

Cover Music Network

A decorative graphic consisting of four stylized musical notes in orange, red, and purple, arranged horizontally to the right of the title.

Dina Mistry
Network Science/Physics
PhD Candidate
MOBS LAB
Northeastern University 2014

Cover Music Network

Musicians connected by cover songs, songs recorded by a musical act whose version was previously recorded by another musical act. Nodes are musical acts or artists and links are the songs being covered. Links go from cover act to original act.

Examples: Joe Cocker's A Little Help From My Friends (The Beatles, 1967)
 The Fugees' Killing Me Softly (Roberta Flack, 1973)

Directed and Multi-edged Network

Data Acquisition & Analysis

- webscraped whosampled.com (user generated data) using [BeautifulSoup4](#) in [Python](#)
- data spans over 4 centuries, includes artist names, record labels, recording years
- analysis: [Networkx](#), [Gephi](#) and other [Python](#) modules

Cover Network Specs

N	28118
L	73743
Songs	32504
GwCC	26389
$\langle k \rangle$	4.59
$\langle s \rangle$	5.25

Top Original Artists

(whose songs have been covered by others)

The Beatles (2495)

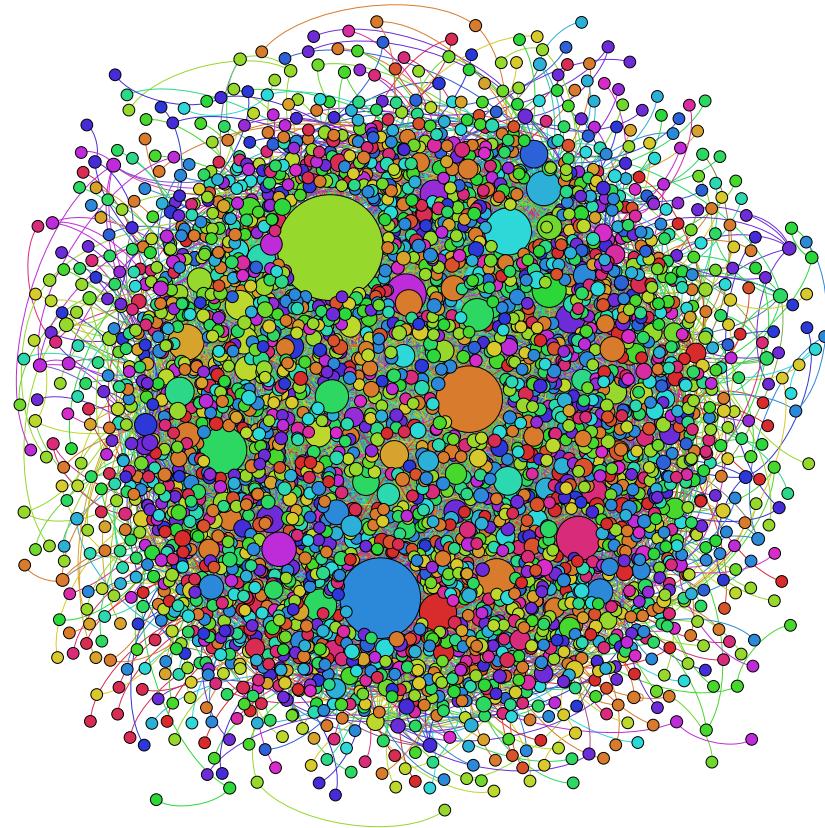
Traditional Folk (1250)

Bob Dylan (537)

Stevie Wonder (483)

Frank Sinatra (446)

The Rolling Stones (413)

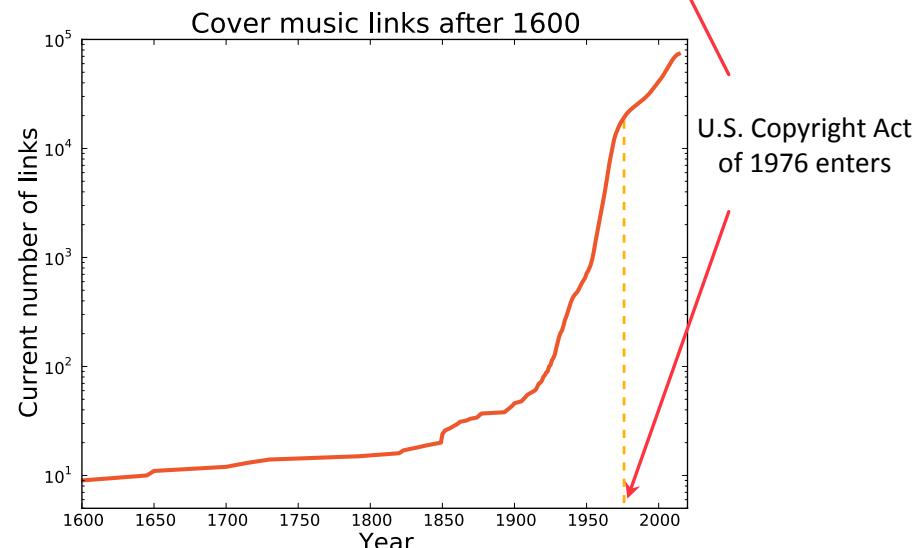
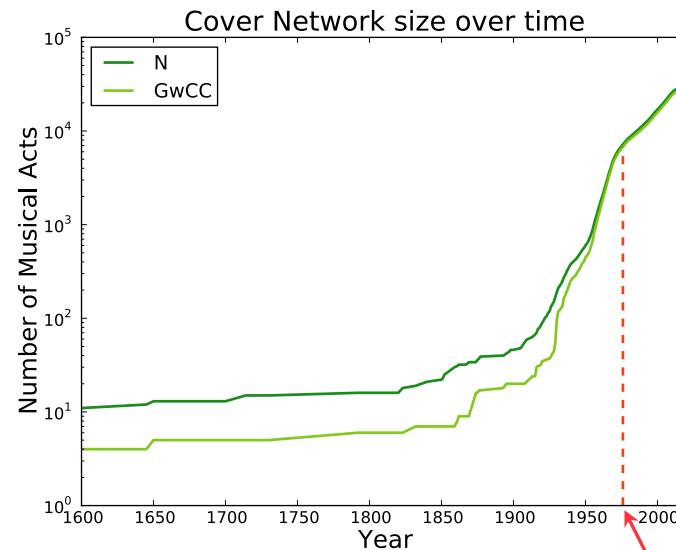


2nd largest community (3425 nodes) in cover network found using Louvain method

Network Dynamics

Cover network spans over 4 centuries. Network growth takes off after the 1930s. Growth slows in mid 1970s, just as U.S. Copyright Act of 1976 enters into law.

Greatest weakly connected component (GwCC) contains almost all nodes by 1954.

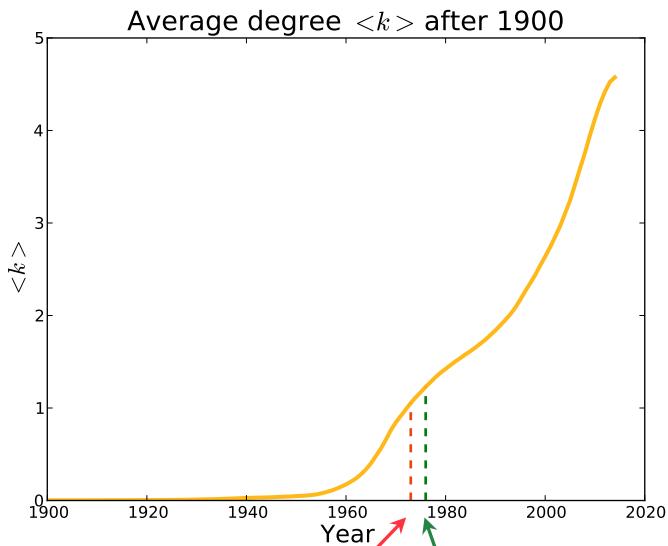
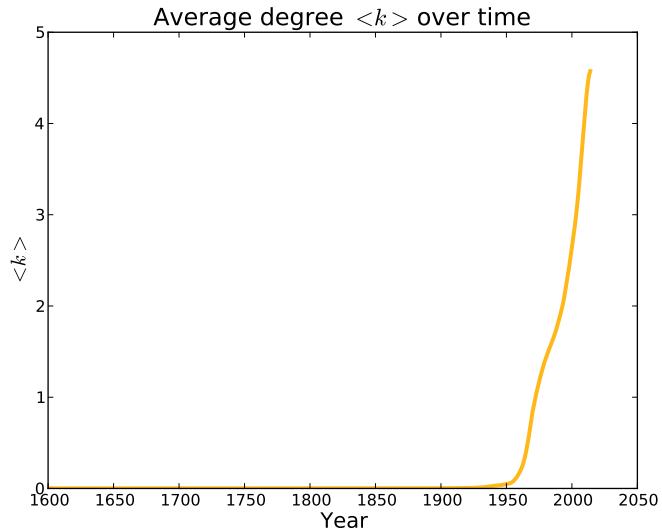


U.S. Copyright Act
of 1976 enters

Network Dynamics

Cover network mostly comprised of pairs, small trees and clusters beginning in 1600. Crosses critical point $\langle k \rangle = 1$ in 1973. Network remains in supercritical regime $\langle k \rangle > 1$ after this.

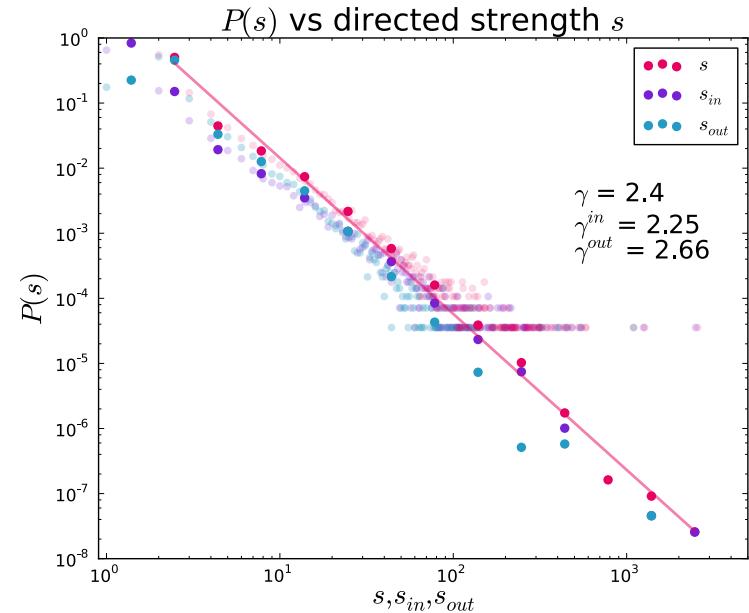
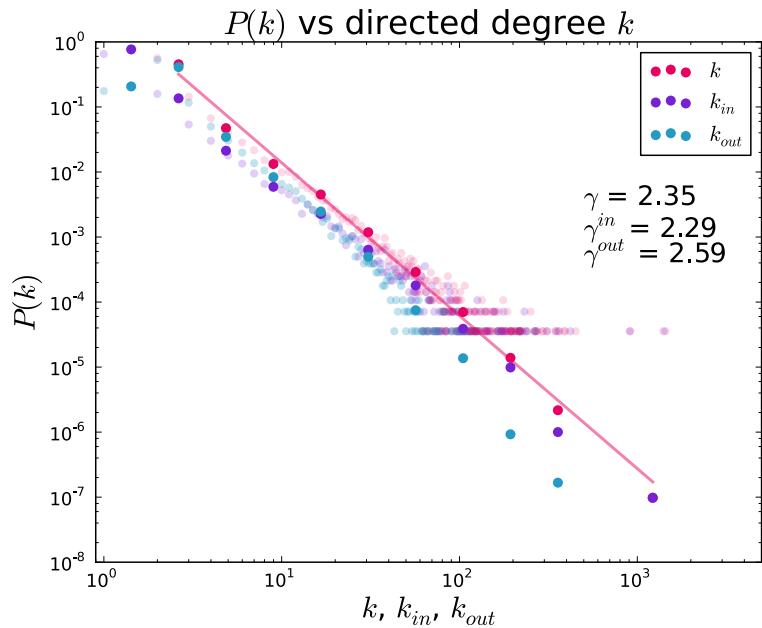
1976: U.S. Copyright Act enters into law. $\langle k \rangle$ growth slows for next two decades until late 1990s. Continues to grow faster into the Internet era.



1973: network enters supercritical regime
 $\langle k \rangle > 1$

Copyright Act of 1976

Probability Distributions: Degree k & Strength s



Results

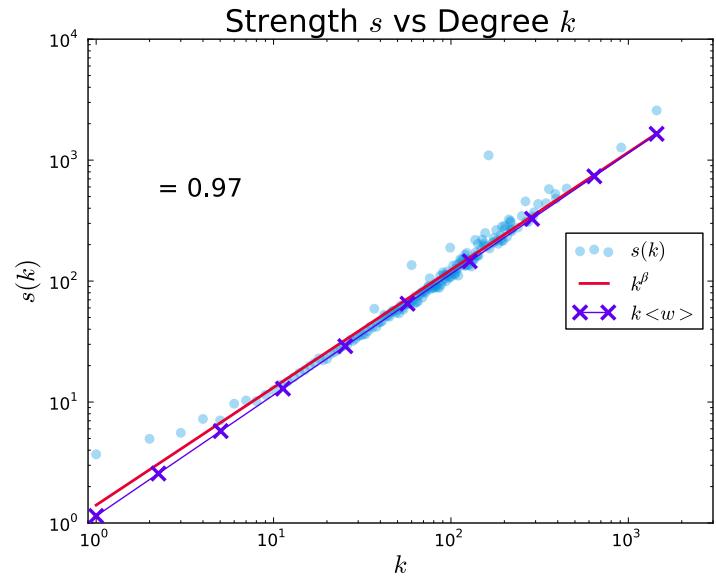
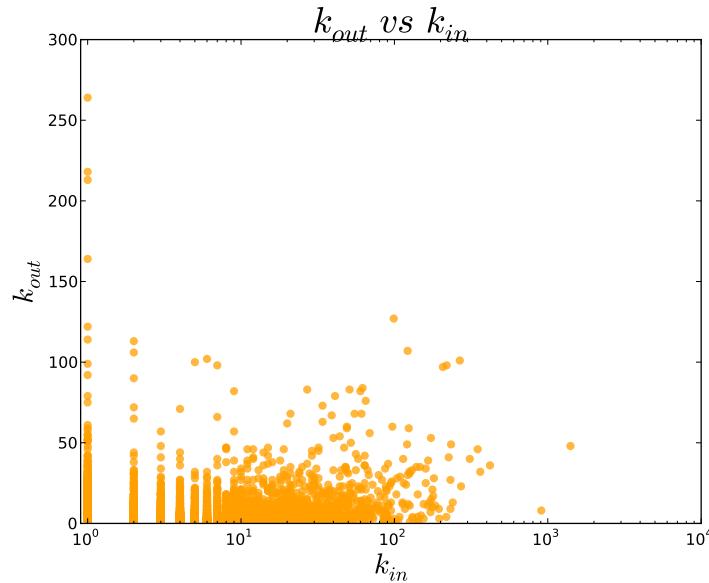
Cover network = Scale free network with $\gamma \approx 2.35$

Theory:

$$k_{\max} = k_{\min} N^{1/(\gamma-1)} = 1 \cdot 28118^{1/1.35} = 1975$$

Real $k_{\max} = 1443$

Degree k and Strength s correlations



Results

High in degree artists have lower out degrees. Well known acts are less likely to cover other artists: they already have an established name and fan base. Newcomers copy others to get recognition and boost their popularity. Strength goes almost linearly with degree – high degree artists are also those with multiple edges connecting them to others.

Degrees & Strengths: Who's important & why?

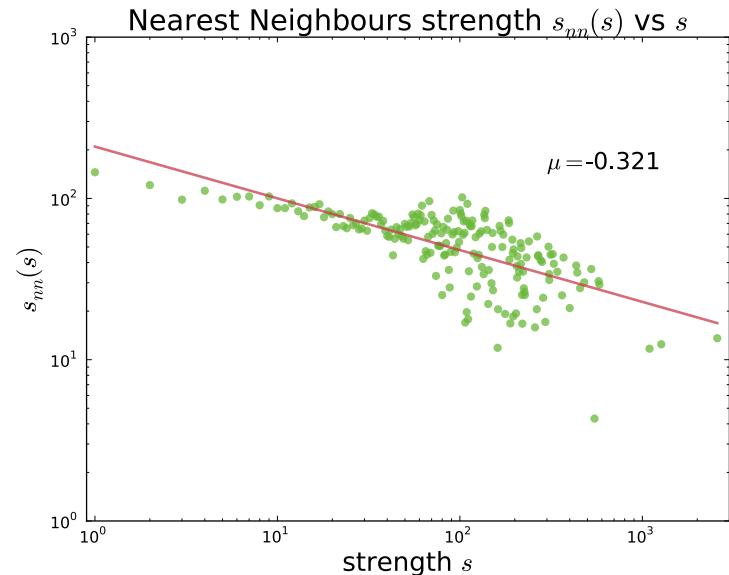
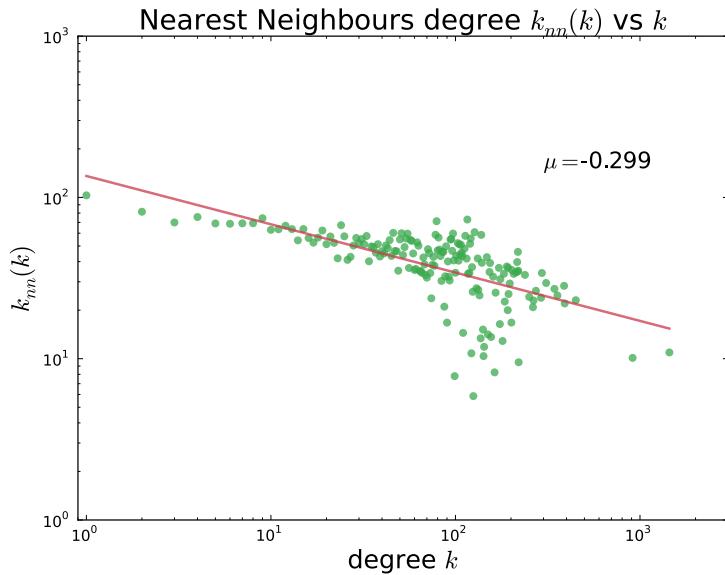
Top 10	Degree	Strength	In Strength	Out Strength
1	The Beatles	The Beatles	The Beatles	Vitamin String Quartet
2	Traditional Folk	Traditional Folk	Traditional Folk	String Tribute Players
3	Bob Dylan	Vitamin String Quartet	Bob Dylan	Kidz Bop Kids
4	Stevie Wonder	Bob Dylan	Stevie Wonder	Rockabye Baby!
5	The Rolling Stones	Frank Sinatra	Frank Sinatra	ApologetiX



Highest strength gives almost the same info as highest in strength – artists are more likely to appear in network when their songs are covered

High out strength artists are unrecognizable

Assortativity



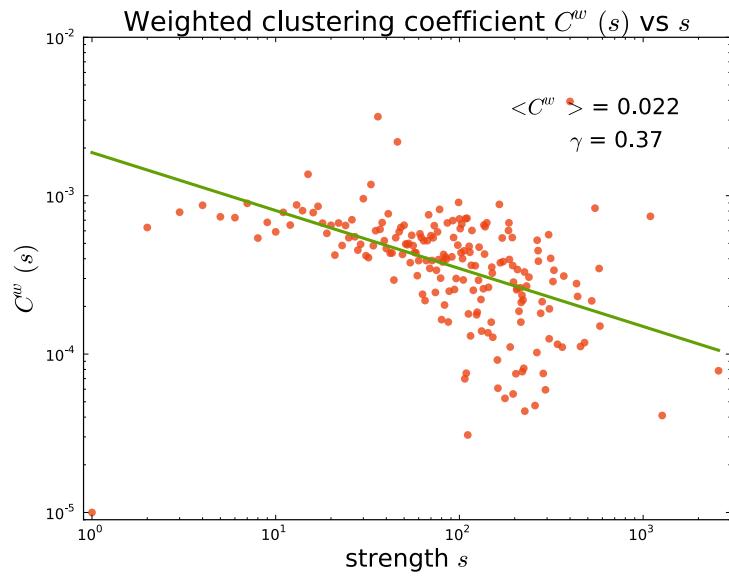
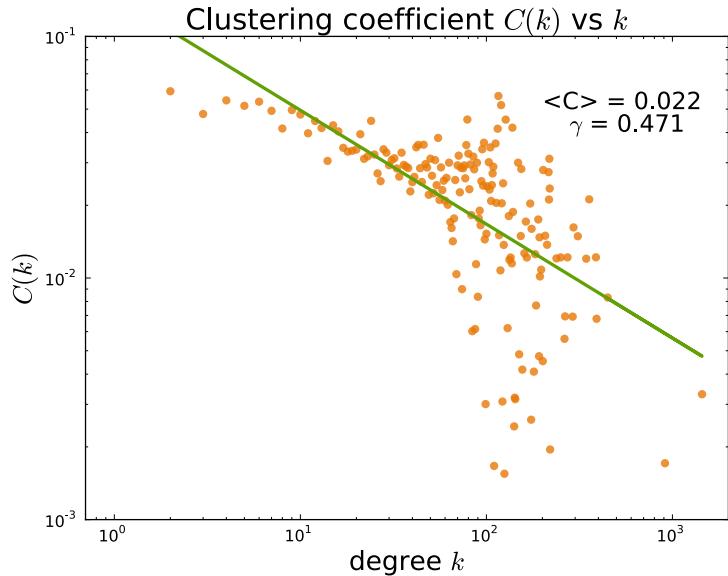
Results

Network slightly disassortative: $\mu < 0$ and $|\mu| < 1$

Hubs are more likely to connect with low degree (strength) nodes.

Why? Highest degree (strength) nodes are highly covered artists and are often covered by artists who are never covered and who only cover a few songs.

Clustering

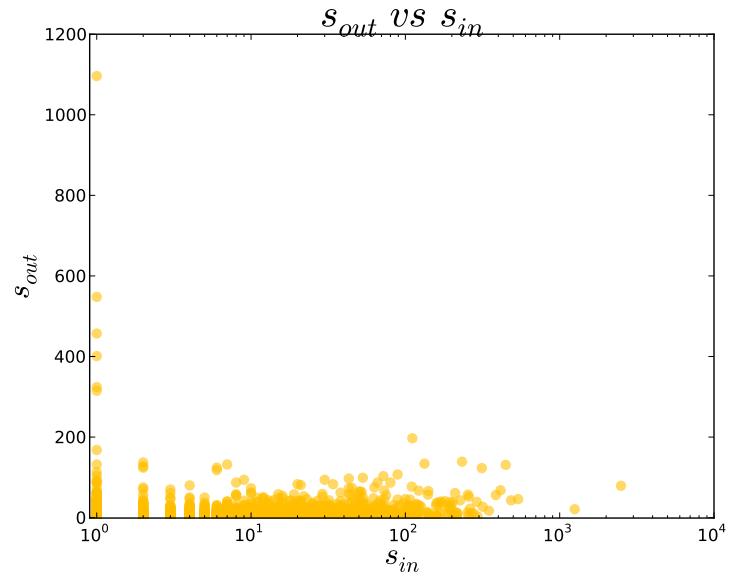
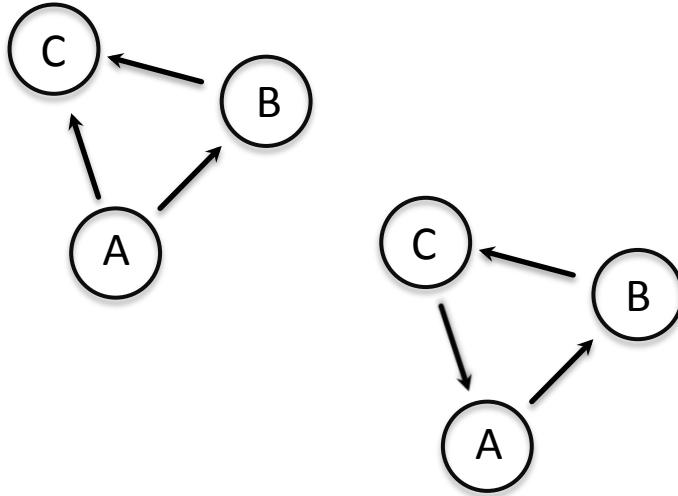


Results

Very low clustering coefficients for both unweighted and weighted network. Since clustering is related to the number of triangles in a network this means there are few of these present in the network. Why?

Clustering Explained

Two scenarios for triangles

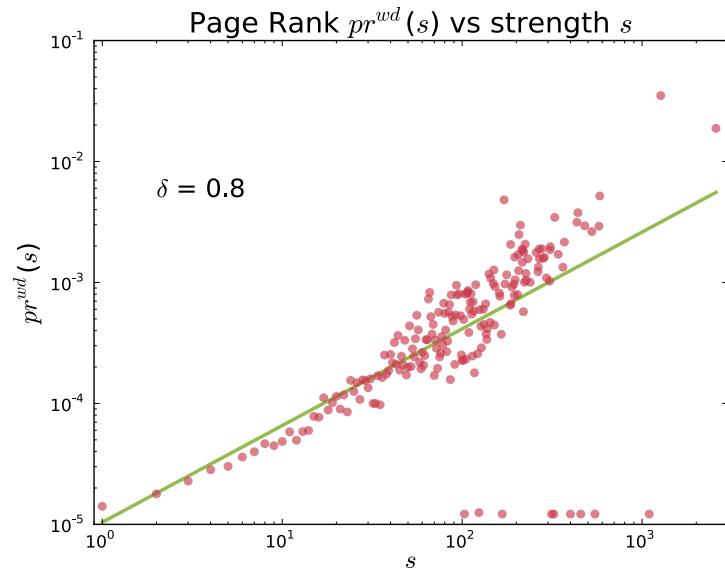
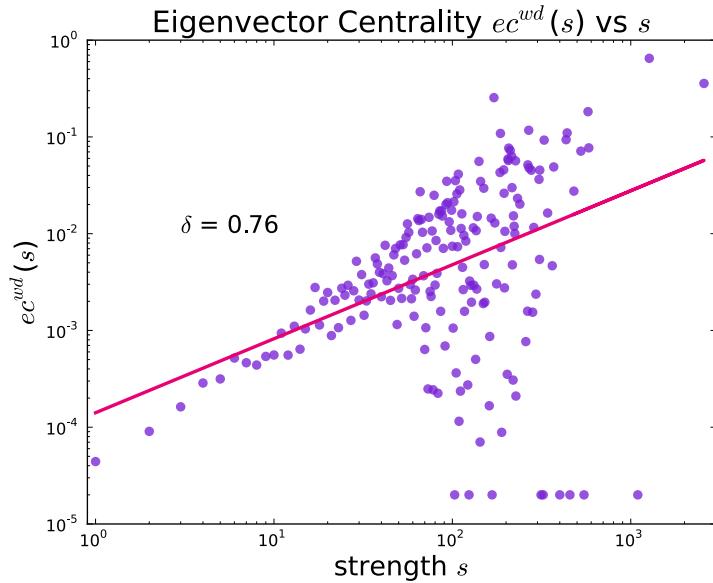


Why so few triangles?

Scenario one: on average A is a low strength artist. If A covers two higher strength artists B and C then either B needs to cover C, or vice versa, to close triangle. B has low out going strength: chances of covering C are low.

Scenario two: A covers B, B covers C, on average B or C has high strength and A has low strength. Chances of C covering A to close triangle are very low.

Centrality Distributions



Results

Eigenvector centrality gives significance to big nodes who are connected to other big nodes in the network ('big fish in a big pond'). Links are directed so some big artists have incoming links from other big artists but reverse link not guaranteed.

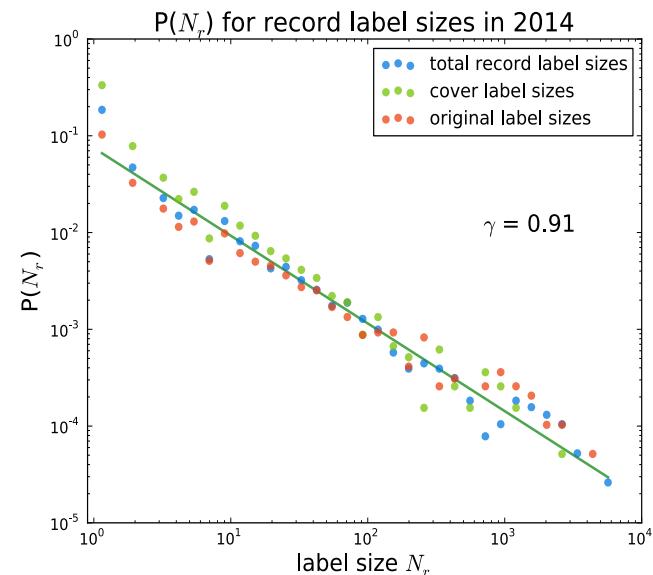
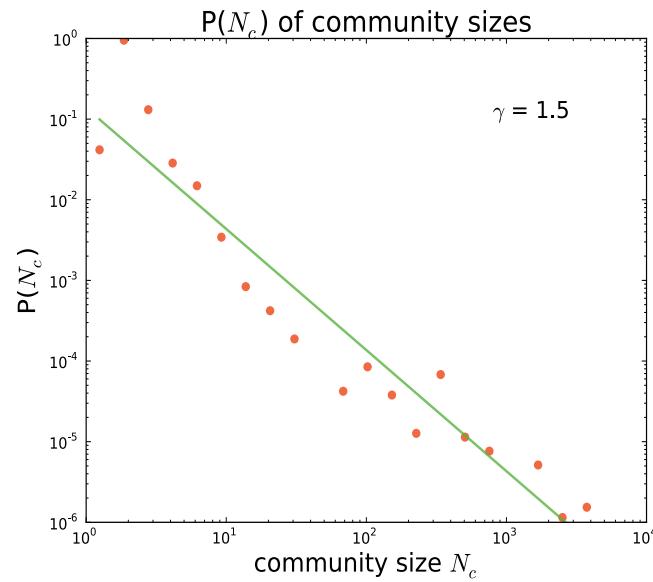
Page rank: highest ranked nodes are those most likely to be encountered while traveling along a random path in the network. Higher strength means nodes have more paths to be a part of.

Communities

Does the network exhibit communities? If so, do these communities correspond to record labels, musical genres, time periods?

Louvain Method:
878 communities (c) detected
Modularity $M = 0.535$
Largest community: 5252
 $\langle N_c \rangle = 32$

Match to record labels?
No, 13180 record labels found!



Randomization: Null Models

To understand what these results mean we need a base network to compare with.

Erdős-Rényi (ER)

same $\langle k \rangle$ and N as
real network

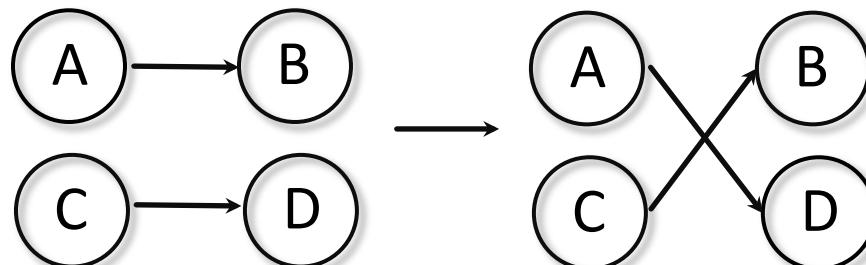
Strength Preserving Randomization (SR)

configuration model preserves
strength (directed) distribution
multi-links allowed

Time-Strength Preserving Randomization (TSR)

configuration model preserves
strength (directed) and causality:
cover artists can only
cover original artists who enter
network before them

Swap edges in pairs



For TSR method only if
 $\text{cover year}_A > \text{original year}_D$
&
 $\text{cover year}_C > \text{original year}_B$

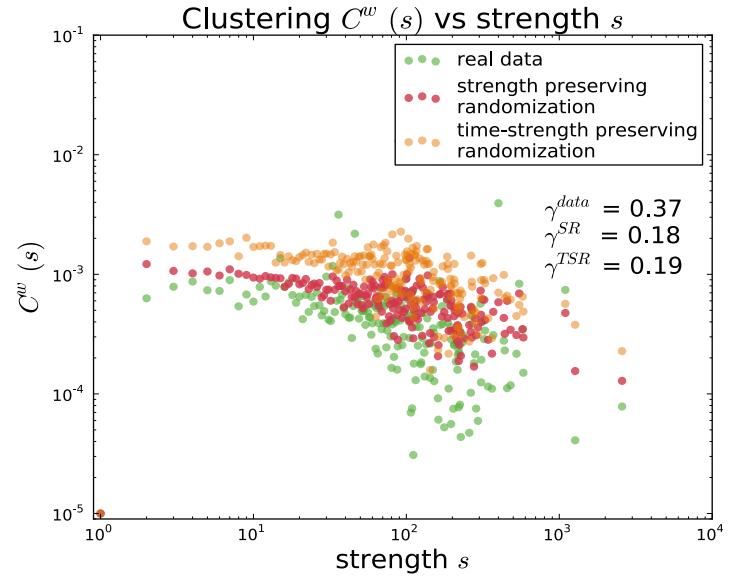
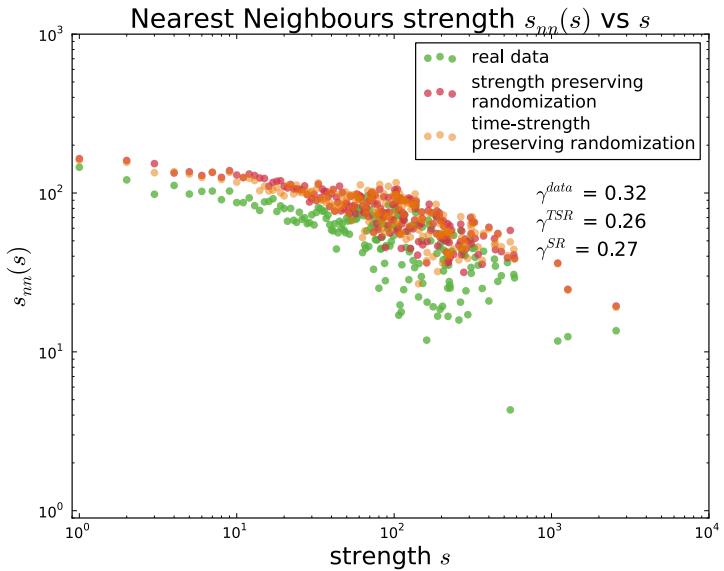
Randomization: Network Measures

	Network	TSR Null Model	SR Null Model	ER Null Model
GCC (weakly)	26389	26950.4	27014.6	27812.2
L	73743	73743	73743	64420.7
$\langle k \rangle$	4.59	5.09	5.12	4.59
$\langle s \rangle$	5.25	5.25	5.25	—
$\langle C \rangle$	2.17×10^{-2}	1.74×10^{-2}	1.55×10^{-2}	1.25×10^{-4}
$\langle C^w \rangle$	3.16×10^{-4}	7.46×10^{-4}	4.73×10^{-4}	—
Pearson's r^d	-5.34×10^{-2}	-5.33×10^{-2}	-5.31×10^{-2}	4.63×10^{-4}
c	878	564	544	60
Modularity M	0.535	0.433	0.415	0.463
$\langle N_c \rangle$	32.03	49.85	51.69	464.40

*c for the number of communities found using Louvain method

** links the same as real data: 6.63% for TSR null model, 4.67% for SR, 0.01% for ER null model

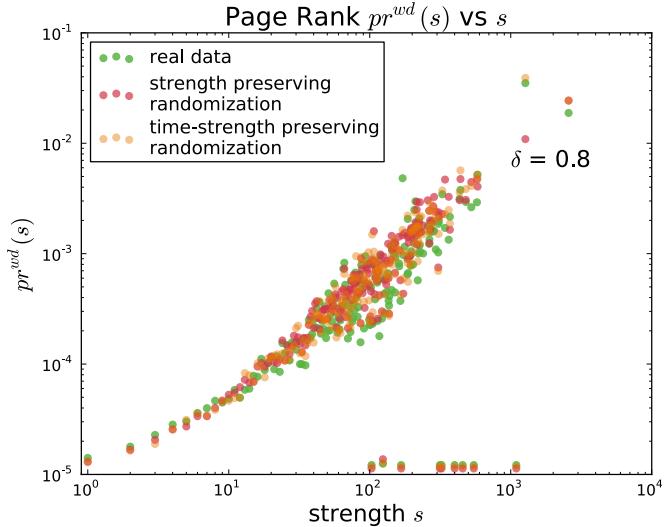
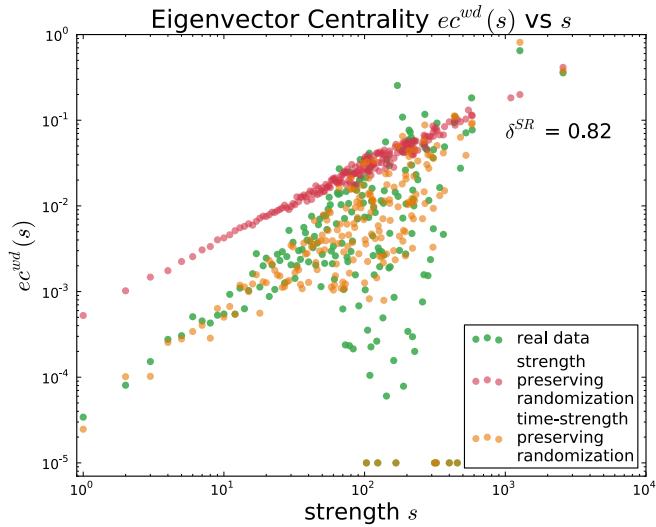
Null model comparisons



Results

Configuration models produce slightly more neutral networks. Original network's disassortativity due to strength distribution. Clustering increases slightly in configuration models – most likely due to multi-edges between artists being reshuffled.

Null model Centralities



Results

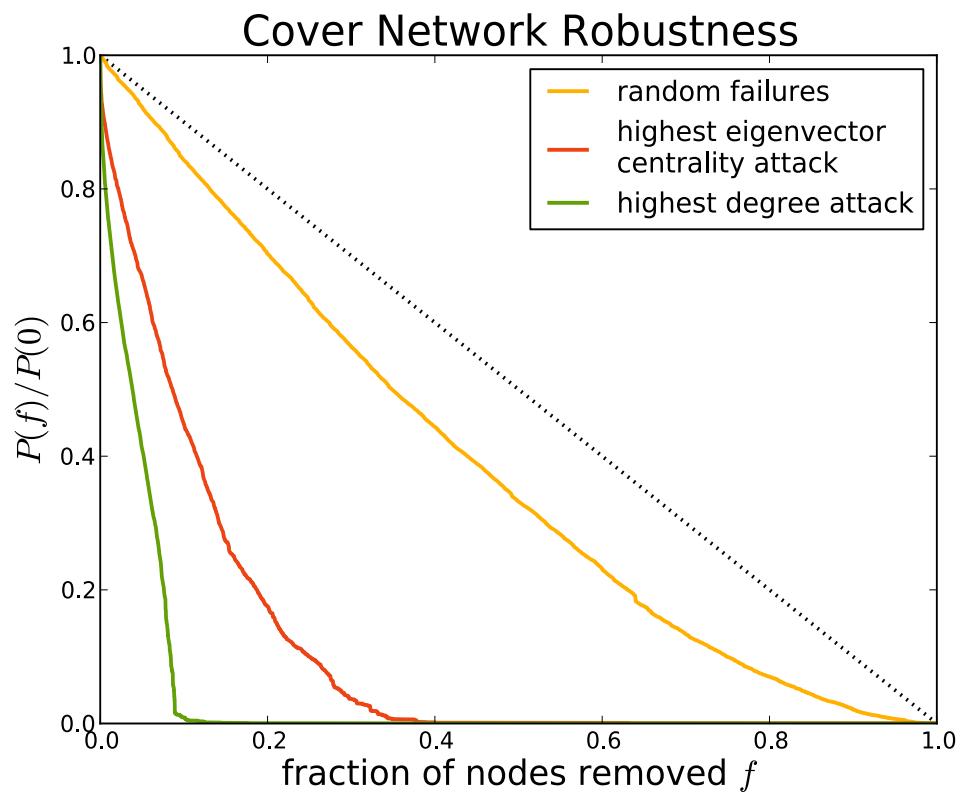
Time and strength preserving model does little to change eigenvector centrality distribution. Strength preserving edge shuffling results in eigenvector centrality increasing with strength. Page ranking remains the same – highest strength nodes are likely to be in more paths thus higher chances of ending up there.

Robustness

Network robust against random failures. Vulnerable to targeted attacks.

Eigenvector Centrality attack:
 $f = 0.35$ network falls apart

Highest degree attack even better:
 $f = 0.09$ network falls apart



Summary

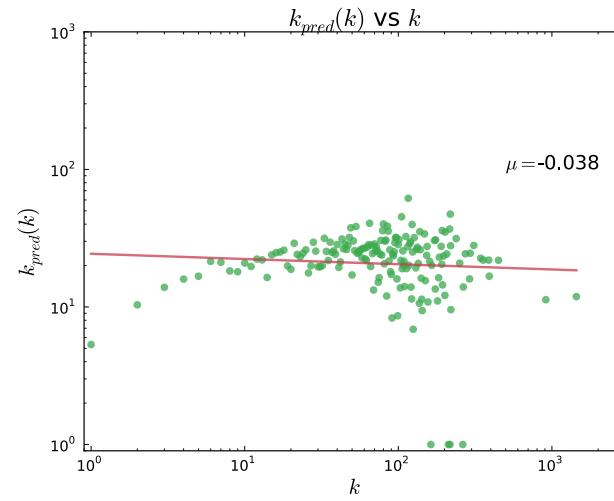
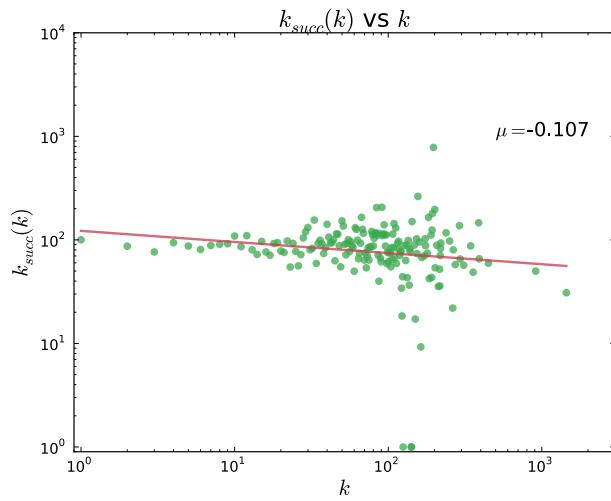
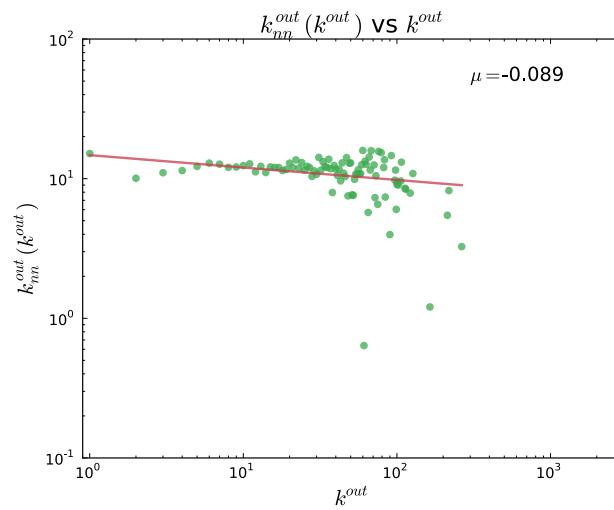
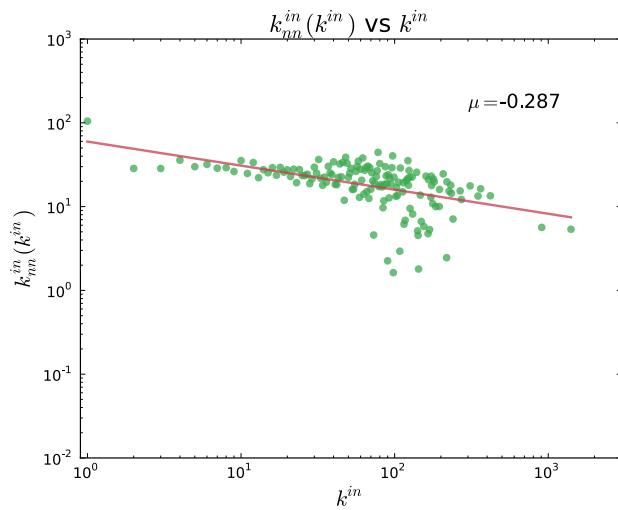
Cover music network is sparse and has very little clustering due to directed links. Centrality measures and strength indicates the network is dominated by popular artists from the late 1950s through late 1970s. More recent artists have yet to break the top 10 covered artists – preferential attachment to old standards rather than newer songs.

Future Work

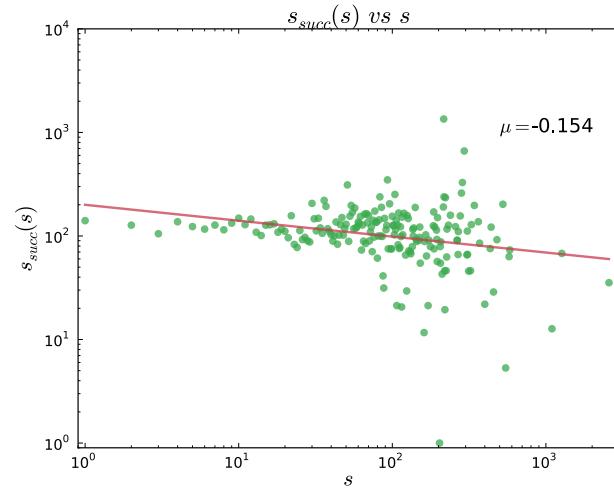
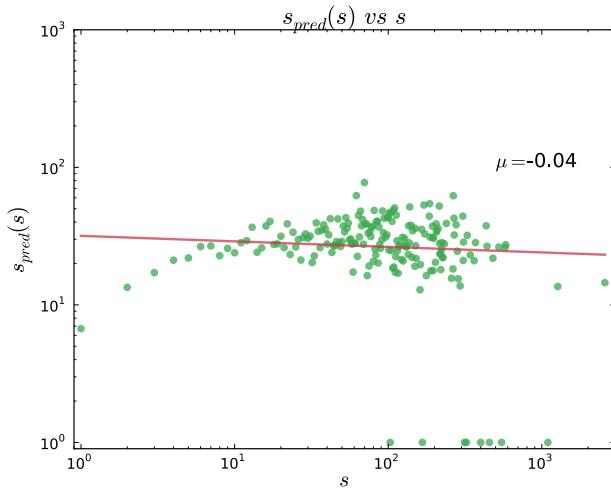
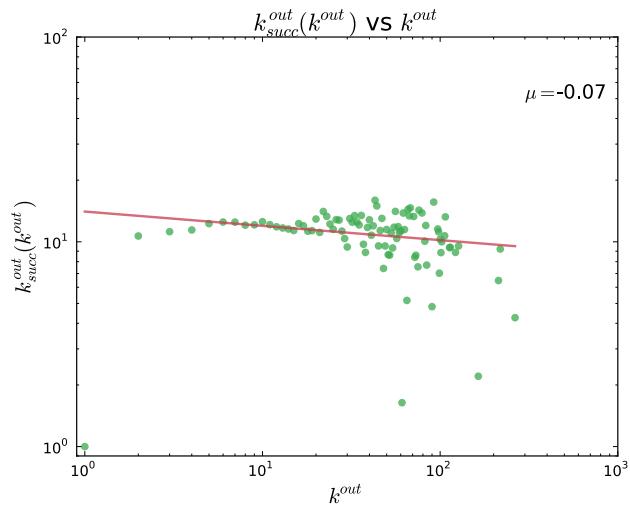
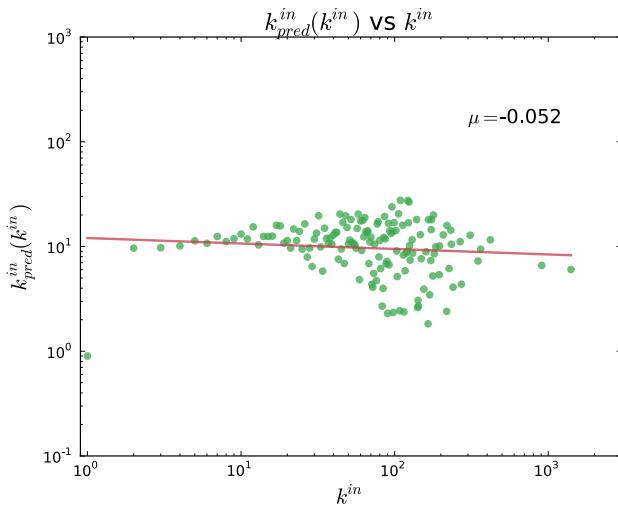
SIR related spreading model to look at adoption and cost of new recording and royalties practices on underlying record label network.

Look for or design community detection method to find different genres and time periods of music. Look for chains of influence: chain of artists covering another artist's work.

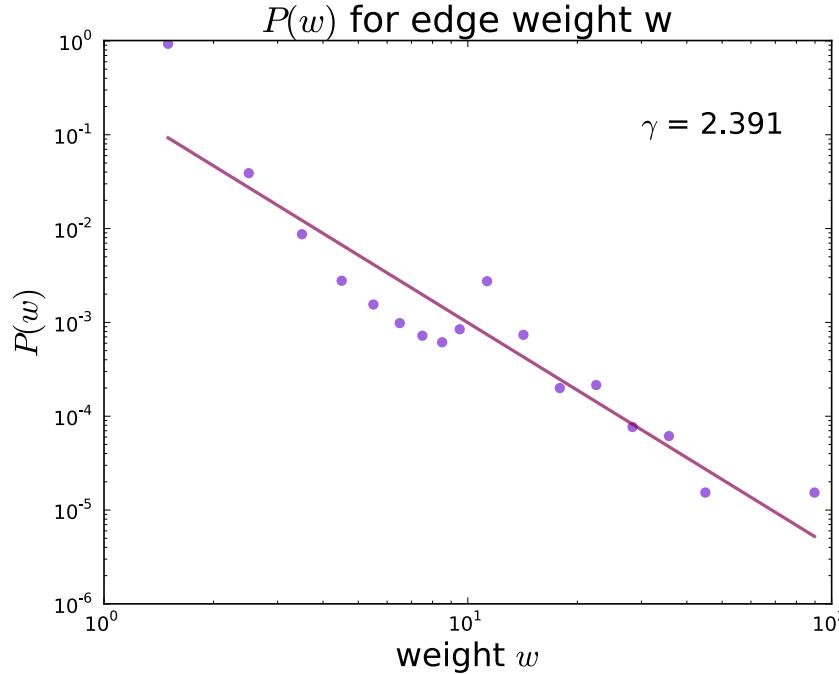
Extra slides: Directed (low) Disassortativity



Extra slides: Directed (low) Disassortativity



Extra slides: Weight Distribution



Results

Not much variability in edge weights – minimum weight = 1, maximum weight = 88. High weight edges go from relatively unknown artists to highly recognizable artists from mid 1950s through late 1970s. Quite often dedicated cover groups like Vitamin String Quartet.

Extra slides: Centralities: Who's important and why?

Top 5 Eigenvector
Centrality^{wd} Artists

Traditional Folk

The Beatles

Chuck Berry

Frank Sinatra

Little Richard

Top 5 Page Rank^{wd}
Artists

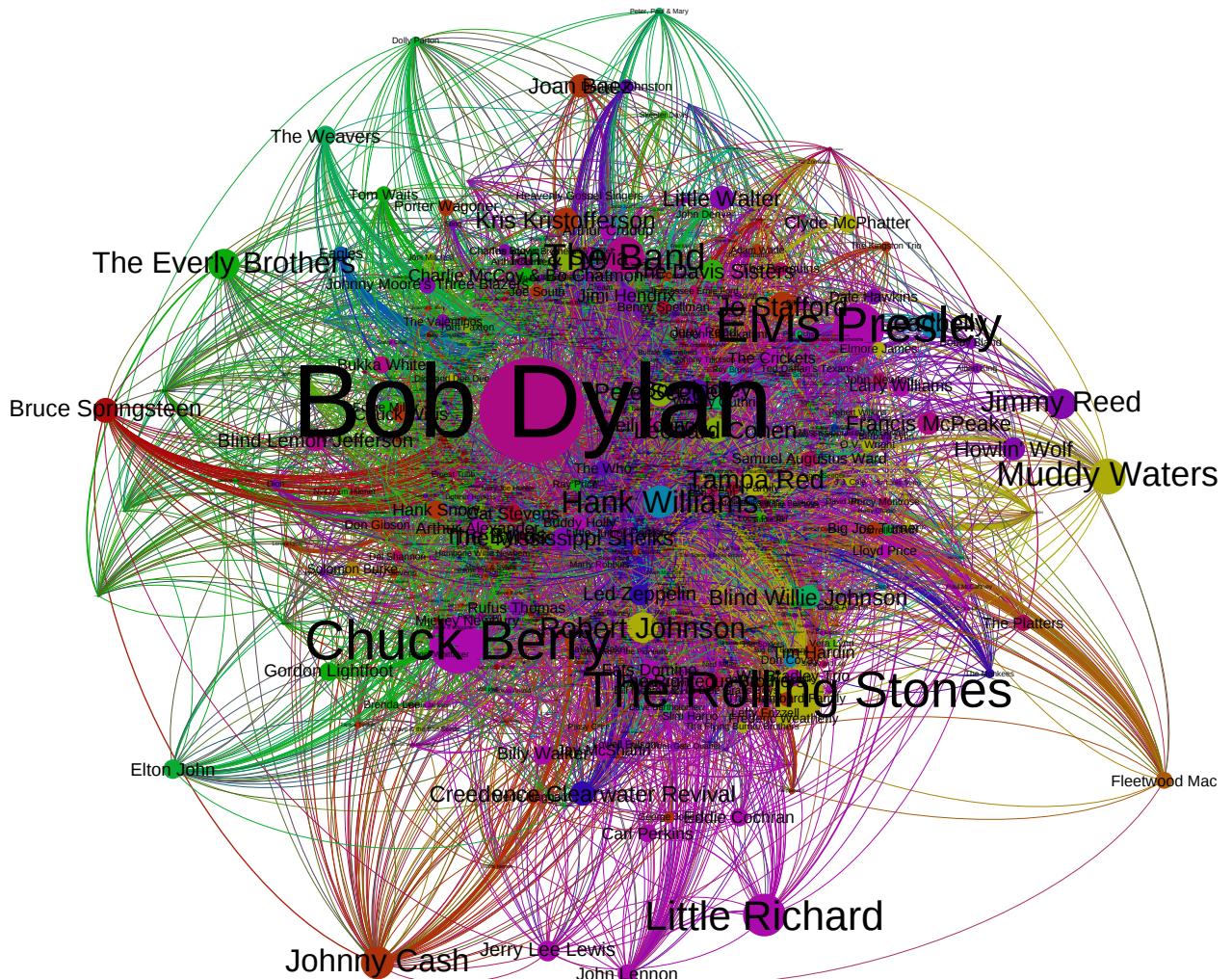
Traditional Folk

The Beatles

Bob Dylan

Marvin Gaye

The Drifters



2nd largest community found through Louvain
method for community detection with size according to
eigenvector centrality and coloured by modularity class

*w refers to weighted, d refers to directed