

Topological properties of music collaboration networks: The case of Jazz and Hip Hop

Lukas Gienapp <lukas_dot_gienapp_at_uni-leipzig_dot_de>, Leipzig University

Clara Kruckenberg <clarakruckenberg_at_gmail_dot_com>, Leipzig University

Manuel Burghardt <burghardt_at_informatik_dot_uni-leipzig_dot_de>, Leipzig University

Abstract

Studying collaboration in music is a prominent area of research in fields such as cultural studies, history, and musicology. For scholars interested in studying collaboration, network analysis has proven to be a viable methodological approach. Yet, a challenge is that heterogeneous data makes it difficult to study collaboration networks across music genres, which means that there are almost only studies on individual genres. To solve this problem, we propose a generalizable approach to studying the topological properties of music collaboration networks within and between genres that relies on data from the freely available Discogs database. To illustrate the approach, we provide a comparison of the genres Jazz and Hip Hop.

1. Introduction: Collaboration in the Music Scene

The enormous increase in machine-readable data has been one of the main drivers of the digital humanities in recent years. This new data enables new, computer-aided approaches [Jockers 2013, 4], both from an exploratory and an empirical perspective. So far, the focus of digitization has been mainly on textual sources. Still, as more and more musical data is becoming available in digital form, we are witnessing an increase in music information retrieval applications [Burgoyne et al. 2016] within the digital humanities. Work in this area is rather heterogeneous, but may be divided into two main branches: (1) 'sound and audio studies', where music is processed and analyzed as an actual audio signal, and (2) 'symbolic music studies', where music is analyzed from a more formalized perspective, as it is written down in some kind of notation system. Typically, the latter branch is dedicated to questions such as digitization of analog music scores (optical music recognition), music representation formats (MusicXML, MEI, etc.), and the quantitative analysis of musical features (e.g., melodic similarity). As symbolic music collections are widely available, this branch of music studies seems to be particularly well suited for quantitative analyses. One specific subcategory of symbolic music is music metadata, which comes in many shapes and sizes, for instance, in the form of audio features (danceability, energy, etc.) via the Spotify API^[1] or as metadata about artists, genres, etc. (Last.fm).

In this work, we propose the use of the Discogs database to study patterns and strategies of music collaboration. The larger theme that motivates this research is the importance of collaborative contexts for problem-solving, creativity, and innovation in general [Mertl et al. 2008], the latter two of which are particularly important in the field of music production. Teitelbaum et al. [Teitelbaum et al. 2008] note that "music is one of the richest sources of interaction between individuals", i.e., music collaboration between artists is a common social phenomenon. However, the concrete strategies of collaboration, such as the role of particular artists as hubs and the formation of communities, may vary widely between genres. This is illustrated by a large body of existing studies on music collaboration across different genres, including *Classical Music* [Mertl et al. 2008] [Park et al. 2015] *Pop Music* [de Lima e Silva et al. 2004] [Park et al. 2007] *Heavy Metal* [Makkonen 2017], *Rap/Hip Hop* [Araújo et al. 2017] [de Lima e Silva et al. 2004] [Smith 2006] and, above all, *Jazz* [Filipova et al. 2012] [Giaquinto et al. 2007] [Gleiser and Danon 2003] [Hannibal 2015] [Macdonald and Wilson 2006] [Patuelli et al. 2011] [Seddon 2005].

Typical research questions that are investigated in the above studies are, for instance, the analysis of structural properties of collaboration networks (e.g., small worlds) [Park et al. 2007] [Teitelbaum et al. 2008] or the study of local scenes / regional communities in the respective genres [Hammou 2014] [Makkonen 2017]. The methods used in these studies are very diverse, as they include qualitative approaches (i.e., interviews) as well as quantitative approaches. When it comes to quantitative approaches, the predominant method is network analysis, which is well known in many related fields of application. Examples include analyzing collaboration between Hollywood actors [Amaral et al. 2000] [Zhang et al. 2006], scientists [Newman 2001], or business alliances [Schilling and Phelps 2007]. Within the sub-genre of network analyses of music collaboration, we also find a multitude of different datasets used in existing studies. However, most of them only collect small amounts of data, are limited to a single timeframe, genre, or geographic area, and work with diverse understandings of the concept of "collaboration". These differences in methods and data make it very difficult to investigate collaboration strategies between different genres of music on a wider range. With a few exceptions [Seddon 2005], there are hardly any studies that compare different genres in terms of their cooperation strategies and networks. Yet comparative studies are particularly exciting, as they allow scholars to work out specific topological patterns of different genres and discuss them in relation to artistic characteristics, especially with regard to creativity and innovation.

4

For this reason, we propose a reusable digital humanities approach for studying topological properties of music collaboration networks within and between genres that may pave the way for further, more comprehensive comparative studies in the field of music cooperation. Since we use metadata from the Discogs database, we have to limit our data set to collaborations between artists that are explicitly documented by a joint release. In this work, we also perform an actual comparison of two genres to illustrate our proposed method. We picked Jazz and Hip Hop for two reasons. The first is the existing scholarship about the two genres, which provides important context for interpreting the results of network analysis. The second is that the two genres are known for sharing a similar approach to collaboration, for single artists are often collaborating in varying constellations. Yet, their collaboration networks are rarely compared. Not only is network analysis a particularly amenable method to studying the style of collaboration that are prominent in Jazz and Hip Hop communities, but the method also provides a way to compare the two genres in new ways.

2. Collaboration in Jazz and Hip Hop

5

To back up our comparison on collaboration in Jazz and Hip Hop, this section introduces some basic characteristics of both genres and also provides an overview of some of the most prominent related work in this area.

2.1 Jazz

6

Studying collaborations in the genre of Jazz seems to be the most prominent stretch of research in existing work on music collaboration. This can mainly be attributed to the extensive amount of collaboration within this genre. Unlike in other genres, such as Rock, where fixed groups are the most common form of collaboration, Jazz musicians usually play and record in constellations. Real-time Jazz performances themselves are highly interactive and collaborative, as musicians typically react to each other and improvise while sticking to a common framework with regard to tempo and harmonics [Macdonald and Wilson 2006]. This fluid nature makes Jazz a prime research interest for social network analysis in the music domain, as the resulting networks tend to be rather complex. In the existing related work on Jazz collaborations, researchers use very different types of data, ranging from archival and biographical sources to existing collections of metadata on the genre.

7

One notable dataset on Jazz collaborations is called *Linked Jazz*. In an attempt to create an in-depth network dataset, Patuelli et al. [Patuelli et al. 2011] construct an ontology from digital archive material on Jazz history, more specifically transcripts of interviews with 12 Jazz musicians. Their resulting network includes 952 connections between those interviewees and 529 other Jazz artists. In more recent versions of their dataset, over 2,000 Jazz musicians with over 3,600 connections are included. While personal recollections of Jazz musicians are probably the most accurate source of data when it comes to making connections that actually represent the real world, this approach has two major disadvantages. First, it is difficult to make generalizable statements, since only a very limited number of musicians are considered. Although such data sets may be interesting from a historical perspective, they are not suitable as a basis for

empirical statements on the structure of collaboration in Jazz. Second, the actual data acquisition entails a lot of work, as the qualitative data source is not readily available in a structured format.

The *Red Hot Jazz Archive*, as used by Gleiser and Danon [Gleiser and Danon 2003], is another example of existing resources on collaborations in Jazz. It contains information about Jazz bands between 1895 to 1929, listing the members and discographies for 1,099 bands. It also provides more detailed biographical information for 192 musicians. The data was compiled manually from a variety of sources, such as oral histories, biographies, and several historical books on the Jazz culture of that era. The *Red Hot Jazz Archive* was created in a similar way to the *Linked Jazz* database, which means it also shares some of the previously described problems, namely the lack of generalizability and the time-consuming preparation of the dataset. The authors use a two-stage process to examine Jazz collaborations based on this archive. First, they infer a network of individual artists, which are connected if they played in the same band. From there on, they create a network of bands, with two bands being connected if they share at least one member. Studying both networks topologically with respect to degree distributions, clustering, and average nearest neighbor degree, they note two major limitations of their network model: artists may be credited with different spellings or under pseudonyms, and relations between artists are not time-stamped, which makes the analysis of the evolution of networks over time impossible.

In addition to the use of historical documents, another popular approach for the generation of collaboration networks is to use metadata collections of music releases. Hannibal [Hannibal 2015] provides an extensive overview of the evolution of Jazz as a genre and the collaboration patterns associated with it based on a network obtained from *Tom Lord's The Jazz Discography*. While this dataset is deemed comprehensive (Phillips and Kim, 2009), as it contains data from 410,000 Jazz releases, it is not license-free and only available at a high price. In the corresponding network, musicians who recorded an album together are connected via an undirected edge. Each edge is tagged with the year of recording. This method of network creation requires very little manual work and yet results in comprehensive and metadata-rich networks. Hannibal [Hannibal 2015] uses the network to explore career success and historical significance of musicians in relation to their position in the network by using different metrics such as centrality, brokerage, and closure, which he relates to measures of career success (e.g., awards and sales). He presents empirical evidence that Jazz musicians who perform in closed groups are less likely to have a successful career than those who maintain a more open structure of collaborative connections.

In another related research paper, Giaquinto et al. [Giaquinto et al. 2007] use the *AllMusic* database to infer a network that connects Jazz musicians according to their similarity as determined by the database, to study which artists were most influential. However, their network size is far from comprehensive, including only 418 artists at most. Finally, Filipova et al. [Filipova et al. 2012] provide a web service called "Map of Jazz", which can be used for an explorative search for Jazz musicians based on collaboration network data.^[2] They aim to provide a tool to make discographies more accessible for researchers, as large-scale discography data can be difficult to analyze in the classic textual form. Their network is based on discographical data that was manually compiled from a multitude of sources, ranging from musical records and magazines to biographical data and monographs. Overall, 11,824 musicians are included. A major difference to other works is that their network is dynamic, which includes metadata about the time and place of connections. However, their work is mostly concerned with tool development and provides little descriptive insight into the network's properties.

2.2 HipHop

Hip Hop is another promising genre when it comes to collaborations. In her book on Hip Hop culture and the role of collaboration, Smith [Smith 2016] notes that "[f]rom its outset in the 1970s, hip-hop was not solely a musical genre, but encompassed dance (b-boying/b-girling) and visual art (graffiti)". In other words, Hip Hop was and is a cultural phenomenon with many facets, of which music is only one aspect. According to Smith [Smith 2016], collaboration in Hip Hop is an essential mechanism for interaction and communication of shared ideologies, attitudes, and behaviors within the scene. Compared to Jazz, there is substantially less work that deals with music collaborations in Hip Hop.

The most comprehensive study of collaboration in Hip Hop is provided by Smith [Smith 2006], who presents an

interesting approach by examining the metadata of over 30,000 songs extracted from lyrics databases. Since all artists who have had a vocal part in a song are listed in the metadata of the songs, a connection can be established between any artists that appear together on a song. Similar to Gleiser and Danon [Gleiser and Danon 2003], name standardization is also a problem in this study. Smith uses a fuzzy search algorithm to partly match different spellings of names. However, if the same artist is credited under a different pseudonym altogether, the standardization fails. Another related work examines the collaboration network of a single artist [Araújo et al. 2017]. More concretely, the authors try to model the evolution of *DJ Khaled* with respect to collaborations by building an incremental network for each of his albums. This network is then used to identify communities as well as influential artists who affected the music of DJ Khaled the most.

2.3 Interim conclusion

The related work presented in this section illustrates the multitude of different data sources that are used for studying music collaboration. These are typically accompanied by a broad range of methods for inferring network structure from the data. Several limitations arise as a result of this: (1) most datasets are only available via inference from noisy data, with associated problems in data cleaning; (2) while comprehensive network and collaboration datasets exist for Jazz, sources for other genres, for instance, Hip Hop, are rather scarce; (3) there is no common method for establishing links between musicians that is applicable to all genres and analytical settings.

13

While datasets used to study music collaboration are very diverse, we found only little variation when it comes to network metrics in the related work studies: the most prominent metrics are degree distribution, degree correlations, average nearest neighbor degree, betweenness centrality, clustering coefficient, and transitivity score. The Girvan-Newman partitioning algorithm seems to be a popular choice to identify (sub-)communities in the networks. As all of the above-mentioned papers provide different results for these metrics, the results cannot be directly compared with each other in order to gain insight about structural differences in collaboration between genres, since the underlying networks are not derived in the same way. Therefore, we will now propose a method for how to address these challenges and hope they open up a path for cross-genre network analysis.

14

3. Methodology

In this section, we present our approach for the generic study of music collaborations across different genres. First, we explain our considerations when modeling the network graph. Next, we give an overview of the network metrics used and a brief explanation of how they can be interpreted for the analysis of music collaborations. Finally, we introduce the underlying Discogs dataset.

15

3.1 Network Modeling

When building a network, it is essential to define what constitutes an entity (node) and what constitutes a connection (edge) between these nodes. The challenge is to find a definition that, on the one hand, is applicable to a variety of research questions and, on the other hand, correctly reflects the underlying connections in the real world. In order to establish a common methodological approach to network structure, the dataset created as part of this work is concerned with connections between individual musicians. This is common in the field [Hannibal 2015] [Park et al. 2007] [Smith 2006], as the individual artist is the "least common denominator" across genres in terms of organizational structure. While Big Band Jazz can have group sizes of up to 170 musicians (e.g., *Paul Whitman and his Orchestra*), Hip Hop, for instance, is much more centered around the individual artist (e.g., *Snoop Dogg*), and larger groupings of artists are the exception (e.g., *The Wu-Tang Clan*^[3]), not the norm.

16

There are several options available to establish connections between nodes, all different with respect to their weighting, direction, and sparseness. When studying collaboration, the edges are supposed to represent a collaborative process shared between the two connected nodes. With music records being the main available data source to infer these connections, one choice is to establish edges between all artists that appear together on a record. The main advantage of a release-centered approach over alternative methods – such as the definition of an edge between two artists being if

17

they are part of the same band or collective [Gleiser and Danon 2003] – is the possibility to assign different weights to the connections. We may well assume that artists who recorded more tracks together have a strong creative influence on each other. Therefore, the network edges are weighted by the number of collaborations.

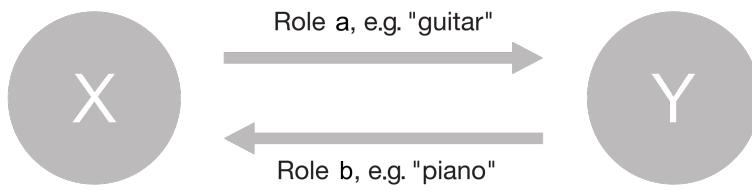
However, an apparent problem here is that a music release is usually a collection of different tracks, and these tracks are not always recorded with the same set of musicians. Therefore, if only album-wide credits are available, not all of the listed musicians can automatically be assumed to have shared a collaborative process, violating the hypothesized definition of edges. The only person that is connected to all other credited artists with a high probability is the main artists of the release. This problem can be tackled in two different ways: (1) create connections between all credited artists, on at the cost of possibly introducing unwarranted ones; or (2) only create connections between credited artists and the main artist of the album, thus only creating warranted connections at the cost of missing some. Several factors should be considered in this decision: first, it can be theorized that the amount of missed connections in the latter case is higher than the number of unwarranted ones in the former since, for a high proportion of releases, a big subset of credited artists appeared on all tracks of an album. Accordingly, the data/noise trade-off seems favorable when adding all possible connections to the network.

18

Another relevant factor is the metrics that will be used to derive conclusions from the network data, especially metrics measuring network centrality. The first option would place higher network importance on primary artists, as they would receive a disproportionately high amount of connections. Also, this approach invalidates metrics that operate on notions of transitivity, such as the clustering coefficient, since the transitive edges are explicitly excluded from the dataset if the credited artists are not interconnected. Therefore, we opt to include all possible connections between the artists associated with a release. This is also the most common approach [Gleiser and Danon 2003] [Hannibal 2015] [Smith 2006]. Since no information can be derived about the direction of collaborative influence from simple co-appearance on tracks, the network can be assumed to be undirected. However, to incorporate relevant information about the concrete role each musician has played on a record (i.e., instruments, vocals, producing roles, etc.), we relied on two parallel directed edges (see Figure 1). These connections can be interpreted as “ x played with y in role a ” and as a parallel edge “ y played with x in role b ”, where x and y are two connected musicians, and a and b are instruments or other musical roles. This information about the specific roles of musicians on a record can later be used to filter the data and create tailored sub-networks that – for instance – might only include edges representing piano players or producers. Besides the option to select specific nodes in the data set via their role annotations, we treat these directed parallel edges as a single undirected edge when we apply our network metrics (see next section) for the analysis of music collaborations.

19

Parallel directed network model



Undirected network model



Figure 1. The two corresponding network models. The parallel directed model is used to annotate metadata in the network, with two edges representing a single connection in the undirected model.

As for the metadata about the connections, various additional collaboration information can be provided. To allow for the creation of dynamic networks and insights into historical questions, each edge is dated with the release year of the record. This has a drawback, as noted by Hannibal [Hannibal 2015], who time-stamped the collaboration with the year of the recording session, not the year of the album release: individual songs may have been released publicly years after the recording took place. However, detailed recording session data may not be publicly available for genres apart from Jazz. In the interest of deriving a common network modeling method across multiple genres, we decided to use the album release date. 20

3.2 Metrics

In the following section, the metrics used for network analysis are described. For each metric, we provide a short description of how the metric is calculated and what insight can be derived from it regarding the collaboration networks at hand. For a more formal and in-depth description of common network metrics, see, for example, Newman [Newman 2010]. 21

3.2.1 Degree

The degree of a node is the number of incoming and outgoing edges a node has. The neighborhood of a node is thus given by all other nodes it is connected to. Applied to the undirected music collaboration networks at hand, the degree is the number of other musicians an individual has collaborated with, and the neighborhood is the set of all those musicians. The average degree and the overall distribution of degrees gives insight into the collaboration frequencies of musicians in that genre. In addition, this work looks at the degree variance and degree distribution in order to see whether the network mostly revolves around highly connected artists or if the whole network tends to be closely connected. 22

3.2.2 Power-law scaling exponent

For numerous types of networks, it can be observed that a small number of nodes have a large number of connections to other nodes, i.e., a high degree, while the majority of nodes have a rather low degree. Networks that exhibit this property are commonly referred to as scale-free networks. In these networks, the number of nodes that have a certain number of links is related to the degree of these nodes, such that the degree distribution of the network is a power-law distribution with a scaling exponent α . Typically, α is in the range of $2 \leq \alpha \leq 3$, although in some cases, values outside 23

this range can be observed [Newman 2010].

Scale-free characteristics can be observed for a wide range of real-world networks, especially social and collaborative networks. It has been hypothesized that they form due to two major properties: growth and preferential attachment. Since these networks grow naturally, nodes that are introduced at an earlier stage have a higher chance of developing connections, hence boosting their degree. Similarly, preferential attachment means that establishing links reinforce a positive bias toward the connected nodes, as a node of high degree (also called "hub") seems to be more attractive to establish connections with, in turn, attracting more future connections [Barabási et al. 2003]. In the collaborative network at hand, these central hubs signify highly influential musicians. By investigating whether the degree distribution follows a power law, we can apply these characteristics to music collaboration networks, allowing for hypotheses to be derived about the origins of success for musicians.

24

3.2.3 Degree correlation

The degree correlation is defined as the Pearson correlation coefficient ρ of the degrees of nodes connected by an edge [Girvan and Newman 2002], specifying if edges tend to be formed between nodes of similar or dissimilar degree. This parameter provides insight into the networks connectivity structure by specifying whether the network is assortative (nodes mainly connect to nodes with a similar degree, $\rho > 0$), disassortative (there is a systematic disparity in degree between connected nodes, $\rho < 0$), or neutral (connections are random, $\rho \approx 0$). In the networks at hand, assortative would mean that musicians with lots of connections to other musicians tend to collaborate with individuals who exhibit the same characteristic. Disassortative signifies that those well-connected musicians tend to collaborate more with less-connected musicians. Neutral would mean that musicians of high and low degrees frequently collaborate in an intermixed fashion, with no apparent general trend.

25

3.2.4 Average nearest neighbor degree

The average nearest neighbor degree (ANND) [Barrat et al. 2004] of a node is the average degree of all nodes in its direct neighborhood. In weighted networks, each neighboring nodes' degree can be weighted by the weight of the edge connecting it. This value is computed individually for all nodes of a degree k , and then averaged over all nodes of that degree, a process that is repeated for all values of k present in the network. This results in the value being a function of k and providing additional information about the assortative properties of the network.

26

In assortative networks, the average neighbor degree is increasing for larger values of k , decreasing in disassortative networks, and constant when the network has neutral mixing. While the degree correlation captures the general network trend in a single value for all nodes, the ANND extends the insight into the local structure by further specifying for which nodes the correlation occurs, as the ANND is a function of k . Furthermore, Pearson's degree correlation has been shown to depend on the size of the network, an issue that the average neighbor degree does not face [Yao et al. 2017]. As we deal with large networks, we compute the average neighbor degree in addition to the degree correlation to complement the interpretation of the assortative properties of music collaboration given by the degree correlation.

27

3.2.5 Mean local clustering coefficient

The mean local clustering coefficient [Watts and Strogatz 1998] is a measure of how complete the nodes in a network are connected to each other. This notion of completeness is based on the idea of triangles forming in the neighborhood of a node: when two nodes in the neighborhood are in turn connected to each other, the network is more tightly knit. The local clustering of a node is thus given by the ratio of triangles that exist to all possible triangles in its neighborhood. The mean local clustering coefficient is then defined as the mean of all local clustering coefficients. Applied to the collaborative nature of our network, a higher clustering coefficient would indicate that musicians that share a common collaborator are likely to collaborate with each other too, facilitating the exchange of creative ideas throughout the genre.

28

3.2.6 Transitivity score

A network's transitivity score, also called global clustering coefficient, is another metric describing the connectedness of

29

the network [Newman 2010]. In contrast to the local clustering coefficient, it characterizes the whole network by applying the notion of complete triangles vs. possible triangles to all edges in the graph on a global, instead of a local level. Similar to the mean local clustering coefficient, this metric provides insight into the transitive properties of the network. In our case, such a transitive relation can be expressed as "if musician A recorded together with musician B, and musician B with musician C, chances are high that A and C also recorded music together". The transitivity scores in social networks tend to be rather high [Newman 2010], especially in collaborative networks: for music, in particular, Smith gives a value of $C = 0.18$; in other applications such as a network of film actor collaborations, we have similar values: $C = 0.20$ [Newman 2003]. For our formalization of a network, these scores (and also the local clustering coefficients) are expected to be even higher. Since we add a fully connected subgraph for each music record (which in itself has a transitivity score of 1), and if the number of artists on a given record is high, the transitivity score of the whole network is boosted in turn.

3.3 Data

The Discogs database is the largest openly accessible collection of crowdsourced metadata of music releases [Bogdanov and Serra 2017], yet it is less commonly used for research purposes. Discogs currently lists over 12 million music releases from 6.7 million artists on nearly 1.5 million music labels. Metadata for existing and new releases is continuously entered, checked, and curated by over 500,000 registered users. A voting system is used to rate the accuracy and completeness of data.^[4] For each item, a vast variety of metadata can be specified, ranging from standard fields such as title, label, genre, and artists to fine-grained information, such as contributing musicians by track, issued versions of a master release or trivia. This allows researchers to describe virtually any type of audio recording ever released in great detail. For many releases, a list of the contributing musicians collaborating on the release, as well as their information about their collaboration role, is provided.

30

Discogs provides complete dumps of their whole database that are updated on a monthly basis. The data is licensed under CC0, so there are no restrictions placed on the use of the data, which makes it a perfect data resource for research. All data in this project was obtained from a Discogs Data dump from October 9th, 2019.^[5] Based on this data, we created a collaboration network by defining an edge between the artist and each collaborator if they appear together on the same release. This process uses the Discogs ID of each person associated with the release, so even if an artist uses a pseudonym, the correct link is created in the network. Only releases which have exactly one genre listed in their metadata were considered for the network. In this way, only intra-genre collaborations are modeled. While there is an immense amount of inter-genre collaboration, the main objective of this study was to search for structural differences within genres.

31

For every entry in the database, Discogs readily provides a label denoting its data quality. We only included releases that have the status *accepted* (indicating duplicates, etc.) and a data quality higher or equal to *correct*, to ensure a high quality in our derived network dataset. All relations that have a non-musical role, such as *cover design*, were omitted. Following the role categories defined by Discogs^[6], only roles from the categories *vocals*, *instruments*, *dj mix*, *remix*, and *beats* were deemed as valid, as these categories include the actual musicians involved in creating the music on the release. Finally, metadata about nodes (artists) and edges (collaborations) was added. Nodes with the Discogs IDs 194 and 355 with all associated edges were deleted manually, as these IDs were only database placeholders and did not link to any artist. For each node, the Discogs artist ID and the name were included. For each edge, the year, role, and Discogs release ID were included. All data was parsed using Python, and the networks were exported to a *GraphML* file [Brandes et al. 2011]. Network metrics were calculated using the *networkx* package for Python [Hagberg et al. 2008]. In addition, the Python package *power-law* was used to calculate the power-law exponent [Alstott et al. 2014]. Finally, *Gephi* [Bastian et al. 2009] was used for graph visualizations.

32

4. Results and Discussion

4.1 Dataset

Table 1 provides an overview of the network details that were derived for the genres Jazz and Hip Hop, using the

33

previously described method on Discogs metadata. We only include releases with the highest tiers of the data quality label provided by Discogs, which indicates correctness and completeness based on a crowdsourced voting system. Thus, we ensure that only verified information is considered. Also, Discogs counts each issue and repress of the same album as a separate release, but provides a meta grouping in the form of a so-called master release, grouping together all the versions of an album. Since the metadata (at least the data relevant to our research) is consistently the same across all versions of such a master release, we based our data extraction on those. These two facts combined account for us using only a seemingly low number of releases, given the immense amount of available data.

Using our proposed method, we derive the most comprehensive, freely accessible network dataset on music collaboration to date, both in terms of spanned timeframe and included artists. Before carrying out analyses and calculating metrics, each network was converted to a single-edge, undirected graph by combining all the parallel edges between two nodes into one weighted edge, with its weight being the sum of merged edges. If six parallel directed edges exist as a result of two musicians collaborating on three records, those are replaced by a single undirected edge of weight six.

34

	Artists	Relations	Parsed releases	Timespan
Hip Hop	23098	338200	12301	1979 - 2019
Jazz	40296	3488662	21808	1917 - 2019

Table 1. Network characteristics for Jazz and Hip Hop as extracted from Discogs metadata

4.2 Jazz Collaboration

Figure 2 summarizes the most important descriptive network statistics for our Jazz network. To put those numbers into context, we also provide the results of the earlier cited study of Gleiser and Danon [Gleiser and Danon 2003], providing comparable statistics for their network analysis. As Gleiser and Danon's data is based on biographical information from 192 musicians, their network is much smaller. The newly generated Discogs network is over thirty times larger in terms of recorded musicians and includes about fourteen times more connections. The data spans nearly 100 years of Jazz history, providing much more comprehensive insight into the historical development of the genre.

35

	n	m	\bar{d}	\bar{c}	C	p	α
Discogs Jazz network	40294	569562	28.27	0.80	0.32	0.26	1.53
Linked Jazz network [Gleiser and Danon 2003]	1275	38326	60.4	0.89	-	0.05	-

Table 2. Comparison of descriptive statistics of different our Discogs Jazz network and the existing Linked Jazz network by [Gleiser and Danon 2003]. n – number of nodes in the network, m – number of undirected edges, \bar{d} – average degree per node, \bar{c} – mean local clustering coefficient, C – transitivity score, p – degree correlation coefficient, α – power-law scaling exponent

As shown in Figure 2, the Discogs Network, with an average degree of 28.27 connections per musician, is considerably denser compared to the Linked Jazz network. This disparity could be explained by the fact that Gleiser and Danon's network is generated from biographical information from the years 1895 to 1929. During this time, Jazz is characterized by larger collectives of musicians, as it coincides with the musical styles of New Orleans Jazz, Dixieland Jazz, and Chicago Jazz. In modern Jazz, a larger variety of group sizes is found, with an inclination towards solo artists and trios [Gioia 1998]. Consequently, a node in the network of Gleiser and Danon is more likely to have more connections, as the band size at the time was comparatively large, introducing a higher number of connections. Deriving data from oral recollections further facilitates that the network consists of one large coherent component. If, in contrast, the network was based on music releases, smaller clusters only representing one album would appear that are not connected to the main network in any way. As a matter of fact, the Discogs Jazz network spans 1,100 separate components, ranging from 1 to 35,128 musicians in size. 75% of these components have a maximum of 6 nodes. Yet, 87% of musicians are included in the largest component.

36

This is corroborated by the degree distribution in Figure 2, which also shows that the Discogs Network includes a large number of musicians with only a few direct connections to other musicians, while nodes with a higher number of connections are increasingly rare. In fact, over 75% of nodes have a degree smaller than the mean. The resulting distribution therefore resembles a power-law distribution.

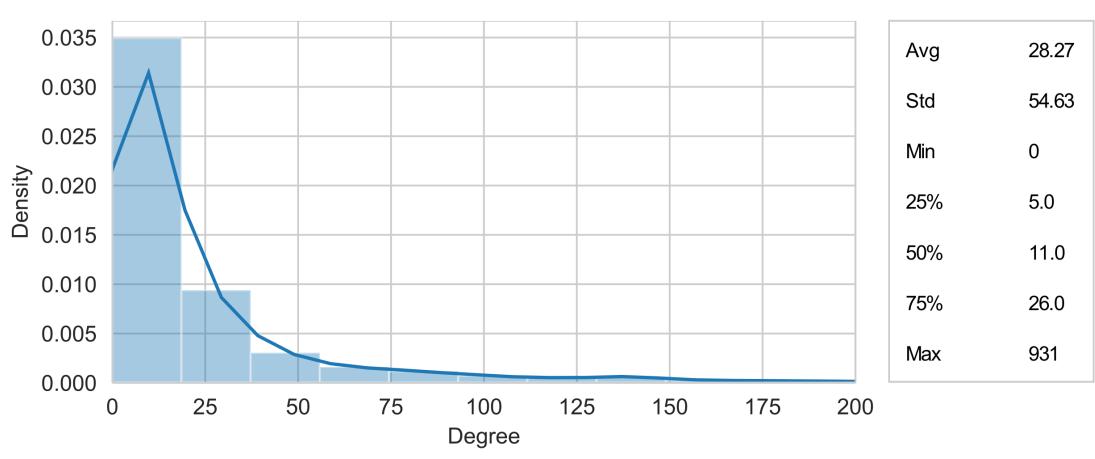


Figure 2. Degree distribution of the Jazz network. Outliers $d > 200$ are omitted from the distribution plot for better legibility.

In fact, with a scaling exponent of $\alpha = 1.53$, the network can be characterized as scale-free. Although the α -value is outside the 2-3 range, which is typical for scale-free networks, such outliers commonly occur and the degree distribution in Figure 2 is very similar to a power-law distribution for degrees greater than approximately 10. Consequently, we can classify the Jazz collaboration network as a scale-free network. Previous work did not calculate a scaling exponent for their networks; hence no comparison can be given in that regard.

Another metric that is influenced by the widely compartmentalized nature of the network is the degree correlation, which once again differs considerably in comparison to existing work. At $\rho = 0.26$, a rather strong degree correlation is present, indicating that musicians tend to collaborate with other artists who have a similar degree. Possible reasons for this correlation effect become apparent in Figure 3, which plots ANND by degree and indicates the general trend with a linear regression.

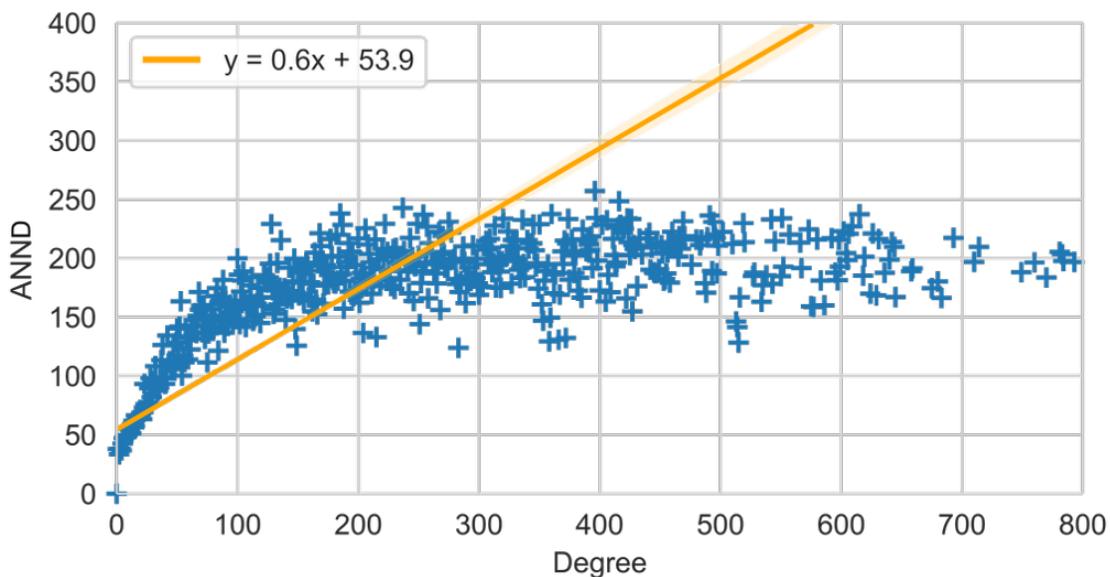


Figure 3. Average nearest neighbor degree (ANND) by node degree in the Jazz Network.

Since the regression line is monotonously rising, the degree correlation effect can be validated. We can further conclude that since 75% of nodes have a degree of 26 or less, most of the correlation effect can be attributed to the mentioned small components coming from only one record or a fixed set of people, as these share the same degree within their component, having little outside connections.

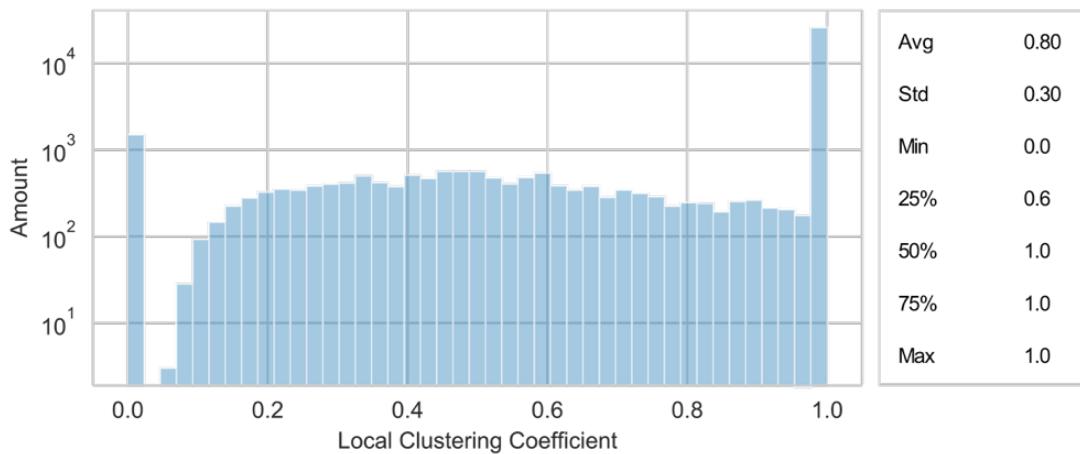


Figure 4. Local clustering coefficient distribution for the Jazz network.

Regarding the overall distribution of the local clustering coefficient of the Discogs network in Figure 4, it can be observed that half of the regarded data have a local clustering coefficient of 1. Another spike occurs at 0, with all other values being represented in a nearly equal fashion. The two spikes at 0 and 1 could once again be due to the high number of smaller components: if a group of musicians has only released one album, or does not collaborate outside their usual setting, all the nodes inside this component have a clustering coefficient of 1. Similarly, solo artists inherently have a clustering coefficient of 0.

In conjunction with the degree distribution and correlation, some assumptions about the overall structure of collaboration in Jazz can be derived. The results regarding degree distribution and correlation allow some assumptions about the overall nature of Jazz collaborations in the regarded network. The new data indicates that in contrast to previous findings, not only one dominant component, which includes almost all of the network, can be found, but smaller, separate groups of artists play a significant role too. A large number of musicians only appear in a fixed set, forming their own network components, bearing no connection to the rest of the artists. However, since most nodes are included in the largest component, the existing interpretation of Jazz as a highly collaborative genre still stands, as expressed by the transitive, closely-knit network structure. As the network is scale-free, it exhibits only a few highly connected hubs. These hubs may be theorized to stand for those artists who have a huge influence on the overall genre.

4.3 Hip Hop Collaboration

Table 3 provides descriptive network statistics for the Discogs Hip Hop network and compares them to the data derived by Smith [Smith 2006]. Again, the networks are generated in different ways. While Smith's network includes only the connections between rappers/singers associated with single song lyrics, the Discogs data provides a much broader range of roles, such as musicians involved in the DJ mix, remix, and beats of an album.

	n	m	\bar{d}	\bar{c}	C	ρ	α
Discogs Hip Hop Network	23099	81700	7.07	0.73	0.63	0.45	1.58
Lyrics Network [Smith 2006]	4433	57972	20.95	0.48	0.18	0.06	3.5

Table 3. Comparison of descriptive statistics of different our Discogs Hip Hop network and the existing lyrics Hip Hop network by [Smith 2006]. n – number of nodes in the network, m – number of undirected edges, \bar{d} – average degree per node, \bar{c} – mean local clustering coefficient, C – transitivity score, ρ – degree correlation coefficient, α – power-law scaling exponent

Here too, the Discogs network is considerably larger in size. There are about five times as many musicians in our network, sharing 1.4 times as many connections. Once again, with 7.07 the average degree is considerably lower in the Discogs data, with an average number of 20.95 neighbors in Smith's network. This could be partly due to the methods used for network inference, as already mentioned for the Jazz networks. Moreover, the granularity of the data could be assumed as a different reason for the Hip Hop genre, since the time frame and general method are similar for both Hip Hop datasets. Smith uses around 30,000 lyrics files to collect the network data, about 4.5 times more releases than were included in our study. A single song, as opposed to a whole record, usually has fewer people involved. Therefore, the number of involved artists per release is much higher in the Discogs data, but the total number of releases used to infer the network is lower. These numbers facilitate a higher ratio of edges to nodes, in turn explaining the divergence in average degrees. 44

In general, the chosen metrics deviate strongly from their counterparts in existing work. Both the average local clustering degree and the global clustering coefficient are much higher in the Discogs network than in Smith's analyses. Following the same line of reasoning as for the Jazz network, some of these observations may be attributed to the overall large number of components included in our network. In total, 3,076 components are present, ranging from 1 to 11,285 in size. Half the network is included in the largest component, while 75% of components consist of four nodes or less. Figure 5 shows the degree distribution for the Hip Hop network, with the majority of nodes showing a rather low degree, quickly decaying for fewer nodes with higher degrees. More concretely, the distribution indicates that three out of four artists have only collaborated with eight people or less. 45

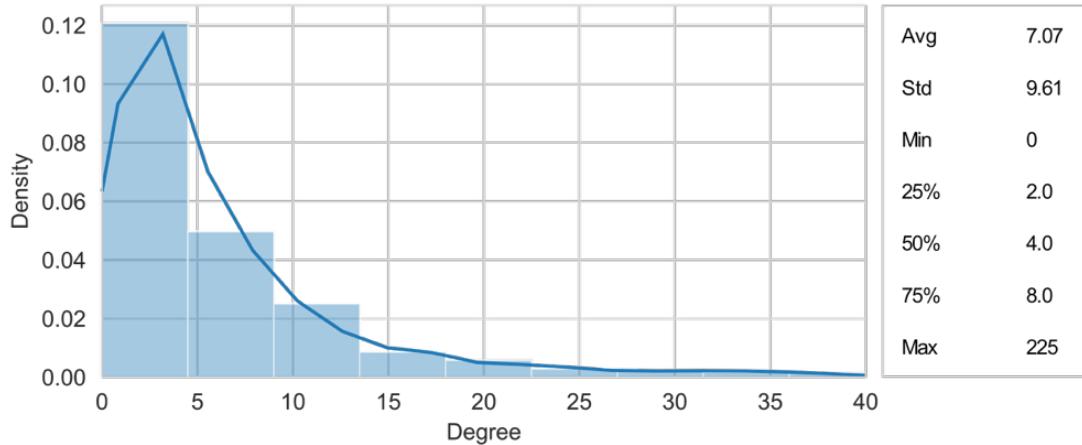


Figure 5. Degree distribution of the Hip Hop network. Outliers $d > 40$ from are omitted from distribution plot for better legibility. 46

As already seen in the Jazz network, the scaling exponent here again α is outside the usual range. In previous work, values of α are reported [Smith 2006]. This difference in our data can mostly be attributed to the network creation approach: as the Discogs network presumably is more sparse with regard to connections (as indicated by the lower average degree), the decay of the power-law function is also smaller. Ultimately, both values lead to the same conclusion, which is to classify the collaboration structure of Hip Hop as being scale-free. The high number of disconnected components in the network, in conjunction with the diameter of these local sub-networks being small,

gives a possible explanation for the high degree correlation of $p = 0.45$.

Figure 6 shows the ANND plotted by node degree, supporting the notion of a high correlation between degrees. The overall trend, as indicated by the linear regression line, is rising, with the effect being most pronounced in lower degrees. This implies that artists with a low degree tend to work together with other artists of low degree, while this trend is much more diverse in higher degrees, where no clear pattern is apparent.

47

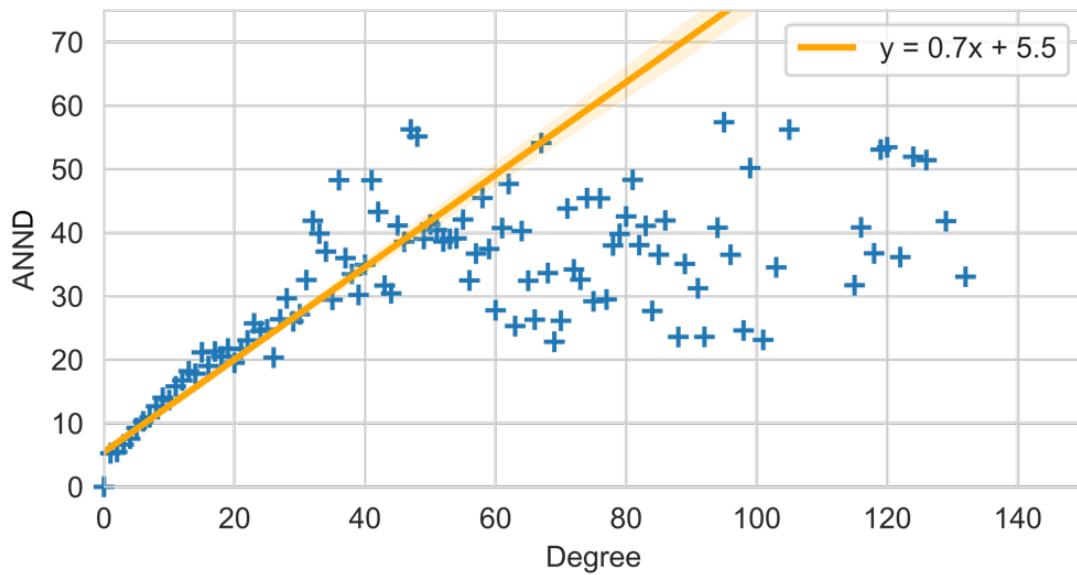


Figure 6. Average nearest neighbor degree (ANND) by node degree in the Hip Hop Network.

With regard to the clustering properties of the network, we can observe a systematic difference in previous findings. Both the average local clustering coefficient and the global clustering coefficient are much higher in the Discogs network than in Smith's analyses.

48

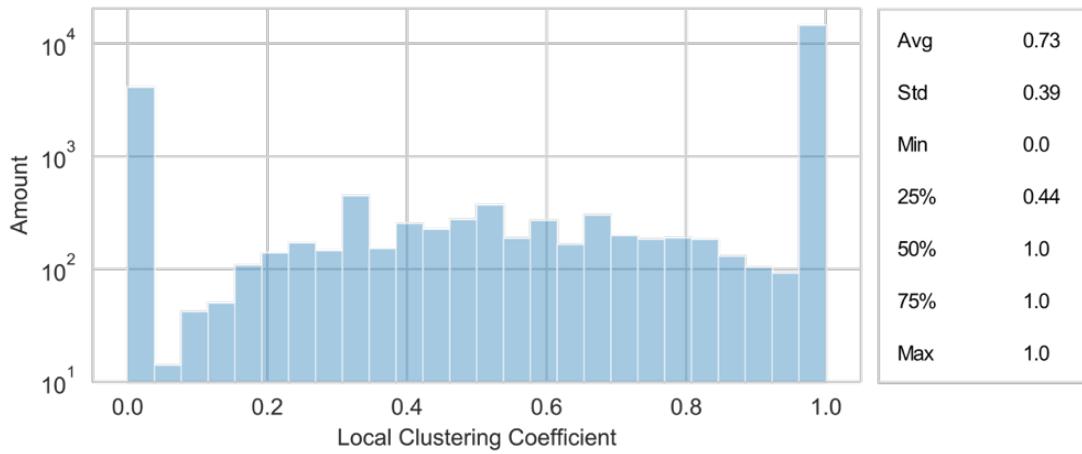


Figure 7. Local clustering coefficient distribution for the Hip Hop network.

The two spikes at 0 and 1 in the distribution of the local clustering coefficient shown in Figure 7 can be attributed to the high number of small components and some solo artists that are being included. The values in between are somewhat evenly distributed throughout the whole range, with the upper range being slightly more represented.

49

Regarding the overall structure of collaboration in Hip Hop, our data indicates that solo artists, as well as small and consistent groups of artists, play a bigger role than previously assumed since a significant number of musicians only

50

appears in a fixed set with no connection to the rest of the artists. Still, a large portion of the community shares highly collaborative properties, forming many connections between different subnetworks. Hubs are, therefore, highly relevant when it comes to brokering between more tightly knit local groupings. This observation is characteristic of scale-free networks, which can be assumed for the case of Hip Hop collaboration.

4.4 Comparison of Genres

Jazz is typically characterized as a highly dynamic and versatile genre. Especially the Jazz community of the 1940s and 1950s is shaped by constant upheavals and change, not only concerning the musical style itself but above all, the constellations of musicians in this time seemed to be in constant flux. Unlike in current Pop or Rock bands, it is rather atypical in Jazz to play in the same constellation of musicians for a longer period of time. Instead, many musicians belong to several formations simultaneously [Hinton et al. 1988, 244 ff.]. This tendency is well reflected in the collaboration networks. In comparison to the Discogs Hip Hop network, the average Jazz musician has more than four times more connections than a Hip Hop artist. The data shows that collaborations in Jazz are more frequent and involve more people than collaboration in Hip Hop. The Hip Hop community is more compartmentalized than the Jazz community, with only 50% of the artists being included in the main component, and 75% in Jazz, respectively.

After all, we can observe that Jazz musicians in the course of their career tend to collaborate with more different musicians than hip hop artists do. A similar picture emerges with regard to the clustering coefficient: While the Discogs Hip Hop network exhibits a total global clustering coefficient of 0.63, the Jazz network is only half as transitive. Thus, while the Jazz network as a whole seems to be highly connected, in Hip Hop, we see a tendency towards the development of smaller, though very closely connected groups. The degree correlation also shows that there seems to be a higher tendency of two actors with similar degrees to work together on a project in the Hip Hop network, which appears to be less the case in Jazz.

In Figures 8-11, we show partial visualizations of both the Jazz and the Hip Hop networks. A complete visualization of the two networks is not possible due to their size. Instead, we opted to show the neighborhood of important hubs in each network. For Hip Hop, we chose *Tupac Shakur (2Pac)* and *Biggie Smalls (Notorious B.I.G.)*, two of the most influential Hip Hop artists of all time. For Jazz, we show the neighborhood of *Louis Armstrong* and *Ella Fitzgerald*. The visualizations illustrate the differences described before: a musician in Jazz is connected with many more other artists than is the case in Hip Hop, Jazz is less transitive, and the other nodes a hub connects to vary more in degree. Also, in Jazz, most nodes are in one large interconnected component, with smaller groups in the periphery, while in Hip Hop, we find smaller, tightly knit clusters.

51

52

53

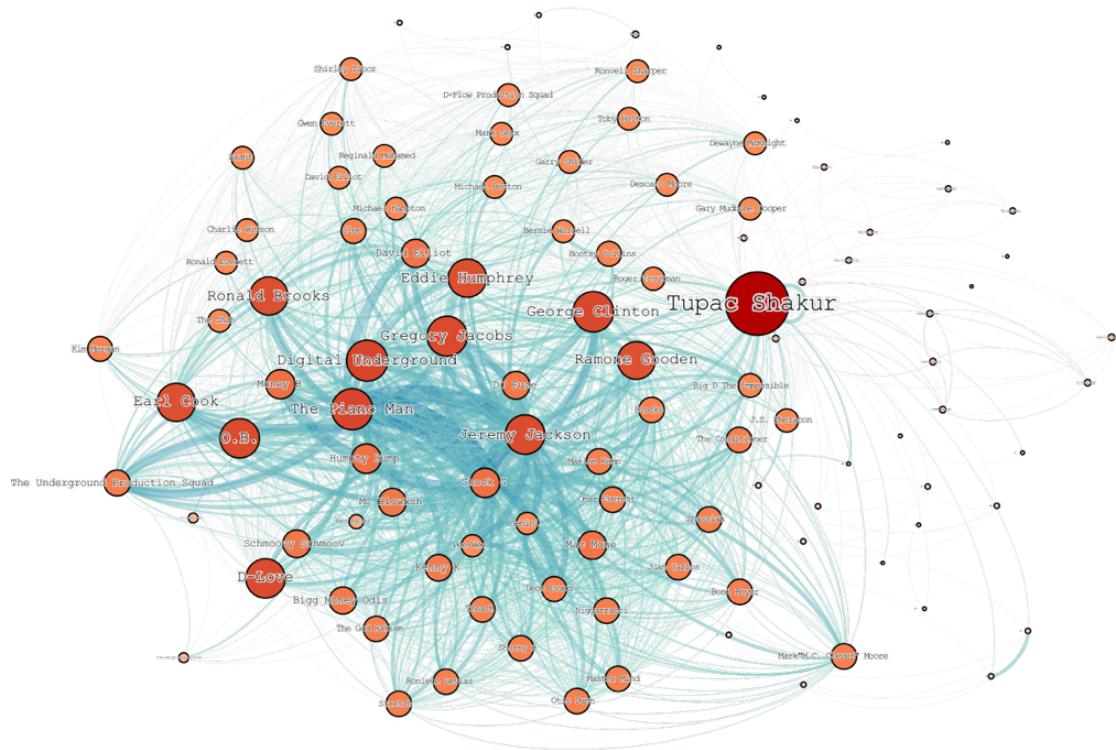


Figure 8. Neighborhood of Tupac Shakur in the Hip-Hop network. Nodes are scaled by degree.

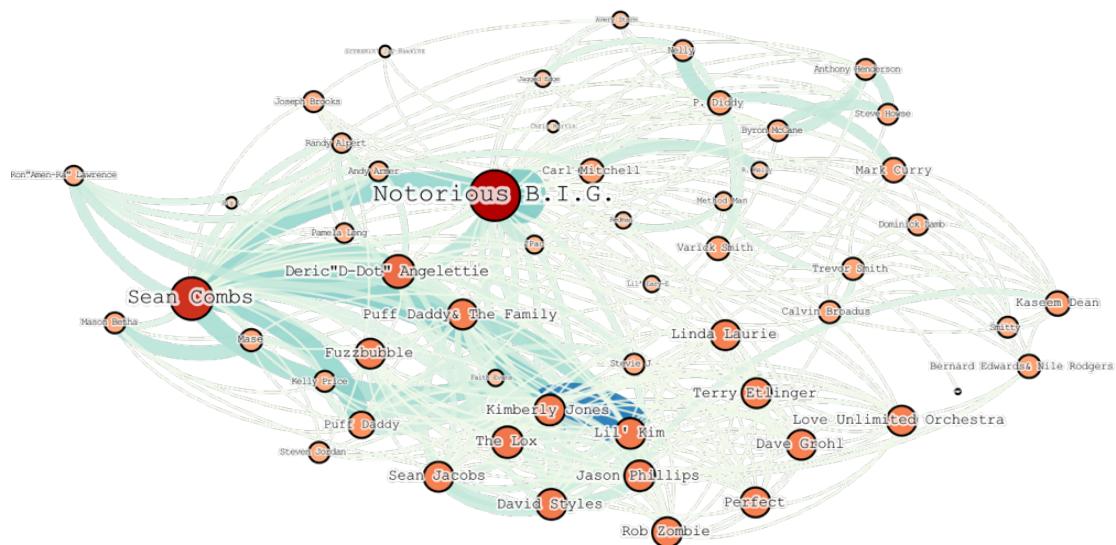


Figure 9. Neighborhood of Notorious B.I.G. in the Hip-Hop network. Nodes are scaled by degree.

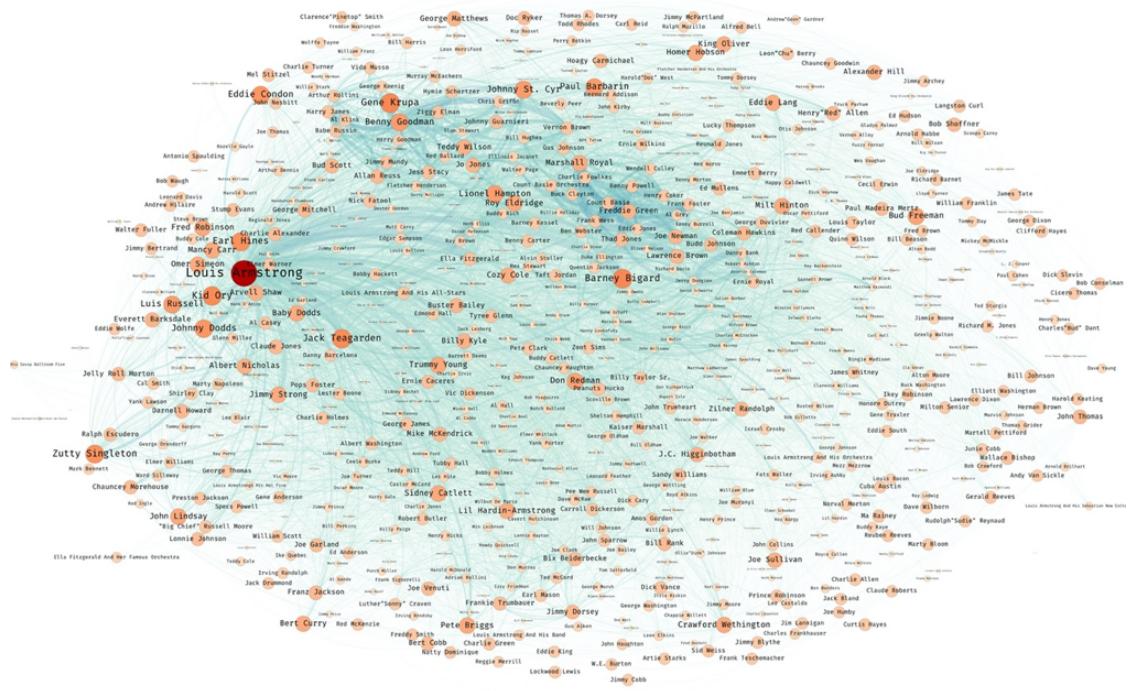


Figure 10. Neighborhood of Louis Armstrong in the Jazz network. Nodes are scaled by degree.

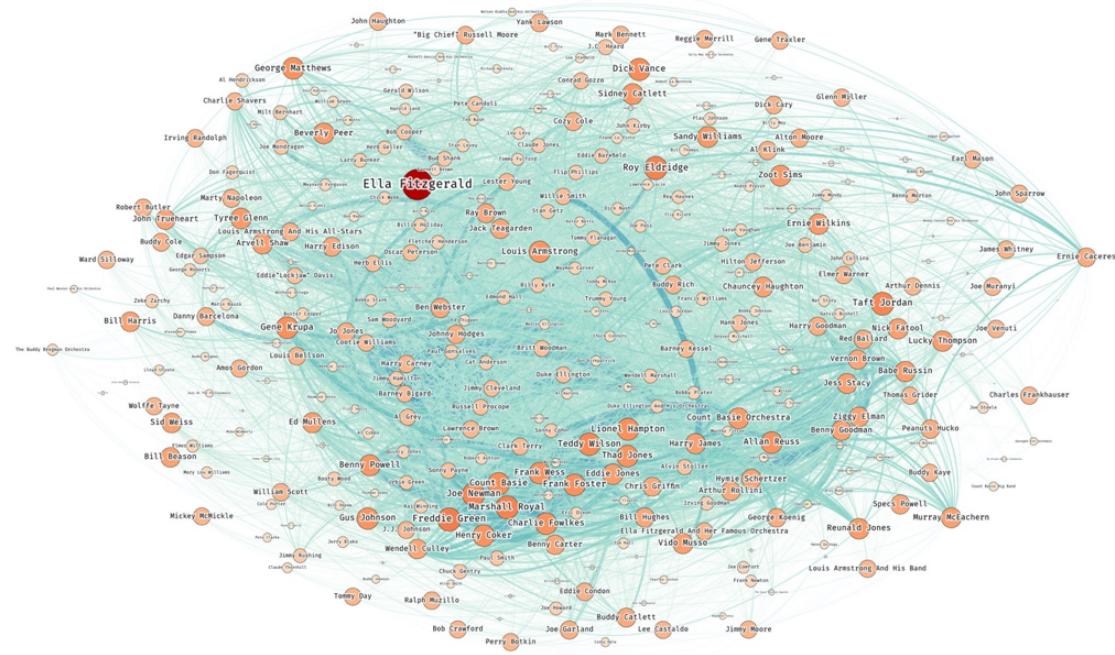


Figure 11. Neighborhood of Ella Fitzgerald in the Jazz network. Nodes are scaled by degree.

5. Conclusion

The study of collaboration patterns in music is a lively branch of musicological, sociological, and historical research. However, most existing research is based on very heterogeneous data collections, making it hard to compare different results to each other. Furthermore, the lack of a common collection of data and of a shared definition of collaboration

makes it almost impossible to compare different genres. Accordingly, existing research on music collaboration is mainly focused on singular genre analyses, with Jazz being the most studied genre by far. To support future inter-genre comparisons, we propose an approach that makes use of the freely available, ever-growing metadata collection Discogs. We further suggest deriving collaborations between musicians by extracting information of shared music releases from Discogs. Whenever two musicians played together on a record, we treat this as an instance of collaboration. We are well aware that this particular view on music collaboration comes with some conceptual limitations, as collaboration might also take place by means of unrecorded live gigs or private jam sessions. Yet, this approach to modeling collaboration can be scaled-up very easily, allowing us to obtain networks that are much more comprehensive in size than any of the data sets of comparable existing studies. These networks further include a rich set of additional metadata, for instance the time and role of each collaborative connection. Furthermore, the data source does not suffer from noisy data, as unique identifiers for artists are used in Discogs to map different spellings and pseudonyms to the same node.

To illustrate the potential of the suggested approach, we also present a case study in the inter-genre comparison of Jazz and Hip Hop, applying a variety of metrics commonly used to describe topological properties in collaborative networks. In contrast to previous work on music collaboration, we were able to demonstrate the influence of smaller artist groups and lesser known musicians. Accordingly, we were able to get some novel insights in this direction, as the collaboration networks seem to consist of many more components than previously assumed. This difference accounts for most of the divergence of metrics when comparing our results to those of previous music collaboration studies. The Discogs data suggests that the networks at hand are scale-free, (although with a very low scaling exponent). In other words, very few artists are very well connected to a lot of other artists, while most of the artists only have very few relations to other artists. The collaboration patterns of the two investigated genres mainly differ in group size, number of collaborations per artist, and overall density of the network. In Jazz, collaboration is more frequent, takes place with more people at a time, and with a greater variety of people overall than in Hip Hop in formal recordings.

Future work could provide a more detailed insight into collaboration patterns, e.g., by interrelating network position (i.e., centrality) with popularity metrics from other sources, as shown in Smith [Smith 2006]. Another possibility would be the inclusion of a more in-depth analysis of the formed cliques using the Girvan-Newman method of betweenness centrality [Girvan and Newman 2002], as previously employed by Gleiser and Danon [Gleiser and Danon 2003]. Other future directions might point toward the exploration of the scale-free properties of the networks in more detail. Central questions here are whether the length of a music career or the overall degree is a reliable indicator for success. Providing an interactive exploration interface similar to Filipova et al. [Filipova et al. 2012] could further enhance the usefulness of our approach for future research, as it allows researchers from all backgrounds to gain insight into the collaborative patterns. An example of such an application based on Discogs data is the *DiscoGraph* project.

Finally, the network data itself could be enhanced in several ways, for instance, by adding more releases (also from lower quality levels) in order to further increase the network size. Metadata about sub-genres of a release would be a valuable addition for historical network analysis, since the evolution of those sub-genres and the musicians associated with them could be studied in more detail. Enhancing the networks to records that are listed with more than one genre would provide the grounds to facilitate research in inter-genre collaborations. On the whole, we hope that the approach described in this article will initiate many further, more detailed studies that will examine collaboration strategies and patterns between artists of different genres, thus adding to the emerging field of computational musicology as part of the digital humanities.

Notes

[1] Spotify API <https://developer.spotify.com/documentation/web-api/reference/tracks/get-audio-features/>; Note: All URLs mentioned in this article were last checked July 7, 2020.

[2] Map of Jazz

[3] Note: All members of the Wu-Tang Clan are also active as solo artists.

[4] See: <https://support.discogs.com/hc/en-us/articles/360005055593-Database-Guidelines-20-Voting-Guidelines>

[5] <https://data.discogs.com>.

[6] <https://www.discogs.com/help/creditslist>

Works Cited

- Alstott et al. 2014** Alstott, J., Bullmore, E., Plenz, D. power-law: a Python package for analysis of heavy-tailed distributions. *PLoS One* 9, e85777 (2014).
- Amaral et al. 2000** Amaral, L.A.N., Scala, A., Barthélémy, M., Stanley, H.E. Classes of small-world networks. *Proc. Natl. Acad. Sci.* 97 (2000): 11149–11152.
- Araújo et al. 2017** Araújo, C.V.S., Neto, R.M., Nakamura, F.G., Nakamura, E.F. Using Complex Networks to Assess Collaboration in Rap Music: A Study Case of DJ Khaled, in: Proceedings of the 23rd Brazillian Symposium on Multimedia and the Web - WebMedia '17. Presented at the 23rd Brazillian Symposium on Multimedia and the Web, ACM Press, Gramado, RS, Brazil (2017): 425-428.
- Barabási et al. 2003** Barabási, A.-L., Bonabeau, E. Scale-free networks. *Sci. Am.* 288 (2003): 60–69.
- Barrat et al. 2004** Barrat, A., Barthelemy, M., Pastor-Satorras, R., Vespignani, A. The architecture of complex weighted networks. *Proc. Natl. Acad. Sci.* 101 (2004): 3747-3752.
- Bastian et al. 2009** Bastian, M., Heymann, S., Jacomy, M. Gephi: An Open Source Software for Exploring and Manipulating Networks, in: *Proceedings of the International AAAI Conference on Weblogs and Social Media* (2009)
- Bogdanov and Serra 2017** Bogdanov, D., Serra, X. Quantifying music tends and facts using editorial metadata from the Discogs database, in: Hu X, Cunningham SJ, Turnbull D, Duan Z. *ISMIR 2017 Proceedings of the 18th International Society for Music Information Retrieval Conference* (2017): p. 89-95.
- Brandes et al. 2011** Brandes, U., Eiglsperger, M., Herman, I., Himsolt, M., Marshall, M.S., 2001. GraphML progress report structural layer proposal, in: *International Symposium on Graph Drawing*. Springer, pp. 501–512.
- Burgoyne et al. 2016** Burgoyne, J.A., Fujinaga, I., Downie, S. Music Information Retrieval, in: Schreibman, S., Siemens, R., Unsworth, J. (Eds.), *A New Companion to Digital Humanities*. Wiley Blackwell, West Sussex (2016).
- Filipova et al. 2012** Filipova, D., Fitzgerald, M., Kingsford, C., Benandon, F. Dynamic Exploration of Recording Sessions between Jazz Musicians over Time, in: 2012 International Conference on Privacy, Security, Risk and Trust and 2012 International Conference on Social Computing, IEEE, Amsterdam, Netherlands (2012): 368–376.
- Giaquinto et al. 2007** Giaquinto, G., Bledsoe, C., McGuirk, B. Influence and similarity between contemporary jazz artists, plus six degrees of kind of blue. PhD Thesis (2007).
- Gioia 1998** Gioia, T. *The History of Jazz*. Oxford University Press, USA (1998).
- Girvan and Newman 2002** Girvan, M., Newman, M. Community structure in social and biological networks. *Proc. Natl. Acad. Sci.* 99 (2002): 7821–7826.
- Gleiser and Danon 2003** Gleiser, P.M., Danon, L. Community structure in jazz. *Adv. Complex Syst.* 6 (2003): 565–573.
- Hagberg et al. 2008** Hagberg, A., Swart, P., S Chult, D. Exploring network structure, dynamics, and function using NetworkX. Los Alamos National Lab. (LANL), Los Alamos, NM, United States (2008).
- Hammou 2014** Hammou, K. Between social worlds and local scenes: Patterns of collaboration in francophone rap music, in: *Social Networks and Music Worlds*. Routledge (2014): 128–145.
- Hannibal 2015** Hannibal, B. The Network Influences of Innovation and Lifetime Career Success in Jazz Musicians between 1945 and 1958. PhD Thesis (2015).
- Hinton et al. 1988** Hinton, M., Berger, D.G., Morgenstern, D., 1988. *Bass line: the stories and photographs of Milt Hinton*. Temple University Press.
- Jockers 2013** Jockers, M. L. (2013). *Macroanalysis: Digital methods and literary history*. University of Illinois Press.
- Macdonald and Wilson 2006** Macdonald, R.A.R., Wilson, G.B. Constructions of jazz: How Jazz musicians present their collaborative musical practice. *Music. Sci.* 10 (2006): 59–83.

Makkonen 2017 Makkonen, T. North from here: the collaboration networks of Finnish metal music genre superstars. *Creat. Ind. J.* 10 (2017): 104–118.

Mertl et al. 2008 Mertl, V., O'Mahony, T.K., Tyson, K., Herrenkohl, L.R., Honwad, S., Hoadley, C. Analyzing collaborative contexts: Professional musicians, corporate engineers, and communities in the Himalayas, in: *Proceedings of the 8th International Conference on International Conference for the Learning Sciences-Volume 3. International Society of the Learning Sciences* (2008): 282–289.

Newman 2001 Newman, M. The structure of scientific collaboration networks. *Proc. Natl. Acad. Sci.* 98 (2001): 404–409.

Newman 2003 Newman, M. The structure and function of complex networks. *SIAM Rev.* 45 (2003): 167–256.

Newman 2010 Newman, M. *Networks: An Introduction*, 1st ed. Oxford University Press (2010).

Park et al. 2007 Park, J., Celma, O., Koppenberger, M., Cano, P., Buldú, J.M. The social network of contemporary popular musicians. *Int. J. Bifurc. Chaos* 17 (2007): 2281–2288.

Park et al. 2015 Park, D., Bae, A., Schich, M., Park, J. Topology and evolution of the network of western classical music composers. *EPJ Data Sci.* 4, 2 (2015).

Patuelli et al. 2011 Patuelli, C., Weller, C., Szablya, G. Linked Jazz: An Exploratory Prototype, in: *International Conference on Dublin Core and Metadata Applications*. Dublin, Ireland (2011): 158–164.

Phillips and Kim 2009 Phillips, D.J., Kim, Y.-K. Why pseudonyms? Deception as identity preservation among jazz record companies, 1920–1929. *Organ. Sci.* 20 (2009): 481–499.

Schilling and Phelps 2007 Schilling, M.A., Phelps, C.C. Interfirm collaboration networks: The impact of large-scale network structure on firm innovation. *Manag. Sci.* 53 (2007): 1113–1126.

Seddon 2005 Seddon, F.A. Modes of communication during jazz improvisation. *Br. J. Music Educ.* 22 (2005): 47–61.

Seddon and Biasutti 2009 Seddon, F., Biasutti, M. A comparison of modes of communication between members of a string quartet and a jazz sextet. *Psychol. Music* 37 (2009): 395–415.

Smith 2006 Smith, R.D. The network of collaboration among rappers and its community structure. *J. Stat. Mech. Theory Exp.*, P02006 (2006).

Smith 2016 Smith, S. *Hip-Hop Turntablism, Creativity and Collaboration*, Routledge (2016).

Teitelbaum et al. 2008 Teitelbaum, T., Balenzuela, P., Cano, P., Buldú, J.M. Community structures and role detection in music networks. *Chaos Interdiscip. J. Nonlinear Sci.* 18, 043105 (2008).

Watts and Strogatz 1998 Watts, D.J., Strogatz, S.H. Collective dynamics of 'small-world' networks. *Nature* 393 (1998): 440.

Yao et al. 2017 Yao, D., van der Hoorn, P., Litvak, N. Average nearest neighbor degrees in scale-free networks. *ArXiv Prepr. ArXiv170405707* (2017).

Zhang et al. 2006 Zhang, P.-P., Chen, K., He, Y., Zhou, T., Su, B.-B., Jin, Y., Chang, H., Zhou, Y.-P., Sun, L.-C., Wang, B.-H., others. Model and empirical study on some collaboration networks. *Phys. Stat. Mech. Its Appl.* 360 (2006): 599–616.

de Lima e Silva et al. 2004 de Lima e Silva, D., Medeiros Soares, M., Henriques, M.V.C., Schivani Alves, M.T., de Aguiar, S.G., de Carvalho, T.P., Corso, G., Lucena, L.S. The complex network of the Brazilian Popular Music. *Phys. Stat. Mech. Its Appl.* 332 (2004): 559–565.