Shannon Entropy, but other metrics such as the Brillouin index and the Gini coefficient have also been proposed to capture similar notions of diversity, each with their own strengths depending on the underlying data and distribution [14].

Building on this body of work, the present study combines network topology, genre classification, and temporal analysis to investigate how collaboration patterns evolve over time. In particular, it emphasizes genre diversity and the structural role of artists, using entropy and centrality metrics to quantify influence.

3 Data

3.1 Spotify Million Playlist Dataset

The dataset used in this study is based on The Spotify Million Playlist Dataset Challenge [15], a publicly available dataset designed to improve music recommendation systems. This dataset consists of one million user-generated playlists collected from Spotify between 2010 and 2017. It was selected for its large size and diverse representation of musical preferences in different time periods and genres. From the playlists, a subset of unique tracks was extracted and additional metadata was acquired by querying the Spotify Web API. In the end, the final processed dataset includes 874762 songs, by 186403 artists. Table 1 presents a sample of the JSON data structure. Each entry is indexed by the track ID and contains the specified fields, with artists represented as a nested object.

Track ID	`290aX99VkV4vLX2ab5sKrJ`
Track Name	Money On The Floor
Album Name	Live From The Underground
Release Year	2012
Duration (ms)	247160
Popularity	41
Explicit	True
Artists	**Name:** Big K.R.I.T. `ID: 0CKa42Jqrc9fSFbDjePaXP`
	Name: 8Ball & MJG `ID: 7iUhmKPNkkPPS6FCQxqtNq`
	Name: 2 Chainz `ID: 17lzZA2AlOHwCwFALHttmp`
Genres	Rap Southern Hip Hop Crunk Memphis Rap

Table 1: Sample Data Structure. The popularity score is an algorithmically determined metric provided by Spotify [16].

It is important to acknowledge certain limitations in the dataset. First, the data is inherently biased towards user-generated playlists, meaning that it does not fully represent all music available on Spotify. Additionally. The popularity scores fluctuate over time, since it has emphasis on recent listening trends. The dataset will only capture a snapshot in time, meaning that values may have changed since data collection. Finally, Spotify's genre annotation is by no means complete, and some tracks may lack assigned genres or be misclassified.

Another important limitation concerns the inconsistency of metadata. While the project aims to study collaborations among active or contemporary artists, the dataset includes many releases by classical composers who deceased centuries ago, but whose works continue to be featured in modern playlists. To mitigate this, the 25 most popular classical composers according to Pentreath were excluded from the data [17]. Additionally, releases involving more than five credited artists were filtered out, as these often represent compilation albums or highly aggregated credits that can introduce artificial noise in the network. These factors should be considered when interpreting the results.

3.2 Music Collaboration Graph

With the dataset, I am ready to define a music collaboration graph. As explored by Arram Bae et. al [4] by modeling the British composer network, the complete list of song data

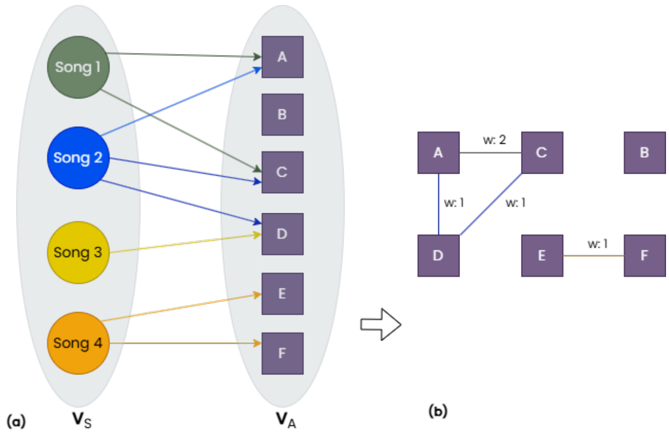

Figure 1: (a) Bipartite network of the songs and artists. (b) One-mode projection of the bipartite network onto the set of artists by connecting artists when they are associated with a common release. Edge weights equals the number of releases, they did together.

can naturally be represented as a bipartite graph with two node classes, V_S and V_A. V_S consists of the songs and V_A represents the artists from the dataset. There exists one or more edges from $v_S \in V_S$ to artists $v_A \in V_A$, if v_A was credited in the song's metadata.

To derive the yearly collaboration networks, I project this bipartite graph onto the artist node set. Specifically, for each year t, I form an undirected edge $(u, v) \in E_t$ between two artists u and v if they both contributed to at least one song released in year t. Figure 1 illustrates an example of the graph creating process.

With the artist-to-artist projection, the temporal music collaboration network is then modeled as a sequence of undirected graphs

$$\mathcal{G} = \{G_{1986}, G_{1987}, \dots, G_{2016}\},$$

where each single snapshot $G_t = (V_t, E_t)$ represents the state of the collaboration network in year t. This means the model is not accumulative or overlapping between years, but reflects discrete yearly views of the network, as described by Coscia [18].

The set of vertices V_t consists of all artists who released at least one record in year t. The set of undirected edges $E_t \subseteq V_t \times V_t$ captures collaborations between artists during the same year.

An edge $(u, v) \in E_t$ exists if artists u and v collaborated on at least one release in year t. Each edge is assigned a weight $w_t(u, v) \in \mathbb{N}$, representing the number of joint releases by u and v in that year.

3.2.1 Genres

Each node and edge in G_t is annotated with genre information derived from the songs of the dataset. For artist nodes, this takes the form of (1) one single primary genre, and (2) a multiset of genres representing all genre labels from the songs the artist participated in during year t, allowing for repeated occurrences of the same genre. Similarly, each edge $(u, v) \in E_t$ stores a multiset of genres corresponding to all collaborative tracks between artists u and v in that year.

From the 673 unique present genres, each artist node is assigned a single primary genre. The motivation for this, is to abstract some of the complexity away and facilitate some more general comparisons across years and genres.

To derive these primary genres, all genre labels were manually grouped into a set of nine supergenres (e.g., *Pop*, *Rock*, *Hip Hop*, etc.). This mapping process was done manually by personal judgment, but I will discuss alternative approaches in the later sections. The distribution among the frequencies of genre appearing in the data, and which super-genre they belong too, are visualized as a tree-map in Appendix 8.1. Once each genre was assigned to a super-genre, the primary genre of a node was determined by a simple majority vote over the node's genre multiset for a given year. In cases of ties, one of the top super-genres was selected at random.

4 Methodology

These sections outline the methodological approaches taken, and how they are implemented in this project from the context of the related works.

4.1 Giant Component

To understand the overall connectivity of music collaborations, this study examines the giant component in yearly snapshots of the network. The giant component is the largest connected subgraph, commonly used to assess whether a network is cohesive or fragmented. In this context, its growth over time is interpreted as a sign of increasing connectivity and broader access to collaboration opportunities. The size of the giant component serves as a tool to understand how centralized or diffuse the collaboration network is in a given year.

4.2 Degree Distribution and Power-Law Exponents

Degree distributions are analyzed for both the full collaboration graph and its giant component, across each year from 1986 to 2016. For each case, the distribution is fitted to a power-law model of the form

$$P(k) \sim k^{-\alpha}$$

Where k is the degree and α is the scaling exponent. The exponent α is estimated using maximum likelihood estimation (MLE), which identifies the value that best fits the observed degree distribution. This is done with the python package *powerlaw* [19]. A lower value of α indicates a heavier tail, meaning that a small number of artists maintain a high number of collaborations, which is typical for scale-free networks. Comparing these values over time provides insight into how collaborations are distributed across the network, whether they remain concentrated among a few central artists or becomes more evenly spread. This also allows us to assess whether the network maintains scale-free characteristics as it evolves.

4.3 Betweenness Centrality

Betweenness centrality is a widely used measure for identifying nodes that serve as bridges or intermediaries within a network. For a given artist, the betweenness centrality score reflects the number of shortest paths between all other pairs of artists that pass through that artist. High betweenness indicates that an artist plays a crucial role in connecting different parts of the network, potentially bridging otherwise disconnected groups or distant parts of the network.

The formal definition of betweenness centrality, is defined as

$$C_B(v) = \sum_{s,t \in V} \frac{\sigma(s,t|v)}{\sigma(s,t)}$$

where $\sigma(s,t)$ is the number of shortest paths between nodes s and t, and $\sigma(s,t|v)$ is the number of those paths that pass through node v.

If $s = t$, then $\sigma(s,t) = 1$. If $v \in \{s,t\}$, then $\sigma(s,t \mid v) = 0$, since a node is not considered to lie on a path between itself and another node.

In this study, edge weights are not considered and all collaborations are treated equally. This means that the shortest path distances are computed purely on the basis of the number of edges and not their strength.

Betweenness centrality is computed annually for all nodes in the collaboration graph from 1986 to 2016. In later years, as the network grows larger and exact computation becomes increasingly resource-intensive, an approximation method is used to estimate scores more efficiently. While these approximations may not yield precise values, they preserve the relative ordering of nodes. Meaning that artists with higher centrality still rank above those with lower centrality. These scores help highlight artists who occupy central structural positions in the network.

4.4 Entropy

Entropy is used here to measure the diversity of genres in an artist's collaboration neighborhood. A higher entropy score, will suggest that an artist collaborates more broadly across genres, and a lower score suggests that artists are more likely to collaborate within the same genres.
To compute an artist's entropy value, I take the same approach as Silva et al. [13]. I construct the histogram $h_a(g)$, representing the frequency of each genre g appearing

among the artists connected to artist a. As described in 3.2, genre labels from these artists are appended to the artist's genre attribute when constructing the graph, allowing duplicates, meaning that genres can contribute multiple times based on the number of collaborations. This also makes edge weights redundant in this case, as including them would introduce the problem of having genre counts scale artificially with the number of collaborators per release.

To avoid the entropy measure being dominated by the most frequent genre labels, each histogram $h_a(g)$ is normalized by dividing by the global frequency of genre g throughout the entire network:

$$\bar{h}_a(g) = \frac{h_a(g)}{\sum_j h_j(g)}$$

The normalized values $\bar{h}_a(g)$ are then transformed into a probability distribution $P_a(g)$:

$$P_a(g) = \frac{\bar{h}_a(g)}{\sum_{g'} \bar{h}_a(g')}$$

Finally, the entropy is computed as:

$$H_a = -\sum_g P_a(g) \cdot \log P_a(g)$$

One of the limitations of this approach is that it does not consider the closeness or relationship between genres, and therefore does not account for semantic or stylistic similarity. Shannon entropy does not capture the natural hierarchical structure of genre labels. For example, Shannon entropy treats *synth pop* and *pop* as equally distinct as *heavy metal* and *christmas*, even though the former are more closely related. As such, while entropy gives a useful indication of variety, it does not directly reflect genre distance or blending across genre hierarchical trees.

5 Results and Analysis

5.1 Graph Summary

As a foundation for examining the evolution of the network, Table 2 provides a yearly overview of the graph snapshots from 1986 to 2016. For each year, the total number of artist nodes and collaboration edges are listed, along with giant component (GC), including its size and average degree.

Year	Nodes	Edges	GC Nodes	GC Edges	Avg. Degree	GC Avg. Degree	Collab %
1986	932	318	30	51	0.682	3.4	11.41
1987	1175	619	64	106	1.054	3.312	14.48
1988	1330	708	58	95	1.065	3.276	16.02
1989	1340	591	58	100	0.882	3.448	13.13
1990	1808	923	157	279	1.021	3.554	16.16
1991	1855	924	52	97	0.996	3.731	13.66
1992	1986	932	135	256	0.939	3.793	13.22
1993	2226	906	91	168	0.814	3.692	11.10
1994	2664	1376	267	471	1.033	3.528	13.84
1995	2821	1303	182	326	0.924	3.582	14.36
1996	3064	1418	71	127	0.926	3.577	13.01
1997	3315	1601	182	340	0.966	3.736	13.95
1998	3648	2072	337	643	1.136	3.816	15.63
1999	4239	2435	452	784	1.149	3.469	15.99
2000	4666	2440	412	720	1.046	3.495	15.97
2001	4845	2473	352	686	1.021	3.898	12.74
2002	5060	2417	449	755	0.955	3.363	12.71
2003	5825	2758	649	1141	0.947	3.516	13.55
2004	6297	2840	487	887	0.902	3.643	12.83
2005	8086	3605	748	1310	0.892	3.503	13.15
2006	9646	4332	977	1794	0.898	3.672	13.32
2007	10549	4528	1058	1830	0.858	3.459	13.41
2008	11955	4884	797	1422	0.817	3.568	12.07
2009	14574	6348	1312	2238	0.871	3.412	13.66
2010	17738	7902	1839	3197	0.891	3.477	13.77
2011	21715	10608	2966	5146	0.977	3.470	15.06
2012	25412	14748	4416	7804	1.161	3.534	18.05
2013	28092	16985	5526	9794	1.209	3.545	18.74
2014	31562	20256	7007	12129	1.284	3.462	21.16
2015	35293	24968	9209	16504	1.415	3.584	23.47
2016	39085	29708	11723	20987	1.520	3.580	25.67

Table 2: Graph summary of the collaboration network. The collaboration percentages are the share of nodes from the bipartite graph with degree greater than one, or simply the songs, with more than one credited artist.

The collaboration percentage represents the proportion of songs in each year that were created by more than one credited artist. In the bipartite graph, this corresponds to the share of song nodes with a degree greater than one. Interestingly, while the collaboration percentage increases significantly over time, from around 11% in the late 1980s to over 25% by 2016—the average degree of the artist nodes grows at a much slower rate. This suggests that although collaborations became more common, they did not necessarily lead to a denser network of artist relationships. One possible explanation is, that collaborative songs are released by the same highly active artists, which inflates the collaboration percentage, as these primarily reinforces existing links without increasing the average number of distinct co-collaborators per artist.

Another interpretation of this, is that while collaboration frequency increases, but the average degree does not grow at the same pace, could be because the artists collaborate more frequently but with the same kind of artists rather than different ones.

Fig. 2 presents a snapshot of the collaboration network in 2016, highlighting the giant component alongside the smaller disconnected subcomponents. The giant component dominates the network and connects the majority of active artists.

5.2 Degree Distribution

To explore this idea further, I examined the degree distribution of the collaboration network over time. Fig. 3 and Fig. 4 shows a plot of the frequencies of degrees for artists over the years. The degree distribution for both the full network and the giant components shows a pattern that, some artists are super collaborators while the majority has very few collaborations. This pattern is typical for social networks, which exhibits scale-free properties.

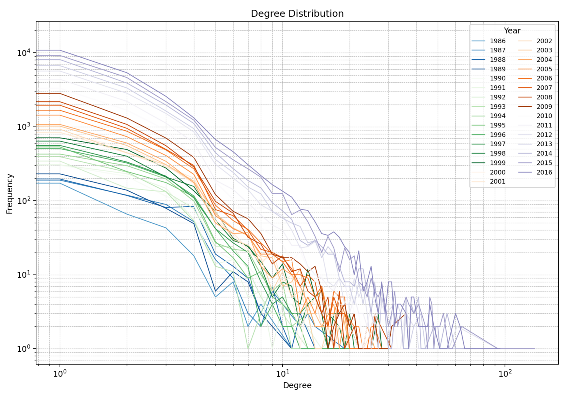

Figure 3: Log-Log plot of the artists' degree distributions over years.

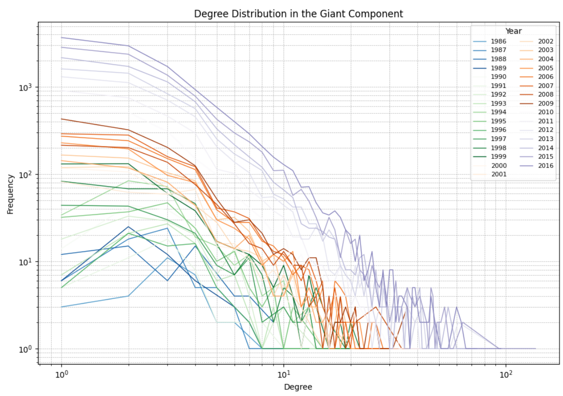

Figure 4: Log-Log plot of the artists degree distribution in the giant component over years.

The power-law exponent α was computed for both the full graph and its giant component each year. As shown in Figure 5, the values of α fluctuate considerably in the early years but gradually stabilize from the late 1990s onward.

High values of α, in particular 1990 and 1993, suggest a network with very few super-collaborators and degrees more evenly spread. These spikes possibly reflect sparsity in the early data, where fewer collaborations are recorded and the network is still fragmented, but could also indicate the presence of disconnected subgroups or more local music communities that had not yet integrated into the giant component.

Figure 2: Snapshot of the 2016 music collaboration network, the giant component and the sub-components. Artist node sizes reflect degree; edge thickness indicates collaboration weight; node colors reflect main genre classification.

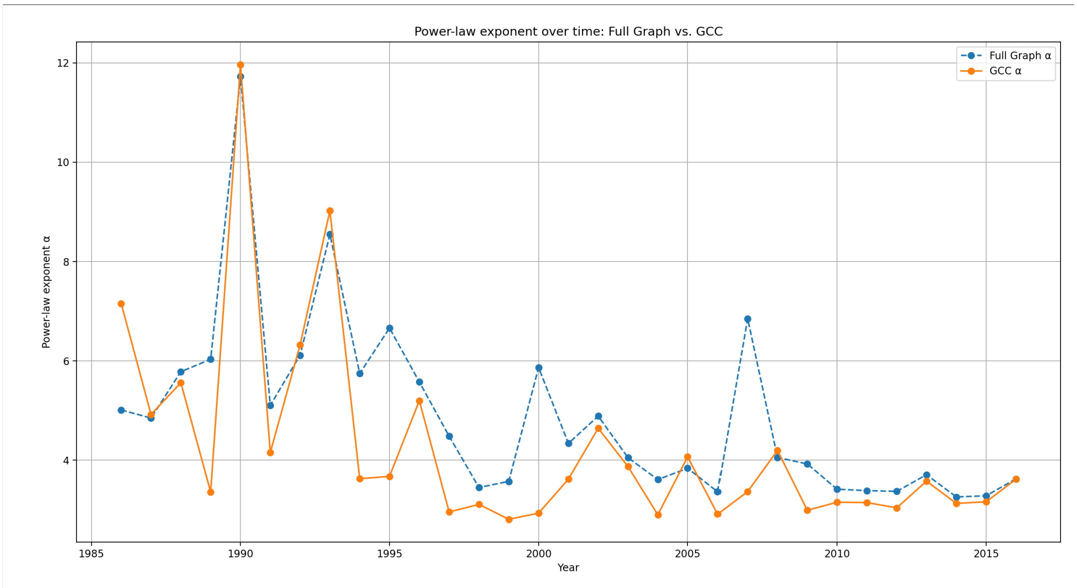

Figure 5: The power-law exponents of the yearly snapshots in the full graph and and the giant components.

As the network matures, α values converge toward a range between 3 and 4. This indicates a more classic structure, that a small number of artists are highly connected, while the majority maintain few connections. These lower α values correspond to heavier-tailed distributions, typical of social and collaboration networks where hubs emerge over time.

One possible interpretation of the convergence between the power-law exponents of the full graph and the giant component is that the overall music collaboration network has become more integrated over time. As the network evolves, more artists appear to join the giant component, meaning their collaborations now link them into the broader, shared network. This could potentially reflect a general trend in the music industry, that the rise of global platforms like Spotify have contributed in facilitating collaborations, making it easier to co-release songs.

5.3 Genre Evolution

Figure 6 shows the evolution of mean betweenness centrality scores by genre from 1996 to 2016. Two key patterns stand out. First, genres like *hip-hop* and *r&b* consistently rank higher than others, particularly from the early 2000s onward. This suggests that artists in these genres frequently act as bridges or as facilitators of collaborating within the network. Their central structural role may reflect a more collaborative culture of these genres. On the other hand, genres like *classical*, *country*, and *folk-world* remain consistently lower in betweenness centrality. These genres may be more self-contained, with fewer collaborations in general, or less represented in global digital platforms.

Secondly, there is a general upward trend in mean betweenness centrality scores across most

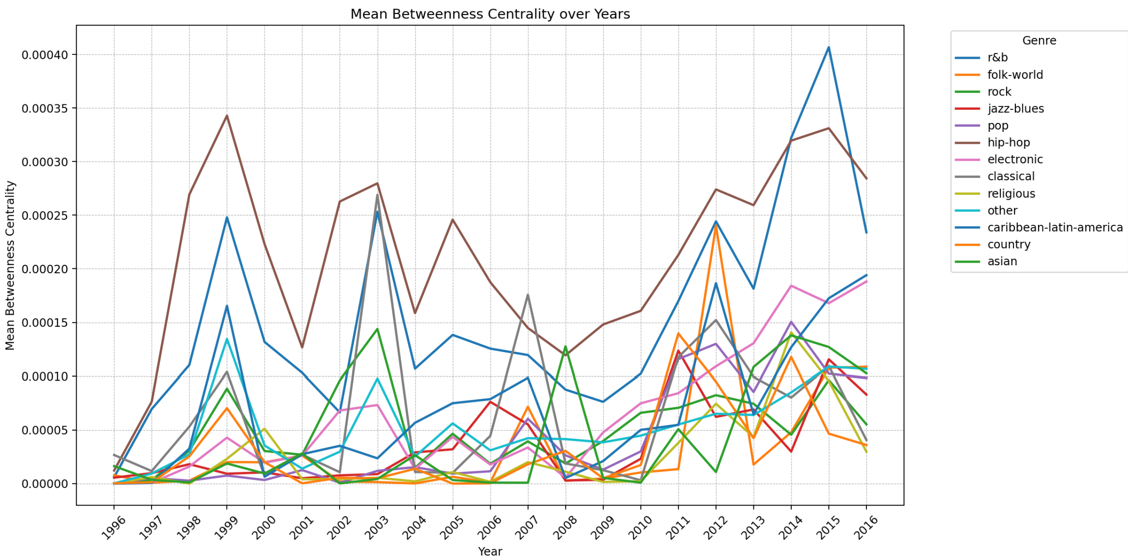

Figure 6: Mean betweenness centrality of the different main genres throughout the years.

genres throughout the time period. This could indicate, that some artists are increasingly occupying structurally important positions. This can happen when the network relies more heavily on specific individuals to link otherwise separate parts of the graph together.

In summary, although collaboration frequency increases over time, as shown by the rising collaboration percentage, this does not result in a proportionate rise in average degree. This suggests that while artists are collaborating more, they often do so within familiar circles, reinforcing existing relationships rather than forming new ones.

At the same time, the growth of the giant component and the convergence of the power-law exponent indicate that the network is becoming more globally integrated. More artists are connected to the broader network, but the mean betweenness centrality suggests that smaller number of them take on increasingly important roles in holding the network together.

So while the network as a whole may not be significantly denser in terms of local connections per artist, it is becoming more structurally centralized, with collaboration more facilitated through key artists who act as bridges. This reflects a shift toward a more hierarchical structure, where a few highly connected artists serve as access points between otherwise distinct groups of artists.

5.4 High Collaboration Artists

This section highlights the most prominent collaborators in the network, identified through high scores in both entropy and betweenness centrality. Appendix 8.2 provides a year-by-year overview of the top 10 artists for each metric. Several artists stood out by ranking in the top 10 for both measures in the same year — indicating not only a high degree of stylistic diversity in their collaborations but also a structurally central position in the network. Those who were among the top 10 artists in both measures include *Donald Fraser* and *Berliner Philharmoniker* (1996), *JAY-Z* (1998), *Xzibit* (2000), *Jadakiss* and *Diddy* (2001), *Linkin Park* (2002), *Snoop Dogg* (2006), *Akon* (2008), *Busta Rhymes* and *Lil Wayne* (2009), *Pitbull* (2010), *Tiësto* (2012), *Dido* and *French Montana* (2013), *Yellow Claw* (2015), *Ty Dolla $ign* and *French Montana* (2016).

Some artists appeared in the top 10 across multiple decades, showing a long-term presence in music collaborations. Among all, Snoop Dogg had the most total appearances, showing up 14 times between 2000 and 2016. Close behind was Indian film composer A.R. Rahman, who appeared 12 times from 1999 to 2016. Figure 7 visualizes the two-hop collaboration network centered around *Snoop Dogg* in 2011. Edge and node colors represents the different main genres of *Snoop Dogg's* neighborhood.

A few others also stood out for their decade-spanning relevance. Shankar Mahadevan and the trio Shankar-Ehsaan-Loy related to the scene of Indian sountrack, appeared in the early 2000s and showed up again in later years. Ludacris made it to the list in the early 2000s and again in the mid-2010s, pointing to a career with sustaining impact. While artists like Too $hort only show up a couple of times, they still span a wide time range, hinting at a lasting, if more occasional, impact on the collaboration scene.

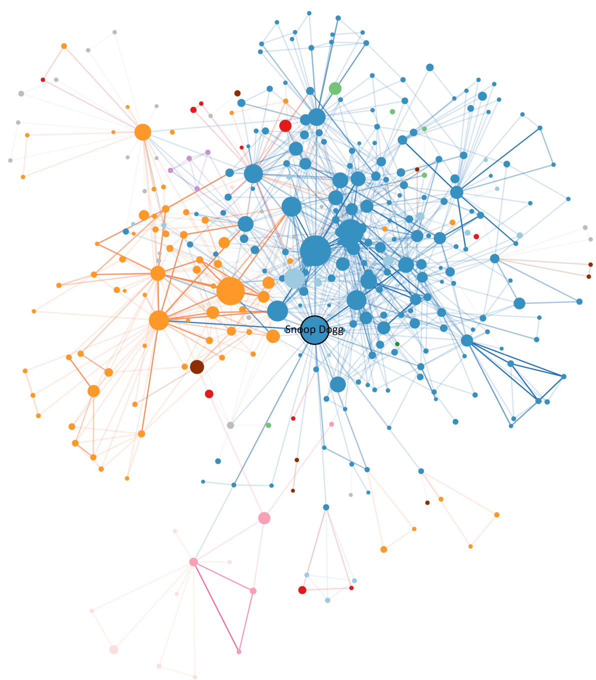

Figure 7: Two-hop collaboration network centered on Snoop Dogg in 2011. Appendix 8.3 has a full size version with artist labels.

6 Discussion

In this discussion, I will reflect on how the results can be interpreted, consider alternative approaches that could have been taken, address limitations and what did not go as planned.

To answer how the structure and volume of collaborations has changed throughout the years, the results show, that collaborations in music have increased significantly over the 30-year period studied. The raw collaboration percentage rose from around 11% in 1986 to over 25% in 2016. However, the average degree for artists did not grow nearly as much, which suggests that while more collaborations are happening, they often take place within the same circles. It might have been helpful to include edge weights in this part of the analysis, as they were in the end not used at all. These could potentially have helped explain, why the overall percentage of collaborations increase but the distinct collaborators of an artist does not.

The growing size of the giant component further indicates that more artists are becoming part of a shared, interconnected collaboration space. Combined with the convergence of power-law exponents between the full graph and the giant component, this suggests that the network is becoming more globally integrated. At the same time, the average betweenness centrality is increasing across nearly all genres. This could indicate that integration is not happening evenly, but rather through a smaller number of key artists who act as bridges between different groups or clusters. The network does not seem to become uniformly denser, instead, it is becoming more structurally centralized around influential artists.

Entropy was used in this project to quantify how genre-diverse an artist's collaboration neighborhood is, and how stylistically broad their connections are. While this gives some insight at the individual artist level, it's still fair to question how well this approach actually works in this context.

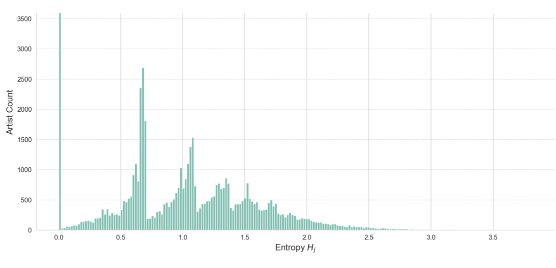

Figure 8: Entropy distribution of the period 1996 to 2016.

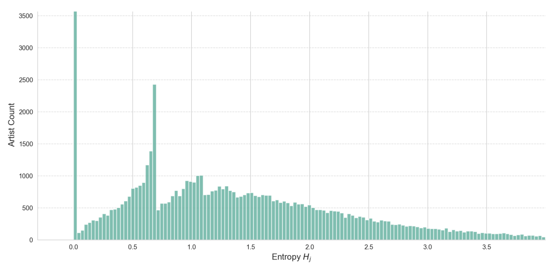

Figure 9: A null model with randomized genres on the nodes.

To investigate this, entropy was calculated on the full network — not as yearly snapshots, in order to get a larger and more stable dataset. Figure 8 shows the distribution of entropy scores for all artists across the full period. The idea to use entropy in this way is inspired by Silva et al. [13], who applied it in academic citation networks to measure the interdisciplinarity of journals. Based on that, I had hoped the distribution would follow a power-law shape, similar to how degree distributions often do in network analysis.

However, the resulting entropy distribution showed clear plateaus, especially around values at 0.69, 1.11, and 1.60. These correspond to log-based values of 1, 2, and 3, which makes sense, since many low-degree artists in the dataset are only annotated with one, two, or three genres. To check whether the entropy calculations themselves were working correctly, Figure 9 shows a null model, where the number of genres per artist was kept the same, but the genres were randomly shuffled across nodes.

The null model shows plateaus around the same values as the actual distribution, but overall follows a smoother curve, more in line with what you would expect from randomized data. This suggests that the entropy was calculated correctly. However, it also suggests that the entropy measure is less effective in this dataset than hoped, mainly because many artists simply do not have enough genre labels to produce a meaningful spread. In contrast, the citation networks studied by Silva et al.[13] involve hundreds of citations per journal, which provides much richer data to work with. Here, the relatively small number of genre annotations per artist limits how much insight entropy can provide on its own.

Despite these issues, ranking artists by betweenness centrality and entropy ended up highlighting several well-known artists, for example, *Snoop Dogg* and *JAY-Z*. So even though there were some challenges with how well the chosen metrics capture diversity or bridge-building roles, it is still satisfying to see artists like these stand out in the results. Both can be argued to have played major roles in shaping the mainstream music scene and have had careers marked by genre shifts and collaborations across musical boundaries.

6.1 Limitations

One important discussion point concerns the dataset itself, which presented several challenges from the beginning due to inconsistent metadata. For example, classical composers were filtered out during preprocessing, as they are long deceased and their inclusion introduced noise in the data. Despite these efforts, a significant amount of unintended content remains. In particular, aggregated Spotify profiles such as *Various Artists* or *Traditional* caused problems. These profiles often have millions of monthly listeners and scored highly on metrics like entropy and betweenness centrality, even though they don't represent a single, real artist, but rather a collection of different contributors. In such cases, it was difficult to identify the actual artist behind a release, which made the interpretation of results more complicated.

Beyond the metadata itself, the size and balance of the dataset across time also raise questions. The earliest years, especially from 1986 to 1996, are heavily underrepresented compared to later periods. This makes it difficult to assess whether patterns observed in those early years, such as spikes or drops in average degree or power-law exponents, reflect real structural properties or are simply due to limited data. Similarly, as the dataset grows larger in the 2000s and 2010s, it becomes harder to say whether the convergence of metrics like the power-law coefficient is driven by actual changes in the collaboration network or just the effect of more data smoothing out variation.

Another limitation relates to how genres were handled. In this project, all genre labels were manually grouped into a smaller set of supergenres based on my personal judgment. But what qualifies as a supergenre is subjective, and the grouping process could have been more robust. One solution could be, to have queried an external service for assistance. For example, McDonald has developed a tool that, for each Spotify genre, provides a ranked list of the most similar genres based on audio features and listening patterns [20].

The dataset also contained several other fields that could have been interesting to explore. For example, the popularity metric ended up not being used at all, but it could have been relevant to study whether there is a structural connection between collaboration patterns and artist success. It might have been possible to compare how different types of collaboration relate to popularity. Popularity is just one of many data points available through Spotify's API. Other features like tempo, speechiness, energy, and acousticness could also have been used to e.g. classify genres in a more data-driven way, rather than relying only on manual grouping.

Another thing worth discussing is the choice of how the network was modeled over time. In this project, each graph snapshot was built per year, meaning that edges only reflect collaborations that happened within that specific year. While this gives a clean and simple view of the network's evolution, it also means some information is lost, for example, longer-term or repeated collaborations across years are not captured. An alternative approach could have been to model the graph in an accumulative way, where collaborations persist across the years. This or other temporal modeling choices, could especially have brought more insights when analyzing the individual artists collaborations, and in general might have given a different picture of connectivity over time.

In summary, this project shows how network-based metrics can reveal patterns in music collaborations. Despite certain methodological and data-related challenges, the study demonstrates use of network analysis to provides a structured way to examine how artists connect, how collaboration evolves over time, and how roles such as influence and diversity can be quantified, highlighting the value of using network science to understand dynamics of music.

7 Conclusion

This project set out to examine the evolution and structure of music collaboration networks from 1986 to 2016 by modeling artist relationships as a temporal graph. Through a combination of structural metrics, such as degree distribution, giant component size and power-law exponents, as well as node-oriented measures such as betweenness centrality and entropy, the analysis provided insight into how collaboration patterns have changed over time.

The results show that the network has become increasingly connected, with more artists joining the giant component and overall collaboration frequency rising. However, this growth has not translated into a uniformly denser network. Instead, the results suggest a shift toward greater structural centralization, where a smaller number of highly connected artists play a key role in holding the network together. Betweenness centrality suggests the emergence of such bridge-building artists, particularly in genres like Hip-Hop and R&B.

Inspired by similar uses of entropy in interdisciplinary research, entropy was here used to explore genre diversity in collaborations of artists. When paired with centrality, it helped identify artists with both stylistic breadth and structural importance.

Overall, this study demonstrates how network science can offer valuable perspectives on artistic collaboration and genre dynamics in musical collaboration networks. While some limitations, especially around data quality and genre handling affected the insights, the project still provides a meaningful approach to studying structural and temporal patterns in music collaboration networks. Network metrics are shown to be useful for reasoning about artists influence, connectivity, and diversity, and they help frame music not only as an artistic product but as a connected, evolving system shaped by patterns of collaboration.

References

[1] Steven J. Mithen. *The Singing Neanderthals: The Origins of Music, Language, Mind, and Body*, page 2. Orion Publishing Co, 2006.

[2] Colin M. Messinger. *How Bebop Came to Be: The Early History of Modern Jazz*. The Cupola, 2013. https://cupola.gettysburg.edu/cgi/viewcontent.cgi?article=1263&context=student_scholarship.

[3] Michele Coscia. *Node Attribute Analysis for Cultural Data Analytics: A Case Study on Italian XX–XXI Century Music*. Appl Netw Sci 9, 2024. https://link.springer.com/article/10.1007/s41109-024-00669-5.

[4] Maximilian Schich Juyong Park Doheum Park, Arram Bae. *Topology and evolution of the network of western classical music composers*. EPJ Data Science, 2015. https://epjdatascience.springeropen.com/articles/10.1140/epjds/s13688-015-0039-z.

[5] Manuel Burghardt Lukas Gienapp, Clara Kruckenberg. *Topological properties of music collaboration networks: The case of Jazz and Hip Hop*. Digital Humanities Quarterly, 2021. https://digitalhumanities.org/dhq/vol/15/1/000504/000504.html.

[6] Danilo B. Seufitelli Anisio Lacerda Mirella M. Moro Gabriel P. Oliveira, Mariana O. Silva. *Detecting Collaboration Profiles in Success-Based Music Genre Networks*. International Society for Music Information Retrieval, 2020. https://archives.ismir.net/ismir2020/paper/000275.pdf.

[7] Vincent Po Daniel Wood Erica Wu Akash Levy, Sunny He. *The Musical Collaboration Graph*. 2017. https://www.academia.edu/30953480/The_Musical_Collaboration_Graph.

[8] Pietro Panzarasa Tore Opsahl. *Clustering in weighted networks*, page 155–163. Social Networks 31, 2009. https://www.sciencedirect.com/journal/social-networks.

[9] Jörn Grahl Pascal Budner. *Collaboration Networks in the Music Industry*. 2016. https://www.researchgate.net/publication/309631728_Collaboration_Networks_in_the_Music_Industry.

[10] Loet Leydesdorf. *Betweenness as an Indicator of the Interdisciplinarity of Scientific Journals*. Journal of the American Society for Information Science and Technology, 2007. https://www.leydesdorff.net/betweenness/betweenness.pdf.

[11] Ulrik Brandes. *A Faster Algorithm for Betweenness Centrality*. 2001. https://snap.stanford.edu/class/cs224w-readings/brandes01centrality.pdf.

[12] Kamesh Madduri David A. Bader, Shiva Kintali and Milena Mihail. *Approximating Betweenness Centrality*. 2007. https://davidbader.net/publication/2007-bkmm/2007-bkmm.pdf.

[13] O.N. Oliveira Jr L. da F. Costa F.N. Silva, F.A. Rodrigues. *Quantifying the interdisciplinarity of scientific journals and fields*, pages 469–477. Journal of Infometrics, Volume 7, Issue 2, 2013. https://www.sciencedirect.com/science/article/abs/pii/S1751157713000096.

[14] Jesper Wiborg Schneider Qi Wang. *Consistency and validity of interdisciplinarity measures.* 2020. https://direct.mit.edu/qss/article/1/1/239/15574/Consistency-and-validity-of-interdisciplinarity.

[15] Johannes Kepler University Spotify, The University of Massachusetts. *Spotify Million Playlist Dataset Challenge.* Spotify, 2020. https://www.aicrowd.com/challenges/spotify-million-playlist-dataset-challenge.

[16] Spotify. *Spotify Web API.* Spotify, 2025. https://developer.spotify.com/documentation/web-api/reference/get-track.

[17] Rosie Pentreath. *The 30 greatest classical music composers of all time.* Classic.fm, 2020. https://www.classicfm.com/composers/greatest-classical-music-history.

[18] Michle Coscia. *The Atlas for the Aspiring Network Scientist*, page 114. 2025. https://arxiv.org/pdf/2101.00863.

[19] Dietmar Plenz. Jeff Alstott, Ed Bullmore. *powerlaw: A Python Package for Analysis of Heavy-Tailed Distributions.* 2014. https://github.com/jeffalstott/powerlaw.

[20] Glenn McDonald. *Every Noice at Once.* 2025. https://everynoise.com/.

8 Appendix

8.1 Super-Genre Tree Map Plot

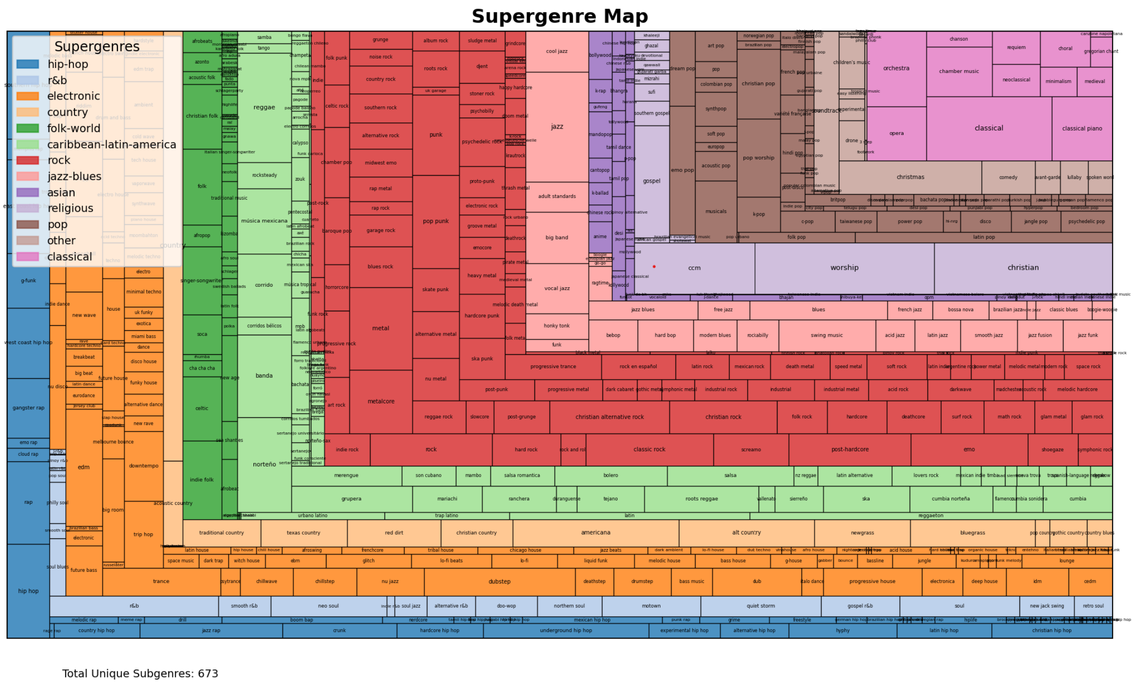

Figure 10: The distribution of genres from the data, mapped to super-genre parents. The box size's area represent the frequency of the genres.

8.2 Top 10 Scoring Artists

Year	Rank	Artist	Betweenness Centrality	Artist	Entropy
1996	1	Donald Fraser	0.000314	Donald Fraser	2.7703
	2	Traditional	0.000300	Björk	2.5096
	3	Wiener Philharmoniker	0.000261	Hariharan	2.3302
	4	Berliner Philharmoniker	0.000202	Berliner Philharmoniker	2.2868
	5	Gil Shaham	0.000163	Traditional	2.2649
	6	John Eliot Gardiner	0.000148	White Zombie	2.1144
	7	Sylvia McNair	0.000133	David Morales	2.1056
	8	JAY-Z	0.000128	Ella Fitzgerald	2.0896
	9	Choir of Paisley Abbey	0.000127	Alka Yagnik	2.0677
	10	André Previn	0.000126	Hollywood Bowl Orchestra	2.0568
1997	1	Busta Rhymes	0.000669	Shankar Mahadevan	2.8162
	2	Mase	0.000646	The Prodigy	2.5397
	3	Too $hort	0.000622	Wyclef Jean	2.5268
	4	SWV	0.000521	Nine Inch Nails	2.4989
	5	D-Shot	0.000508	Nadeem Shravan	2.4984
	6	JAY-Z	0.000503	Instant Karma	2.4409
	7	E-40	0.000482	Mahalakshmi Iyer	2.4272
	8	Redman	0.000460	B.B. King	2.3736
	9	Babyface	0.000454	Traditional	2.3718
	10	The Notorious B.I.G.	0.000453	Slash	2.3107
1998	1	Inspectah Deck	0.001754	Jatin-Lalit	2.3358
	2	JAY-Z	0.001720	Charlotte Church	2.3304
	3	DJ Clue	0.001702	Too $hort	2.3223
	4	Big Pun	0.001663	Sian Edwards	2.3072
	5	Kurupt	0.001432	Jermaine Dupri	2.2992
	6	Too $hort	0.001395	Alka Yagnik & Arvind Hasabnish	2.2454
	7	London Symphony Orchestra	0.001022	Bounty Killer	2.2014
	8	Daz Dillinger	0.001013	Vairamuthu	2.1845
	9	Snoop Dogg	0.000995	Silkk The Shocker	2.1693
	10	DMX	0.000983	JAY-Z	2.1265
1999	1	The Notorious B.I.G.	0.003927	Udit Narayan	2.6668
	2	Snoop Dogg	0.001922	A.R. Rahman	2.6081
	3	Traditional	0.001841	Alka Yagnik	2.5419
	4	Dr. Dre	0.001410	Hema Sardesai	2.5411
	5	Missy Elliott	0.001329	Handsome Boy Modeling School	2.4999
	6	Mary J. Blige	0.001316	Choir of King's College, Cambridge	2.4827
	7	Bob Marley & The Wailers	0.001300	Stephen Cleobury	2.4827
	8	Method Man	0.001091	Kavita Krishnamurthy	2.4637
	9	Lisa Harris	0.001088	Lisa Harris	2.3912
	10	Nas	0.001045	Vienna Boys' Choir	2.3862
2000	1	Snoop Dogg	0.002338	A.R. Rahman	2.9431
	2	Xzibit	0.001619	Veturi	2.8559
	3	Eminem	0.001556	Traditional	2.6694
	4	Mary J. Blige	0.001433	Yo-Yo Ma	2.6248
	5	Guru	0.001259	Xzibit	2.5370
	6	DJ Clue	0.001244	Sackcloth Fashion	2.5261
	7	Junior Reid	0.001068	Mark O'Connor	2.5115
	8	Ghostface Killah	0.000983	Shankar Mahadevan	2.4844
	9	Method Man	0.000888	S. P. Balasubrahmanyam	2.4673
	10	Busta Rhymes	0.000777	Edgar Meyer	2.4529
2001	1	Missy Elliott	0.001214	Udit Narayan	2.7244
	2	Jadakiss	0.000924	A.R. Rahman	2.5879
	3	Ja Rule	0.000815	Sukhwinder Singh	2.4958
	4	2Pac	0.000728	Alka Yagnik	2.4591
	5	Diddy	0.000676	Anu Malik	2.4282
	6	Busta Rhymes	0.000636	Jadakiss	2.4171
	7	Fat Joe	0.000635	Diddy	2.3186
	8	Lil' Mo	0.000635	Trinity Church Choir, New York	2.3161
	9	Ludacris	0.000575	E-40	2.2888
	10	UGK	0.000569	8Ball	2.2783

Year	Rank	Artist	Betweenness Centrality	Artist	Entropy
2002	1	JAY-Z	0.003140	Nitty Gritty Dirt Band	2.5426
	2	Nate Dogg	0.002083	Talib Kweli	2.5361
	3	Pharoahe Monch	0.001528	Alison Krauss	2.5135
	4	Daz Dillinger	0.001211	Traditional	2.4895
	5	Snoop Dogg	0.001203	Doc Watson	2.4845
	6	Linkin Park	0.001165	Jazzanova	2.4444
	7	2Pac	0.001096	Taj Mahal	2.4051
	8	Sean Paul	0.001063	Linkin Park	2.3557
	9	Cam'ron	0.000984	The Band	2.3497
	10	Nas	0.000949	Twista	2.2211
2003	1	Eve	0.005068	A.R. Rahman	2.7506
	2	Mary J. Blige	0.005062	Ludacris	2.5151
	3	Sting	0.005045	Shankar-Ehsaan-Loy	2.4699
	4	Nick Ingman	0.004909	Karthik	2.4370
	5	Georges Bizet	0.004828	Daft Punk	2.3686
	6	London Symphony Orchestra	0.004589	George Gershwin	2.3450
	7	Jadakiss	0.004021	Tippu	2.3343
	8	Gustav Holst	0.003376	Ira Gershwin	2.2723
	9	Alexander Borodin	0.003191	Traditional	2.2615
	10	Joshua Bell	0.002484	Chitra Sivaraman	2.2556
2004	1	Jadakiss	0.001594	Shankar Mahadevan	2.7798
	2	Kanye West	0.001280	Karthik	2.6903
	3	Snoop Dogg	0.001127	A.R. Rahman	2.6350
	4	Pharrell Williams	0.000835	Handsome Boy Modeling School	2.5780
	5	2Pac	0.000830	Toots & The Maytals	2.5729
	6	Bootsy Collins	0.000792	Twista	2.5690
	7	Elton John	0.000758	Willie Nelson	2.5439
	8	Toots & The Maytals	0.000755	Udit Narayan	2.5320
	9	Lil Jon	0.000712	Shankar-Ehsaan-Loy	2.5112
	10	Ray Charles	0.000671	Nelly	2.4434
2005	1	The Notorious B.I.G.	0.001854	A.R. Rahman	2.7752
	2	Slim Thug	0.001373	Billy Joel	2.4848
	3	Three 6 Mafia	0.001238	Choir of King's College, Cambridge	2.4630
	4	CeeLo Green	0.001174	Anthony Hamilton	2.4398
	5	DANGERDOOM	0.001069	Sir David Willcocks	2.4327
	6	Jazze Pha	0.000970	Baby Bash	2.4240
	7	The Game	0.000940	Ying Yang Twins	2.4224
	8	Eminem	0.000854	Wu-Tang Clan	2.4114
	9	MF DOOM	0.000840	Royal Philharmonic Orchestra	2.4041
	10	Pitbull	0.000823	Twista	2.3831
2006	1	Snoop Dogg	0.002775	Massive Attack	2.8432
	2	Andrea Bocelli	0.001463	A.R. Rahman	2.7576
	3	Stevie Wonder	0.001428	Pharrell Williams	2.5481
	4	David Foster	0.001167	Beyoncé	2.4962
	5	Young Buck	0.000914	Lloyd Banks	2.4931
	6	Sérgio Mendes	0.000838	Shankar-Ehsaan-Loy	2.4713
	7	Too $hort	0.000837	Shankar Mahadevan	2.4656
	8	Juelz Santana	0.000832	Young Buck	2.4388
	9	Don Omar	0.000792	Lil Jon	2.4205
	10	John Legend	0.000763	Snoop Dogg	2.3872
2007	1	Miguel Bosé	0.003696	Alison Krauss	2.8918
	2	Andrea Bocelli	0.003679	Robert Plant	2.7840
	3	Laura Pausini	0.003592	Akon	2.5576
	4	Shakira	0.003352	Busta Rhymes	2.4603
	5	Lang Lang	0.003349	Choir of King's College, Cambridge	2.4506
	6	Christoph Eschenbach	0.003338	Philip Ledger	2.4425
	7	Philadelphia Orchestra	0.003316	Traditional	2.4006
	8	Wyclef Jean	0.003187	Stephen Cleobury	2.3629
	9	Nikolai Rimsky-Korsakov	0.003172	Guru	2.3311
	10	Katherine Jenkins	0.002892	Solar	2.3311

Year	Rank	Artist	Betweenness Centrality	Artist	Entropy
2008	1	Akon	0.001112	London Philharmonic Orchestra	2.6009
	2	The Pussycat Dolls	0.001058	Harris Jayaraj	2.5786
	3	A.R. Rahman	0.001023	Lil Wayne	2.5561
	4	Benny Dayal	0.001005	Canadian Brass	2.5528
	5	Harris Jayaraj	0.000940	Eric Robertson	2.5528
	6	Lil Wayne	0.000784	Don Jackson	2.4840
	7	Snoop Dogg	0.000768	David Banner	2.4651
	8	Michael Jackson	0.000736	Yo-Yo Ma	2.4177
	9	E-40	0.000726	Akon	2.3937
	10	Kanye West	0.000633	The Game	2.3844
2009	1	E-40	0.001353	Busta Rhymes	2.5481
	2	T-Pain	0.001211	Major Lazer	2.5164
	3	Lil Wayne	0.001208	Lil Wayne	2.4204
	4	Gucci Mane	0.001193	Harris Jayaraj	2.3639
	5	Kanye West	0.001049	N.O.R.E.	2.2989
	6	Calvin Harris	0.001024	N.A.S.A.	2.2955
	7	Wale	0.000966	Snoop Dogg	2.2943
	8	Busta Rhymes	0.000913	Various Artists	2.2938
	9	David Guetta	0.000841	Irie Ites	2.2878
	10	Flo Rida	0.000811	50 Cent	2.2828
2010	1	Snoop Dogg	0.001990	Shreya Ghoshal	2.8648
	2	T.I.	0.001851	AFROJACK	2.8202
	3	Armin van Buuren	0.001783	A.R. Rahman	2.7925
	4	Lil Wayne	0.001662	Shankar-Ehsaan-Loy	2.7777
	5	Justin Timberlake	0.001345	Shankar Mahadevan	2.7117
	6	Pitbull	0.001122	Rahat Fateh Ali Khan	2.6734
	7	Rick Ross	0.001065	Pitbull	2.6331
	8	Akon	0.001045	Adnan Sami	2.6128
	9	Wisin & Yandel	0.001031	Ludacris	2.5905
	10	Richard Clayderman	0.001012	Los Pericos	2.5799
2011	1	Tony Bennett	0.003315	Snoop Dogg	3.0345
	2	Snoop Dogg	0.002741	T-Pain	2.9458
	3	Lady Gaga	0.002449	Pitbull	2.9415
	4	Andrea Bocelli	0.002146	Ludacris	2.8155
	5	Statik Selektah	0.002070	Kill The Noise	2.6361
	6	DEV	0.001864	Downlink	2.6327
	7	Lecrae	0.001685	Excision	2.5961
	8	Gabriel Fauré	0.001664	Akon	2.5819
	9	Ludacris	0.001597	A.R. Rahman	2.5797
	10	AFROJACK	0.001582	Mohit Chauhan	2.5729
2012	1	Rhythms Del Mundo	0.004613	Various Artists	3.1442
	2	Mumford & Sons	0.004492	Rick Ross	2.8874
	3	Paul Simon	0.004451	Skrillex	2.7877
	4	Traditional	0.004387	Tony Bennett	2.7401
	5	Lecrae	0.003718	Pritam	2.7338
	6	Various Artists	0.003250	Kill The Noise	2.7102
	7	Tiësto	0.003079	Shreya Ghoshal	2.6971
	8	Eminem	0.003058	Foreign Beggars	2.6942
	9	Big K.R.I.T.	0.002926	Tiësto	2.6727
	10	Pitbull	0.002644	Miguel Bosé	2.6690
2013	1	Joshua Bell	0.004274	2 Chainz	2.9326
	2	Straight No Chaser	0.003481	Vishal Dadlani	2.9141
	3	French Montana	0.002986	A.R. Rahman	2.8968
	4	2 Chainz	0.002939	French Montana	2.8662
	5	A.R. Rahman	0.002710	Shreya Ghoshal	2.8624
	6	CeeLo Green	0.002681	Major Lazer	2.8043
	7	Kendrick Lamar	0.002589	Shalmali Kholgade	2.8017
	8	The Piano Guys	0.002469	Vishal-Shekhar	2.8001
	9	Dido	0.002462	Dido	2.7313
	10	Mariah Carey	0.002256	Shankar-Ehsaan-Loy	2.7287

Year	Rank	Artist	Betweenness Centrality	Artist	Entropy
2014	1	2 Chainz	0.006221	Snoop Dogg	3.0186
	2	Smokey Robinson	0.004518	Skrillex	2.9265
	3	Pitbull	0.004088	Sunidhi Chauhan	2.9244
	4	Lecrae	0.004080	Wizkid	2.8687
	5	Chris Brown	0.003935	A.R. Rahman	2.8337
	6	Elton John	0.003781	Tyga	2.7822
	7	Diplo	0.003671	Santana	2.7663
	8	Juicy J	0.003491	Zeds Dead	2.7632
	9	Tiësto	0.003486	Shreya Ghoshal	2.7322
	10	Jessie J	0.003398	Vishal-Shekhar	2.7129
2015	1	Ty Dolla $ign	0.005973	Ty Dolla $ign	3.2917
	2	French Montana	0.003619	Pusha T	3.0337
	3	Paul McCartney	0.003058	Diplo	2.9863
	4	John Lennon	0.003035	Snoop Dogg	2.9355
	5	Pentatonix	0.003002	Wisin	2.9135
	6	Penn Masala	0.002931	Major Lazer	2.8791
	7	Juicy J	0.002820	Yellow Claw	2.8069
	8	Jonita Gandhi	0.002757	Bright Lights	2.7644
	9	Traditional	0.002707	KK	2.7542
	10	Yellow Claw	0.002605	Good Times Ahead	2.7491
2016	1	Ty Dolla $ign	0.006116	Major Lazer	3.0501
	2	Vijay & Sofia Zlatko	0.004217	Ty Dolla $ign	2.9070
	3	Otter Berry	0.003738	Yellow Claw	2.9042
	4	G-Eazy	0.003636	The Chainsmokers	2.8648
	5	Vijay	0.003620	Dillon Francis	2.8643
	6	Steve Aoki	0.003577	Jonita Gandhi	2.8095
	7	DJ Snake	0.003355	French Montana	2.7811
	8	French Montana	0.003277	Justin Bieber	2.7696
	9	Cheat Codes	0.003260	Pitbull	2.7559
	10	Tritonal	0.002867	Shanahan	2.7437

Table 3: Top 10 artists ranked by betweenness centrality and entropy.

8.3 2-Hop Neighborhood

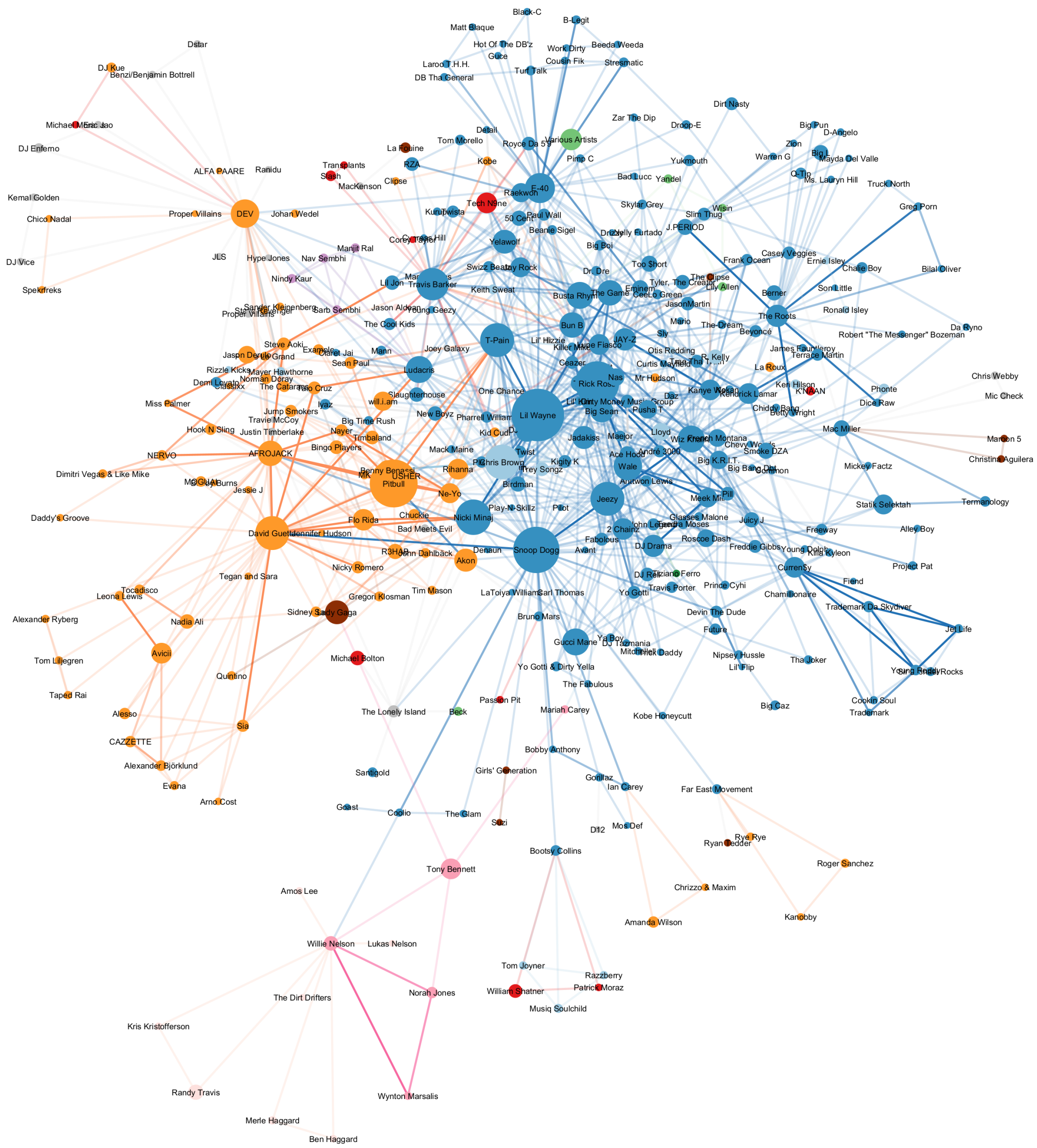

Figure 11: Two-hop collaboration network centered on Snoop Dogg in 2011. Node sizes reflect degree; edge thickness represents collaboration weight; node colors indicate genre.