

# Exploring J-DISC: Some Preliminary Analyses

Yun Hao, Kahyun Choi, J. Stephen Downie

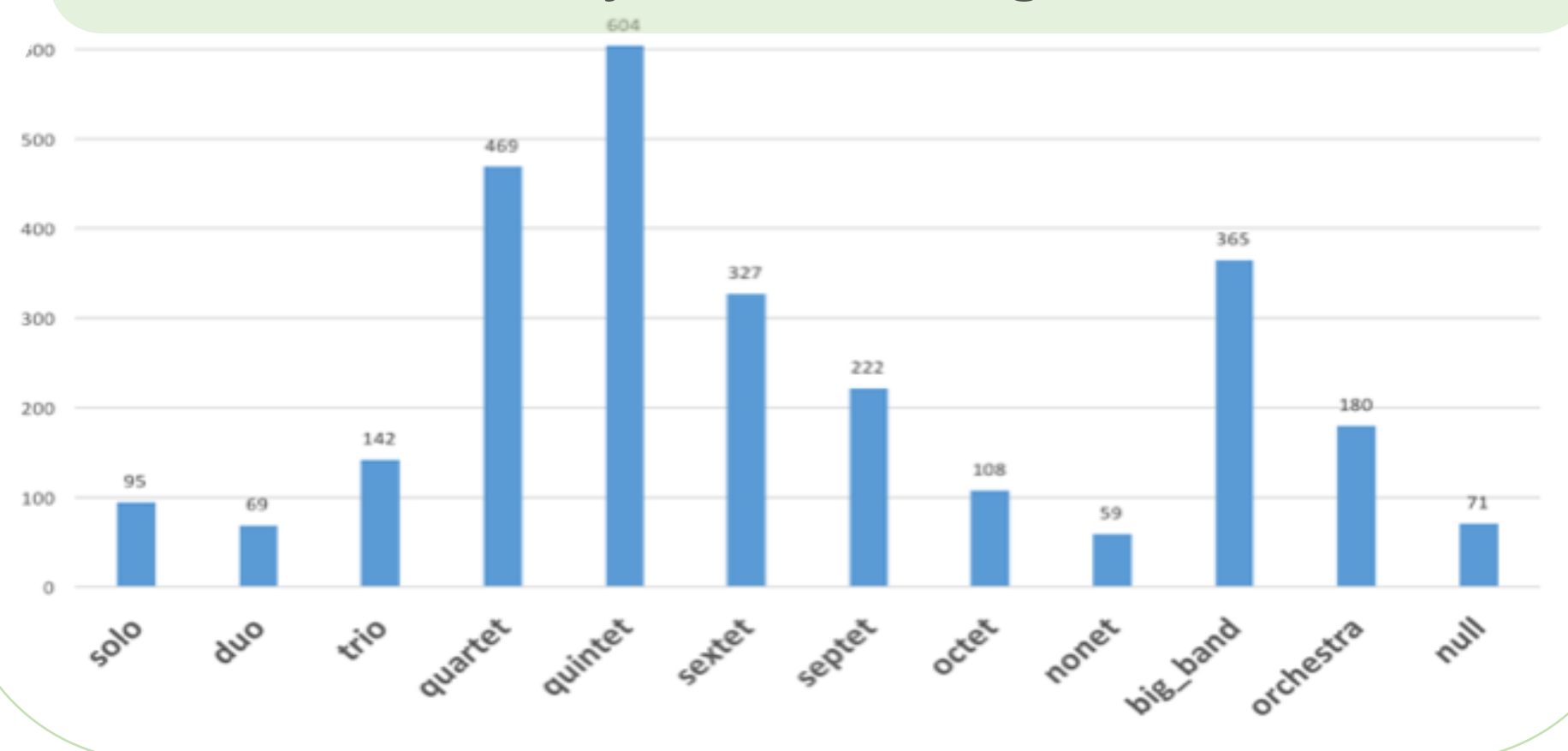
## INTRODUCTION

J-DISC is a specialized digital library for information about jazz recording sessions that includes rich structured and searchable metadata. It is created by the Center for Jazz Studies at Columbia University. The data was collected with the foundation of 75 “core” artists determined by the J-DISC team in consultation with music scholars, and is recorded using a unique metadata structure designed to better capture recording session-related data regarding the 2,711 unique recording sessions associated with 3,744 distinct artists. Key entities recorded in J-DISC include: *person*, *skill*, *session*, *track*, *composition*, and *issue*. Our work presented here includes both informetric and network analyses across this dataset to better illustrate the kinds of uses that musicologists could make of this collection.

## SESSION

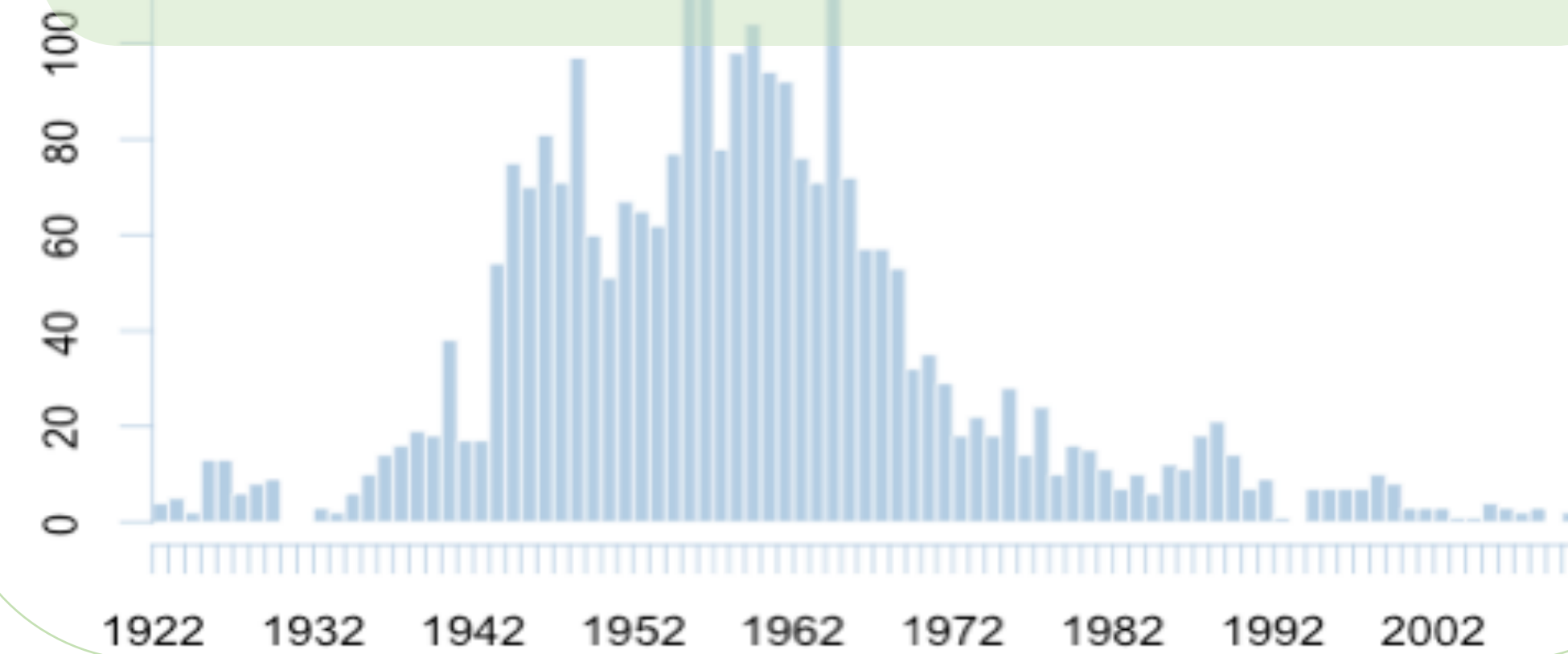
### Ensemble Sizes

Sessions are not only musical events, but also social events from which interesting findings of social network of jazz artists might be reached.



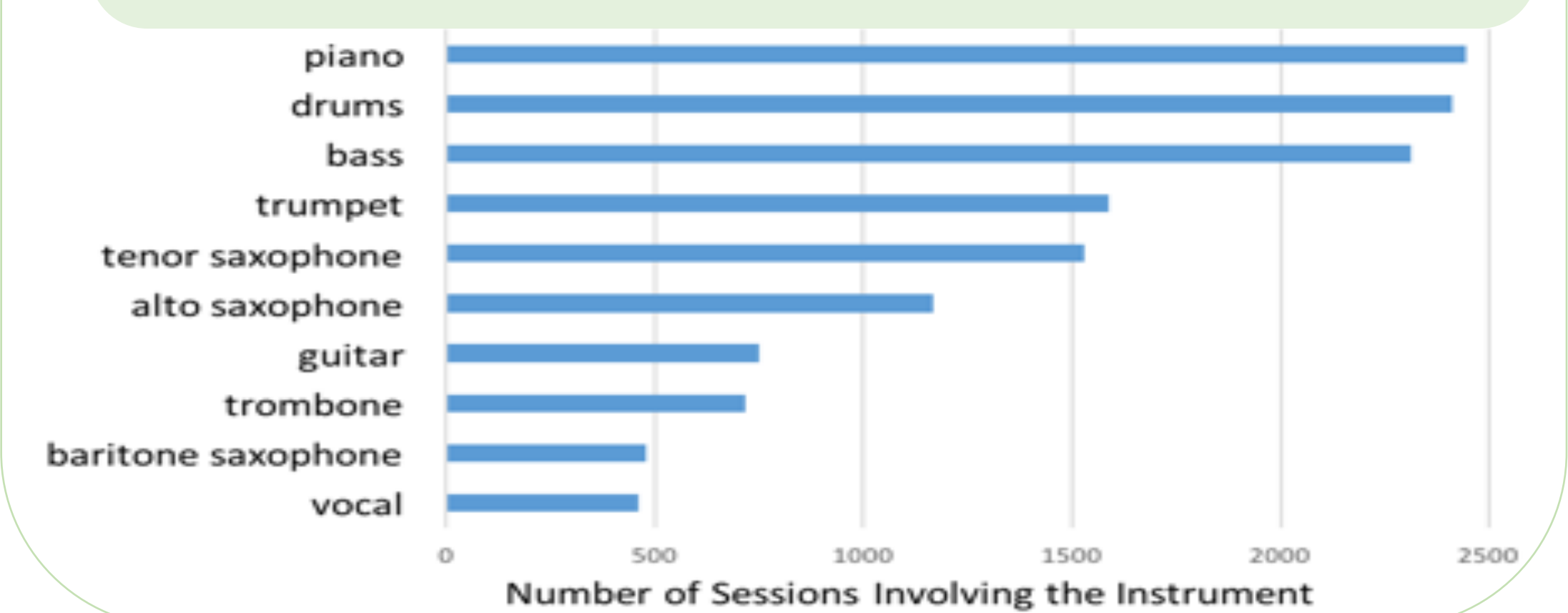
### Time Span

The earlier sessions occurred around 1923 led by, for example, Jelly Roll Morton. The later sessions occurred around 2011 led by, for example, Paulo Moura and Gerald Wilson. Most of the sessions in J-DISC were recorded between 1940 to 1970.

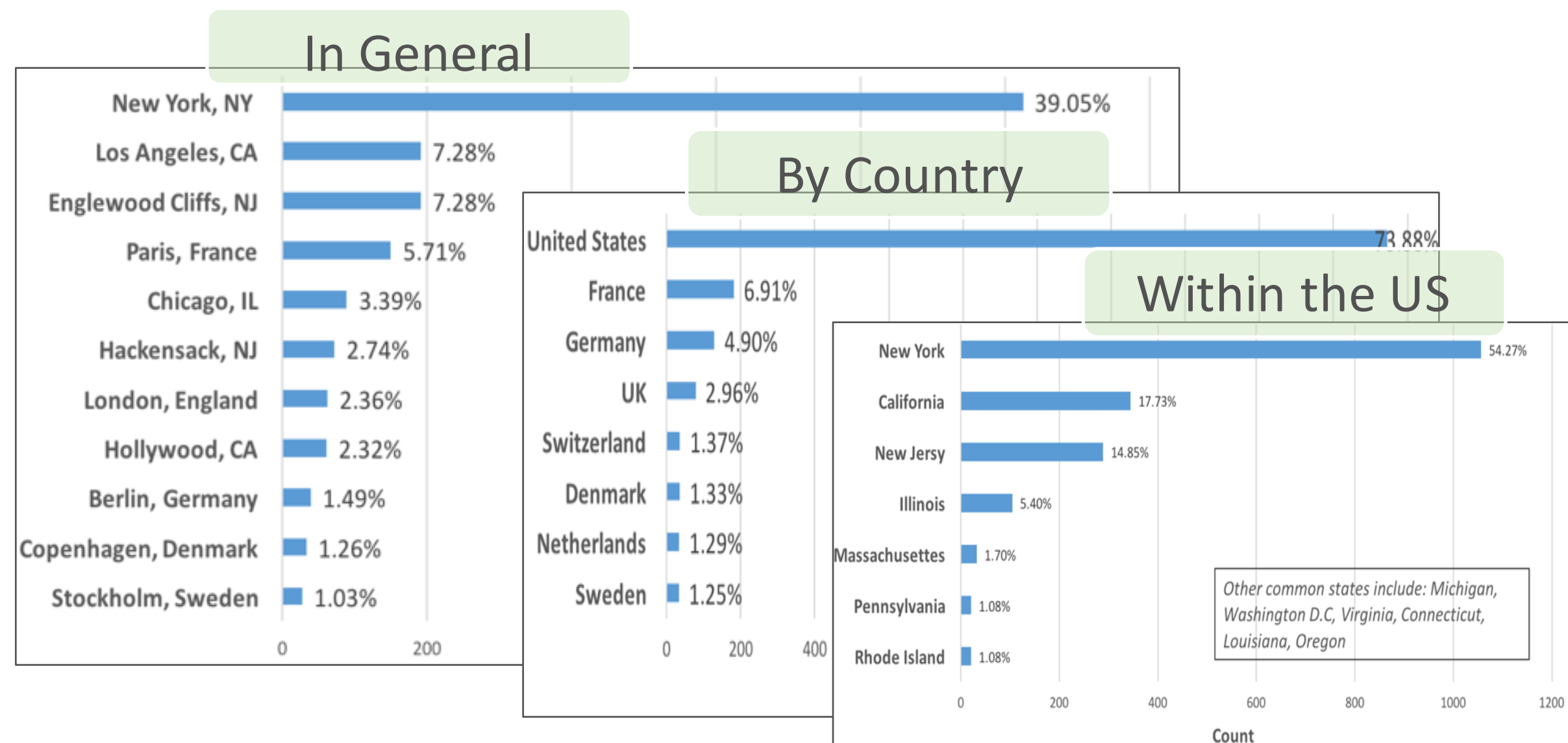


### Instrument

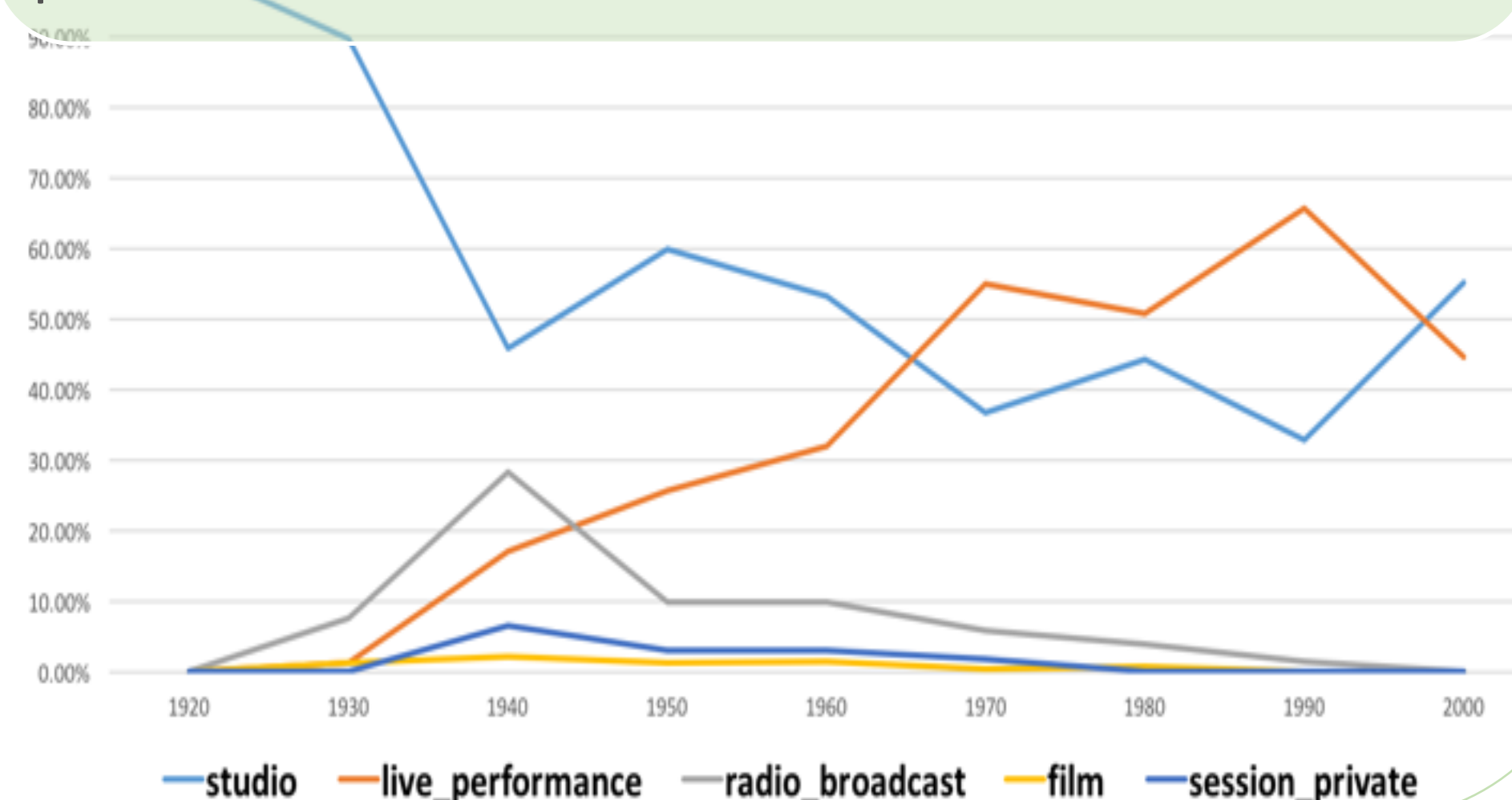
Around 150 instrument types in total. On average, each session involves 6 different instrument types. Piano, drums and bass are the three most important instruments in the J-DISC session.



### Venues



Venues where jazz music is played include clubs, dancehalls, theaters, studios, churches, etc. Studio sessions were the dominating type until mid-1960s. It is interesting to note the growth of live performance in the J-DISC data after the mid-1960s.



### Top Artists and Session Pairings

Table 1. Artists who frequently performed together

Rank	Artist_1	Artist_2	Cnt
1	John Coltrane	McCoy Tyner	85
2	Jimmy Lyons	Cecil Taylor	84
3	McCoy Tyner	Elvin Jones	82
4	John Coltrane	Jimmy Garrison	80
5	John Coltrane	Elvin Jones	77
6	Freddie Green	Count Basie	71
7	Dizzy Gillespie	James Moody	66
8	McCoy Tyner	Jimmy Garrison	63
9	Don Byas	Kenny Clarke	59
10	Jimmy Garrison	Elvin Jones	58

## NETWORK ANALYSES

### Basic Metrics

Betweenness centrality, eigenvector centrality and the number of triangles with the artist being one vertex are calculated to discover the most salient artists. It turns out that there is an overlap of four artists in the three lists, with Dizzy Gillespie ranking first in all three.

Table 2. Top 10 artists with the largest betweenness centrality

Rank	Artist	Btw.	Rank	Artist	Btw.
1	Dizzy Gillespie	1481544	6	Billie Holiday	307284
2	Don Byas	995781	7	Wardell Gray	279149
3	Charlie Parker	480710	8	Jelly Roll Morton	236322
4	Cecil Taylor	386678	9	Mary Lou Williams	207149
5	Kenny Dorham	316248	10	Lucky Thompson	166597

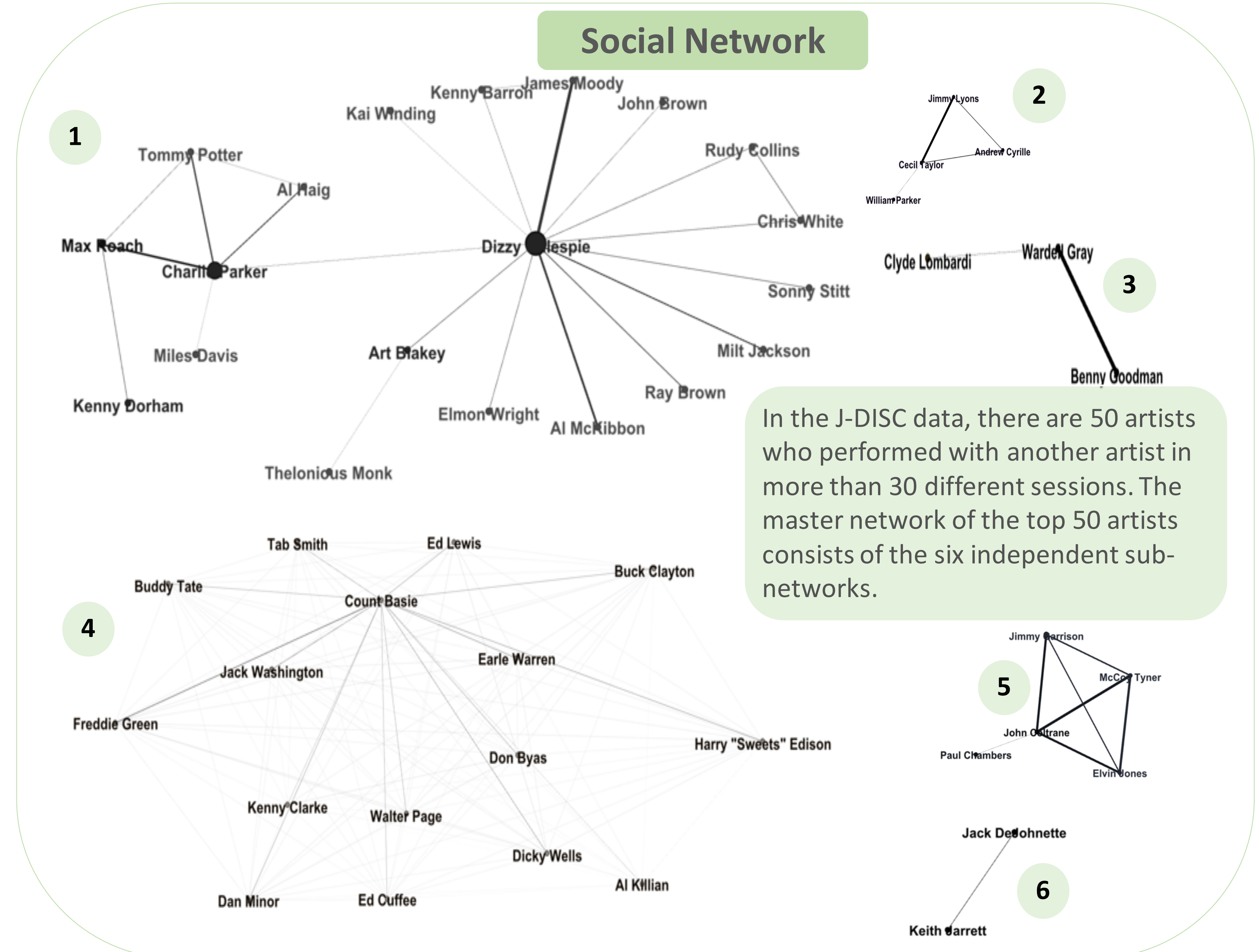
Table 3. Top 10 artists with the largest eigenvector centrality

Rank	Artist	Eig.	Rank	Artist	Eig.
1	Dizzy Gillespie	0.0085	6	James Moody	0.0032
2	Charlie Parker	0.0040	7	Wardell Gray	0.0031
3	Don Byas	0.0039	8	Jerome Richardson	0.0030
4	Kenny Dorham	0.0035	9	Billie Holiday	0.0030
5	Clark Terry	0.0032	10	Kenny Clarke	0.0027

Table 4. Top 10 artists with the most triangles

Rank	Artist	# of Tri.	Rank	Artist	# of Tri.
1	Dizzy Gillespie	911	6	Don Byas	288
2	John Coltrane	496	7	Charlie Parker	277
3	Cecil Taylor	385	8	Kenny Clarke	204
4	Keith Jarrett	332	9	McCoy Tyner	194
5	Kenny Dorham	310	10	Elvin Jones	181

### Social Network



In the J-DISC data, there are 50 artists who performed with another artist in more than 30 different sessions. The master network of the top 50 artists consists of the six independent sub-networks.

## CONCLUDING REMARKS

One of the powerful aspects of the J-DISC dataset is its richness in recording session-based data, and thus is of great value for musicological work especially those focused on social network aspects of jazz creation and production. Other potential uses of the J-DISC session-related metadata include building ground truth for MIR research, enabling new visualizations or developing interfaces for novel jazz music streaming services, etc. When associated with jazz audio, J-DISC data can also be used to create a comprehensive digital library and could be the foundation of new avenues of research of computational musicology. Although the data itself is limited in completeness, in the future J-DISC could still contribute to enhancing accessibility of jazz resources by becoming part of the Linked Open Data network to link to other jazz resources that are available online.