Visualizing the Knowledge Space of Music

Glenn Henshaw, Fan Yun

April 24, 2017

Goal: Map the knowledge space of a given field

- How can the field 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

- Although Experts have deep insight into their specialty, they tend to be bias towards their own subfield. Also, expert surveys tend to be costly endeavors.

Goal: Map the knowledge space of a given field

- **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:** detailed of each cluster
- **Strategic Diagram:** Local and global view of the clusters

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*

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

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

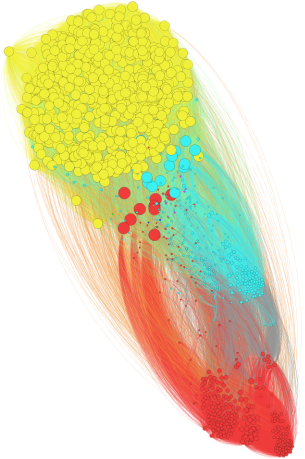
Example

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

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

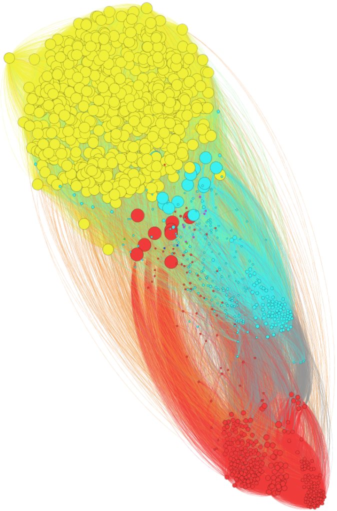
From matrix to graph using the matrix C

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



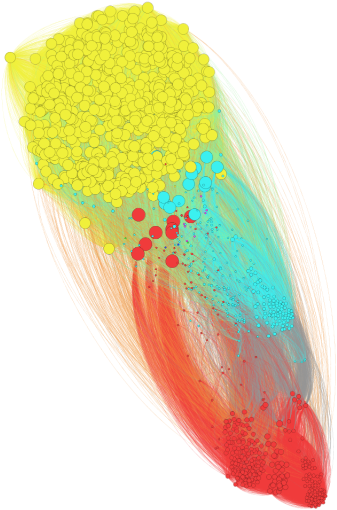
From matrix to graph using the matrix C

- 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.



Term-vectors and cosine similarity

Definition

- Fix an arbitrary ordering on the $59,908$ articles under study.
- Associate each term, with a $59,908$ dimensional vector of 1s and 0s.
  - If the term indexes the ith article then the ith entry is a 1.
  - Otherwise the ith entry is 0.
- Two terms with vectors i, j have cosine-similarity (CS),

$$\mathrm{CS}(i,j) = \frac{i \cdot j}{\sqrt{i \cdot i}\sqrt{j \cdot j}}. \tag{1}$$

Term-vectors and cosine similarity

Example

	Terms
Article 1	animal, food
Article 2	food, death and dying
Article 3	film and videos

'Animal' $= (1, 0, 0)$
'food' $= (1, 1, 0)$
'death and dying' $= (0, 1, 0)$
'film and videos' $= (0, 0, 1)$

Term-vectors and cosine similarity

Example

'animal' $= (1, 0, 0)$
'food' $= (1, 1, 0)$
'death and dying' $= (0, 1, 0)$
'film and videos' $= (0, 0, 1)$

$$
\begin{aligned}
\text{CS ('food','animal')} &= \frac{(1,1,0)\cdot(1,0,0)}{\sqrt{(1,1,0)\cdot(1,1,0)}\sqrt{(1,0,0)\cdot(1,0,0)}} \\
&= \frac{1\cdot 1 + 1\cdot 0 + 0\cdot 0}{\sqrt{1\cdot 1 + 1\cdot 1 + 0\cdot 0}\sqrt{1\cdot 1 + 0\cdot 0 + 0\cdot 0}} \\
&= \frac{1}{\sqrt{2}}
\end{aligned}
$$

Another way to calculate cosine similarity

$$
\begin{aligned}
\mathrm{CS}(i,j) &= \frac{i \cdot j}{\sqrt{i \cdot i}\sqrt{j \cdot j}} \\
&= \sqrt{\frac{C_{ij}^2}{C_{ii} C_{jj}}} \\
&= \sqrt{\frac{C_{ij}}{C_{jj}} \frac{C_{ji}}{C_{ii}}} \\
&= \sqrt{P(i \mid j) P(j \mid i)}
\end{aligned}
$$

Where $P(i \mid j)$ is the probability of finding term i in an article given that it had term j.

Example

Let

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

Then,

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

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:

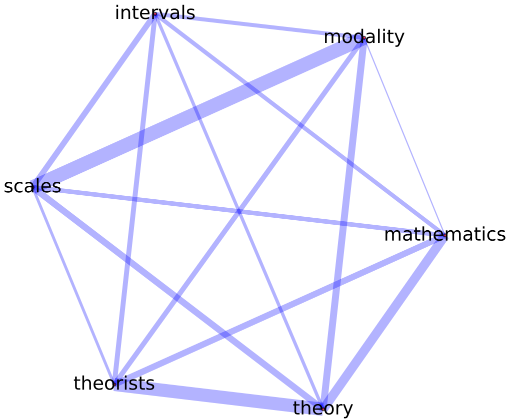
- Find two clusters of minimal distance, i.e. $\min D(c_i, c_j)$.
- Merge the clusters c_i and c_j into a cluster $c_{i\&j}$.
- Delete the clusters c_i and c_j.

We visualize Hierarchical clustering with a dendrogram.

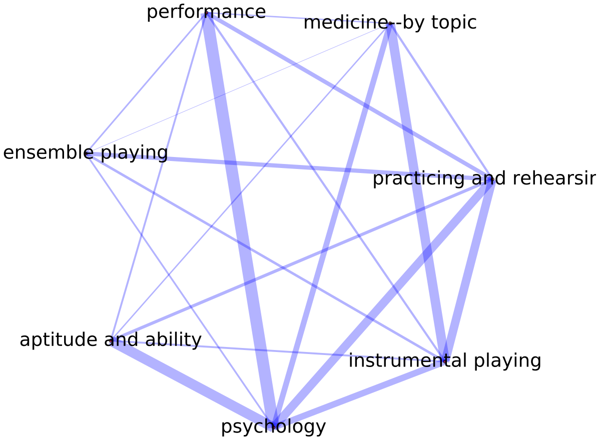
Dendrogram


nature
symbolism
visual and plastic arts
instruments
iconography
instruments--wind (woodwind reed family)
instruments--wind (brass)
instruments--wind (woodwind flute family)
instrumentation and orchestration
timbre
Japan
instruments--string (lute family)
instruments--string (zither family)
Asia
Europe
Korea
historiography
Western music
language
terminology
instruments--wind (free reed family)
ceremonies
instruments--percussion (drum)
India
dance music
dance
instruments--percussion (bells)
antiquity
temperament and tuning
Greece
instruments--collections
instruments--mechanical
dramatic arts
China
congresses conferences symposia
musicology
competitions awards and prizes
festivals
Russian Federation
performance venues
performers--violin
instruments--string (violin family)
performing groups--small ensembles
choral music
polyphony
religion and religious music--Christianity
religion and religious music--Christianity (Roman Catholic)
Italy
Austria
instruments--string (guitar family)
Spain
identity--collective
Brazil
Czech Republic
nationalism
religion and religious music--Christianity (Protestant)
United Kingdom--England
song--general
semantics
twenty or twenty-first-century music
semiotics
genre studies
vocal music
instrumental music
catalogues and indexes--by name
discographies
song--by composer
Bartók Béla
Hungary
ethnomusicologists
ethnomusicology
Lithuania
song--popular and traditional
correspondence
libraries museums collections
periodicals
criticism
France
reception
wars and catastrophes
politics
cultural policies
Germany
Canada
societies associations fraternities etc.--nat.
performers--organ
performers--traditional and neotraditional music
performers--voice
musicologists
obituaries
performers--piano
pedagogues
academic institutions
anecdotes reminiscences etc.
jazz
performers--jazz and blues
improvisation
libretto--by librettist
literature
premieres
musical theater--popular
Čajkovskij Pëtr Il'ič
choreographers
dramaturgy
Wagner Richard
Verdi Giuseppe
opera
performers--conducting
performing organizations
directors producers etc.
film music and television music
film and video
ensemble playing
practicing and rehearsing
instrumental playing
medicine--by topic
performance
psychology
aptitude and ability
conducting
movement and gesture
amateurs
Australia
audiences
careers
pedagogy
children
therapy
voice
anatomy and physiology
acoustics
singing
pitch
perception
motive and theme
form
texture
tonality
harmony
melody
rhythm and meter
text setting
poetry
scales
modality
intervals
theory
theorists
mathematics
serialism
Schoenberg Arnold
Stravinsky Igor
creative process
counterpoint
computer applications
analysis
philosophy
aesthetics
science
Debussy Claude
time
mixed media
sonority
electronic music and computer music
composition
instruments--keyboard
Bach Johann Sebastian
tempo
performance practice
Haydn Joseph
Mozart Wolfgang Amadeus
Beethoven Ludwig van
Liszt Franz
arrangements
Chopin Frédéric
interpretation
accompaniment

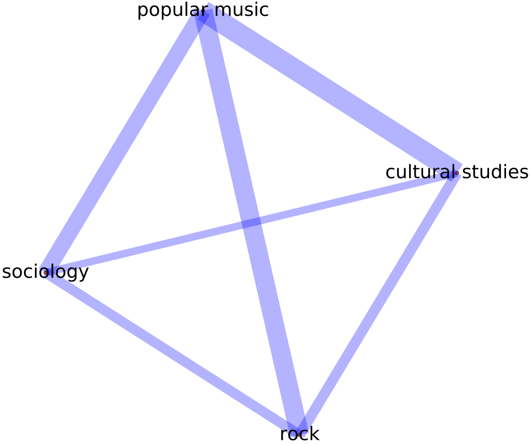
Clusters: Theory



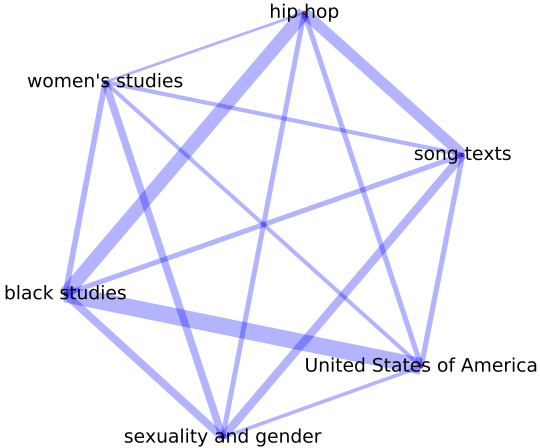
Clusters: Performance



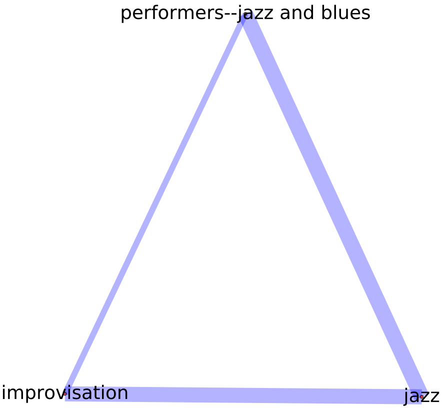
Clusters: Rock



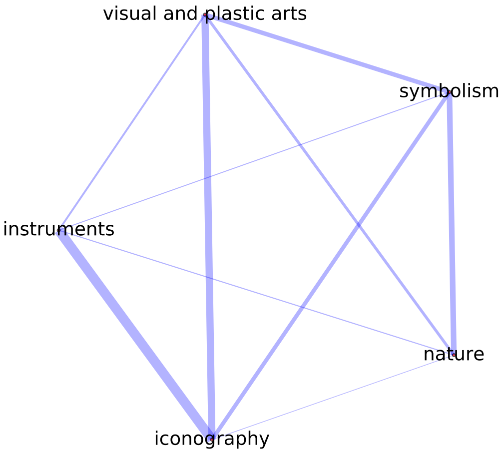
Clusters: Black Studies and United States



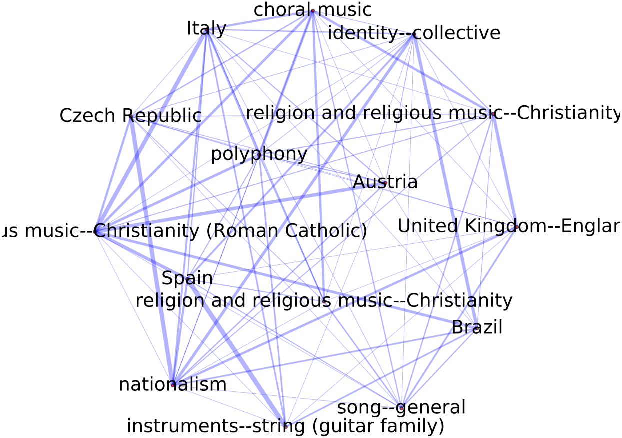
Clusters: Jazz



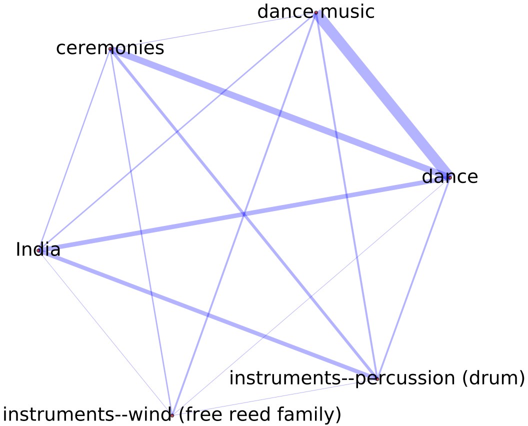
Clusters: Iconography



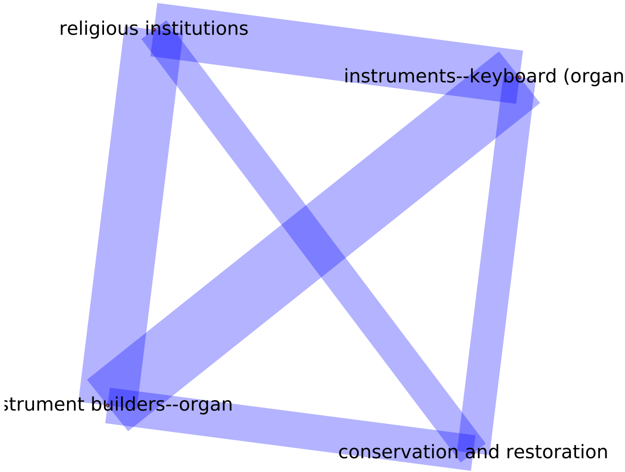
Clusters: Religious Music



Clusters: Dance



Clusters: Organ Music and Restoration

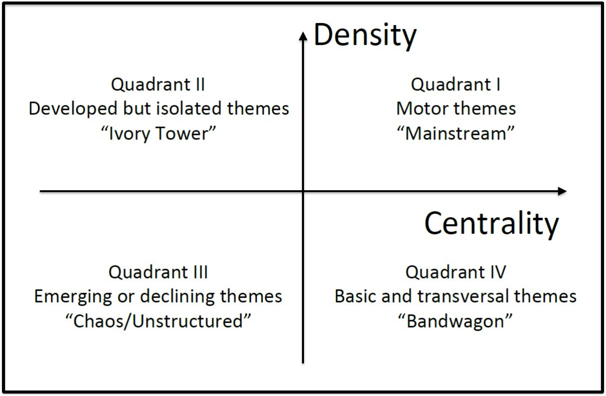


Density and Centrality

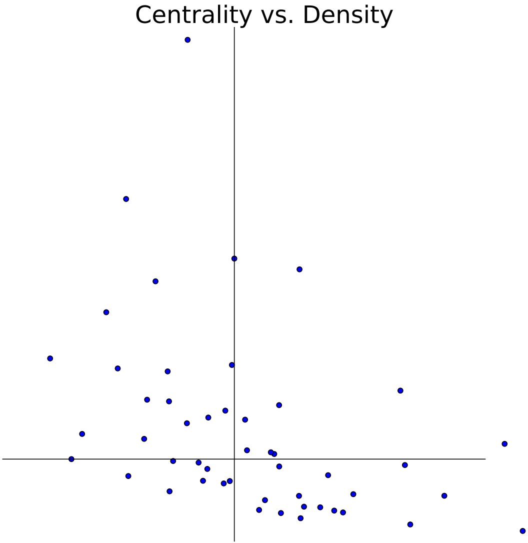
For each cluster we calculate:

- **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

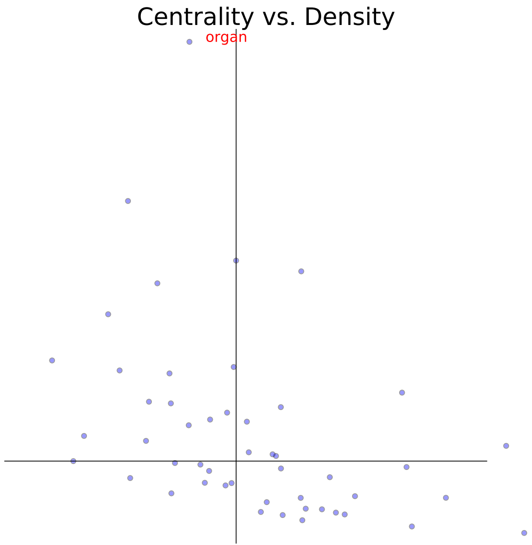
Strategic Diagram



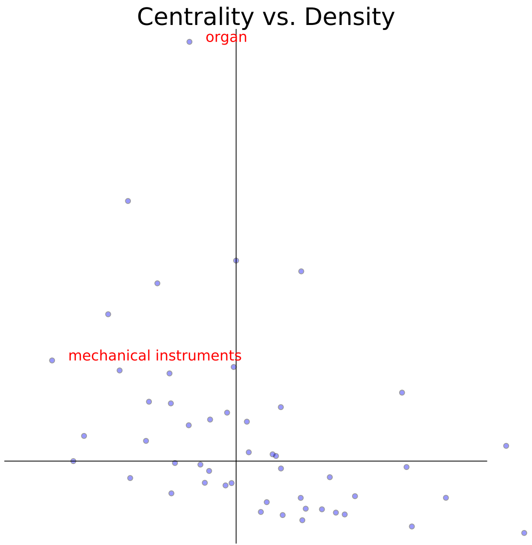
Strategic Diagram



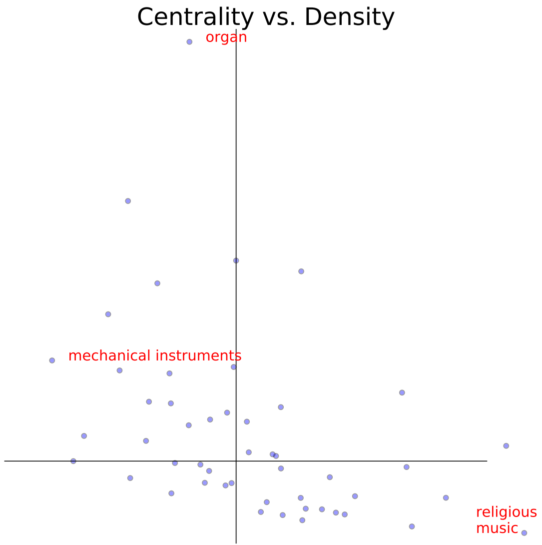
Strategic Diagram



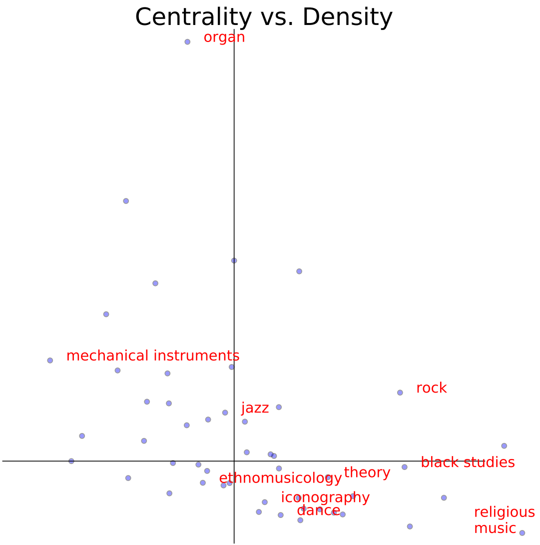
Strategic Diagram



Strategic Diagram



Strategic Diagram



Thank You!