Network Analysis of the Danish and Italian Music Industries

Lisandro Marco Benetti   Daniel Fejerskov-Quist
Supervisor: Michele Coscia

{limb, dfej}@itu.dk

Code: KISPECI1SE
Examination Group: S25KISPECI1SE024
June 2, 2025

Abstract

Cultural phenomena can be analyzed through various approaches. This thesis explores the structure and dynamics of the Danish music industry through the analytical framework of network science. Building on the methodology developed in *Node Attribute Analysis for Cultural Data Analytics: a Case Study on Italian XX-XXI Century Music* by Michele Coscia (2024) we construct a bipartite network of Danish artists and bands, using data extracted primarily from Discogs. By applying node attribute analysis we investigate genre specialization, temporal patterns, and the role of record labels in shaping collaboration between bands. In addition to analyzing the Danish network in isolation, we extend the research by comparing it to its Italian counterpart and constructing a merged international network to explore cross-country interactions. Our findings reveal distinct patterns of collaboration and genre clustering in each country, highlighting the Danish scene's relatively low genre specialization and greater cross-genre interactions. We also identify key bands and artists that serve as central connectors in both national and international contexts. This research not only provides a detailed overview of structural properties in the Danish music industry but also contributes to the growing field of cultural data analytics by demonstrating how network analysis can uncover hidden dynamics in musical collaboration.

Dedicated to Alex Puddu

Contents

1 Introduction

Node attribute analysis in network science involves examining the properties or characteristics assigned to nodes, such as demographic, categorical, or numerical attributes, to better understand their roles and influence within the network structure. By combining node attributes with structural information, it is possible to explore how these features correlate with connectivity patterns, community membership, or node centrality. This integrated approach provides more insight than purely structural analysis alone.

This type of analysis becomes especially interesting when applied to networks that represent social or cultural phenomena. For example, homophily studies have shown how race, ethnicity, age, gender, education, and religion can generate segregation within social networks composed of individuals who share similar characteristics [1]. Other research has investigated how cultural and racial attributes affect friendship networks in schools [2], or how to measure ideological polarization within social networks [3].

Cultural analytics has also been applied extensively to music. Studies have analyzed collaboration networks among jazz musicians [4], used statistical and network tools to track musical structure over time [5], employed bipartite graphs of users and artists from music streaming data to detect genre structures [6], and investigated the structural organization of music similarity networks with a focus on community detection and node role classification [7], among many others.

In this paper, we build on the work presented in *Node Attribute Analysis for Cultural Data Analytics: A Case Study on Italian XX–XXI Century Music* by Michele Coscia [8][1]. In this paper, the author introduces an innovative approach for understanding the Italian music industry from 1902 to 2024 by constructing a bipartite network that connects musicians to the bands they have played in. Coscia then examines how node attributes, such as genre, year, and region, interact with the network structure.

In 2024 we began developing a Danish network as part of the research project *"Building a Band Artist Network – For the Danish Music Industry"* [9], using similar methods and the same data structure as in **Coscia(2024)**. Since the author made the Italian network publicly available, we are able to compare them.

Like the Italian network, our dataset includes *Genre* and *Year* as node attributes. However, we chose to omit the *Region* attribute, as it is less relevant in the Danish context. Instead we incorporated *Record Label* as a new attribute.

In this paper we follow a similar line of analysis as **Coscia(2024)**, applying it to our Danish network, and we extend the work by conducting a systematic comparison between the Danish and Italian music industries. We also broaden the scope of analysis by incorporating data from record labels that is applicable to both networks.

Although part of the study focuses on comparing the networks in isolation, we also conduct a more integrated analysis by merging both networks into a single connected component to examine cross-national interactions.

[1]Given the amount of times this paper is mentioned, from this point onwards we reffer to it as **Coscia(2024)**

Some of the research questions that we aim to address include:

- Do music genres relate to time periods in similar ways in Denmark and Italy?
- Do the Italian and Danish music industries exhibit similar clustering patterns by genre?
- To what extent can publishing under a specific record label explain collaborations between bands?
- Do record labels and genres exhibit patterns of nestedness?
- Which bands are most important in connecting Denmark and Italy, and how can their importance be measured and explained?

We begin by describing our data model and summarizing how we collected, processed, and cleaned the data, documenting key decisions and procedures. We then present the methods used to analyze the network.

Once the theoretical framework is established, we perform an in-depth analysis of the Danish network and compare it with the Italian network. Following this, we examine the role of record labels, focusing in particular on their relationship with musical genres. Finally, we conclude with an analysis of the merged network that integrates both the Danish and Italian data.

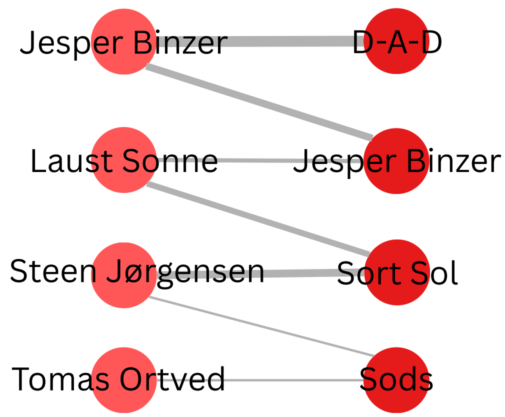
2 Data

In this section we will go through the rules and process of extracting the data for our network. It will contain the most important points from our research project [9].

2.1 Data Model

Our data model follows the framework introduced in **Coscia(2024)**. Specifically, we model a bipartite network $G = (V_1, V_2, E)$, where V_1 and V_2 represent two distinct classes of nodes: *artists* and *bands*, respectively. Figure 1 shows a small example of the network.

Figure 1: Small example of the network with artist nodes on the left and band nodes on the right. The connections shown are not necessarily true in the database.



2.1.1 Artists And Bands

It is important to emphasize that the terms *artist* and *band* are not used arbitrarily, but carry specific meanings within the context of our research. An *artist* refers to an individual human being who has contributed to a release, in a role considered relevant according to the criteria detailed in Section 2.3.3. A *band*, on the other hand, is understood as an abstract entity, a musical project typically composed of multiple artists collaborating to create and publish music.

While it is straightforward to distinguish that the Danish rock band *D-A-D* is a *band* and the Danish singer and guitarist Jesper Binzer is an *artist*, the distinction becomes less clear in cases where artists and bands share the same name. It is true that artists can release music under their own names, but in our model, the entity that publishes the release is classified as a *band*, while the individual contributing to the release is identified as an *artist*.

For example, Alex Puddu is a multi-instrumentalist included in our network as an *artist*, but he also maintains a musical project under the name *Alex Puddu*, which is treated as a separate *band* entity in our model. In this case, Alex Puddu, the *artist*, collaborates with the *band Alex Puddu*, representing a musical project distinct from the individual himself.

This distinction is also crucial for understanding how we construct both the artist and the band projections from the bipartite network, as discussed in Section 2.5.

There is one practical exception to this principle: orchestras. Although many orchestras are extracted as *bands* in accordance with our data model, several also appear as *artists*. Ideally, an orchestra should be credited by listing all its individual members involved in a given release, capturing each *artist* who contributed. However, this level of detail is rarely present in releases. In most cases, only the orchestra's name is credited.

Since identifying the full composition of orchestras at specific points in time proved to be challenging, and in most cases impossible, we decided to include full orchestras as *artists* as well. This was the only feasible way to capture collaborations involving them. This decision is particularly important for the early decades of the twentieth century, where many artists in our network are connected through orchestras and have no other individual credits beyond the orchestra name.

From this point forward, the terms *artist* and *band* are used according to the definitions outlined above, unless otherwise specified. For readability, these terms will no longer be set in italics.

2.1.2 Edges

If an artists has collaborated with a band they are connected by an edge (v_1, v_2, t) where $v_1 \epsilon V_1, v_2 \epsilon V_2$. An edge has a single attribute, which is the year the collaboration happened, and thus our network is temporal. This also means that an artist can be connected to the same band multiple times if they have collaborated in different years.

2.1.3 Node Attributes

As previously mentioned, our nodes contain three attributes. From Discogs, we extracted a total of 322 distinct genres and styles that are associated with at least one band in our network. This represents 155 fewer genres than those found in the Italian network. We believe this is due to the difference in size and genre specialization between the two countries. This is further explored in Section 4.1.

For each release published by a band, we record the associated genres and styles, maintaining a count for each. This count forms the first node attribute in our model.

The second attribute associated with each node is the set of record labels under which a band has published music. From Discogs, we extracted a total of 2,491 unique labels linked to at least one band in Denmark, compared to 14,871 labels in the Italian network. These labels are counted in the same manner as the genres and styles, based on their frequency across a band's releases.

The third and final attribute is the year of release. We represent this using a one-hot encoded matrix, where each column corresponds to a year. A value of one indicates that a band has a release in that year.

Examples of how this data is encoded can be found in the Appendix 11.2.

2.2 Data Sources

The vast majority of our data is sourced from the website Discogs[2]. In the event that Discogs did not have any useful information we checked Wikipedia[3] and in some cases AllMusic[4]. For some releases it was impossible to find any useful credits meaning we regrettably had to leave them out. Discogs also releases a monthly XML[5] dump of their entire database

[2] https://www.discogs.com/
[3] https://www.wikipedia.org/
[4] https://www.allmusic.com/
[5] We worked with the data released on February 2025

which we use to extract data on genres and labels for each record. It is important to note that Discogs is highly user curated which can give rise to inconsistencies. The way Discogs users in different countries tag genres might vary. As an example, the album *Aquarium* of the Danish band *Aqua* is tagged as as "Electronic" and "Eurohouse" for the 1997 Danish release, but it also includes "Pop" and "Europop" for the European release of the same year.

This type of inconsistencies are beyond our control and might have an influence on our analysis. Discogs started as a database for electronic music but covers all possible genres today, which can lead to a bias towards that genre. Furthermore our sources have a bias towards English-speakers which increases the likelihood of some Danish bands missing credits.

2.3 Data Collection Rules

The data collection took place between September 2024 and March 2025. In this section we describe the criteria that we used when deciding what data to extract and what not to extract. Having a set of rules for data extraction helps us obtain a coherent network. We define what we consider a published record, what counts as a Danish band and what counts as a record credit that can generate a link in the bipartite network. As we want to keep things comparable between the Danish and the Italian network our criteria are similar to the ones outlined in **Coscia(2024)**.

2.3.1 Published Record

Our first criterion is to only include records published in the music market. This means we did not include any releases for movie soundtracks, as these follow different dynamics than records released as stand-alone products. Including them would have added noise to the interpretation of our results.

When deciding what counts as a record we chose to be flexible, given how publication formats have varied across genres and time periods. For instance, while mainstream music has traditionally focused on album releases, there is a trend of releasing a lot of singles in more contemporary electronic and popular music, especially since the rise of streaming platforms. In Denmark this trend is particularly prominent in the modern Pop and Rap scenes. As an example, Tobias Rahim is strongly connected with artists like Burhan G, Jung, and Gilli but none of these collaborators appear on any of his two albums. His connections in our network are derived from single releases, and ignoring singles would have meant overlooking significant collaborations.

A modification we made to the original criteria was to include bands with only a single album release. In **Coscia (2024)**, such projects are not considered valid collaborations, but we chose to define them as meaningful, especially when the involved artists have worked together on other occasions. For instance, while Kim Larsen & Jungledreams released only one album under that name, the artists have collaborated multiple times, justifying its inclusion.

Finally, we excluded records with only one credited participant, since they can not generate any links in the network. While such records may still be relevant for genre-based analysis, they play no structural role in defining musical collaboration.

2.3.2 Danish Record

We consider a Danish record to be any release of a Danish band. This simple definition hides the real challenge: defining what a Danish band is. When deciding what Danish band is, we first look trough the records that are published in Denmark. If a band does not publish their records in Denmark, it would be difficult for the band to target a Danish audience, and therefore would not be considered a Danish band. There is a strong correlation between these two factors. However, this criteria alone is not enough, as not every band that publishes in Denmark can be considered Danish. For example, the internationally famous Swedish band ABBA has several releases in Denmark, but they cannot be considered a Danish band.

Thus we consider other factors when defining a Danish band, such as the members' origins, the band's place of formation, and its target audience. As an example the band Aqua consists of three Danish artists and one Norwegian, formed in Copenhagen and with 21 releases in Denmark. This band is, of course, considered Danish.

A significant difference between the Italian and Danish music industries is the number of artists singing in their native tongue. For cultural reasons, the Italian language is predominant in the Italian music industry, whereas in Denmark, we find a considerable number of artists who sing in English and target international audiences. In this context, the members' origin and the place of formation play a crucial role. This made language a less important factor to consider when defining a Danish band. For instance, D-A-D is considered a Danish band because, despite singing in English and targeting a broader audience beyond Denmark, all of the band members are Danish, and the band has its roots in Denmark.

2.3.3 Record Credits

When deciding what types of credits to include we are only interested in genuine collaborations that affect the final musical product and which happen in collaboration with the band. Thus the first decision is to include anyone involved that plays an instrument. We also include producers, conductors and programmed instruments.

One thing we did different from the original paper was to extract *"Mixed by"* credits but not *"Mastered by"*. We did not consider mastering as a valid collaboration as mastering can be done at any point in the future while the mixer is usually working with the band. We did not extract credits related to lyrics and songwriting, because those credits show up when bands e.g. cover songs and we can not distinguish this collaboration. An exception to this is credits such as *“Composed by”*, which is widely used in modern pop and rap releases to credit artists who participate in writing beats and melodies for songs. We made sure to only extract that credit on those types of releases.

If a band released more than one album in a given year, we generally chose only one of them to include. This decision was necessary because the albums likely featured different collaborators. Including both albums would create connections between artists who had not actually collaborated, which would affect the projected artist network (see Section 2.5). However, whenever possible, we extracted both albums by assigning one of them to a later, but close, year. Although this adjustment means the year of release is not entirely accurate, we believe the resulting connections between artists are of higher analytical value.

2.3.4 Record Label

On Discogs each release is related to a specific record label, a company that helps musicians produce, distribute, and promote their music.

Discogs again provides its data in an XML format via their *official data repository*. From all released albums in the dataset, we filtered those associated with bands from our network, yielding a list of labels that had released at least one album by one of these bands.

As mentioned earlier, Discogs is a user-generated platform, which introduces challenges related to inconsistent or duplicated information. These issues are particularly evident when it comes to labels. During data extraction, we observed frequent duplication of label entries or incorrect associations between labels and albums. Larger labels often include warnings on their pages about proper usage, highlighting the complexities of maintaining accurate records in a user-driven database. For instance, *Universal Music Group* includes such a warning.

Another common inconsistency is the duplication of labels under slightly different names, such as one using a commercial name and another using the legal or registered name. A typical example is *Al Bano Carrisi Production*, which also appears as *Al Bano Carrisi Productions Sas*, where 'Sas' refers to 'Società in accomandita semplice', the Italian equivalent of a limited partnership.

Discogs also allows the definition of label relationships such as parent labels and sub-labels. Initially, we aimed to construct a hierarchical tree of labels, but two main issues made this objective too time consuming. First, we only included labels that had released albums by bands in our network. This limitation sometimes omitted intermediate connections, making the tree incomplete. For example, *Emi Music Sweden* was not connected to *Emi Music International* because no band in our dataset had a release with *Emi Music Scandinavia*, which would have served as the intermediary.

Second, we encountered contradictory and illogical relationships between parent labels and sub labels. For instance, the Italian label *GDM* is listed as both the parent and sub-label of *GDM Music srl*. Similarly, *RDM Edition* is recorded as its own parent and sub-label.

Additionally, we encountered cases of releases attributed to *Not On Label* or self-released albums, which cannot be considered proper labels. Notably, *Not On Label* became the 16^{th} most frequently used label in Denmark. An example of such an entry is *Not On Label (Alex Puddu Self-released)*.

We believe that, from the users' perspective, labels are generally less important than artists or releases, which leads to lower data quality and less incentive to maintain consistency. Since correcting all these inconsistencies would require extensive manual effort and would be inefficient, we chose to proceed with our label analysis while acknowledging the presence of potential errors.

To perform the analysis described in Section 4.6 and Section 5, we excluded all labels that contain the phrase "Not On Label" in their name, as well as self-releases, since these cannot be categorized as actual record labels.

2.4 Data Cleaning

To ensure the quality and reliability of our database and, therefore, the validity of our results, it was essential to thoroughly clean the data after extraction. In this section, we outline the most important steps taken during the data cleaning process.

2.4.1 Errors On Nodes And Edges

We encountered several types of potential errors in the artist's names that we extracted, including homonyms, pseudonyms, and spelling variations. To identify homonyms, artists who share the same name but are actually different individuals, we typically had to manually inspect the network, looking for connections between bands that appeared to have no real association. Most of these issues were caught during data extraction. For instance, our network includes the well-known Danish musician Kim Larsen, but we also identified a lesser-known guitar player with the same name. To avoid confusion, we labeled the latter as Kim Larsen2.

To detect pseudonyms, we manually reviewed the aliases listed on each artist's Discogs page and checked whether any of them appeared in our database. This step was especially important for genres like hip-hop, where artists often change or use multiple names. In such cases, we selected a single canonical name to represent the artist, trying always to maintain the main name used by Discogs.

For smaller spelling variations, we applied the Ratcliff-Obershelp string similarity algorithm [10] to detect near-matching names. When the algorithm flagged similar names, we manually reviewed them to determine whether they referred to the same individual or to two distinct artists. For example, we found multiple spellings of Gustaf Ljunggren, which were ultimately consolidated under one canonical name.

The errors mentioned above can affect the edges in the network, as they may lead to incorrect connections between bands that are not truly related and also miss true connections. Although it is impossible to catch every instance of these issues, the impact of such noise is mitigated by the fact that we are working with a projection of the network. In particular, the Noise-Corrected (NC) backboning method we apply plays a key role in filtering out spurious correlations and unreliable links. This method helps to preserve only the most statistically significant edges, reducing the influence of noise on our final network structure. We discuss this approach in more detail in Section 3.7.

2.5 Bipartite Projections

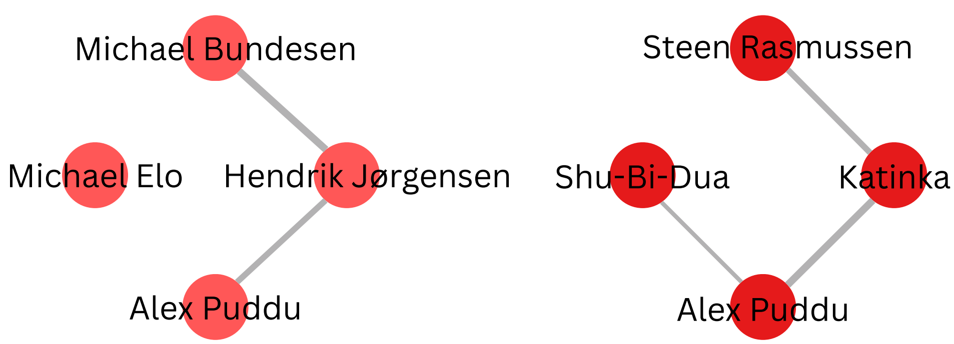
We project the bipartite network into a band projection and an artist projection, although our analysis mainly focuses on the band projection. There are different ways to count edge weights in bipartite networks[11] but we will use simple counts as edge weights. This is because we will be using using noise corrected backboning to correct edge errors, which we talk about in section 3.7, and this allows us to match the null expectation of the backboning method with the edge generating process of the projection [12]. For the artist projection the edge weight is simply a count of the number of records where two given artists have both appeared. Figure 2 shows a small example of the bipartite network projections. In the artist projection example Alex Puddu and Hendrik Jørgensen are connected because they have collaborated on the same record. But Michael Elo and Hendrik Jørgensen are not connected, even though both of them have played with Shu-bi-dua, because they have not played on the same record. In the band projection all the bands are connected because they share at least one artist.

For the band projection the edge weights need to be counted differently. Two bands should still be connected by the number of artists they share, but there is a difference in sharing an artist that only appeared on one record of both bands and an artist who appeared on ten records, as that artist will be a key member and not just a featured artist. The way we do this is with the minimum importance overlap criterion [12]. If artist a was active in w_{a,b_1} years for band b_1 and for w_{a,b_2} years in band b_2. The strength with which artist a brings b_1 and b_2 together is the minimum number of years in which artist a was part of either, $min(w_{a,b_1}, w_{a,b_2})$. If we say that the set of shared artists between b_1 and b_2 is $b_1 \cap b_2$, then the weight of the edge between the two bands in the band projection becomes:

$$w_{b_1,b_2} = \sum_{a \epsilon b_1 \cap b_2} min(w_{a,b_1}, w_{a,b_2})$$

this matches the intuition that bands have stronger edges only if they share many artists and those artists have contributed to many of the bands records.

Figure 2: Small example of the bipartite projections. Artist projection left & band projection right



(a) (b)

3 Methods

In this section we will explain the most important theory behind the methods used for analyzing the network. These methods are also used in **Coscia(2024)**.

3.1 Node Attribute Variance

Variance [13] is a measure of how much a variable's values disperse or in other words how far they are spread around the mean. For a variable x the variance is calculated as

$$var_x = \frac{1}{2} \sum_{u,v} x_u x_v d_{uv}^2$$

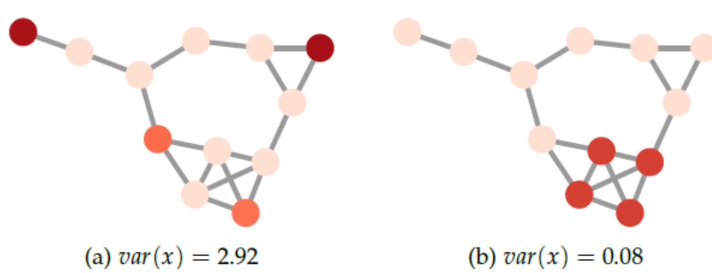
Here *d* is the Euclidean distance between two observations *u* and *v*. The variance of a network can be estimated by replacing the Euclidean distance with a proper metric. The best choice for this is the effective resistance matrix Ω [14]. The effective resistance matrix uses random walks to estimate distance, which makes it more stable to small fluctuations in the graph structure, like removing or adding en edge, than e.g. using the shortest path distance as a metric. By substituting in Ω we get

$$var_{x,G} = \frac{1}{2} \sum (x \otimes x) \Omega^2$$

We provide an example in the context of our network. The nodes carry the attributes of year, genre and label which can be thought of as a signal distributed over the graph. We can measure the dispersion of this signal across the network by using the graph-based definition of variance. By using the effective resistance Ω instead of the Euclidean distance we can measure how "far apart" two bands are based on all paths connecting them. Thus a high variance measured with $var_{x,G}$ indicates that important bands, e.g. for a genre, are scattered across the network while a low variance indicates more clustering of a given attribute.

The illustration in Figure 3 can help to understand network variance[6]. The dark nodes represent a high value of a given node attribute and the variances are calculated using the effective resistance matrix Ω. Intuitively it makes sense for figure a to have a high variance and for figure b to have a low variance, which is also what the measure finds.

Figure 3: An example of network variance



3.2 Node Attribute Distance

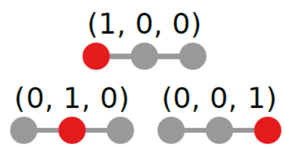
In a normal flat Euclidean space the distance between two vectors x and y can be found as:

$$\delta_{x,y} = \sqrt{(x-y)^T I (x-y)}$$

[6]This Figure and example is taken from *The Atlas For The Aspiring Network Scientist* [15]

Here I is the identity matrix and using that to transform the vector space means that we are treating each dimension as equally important and independent. But in our network the nodes correspond to bands and they are connected based on meaningful structural relationships. Therefore we would like to know how different the nodes are along the structure of the graph. Figure 4 gives an example using three vectors to help illustrate this [7]. Using the Euclidean distance the vectors (0,1,0) and (0,0,1) would both have the same distance to (1,0,0). However, we are not interested in this behavior as it does not consider the connection of the nodes. Looking at the vectors (1,0,0) and (0,0,1) the left and right node are not connected, so we would like for this to be a longer distance than the distance between (1,0,0) and (0,1,0).

Figure 4: Example Of The Euclidean Distance



So, by replacing I with the Moore–Penrose pseudoinverse of the Laplacian, $L^{\dagger}$ [16] we get

$$\delta_{x,y,G} = \sqrt{(x-y)^T L^{\dagger}(x-y)}$$

which is the generalized Euclidean distance. It is still a measure of how different the vectors x and y are, but it takes into account how the nodes are connected. If two nodes are tightly connected the difference in their attribute values are less important. In other words, when using the generalized euclidean measure, more importance is put on the edge weights, meaning how the nodes are connected. An example in terms of our network follows here: We can think of two bands that might look like they are close because they share attributes such as genre, year and label. But if they are only weakly connected, or not connected at all, then the distance between them is actually much greater than between two bands who do not share attributes but have a lot of collaboration.

3.3 Node Attribute Clustering

Defining a distance metric between numerical vectors on a graph enables us to make use of unsupervised learning approaches that leverage the network structure. When performing clustering analysis the goal is to put similar data points into distinct groups, or clusters. The elements within the same clusters should, on average, ideally be more similar to each other than to those outside the cluster. The distance metric for node attributes described in section 3.2 gives us a way to identify clusters based on how the node attributes are distributed across the network.

However, there are some challenges to this as each node in the network corresponds to a different analytical dimension. Networks often consist of a large number of nodes, which translates to a high number of dimensions. Even a relatively small network like the Danish band projection, with |V| = 416 nodes, will have a high dimensional space[12]. This leads to

[7]This Figure and example is taken from *The Atlas For The Aspiring Network Scientist* [15]

the well known "curse of dimensionality" [17], where increasing the number of dimensions causes observations to become more distant from one another, making it harder to identify meaningful clusters.

To address this issue, dimensionality reduction techniques are often employed. Each node attribute, initially represented as a vector in a space with |V| dimensions, can be transformed into a lower dimensional space, typically two or three dimensions. We use t-SNE[18] for the dimensionality reduction. After the data is transformed into a lower dimensional space, we apply a clustering algorithm, specifically hierarchical agglomerative clustering using the Ward linkage method [19].

Since standard dimensionality reduction techniques typically assume that the nodes live in an Euclidean space, it's essential to incorporate the distance method defined in section 3.2 when working in the non-Euclidean space induced by the network. Fortunately, many reduction methods support custom distance metrics, enabling us to preserve the network's structural information during the transformation. As a result, the reduced embeddings reflect the topology of the graph and are more suitable for clustering node attributes effectively.

3.4 Modularity

Modularity [20] is a metric used to evaluate the quality of a division of a network into predefined communities. The key idea behind modularity is to compare the actual number of edges within communities to the expected number of such edges if the edges were placed at random, but keeping the node degrees the same. A high modularity value indicates that the network has a strong community structure, meaning that many edges fall within communities and few edges connect different communities.

The modularity formula is defined as:

$$M = \frac{1}{2|E|} \sum_{u,v \in V} \left[A_{uv} - \frac{k_v k_u}{2|E|} \right] \delta(c_v, c_u)$$

where:

- M: Is the modularity score, a value between -1 and 1.
- $|E|$: The total number of edges in the network.
- A_{uv}: The adjacency matrix element: 1 if there is an edge between node u and node v, 0 otherwise.
- k_u, k_v: The degree of node u and node v,
- $\delta(c_v, c_u)$: The Kronecker delta function:

  $$\delta(c_v, c_u) = \begin{cases} 1 & \text{if } c_v = c_u \text{ (i.e., same community)} \\ 0 & \text{otherwise} \end{cases}$$

  This ensures the sum only considers pairs of nodes in the same community.

- $\frac{k_v k_u}{2|E|}$: This represents the expected number of edges between nodes u and v if edges were distributed at random but respecting the degree distribution.

The modularity score is used in Section 4.1 with the five biggest genres of the Danish network as communities.

3.5 Assortativity

We used attribute assortativity to calculate assortativity within bands of the five biggest genres in Section 4.1, and within nationality in Section 6. We also use numeric assortativity with average year of activity in Section 4.1.

Numeric assortativity [21] measures the Pearson correlation of a numerical node attribute across edges. The numeric assortativity coefficient r is defined as:

$$r = \frac{\sum\limits_{(u,v)\in E} (x_u - \bar{x})(x_v - \bar{x})}{\sqrt{\sum\limits_{(u,v)\in E} (x_u - \bar{x})^2} \cdot \sqrt{\sum\limits_{(u,v)\in E} (x_v - \bar{x})^2}}$$

where:

- r it the assortativity coefficient, a value between -1 and 1..
- E is the set of edges in the graph,
- x_u, x_v are the numeric attributes of nodes u and v, respectively. We use average year of activity in our case.
- $\bar{x}$: Mean of the attribute x over all nodes.

The attribute assortativity coefficient[21] r is defined as:

$$r = \frac{\sum_i e_{ii} - \sum_i a_i^2}{1 - \sum_i a_i^2}$$

where:

- r it the assortativity coefficient, a value between -1 and 1..
- e_{ij} is the fraction of edges connecting nodes of type i and j
- $a_i = \sum_j e_{ij}$ is the fraction of edges connected to type i.

3.6 Building The Eras Dendrogram

We can discover distinct eras in our network by using the temporal node attributes to make clusters. The clustering method is close to the description from section 3.3 but with some additional constraints. We limit the clustering technique to be agglomerative hierarchical clustering, meaning that we begin by treating each node as its own cluster and then merge the ones that are closest to each other. We only merge clusters that are adjacent to each other with respect to time, meaning that we wont merge a cluster containing 1986 with 1988 unless 1987 is part of one of those clusters. Finally we treat merged nodes (clusters) as a new node with an attribute vector containing the average attributes of the nodes inside that cluster. However, the last constraint means that the cluster of 1987-1988 could be closer to 1986 than 1987 was originally. Because of this we use the following distance function:

Suppose we are trying to merge clusters X and Y. Suppose also that Y was built merging Y_1 with Y_2, at distance $\delta^{'}_{Y_1,Y_2,G}$ Then, the merging distance of X and Y is: $\delta^{'}_{X,Y,G} = max(\delta_{X,Y,G}, \delta^{'}_{Y_1,Y_2,G})$,as described [12]. This ensures that $\delta^{'}$ is a monotone function. If X and Y both only contain one element, then $\delta^{'} = \delta$.

The clustering works as follows: To start, we produce a binary matrix that encodes a 1 if a band released an album in a given year and 0 otherwise. Then the distances between all adjacent year vectors are calculated. The years with the shortest distances are merged into clusters and the new distances are calculated. Once all the clusters are finalized we join with data containing a count of how many albums each band has released in a given genre. From this we can again calculate the distance from each genre to a given cluster/era.

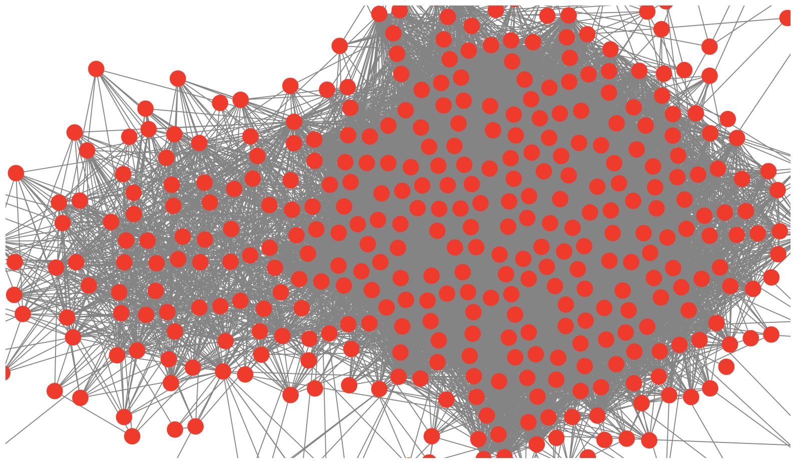
We also have to choose how to cut the dendrogram. Normally this would be done by selecting a threshold that maximizes some quality function. However, given that our data is sparse and noisy we do not want to use the same threshold for all clusters, as this could lead to too many singletons or to large clusters. The following criteria are used to ensure an adaptive threshold. We always merge singletons to any cluster. We reject merging clusters if both of them contain more than ten years. If a merge is rejected the clusters are final.

3.7 Network Backboning

Network backboning refers to the process of simplifying a network that is excessively dense by removing edges that are likely insignificant or weak. The goal of this reduction is to reveal the meaningful underlying structure of the network by eliminating noisy or spurious connections. A backbone allows for clearer visualizations and more accurate application of standard network analysis techniques.

In our case, the Danish network exhibits a “hairball” structure (see Figure 5), making interpretation and analysis difficult. Therefore, sparsification is necessary to make the network more manageable and informative.

Figure 5: Danish network will all it's edges



There are several approaches to network backboning. In this paper, we adopt the Noise-Corrected (NC) method [22].

A primary reason for using this technique is purely practical: to maintain consistency with the methodology employed in **Coscia(2024)**. To ensure that our results are comparable to those presented in that study, it is essential to adopt similar analytical approaches and network structures. Although the full, unsparsified database is available, the purpose of this paper is not to redo Coscia's analysis using different techniques, but rather to explore and understand the Danish and Italian music industries using the same parameters and framework.

The second reason, which is more technical and adds up to the first one, is that we agree with the justification presented in the supplementary material of **Coscia(2024)** [12] for adopting the Noise-Corrected (NC) backboning method.

As mentioned in Section 2.2, our main source of information is the website Discogs. Being a user-generated database it is inherently prone to errors, misspellings, and missing data. Although extensive data cleaning was performed, as detailed in Section 2.4, we cannot assume the absence of mistakes or noise. In fact, we assume the opposite.

Consequently, when choosing a general purpose backboning method, we must avoid techniques that assume perfectly clean data, particularly regarding edge weights. Methods such as the High Salience Skeleton [23] or the Doubly-Stochastic Transformation [24] are sensitive to noise and do not assess whether an edge is significantly stronger than what might be expected by chance. Therefore, we must discard them as suitable methods for our specific context.

Among the methods that do account for noise and prioritize statistically significant connections, the Disparity Filter (DF) [25] may initially seem appropriate. However, we opted not to use it for a key reason.

The Disparity Filter evaluates each edge locally by assuming that a node distributes its total weight uniformly across its neighbors. While this may be suitable in some networks, it is too simplistic for our context. Large bands with extensive discographies and collaborations naturally exhibit skewed distributions of connections, making the assumption of uniform distribution unrealistic.

In contrast, the Noise-Corrected (NC) method evaluates edge weights against a global null model that considers the activity levels of both nodes involved. This makes it better suited to handle unbalanced relationships, such as collaborations between large and small bands. The NC method provides a more nuanced and statistically robust approach to backboning: it treats asymmetric relationships fairly and does not rely on overly simplistic assumptions about interaction patterns. For these reasons, we chose the Noise-Corrected method as the most appropriate technique for our analysis.

When creating a projection from the artist–band bipartite network into a band–band unipartite network, bands become connected by weighted edges that reflect the number of shared artists. However, this projection process is particularly vulnerable to noise. Inaccurate artist–band associations can artificially inflate edge weights, while missing artist credits can reduce them, potentially eliminating otherwise valid connections.

To mitigate this, the Noise-Corrected (NC) backbone method applies a statistical test that estimates the likelihood that an observed edge weight could have occurred by chance. If a band–band connection results from a spurious or accidental link in the bipartite network, its weight is more likely to fall within the expected range of a null model and will be filtered out during the pruning process.

Conversely, real connections that are weakened due to incomplete data may also risk being discarded. To address this, the method uses a threshold for statistical significance. This ensures that only the most robust and well-supported links are retained, even when affected by a limited degree of missing or noisy data.

We apply a threshold that produces a backbone retaining 97.10% of the original nodes within the largest connected component. This means that fewer than 3% of the nodes must be reconnected to ensure the network forms a single connected component. To achieve this, we examine all pairs of nodes and iteratively add the edge with the highest statistical significance that connects two previously disconnected components. This process repeats until the entire network is connected as one component.

Crucially, this reconnection step introduces minimal distortion to the overall network structure because it only involves adding edges that fall below the original significance threshold and affects less than 3% of the nodes. As a result, the backbone preserves the meaningful connections identified by the Noise-Corrected method while ensuring network connection for subsequent analysis.

4 Network Analysis

In this section we will perform an analysis of the Danish band-artist network similar to the analysis done in **Coscia(2024)** and compare our findings for the Danish network with the Italian network. We perform a new analysis by looking at the correlation of the genre variances between the two countries.

4.1 Summary Statistics

Table 1 shows basic summary statistics of both the Danish and Italian network. We can see that Denmark has 21.55 unique artists per band, which is almost 10 more than Italy which has 11.85 artists per band. This hints at more collaboration happening in the Danish music scene. We also see that Danish bands are active for 7.8 years on average, almost a year more than in Italy.

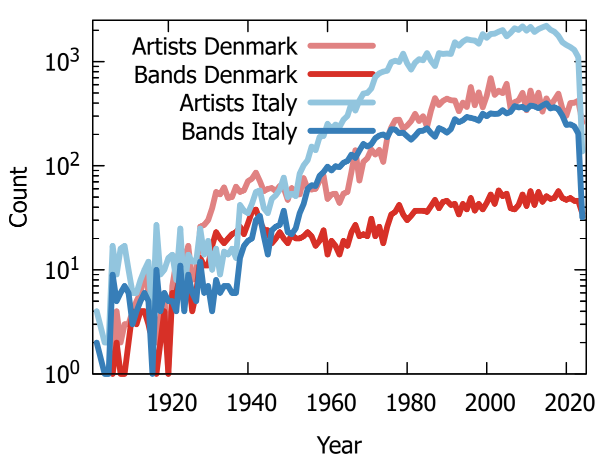
Table 1: Summary statistics for the bipartite networks of Denmark and Italy.

Denmark		Italy	
Variable	Value	Variable	Value
# Bands	416	# Bands	2,447
# Artists	8,965	# Artists	29,014
# Edges	26,176	# Edges	101,441
# Years	119	# Years	122
# Band AVG Year	7.8	# Band AVG Year	6.9

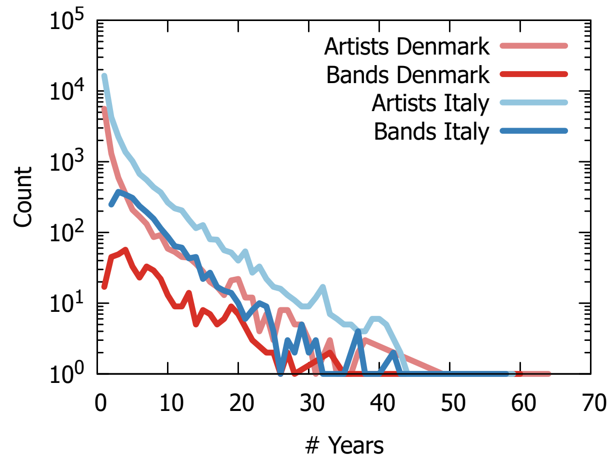
Figure 6 shows the evolution of both networks throughout the years. Figure 6a shows the amount of active artist and bands in a given year and 6b shows the amount of artists and bands with a given number of active years. We can see that both countries exhibit similar temporal patterns. It is the case for both countries that there is not much activity before the Second World War, as this is the beginning of physical recordings, and after the end of the war there is a steady increase in the amount of bands and artists in both countries. This trend continues until the 2020's when there is a sharp decline. However, we can see that the drop is much larger for Italy meaning that in the Danish database there is a higher sample of newer artists. One reason for this difference is that we included relatively modern bands with only one album, in contrast to **Coscia(2024)**, who excluded one-album projects. Secondly, Discogs is a user-generated database, so if no user had uploaded a new release by the time our data was extracted, that release would be missing. The likelihood of a release being registered on Discogs increases with the amount of time that has passed between the release date and the extraction date. This implies that during our extraction we had a higher probability of finding new releases on Discogs than **Coscia(2024)**, for the same years.

Figure 6: Temporal evolution of the networks

(a)



(b)



The amount of years that bands and artist are active is quite diverse. For both Denmark and Italy we can see from figure 6b that most artists are only active for one year and bands are only active for a couple more. In the Danish network, the artist active in the highest number of distinct years is *Det Kongelige Kapel* (The Royal Danish Orchestra), with activity in 60 different years. When we explained our data model (see Section 2.1), we justified why an orchestra like *Det Kongelige Kapel* was exceptionally considered an artist. However, if we adhere to the formal definition of an artist, the one active in the highest number of distinct years is the Danish jazz violinist *Svend Asmussen*, famously known as *"The Fiddling Viking"*. While “fiddle” is a synonym for violin, according to the Oxford Learner’s Dictionary[26], “fiddling” as an adjective means *"small, unimportant, and often annoying"*. We cannot confirm whether Svend was truly a Viking, but we can certainly say he was not a fiddling member of the Danish music industry, with activity in 49 different years.

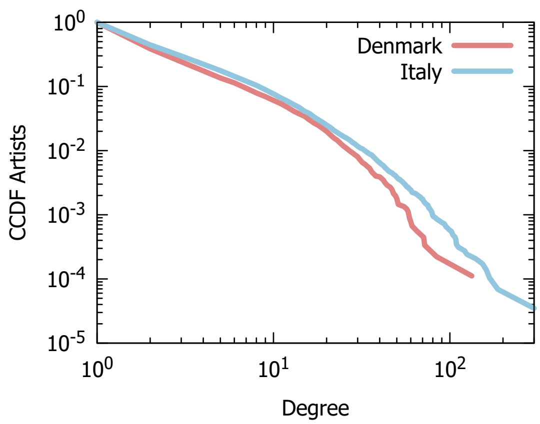
To maintain consistency and ensure comparability, we exclude all bands that have been active for less than two years, following the approach of **Coscia (2024)**. In Denmark, the mode number of active years for bands is four, one more than in Italy. The Danish band with the highest number of active years is *Det Kongelige Kapel*, appearing in 64 distinct years. This result can be explained by the fact that, unlike typical bands, orchestras are usually affiliated with formal institutions, follow established hierarchies and roles, and maintain a stable repertoire and personnel structure. Moreover, their continuity is supported by the consistent replacement of individual members over time, allowing the ensemble to remain active across decades regardless of changes in its lineup and making orchestras less constrained by time.

From figure 7 we can see that there is also high diversity in the amount of collaboration links. Both the degree distributions for bands and artists are skewed. We can see that bands in Denmark on average have a higher degree than bands in Italy, but Italy has some bands with a degree of 500 and higher, which Denmark does not. Conversely we can see that the degree of artists is generally lower in Denmark. The average degree for Danish bands is 62.92 and the band with the highest degree is Sebastian with a degree of 478. The degree here is the count of band-artist pairs meaning that we count the same artist twice if they played multiple times with the same band.

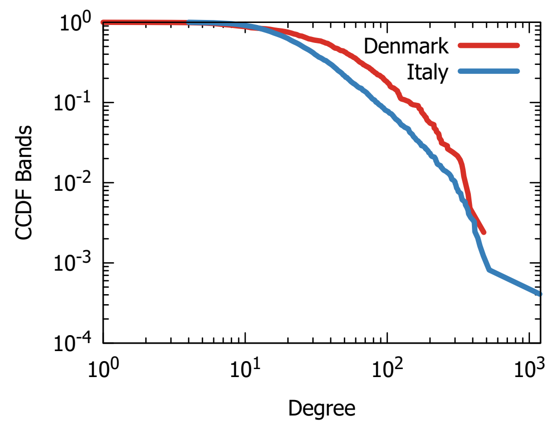
The degree distributions for the artists are lower with an average degree of 2.92 for the Danish network. The average band degree is higher in both networks because we filter out bands with less than 2 years of activity. The Danish artist with the highest amount of connections is percussionist Jacob Andersen with 133 connections.

Figure 7: The probability of a node to have a given degree or higher in the bipartite network

(a) Artists



(b) Bands

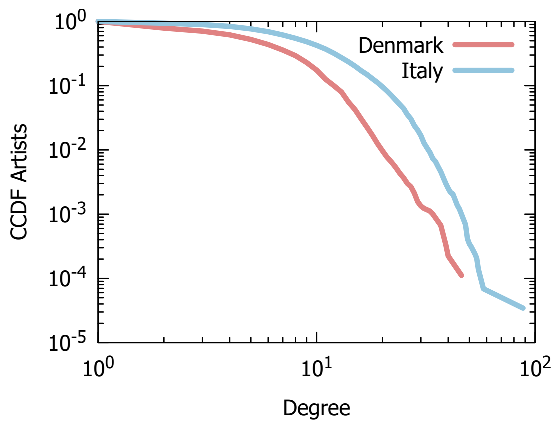


In Table 2 the summary statistics for the artist and band projections for both Denmark and Italy are shown. The Danish band projection has a low density, but a quite high average degree, clustering coefficient and modularity score. This indicates that there are meaningful communities in the network, which is to be expected in a network of artists collaborating with each other. The results between the two networks are overall very similar, except Denmark has a high average band degree and low artists degree, while the opposite is true for the Italian network. Figure 8 shows the same as Figure 7 but for the network projections.

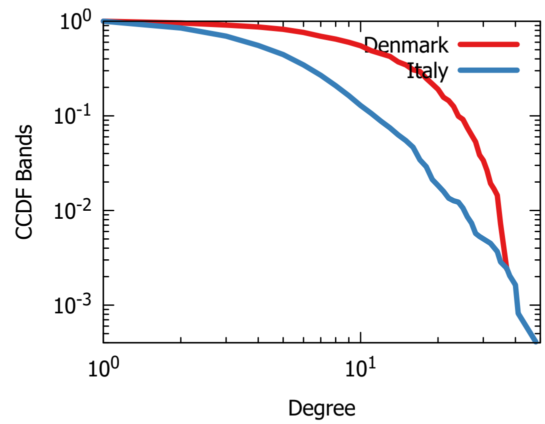
Table 2: Summary statistics for the projected networks of Denmark and Italy

	Denmark		**Italy**	
Variable	Artist	Band	Artist	Band
# Nodes	6,944	414	29,014	2,447
# Edges	16,576	2,510	141,712	6,512
Avg Deg	4.8	12.1	9.8	5.3
Density	0.0007	0.0294	0.0003	0.0022
Clustering	0.5928	0.3457	0.5567	0.4160
Modularity	0.9596	0.5918	0.8812	0.8437

Figure 8: The probability of a node to have a given degree or higher in the projected network



(a) 4a



(b) 4b

Table 3 shows the top ten bands in the Danish band projection according to three measures of centrality: degree, closeness centrality, and betweenness centrality. Degree reflects importance in a band having many connections, closeness centrality reflects that the band is in the core of the network and betweennes centrality reflects that the band is important in keeping the network connected. If a band or artist is present in the top of multiple of these measures it is a good indicator that they are important for the network. The measure of degree is mostly populated by artists from the beginnings of the Danish recording industry. That is likely due to the small scale of the industry at that time, leading to artists collaborating together with the same orchestras and in the same operas. In the measures for Closeness and Betweenness centrality we see some rock bands (Volbeat, Illdisposed) and classical orchestras (Det Kongelige Kapel, Statsradiofoniens Symfoniorkester). As for the classical orchestras, they combine the first half of the 1900's with the second half, which explain their central position. That is also related to the fact that orchestras are ever evolving entities of musicians, as it was explained before. It is also reasonable that certain rock bands obtain high scores in the analyzed metrics, given that the genre is extensively distributed throughout the network as will be shown later in Section 4.1.2. As a result, rock bands with a large number of connections tend to occupy highly central positions within the network.

Taking a more genre-specific perspective, *Illdisposed* is a death metal band [8]. This illustrates that niche genres continue to hold significance within Denmark, reflecting a broad inclusion of diverse musical styles in the Danish music industry.

Table 4 shows the same as Table 3 but for the artist projection. There are many artists that are present in two of the measures making it difficult to assess who might be the most important Danish artist as there seems to be a core of important artists.

[8] Proving, as Shakespeare famously noted, that there truly is something rotten in the state of Denmark. That observation is shared by Illdisposed, according to their third studio album, There's Something Rotten... In The State Of Denmark (1997)

Table 3: The top 10 bands in the band projection according to different centrality measures. In bold we have nodes central in multiple measures.

#	Degree	Closeness	Betweenness
1	Teddy Petersen Og Hans Orkester	**Volbeat**	**Det Kongelige Kapel**
2	Christian Arhoff	**Det Kongelige Kapel**	**Volbeat**
3	Liva Weel	**Illdisposed**	**Statsradiofoniens Symfoniorkester**
4	**Gustav Winckler**	**Statsradiofoniens Symfoniorkester**	Papa Bue's Viking Jazz Band
5	The Okey-Dokies	Frans Andersson	Aksel Schiøtz
6	Elsa Sigfúss	Magtens Korridorer	**Herman D. Koppel**
7	Dirch Passer	Baby Woodrose	**Illdisposed**
8	Raquel Rastenni	**Swan Lee**	**Gustav Winckler**
9	Peter Sørensen	**Herman D. Koppel**	Usipian
10	Valdemar Davids	The Old Man & The Sea	**Swan Lee**

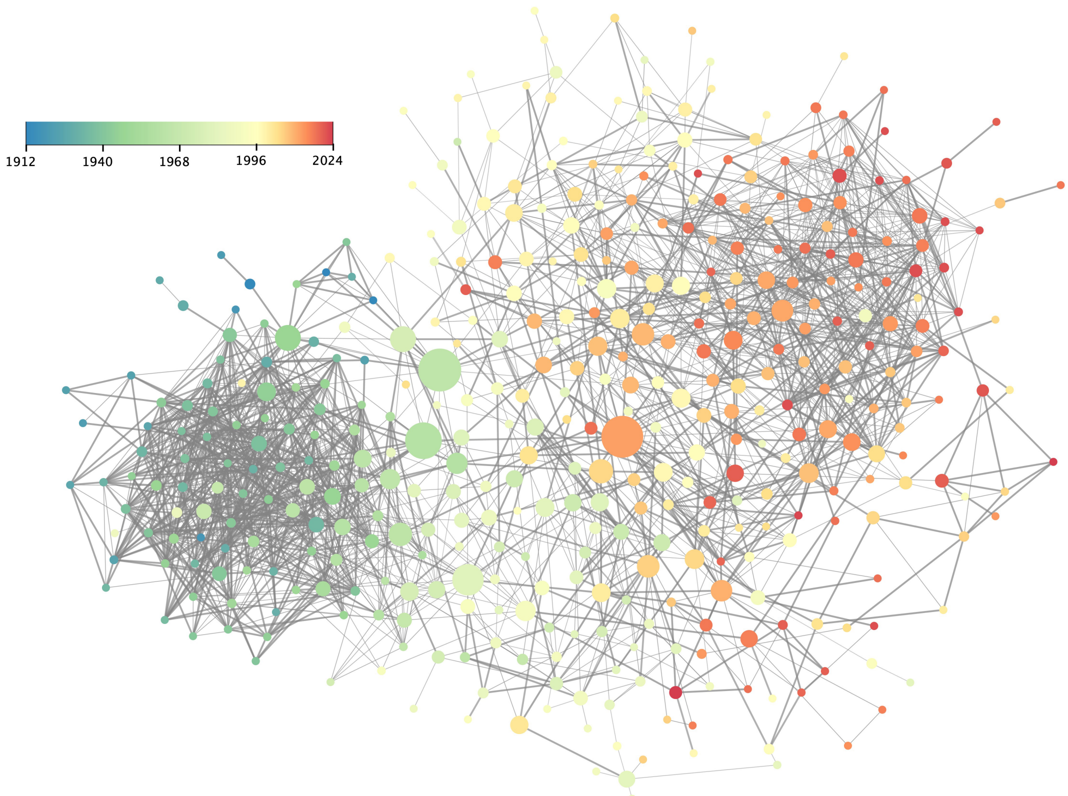
Table 4: The top 10 artists in the artist projection according to different centrality measures. In bold we have nodes central in multiple measures.

#	Degree	Closeness	Betweenness
1	Alisa Xayalith	**Poul Krebs**	**Peter Sommer**
2	Aarhus Symfoniorkester	**Nils Henriksen**	**Nils Henriksen**
3	Bjarke Mårum	Michael Falch	**Michael Friis**
4	Andrew Barranca	Troels Skjærbæk	**Per Frost**
5	Alwyn Wright	**Per Frost**	**Poul Krebs**
6	**Jacob Hansen**	Lis Sørensen	Uffe Savery
7	Ben Muhlethaler	**Palle Hjorth**	**Jacob Hansen**
8	Anders Vestergaard	**Peter Sommer**	Morten Friis
9	Agneta Bengtsson	**Michael Friis**	Burhan G
10	Jon Larsen	Sebastian	**Palle Hjorth**

4.1.1 Temporal Component Of The Band Projection

Like in **Coscia (2024)** we produce two visualizations for the band projection to highlight the most important attributes behind edge creation. The first one is the temporal component which is visualized in figure 9. We have rotated the network such that the years are displayed from oldest to newest, from left to right. It is clear to see from the figure that the network has a temporal dimension.

Figure 9: The temporal component of the band projection. Each node represents a band, and edges indicate a statistically significant overlap in shared artists. Edge color reflects statistical significance, ranging from bright (less significant) to dark (more significant), while edge thickness corresponds to the weight of the overlap. Node size is proportional to betweenness centrality, and node color indicates the average active year of the band, blue for earlier years and red for more recent ones.



Visually, two large clusters can be identified: one encompassing bands from the early twentieth century up until around the 1960s, and another consisting of more modern bands from the 1970s onward. A possible explanation for this phenomenon could be the shift that occurred with the electrification of music during the 1960s. Bands that previously relied on acoustic instruments tended to collaborate among themselves, forming the first cluster, while those incorporating electric instruments began forming connections within the newer cluster. However, since our network does not capture information about the types of instruments played, this theory remains speculative.

Using the average year of activity of each band, we also calculated the year assortativity, that measures the extent to which bands are connected to others that were active around the same time. In our network, the year assortativity coefficient is 0.928, which indicates an exceptionally strong positive correlation. This means that bands tend to share collaborators primarily with other bands that were active in the same time period.

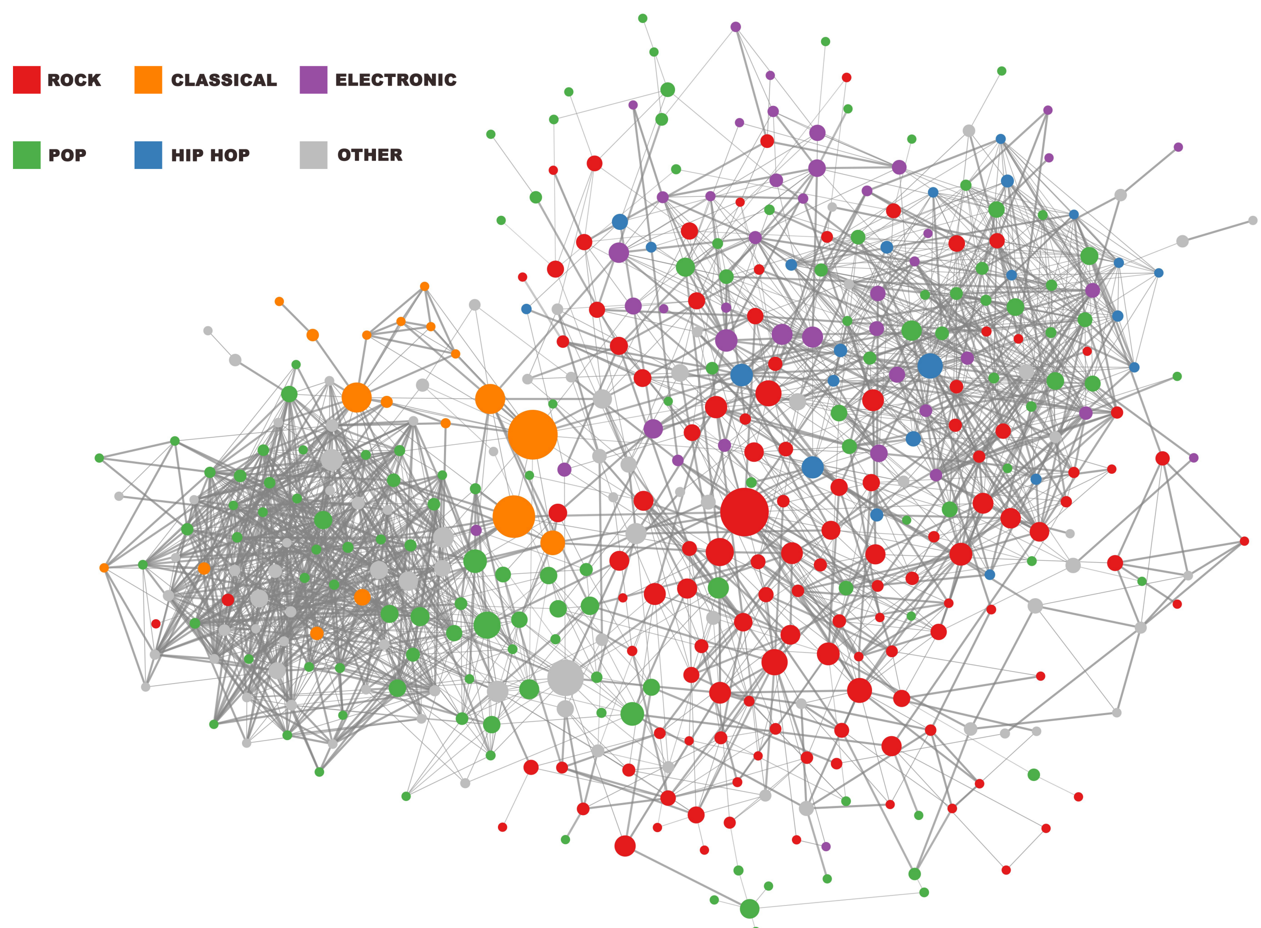
Such a high value is not surprising. Musicians, producers, and other collaborators usually operate within specific time frames due to career lifespan, trends in music production, and availability. Therefore, connections formed through shared collaborators are naturally temporally constrained.

The fact that the Italian network exhibits a similar value (0.91) reinforces this interpretation. Despite the differences in industry size or genre specialization between the Danish and Italian music, that will be discussed next, both show a high degree of temporal assortativity. This suggests that temporal proximity is a strong universal organizing principle in music collaboration networks, regardless of national context.

4.1.2 Genre Component Of The Band Projection

The second node attribute is the genre component, which can be seen in figure 10. It would be expected that genre plays a role in creating clustering within the network, as bands playing the same genre naturally tend to collaborate more with each other. Following the approach of **Coscia (2024)**, we chose to highlight the four most prominent genres in the Danish network: *pop*, *rock*, *electronic*, and *hip-hop*. Additionally, we include a fifth genre, *classical*, due to the presence of a distinct cluster containing nodes with high betweenness centrality.

Figure 10: The genre component of the band projection. Same legend as Figure 9, except for the node's color. Here, color encodes the dominant genre among pop (green), rock (red), electronic (purple), hip hop (blue), classical (orange) and other (gray).



To assign a genre to each band, we consider the genre associated with the highest number of releases. If a band's main genre is not among the five highlighted, we label them as *other*.

In contrast to the Italian network, which exhibits clear visual genre-based clusters, the Danish network appears more scattered. *Pop* is dispersed throughout the network, with significant representation both in the early 20th century and in recent decades. *Rock* is also distributed non-uniformly but tends to cluster more toward the center of the network. While electronic bands show signs of collaboration, neither *hip-hop* nor *pop* artists form clearly defined clusters.

The modularity of a network quantifies how well it can be divided into distinct communities, in this case, genre-based clusters. High modularity indicates that most connections occur within defined groups (e.g., genres), with relatively few links between them.

Using this genre-based community division from Figure 10, the Danish network yields a modularity of 0.2579, which is substantially lower than the 0.524 observed in the Italian network. We hypothesize that this difference may be attributed to the relative size of the two music industries. In competitive and saturated markets such as Italy's, artists may be pressured to specialize, creating specific niches in order to stand out. In contrast, in a smaller market like Denmark's, with fewer artists competing for the same audience, broad appeal may be sufficient, leading to less genre specialization.

This is closely related to the difference in the assortativity coefficient. The genre assortativity measures the extent to which bands are connected to others of the same genre. A high value indicates that collaborations tend to occur within genres whereas a lower value suggests more inter-genre interaction. While the Italian network has a high assortativity of 0.689, the Danish network shows a significantly lower value of 0.342. In more niche and specialized music industries, where it is statistically likely that different bands produce similar styles even within specific genres, intra-genre collaborations are more common than inter-genre collaborations. On the other hand, in smaller industries, inter-genre collaborations may occur more frequently, as the boundaries of what constitutes a "similar" band are more blurred. For example, in the Italian network, *hip-hop* forms an extremely intra-connected cluster, whereas in the Danish network, *hip-hop* artists more often collaborate with *pop* singers. This tendency can also be observed among producers, who are often less specialized in a single genre. Their broader range of collaborations facilitates inter-genre connections and consequently contributes to a lower assortativity in the network.

4.2 Genre Specialization

We get an idea of how dispersed the different genres are by calculating the network variance of the genre attributes of the nodes. A high variance indicates that a genre is spread across the network and that the bands playing that genre do not share members. The opposite is true for a low variance, indicating a cluster of bands collaborating in that genre. The five highest and lowest variance genres can be seen in Table 5. To keep our results comparable with **Coscia (2024)** we only keep genres where at least 1% of the bands have released one or more records in that genre. Like in **Coscia (2024)** the network variance values are compared with a null version of the genre, as the values themselves do not say anything about significance.

The p-value related to the variance values are thus calculated by bootstrapping, giving us pseudo p-values[8]. We find the p-values by going through the following steps. Let's assume that S is a |V| × |S| genre matrix. Then $S_{v,s}$ is an entry that tells us how many records with a genre (s) a band (v) has published. We can then create a randomized null version of S which we call S'. We ensure that each null genre in S' has the same number of records as in S by extracting with replacement at random $\sum_{v \epsilon V} S_{v,s}$ bands for genre s. The extraction is not uniform as each band has a probability of being extracted proportional to $\sum_{s \epsilon S} S_{v,s}$. This ensure that S' has the same column sum and similar row sum as S. Thus S is randomized and the popularity of each genre and each band is preserved. Then, we can count the number of such random S's in which the null genre has a higher (lower) variance than the observed genre and see if our results in S have a meaningful structure compared to S'.

Table 5 shows that the genre Lo-fi has the highest variance in Denmark and that it is significant. This is an indication that bands playing Lo-Fi have low genre specialization, likely because it is a genre that many unrelated bands try out. On the other hand we see a genre like Gangsta (a style of rap music) has very low variance, indicating that rap is a niche genre in Denmark being played mostly by the same artists.

Table 5: The genres with the five highest and lowest variance in the band projection network. $^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$.

Danish

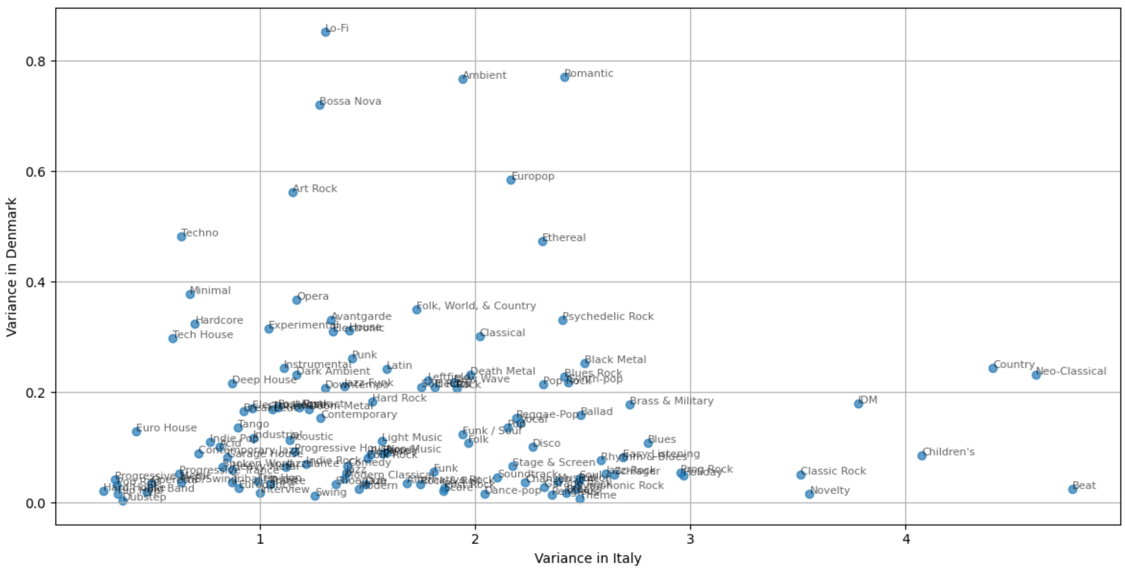
Genre	Variance
Lo-Fi	0.853^{**}
Romantic	0.772^{*}
Ambient	0.768^{**}
Bossa Nova	0.721^{**}
Europop	0.584^{**}
...	...
Theme	0.008^{***}
Ragtime	0.006^{**}
Dubstep	0.004^{***}
Gangsta	0.003^{***}
Dixieland	0.002^{***}

Italian

Genre	Variance
Stoner Rock	4.954^{**}
Beat	4.772^{***}
Neo-Classical	4.605^{***}
Country	4.403^{*}
Post-Modern	4.359^{*}
...	...
Happy Hardcore	0.249^{***}
Power Metal	0.198^{***}
Eurobeat	0.161^{***}
Gabber	0.155^{***}
Trap	0.105^{***}

To compare the results we calculated the spearman correlation (r_s) between the Danish and Italian genre variances which can be seen in Figure 11. We use the Spearman correlation because we are interested in knowing whether the same genres appear at the top and bottom of the ranking, rather than comparing the absolute variance values themselves. The Spearman correlation is rank based and less sensitive to outliers, making it more appropriate for this purpose than Pearson. We find $r_s = 0.0117$ and a P-value of 8.9781e-01. This indicates that the result is not significant and that genres that have high variance in Denmark are not the same that have high variance in Italy. It is also interesting to see in the Figure that we have no genres with high variance in both Denmark and Italy. This tells us that genres that are more consistent (in terms of the same bands playing that genre) in one country, are highly variable in the other country. The genre Lo-Fi for example has very high variance in Denmark but a low variance in Italy. We can see this as an indication that Lo-Fi is a niche genre in Italy while it might have many different sub-groups in Denmark or just many artists trying it out.

Figure 11: Spearman correlation between the variances in genre



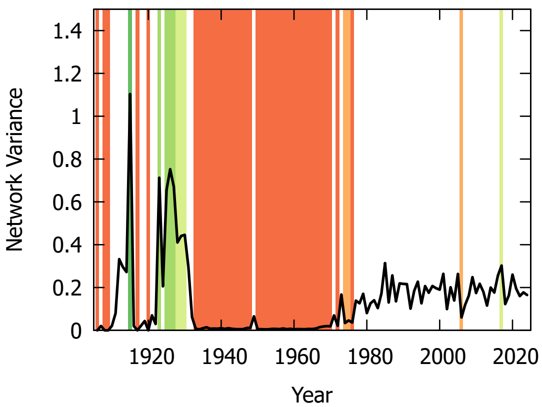
4.3 Temporal Variety

In the same way that we calculated the variance for the genres we can also calculate the variance for the years. Just like the genre variance tells us how diverse the set of bands playing it is, the variance will tell us something about how diverse a year was. To test the statistical significance of the observed variance we do the following: we shuffle the values of the node attribute for a given year a number of times and then test whether the observation is significantly higher, lower, or equal to this expectation. Figure 12 shows how the variance evolves through the years. The x-axis shows the years and the y-axis the variance. The background colors indicate significance: red means lower than expected, green means higher than expected and white means not significantly different than expected.

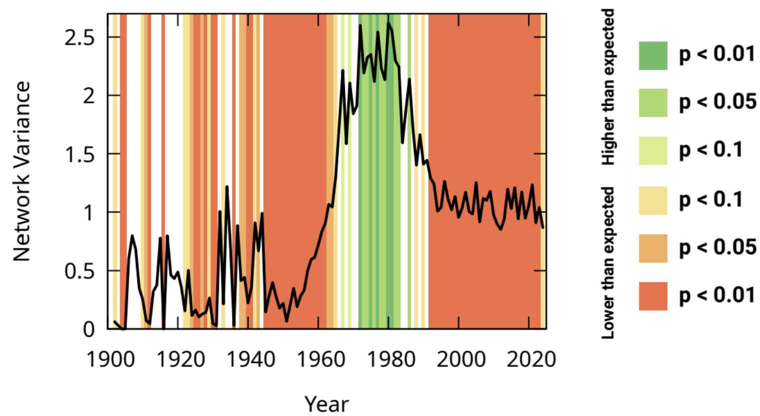
It is interesting to see the differences between the two networks. We see that Italy has two transitions, starting with low variance and low activity, into a high variance and high activity phase beginning in the 1960's and ending around the 90's, where the modern times are characterized by high activity but low variance.

This is very different from Denmark which starts with very high activity at about the mid 1910's, dropping to very low activity in the later 1910's. There is then a period of high activity and also high variance from around 1920 until around 1930 followed by a period of very low variance and activity that lasts until around the 1970's. We then see the activity starting to increase in the 1980's lasting until today, all the while there is no significant difference in the variance, indicating that the Danish music scene has been as varied as can be expected.

Figure 12: Network variance pr. decade



(a) Denmark



(b) Italy

4.4 Genre Clusters

Figure 13 shows a similarity matrix and dendrogram, giving a view of the hierarchical clustering of genres in the Danish network where lighter colors indicate higher similarity. It is made by computing the pairwise distances between all node attributes in the network, as it is done in **Coscia (2024)** for the Italian network. We have arbitrarily selected some clusters to highlight. These have been marked with colors in Figure 13 and the genres in those clusters can be seen in Table 6. For the Danish network we make similar observations as for the Italian Network. We can clearly see that similar genres and styles cluster together. From the Figure we also see a hierarchical structure of music styles and that these have clear clusters-within-clusters. The fact that we can make similar observations for the Danish network, as in the Italian network, supports the use of network based clustering to help guide definitions of new musical groups.

Figure 13: Genre Clusters Denmark

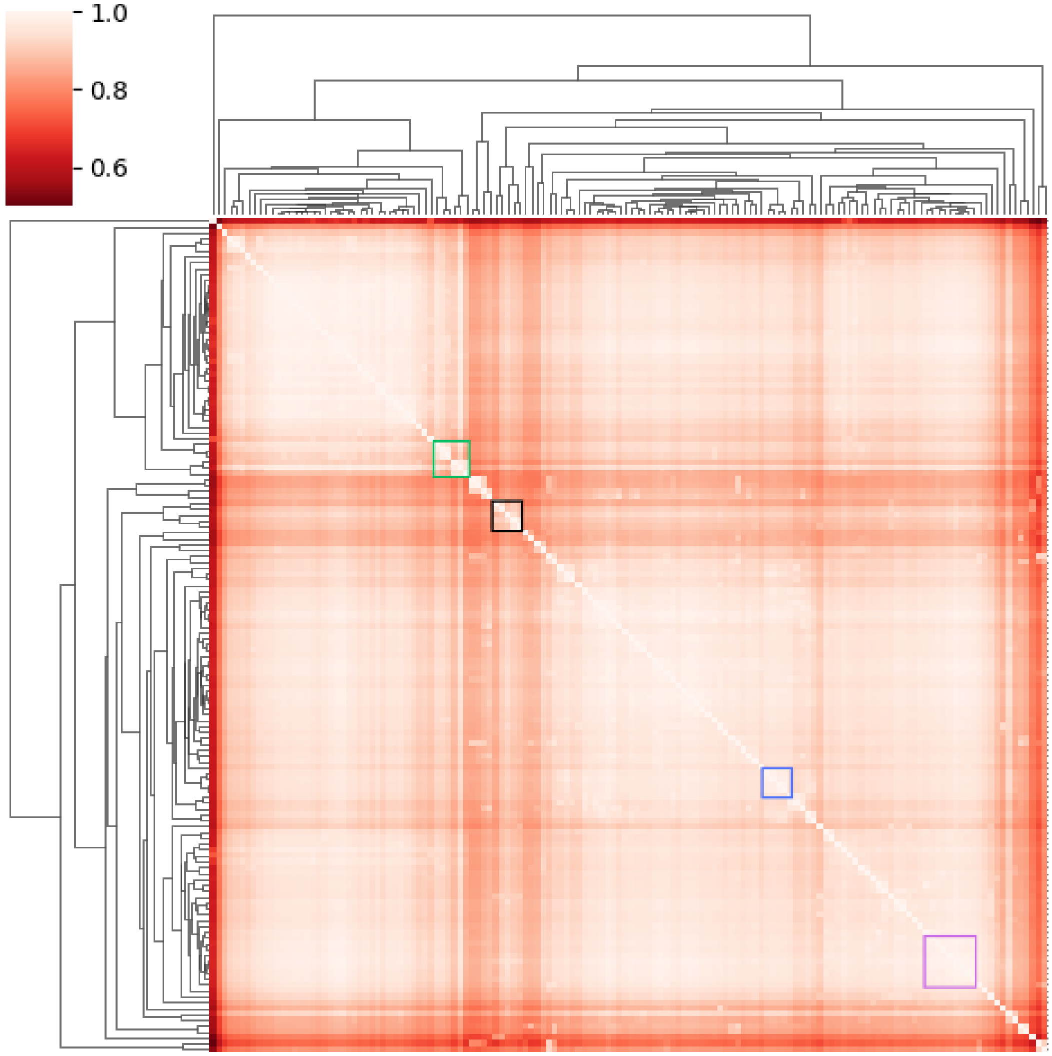


Table 6: Genre Clusters

Color	Genres
Black	Black Metal, Doom Metal, Power Pop, Punk, Hardcore
Blue	Hard House, Trance, Eurodance, Euro House, Electro House, Tech House, Techno, Garage House
Green	Modern, Contemporary, Neo-Classical, Opera, Classical, Romantic
Purple	Classic Rock, Folk Rock, Prog Rock, Rock & Roll, Rock, Pop Rock, Soft Rock Acoustic, New Wave, Symphonic Rock, Progressive Metal

4.5 Temporal Clusters

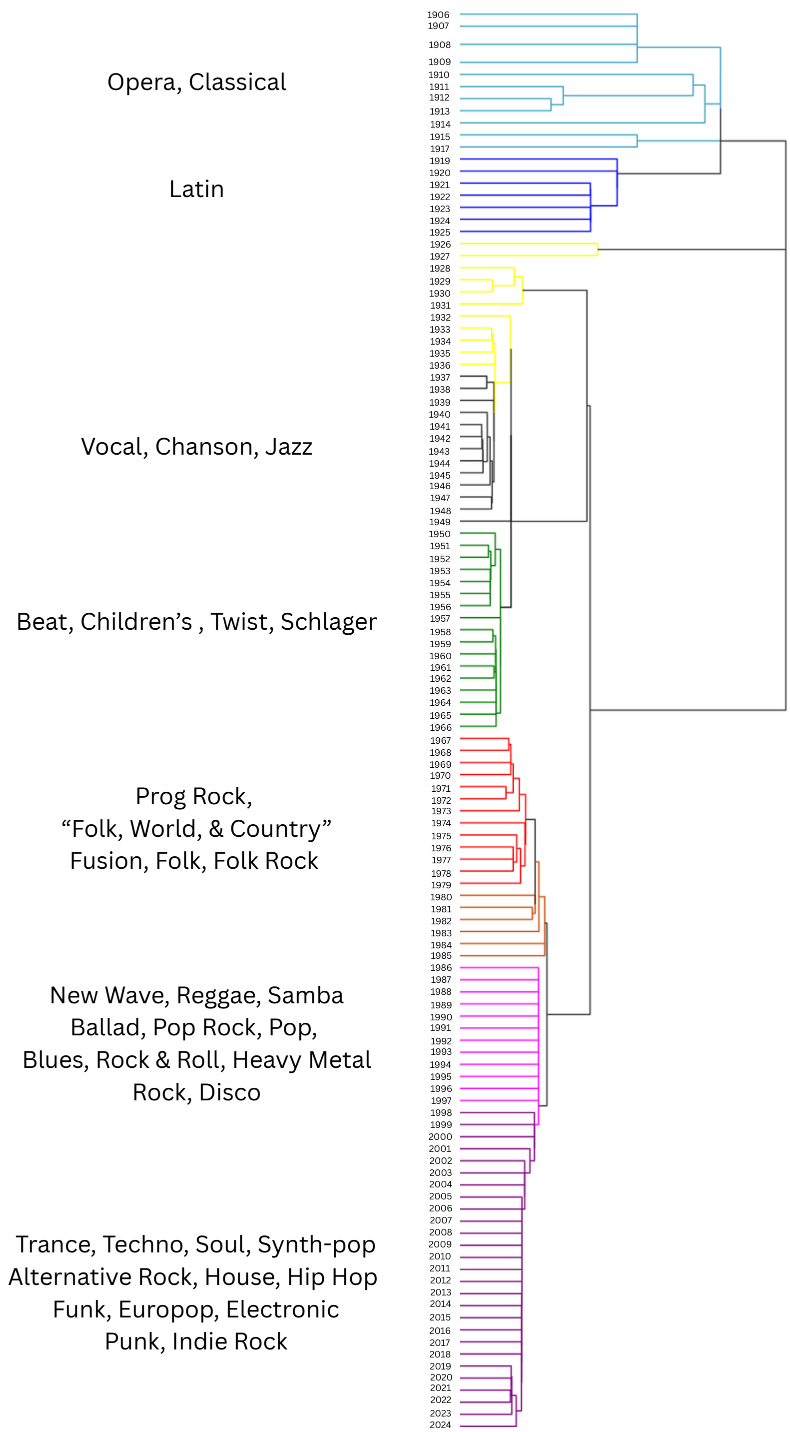
To gauge the different eras in the Danish music industry, we take a look at Figure 14. This figure shows a dendrogram where the years are clustered based on their network similarity in terms of musical activity and shared genre characteristics. The branches have been colored to highlight the distinct eras.

In the Italian dendrogram it was observed that part of the early and late periods suffered from data sparsity. This was not observed for the Danish dataset, partly because we extracted data at a later date. The Italian network is also noted to be very fragmented in the years 1938-1945 (during the Second World War) which is not surprising given the political situation in the country at the time. However, this is not something we observe in the Danish network, suggesting that Denmark was overall less disrupted by the war.

To interpret the dendrogram we follow the same logic that we used for detecting the eras. The standard approach would be to compare counts of activities across clusters, but this does not make use of the network structure. Instead we calculate the network distance between a node attribute representing the era and each genre. The era's node attribute gives us a normalized count of records that every band has released within that era. The attributes are normalized so that each era attribute sums to one, this is to avoid having the most active eras dominate the analysis.

The genres that are most closely associated with each era are displayed to the left of the dendrogram. These genres are not necessarily the most dominant in terms of overall popularity, but a characterization of what genres are unique to each era. For example, the earliest Danish era is closely linked with opera and classical music, while the middle of the century is linked to beat, schlager, and folk genres. We see that the 1980's and 1990's, are associated with disco, heavy metal, and trance and that transitions into hip hop, house, and alternative rock for the year 2000 until today. The most recent era also shows proximity to punk, indie rock, and electronic subgenres, giving an indication of a diverse musical scene in Denmark.

Figure 14: Eras Dendrogram



4.6 Explaining The Network

In **Coscia (2024)** the author performs two linear regression analyses to provide some evidence in the favor of the homophily assumption. The first regression aims to explain if nodes are connected based on their shared attributes. In other words, do bands that share the same genre, label and years of activity tend to connect. The second regression aims to explain the strength of the connection between two nodes, given that there is one, based on the node attributes. So, do bands that share the aforementioned attributes tend to connect more. For the genre and temporal clusters to be meaningful, it is important that bands, that share genres and active years, actually collaborate together. In **Coscia (2024)** the author also includes data for regions in the analysis as it should be more likely for bands from the same area to share members. Because Denmark is a much smaller country with only four regions we forego this analysis and instead include data on record labels. Thus our analysis only differs in that it includes labels instead of regions. If two bands have released at least one album on the same label, the label similarity is set to be 1.

An important note here is that we tried to replicate the original regression analysis for Italy using region data and we get a slightly lower R^2 value of about 4 percentage points. We are not sure as to why this happens, but decided to move ahead with the results we got in our project to keep things as comparable as possible.

We perform the regressions for both Denmark and Italy. The first regression tries to explain the likelihood of an edge to exist in the network. Just as in **Coscia (2024)** it is defined by the following equation:

$$Y_{u,v} = \beta_0 + \beta_1 G_{u,v} + \beta_2 L_{u,v} + \beta_3 T_{u,v} + \epsilon$$

- $Y_{u,v}$ is a binary variable, equal to 1 if bands u and v shared at least one member, and zero otherwise.
- $G_{u,v}$ is the genre similarity, which is the cosine similarity between the vectors recording how many records of a given genre bands u and v have published
- $L_{u,v}$ is the label similarity, equal to 1 if the bands have published a record on the same label, and zero otherwise
- $T_{u,v}$ is the temporal similarity, in which we take the logarithm of the number of years in which both bands released a record, plus one to counter the issue when the bands did not share a year. We take the logarithm because having at least the one collaboration is more important than consecutive collaborations
- β_0 and ϵ are the intercept and the residuals

Note that $Y_{u,v}$ contains all links with a weight of at least one, which includes those that are not statistically significant and where dropped with NC backboning for the visualizations and previous analyses. Thus $Y_{u,v}$ also contain all non-links. The non-links are included because the aim of the regression is to distinguish between links and non-links. However, given that the network is sparse, it is not feasible to have all non-links in the regression. To give an idea of the magnitude: Table 1 shows that the Danish network has 26,176 links. However, the amount of potential links, where n is the amount of nodes, is

$\frac{n*(n-1)}{2} = \frac{416*(416-1)}{2} = 86,320$, which is more than 3 times as many non-links than links. While it might be possible to include all non-links for the Danish network as it is now, it would quickly increase if we included more bands. The Italian network has 2,992,681 total possible links which is not feasible to include.

Thus, we perform a balanced negative sampling. For each link that exists we sample and include in $Y_{u,v}$ a link that does not. For $G_{u,v}$ we only consider the most popular 90 genres, since less popular genres would make bands more similar than what they would otherwise be.

Table 7: The regression results from the two models predicting the existence of a link (Exists) and its weight (Size)

Dependent variable	**Exists (DK)**	**Size (DK)**	**Exists (IT)**	**Size (IT)**
Genre	0.482***	0.842***	0.696***	0.721***
	(0.022)	(0.044)	(0.008)	(0.014)
Label	0.092***	0.163***	0.147***	0.227***
	(0.010)	(0.021)	(0.003)	(0.006)
Year	0.248***	0.366***	0.256***	0.307***
	(0.005)	(0.011)	(0.002)	(0.003)
Constant	0.296***	0.277***	0.260***	0.133***
	(0.005)	(0.013)	(0.001)	(0.004)
Observations	19,032	9,516	167,516	83,758
R^2	0.176	0.154	0.243	0.156
Adjusted R^2	0.175	0.154	0.243	0.156
Residual Std. Error	0.454	0.767	0.435	0.703
	(df = 19028)	(df = 9512)	(df = 167512)	(df = 83754)
F Statistic	1,351.095***	578.164***	17,881.160***	5,179.265***
	(df = 3; 19028)	(df = 3; 9512)	(df = 3; 167512)	(df = 3; 83754)

Note: *p < 0.1; **p < 0.05; ***p < 0.01

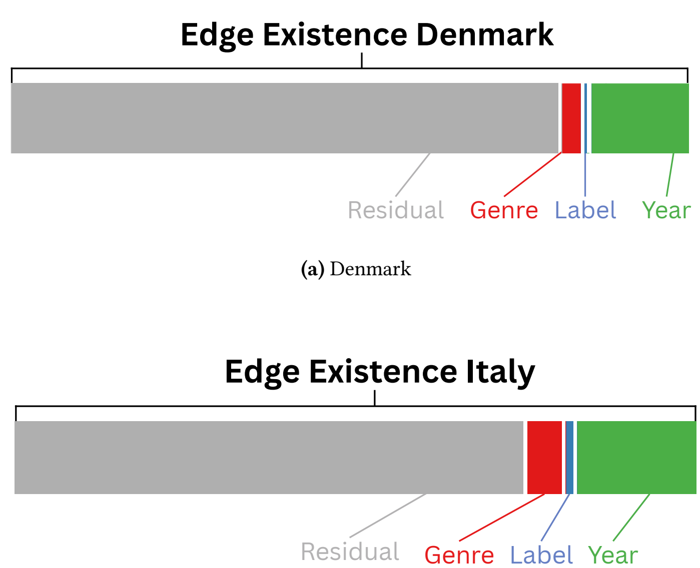
The results for the first regression can be seen in column one for Denmark and column three for Italy in Table 7. The first thing to notice is that we find an R^2 of 17.6% for Denmark and an R^2 of 24.3% for Italy. It is interesting that we can explain less of the variance in the Danish network. Compared to the Italian network we can explain 6.7 percentage points less of the variance. Some possible explanations follow here. It could simply be that the Italian network explains more variance because it is around six times bigger and therefore captures more information. It could also be related to the size of the countries. Denmark is a rather small country and the majority of musical activity will take place in a few large cities such as Copenhagen, Aalborg, Aarhus and Odense. Because there are fewer hotspots it could be that musicians collaborate more across genres, making it harder to predict connections. This also relates to what was discussed in Section 4.1.2 about genre specialization being more prominent in Italy.

We see less variance explained by the labels than we anticipated (see Figure 15), which could be because the effect from them is already contained in the other explanatory variables. It is likely that at least some of the variance explained by labels is already contained in genres and time, given that some labels focus on a selection of genres and that bands have to exist at the same time to collaborate.

Since most bands also release music with big labels it might be that we just over estimated the likelihood of playing together given you are on the same label, as big labels release music from many bands of different genres. Even if a black metal band and a pop band are on the same label it is unlikely that they would collaborate. And even though we see in the nestedness and cluster analyses that e.g. labels focusing on metal genres exist (see Section 5), there are still many labels focusing on specific genres, meaning that bands with a high degree of collaboration might still release under different labels.

Although we are unable to explain 82.4% of the variance in the Danish network with our dataset, and 75.7% of the variance in the Italian network, we still find some evidence for the homophily assumptions, like in **Coscia (2024)**. As shown in Table 7 all estimated coefficients are positive and statistically significant. In **Coscia (2024)** the author does a decomposition of the explanatory variables contributions to the R^2 to estimate their relative importance [27] [28]. We have done the same which can be seen in Figure 15. From this comparison it is interesting to see, that although the labels still explain very little variance they do explain more in the Italian network than the Danish one. The estimated coefficient for the label variable is even slightly larger than the estimated coefficient for the region variable in the original paper, indicating that being on the same label increases the likelihood of collaboration more than being from the same region.

Figure 15: Edge existence in Denmark and Italy



(a) Denmark

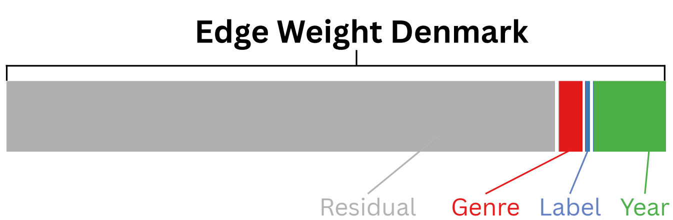
(b) Italy

The second regression tries to predict the strength of a connection between two bands, which is quantified by the edge weight in our network.

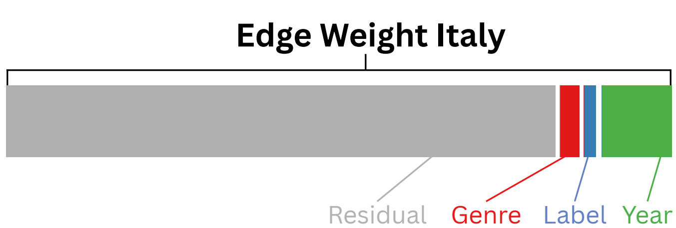
$$log(W_{u,v}) = \beta_0 + \beta_1 G_{u,v} + \beta_2 L_{u,v} + \beta_3 T_{u,v} + \epsilon$$

The explanatory variables remain the same as before but the dependent variable is now the logarithm of the edge weight. This model only cares about edges with a weight higher than zero as the previous model already tries to predict edge existence. The results can be seen in Table 7 column two and four. Again all estimated coefficients are positive and significant, so we expect bands to collaborate more if they are active in the same years, play the same genre and release under the same label. For Italy, we see the same relative drop in R^2 as in **Coscia (2024)** and the same increase in coefficients. Interestingly, and contrary to the Italian network, the R^2 for Denmark only falls by 2.2 percentage points, meaning we have almost the same predictive power for $log(W_{u,v})$ and $Y_{u,v}$ in the Danish network. Figure 16 shows the R^2 decomposition for $log(W_{u,v})$. For both countries all explanatory variables are reduced. For Denmark, genre and label similarity explain more variance here relative to the temporal variable. For Italy we see the label similarity explaining more variance relative to the other variables. When comparing these results with the original results for Italy, it seems that being on the same label matters more than being from the same region.

Figure 16: Edge weight in Denmark and Italy



(a) Denmark



(b) Italy

While we demonstrate here that overlap in node attributes such as genre, year and label can explain some of the structure in collaborations, it can not capture all of it. Earlier research on networks show that social factors that are not captured by such variables also play a big role. One paper worthy of mention is *Community Structure In Jazz* by Gleiser and Danon (2003)[4] where they analyze American jazz bands. They show that there is a clear divide e.g. between different racial groups and different cities. This suggests that even though the bands in the jazz network share some attributes like genre and year there are still other social structures that play a role in determining collaboration.

5 Label Analysis

5.1 Genre And Label Nestedness

To get a better understanding of how the music industry is structured we can look at our data on labels and genres. By creating a binary matrix that encodes a 1 if a label has released an album under a given genre we can gauge at the amount of specialization among labels. Any label with less than five releases is removed as we decided to not consider these as significant labels. To quantify this specialization in a measure we use nestedness. The measure of nestedness describes the degree with which the nodes in a network are subsets of each other. In the interaction between record labels and the genres they publish, the nestedness measure can tell us if smaller labels output the same genres as big labels or if they tend to specialize and do something different.

The two most used measures for nestedness are the Nestedness Temperature Calculator (NTC) [29] and the Nestedness metric based on Overlap and Decreasing Fill (NODF) [30]. The measure we chose to use is the NODF because it is less sensitive to matrix size and density than NTC, so we can more easily compare results between Denmark and Italy. We compute the nestedness by first making a binary matrix with genres on the x-axis and labels on the y-axis. If a label has a release containing a genre we put a 1 otherwise a 0. The rows and columns are then sorted in descending order by their sum. On the website bimat.github.io [9] NODF is defined as:

$$N_{\text{NODF}} = \frac{\sum_{ij} M_{ij}^{\text{row}} + \sum_{ij} M_{ij}^{\text{col}}}{\frac{m(m-1)}{2} + \frac{n(n-1)}{2}}$$

$$M_{ij}^{\text{row}} = \begin{cases} 0, & \text{if } k_i \leq k_j \\ \frac{n_{ij}}{\min(k_i, k_j)}, & \text{if } k_i > k_j \end{cases}$$

where:

- m is the number of rows (e.g., record labels)
- n is the number of columns (e.g., genres).
- k_i is the total number of connections (sum of ones) in row i
- n_{ij} is the number of shared connections between row i and row j

[9] https://bimat.github.io/alg/nestedness.html

We obtain the following nestedness measures between labels and genres for Denmark and Italy:

Denmark: 0.34

Italy: 0.31

This tells us that both countries only have a moderate amount of nestedness meaning that some smaller labels focus on a subset of the genres that bigger labels do, but also that there are labels that focus on unique genres, breaking up the nestedness pattern.

Figure 17: Genre-Label nestedness for Denmark. The full matrix has dimensions 528*270, this is a zoomed in look

Figure 17 shows a zoomed in view of our nestedness matrix where we can see the gradient with the most present label at the top of the y-axis and the most present genre at the start of the x-axis.

To test whether our results are statistically significant we created 1,000 matrices with the same dimensions and density as ours. These were filled randomly with 1's and 0's and the nestedness measure for all of these were calculated. We could have done a more robust test by ensuring that the row and column sums were constant and the same in the random matrices, as in our original matrix. But this randomized approach still provides a useful baseline. We find that the average nestedness measure for all of these is close to zero indicating that our results are meaningful. We also perform a one sided t-test were we compare the observed nestedness for Denmark (0.31) against the distribution of nestedness values that we get from the randomized matrices. For Denmark, we get a p-value of 0.00, meaning that the observed nestedness is statistically significant when compared to the random nestedness distribution.

Traditionally the measure of nestedness has been applied mostly to ecological and biological networks. Studies on biological networks have shown nestedness to be a recurring pattern that indicates stability and resilience in a network. By comparing the nestedness in our network with those found in biological ones we can determine if cultural and biological networks exhibit some of the same underlying structures. One such example is the paper *Nestedness across biological scales* [31]. In that study nestedness was analyzed in different biological and ecological networks such as food webs and molecule structures. They report NODF measures in the range 0.2 to 0.6 depending on the networks. Compared to the label-genre matrix a nestedness of 0.31 suggests a moderate hierarchical structure with a mix of generalists and specialists. This aligns with findings in ecology, where nestedness reflects stability and structured interactions and might indicate that similar self organizing principles underlie the evolution of both cultural and ecological systems.

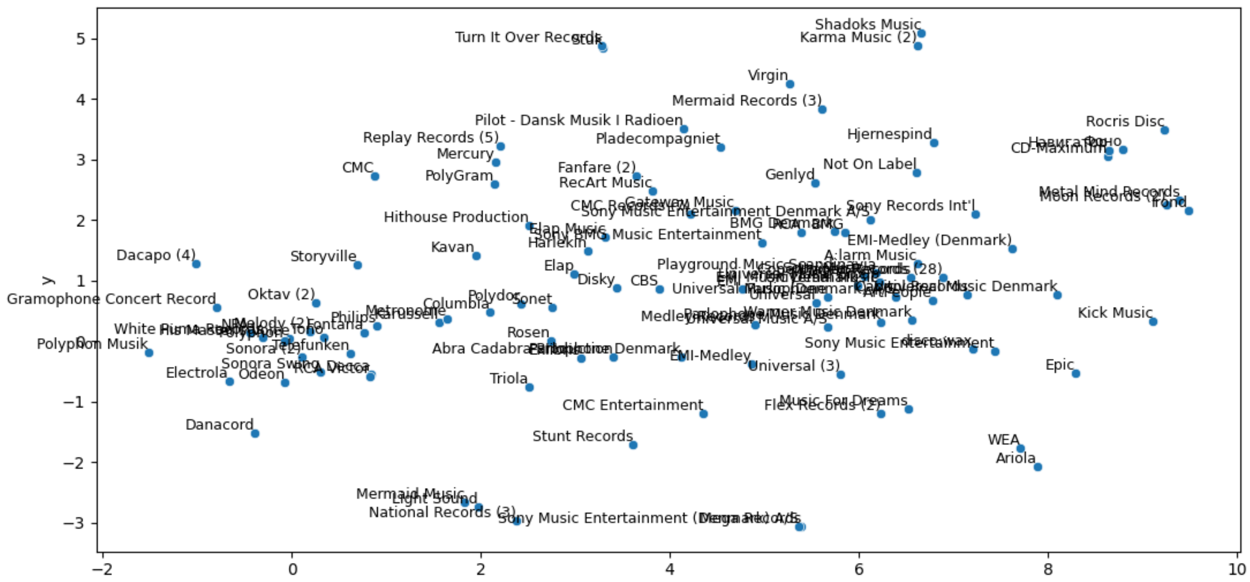
5.2 Label Clusters

To further explore patterns among record labels, we also analyzed the pairwise similarity between labels based on the bands that have released records under them. We made a matrix that counts how many times a band has released a record under each label. This matrix was normalized to avoid having big bands and labels dominate as they will have the highest absolute release counts. Furthermore we remove all bands that have less than five releases as we consider them too small to be important.

Next, we projected the labels and bands onto a graph, where the labels are connected if they have released a record from the same band. To gauge how similar two labels are we calculate a label to label distance matrix using the method described in Section 3.2. This matrix cannot be visualized so we reduce it to two dimensions using t-SNE [18]. The result can be seen in Figure 18. Each point in the projection represents a label, with closer points corresponding to more similar labels.

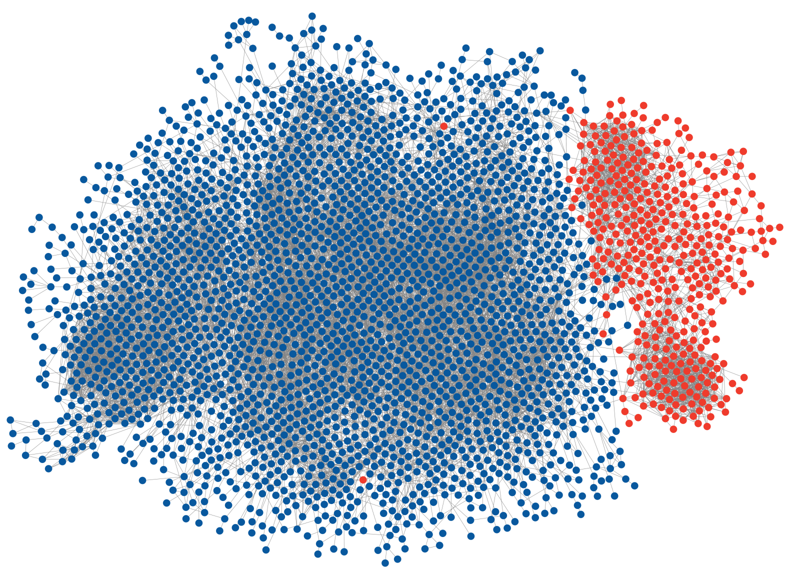
We do not see major clustering tendencies between the labels, but still some clusters emerge, mainly two, with labels scattered around them. First we notice a rather big cluster approximately in the middle of the Figure. This cluster consists mainly of big labels (or sub labels hereof) like Universal, EMI Music and Sony. In the left side of the Figure we see a small cluster which consists of early 1900's record labels like His Masters Voice, Odeon and Polyphon, again showing the temporal aspect of our network. There is also a tendency for small groups of 3 to 4 labels that specialize in more niche genres like heavy metal to be close together in our two dimensional space.

Figure 18: Label to label clusters using t-sne



6 Analysis Of The Combined Danish-Italian Network

Figure 19: Network with Italian and Danish bands. Italian bands are blue and Danish bands are red



The Danish and Italian networks share the same data model, as explained in Section 2.1. This shared model offers a key advantage: the two networks can be combined if at least one artist has contributed to both a Danish and an Italian band. The resulting network becomes a single connected component, revealing new patterns and offering deeper insights. It enables an analysis that goes beyond national boundaries, highlighting inter-country relationships and supporting a more general, less country-specific exploration.

Fortunately, the Italian and Danish networks share not just one, but 273 artists, distributed across multiple bands. This means we are not dealing with a single cut-vertex holding the entire network together, but rather a well-connected structure. Although the merged network still exhibits two distinct clusters, it remains well integrated through several shared nodes.

To obtain a final and manageable network, we applied the Noise-Corrected (NC) backboning method, as described in Section 3.7. For this merged network, we used a higher threshold than in the Danish network, due to the greater number of potential nodes and edges. After applying the filtering, we identified 89 bands from one country that are connected to at least one band from the other. Although this represents less than 0.5% of the full network, these bands play a crucial role as bridges between the Italian and Danish music industries.

When dividing the network into the two communities shown in Figure 19 and calculating the modularity based on this partition, we obtain a modularity score of 0.2596. This indicates the presence of a community structure, though not a particularly strong one. The existence of international connections helps explain the moderately low modularity. However, it is visually evident that the two music industries do not form a fully integrated network. The merged network exhibits an assortativity coefficient of 0.977, indicating a very strong tendency for bands to collaborate predominantly within their own national communities. This suggests that the Danish and Italian music industries operate largely independently, with only limited cross-national integration. In this section, we focus on understanding these international relations.

There are two Danish bands in Figure 19 whose nodes are visually distinct, as they are located within the Italian cluster. As explained in Figure 20, these artists are the house and techno DJ and producer René Kristensen, also known as Noir, and the Danish singer and actress Gitte Hænning.

René Kristensen has collaborated on several occasions with the Italian DJ and producer Simone Vitullo, which links Noir to him and to two other Italian DJ's. He has only one connection to another Danish band: the DJ and producer Trentemøller. This explains Noir's position in the network. However, due to his low degree, we consider him an under-representative node, as discussed in Section 6.3.

Gitte Hænning, on the other hand, presents a particularly interesting case, which we explore in more detail in Section 6.3.1.

Figure 20: Highlight of two Danish nodes that lay within the Italian cluster.**(1)** corresponds to Noir and **(2)** corresponds to Gitte Hænning

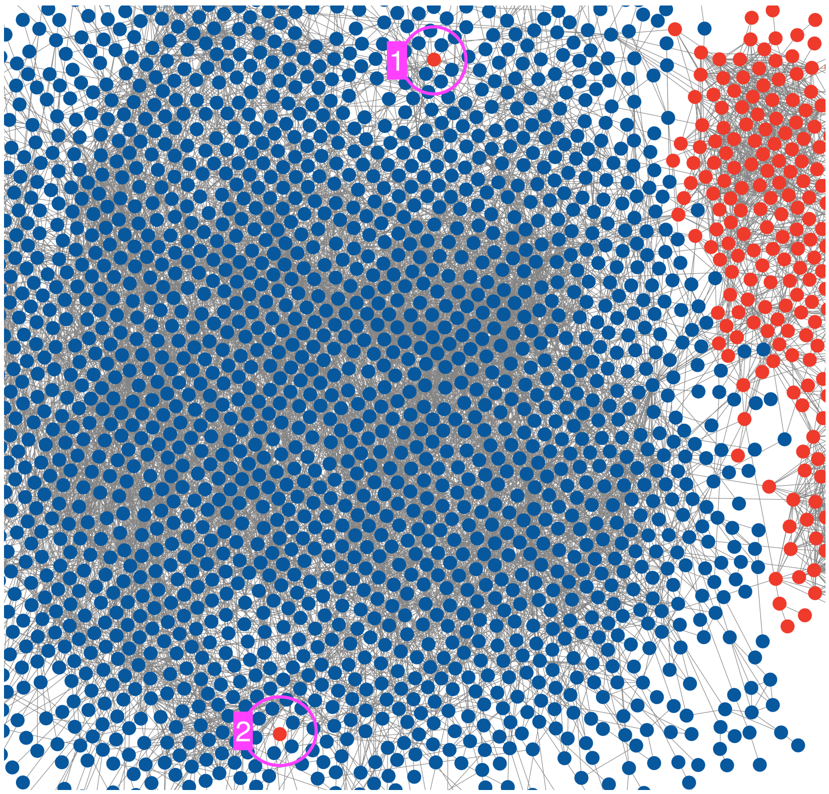


Table 8 highlights the top 10 bands based on degree, closeness, and betweenness centrality. The most prominent node in the network is easily identifiable: the Italian heavy metal band Temperance. Formed in 2013, Temperance is known for their modern melodic power metal sound, blending heavy guitar riffs, electronic elements, and folk influences into powerful, melodic compositions[32].

Table 8: Top 10 bands in the band projection (Italian and Danish) according to different centrality measures. Bands appearing in multiple measures are shown in bold.

#	Degree	Closeness	Betweenness
1	**Temperance (ITA)**	**Temperance (ITA)**	**Temperance (ITA)**
2	Tino Vailati (ITA)	Hautville (ITA)	Robertino (ITA)
3	**Genius (ITA)**	**Gli Avvoltoi (ITA)**	Alex Puddu (ITA)
4	Odyssea (ITA)	**Genius (ITA)**	Efterklang (DK)
5	Beniamino Gigli (ITA)	Lantern (ITA)	Dino Lenny (ITA)
6	Oscar Carboni (ITA)	Confusional Quartet (ITA)	Volbeat (DK)
7	Death SS (ITA)	**Armonite (ITA)**	Iceage (DK)
8	Marnero (ITA)	Secret Sphere (ITA)	**Gli Avvoltoi (ITA)**
9	Radio Boys (ITA)	**Stormo (ITA)**	**Armonite (ITA)**
10	**Stormo (ITA)**	Dino Lenny (ITA)	Christian Arhoff (DK)

Temperance's high centrality can largely be attributed to their collaborations with well-connected Italian artists such as Simone Mularoni and Alessandro Conti, both of whom have worked extensively with Italian metal bands, as well as the Danish producer Jacob Hansen, who has collaborated with numerous Italian and Danish bands. These connections have earned Temperance a total of 39 links: 29 with Italian bands and 10 with Danish bands. This unique position places them at the core of the network and justifies their high betweenness centrality, as they effectively act as a key bridge between the two national music scenes.

Interestingly, several Italian rock bands with no direct connection to the Danish network such as Hautville, Gli Avvoltoi, Genius, Lantern, Armonite, and Secret Sphere, appear as central nodes in the network. Their centrality is largely due to their connection with Temperance, which serves as their gateway to the rest of the network.

In this way, Temperance's role as a frequent shortest path between otherwise distant bands increases not only its own centrality metrics but also boosts the closeness and betweenness centrality of its neighbors, even if those neighbors don't directly link the two national networks themselves.

Other important artists, such as Robertino (see Section 6.3.1) and Alex Puddu (see Section 6.3.2), are discussed in more detail in the sections below.

To better understand how the Italian and Danish networks connect and interact, we introduce two key measures: the Cross-Country Score and the Shannon Entropy Score [33]. These metrics are explained in the following sections and help to quantify the degree and diversity of cross-national collaborations within the network.

6.1 Cross-Country Score

This metric captures the proportion of a band's connections that link it to bands of the opposite nationality. A higher score indicates a stronger connection with bands from the other country, highlighting the band's potential role in bridging national music scenes.

The cross-country score of a band is defined as:

$$S_b = \frac{|E_b^{\mathrm{opp}}|}{|E_b^{\mathrm{opp}}| + |E_b^{\mathrm{same}}|}$$

where:

- S_b is the score of band b that ranges between 0 and 1.
- $|E_b^{\mathrm{opp}}|$ is the number of edges connecting band b to nodes of the opposite nationality.
- $|E_b^{\mathrm{same}}|$ is the number of edges connecting band b to nodes of the same nationality.
- $|E_b^{\mathrm{opp}}| + |E_b^{\mathrm{same}}|$ is equal to $\deg(b)$

In simpler terms, S is just the proportion (normalized percentage) of edges that connect a band b to bands of the opposite nationality.

The score ranges from 0 to 1. A value close to 0 indicates that the band is mostly connected to others from the same country, while a value close to 1 means that most of its connections are cross-national.

By examining bands with high cross-country scores, we can identify key nodes in the network that promote integration and connectivity. This also helps in understanding how closely knit the Italian and Danish music industries are, and which bands act as hubs in this cross-border collaboration.

It is important to note that the cross-country score is designed specifically for binary nationality networks since it assumes that bands belong to one of only two national categories. As such, it effectively captures the extent to which a band connects across this binary divide, providing meaningful insights into the role of a band in linking, for example, the Italian and Danish music scenes.

However, this metric does not generalize well to networks that include more than two nationalities. If a third country were introduced into the network, the cross-country score would no longer be able to express, in a single value, the extent to which a band bridges multiple national contexts. In such cases, it becomes insufficient, as it can only compare two categories at a time. It fails to account for the more complex, multi-dimensional connectivity that arises in a multi-national setting.

While our current network is binary, containing only Italian and Danish bands, it is valuable to consider measures that can scale beyond this dual-country context. To facilitate future inclusion of additional nationalities, we also compute a more general measure: the Shannon entropy.

6.2 Shannon Entropy

This measure allows us to quantify how evenly a band's connections are distributed across multiple nationalities, offering a scalable approach for analyzing inter-country connectivity in more complex networks.

As it was said, our network is binary, but it can still use the Shannon entropy formula. The resulting score represents the diversity in the nationality of a bands connected peers (Danish vs. Italian), where low values mean connections are country specific and high values mean connections are balanced across both countries.

The Shannon entropy is defined as:

$$H(X) = -\sum_{i=1}^{n} p_i \log_n p_i$$

where:

- $H(X)$ denotes the Shannon entropy of a band X
- n is the number of distinct band nationalities, $n = 2$ in our case.
- p_i is the proportion of connections from band X to bands of nationality i.

The metric reflects the distribution of a band's connections across nationalities. An entropy value of 0 indicates that all connections are with artists of a single nationality, signifying no diversity. A value of 1 reflects an even split between nationalities, indicating a highly diverse and internationally integrated network.

In this context, Shannon entropy captures the degree of cross-national engagement exhibited by each band within the network. Higher entropy values indicate a more balanced distribution of connections across the Danish and Italian music scenes, suggesting stronger international integration. Conversely, lower values reflect a concentration of collaborations within a single nationality, pointing to a more locally focused network structure.

As previously mentioned, our analysis is limited to two nationalities: Danish and Italian. However, if additional nationalities were to be incorporated into the network, the Shannon entropy could still be computed using the same formula. By increasing the value of n, the number of nationalities, in the $\log_n(p_i)$ term, the entropy measure remains normalized within the interval $[0, 1]$, ensuring comparability across networks with different numbers of nationalities.

6.3 The Connections

Since the average degree of the network is ≈ 9, we decided to apply a degree threshold of 5, about half the average, for nodes to be considered in our analysis. This threshold helps to filter out unrepresentative nodes that might appear disproportionately important due to their low connectivity.

For instance, the Italian band *Al Bano & Romina* has a degree of only 2, with one connection to an Italian band and one to a Danish band. As a result, it receives a Shannon entropy score of 1, suggesting a perfectly balanced and highly significant cross-national position. However, this is misleading: the score appears high simply because of the very limited number of connections, not because of any substantial bridging role in the network. By setting a degree threshold, we aim to reduce such distortions and focus our analysis on more structurally relevant bands.

Table 9 presents the top ten bands with the highest scores in both Cross-Country Rank and Shannon Entropy.

Table 9: Top 10 bands ranked by Cross-Country Score and Shannon Entropy.

#	Cross-Country	Shannon Entropy
1	Pathosray (ITA) – 0.92	Noha (ITA) – 0.97
2	Gitte Hænning (DK) – 0.83	Efterklang (DK) – 0.90
3	Fleshgod Apocalypse (ITA) – 0.77	Campo di Marte (ITA) – 0.86
4	Noha (ITA) – 0.60	Temperance (ITA) – 0.82
5	Efterklang (DK) – 0.31	Soulmagic (DK) – 0.81
6	Campo di Marte (ITA) – 0.29	Fleshgod Apocalypse (ITA) – 0.78
7	Temperance (ITA) – 0.26	Alex Puddu (ITA) – 0.78
8	Soulmagic (DK) – 0.25	Robertino (ITA) – 0.78
9	Alex Puddu (ITA) – 0.23	Mercyful Fate (DK) – 0.76
10	Robertino (ITA) – 0.23	Pretty Maids (DK) – 0.75

Denmark and the Nordic countries more broadly are well known for their strong rock and heavy metal scenes. While Italy has approximately 7.76 metal bands per 100,000 inhabitants, Denmark has a significantly higher rate of 12.65, which is 62% greater [34]. This highlights the cultural and industrial importance of metal music in Denmark and may contribute to Danish metal artists being more internationally active and visible.

Key figures in the Danish metal industry, such as producers Tommy Hansen and Jacob Hansen, have collaborated with several Italian rock and metal bands (mainly in production roles), including Pathosray, Fleshgod Apocalypse, and Temperance. These collaborations serve as bridges between the Danish and Italian music scenes, helping to explain some of the cross-national connections observed in our network.

6.3.1 Under The Wing Of Sejr Volmer-Sørensen

The Danish actress and singer Gitte Hænning, who rose to fame as a child star in the 1950s, has a high cross-country score, as most of her connections are with Italian bands. She gained early popularity in Scandinavia and later achieved significant success in Germany, Italy, and across the rest of Europe. She recorded songs in multiple languages, including Italian, which contributed to her broad international appeal.

Another artist who appears prominently in both rankings is the Italian singer Robertino. He was discovered at the age of 13 while singing in Roman restaurants to help support his family. His talent caught the attention of a television producer, who would become his mentor.

These two artists share a significant connection: Sejr Volmer-Sørensen. He not only produced and wrote songs for Gitte Hænning such as *Jeg Snakker Med Mig Selv*, but also played a pivotal role in launching Robertino's international career. According to their biographies, it was Volmer-Sørensen who met Robertino while on holiday in Italy with his wife and subsequently brought him to Copenhagen to record and perform. Meanwhile, Gitte Hænning herself was promoted in Italy and even performed in Italian [35] [36].

Interestingly, although Sejr Volmer-Sørensen does appear in our network, he is not directly connected to either of the two artists. This may be due to a lack of properly credited relationships in the data, particularly given that these connections date back to the 1950s. However, the shortest path between Gitte Hænning and Robertino in the network is just two steps, indicating that despite the missing direct link, the underlying structure of the network still captures the proximity of their relationship.

The case of Gitte Hænning and Robertino is a perfect example of how the network, and the scoring methods we applied, effectively reflects real-world interconnections even when historical data is incomplete. It is also important to highlight Robertino's significant role in the network, as evidenced by his second-place ranking in betweenness centrality (see Table 8).

6.3.2 The Alex Puddu Case

Alex Puddu is one of the most interesting nodes in our network, as his biography represents a textbook case of an Italian–Danish musician.

Born in Rome, Italy, in 1967, Puddu relocated to Copenhagen in 1987. A multi-instrumentalist, he writes, arranges, and produces his own music, blending elements of jazz, soul, funk, disco, lounge, Afrobeat, and Latin rhythms into a unique and eclectic style.

Given his many years living in Denmark, frequent collaborations with Danish artists, and numerous releases within the Danish music scene, it was initially difficult to determine whether Alex Puddu should be classified as part of the Italian or Danish network. Professionally, he has built his career in Denmark, often recording albums and working with Scandinavian musicians. However, his musical DNA is unmistakably Italian, rooted in the lush soundtracks of 1960s and 1970s Italian cinema, drawing clear inspiration from composers such as Ennio Morricone and having worked with the famous Italian singer Edda Dell'Orso. [37]

In fact, Puddu appears in both networks: he was originally included in the Italian network created by **Coscia (2024)** and also appears in the Danish network built by us [9]. Since our methodology required assigning a single nationality to each node for statistical and structural consistency, we faced the challenge of choosing one.

To resolve this ambiguity, we contacted the artist directly via social media. To our delight, he kindly responded and shared his own view: “I see myself as an Italian–Danish artist.” While this reinforced his dual identity, we still needed to assign him a single nationality for the purposes of network construction. Based on his additional comment: “When I listen to my latest works in Italian, there’s no doubt that musically and lyrically—if I can put it that way—I’m more part of the Italian funk and nu-disco niche”, we ultimately chose to classify him as an Italian artist. The full message from Alex Puddu can be found in the Appendix11.1.

We acknowledge that a single-nationality classification oversimplifies complex identities like Alex Puddu’s. Nevertheless, given that such cases are rare within our dataset, we opted for a pragmatic, if slightly arbitrary, decision rather than redesigning the network model entirely. Alex Puddu’s case, however, highlights the richness and nuance of cross-cultural musical identities and how they can challenge rigid classification systems.

As an Italian band, Alex Puddu has a Cross-Country score of 0.23 and ranks third in betweenness centrality (see Table 8), highlighting both his significant connections with Danish bands and his crucial role within the overall network structure. He also holds a Shannon Entropy score of 0.78, one of the highest in the entire network. As discussed in the Shannon Entropy Section 6.2, this high score indicates considerable diversity in his connections and a significant level of uncertainty when predicting the nationality of a potential new edge. In other words, if a new link were to form with Alex Puddu, it would be statistically more difficult to classify that connection as either Italian or Danish.

This stands in contrast to other artists such as, for example, the Danish singer Birthe Wilke, who has a much lower Shannon Entropy score of 0.24. In her case, the nationality of a potential new connection would be easier to predict, likely remaining within her own country’s network.

Since Alex Puddu’s most prominent links are already divided between the Italian and Danish music industries, we believe that incorporating additional nationalities into the network would likely lower his Shannon Entropy score. This is because the inclusion of more countries would increase the theoretical maximum entropy, making his existing diversity appear less significant in comparison.

7 Discussion

While we believe our work to be meaningful our analysis and methodology is not without issues. As mentioned, the data we work with is user generated. This leads to a lot of noise in the data as there is nothing to safeguard against mistakes, misleading entries, and incompleteness. We mitigated the issue with thorough data cleaning but we can not ensure the absence of inconsistencies. Instead, we assume it. We found this to be especially true for the data on record labels, so our results should be carefully interpreted. Our results might also be inconclusive as our database is not complete. Due to the difficulty of finding a complete list of all existing Danish bands and the fact that we collect data semi-manually, we have not extracted all the data that would capture the entirety of the Danish music industry. This could lead to niche genres and small artists being underrepresented in the network. We also recognize that the methods we use require rich metadata [8] and that they would not be useful for scarce networks. However, cultural networks are often rich in metadata. Many of the methods for analyzing node attributes also require some correlation between the network structure and the attributes to produce richer results. If the node attributes do not affect the probability of two nodes connecting (the homophily assumption) then the results from clustering genres and detecting musical eras are less meaningful. However, we find support for the homophily assumption as explained in the Section 4.6.

We could have further expanded on our results by quantifying some of our hypotheses in the following ways. We theorize that bands in Denmark collaborate more across genres due to the country's smaller size. This could have been further explored by calculating the average genre information entropy per artist and see if this is higher in Denmark than for Italy. When looking at the genre clusters we could quantify the agreement of genre clusters in both countries. The clusters form due to some threshold. If we establish such a threshold to cut the dendrogram then we could calculate the normalized mutual information (NMI) between the clusters from Denmark and Italy. A high NMI value would indicate that Denmark and Italy group genres in a similar way. The same could be done for the eras discovery. If we get a low NMI it would indicate that the musical history of the two countries developed in different ways. However, this is constrained by the fact the we always cluster nearby years, which would likely give us a high NMI.

Future research could expand by adding more networks for other countries. This could give more insight into how genres cluster and evolve through time among different countries. It could also lead to more interesting analyses across countries, seeing which (if any) artists bridge more than two countries and if e.g. Denmark's music industry is closer to Norway's than Italy's. In this paper we also extended the original analysis by including record label data as a node attribute. Future research could explore even more node attributes such as: Instruments, streaming metadata and lyrical content. As an examples, it could be explored if some instruments provide grounds for more collaboration and connecting different genres. By including a bands monthly listeners on streaming services the homophily assumption could be further explored (do popular bands connect more with other popular bands?). A bands lyrical content could be encoded and it could be explored how that affects connections.

8 Concluding Remarks

This thesis explored the structure and evolution of the Danish music industry. It built on the work done in **Coscia(2024)** [8] by applying the same methods to a new dataset and it extended its perspective by including data on record labels and by connecting the Danish and Italian networks. This was done by constructing a bipartite band-artist network and applying methods for graph and node attribute analyses.

We found that, in some aspects, the Danish music network differs significantly from the Italian network, especially when it comes to genre specialization. We discovered that the bands in the Danish network exhibit lower genre modularity and assortativity, indicating a higher tendency for collaboration across genres and a more fluid and interconnected musical ecosystem. However, we still see some tendency for bands to cluster based on their attributes like genre and specially by year of activity. The high temporal assortativity that we found confirms that time is an universal organizing principle in music collaboration. The temporal analysis highlighted distinct eras in the Danish music scene, with clear transitions between dominating genres.

The data we extracted on labels provided a new layer of insight. In our analysis we found that the relationship between labels and genres exhibit a nested structure that is otherwise mostly found in ecological networks. It also provided insight when explaining why bands collaborate together when used in the regression analysis. Although the label data explained less of the variance than we expected it interestingly explains more of the variance on edge weights for Italy than the regions did.

By combining the Danish and Italian networks, we got a unique view into international music collaborations, identifying central bands that bridge the two countries.

More than descriptive findings, it is our belief that this thesis also contributes to reinforce the use of network analysis methods such as backboning, node attribute variance, and node attribute distance as useful tools for cultural analysis.

9 Acknowledgments and Database Repository

For this paper we have used Grammarly (https://www.grammarly.com/) as a tool to review spelling, grammar, and punctuation errors. It was also used to suggest improvements in sentence structure and word choice. We reviewed the final text to ensure it met academic standards.

The databases created and used for this project can be found in this *repository.*

10 References

[1] Miller McPherson, Lynn Smith-Lovin, and James M Cook. "Birds of a feather: Homophily in social networks". In: *Annual review of sociology* 27.1 (2001), pp. 415–444.

[2] James Moody. "Race, School Integration, and Friendship Segregation in America". In: *American Journal of Sociology* 107.3 (2001), pp. 679–716. DOI: `10.1086/338954`. URL: `https://doi.org/10.1086/338954`.

[3] Marilena Hohmann, Karel Devriendt, and Michele Coscia. "Quantifying Ideological Polarization on a Network Using Generalized Euclidean Distance". In: *Science Advances* 9.9 (2023), eabq2044. DOI: `10.1126/sciadv.abq2044`. URL: `https://doi.org/10.1126/sciadv.abq2044`.

[4] Pablo Gleiser and Leon Danon. "Community Structure in Jazz". en. In: *Advances in Complex Systems* 06.04 (Dec. 2003). arXiv:cond-mat/0307434, pp. 565–573. ISSN: 0219-5259, 1793-6802. DOI: `10.1142/S0219525903001067`. URL: `http://arxiv.org/abs/cond-mat/0307434` (visited on 05/06/2025).

[5] Joan Serrà et al. "Measuring the Evolution of Contemporary Western Popular Music". In: *Scientific Reports* 2 (2012), p. 521. DOI: 10.1038/srep00521. URL: https://doi.org/10.1038/srep00521.

[6] R. Lambiotte and M. Ausloos. "Uncovering Collective Listening Habits and Music Genres in Bipartite Networks". In: *Physical Review E* 72.6 (2005), p. 066107. DOI: 10.1103/PhysRevE.72.066107. URL: https://doi.org/10.1103/PhysRevE.72.066107.

[7] Tomas Teitelbaum et al. "Community structures and role detection in music networks". In: *Chaos: An Interdisciplinary Journal of Nonlinear Science* 18.4 (2008), p. 043109.

[8] Michele Coscia. "Node attribute analysis for cultural data analytics: a case study on Italian XX–XXI century music". en. In: *Applied Network Science* 9.1 (Sept. 2024), p. 56. ISSN: 2364-8228. DOI: 10.1007/s41109-024-00669-5. URL: https://appliednetsci.springeropen.com/articles/10.1007/s41109-024-00669-5 (visited on 05/06/2025).

[9] Lisandro Marco Benetti and Daniel Fejerskov-Quist. *Building A Band Artist Network*. en. 2024.

[10] John W. Ratcliff and David E. Metzener. "Pattern Matching: The Gestalt Approach". In: *Dr. Dobb's Journal* 13.7 (July 1988), p. 46. URL: https://www.drdobbs.com/database/pattern-matching-the-gestalt-approach/184407970?pgno=5.

[11] Tao Zhou et al. "Bipartite network projection and personal recommendation". en. In: *Physical Review E* 76.4 (Oct. 2007), p. 046115. ISSN: 1539-3755, 1550-2376. DOI: 10.1103/PhysRevE.76.046115. URL: https://link.aps.org/doi/10.1103/PhysRevE.76.046115 (visited on 05/25/2025).

[12] Michele Coscia and CS Department. *Supplementary Material for "Node Attribute Analysis for Cultural Data Analytics: a Case Study on Italian XX-XXI Century Music"*. en.

[13] Karel Devriendt, Samuel Martin-Gutierrez, and Renaud Lambiotte. "Variance and Covariance of Distributions on Graphs". en. In: *SIAM Review* 64.2 (May 2022), pp. 343–359. ISSN: 0036-1445, 1095-7200. DOI: 10.1137/20M1361328. URL: https://epubs.siam.org/doi/10.1137/20M1361328 (visited on 05/06/2025).

[14] Karel Devriendt. "Effective resistance is more than distance: Laplacians, Simplices and the Schur complement". In: *Linear Algebra and its Applications* 639 (Apr. 2022). arXiv:2010.04521 [math], pp. 24–49. ISSN: 00243795. DOI: 10.1016/j.laa.2022.01.002. URL: http://arxiv.org/abs/2010.04521 (visited on 05/06/2025).

[15] Michele Coscia. *The Atlas for the Aspiring Network Scientist*. 2021. DOI: 10.48550/arXiv.2101.00863. arXiv: 2101.00863 [cs.CY]. URL: https://arxiv.org/abs/2101.00863.

[16] Michele Coscia. "Generalized Euclidean Measure to Estimate Network Distances". en. In: *Proceedings of the International AAAI Conference on Web and Social Media* 14 (May 2020), pp. 119–129. ISSN: 2334-0770, 2162-3449. DOI: 10.1609/icwsm.v14i1.7284. URL: https://ojs.aaai.org/index.php/ICWSM/article/view/7284 (visited on 05/06/2025).

[17] Richard Bellman. *Dynamic Programming*. Reprint of the 1957 edition. Mineola, NY: Courier Corporation, 2013. URL: https://books.google.dk/books/about/Dynamic_Programming.html?id=CG7CAgAAQBAJ.

[18] Laurens van der Maaten and Geoffrey Hinton. “Visualizing Data using t-SNE”. In: *Journal of Machine Learning Research* 9.86 (2008), pp. 2579–2605. URL: http://jmlr.org/papers/v9/vandermaaten08a.html.

[19] Joe H. Jr. Ward. “Hierarchical Grouping to Optimize an Objective Function”. In: *Journal of the American Statistical Association* 58.301 (1963), pp. 236–244. DOI: 10.1080/01621459.1963.10500845.

[20] M. E. J. Newman. “Modularity and community structure in networks”. In: *Proceedings of the National Academy of Sciences* 103.23 (2006), pp. 8577–8582. DOI: 10.1073/pnas.0601602103.

[21] M. E. J. Newman and M. Girvan. “Finding and evaluating community structure in networks”. In: *Physical Review E* 69.2 (2004), p. 026113. DOI: 10.1103/PhysRevE.69.026113.

[22] Michele Coscia and Frank Neffke. *Network Backboning with Noisy Data*. en. arXiv:1701.07336 [physics]. Jan. 2017. DOI: 10.48550/arXiv.1701.07336. URL: http://arxiv.org/abs/1701.07336 (visited on 05/06/2025).

[23] Daniel Grady, Christian Thiemann, and Dirk Brockmann. “Robust classification of salient links in complex networks”. en. In: *Nature Communications* 3.1 (May 2012). Publisher: Nature Publishing Group, p. 864. ISSN: 2041-1723. DOI: 10.1038/ncomms1847. URL: https://www.nature.com/articles/ncomms1847 (visited on 05/23/2025).

[24] Paul B. Slater. “A two-stage algorithm for extracting the multiscale backbone of complex weighted networks”. In: *Proceedings of the National Academy of Sciences* 106.26 (June 2009). arXiv:0904.4863 [physics]. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.0904725106. URL: http://arxiv.org/abs/0904.4863 (visited on 05/23/2025).

[25] M. Angeles Serrano, Marian Boguna, and Alessandro Vespignani. *Extracting the multiscale backbone of complex weighted networks*. en. Apr. 2009. DOI: 10.1073/pnas.0808904106. URL: https://arxiv.org/abs/0904.2389v1 (visited on 05/21/2025).

[26] Oxford University Press. *fiddling (noun) – Oxford Learner’s Dictionaries*. Accessed: 2025-05-31. 2025. URL: https://www.oxfordlearnersdictionaries.com/definition/english/fiddling?q=Fiddling.

[27] Barry E. Feldman. “Relative Importance and Value”. en. In: *SSRN Electronic Journal* (2005). ISSN: 1556-5068. DOI: 10.2139/ssrn.2255827. URL: http://www.ssrn.com/abstract=2255827 (visited on 05/06/2025).

[28] Ulrike Grömping. “Relative Importance for Linear Regression in *R* : The Package **relaimpo**”. en. In: *Journal of Statistical Software* 17.1 (2006). ISSN: 1548-7660. DOI: 10.18637/jss.v017.i01. URL: http://www.jstatsoft.org/v17/i01/ (visited on 05/06/2025).

[29] Wirt Atmar and Bruce D. Patterson. “The measure of order and disorder in the distribution of species in fragmented habitat”. In: *Oecologia* 96.3 (Dec. 1993), pp. 373–382. ISSN: 1432-1939. DOI: 10.1007/BF00317508. URL: https://doi.org/10.1007/BF00317508.

[30] Mário Almeida-Neto et al. “A consistent metric for nestedness analysis in ecological systems: reconciling concept and measurement”. In: *Oikos* 117 (2008), pp. 1227–1239. URL: https://api.semanticscholar.org/CorpusID:9050949.

[31] Mauricio Cantor et al. “Nestedness across biological scales”. en. In: *PLOS ONE* 12.2 (Feb. 2017). Ed. by Irene Sendiña-Nadal, e0171691. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0171691. URL: https://dx.plos.org/10.1371/journal.pone.0171691 (visited on 05/06/2025).

[32] Temperance. *Biography*. Accessed: 2025-06-01. 2023. URL: https://www.temperanceband.com/biography.

[33] C. E. Shannon. “A mathematical theory of communication”. In: *The Bell System Technical Journal* 27.3 (July 1948), pp. 379–423. ISSN: 0005-8580. DOI: 10.1002/j.1538-7305.1948.tb01338.x. URL: https://ieeexplore.ieee.org/abstract/document/6773024/citations (visited on 05/23/2025).

[34] Caitlin Dempsey. *Interactive Map of Heavy Metal Bands By Country Per 100,000 People*. en-US. Mar. 2012. URL: https://www.geographyrealm.com/map-of-heavy-metal-bands-by-country-per-capita/ (visited on 05/20/2025).

[35] Manuel Ghilarducci. “Soviet Estrada and VIAs in Italian Boots (1960s–1970s)”. In: *Apparatus. Film, Media and Digital Cultures of Central and Eastern Europe* 13 (2021), pp. 36–54. DOI: 10.17892/app.2021.00013.272. URL: https://www.apparatusjournal.net/index.php/apparatus/article/view/272.

[36] *Gitte Hænning*. en-GB. URL: https://www.eurovisionuniverse.com/encyclopedia/gitte-haenning/ (visited on 05/20/2025).

[37] *The Mover feat. Joe Bataan, by Alex Puddu Soultiger*. en. URL: https://parisdjs.bandcamp.com/track/the-mover-feat-joe-bataan (visited on 05/20/2025).

11 Appendix

11.1 A.1: A Message From Alex Puddu

ciao

beh in effetti non e’ cosi facile ! nel senso dal punto di vista di come si vedono le cose! io sono italiano nato e cresciuto a roma, pero’ sono cresciuto e diventato uomo un danimarca , anche se le primissime esperienze musicali sono iniziate in italia , la mia carriera cosi come tutta la mia discografia e’ partita dalla danimarca, ma sono un artista internazionale i miei dischi si trovano anche in australia o in sudamerica dunque io mi vedo cone un artista italiano danese e’ chiaro che se poi ascolto i miei ultimi lavori in lingua italiana, non ce’ dubbio che sia musicalmente e liricalmente parlando se si puo dire, faccio piu parte della nicchia italiana funk e nu disco! in danimarca la mia musica e’ poco considerata anche se sanno chi sono e cosa ho fatto nel passato in tutti i casi mi piace l idea come la mia doppia nazionalita’ che rappresento oggi sia l italia che la danimarca !

Translation:

Hi,

well, actually it’s not that easy! I mean, in terms of how things are seen! I’m Italian, born and raised in Rome, but I grew up and became a man in Denmark. Even though my very first musical experiences started in Italy, my career—as well as all my discography—began in Denmark. But I'm an international artist—my records can even be found in Australia or South America. So I see myself as an Italian-Danish artist. Of course, when I listen to my latest works in Italian, there's no doubt that musically and lyrically—if I can put it that way—I'm more part of the Italian funk and nu-disco niche! In Denmark, my music isn't given much attention, even though people know who I am and what I've done in the past. In any case, I like the idea that today my dual nationality—representing both Italy and Denmark—is something I stand for!

11.2 A.2: Example Of How We Encode The Data

Table 10: Genres and Labels per release

band	release_id	artist_id	label_id	label_name	genre1	genre2	style1	style2
Alex Puddu	3471062	527548.0	399500.0	Al Dente	Rock	Funk / Soul	Funk	Experimental
D-A-D	3468330	272991.0	6580.0	Liberation Records	Rock		Country Rock	Hard Rock
Volbeat	4174860	537021.0	81846.0	EMI Music	Rock		Rock & Roll	Heavy Metal

Table 11: Links in the network

artist	band	year	role
Alex Puddu	Puddu Varano	2001	Electric Guitar
Alex Puddu	Filur	2000	Producer
Alex Puddu	Alex Puddu	2022	Electric Guitar, Synthesizer

Table 12: Encoding of Genre as a Node Attribute

artist_id	Rock	Pop	Tango	Electronic	Death Metal	Math Rock	Goregrind
2488562	4	1	0	0	1	6	1
1032769	2	2	0	1	0	2	0
377590	0	7	1	0	0	1	0

Table 13: Encoding of Label as a Node Attribute

artist_id	Al Dente	Reek Of Death Records	EMI Music Denmark	Medley Records	Columbia	EMI	Musik Venner
2488562	9	0	1	0	1	3	1
1032769	1	0	8	1	0	3	0
377590	0	7	1	2	0	0	9

Table 14: Encoding of years as a Node Attribute

artist_id	1906	1907	1908	1909	1910	1911	1912
2488562	0	1	0	0	1	0	0
1032769	1	0	0	1	0	0	0
377590	0	0	1	0	0	1	0