

Visualizing the Knowledge Space of Music

IAML 2017, the 66th IAML Congress

Glenn Henshaw, Fan Yun

June 22, 2017

What is music scholarship?

- ▶ Music scholarship is hard to define
 - ▶ Deeply interdisciplinary.
 - ▶ It's a young field.

What is music scholarship?

Examples:

- ▶ **Musicology:** *Richard Wagners opposition to animal experimentation: A visionary social critic*
- ▶ **Ethnomusicology:** *A bird tradition in the west of the Balkan Peninsula*
- ▶ **Music pedagogy:** *From Mississippi hot dog to Arizona cactus*
- ▶ **Music therapy** *The use of music with chronic food refusal: A case study*
- ▶ **Popular music studies** *Crossing cinematic and sonic bar lines: T-Pains Cant believe it*

Goal: Map the knowledge space of music

- ▶ How can the music be divided into subfields?
- ▶ What are the principle subfields?
- ▶ What are the emerging subfields?
- ▶ How do these subfields connect/interact?
 - ▶ Internally
 - ▶ externally

Goal: Map the knowledge space of a given field

- ▶ Inskip and Wiering Conducted surveys of Musicological scholarship.
- ▶ Leech-Wilkinson considered community behavior from a historical perspective.

Goal: Map the knowledge space of a given field

- ▶ Expert surveys tend to be costly and slow endeavors.
- ▶ Expert bias.

Goal: Map the knowledge space of music

- ▶ **Co-Word Analysis** attempts to map the knowledge space of a given field by measuring and analyzing the strength of the associations between terms (keywords, indexed terms, or words from a corpus)
- ▶ The **strength** of the association between terms is based on how frequently they co-appear in documents.
- ▶ We use the results of co-word analysis to hierarchically cluster the terms.
- ▶ We visualize the clusters with **dendrograms**, **weighted graphs**, and **strategic diagrams**.

Visualization techniques

- ▶ **Dendrograms:** detailed information on the relationship between terms and clustering.
- ▶ **Weighted graphs:** details of each cluster
- ▶ **Strategic Diagram:** Local and global view of the clusters

What is RILM?

- ▶ A comprehensive music bibliography featuring
 - ▶ abstracts and citations
 - ▶ 143 languages
 - ▶ 1967 – present
 - ▶ 875,000 records
- ▶ RILM indexing represents hierarchical relationships with broader and narrower topics.

The Data

	year	ac	class1	lvl1	cat1	lvl2	cat2	lvl3	cat3	lang	code
0	2013	6647	29	performing organizations	T	Germany	G	Berlin	G	Russian	2013-6647
1	2013	6648	29	Kopatchinskaja Patricia	N	interviews	M	NaN	NaN	Russian	2013-6648
2	2013	6648	29	performers--violin	T	Kopatchinskaja, Patricia	N	NaN	NaN	Russian	2013-6648
3	2013	6651	29	Gounod Charles	N	performances	M	<Faust>	W	Russian	2013-6651
4	2013	6651	29	performing organizations	T	Russia	G	Sankt-Peterburg	G	Russian	2013-6651

- ▶ Almost all academic music articles (2000-2015)
- ▶ 263,656 Rows
- ▶ 59,908 Articles
- ▶ 25,297 distinct terms (lvl1)

15 Most common terms

	count
pedagogy	5499
China	5084
instruments--keyboard (organ family)	3011
popular music	2388
aesthetics	2037
singing	2024
song--popular and traditional	1886
pedagogues	1844
performing organizations	1836
religious institutions	1788
instrument builders--organ	1773
sound recordings	1734
academic institutions	1657
instruments--keyboard (piano family)	1639
psychology	1608

Co-occurrence

Definition

For two terms, i and j , their co-occurrence is defined as

$C_{i,j}$ = How often i and j appear in the same article

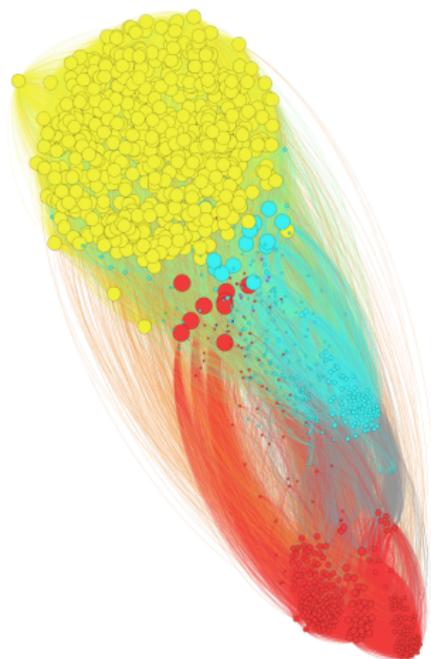
Example

The terms 'aesthetics' and 'popular music' appear together in 123 articles.

$$C_{\text{'aesthetics'}, \text{'popular music'}} = 123$$

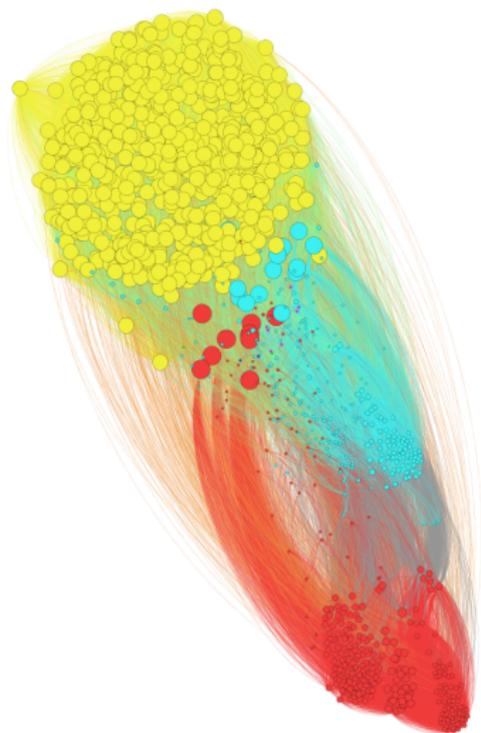
From matrix to graph using the matrix C

- ▶ Dots (nodes) are terms.
- ▶ Lines between terms indicate that $C > 4$.
- ▶ Colors indicate terms that are highly interconnected (clusters).



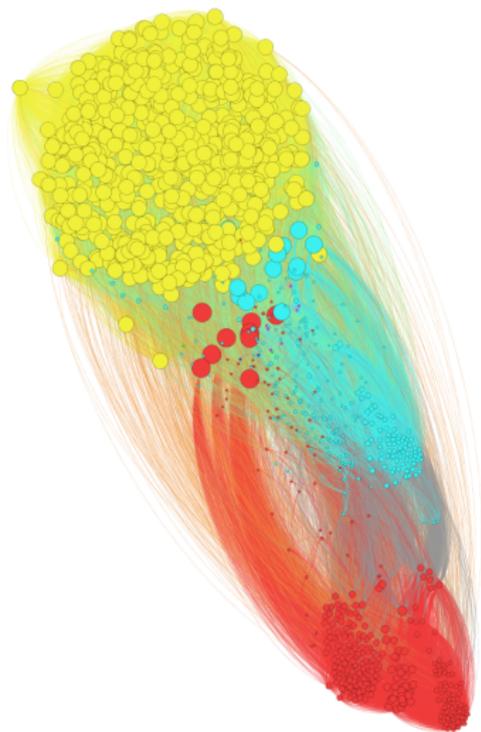
From matrix to graph using the matrix C

- ▶ Despite its beauty, we could not obtain valuable information from this graph!



From matrix to graph using the matrix C

- ▶ High frequency terms dominate the connections.



Cosine similarity

Cosine similarity (CS) is defined as

$$CS(i, j) = \sqrt{\frac{C_{ij}^2}{C_{ii}C_{jj}}}$$

Example

Let

$$\begin{aligned}i &= \text{'China'} & j &= \text{'popular music'} \\k &= \text{'mathematics'} & l &= \text{'scales'}.\end{aligned}$$

Then,

$$\begin{aligned}C_{ij} = 151 & \quad CS(i, j) = \sqrt{151^2 / (5084 \cdot 2388)} = .04 \\C_{kl} = 15 & \quad CS(k, l) = \sqrt{15^2 / (309 \cdot 303)} = .05.\end{aligned}$$

Hierarchical Clustering (a rough sketch)

- ▶ Distance between individual terms: $d(i, j) = 1 - CS(i, j)$.
- ▶ At the start, each term, t_i , is contained in its own cluster $c_i = \{t_i\}$.
- ▶ Define distance between clusters $D(c_i, c_j)$.
 - ▶ Many ways to do this

Repeat until there is only one cluster:

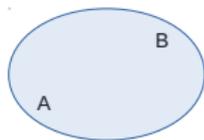
- ▶ Find two clusters of minimal distance, i.e. $\min D(c_i, c_j)$.
- ▶ Merge the clusters c_i and c_j into a cluster $c_{i\&j}$.
- ▶ Delete the clusters c_i and c_j .

We visualize Hierarchical clustering with a dendrogram.

Hierarchical Clustering (a rough sketch)



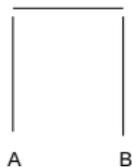
Hierarchical Clustering (a rough sketch)



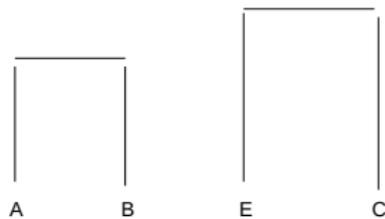
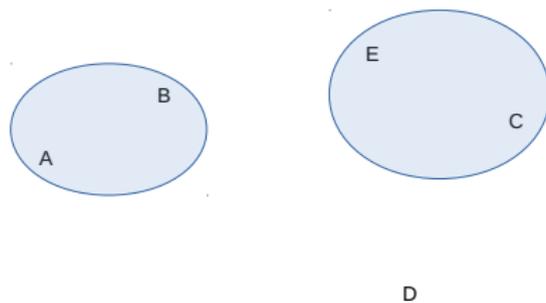
E

C

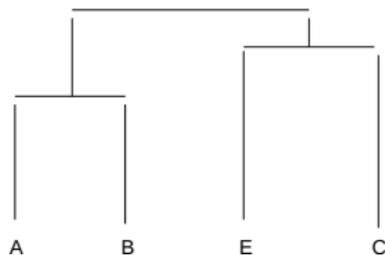
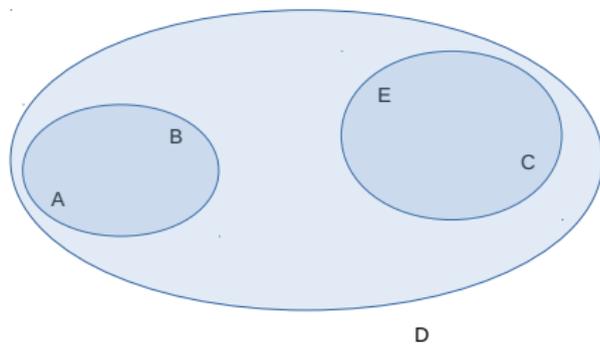
D



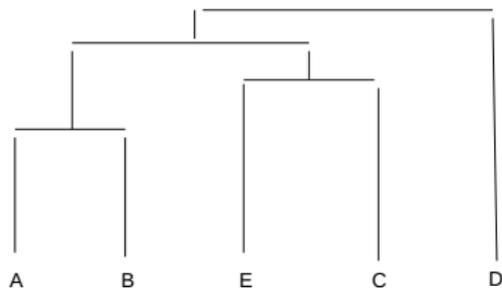
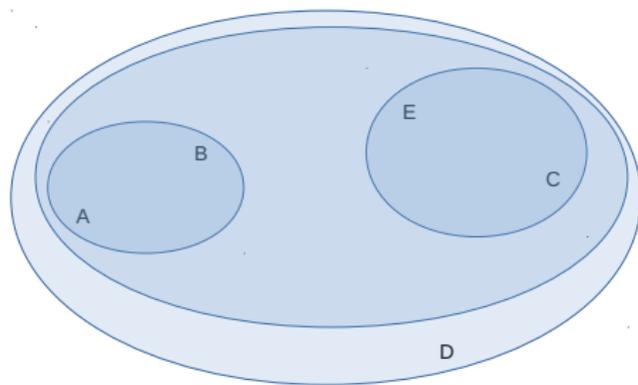
Hierarchical Clustering (a rough sketch)



Hierarchical Clustering (a rough sketch)



Hierarchical Clustering (a rough sketch)

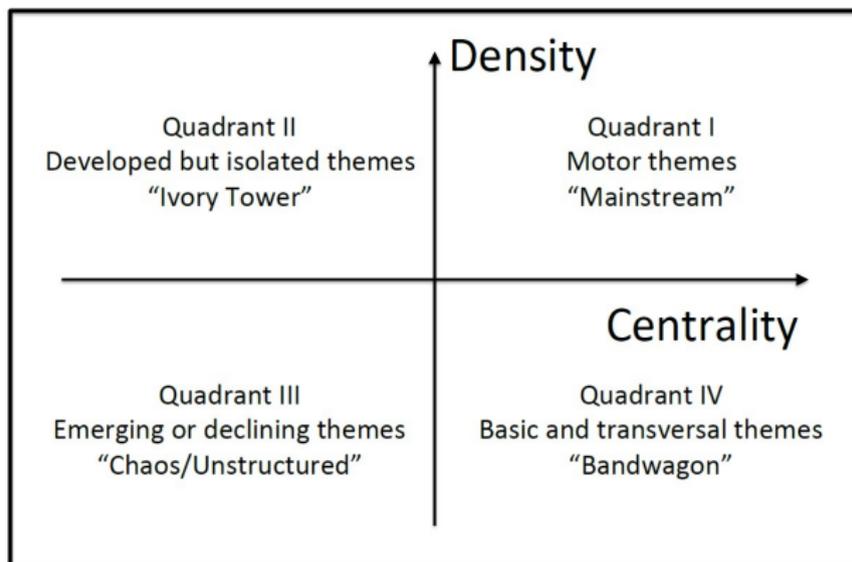


Density and Centrality

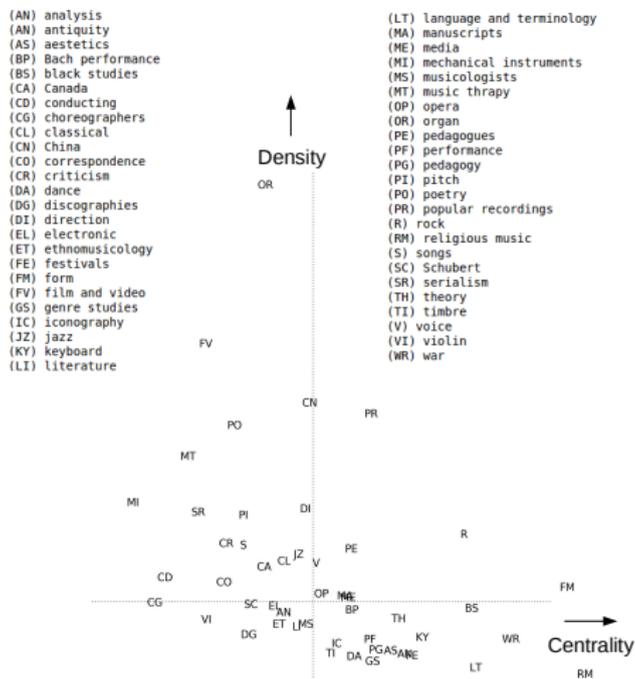
For each cluster we calculate:

- ▶ **Density** - Average strength of all the connections within a cluster
- ▶ **Centrality** - The square root of the sum of the squares of all connections to outside clusters

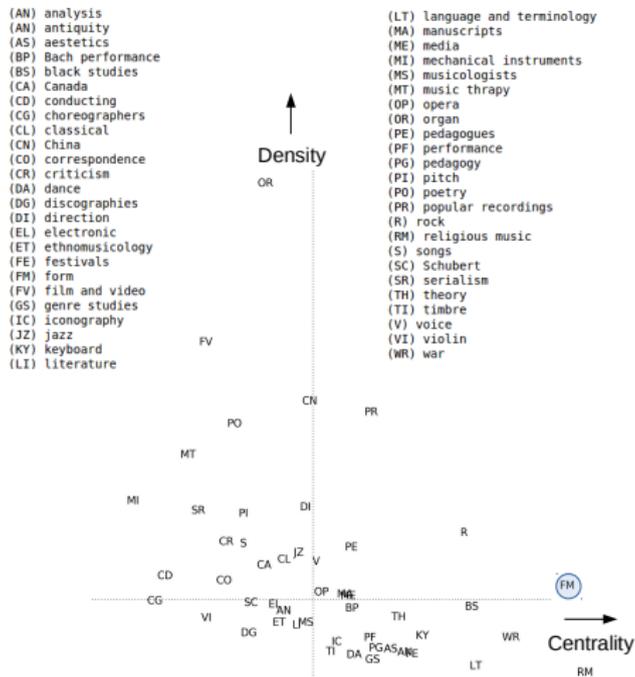
Strategic Diagram



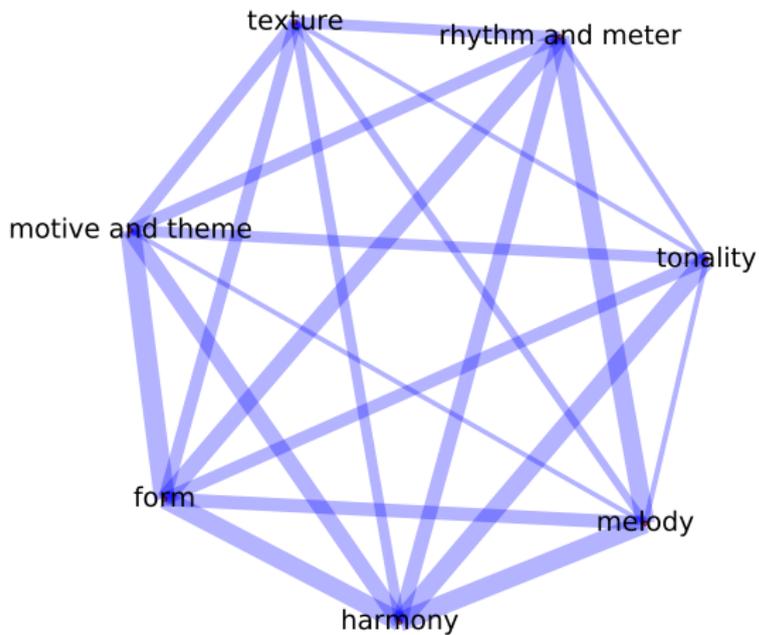
Strategic Diagram



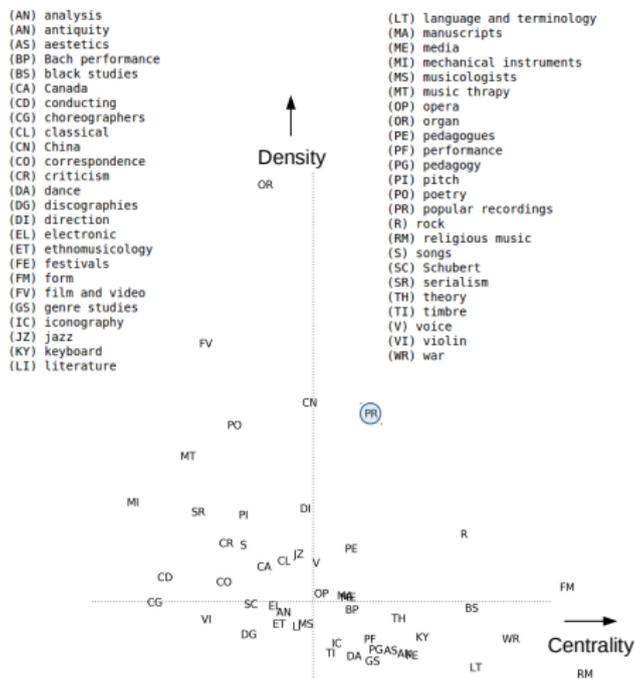
Strategic Diagram



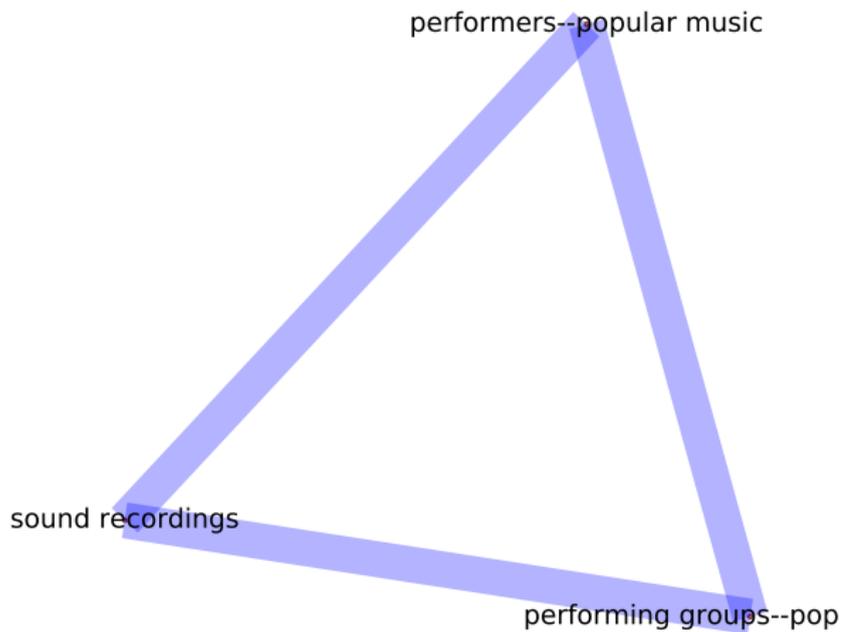
Cluster: Form



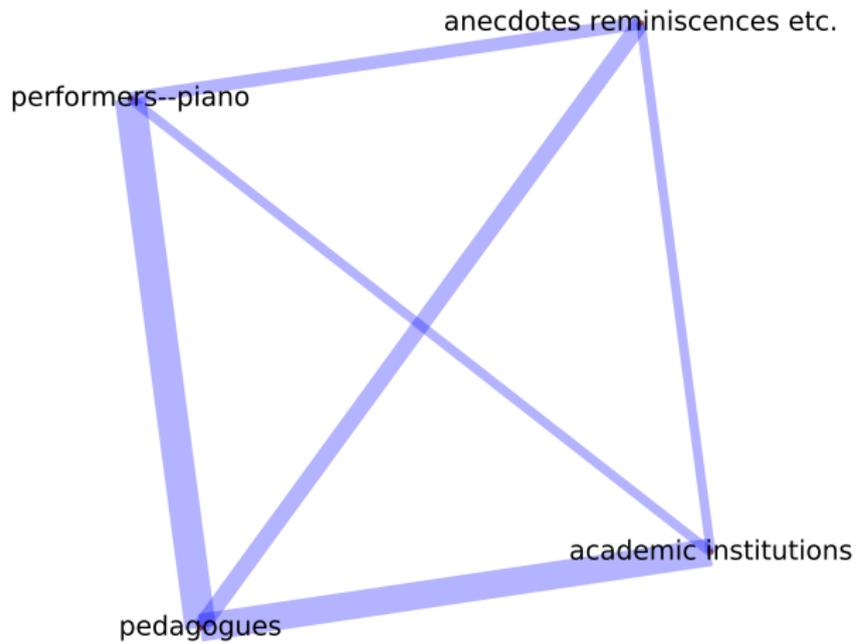
Strategic Diagram



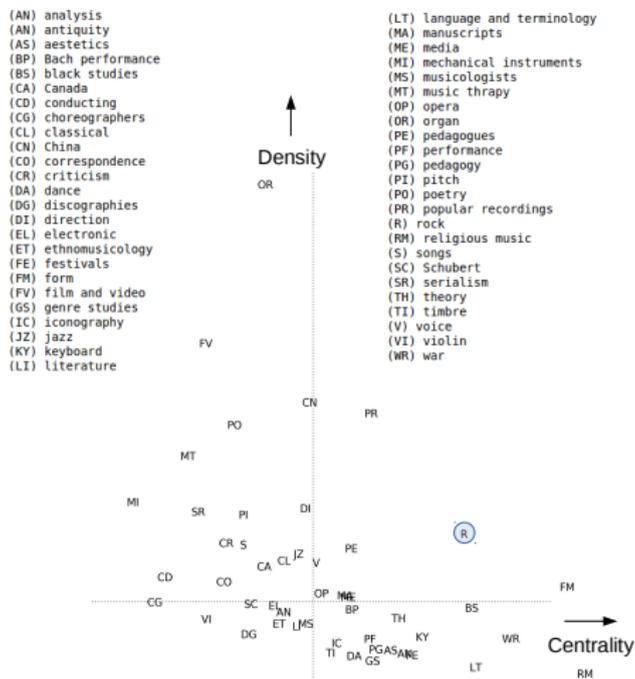
Cluster: Popular Recordings



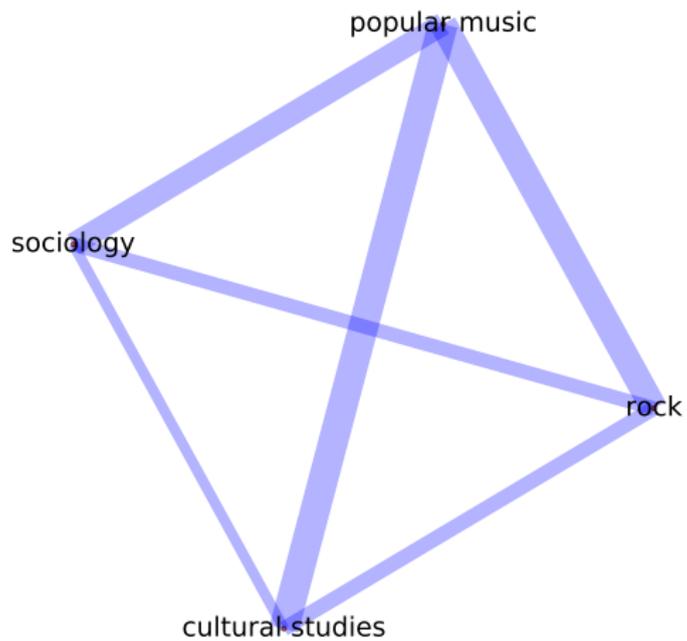
Cluster: Pedagogues



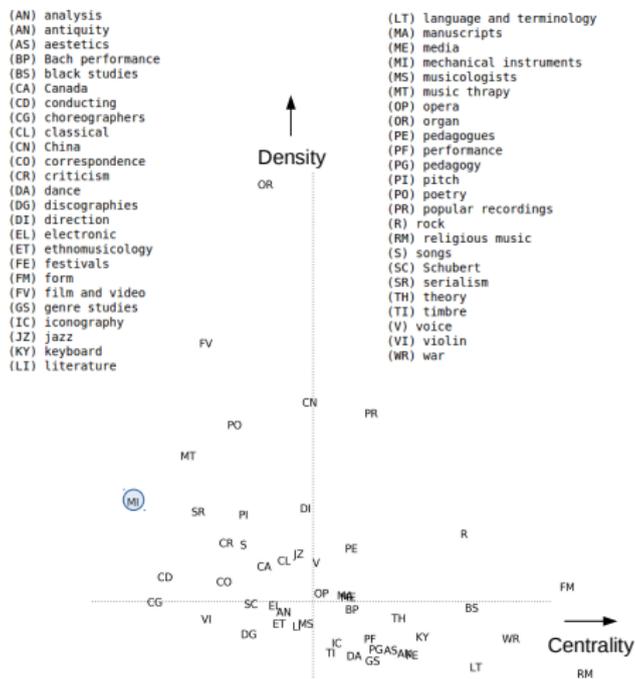
Strategic Diagram



Cluster: Rock



Strategic Diagram

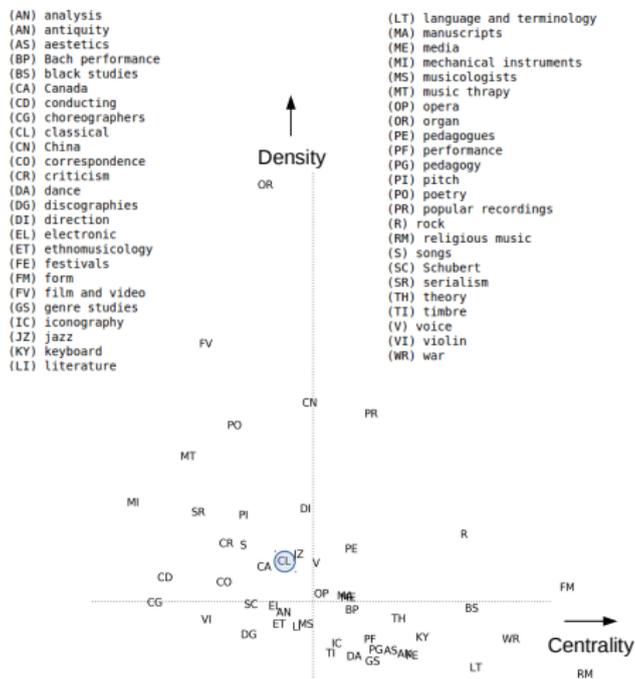


Cluster: Mechanical instruments

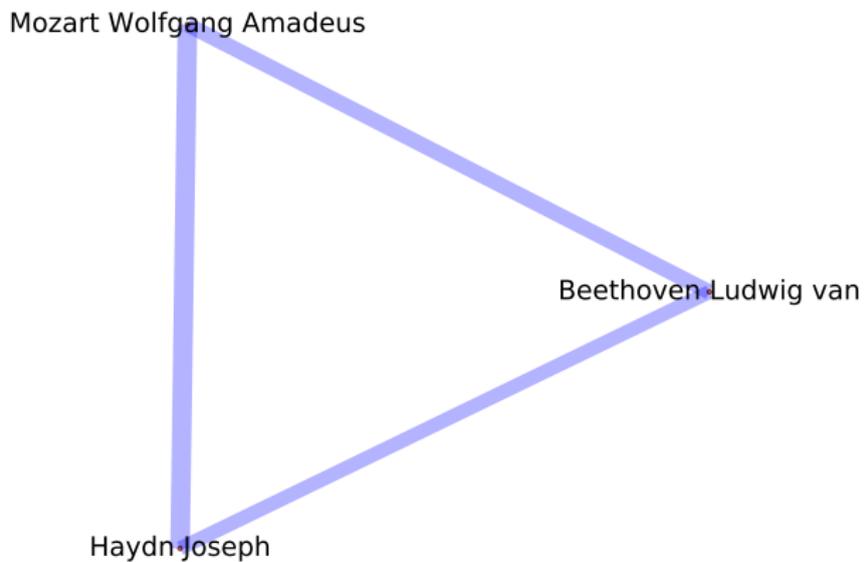
instruments--mechanical

instruments--collections

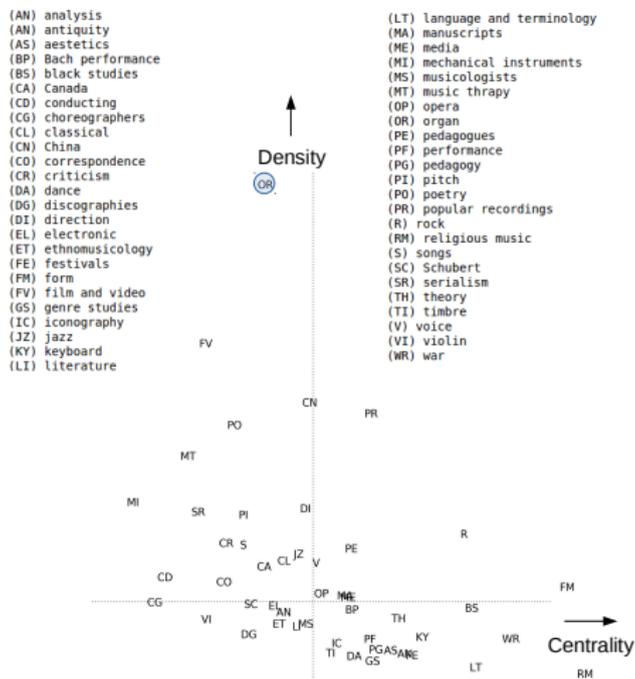
Strategic Diagram



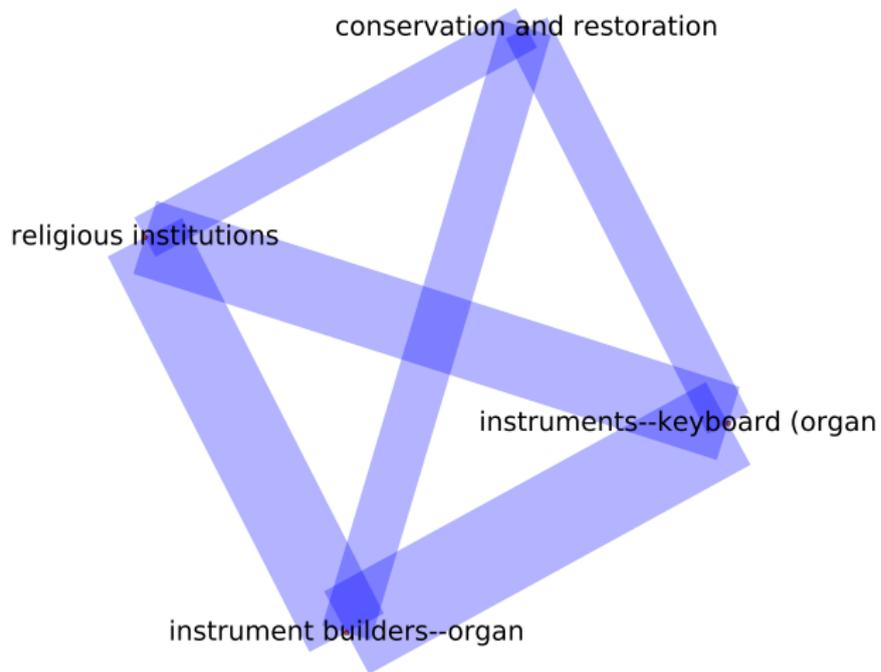
Cluster: Classical



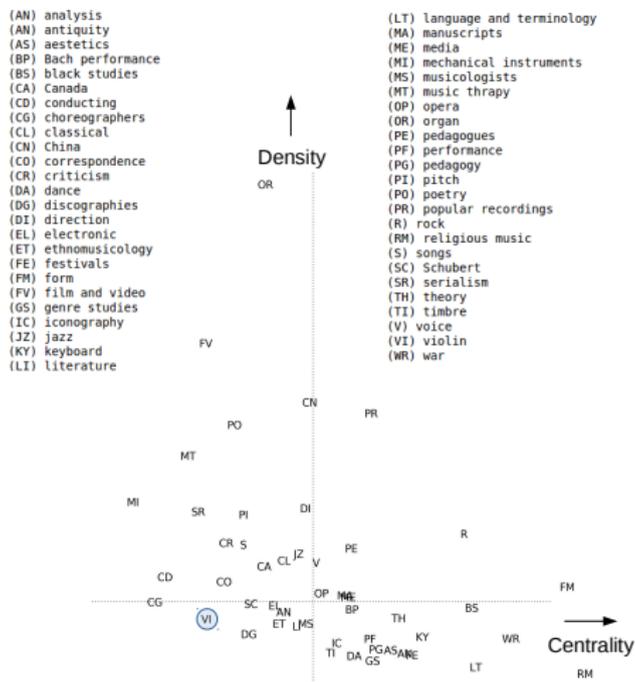
Strategic Diagram



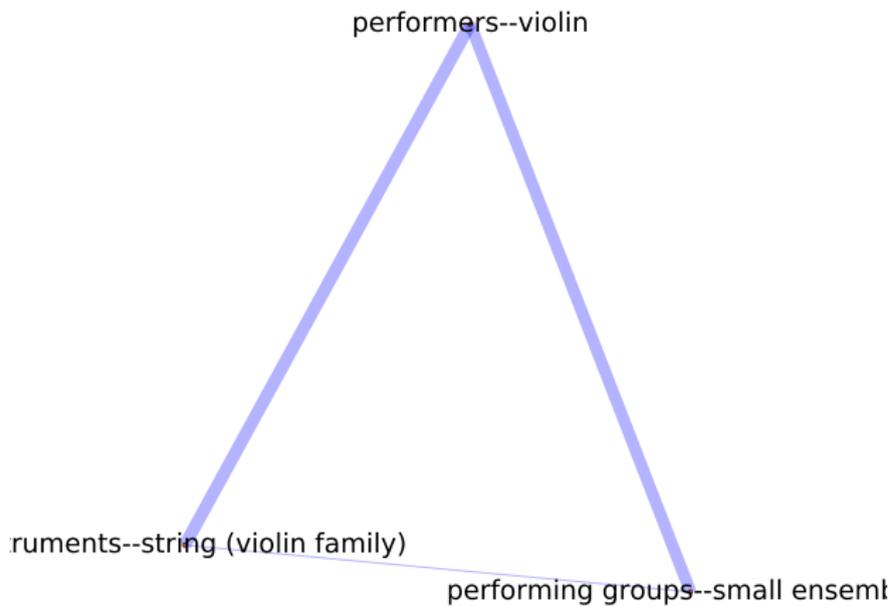
Cluster: Organ



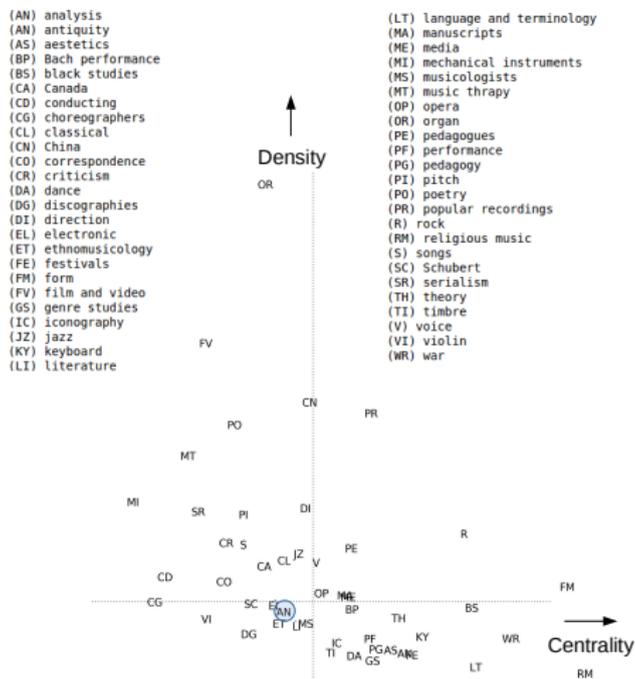
Strategic Diagram



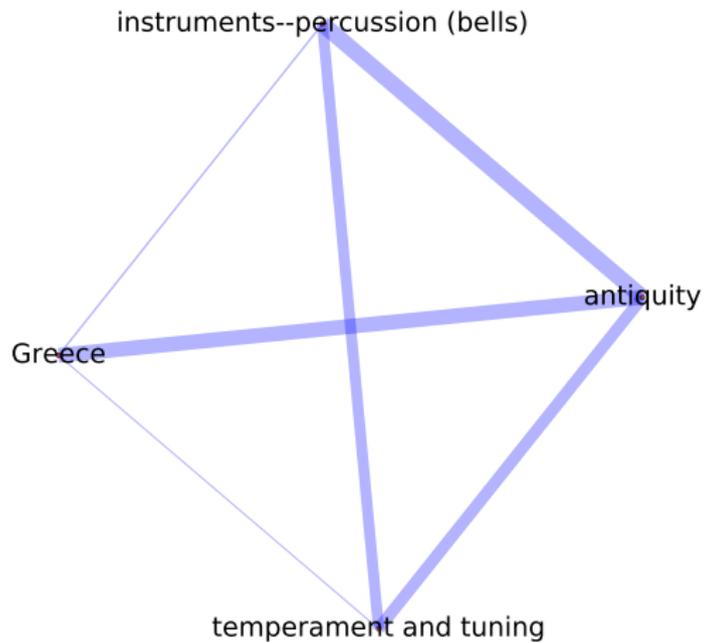
Cluster: Violin



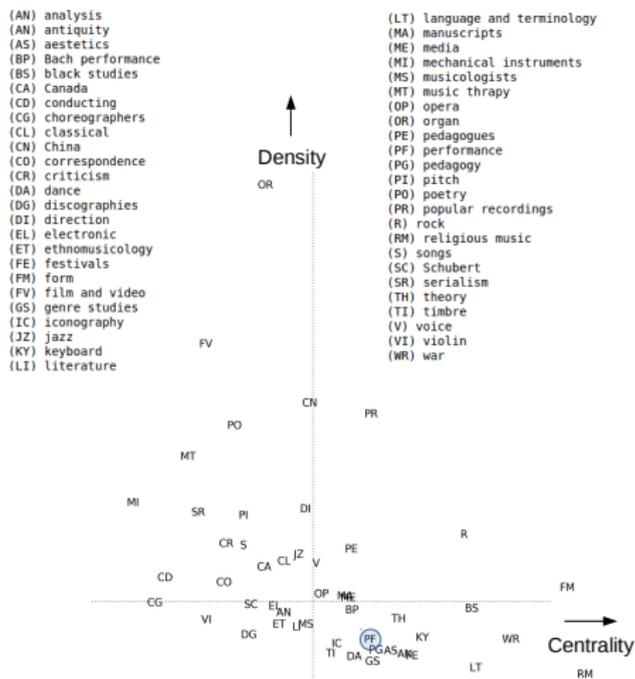
Strategic Diagram



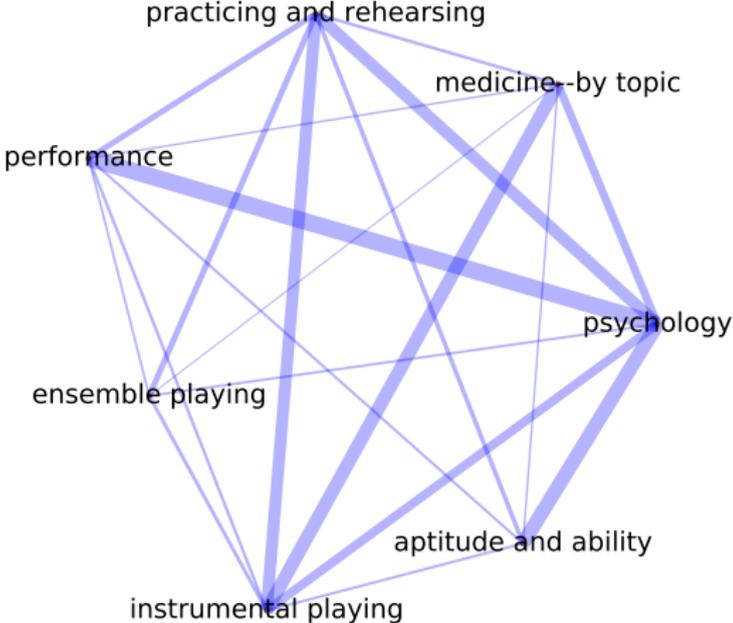
Cluster: Antiquity



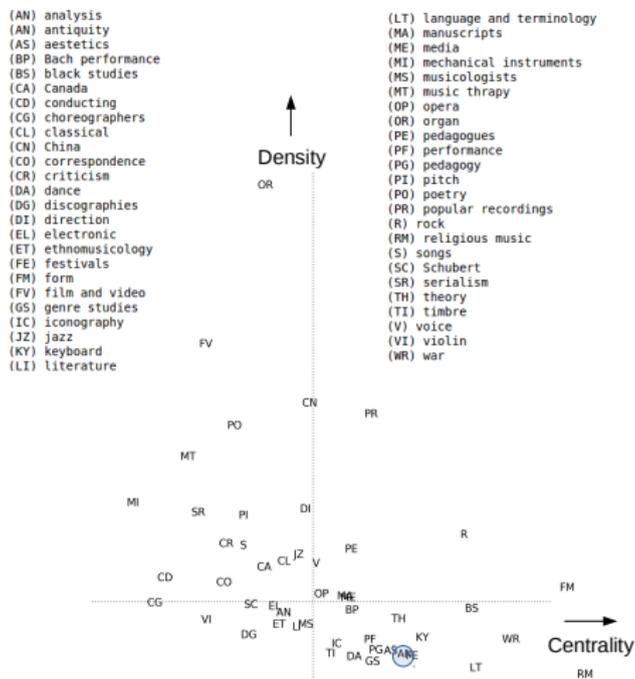
Strategic Diagram



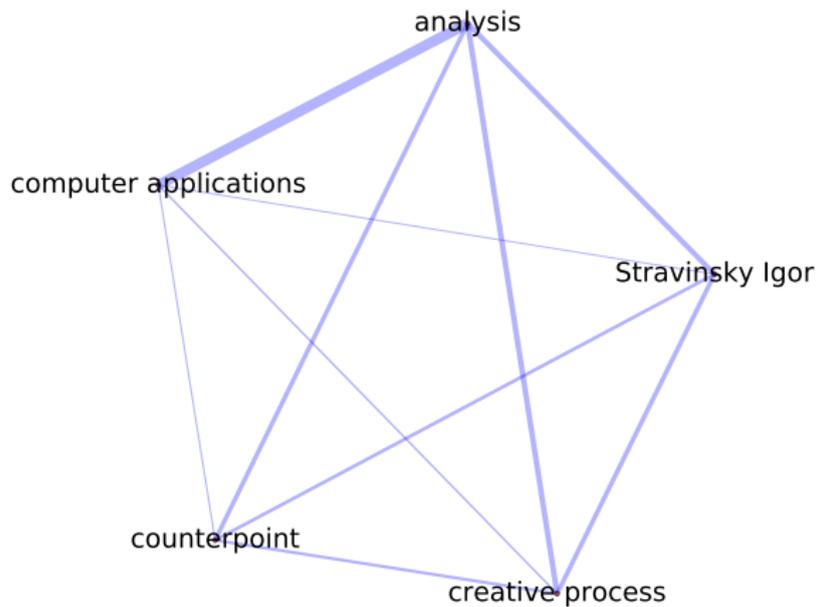
Cluster: Performance



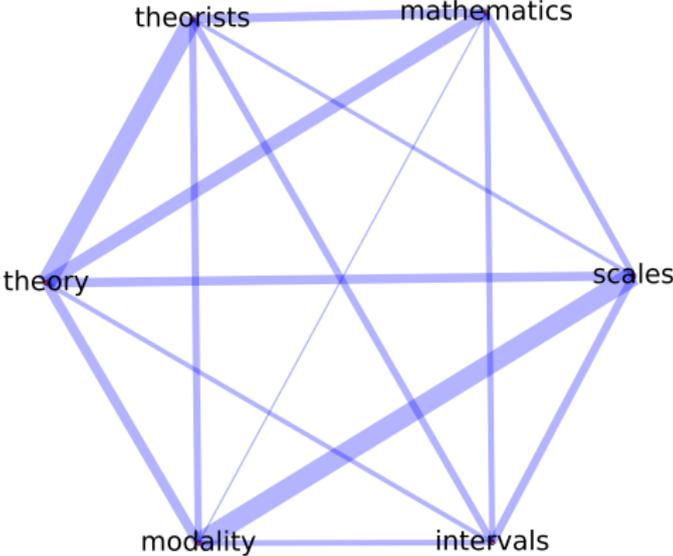
Strategic Diagram



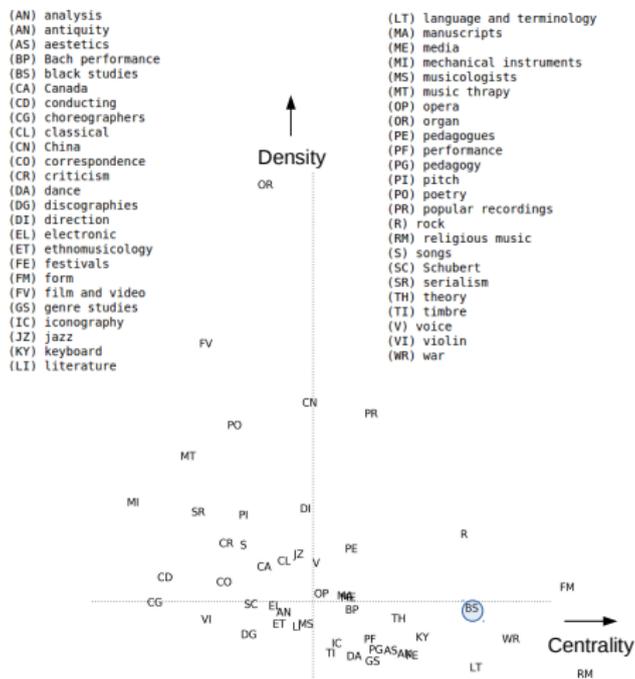
Cluster: Analysis



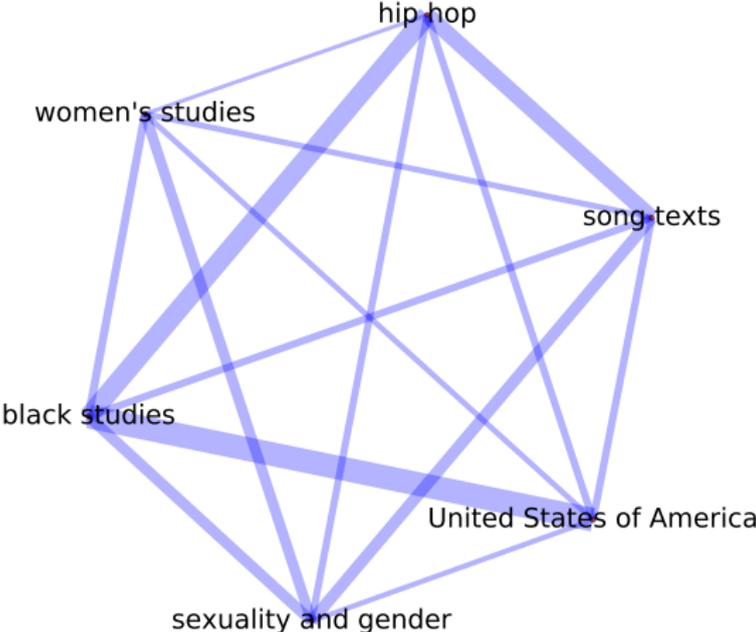
Cluster: Theory



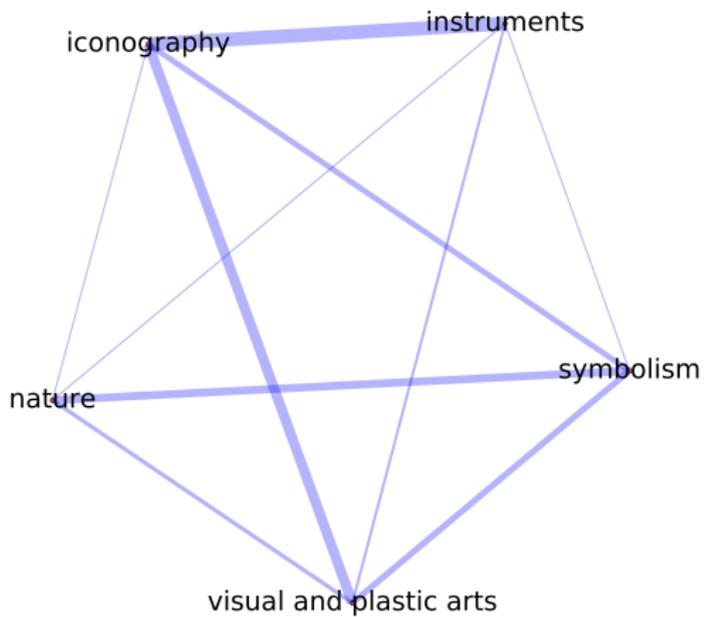
Strategic Diagram



Cluster: Black studies



Cluster: Iconography



Future Work

- ▶ Use more of the indexing string
- ▶ User data
- ▶ Mapping the movement of clusters through the strategic diagram over time
- ▶ Produce an online tool for scholars to browse

Thank You!