Statistical Analysis of a Jazz Musicians Network

Samuel Ritchie, Antoine Roy

May 18, 2018

Abstract

In this work, we aim to understand how the network of jazz musicians is structured, based on our intuition that jazz music presents a specificity compared to other styles of music, due to its history strongly based on improvisation, which remained among the years and its progressive division in many sub-genres. Among various conclusions, we especially found evidence of the presence of homophily in the formation of connections in the network of jazz musicians.

Introduction

Jazz is a music genre that originated in the African-American communities of New Orleans, United States, in the late 19^{th} and early 20^{th} centuries, and developed from roots in blues and ragtime. As jazz spread around the world, it drew on different national, regional, and local musical cultures, which gave rise to many distinctive styles. Even if this progressive diversification of jazz and its induced subgenres make it relatively difficult to define, one of its key elements undoubtedly is improvisation, which closely related to its origins in blues and work-songs of African-American slaves in plantations.

This element remained across jazz evolution. A natural consequence of this improvisation aspect is that jazz musicians have often played with a large number of their fellows, changing bands during their careers or also occasionally joining friends in orchestras or for jam[1] sessions in bars. This high connectivity between artists appears to be a specificity in jazz, as compared to Rock music where the identification to a band is often very strong.

A natural representation of jazz community to efficiently render this alleged connectivity among jazz-men thus appears to be as a network, where each vertex corresponds to a musician, and two musicians are connected if they have collaborated during their careers. In this work, we will precisely study the community structure of the collaboration network of jazz musicians and try to answer to some specific questions. Who are the central artists ? What are the main drivers of the composition of the band (living area, instrument, style) ? Can we isolate specific communities ? In order to answer these questions, we will add some metadata for each musician, describing their personal characteristics. The whole analyzing process, from data scrapping to data cleaning and eventually descriptive and predictive analysis can be found on the repository GitHub created for this work at the adress `https://github.com/SamuelRitchie/SAND`.

Our work will divide in three parts organized as follows. After providing some introductory insights on data sources and cleaning, we will lead a first descriptive analysis of the network to identify its global structure, as well as the most central / influential artists. In a second part, we will implement a clustering approach in order to find potential sub-communities in the network and understand their drivers. To finish, in a Section 3 we will introduce some predictive analysis tools, to predict on the one side the probability two jazz-men would have to collaborate given their characteristics, and on the other side try to recover missing information of a jazz-man thanks to his connectivity in the network.

[1] Explicitly, to "jam" is to improvise music without extensive preparation or predefined arrangements, except for when the group is playing well-known jazz standards

1 Data

1.1 Sources

The main source that helped us to build the network comes from the *Linked Jazz* project (https://linkedjazz.org/). The project draws on jazz history materials in digital format to expose relationships between musicians and reveal their community network. Researchers utilized over 50 transcripts of oral history interviews. The original interviews come from the Hamilton College Jazz Archive, Rutgers Institute for Jazz Studies Archives, Smithsonian Jazz Oral Histories, UCLA's Central Avenue Sounds Series, and the University of Michigan's Nathaniel C. Standifer Video Archive of Oral History. Connections between musicians are decided thanks to Linked Jazz 52nd Street, a crowd-sourcing tool that allows jazz experts and enthusiasts to assist Linked Jazz project team in deciding what type of relationship two individuals share based on interview transcripts. Several levels of relationship exist, from *"Knows"* to *"Collaborated with"*. As the network is still being constructed (analyzing audio and video transcripts takes some time, approximately 50% has been analyzed at the time this report was written), we kept the highest level of relationship between jazz-men, i.e. *"Knows"*, the dataset not being big enough for relations *"Collaborated with"*. Consequently, a connection in our network *does not* necessarily mean they have played together, which is an important fact. However, we will sometimes refer to the "reduced" dataset, which designs the network of musicians who have effectively played together during their career. This subgraph will enable us to try to understand the differences in relations between *knowing* another jazz-man, and having played with him.
To annotate the network, we scrapped DBPedia, the Linked Open Data version of Wikipedia, in order to fetch additional personal information on musicians such as Birth Place, Birth Year, Death Place, Death Year, Jazz Style and Instrument Played. An important data cleaning process was led before the network analysis. The output of this data cleaning step led to four main characteristics per jazz-man :

1. the (main) instrument played
2. the (main) jazz-style of the artist
3. the living area, that we derived from the death and birth city (priority was given to death city) and divided into categories New-Orleans, Chicago, New-York-City (the main cities were jazz developed) and the four regions of United Stats if not in the latter.
4. the birth decade

1.2 Descriptive Analysis

1.2.1 Presentation of the network

The network consists in the connection between 980 American musicians born during the 20^{th} century. We count 2876 connection between them. Here is a a representation of the network, each dote is a musician.

Figure 1: Whole Network

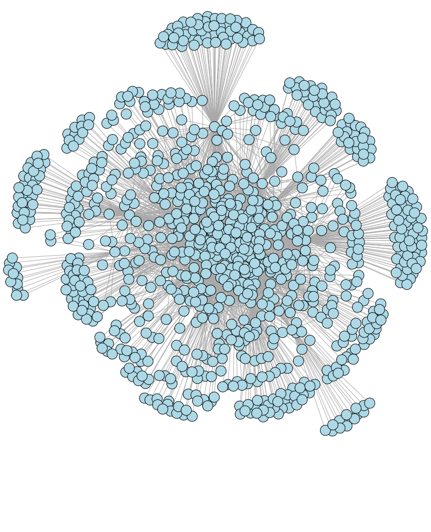
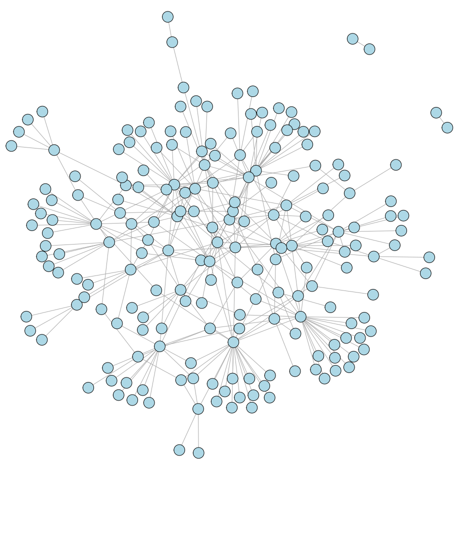


Figure 2: Musicians who played together



We can see a very dense subgraph in the middle and many edges around that are linked to only one other musician. This structure is partly due to the way the data was collected and the lack of information for some musicians. However, the structure is also representative for the jazz-men network with highly connected musicians who are likely to be the best and most famous and many other musicians who navigate around this dense subgraph. This structure is common in the art communities as few artists are very well known and very connected and others are connected only to one or two of these artists.

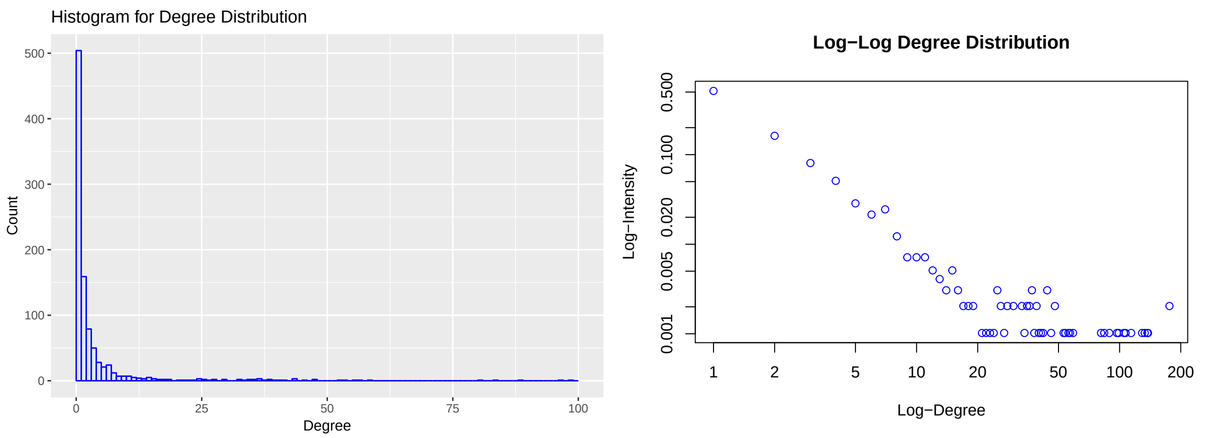
In our case, the centre of the graph is dense meaning that there are more than few people who have many connections. As said in the introduction, jazz musicians are used to play in many different bands which can favour the development of such a network with many musicians playing with many other musicians. Thus we need to inspect the structure of the network in more details to understand the underlying mechanism of the connection of musicians.

To complete our analysis, we used an additional subgraph made up of the musicians who played together. We did not used it as our main network for our analysis as it is very incomplete and limited in terms of interpretation. However, it might be useful to understand some underlying mechanisms in link connection as the level of homophily might be higher than in the whole network (for example musicians who play together are more likely to be in the same city). This network is composed of 199 musicians with 309 links and a mean degree of 3.06.

1.2.2 Degree

Firstly, we are going to inspect the degree of the musicians, i.e. the number of connections for each musician within the network. It gives us an idea about the density of the network and whether musicians are isolated or connected in the graph.

Figure 3: Degree Distribution



As expected given the representation of the network, many musicians are connected to only one other musician (these are the clusters we can see at the edge of the graph). The distribution is unequal and we can see that some musicians have many connections. In average, musicians have about 6 connections but, almost 50% of them have only one connection. Thus, the variance of the degree distribution is 18. As a result, the network is fragile, 44 musicians can split the graph into two subgraphs.

The following table presents the degree distribution for the most connected musicians, the graph represents the whole network and the network of the most connected musicians.

Table 1: Most connected musicians

Name	**Degree**
Benny Golson	176
Louie Bellson	176
Billy Taylor	138
Danny Barker	137
J. J. Johnson	133
Roy Haynes	129
Chico Hamilton	114
Artie Shaw	107
Gerald Wiggins	105

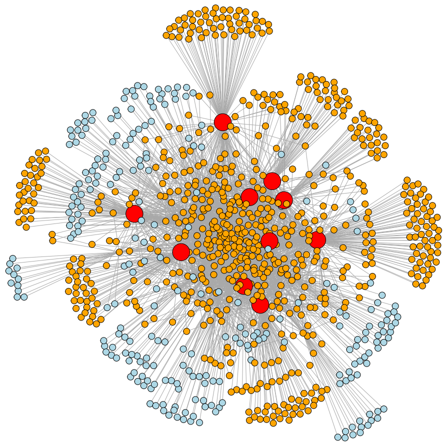


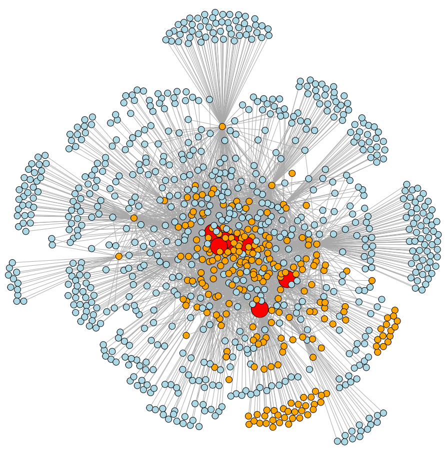
Figure 4: Representation of the Network *(Most connected artists in red, their 1st-degree network in orange)*

We can see that the most connected musicians have many connections. However, most of them are connected to musicians for whom they are the only connection. Thus, we can see that large part of their 1st-degree network consists in clusters of musicians connected uniquely to them. Moreover, we can see that they are not in the centre of the graph. Thus we can expect that their role is not so central in the network although their connections cover almost all the network.

1.2.3 Connectivity

In this section, we want to describe the structure of the network and the way musicians are connected between them. This can give us information about the formation of bands but also about the way that musicians can influence each other.

Figure 5: Representation of the Network *(Most influential artists in red, their 1st-degree network in orange)*



Name	**Rank** *(Closeness Centrality Index)*	**Degree**
Duke Ellington	1	37
Charlie Parker	2	36
Louis Armstrong	3	28
Count Basie	4	35
Dizzy Gillespie	5	30
Benny Golson	6	176
Benny Goodman	7	26
Lester Young	8	27
John Coltrane	9	22
Quincy Jones	10	89

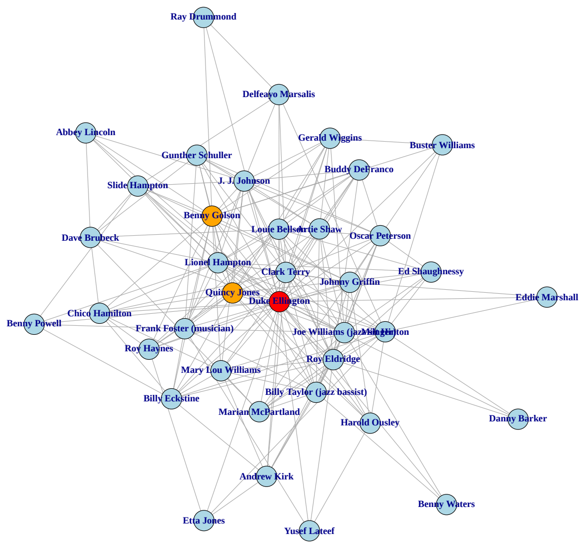
Table 2: Most influential musicians

The closeness centrality measures attempt to capture the notion that a vertex is ‘central’. The most central artists are those who are the closest to all the other musicians of the network. Thus, in our case, it is an indicator of the capacity of a musician to influence others. Most influential artists are different from the one with the highest degree. This means that the number of connection is not a necessary condition for being an influential musician.

The central component of the graph is in fact mostly composed of the network of the most influential artists. As opposed to before, their 1st-degree network is almost not composed of the clusters with single-linked musicians. Thus, they are characterized by the fact that they connect with everybody with a short path.

We can have a look at Duke Ellington Network who is the musician the most influential in our network, which is consistent with the jazz history as his orchestra was one of the most notorious at his time. We can see that he does not have so many connections (34), but his network is very dense, all the musicians he is connected with are connected with at least 3 musicians within his network. He is also connected with two of 10 most influential jazz-men (*in orange*). His network is central in the constitution of the whole network. That is why we can consider him as the most important actor of this network.

Figure 6: Duke Ellington Network



In Duke Ellington's network, we can recognize musicians with whom we formed different bands and for whom he composed some music. We can identify cliques that can be associated to some bands. Thus, the table below gives use the distribution of the cliques depending on the number of musicians considered. We can also observe that the maximum size of a clique in this network is 6, this is also often the maximum size of a band.

We can interpret cliques as band of musicians but also as group of musicians who know each other and may have played together in a jam for example. Given the structure of our data, we are not able to give a clear sense of what is a clique.

Table 3: Distribution of the Cliques

Cliques	1	2	3	4	5	6
Number	980	2876	3050	1649	363	28

2 Clustering

2.1 Description

Previous analysis of the network has shown some clustering in the structure of the network. Thus, it might be useful to conduct a deeper analysis of the clustering within the network using a hierarchical approach to detect communities. This clustering algorithm produces 11 groups for which we gives the distribution below. The allocation is balanced, between 50 and 80 musicians in average in each cluster, except for cluster 1 (191 musicians) and cluster 8 (20 musicians).

Table 4: Cluster Allocation

Cluster number	1	2	3	4	5	6	7	8	9	10	11
Number of Musicians	191	48	116	78	56	48	167	20	112	77	67

The graphical representation in Appendix shows that most isolated groups of musicians with only one connection are grouped in different clusters depending on the musicians they are connected to. Then in the middle of the graph, we can see that many clusters coexist. Thus, we need to analyze in details the clusters to see whether they are homogeneous.

2.2 Adding attributes

In order to study the homogeneity of the cluster created thanks to the algorithm, we inspect the distribution of the different variables within each cluster. Difference in distribution between each cluster can indicate that some variables may have driven the connections of musicians within the network. In the table below we show the most relevant results (the distribution of the variables within those clusters is different from the whole sample). Table in Appendix with results from all 11 clusters is available for the interested reader.

Table 5: Three most frequent labels for each variable within clusters 4, 5 and 11

	Total	**cluster4**	**cluster5**	**cluster11**
Area 1	*New-York-City (30.2%)*	New-Orleans (25.4%)	New-York-City (30.2%)	Chicago (30%)
Area 2	*South (19.8%)*	New-York-City (23.8%)	South (27.9%)	New-York-City (26%)
Area 3	*Chicago (14%)*	South (20.6%)	West (20.9%)	South (24%)
Style 1	*Jazz (37.4%)*	Jazz (54.8%)	Jazz (34.2%)	Jazz (30.2%)
Style 2	*Swing_music (14.3%)*	Dixieland (9.5%)	Bebop (13.2%)	Jazz_fusion (16.3%)
Style 3	*Bebop (11.4%)*	Rhythm_and_blues (7.1%)	Cool_jazz (10.5%)	Hard_bop (9.3%)
Birth Decade 1	*1920s (26.4%)*	1900s (32.9%)	1920s (42.6%)	1930s (29.7%)
Birth Decade 2	*1930s (19.7%)*	1910s (19.7%)	1930s (22.2%)	1940s (28.1%)
Birth Decade 3	*1910s (15.1%)*	1890s (18.4%)	1940s (18.5%)	1920s (10.9%)
Instrument 1	*Other (20.8%)*	Other (23.1%)	Other (19.6%)	Other (28.4%)
Instrument 2	*Piano (16.3%)*	Trumpet (15.4%)	Drums (17.9%)	Piano (19.4%)
Instrument 3	*Saxophone (11.1%)*	Guitar (11.5%)	Double Bass (10.7%)	Saxophone (17.9%)

Cluster 4 is an extremely interesting cluster that could be defined as the "historical" cluster. Indeed, it is composed of jazz-men who come from New-Orleans (25%), where blues (which partially led to jazz) was created. Due to the closing of the seaport in New Orleans, musicians were forced to travel up the Mississippi to find work. Two of the cities most affected by this move were Chicago and New York, which is the reason why we included them in the "main" cities regarding geographical area. It is also composed of the oldest musicians, since half of them were born before 1910.
In cluster 5, musicians differ from the whole sample because they are younger (as in cluster 11) and many of them play the drums. In cluster 11, many musicians come from Chicago (30%) while it is only 14% in the whole sample.

There might be a correlation between some variables and connections within the jazz musicians network. Clusters only based on link formation display some homogeneity depending on other variables. This calls for a deeper analysis of link formation and the way some attributes can drive connections.

3 Predictive analysis

From previous analysis, we have concluded that some variables may be correlated with link formation in the jazz musicians network. To complete our analysis, we want to go further in the analysis of this correlation through different methods.

3.1 Network formation

3.1.1 Assortativity Mixing

To measure the extent of assortative mixing (selective linking among vertices according to a certain characteristic), we used an "assortativity coefficient" which interpretation is similar to a classic correlation coefficient. We applied on all labels of the variables so that we can identify the labels which are correlated with link formation.
The results are disappointing and not in line with the conclusions we had drawn from the clustering. However, this analysis remains very limited as it is univariate and does not take into account the covariance with other variables. Below are presented the results for the instrument variable for both the full database and reduced database. Unsurprisingly, we observe that the instrument played does not impact on the fact two musicians will be linked. This is even more the case for the reduced database (see 3.2 for a more detailled justification). Instruments that have the highest coefficient in the reduced database are *Piano* and *Singer*, which seems logical insofar as these musicians are more likely to play in a jazz band or orchestra than instruments such as saxophone or guitar where there is often a unique player.
Regarding other variables, the only one presenting interesting positive coefficients is the birth decade : musicians know better their comtemporary fellows than other ones.

Table 6: Asasortativity coefficients for instrument variable

Instrument	Full database	Reduced database
Clarinet	0.02	-0.06
Composer	-0.02	-
Cornet	-0.001	-0.01
Double bass	0.56	0.05
Drums	0.16	-0.05
Guitar	0.06	0.1
Piano	0.03	0.09
Saxophone	0.06	0.05
Singer	0.09	0.05
Trombone	0.06	0.03
Trumpet	0.001	-0.01
Vibraphone	-0.01	-0.03

In order to be able to draw more robust conclusion we would need an econometric model to correct this bias, which we detail in the following subsection.

3.1.2 Explaining network formation

Graham (2014) developed a model taking into account these elements. He explains the link probability with a homopholy indicator based on covariates. The model has a logit structure and is as follow :

$$P(W_{ij} = 1|X_i = x_i, X_j = x_j) = \frac{exp(\beta_0||x_i - x_j|| + \nu_i + \nu_j)}{1 + exp(\beta_0||x_i - x_j|| + \nu_i + \nu_j}$$

where, X is a set of covariates (instrument, area, style and birth decade in our case), W_{ij} indicates whether i and j are connected, ν is a fixed effect for each individual reflecting its preference to form links. Then, $||x_i - x_j|| + \nu_i + \nu_j$ is a term reflecting homophily. The estimated coefficient $\hat{\beta_0}$ tells us whether a variable favours link formation when shared by two actors (when $\hat{\beta_0} < 0$).

The techniques to an estimator of β_0 are developed in Graham (2016) and in the associated Python code, available at https://github.com/bryangraham/netrics.

We implemented the code on our dataset using the instrument, the style, the birth decade and area as covariates in the homophiliy term. Unfortunately the method was too greedy and we did not manage to implement it on our personal laptops. Thus we are not able at this stage to implement a more precise analysis and our conclusions are limited to the findings of our descriptive analysis and simple correlation.

3.2 Predicting missing values

Besides trying to predict links and understand the drivers of the association between two musicians, another predictive analysis that is of interest when it comes to network inference is predicting possible missing values. This is particularly relevant regarding our dataset insofar as the scrapping of metadata led on DBPedia generated a high number of missing values, or similarly values that could be assimilated to missing. For instance, 37% of musicians were associated to the jazz-style *jazz* which is obviously not precise enough, and 21% of them were assigned to the instrument *Other*. The aim in this subsection is to develop a predictive algorithm which learns on the vertexes for which we know the attributes in order to infer unknown attributes of vertexes thanks to their position in the network. Assuming a good homogeneity among the vertex labels (i.e. neighbors frequently sharing the same attributes), prediction based on local similarities should provide some first satisfying results.
A simple, but often quite effective, method for producing local predictions is the nearest-neighbor method. For networks, the nearest-neighbor method centers on the calculation, for a given vertex $i \in V$, of the nearest-neighbor average

$$\frac{\sum_{j \in \mathcal{N}_i} x_j}{|\mathcal{N}_i|}$$

for quantitative attributes, or taking the most occurring label in the qualitative setting.

Our KNN algorithm is consequently trained on the complete dataset. The parameter it takes in our case is the order relatively to the vertex it wants to predict. Results achieved on the full dataset give average results (but not catastrophic given we are predicting categorical variables with at least 5 categories). However these results are biaised insofar as we are considering the highest level relationship in the dataset, as mentioned in data sources. Consequently, we led the same analysis on the reduced dataset of musicians that have *played together* during their career. When it comes to predicting the living area, the birth decade, or the jazz-style, it is unsurprisingly that we observe higher accuracy rates, excepted for one variable : *Instrument*. Once again, this seems natural : it is highly unlikely that two musicians playing the same instrument are in the same jazz band. Indeed, a jazz band is often either a trio composed of a piano player, a bass player and a drummer[2], or a quatuor composed of the latter plus a horn[3]. The only way two same instruments could play together is among larger ensembles such as orchestras, for example the notorious American Jazz Orchestra founded in New York City, active from 1986 to 1993.

Table 7: Accuracy of KNN prediction

Variable	Full dataset	Reduced dataset
Style	66.4%	74.7%
Instrument	49.7%	38.3%
Geographic Area	56%	66.3%
Birth decade	52.9%	65.9%

[2]Or a Hammond organ player, a drummer, and a third instrumentalist (either a saxophone player or an electric jazz guitarist) from 1950s onwards

[3]Horn is the generic jazz name for saxophones, trombones, trumpets, or any other wind or brass instrument commonly associated with jazz

4 Conclusion

Through this project, we aimed to understand how the network of jazz musicians was structured, based on our intuition that jazz music presented a specificity compared to all other styles of music, due to its history strongly based on improvisation, which remained among the years and its progressive division in many sub-genres. Using usual network analysis approaches appeared to be a good solution to understand this field, since one can easily represent jazz actors in a network with connections based on the fact they were brought to meet, exchange, or play together throughout their careers.

The first challenge to do so relied in data collection and cleaning. In order to have a sufficient amount of data, we decided to study all relationships between jazz-men, and not only the fact they have played together, which we modeled as a sub-graph. Metadata for each actor of the network was collected independently, with a consequent bottom-line medium quality data. In spite of this relatively poor quality data, we managed to draw several interesting conclusions from the analysis of the jazz musicians network.

First, we observed that the network was characterized by a very dense sub-graph at its centre and many musicians at its edge with only one connection. It is representative for an artists network, where we often find the most influential artists in the center, as they are able to easily connect with the rest of the musicians. Unsurprisingly, we found that that popular artists such as Duke Ellington, Charlie Parker or Louis Armstrong found themselves extremely influential in a statistical point of view, with the highest centrality indexes of the network.
Clustering analysis then highlighted several groups of musicians based on their connections, and was able to identify historical groups such as the original New-Orleans movement, directly issued from blues at the end of the 19^{th} century. It also showed that the proximity of jazz-men in terms of knowing each other logically depended on the instrument they played.
The poor quality of the labeling data was particularly visible when it came to implement predictive analysis algorithms. Assortative mixing did not show interesting correlations relatively to the variables, but confirmed that when studying the reduced dataset of jazz-men having played together, instruments were clearly not a driver of association, as opposed to the birth decade for example.
Nonetheless, when working on the prediction of missing labels, we have showed that a 1-nearest-neighbor algorithm achieved rather efficient results in terms of accuracy prediction. This performance shows that large part of the first-order network are similar in characteristics (more especially for the reduced dataset than for the full one), which the assortativity mixing coefficient did not manage to capture.

Finally, most of our results are in line with our initial thinking and assumptions and show some homophily between musicians who are connected. The idea that a band composition is essentially driven by music style, birth area and decade, even if not completely proven, was not contradicted by our analysis. Additional metadata as well as more precise information on the exact connections between artists would have been helpful to get more robust results. Besides, studying another music style such as Rock music could be interesting in order to confirm Jazz network indeed is specific compared to other musical networks.

Contributions

Samuel was in charge of web scrapping and data collection, he conducted the predictive analysis on missing labels.
Antoine was in charge of descriptive statistics, clustering and correlation analysis.

References

Graham, Bryan S. (2016). "Introduction to netrics: a Python 2.7 module for the econometric analysis of network data," (Version 1.0) [Computer program]. Available at http://bryangraham.github.io/econometrics/

Graham, Bryan S. (2014). “An econometric model of link formation with degree heterogeneity,” NBER Working Paper No. 20341.

Gleiser P. M., Danon L. (2003). “Community Structure in Jazz”, Advances in Complex System Vol. 6, No. 4

Kolaczyk E. D., Csárdi G. (2014). Statistical Analysis of Network Data with R. Springler

Appendix

Figure 7: Network with clusters

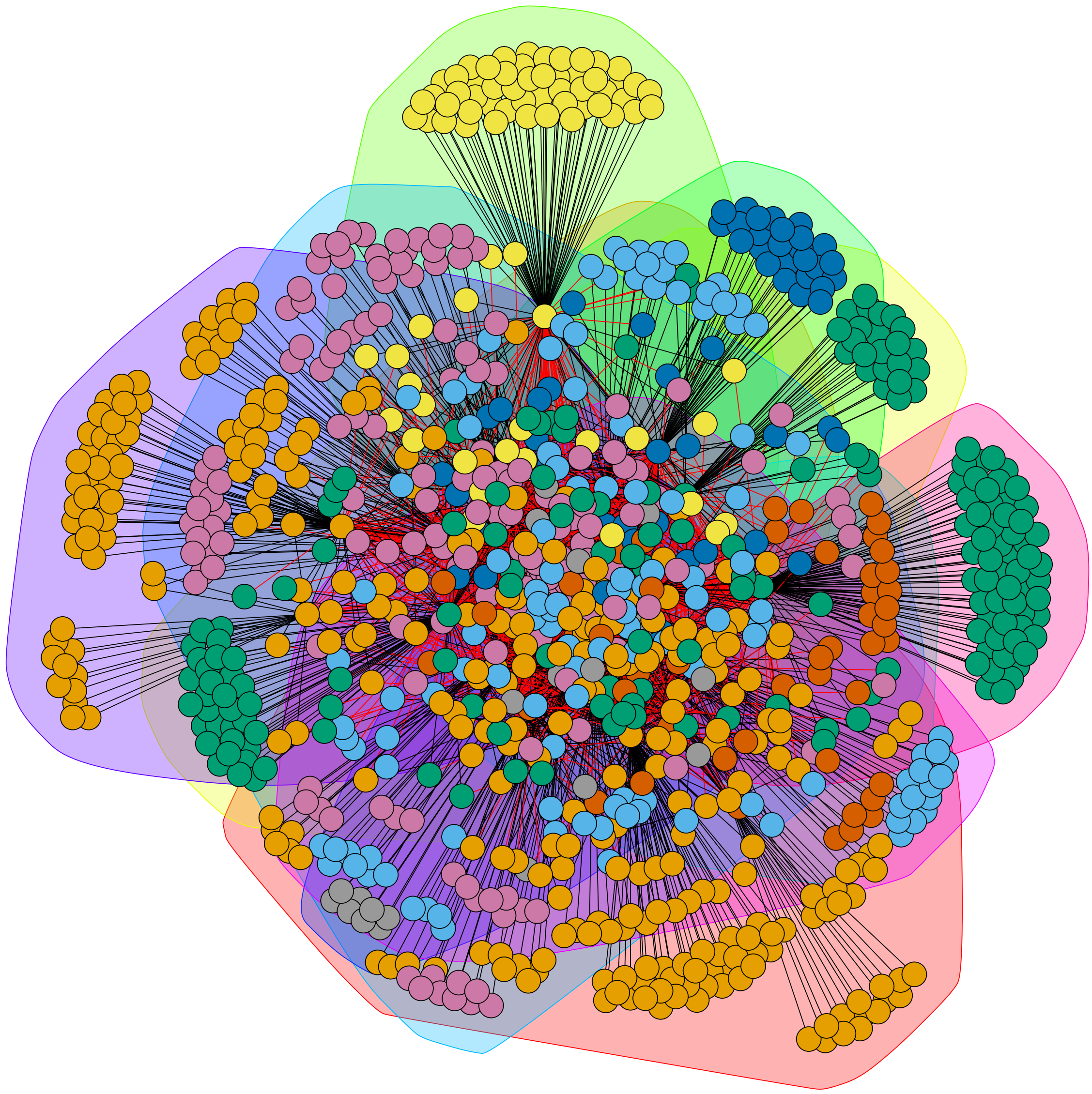


Table 8: Three most frequent labels for each variable within each cluster

	Total	**cluster1**	**cluster2**	**cluster3**	**cluster4**	**cluster5**
Area 1	New-York-City (30.2%)	New-York-City (30.1%)	New-York-City (28.6%)	New-York-City (27.8%)	New-Orleans (25.4%)	New-York-City (30.2%)
Area 2	South (19.8%)	South (21.7%)	South (22.9%)	South (21.1%)	New-York-City (23.8%)	South (27.9%)
Area 3	Chicago (14%)	NorthEast (18.9%)	MidWest (17.1%)	NorthEast (20%)	South (20.6%)	West (20.9%)
Style 1	Jazz (37.4%)	Jazz (40.6%)	Jazz (51.7%)	Jazz (35.7%)	Jazz (54.8%)	Jazz (34.2%)
Style 2	Swing_music (14.3%)	Bebop (22.6%)	Swing_music (20.7%)	Jazz_fusion (16.7%)	Dixieland (9.5%)	Bebop (13.2%)
Style 3	Bebop (11.4%)	Hard_bop (11.3%)	Blues (6.9%)	Bebop (11.9%)	Rhythm_and_blues (7.1%)	Cool_jazz (10.5%)
Birth Decade 1	*1920s (26.4%)*	1930s (29.8%)	1920s (38.3%)	1920s (25.4%)	1900s (32.9%)	1920s (42.6%)
Birth Decade 2	*1930s (19.7%)*	1920s (21.3%)	1910s (31.9%)	1930s (21.1%)	1910s (19.7%)	1930s (22.2%)
Birth Decade 3	*1910s (15.1%)*	1940s (20.7%)	1900s (8.5%)	1940s (20.2%)	1890s (18.4%)	1940s (18.5%)
Instrument 1	*Other (20.8%)*	Piano (18.8%)	Other (33.3%)	Drums (28.4%)	Other (23.1%)	Other (19.6%)
Instrument 2	*Piano (16.3%)*	Saxophone (18.3%)	Piano (22.9%)	Other (13.8%)	Trumpet (15.4%)	Drums (17.9%)
Instrument 3	*Saxophone (11.1%)*	Other (15.7%)	jazz (8.3%)	Piano (13.8%)	Guitar (11.5%)	Double Bass (10.7%)
	cluster6	**cluster7**	**cluster8**	**cluster9**	**cluster10**	**cluster11**
Area 1	*New-York-City (33.3%)*	New-York-City (36.6%)	New-York-City (29.4%)	New-York-City (29.5%)	New-York-City (29.2%)	Chicago (30%)
Area 2	*West (20%)*	South (16.2%)	South (23.5%)	West (21.6%)	South (21.5%)	New-York-City (26%)
Area 3	*Chicago (16.7%)*	Chicago (13.4%)	Chicago (17.6%)	NorthEast (19.3%)	NorthEast (16.9%)	South (24%)
Style 1	*Jazz (57.1%)*	Jazz (29.3%)	Bebop (23.5%)	Jazz (41.2%)	Jazz (28.8%)	Jazz (30.2%)
Style 2	*Swing_music (10.7%)*	Swing_music (29.3%)	Hard_bop (17.6%)	Swing_music (30.9%)	Bebop (15.4%)	Jazz_fusion (16.3%)
Style 3	*Bebop (7.1%)*	Rhythm_and_blues (10.6%)	Jazz (17.6%)	Bebop (5.9%)	Rhythm_and_blues (11.5%)	Hard_bop (9.3%)
Birth Decade 1	*1920s (39.1%)*	1920s (30.7%)	1910s (25%)	1920s (29.4%)	1920s (31.6%)	1930s (29.7%)
Birth Decade 2	*1930s (19.6%)*	1910s (22.9%)	1920s (25%)	1910s (22.9%)	1930s (26.3%)	1940s (28.1%)
Birth Decade 3	*1910s (13%)*	1900s (19.3%)	1930s (20%)	1900s (14.7%)	1940s (13.2%)	1920s (10.9%)
Instrument 1	*Other (29.2%)*	Piano (22.2%)	Singer (25%)	Other (23.2%)	Other (24.7%)	Other (28.4%)
Instrument 2	*Double Bass (10.4%)*	Other (19.8%)	Piano (15%)	Piano (13.4%)	Piano (15.6%)	Piano (19.4%)
Instrument 3	*Trombone (10.4%)*	Singer (16.8%)	Saxophone (15%)	Trumpet (11.6%)	Trumpet (11.7%)	Saxophone (17.9%)