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

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

<table>
<thead>
<tr>
<th>Term</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>pedagogy</td>
<td>5499</td>
</tr>
<tr>
<td>China</td>
<td>5084</td>
</tr>
<tr>
<td>instruments--keyboard (organ family)</td>
<td>3011</td>
</tr>
<tr>
<td>popular music</td>
<td>2388</td>
</tr>
<tr>
<td>aesthetics</td>
<td>2037</td>
</tr>
<tr>
<td>singing</td>
<td>2024</td>
</tr>
<tr>
<td>song--popular and traditional</td>
<td>1886</td>
</tr>
<tr>
<td>pedagogues</td>
<td>1844</td>
</tr>
<tr>
<td>performing organizations</td>
<td>1836</td>
</tr>
<tr>
<td>religious institutions</td>
<td>1788</td>
</tr>
<tr>
<td>instrument builders--organ</td>
<td>1773</td>
</tr>
<tr>
<td>sound recordings</td>
<td>1734</td>
</tr>
<tr>
<td>academic institutions</td>
<td>1657</td>
</tr>
<tr>
<td>instruments--keyboard (piano family)</td>
<td>1639</td>
</tr>
<tr>
<td>psychology</td>
<td>1608</td>
</tr>
</tbody>
</table>
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' appear together in 123 articles.

$$C_{'aesthetics','popular music'} = 123$$
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.
Cosine similarity

Cosine similarity (CS) is defined as

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

Let

\[ i = 'China' \quad j = 'popular music' \]
\[ k = 'mathematics' \quad l = 'scales'. \]

Then,

\[ C_{ij} = 151 \quad CS(i, j) = \sqrt{\frac{151^2}{5084 \cdot 2388}} = .04 \]
\[ C_{kl} = 15 \quad CS(k, l) = \sqrt{\frac{15^2}{309 \cdot 303}} = .05. \]
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)
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

- Quadrant II: Developed but isolated themes “Ivory Tower”
- Quadrant I: Motor themes “Mainstream”
- Quadrant III: Emerging or declining themes “Chaos/Unstructured”
- Quadrant IV: Basic and transversal themes “Bandwagon”
Strategic Diagram

(AN) analysis
(AN) antiquity
(AS) aesthetics
(BP) Bach performance
(BS) black studies
(CA) Canada
(CD) conducting
(CG) choreographers
(CL) classical
(CN) China
(CD) correspondence
(CR) criticism
(DA) dance
(DG) discographies
(DI) direction
(EL) electronic
(ET) ethnomusicology
(FE) festivals
(FM) form
(FV) film and video
(GS) genre studies
(IC) iconography
(JZ) jazz
(KY) keyboard
(PT) literature

(LT) language and terminology
(MA) manuscripts
(ME) media
(MI) mechanical instruments
(MS) musicologists
(MT) music therapy
(OP) opera
(OR) organ
(PF) pedagogues
(PF) performance
(PG) pedagogy
(PI) pitch
(PO) poetry
(PR) popular recordings
(R) rock
(RM) religious music
(S) songs
(SC) Schubert
(SR) serialism
(TH) theory
(TI) timbre
(V) voice
(VI) violin
(WR) war
Strategic Diagram

Centrality

Density

(FV) film and video
(GS) genre studies
(IC) iconography
(JZ) jazz
(KY) keyboard
(LI) literature
(AN) analysis
(AN) antiquity
(AS) aesthetics
(BP) Bach performance
(BS) black studies
(CA) Canada
(CD) conducting
(CG) choreographers
(CL) classical
(CN) China
(CC) correspondence
(CR) criticism
(DA) dance
(DG) discographies
(DI) direction
(EL) electronic
(ET) ethnomusicology
(FE) festivals
(FM) form
(FV) film and video
(GS) genre studies
(IC) iconography
(JZ) jazz
(KY) keyboard
(LI) literature

(LT) language and terminology
(MA) manuscripts
(ME) media
(MI) mechanical instruments
(MS) musicologists
(MT) music therapy
(OP) opera
(OR) organ
(PE) pedagogues
(PF) performance
(PG) pedagogy
(PM) pitch
(PQ) poetry
(PR) popular recordings
(R) rock
(RM) religious music
(S) songs
(SC) Schubert
(SR) serialism
(ST) theory
(TI) timbre
(V) voice
(VI) violin
(WR) war
Cluster: Form

- texture
- rhythm and meter
- motive and theme
- tonality
- form
- melody
- harmony
Strategic Diagram
Cluster: Popular Recordings

performers--popular music

sound recordings

performing groups--pop
Strategic Diagram
Cluster: Rock

- Sociology
- Cultural studies
- Popular music
Cluster: Mechanical instruments
Cluster: Classical

Mozart Wolfgang Amadeus

Beethoven Ludwig van

Haydn Joseph
Cluster: Organ

- conservation and restoration
- religious institutions
- instruments--keyboard (organ)
- instrument builders--organ
Strategic Diagram

(AN) analysis
(AN) antiquity
(AS) aesthetics
(BP) Bach performance
(BS) black studies
(CA) Canada
(CD) conducting
(CG) choreographers
(CL) classical
(CN) China
(CD) correspondence
(CR) criticism
(DA) dance
(DG) discographies
(DI) direction
(EL) electronic
(ET) ethnomusicology
(FE) festivals
(FM) form
(FV) film and video
(GS) genre studies
(IC) iconography
(JZ) jazz
(KY) keyboard
(LI) literature

(LT) language and terminology
(MA) manuscripts
(ME) media
(MI) mechanical instruments
(MS) musicologists
(MT) music therapy
(OP) opera
(OR) organ
(PE) pedagogues
(PF) performance
(PG) pedagogy
(PI) pitch
(PQ) poetry
(PR) popular recordings
(R) rock
(RM) religious music
(S) songs
(SC) Schubert
(SR) serialism
(TH) theory
(TI) timbre
(V) voice
(VI) violin
(WR) war
Cluster: Antiquity

- instruments--percussion (bells)
- Antiquity
- Greece
- temperament and tuning
Cluster: Performance

- Ensemble playing
- Performance
- Practicing and rehearsing
- Instrumental playing
- Psychology
- Medicine--by topic
- Aptitude and ability
Strategic Diagram

Centrality

Density

(AN) analysis
(AN) antiquity
(AS) aesthetics
(BP) Bach performance
(BS) black studies
(CA) Canada
(CD) conducting
(CG) choreographers
(CL) classical
(CN) China
(CO) correspondence
(CR) criticism
(DA) dance
(DG) discographies
(DI) direction
(EL) electronic
(ET) ethnomusicology
(FE) festivals
(FF) form
(FV) film and video
(GS) genre studies
(IC) iconography
(J2) jazz
(KY) keyboard
(LI) literature

(LT) language and terminology
(MA) manuscripts
(ME) media
(MI) mechanical instruments
(MS) musicologists
(MT) music therapy
(OP) opera
(GR) organ
(PE) pedagogues
(PF) performance
(PG) pedagogy
(PI) pitch
(PR) poetry
(PR) popular recordings
(R) rock
(RM) religious music
(S) songs
(SC) Schubert
(SR) serialism
(TH) theory
(TT) timbre
(V) voice
(VI) violin
(WR) war
Cluster: Analysis

- Analysis
- Computer applications
- Counterpoint
- Creative process
- Stravinsky Igor
Strategic Diagram
Cluster: Theory

theorists  mathematics

theory  scales

modality  intervals
Cluster: Black studies

- Black studies
- Sexuality and gender
- Women's studies
- Hip hop
- United States of America
- Song texts
Cluster: Iconography

- iconography
- instruments
- nature
- symbolism
- visual and plastic arts
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!