

Co-Word Analysis of Chinese Ethnomusicology 2002-2015

Glenn Henshaw, Fan Yun

August 15, 2016

Goal: Map the knowledge space of a given field

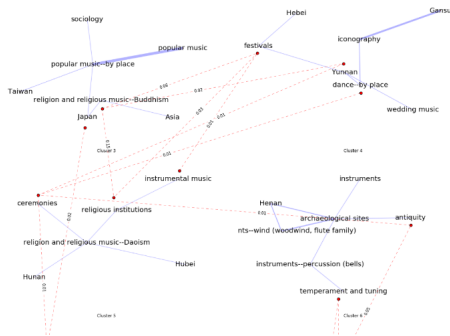
- ▶ How can the field be divided into subfields?
- ▶ What are the principle subfields?
- ▶ What are the emerging subfields?
- ▶ How do these subfields connect/interact?

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.



What is Ethnomusicology?

Ethnomusicology is the study of music in its cultural context. Ethnomusicologists approach music as a social process in order to understand not only what music is but why it is: what music means to its practitioners and audiences, and how those meanings are conveyed.

-Society for Ethnomusicology

What is RILM?

- ▶ RILM Abstracts of Music Literature is a comprehensive music bibliography features citations, abstracts, and subject indexing of all formats of music scholarship in any language. RILM indexing represents subjects covered in a given publication and their hierarchical relationships with broader and narrower topics. By June 2016, there are more than 51,000 Chinese language records in RILM covering all types of publications from Mainland China, Taiwan, and Hong Kong.

The Data

	RY	AC	lvl1	lvl2	cat1
accID					
185886	2002	5580	Latin America	Caribbean region	G
185886	2002	5580	instruments--percussion	percussion instruments	I
185886	2002	5580	dance--by place	Caribbean region	T
185886	2002	5580	performers--popular music	Caribbean region	T
112116	2002	5586	ethnomusicologists	Yang, Yinliu	T
112116	2002	5586	Yang, Yinliu	writings	N
79733	2002	5598	China	religious music	G
79733	2002	5598	cultural policies	China	T
199502	2002	5599	Japan	traditional music	G
199502	2002	5599	China	traditional music	G

- ▶ Almost all Chinese ethnomusicology articles (2002-2015)
- ▶ 6778 Rows
- ▶ 1998 Articles
- ▶ 872 distinct terms

10 Most common terms

	Counts	Terms
0	1604	China
1	264	ethnomusicology
2	214	dramatic arts
3	192	song--popular and traditional
4	148	antiquity
5	129	ceremonies
6	97	cultural studies
7	91	melody
8	87	instruments--string (zither family)
9	72	modality

10 Most common terms without “China”

	Counts	Terms
0	264	ethnomusicology
1	258	traditional music
2	216	dramatic arts
3	197	history of music
4	192	song--popular and traditional
5	156	antiquity
6	129	ceremonies
7	97	cultural studies
8	91	melody
9	87	instruments--string (zither family)

Coincidence Matrix

Definition

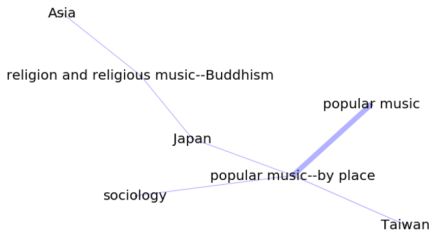
For two terms, i and j , C_{ij} is the number of times they appear together as terms of a paper. C is the $N \times N$ matrix obtained in this way.

Example

For the most common 10 terms, we have the following coincidence matrix.

	0	1	2	3	4	5	6	7	8	9
0	264	35	4	3	10	2	8	21	2	0
1	35	258	44	4	35	3	9	14	28	15
2	4	44	216	10	8	6	8	7	41	0
3	3	4	10	197	8	9	19	4	3	10
4	10	35	8	8	192	6	7	8	15	2
5	2	3	6	9	6	156	15	8	4	7
6	8	9	8	19	7	15	129	7	2	0
7	21	14	7	4	8	8	7	97	0	4
8	2	28	41	3	15	4	2	0	91	2
9	0	15	0	10	2	7	0	4	2	87

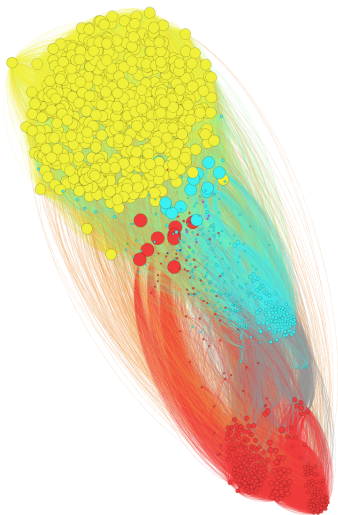
From matrix to graph



- ▶ The nodes of the graph are terms
- ▶ The thickness of the lines between terms is proportional to the values of whatever matrix that is being used.

From matrix to graph using the matrix C

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



The strength matrix, S : a more refined approach

Definition

For terms i and j ,

$$S_{ij} = \frac{(C_{ij})^2}{C_{ii}C_{jj}}.$$

S_{ij} can be interpreted probabilistically as $P(i | j) \cdot P(j | i)$

The strength matrix, S : a more refined approach

Definition

For terms i and j ,

$$S_{ij} = \frac{(C_{ij})^2}{C_{ii} C_{jj}}.$$

S_{ij} can be interpreted probabilistically as $P(i | j) \cdot P(j | i)$

Example

Let

i = 'ethnomusicology' j = 'traditional music'
 k = 'temperament and tuning' l = 'theorists'.

Then,

$$\begin{array}{lll} C_{ij} = 35 & S_{ij} = 35^2 / (264 \cdot 258) & = .019 \\ C_{kl} = 11 & S_{kl} = 11^2 / (56 \cdot 23) & = .09. \end{array}$$

The strength matrix

- What is the strongest connection?

	0	1	2	3	4	5	6	7	8	9
0	1.000000	0.017985	0.000281	0.000173	0.001973	0.000097	0.001879	0.017221	0.000167	0.000000
1	0.017985	1.000000	0.034740	0.000315	0.024729	0.000224	0.002434	0.007832	0.033393	0.010024
2	0.000281	0.034740	1.000000	0.002350	0.001543	0.001068	0.002297	0.002339	0.085521	0.000000
3	0.000173	0.000315	0.002350	1.000000	0.001692	0.002636	0.014205	0.000837	0.000502	0.005835
4	0.001973	0.024729	0.001543	0.001692	1.000000	0.001202	0.001978	0.003436	0.012878	0.000239
5	0.000097	0.000224	0.001068	0.002636	0.001202	1.000000	0.011181	0.004229	0.001127	0.003610
6	0.001879	0.002434	0.002297	0.014205	0.001978	0.011181	1.000000	0.003916	0.000341	0.000000
7	0.017221	0.007832	0.002339	0.000837	0.003436	0.004229	0.003916	1.000000	0.000000	0.001896
8	0.000167	0.033393	0.085521	0.000502	0.012878	0.001127	0.000341	0.000000	1.000000	0.000505
9	0.000000	0.010024	0.000000	0.005835	0.000239	0.003610	0.000000	0.001896	0.000505	1.000000

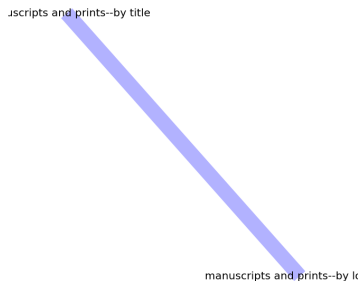
Clustering Algorithm (a rough sketch)

Initial choices:

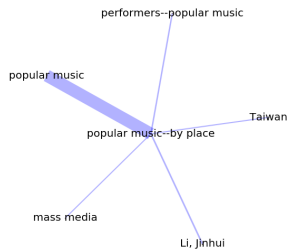
- ▶ Number of top terms to use.
 - ▶ Minimal co-occurrence
 - ▶ Minimal strength for a cluster
 - ▶ Number of clusters
1. Pick the connection with the highest strength.
 2. From those terms, add all the connections that meet the minimal strength.
 3. Add the new terms to a list.
 4. Go to step 2
 5. Steady state, go to step 1

Clusters

Manuscripts

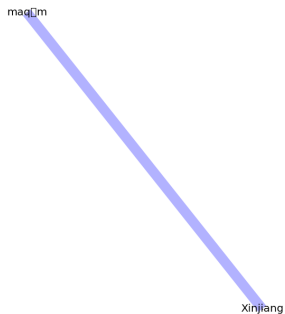


Taiwan popular music

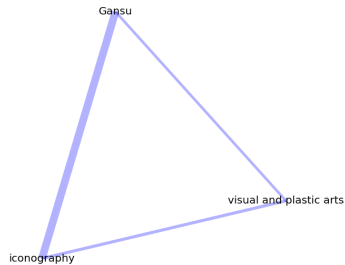


Clusters

Xinjiang Maqam

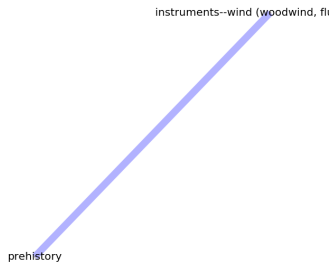


Iconography

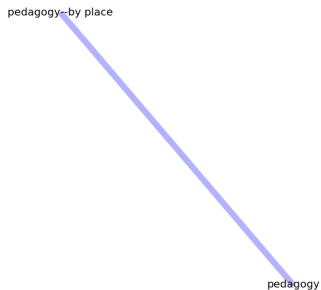


Clusters

Prehistory flute

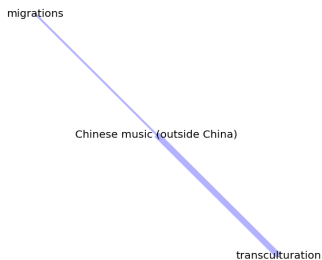


Pedagogy



Clusters

Transcuration

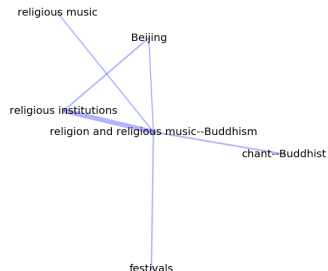


Tibetan dance

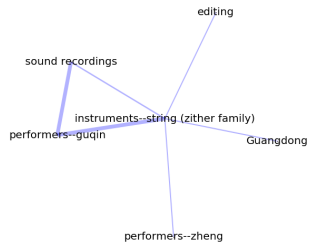


Clusters

Religious music



Guangdong Zheng music



Clusters

Yunnan dance

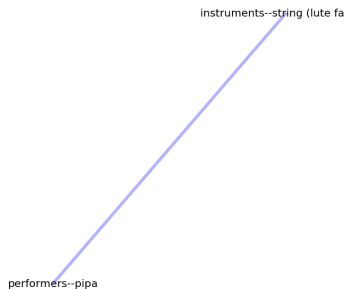


Tonality and scale studies

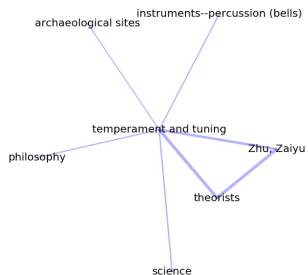


Clusters

Pipa

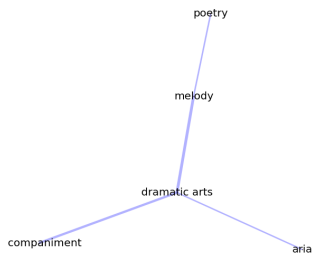


Bianzhong studies



Clusters

Dramatic arts

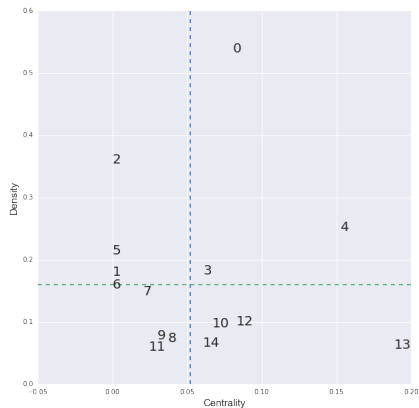


Density and Centrality

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



0 Manuscripts
1 Taiwan popular music
2 Xinjiang Maqam

3 Iconography
4 Prehistory flute
5 Pedagogy

6 Transculturation
7 Tibetan dance
8 Religious music

9 Guangdong Zheng
10 Yunnan dance
11 Tonality and scale

12 Pipa
13 Bianzhong
14 Dramatic arts

Thank You!