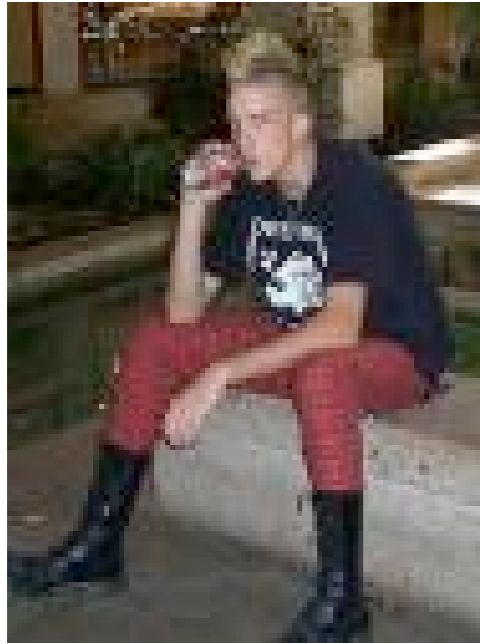


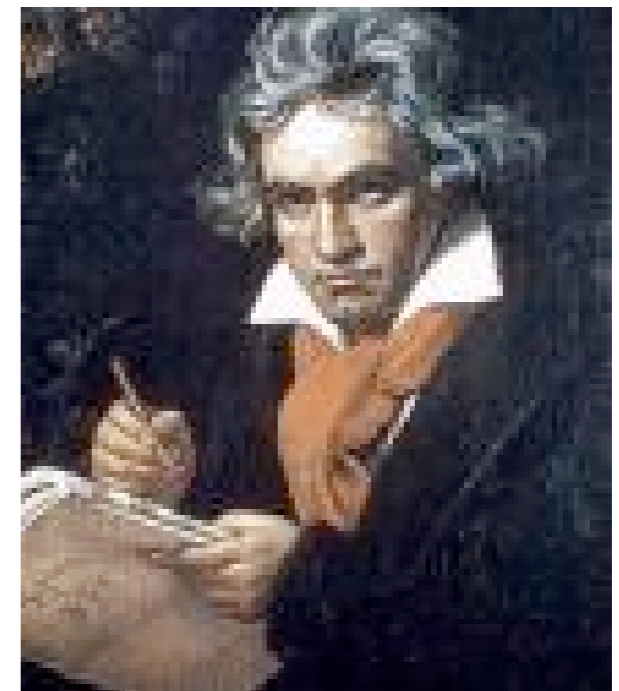
# Uncovering collective listening habits and music genres in bipartite networks

R. Lambiotte, M. Ausloos  
(Supratecs, ULG, Belgium)





What kind of music do  
you listen to?



- Short words on complex networks
- Methodology (collaborative web sites)
- Analysis of the bipartite network (listener/music group)
- Correlation matrix and network projection
- Uncovering collective behaviours (listeners communities/music trends) by:
  - o Percolation idea-based methods (PIB)
  - o Random walk exploration of the network
- Reality of usual music subdivisions?
- Personal Identification and Community Imitation (PICI) model

# I) Short words on complex networks

## In social science

The use of small (some hundreds nodes) graphs is common to represent interactions between people:

- family tree
- network of acquaintances
- network of sexual relations

....

Chains of Affection: The Structure of Adolescent Romantic and Sexual Networks:  
Peter S. Bearman, James Moody  
and Katherine Stovel

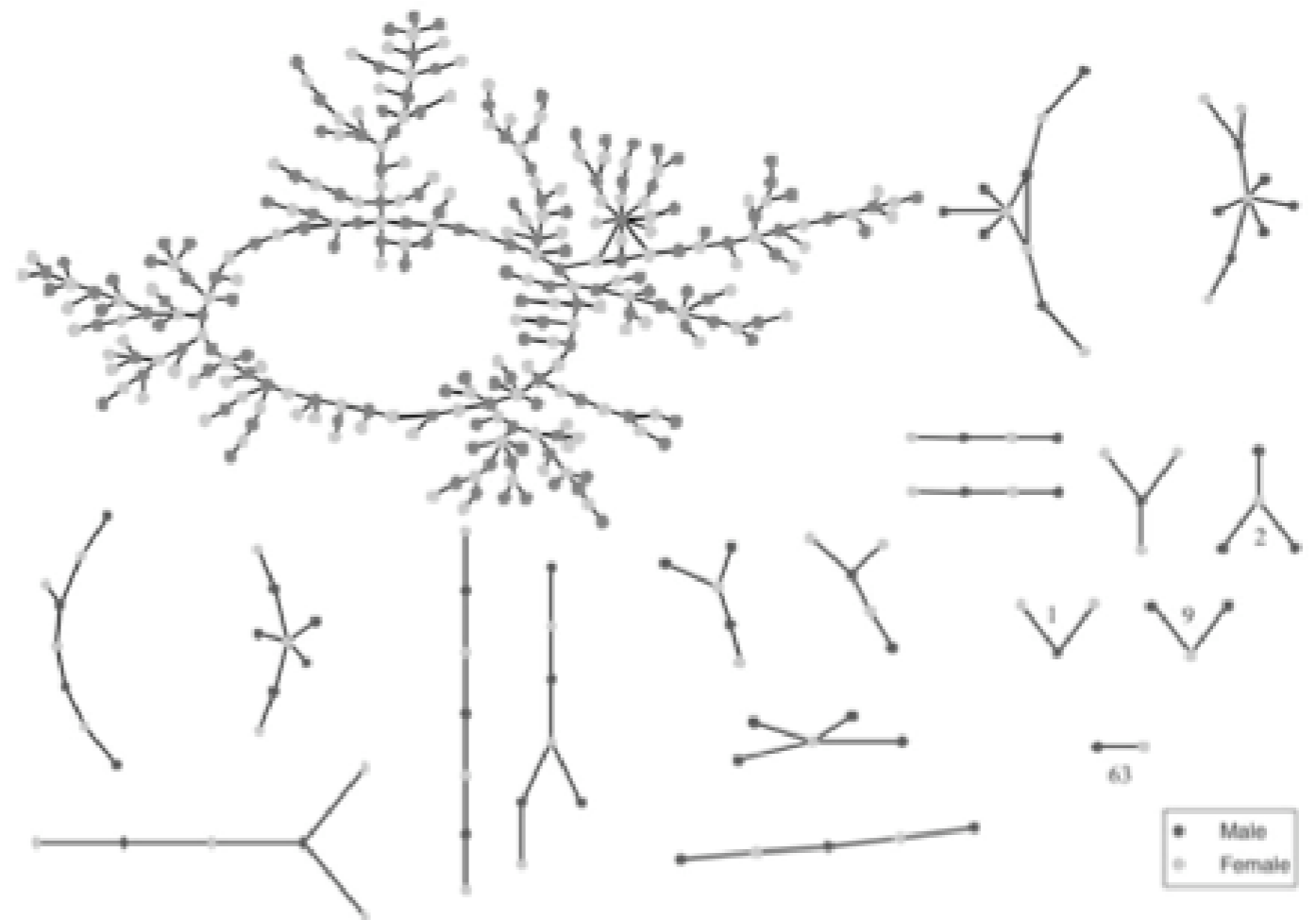


FIG. 2.—The direct relationship structure at Jefferson High

## In physics

Also a long history of graph representation: Feynman diagrams, Mayer expansion...

Only recently, physicists have focused on random networks, and in their realisation in complex systems (Erdos-Renyi).

**Internet**

**Transport networks**

**Power grids**

**Protein interaction networks**

**Food webs**

**Metabolic networks**

**Social networks**

Complex networks  
composed by a large number  
of nodes and links



**Statistical approach**

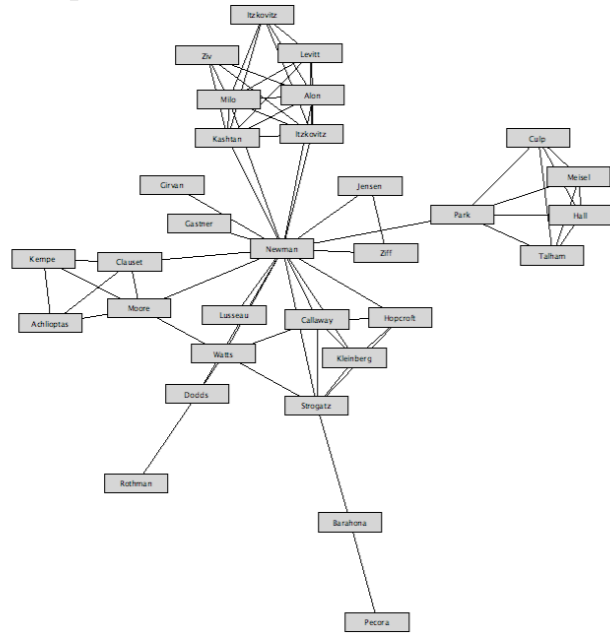
Many studies to understand the structuring of the network, as well as dynamical processes on the networks

# Standard properties

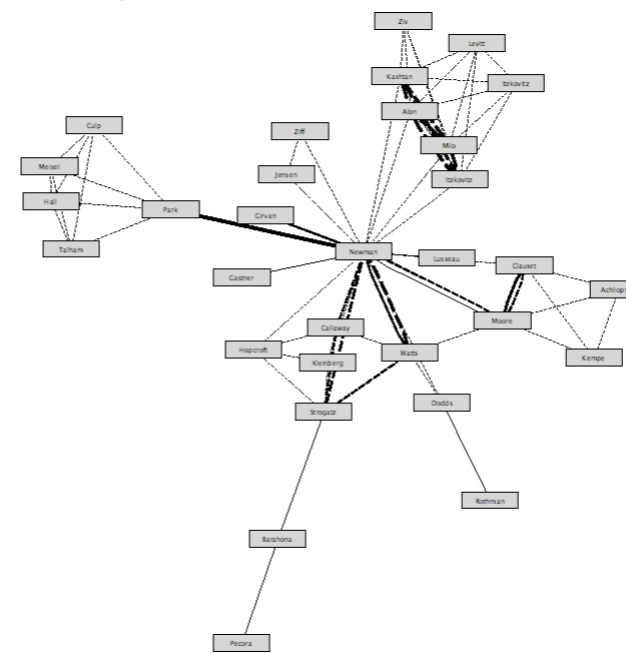
- Degree (number of links/node) distributions: power-law, exponential, stretched exponential...
- Clustering coefficient: density of triangles in the network  
=> measures how the friend of a friend is a friend
- Assortativity: degree correlations between adjacent links  
(Are large sites connected with large sites?)
- Attempts to describe long-range interactions/large scale structures

# More and more general structures:

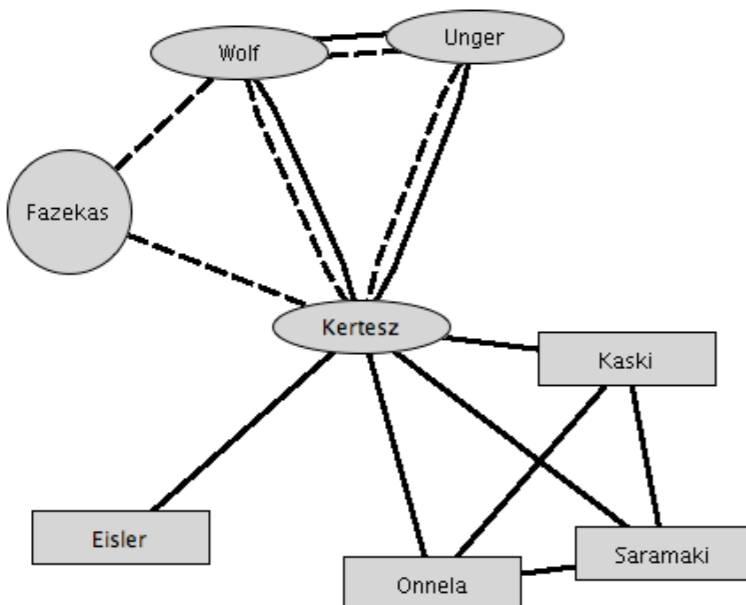
## Unipartite network



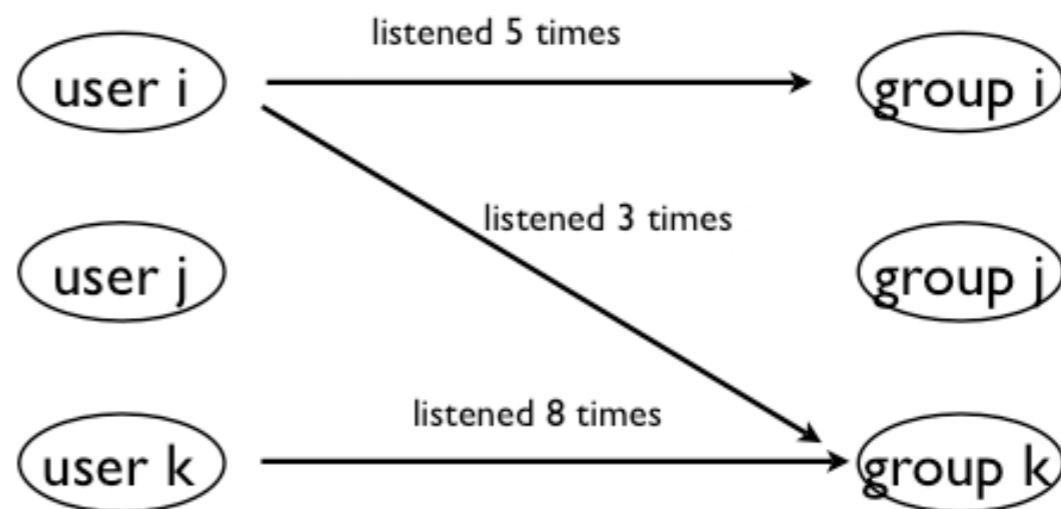
## Weighted network



## Coupled networks

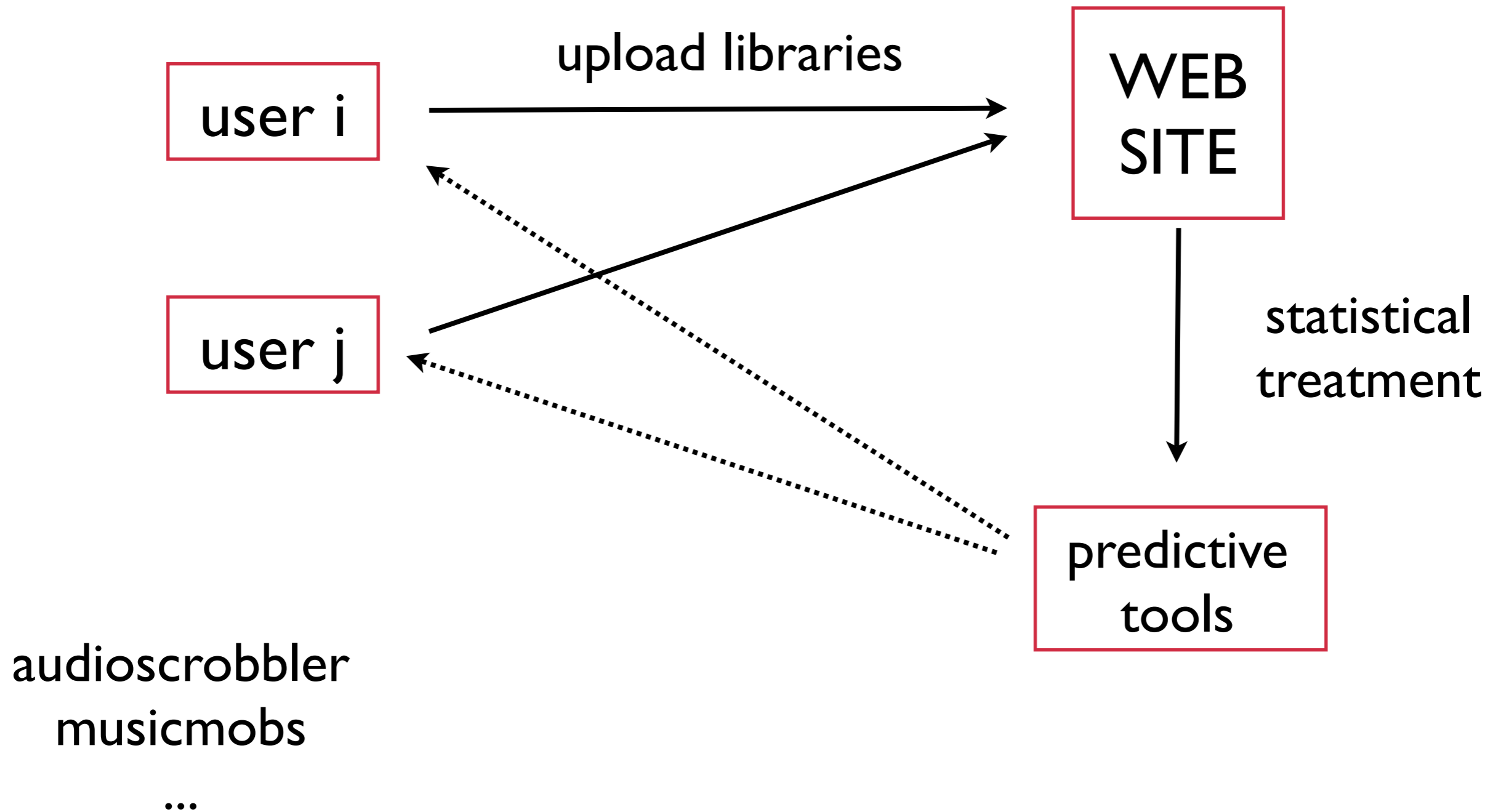


## Bipartite network



## 2) Methodology (Collaborative filtering)

Recently (1 year), new free services on the web:



# Data analysis

90	46	4
90	227	11
90	737	1
90	787	7
90	798	41
90	938	5
90	1208	39
90	1435	40
90	3008	17
90	3023	6
90	3197	1
90	3202	2
90	3625	1
90	4586	3
90	4699	1
90	5238	23
90	5360	3
90	1006894	2
90	1007236	2
90	1017629	1
90	1051293	1
90	1233975	1
90	6674638	1
90	6674664	2

#listener

#group

#playcount

2102323	Motorpsycho/Jaga Jazzist Horns
2102324	04 Cactus Jack
2102325	arcatera
2102326	12 air liquide
2102327	03 - Pulsedriver
6794080	Lenky and Sean Paul
2102328	Dr.Album
2102329	docProst
2102331	Nes Vs. Mr. Peculiar
2102332	D12, B-Real
2102334	John Hiatt and Loudon Wainwright
2102335	rage agenst the machine
2102336	Nakatomi - [EMG]
2102339	02 ATB
2102340	I Gult and Bla*tt
2102341	Ingmar Stenmark
2102345	0146 Jefferson Airplane feat. Grace Slick

#group

#data name

From audioscrobbler (January 2005), a data set with:

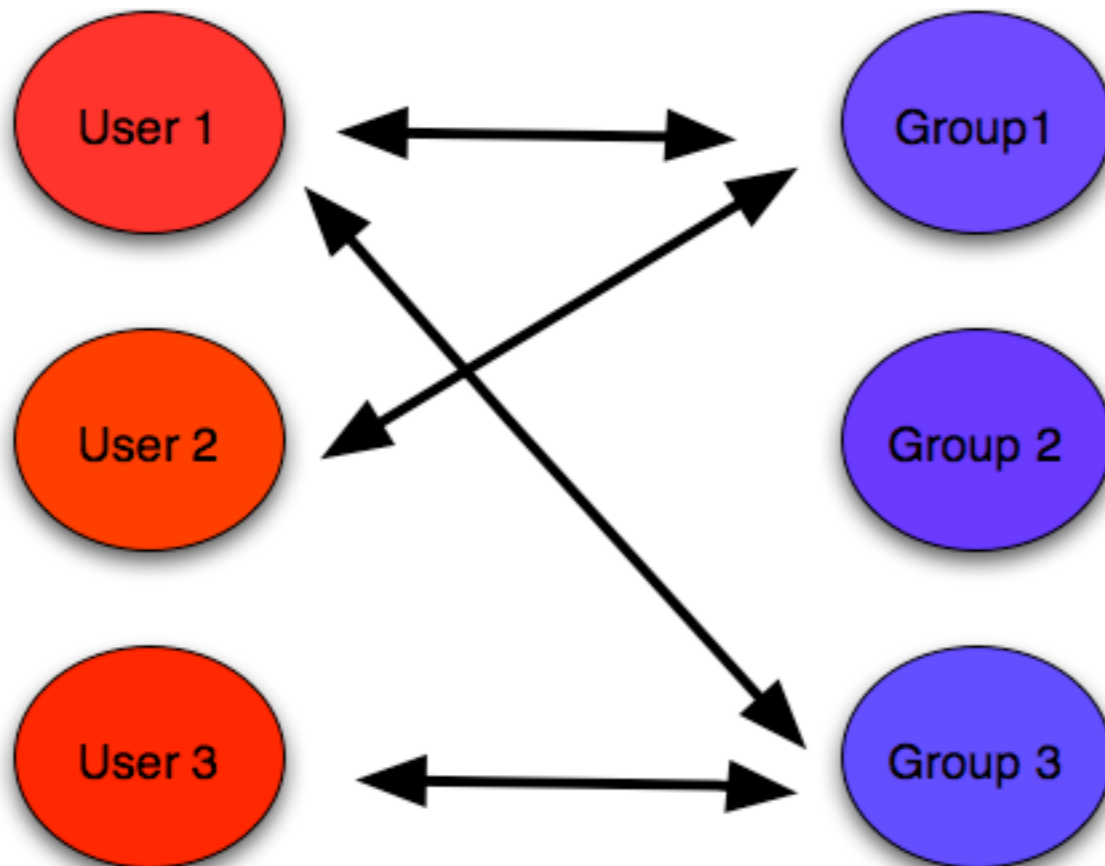
- 35916 users
- the music library of each user + the number of times they listen to each group

There is a total of 617900 *different* music groups:



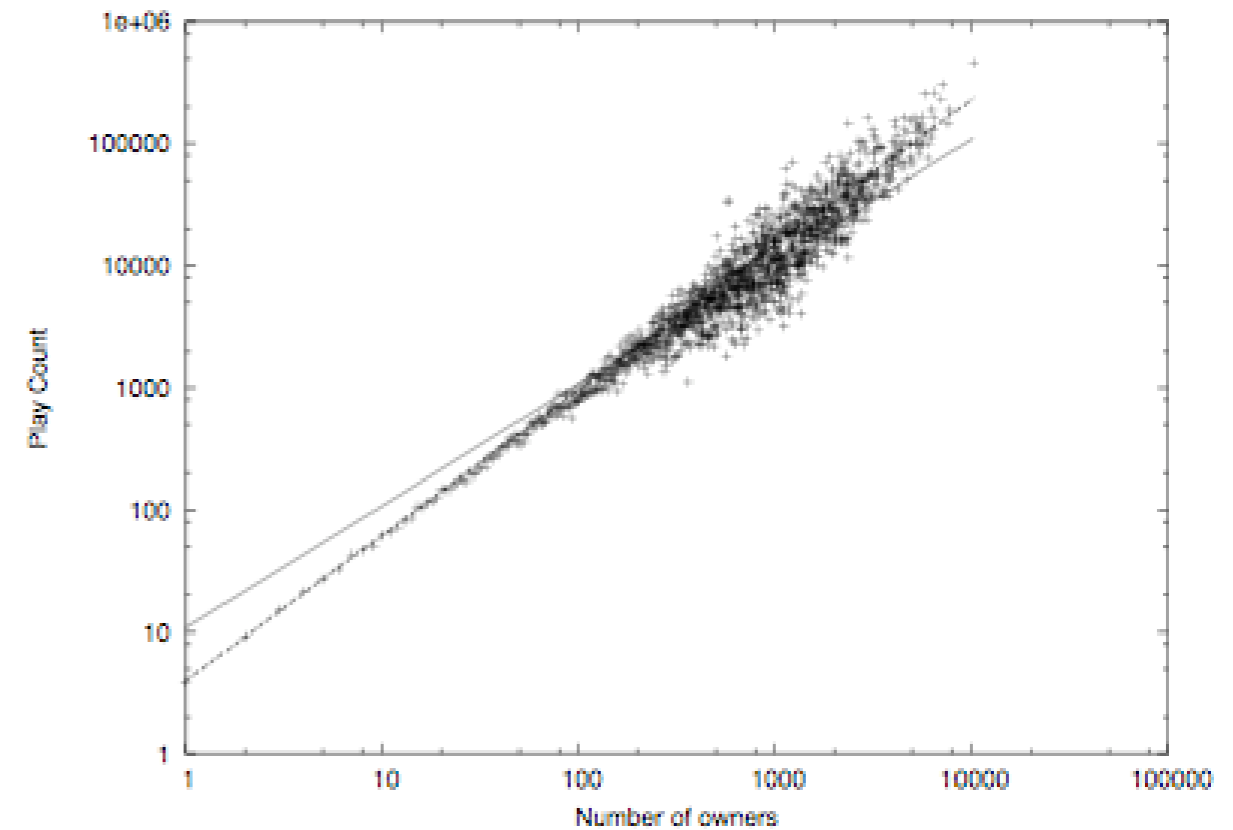
Mozart: 1468 users  
Wolfgang Amadeus Mozart: 539 users  
Amadeus Mozart: 17 users  
Mozart Wolfgang Amadeus: 7 users  
Wolfgang Amadeus Mozart: 8 users  
...

Natural bipartite structure




# Top ten groups

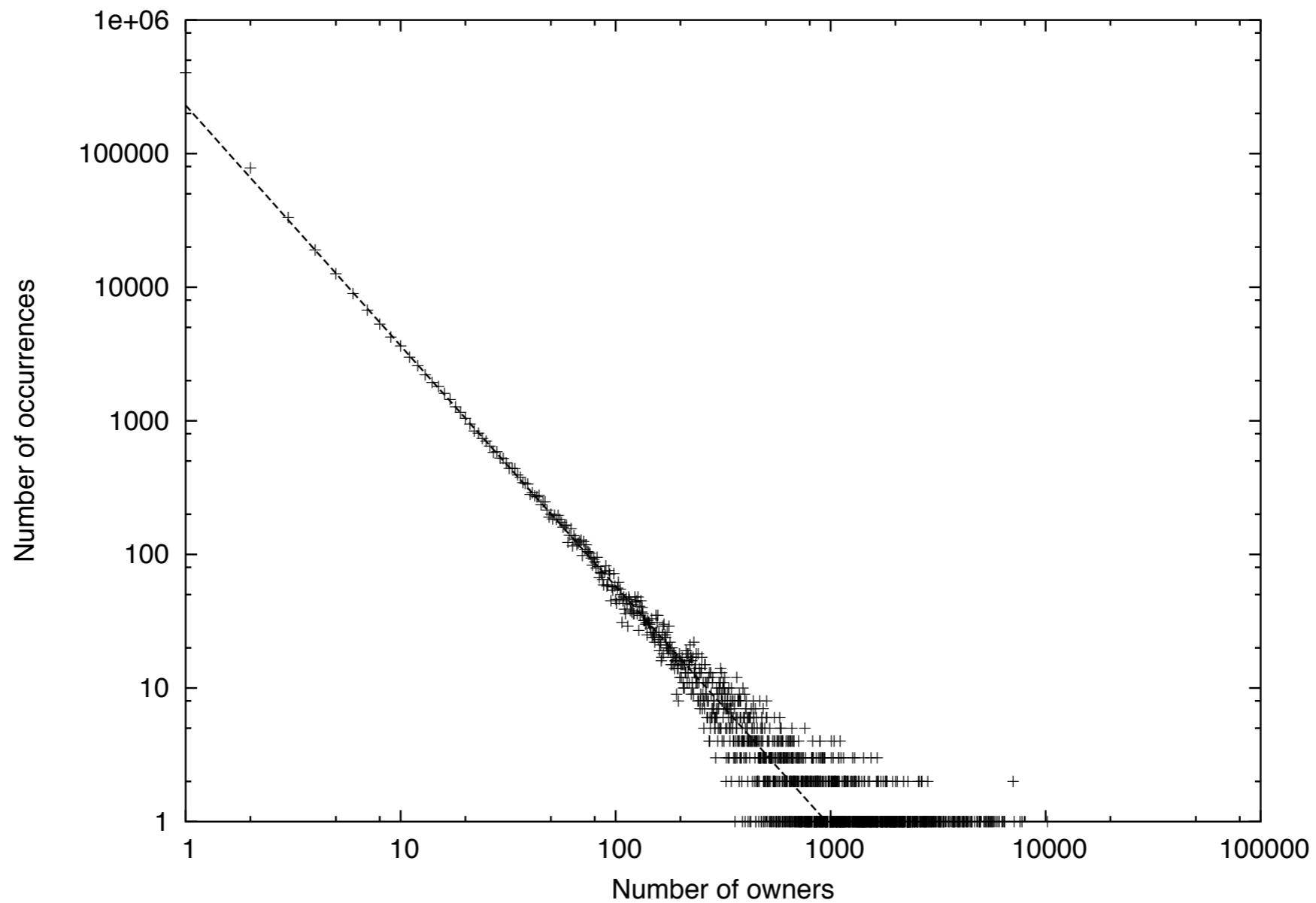
	Play Count		Audience	
1	Radiohead	462235	Radiohead	10194
2	The Beatles	309781	Nirvana	7757
3	Metallica	304203	Coldplay	7603
4	Pink Floyd	263674	Metallica	7135
5	Modest Mouse	255689	The Beatles	7057
6	Linkin Park	188624	Red Hot Chili Peppers	7057
7	Nirvana	188545	Pink Floyd	6456
8	Incubus	166480	Green Day	6432
9	Muse	162774	Weezer	6381
10	Tool	162612	Linkin Park	6290



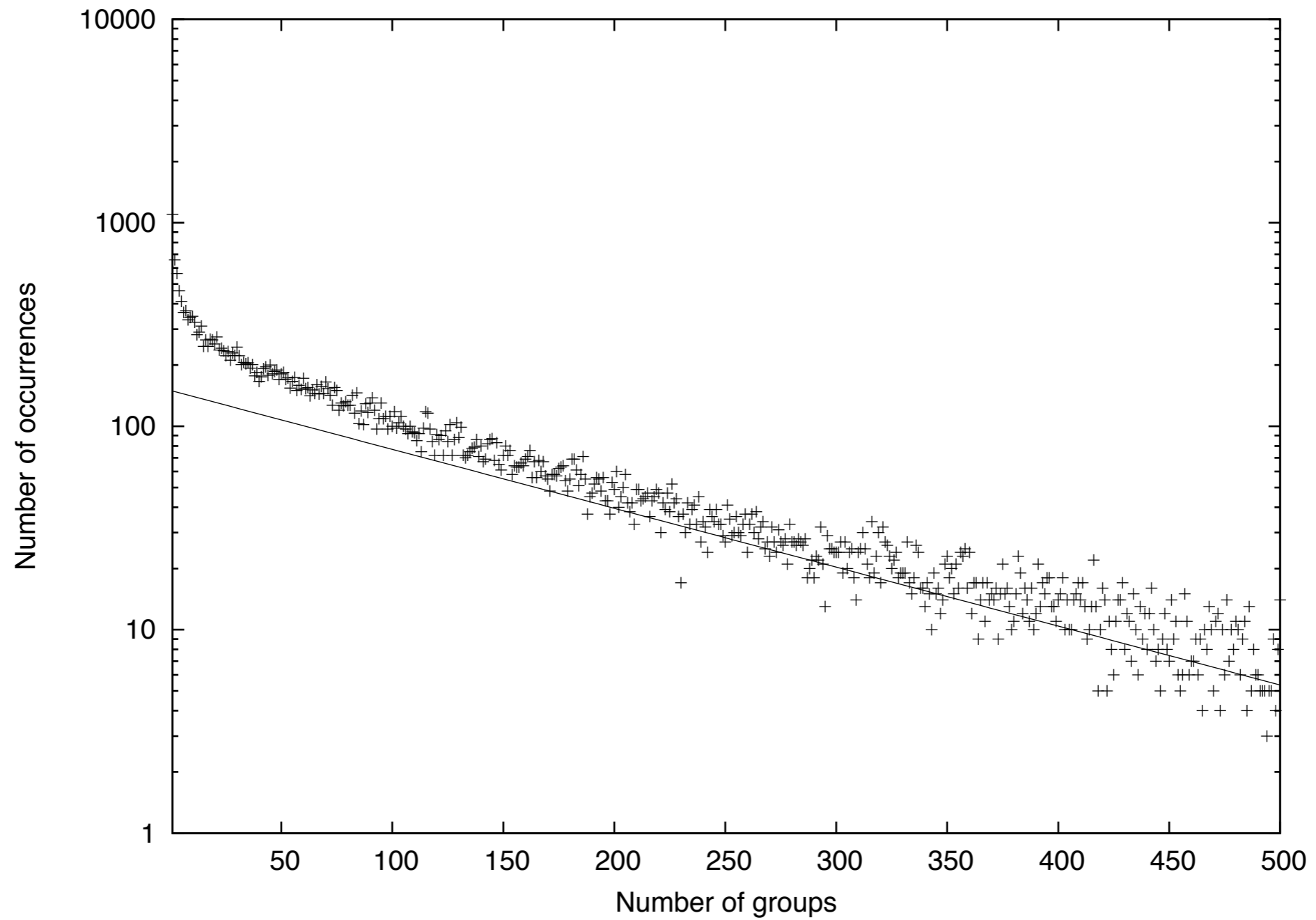
### 3) Analysis of the bipartite network

- 35916 listeners
- 617900 music groups
- In the bipartite graph, there are 5028580 links, and the total number of playcounts is 54386834
- On average, each user owns 140 music groups in his/her library
- Each group is owned by 8 persons.
- The listener with the most groups possesses 4072 groups (0.6% of the total music library)
- The group with the largest audience, Radiohead, has 10194 users (28% of the user community).

 Asymmetry in the bipartite network

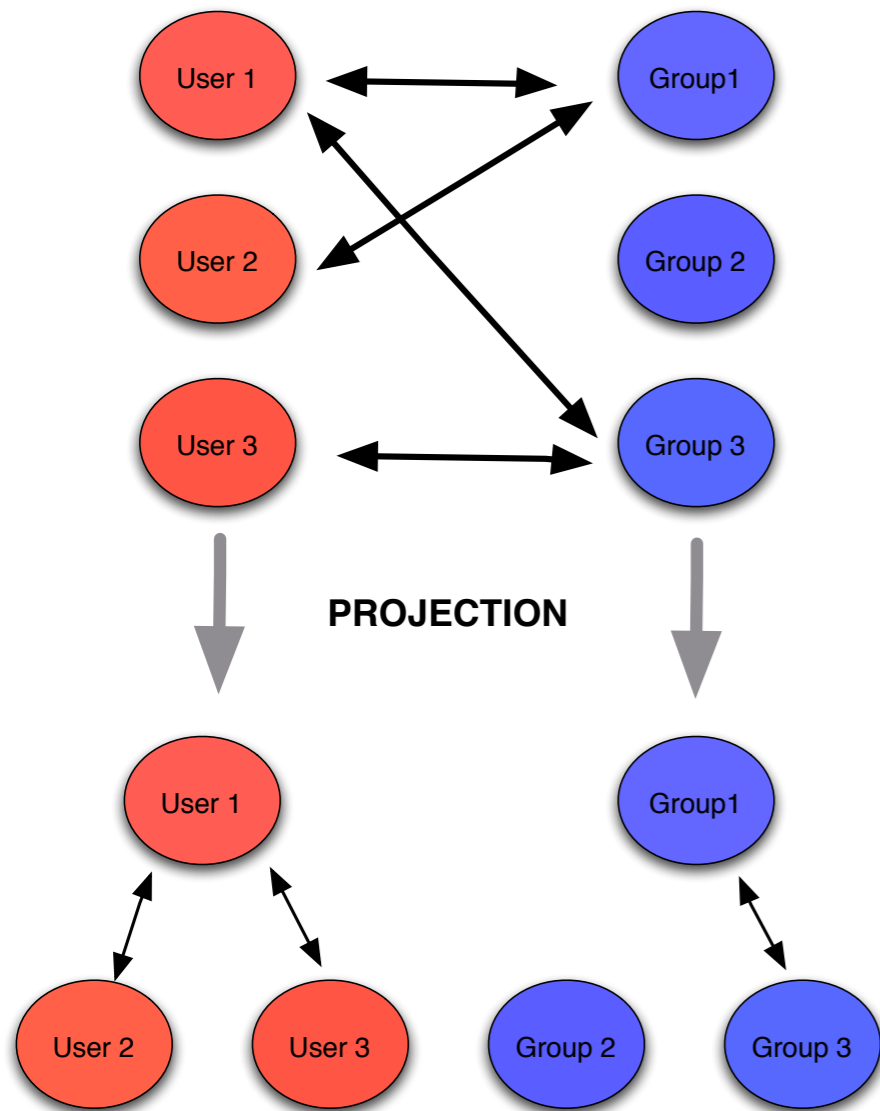


Distribution of the number of listeners per group. This distribution is shown to behave like the power-law  $\sim n^{-1.8}$ .



Distribution of the number of music groups per user, exponential tail.

## 4) Network projection



Structure of music trends, genres



Sociological structure of listeners

Newman, Watts, Strogatz, Physical Review E, 64, 026118 (2001).

But, the unipartite network is almost fully connected, i.e. most of the listeners are linked in the unipartite representation (mainstream music groups).

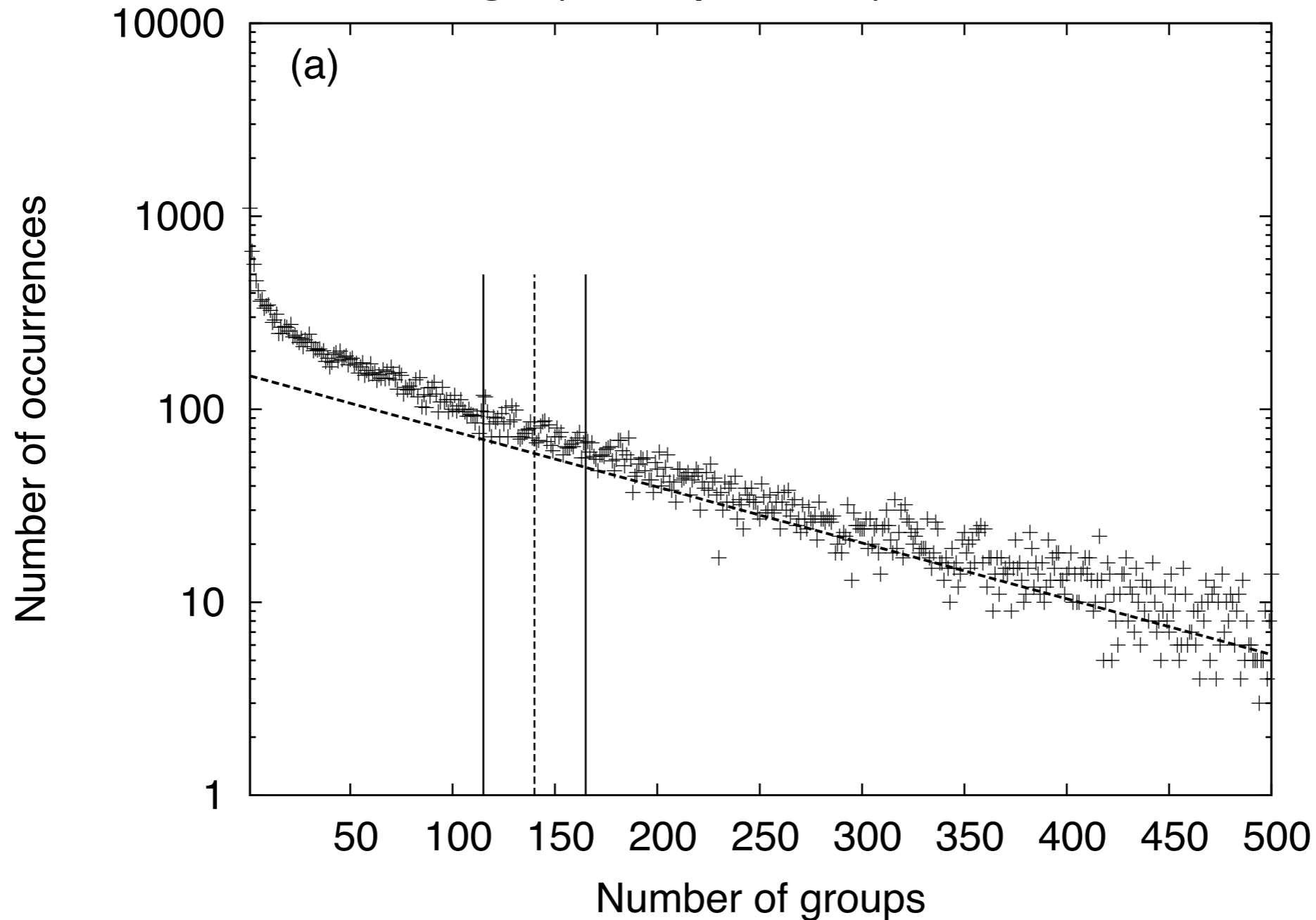
For instance, Radiohead fully connects 28% of the user community whatever the rest of their music library contents.



Oversimplified and useless representation.

# Listener network

In the following, we focus on a sample of listeners, around the average (3806 persons).



For each person, we define a 617900 vector, with 1 if he owns group  $i$ , and 0 if not.  $(1, 0, 0, \dots, 0, 1, \dots, 1)$



Individual Music Signature (IMS)

For each pair of group, we calculate the cosine between their 2 vectors:

$$C_{ij} = \frac{\mathbf{V}_i \cdot \mathbf{V}_j}{|\mathbf{v}_i| |\mathbf{v}_j|}$$

→ Symmetric measure of correlations, in  $[0, 1]$

→ Symmetric 3806X3806 matrix

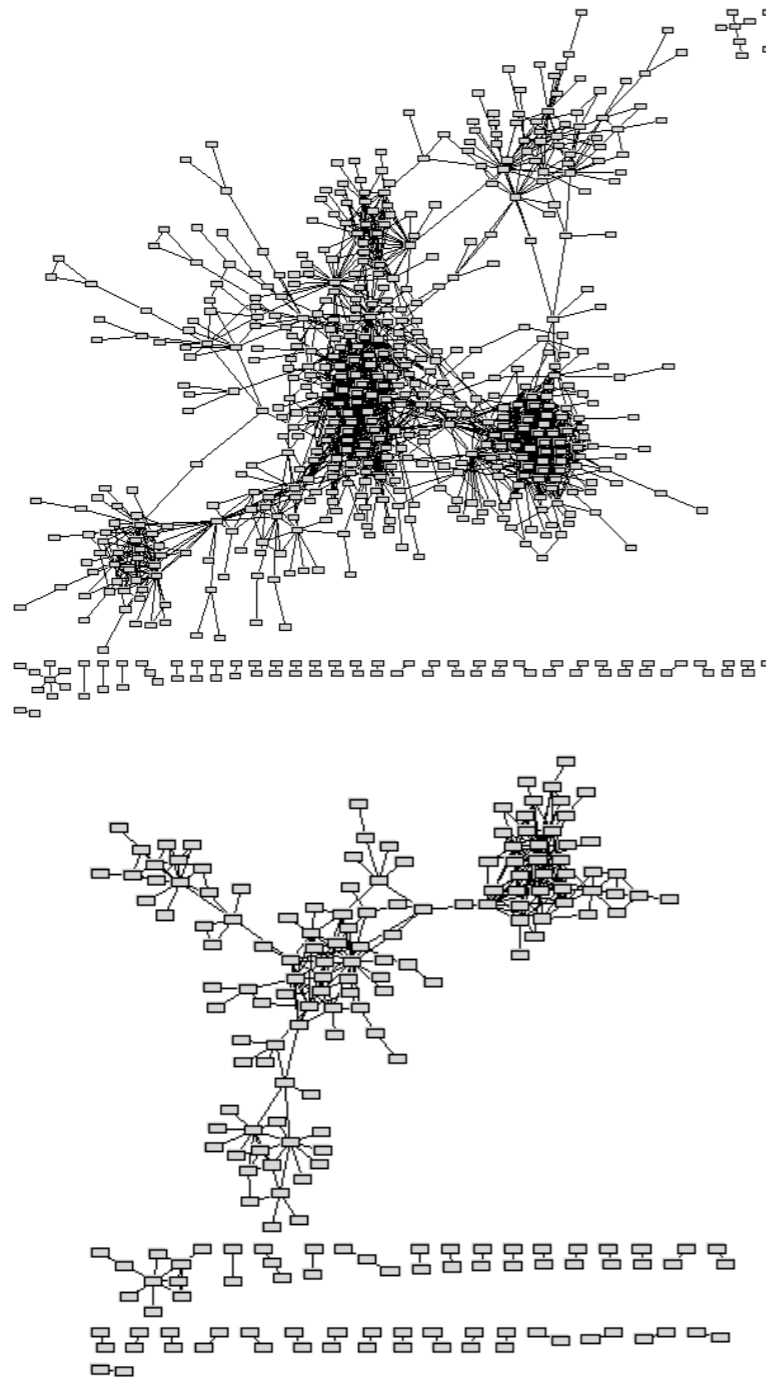
To reveal structures, we filter the matrix

If  $C_{ij} > h$  → link between i and j

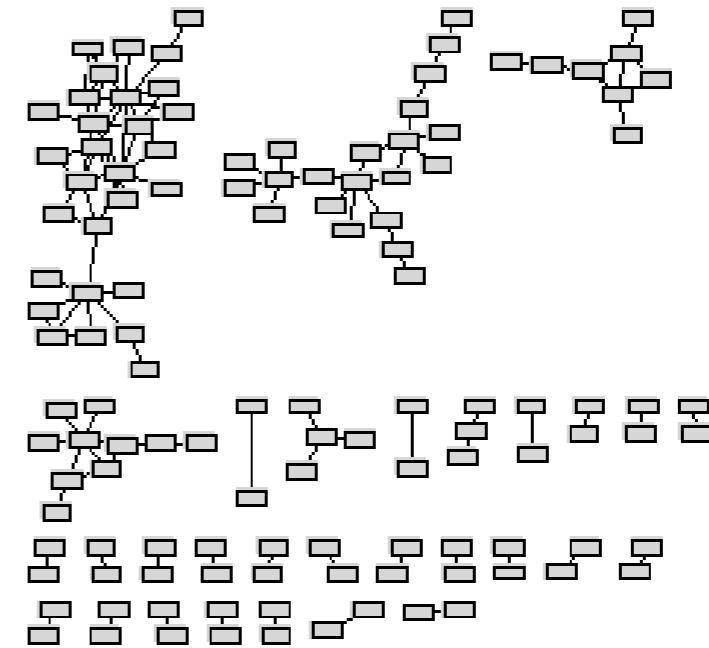
Else → no link between i and j

Starting at  $h=0$ : fully connected

- ➔ Increasing  $h$ , emergence of structures: peninsulae
- ➔ The peninsulae may eventually detach from the main island.

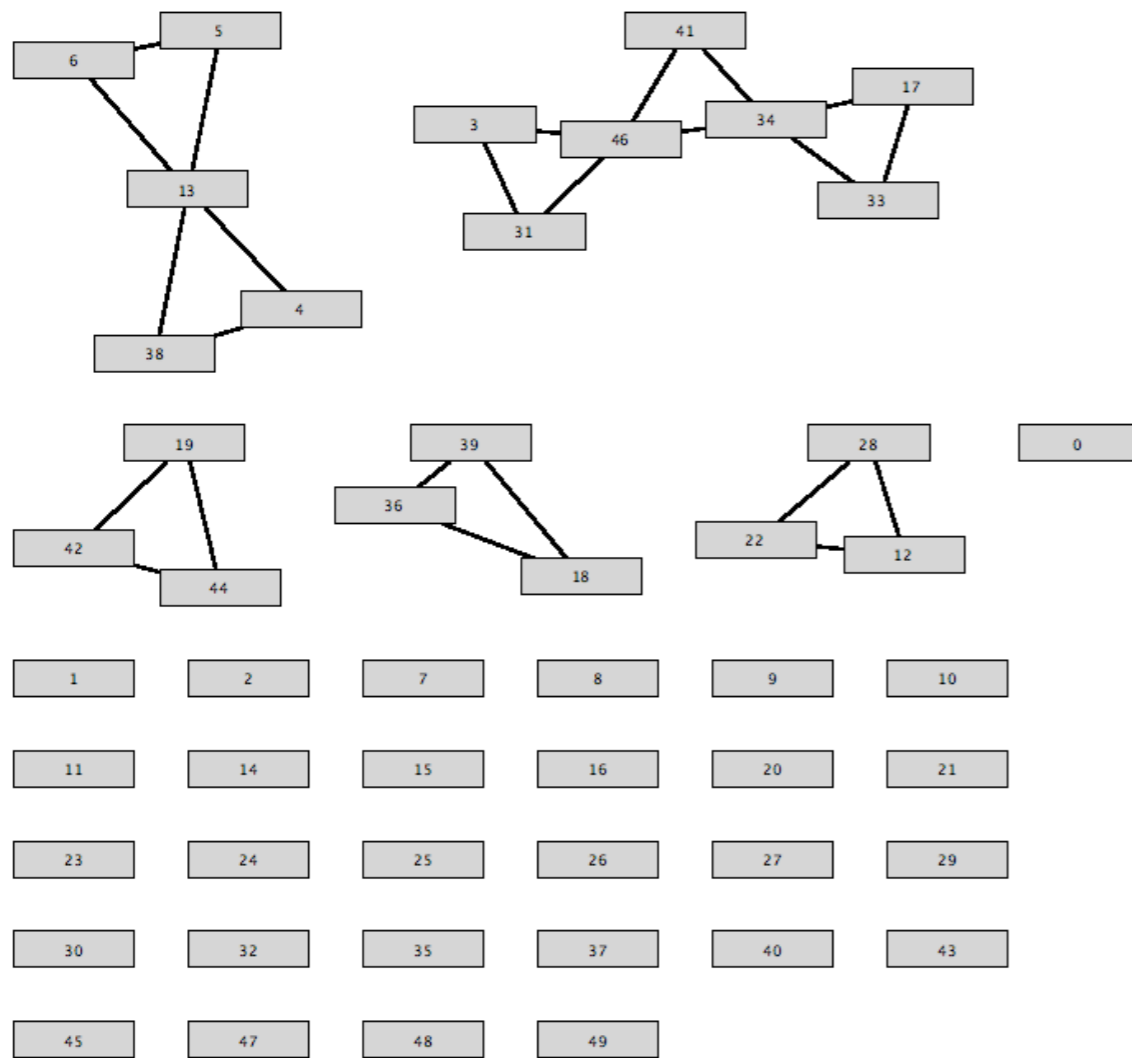


➔ We identify a community with a disconnected island



# Percolation idea based projection (PIB)

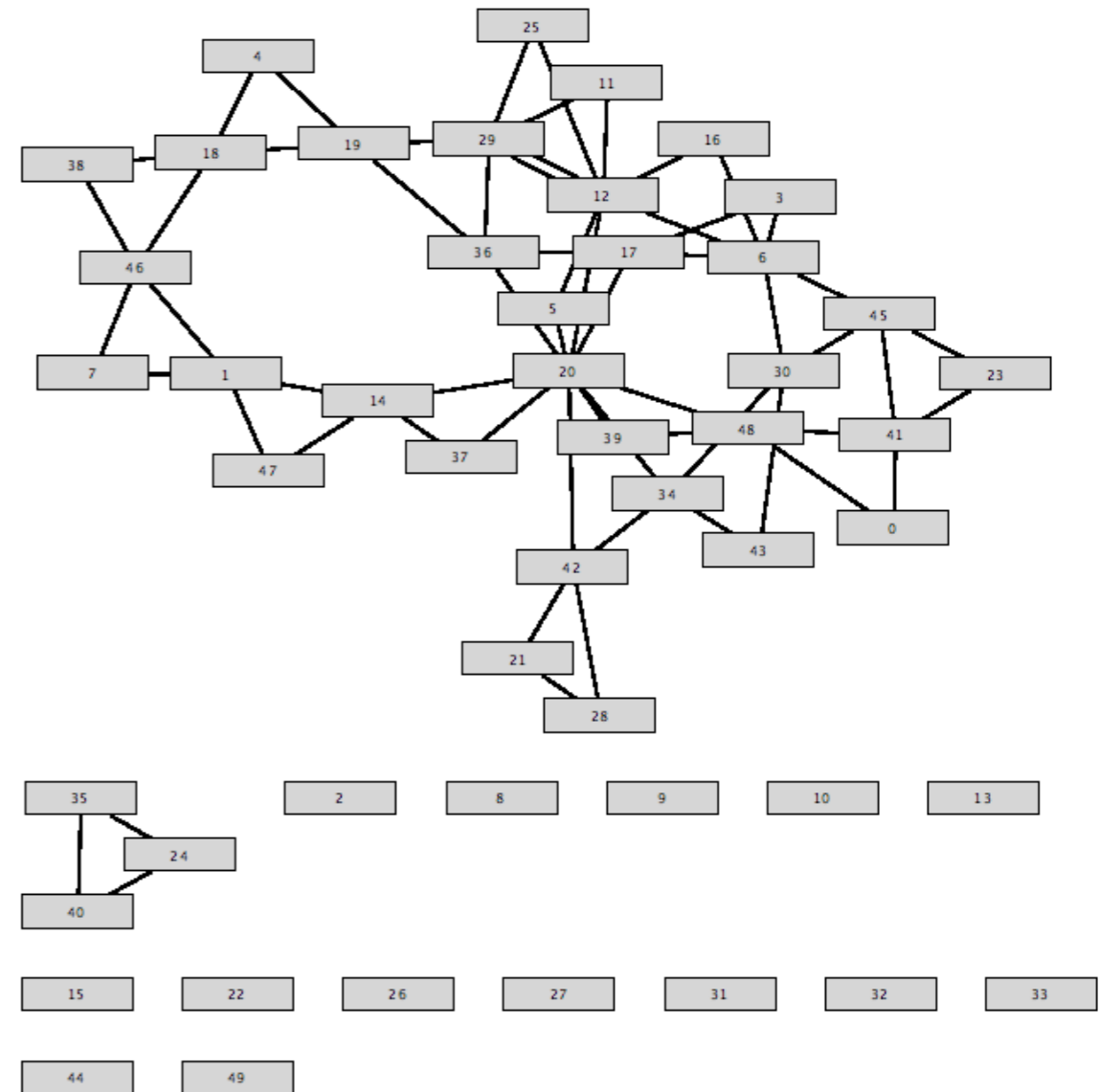
High filtering



Several small islands



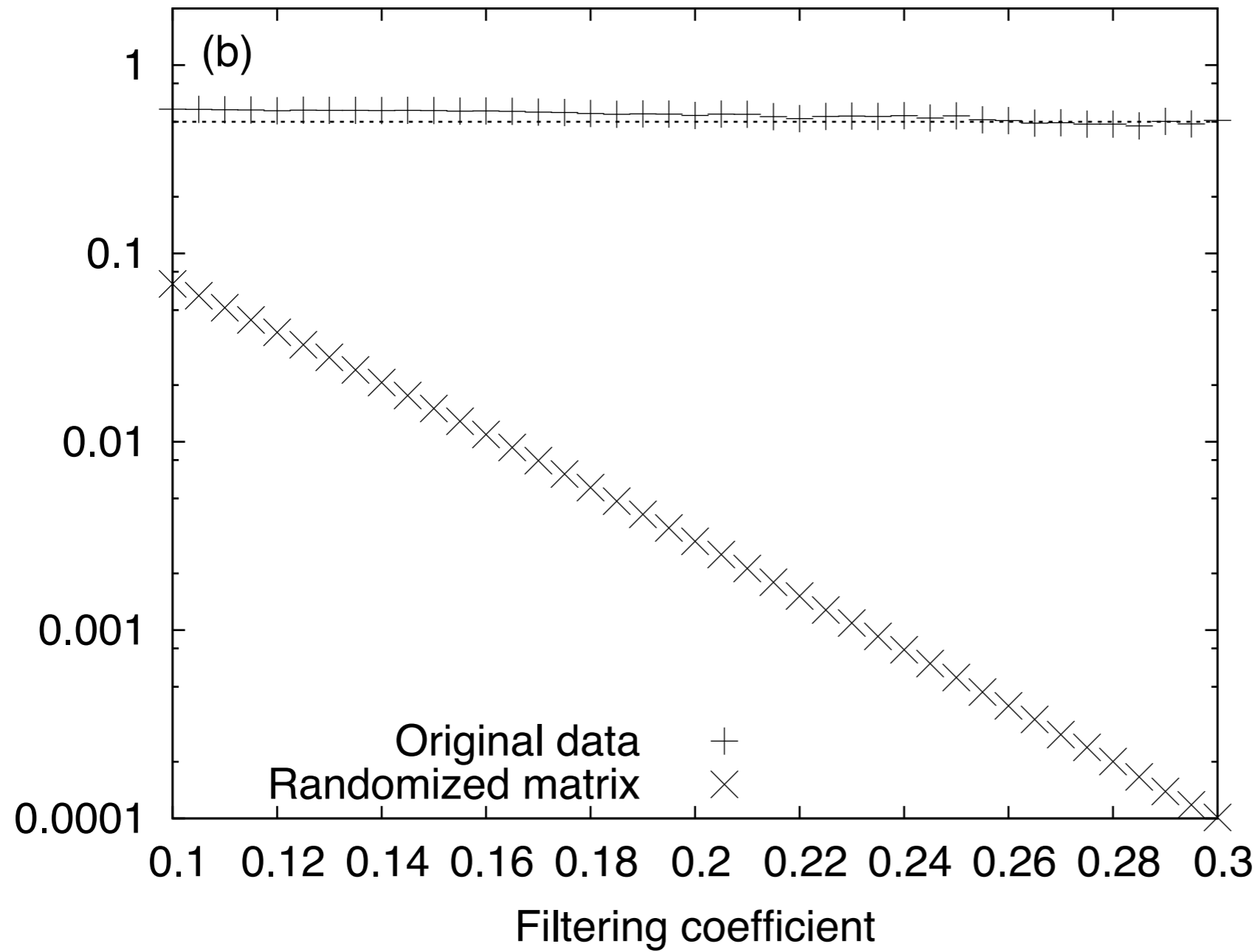
Small filtering



One large percolated island



# Dependence of the clustering coefficient



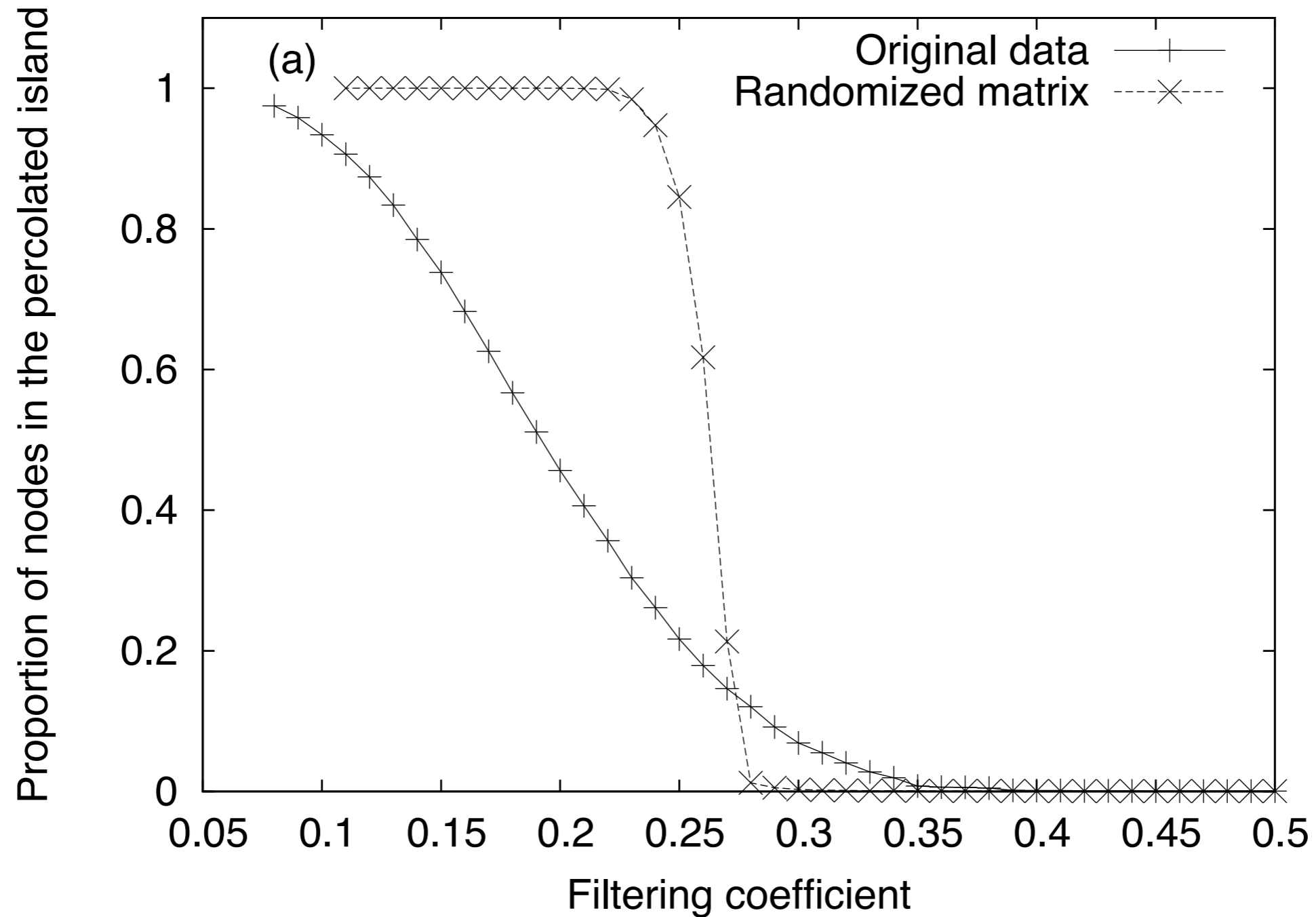
social mechanisms

high structuring of the network

# Percolation transition

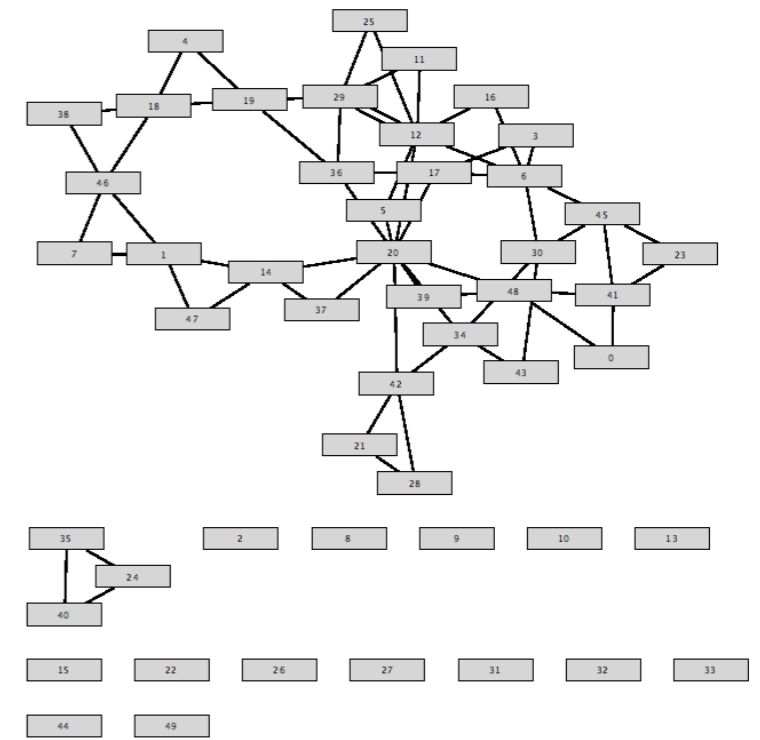
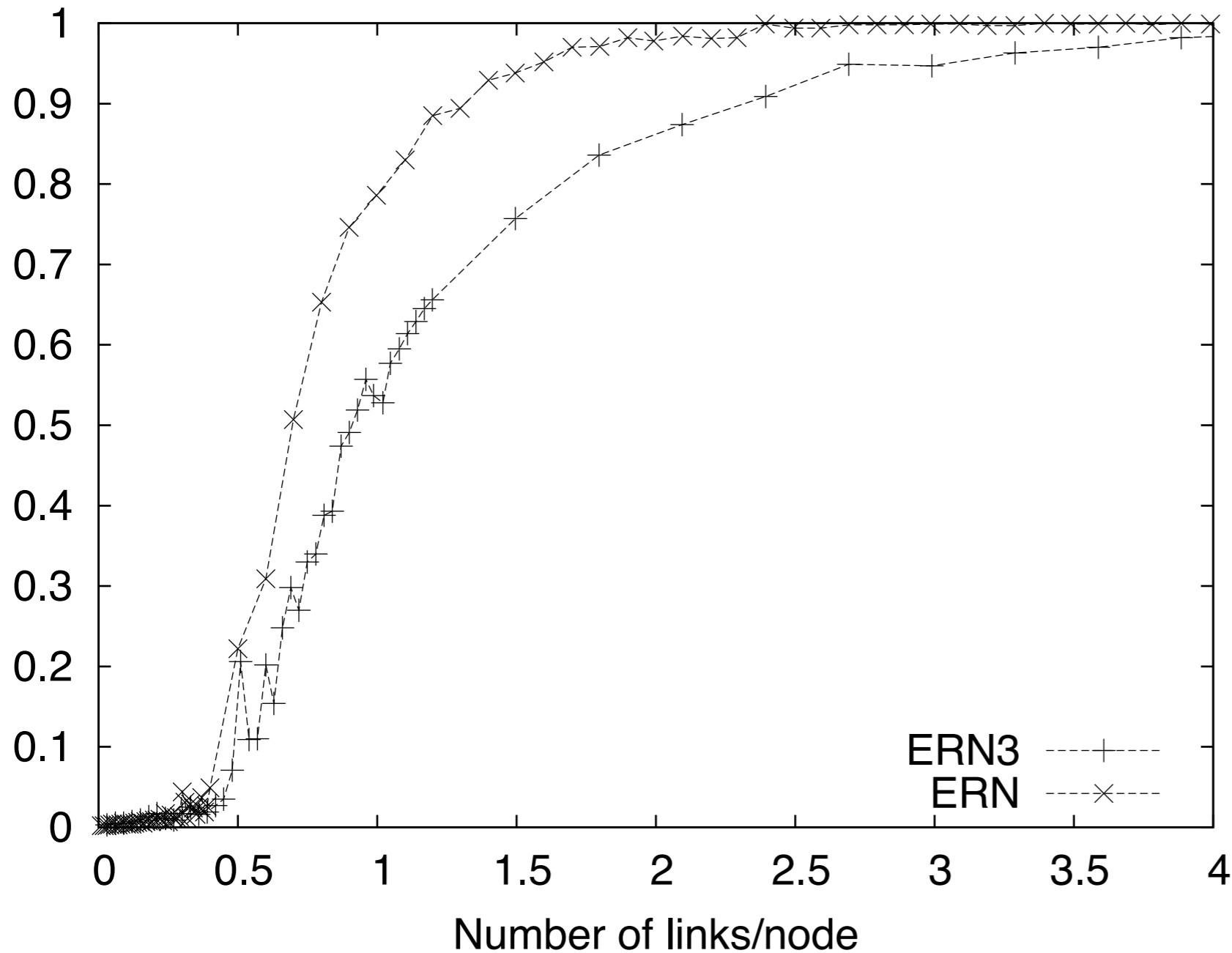
Compare with the randomised matrix (RM)

Structures broaden and displace the percolation transition

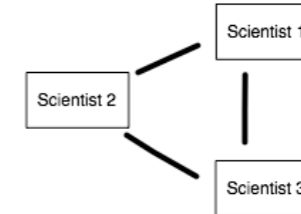


# Effects of correlations and structuring on percolation

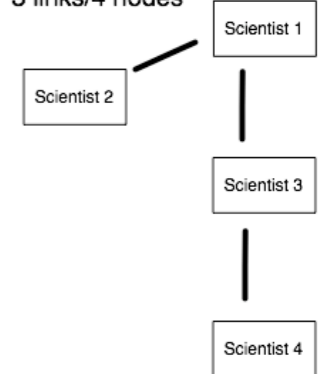
## Triangular Erdos-Renyi



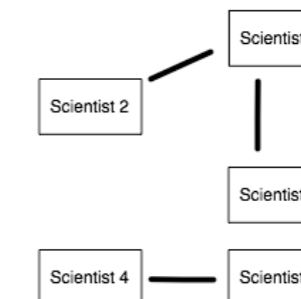
3 links/3 nodes



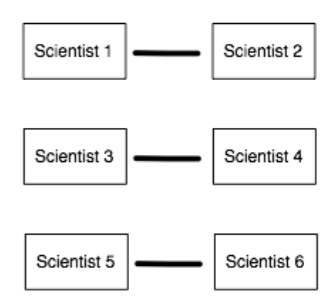
3 links/4 nodes



3 links/5 nodes



3 links/6 nodes

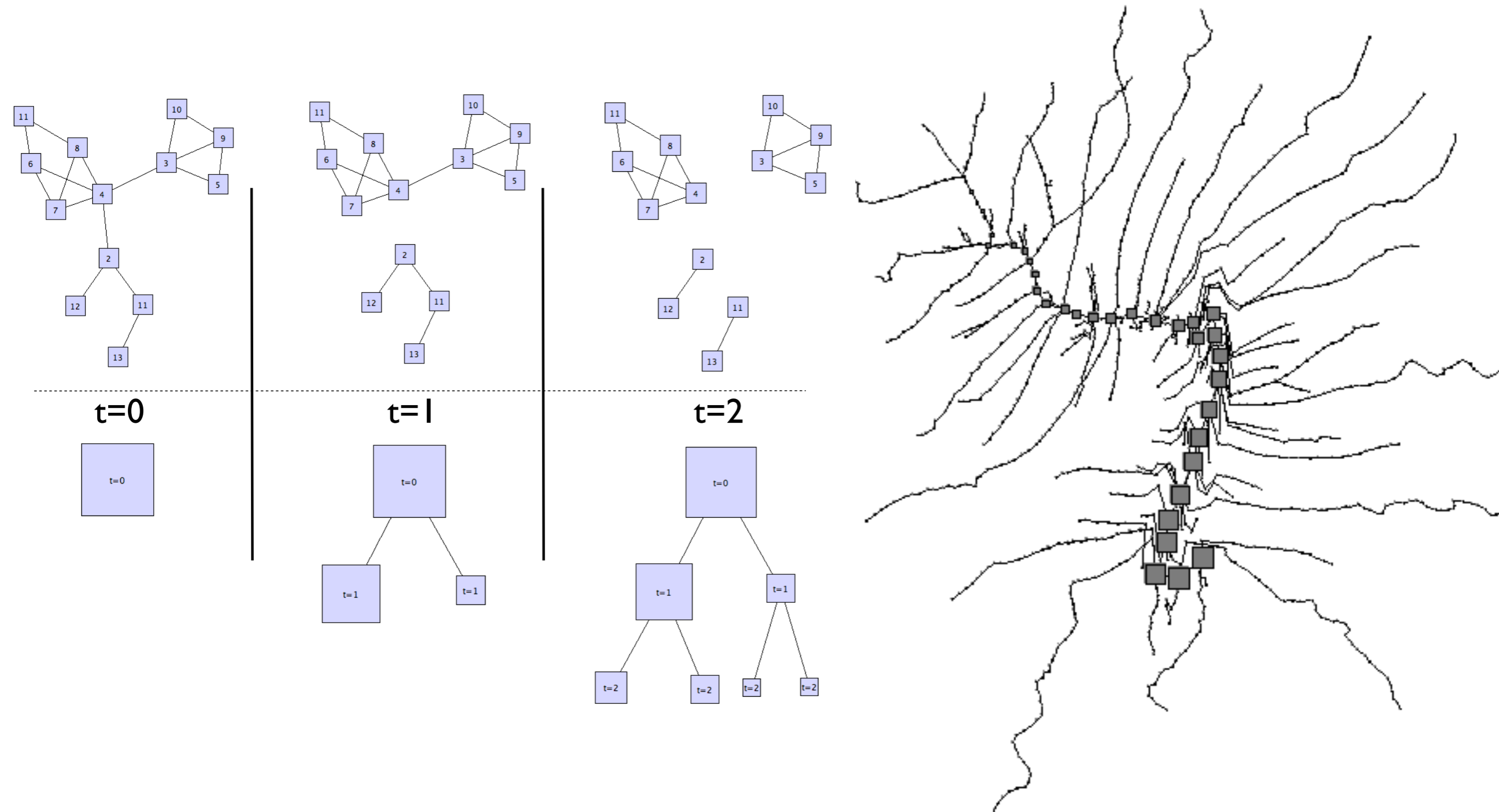


# Tree representation

Start at a low filtering constant value  $h$ .

Represent disconnected islands by squares (surface  $\sim$  nodes)

Increase  $h$ . Draw a link between parent nodes.



# Reality of usual music subdivisions?

sis+ PJ Harvey+ **Manic Street Preachers**+ *Roots Manuva*+  
*Unkle*+ **Linkin Park**+ *Atari Teenage Riot*+ **Kula Shaker**+ *The  
Police*+ *James Iha*+ **Semisonic**+ **Weezer**+ *Anastacia*+ **Rob  
Dougan**+ *Eels*+ *Fatboy Slim*+ **Green Day**+ **Lostprophets**+  
**System of a Down**+ *U.N.K.L.E.*+ *El-P*+ **Bee Gees**+ **Du-  
ran Duran**+ **Therapy?**+ *The Prodigy*+ **Foo Fighters**+ **JJ72**+  
*Alkaline Trio*+ **The Beatles**+ **Incubus**+ *Prodigy*+ **Muse**+  
*And You Will Know Us By The Trai*+ **Jimmy Eat World**+  
**Ash**+ **Rival Schools**+ *Cher*+ *At The Drive-In*+ *Johnny Cash*+  
*Mansun*+ *Queens of the Stone Age*+ **Basement Jaxx**+ *Dave  
Matthews Band*+ *Dj Tiesto*+ *Cast*+ **The Strokes**+ *Anthrax*+  
*Ian Brown*+ **Saves The Day**+ *Morrissey*+ *Police*+ *Modest  
Mouse*+ *Interpol*+ *St Germain*+ **The Beach Boys**+ *Bonnie  
Tyler*+ *Theme*+ **Fenix\*TX**+ **Snow Patrol**+ *The Cooper Tem-  
ple Clause*+ *Buddy Holly*+ **Nada Surf**+ *onlinedrawing*+ *Michael  
Kamen*+ *Remy Zero*+ *Ernie Cline*+ **Quicksand**+ *Olivia New-  
ton John*+ **Polar**+ *Ikara Colt*+ *Keiichi Suzuki*+ *Rivers Cuomo*+  
*Paddy Casey*+ **Billy Talent**+ *Mireille Mathieu*+ *Jack Dee*+ *To-  
moyasu Hotei*+ *Daniel O'Donnell*+ *Hope Of The States*+ **Franz  
Ferdinand**+ *The Shadows*+ **THE STILLS**+ **The RZA**+ *The  
Mamas and the Papas*+ *Melissa Auf Der Maur*+ *Barron Knights*+  
**The Killers**+ **R.E.M.**+ **Jay-Z** **DJ Danger Mouse**+ *Pras  
Michel Feat ODB and Maya*+ *The Monks Of Roscrea*

What kind of music do you listen to?

Mixture of usual music genres:  
Rock/Electro/Hip-Hop/Pop...

Inadequacy of such music subdivisions  
to characterise individual and  
collective listening habits.

# Music group network

Alternative way to classify music, different from the usual genre-fications, based on the correlations between their audience

For each group, we define a 35916 vector, with 1 if the the user  $i$  owns it, and 0 if not.

Signature of the group = his audience

We focus on the top 5000 groups



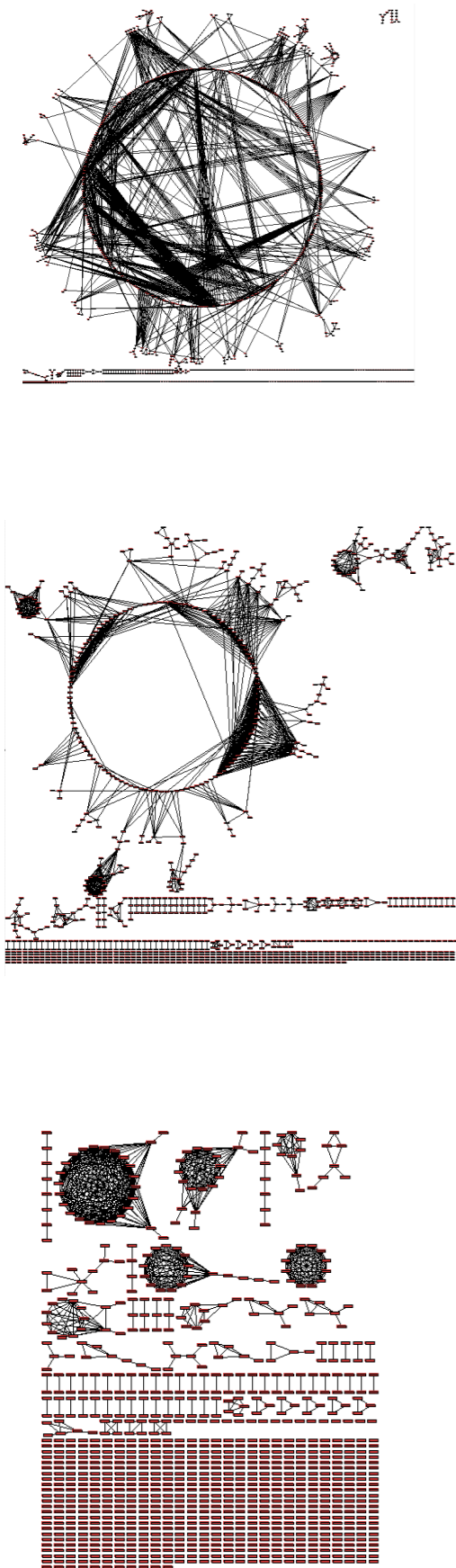
$(1, 0, 0, \dots, 0, 1, \dots, 1)$

$$C_{ij} = \frac{\mathbf{v}_i \cdot \mathbf{v}_j}{|\mathbf{v}_i| |\mathbf{v}_j|}$$

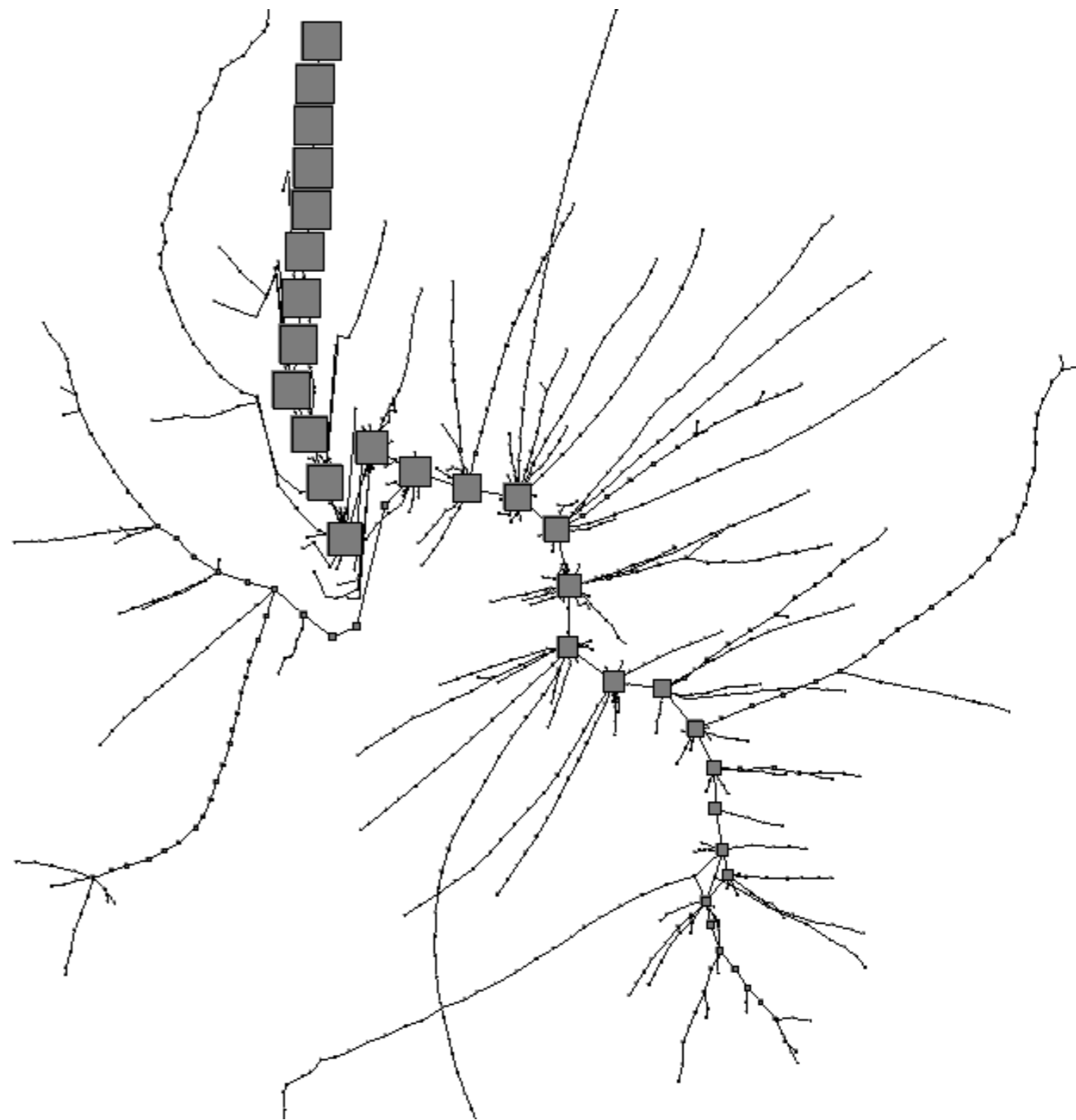


We construct the graph by filtering the matrix.

# Percolation



# Tree representation



# Following one branch

[Talvin Singh, The Cinematic Orchestra] 2

[Talvin Singh, The Cinematic Orchestra] 2

[Plaid, The Cinematic Orchestra, Talvin Singh] 3

[Plaid, The Dining Rooms, Talvin Singh, The Cinematic Orchestra, Nightmares on Wax] 5

[Plaid, The Dining Rooms, Talvin Singh, The Cinematic Orchestra, Savath and Savalas, Nightmares on Wax] 6

[Plaid, Nightmares on Wax, Savath and Savalas, The Dining Rooms, Talvin Singh, The Cinematic Orchestra, Kid Loco, Chick Corea, Telefon Tel Aviv, Herbert, Gimmik] 11

[Plaid, Nightmares on Wax, Savath and Savalas, The Dining Rooms, Talvin Singh, The Cinematic Orchestra, Kid Loco, Jaga Jazzist, Chick Corea, Telefon Tel Aviv, Herbert, Gimmik] 12

[Plaid, Nightmares on Wax, Savath and Savalas, The Dining Rooms, Talvin Singh, The Cinematic Orchestra, Kid Loco, Jaga Jazzist, Chick Corea, Telefon Tel Aviv, Herbert, Gimmik] 12

[Plaid, Nightmares on Wax, Savath and Savalas, The Dining Rooms, Talvin Singh, The Cinematic Orchestra, Kid Loco, Jaga Jazzist, Chick Corea, Telefon Tel Aviv, Herbert, Gimmik] 12

[T.A.T.U, Train, t.A.T.u., Avenue Q, die trying, Stephen Lynch, Lost in Translation, Elvis Costello and The Attractio, Three 6 Mafia, Razorlight, The Thermals, Fischerspooner, Simon and Garfunkel, Beethoven, Snog, Elefant, The Libertines, Stiff Little Fingers, Calla, Jesus and Mary Chain, All-Time Quarterback, Scarface, The Coup, Hanzel und Gretyl, Lou Reed, Inkubus Sukkubus, April March, Paris, goblin, Velvet Underground, Canibus, Blackalicious, Luna, Indigo Girls,

## Branches correspond to well-defined genres, but also reveal unexpected groupings

dard, homogenous style groupings. Amongst these homogeneous cliques, there are *[George Strait, Faith Hill, Garth Brooks, Clint Black, Kenny Chesney, Shania Twain, Alan Jackson, Martina McBride, Alabama, Tim McGraw, Reba McEntire, Diamond Rio, John Michael Montgomery, SheDaisy, Brooks and Dunn, Clay Walker, Rascal Flatts, Lonestar, Brad Paisley, Keith Urban]*, *[Kylie Minogue, Dannii Minogue, Sophie Ellis Bextor]*, *[Serge Gainsbourg, Noir Désir]*, *[Billie Holiday, Glenn Miller, Benny Goodman]*, *[Morrissey, Faith No More, Machine Head, The Smiths, Rammstein, Smashing Pumpkins, Slipknot, Tomahawk, Mr. Bungle]*, that are country, dance pop, geographically localised i.e. France, swing jazz and rock groupings respectively.

for example. But other groupings defy monolithic style categorization, like: *[The Jon Spencer Blues Explosion, Yello, Galaxie 500, Prince and the Revolution, Ultra Bra, Uriah Heep, Laurent Garnier]*, *[Crosby, Stills, Nash and Young, Orb, Zero 7, Royksopp, Stan Getz]*. The latter include unexpected mixtures of Indie Rock/Funk/Hard Rock/Dance, and Folk/Electro/Jazz respectively.

*[Isaac Hayes, Bernard Herrmann, The RZA, Al Hirt, Charlie Feathers, zamfir, Luis Bacalov, Neu!, Tomoyasu Hotei, Santa Esmeralda, Nancy Sinatra, Meiko Kaji, The 5.6.7.8's, Ennio Morricone, Shivaree, Malcolm McLaren]*

Thanks to G. D'Arcangelo

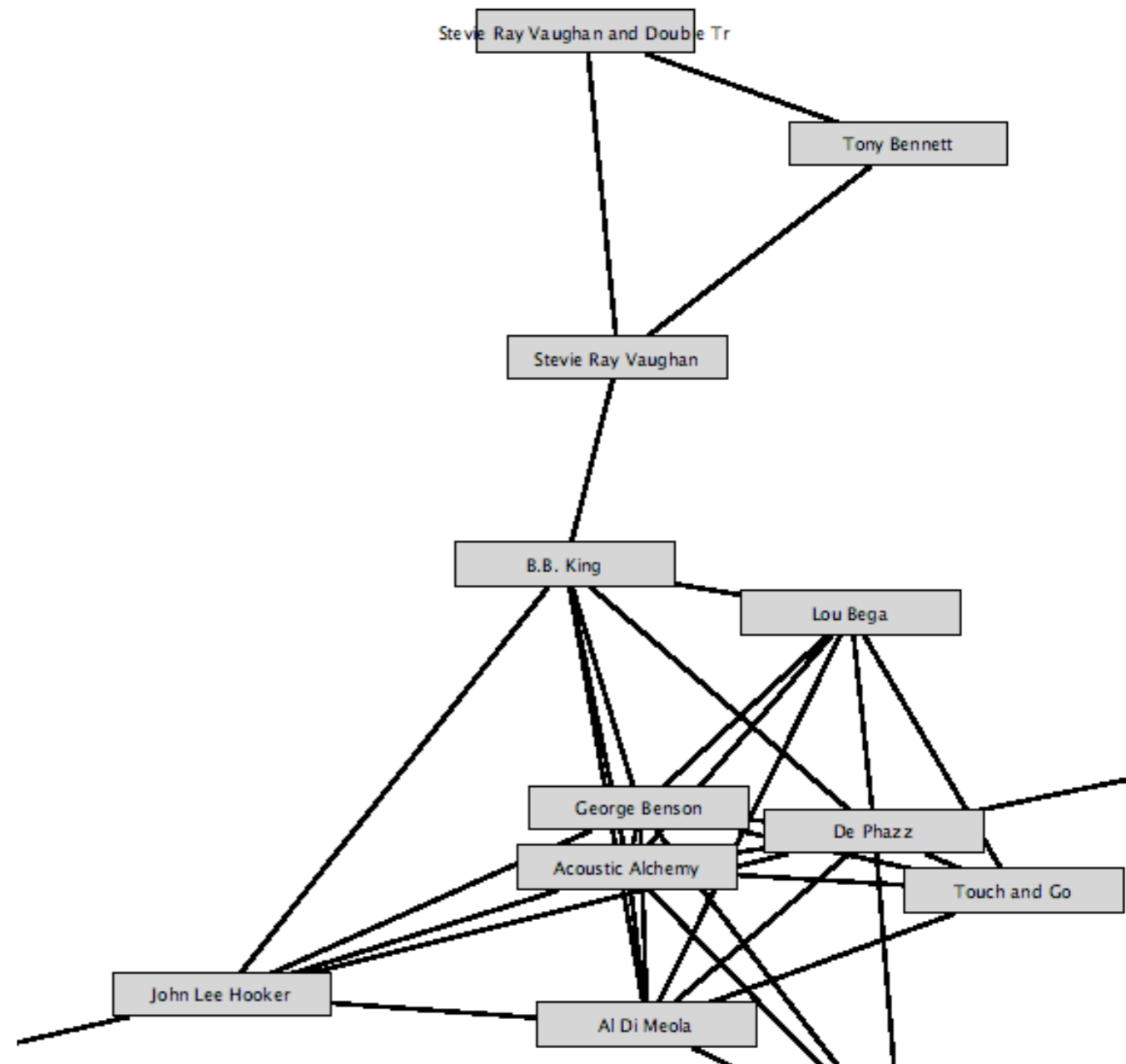
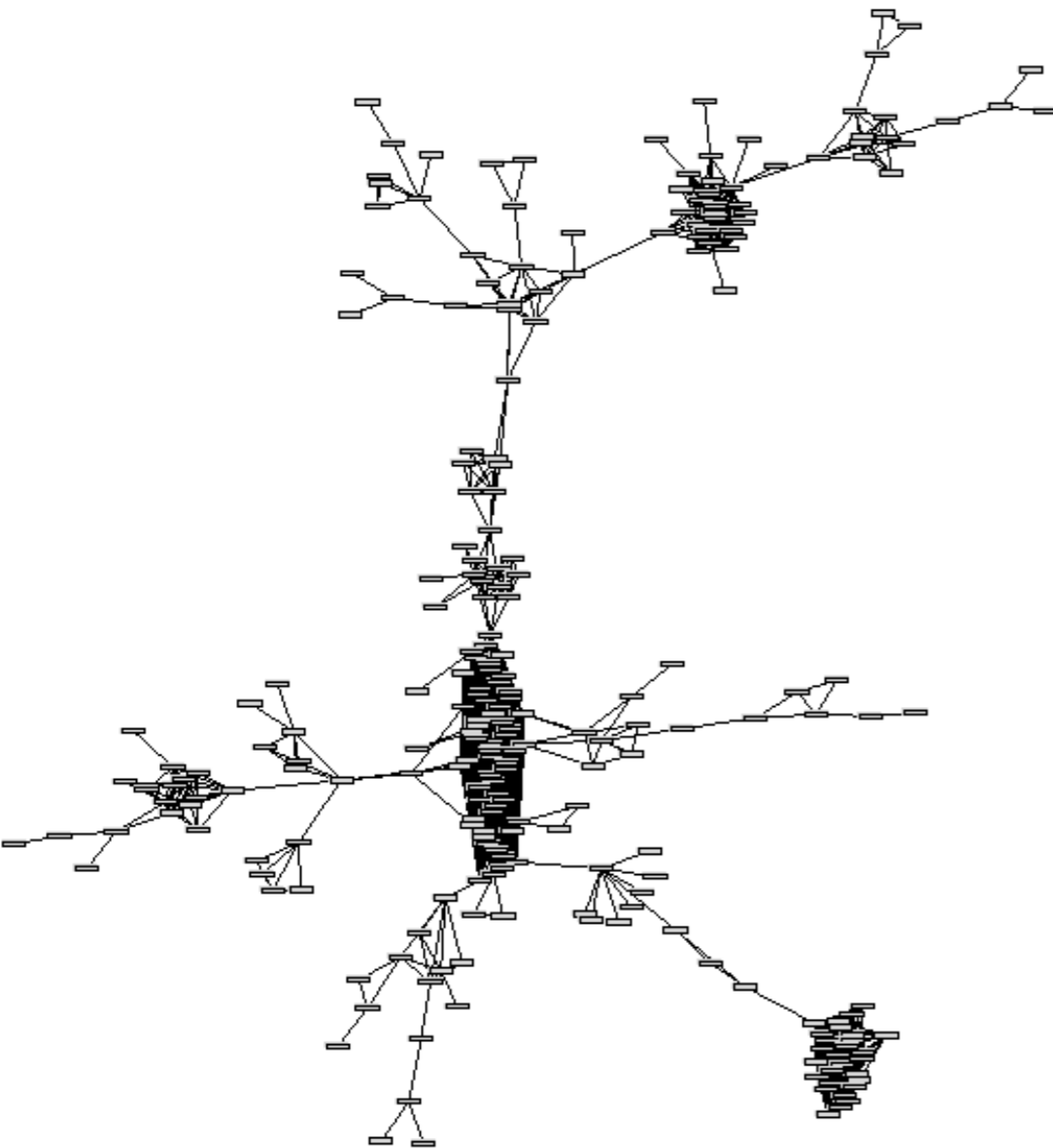
## 5) Random walk exploration

PIB: identifying "music genres"/"listener communities" with isolated islands.

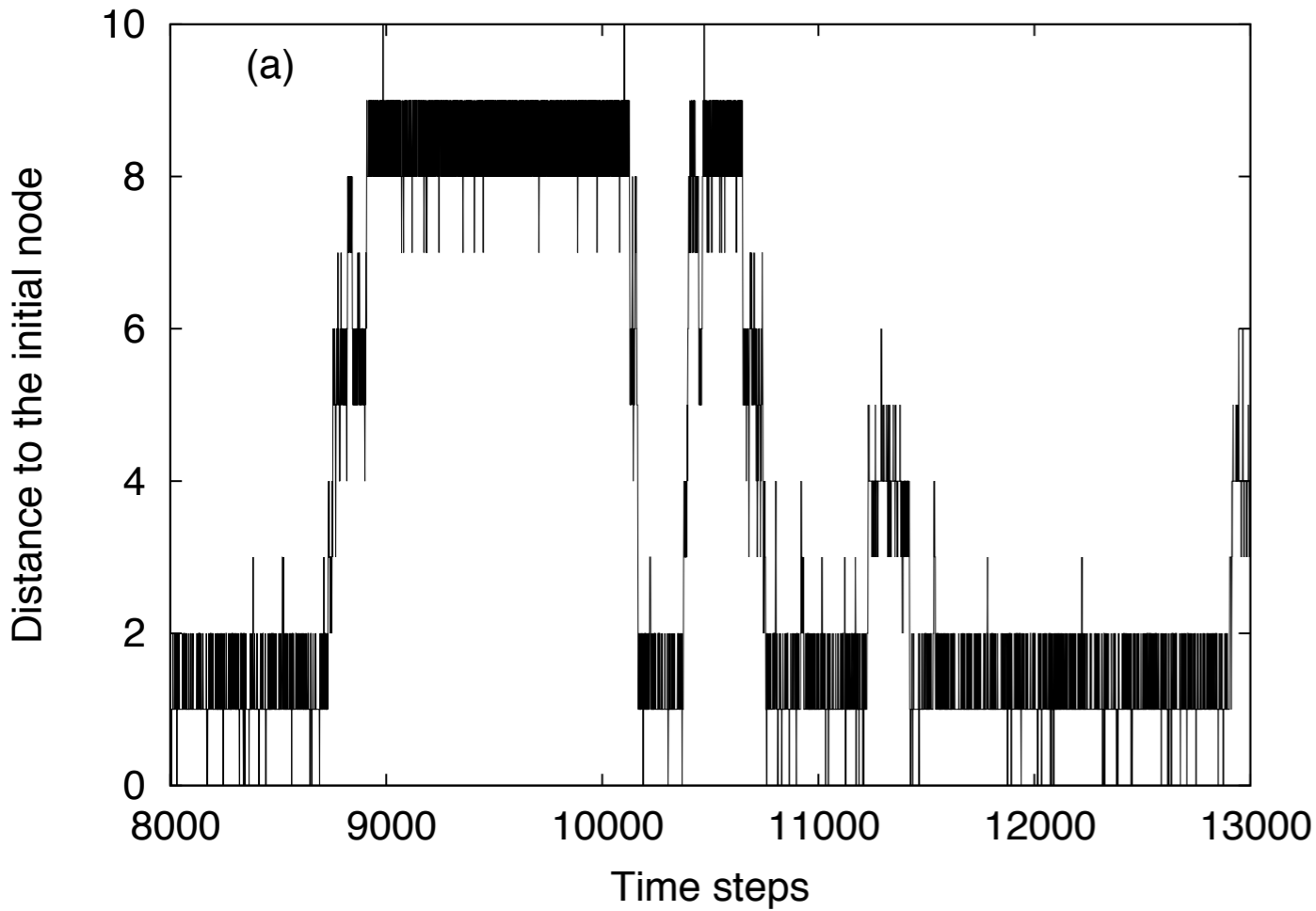
It is obviously a drastic simplification that may lead to the neglect of pertinent structures, and therefore requests a more detailed exploration of the network structure.

➔ Hierarchical characterisation, RW

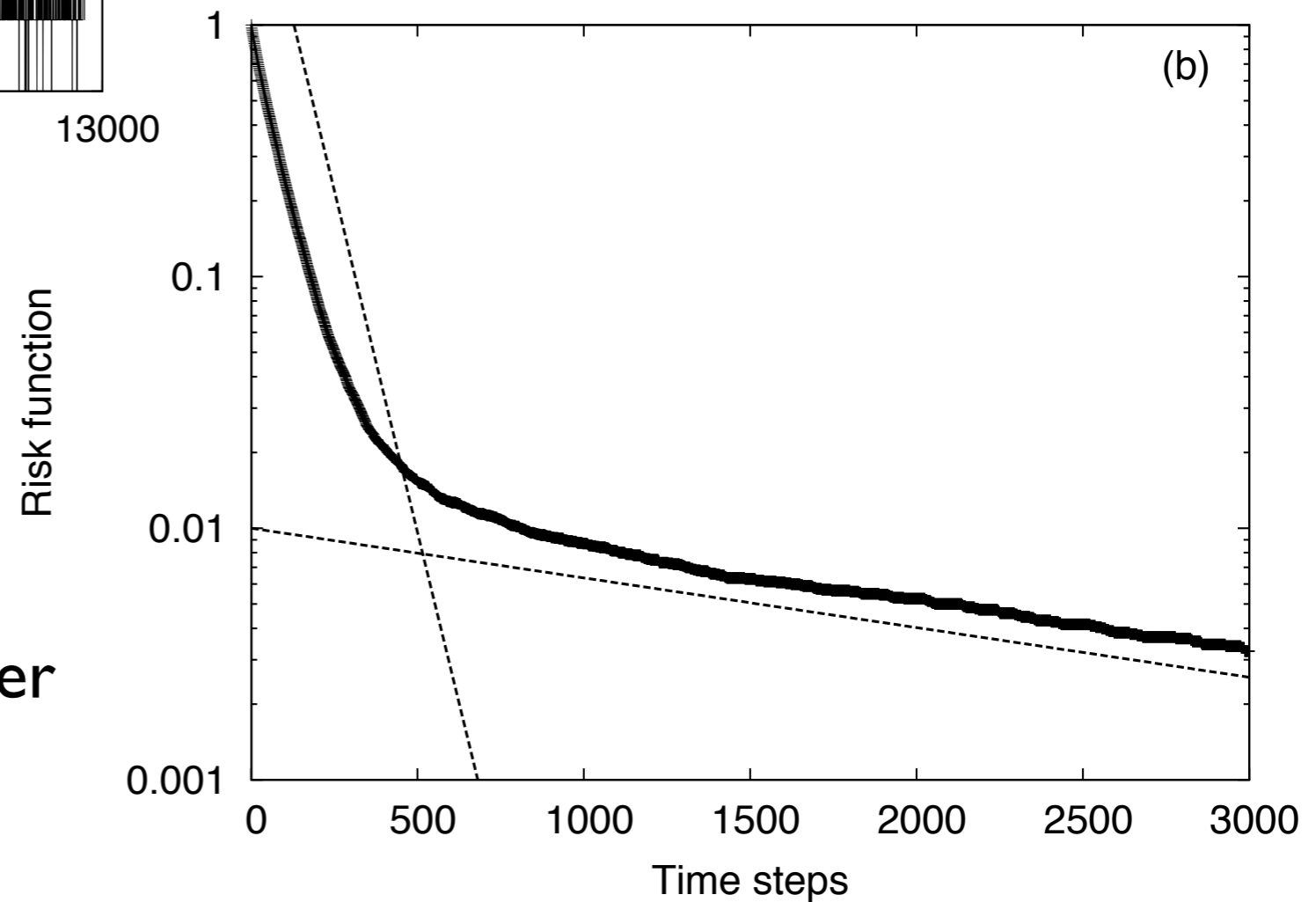
Characterise the network by a Random walk exploration of the percolated island



Start at the central node (minimises the average distance)  $\langle d \rangle = \frac{1}{(n_I - 1)} \sum_{i \neq c}^{n_i} d_{ci}$



2 time scales:  
 - fluctuations in one cluster  
 - passage from one cluster to the other



## 6) Personal identification - community imitation (PCI)

- Growing bipartite network (PA for the groups)
- Opinion formation-like model to describe taste formation

Taste builds through two different processes



Collective effects:  
social interactions



Music exchange between  
alike individuals



Individual choices



Random choices

## Model consists in 3 processes



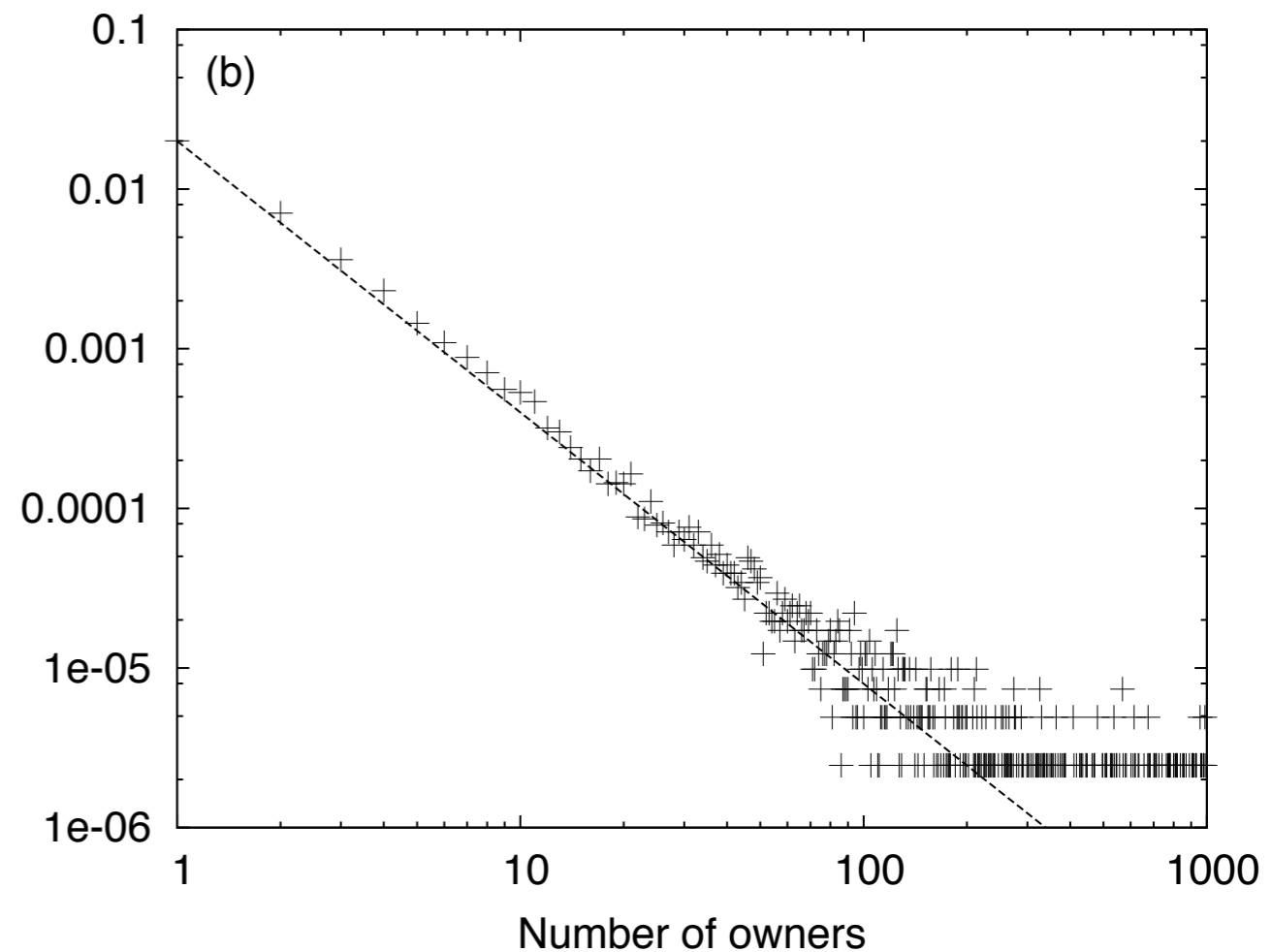
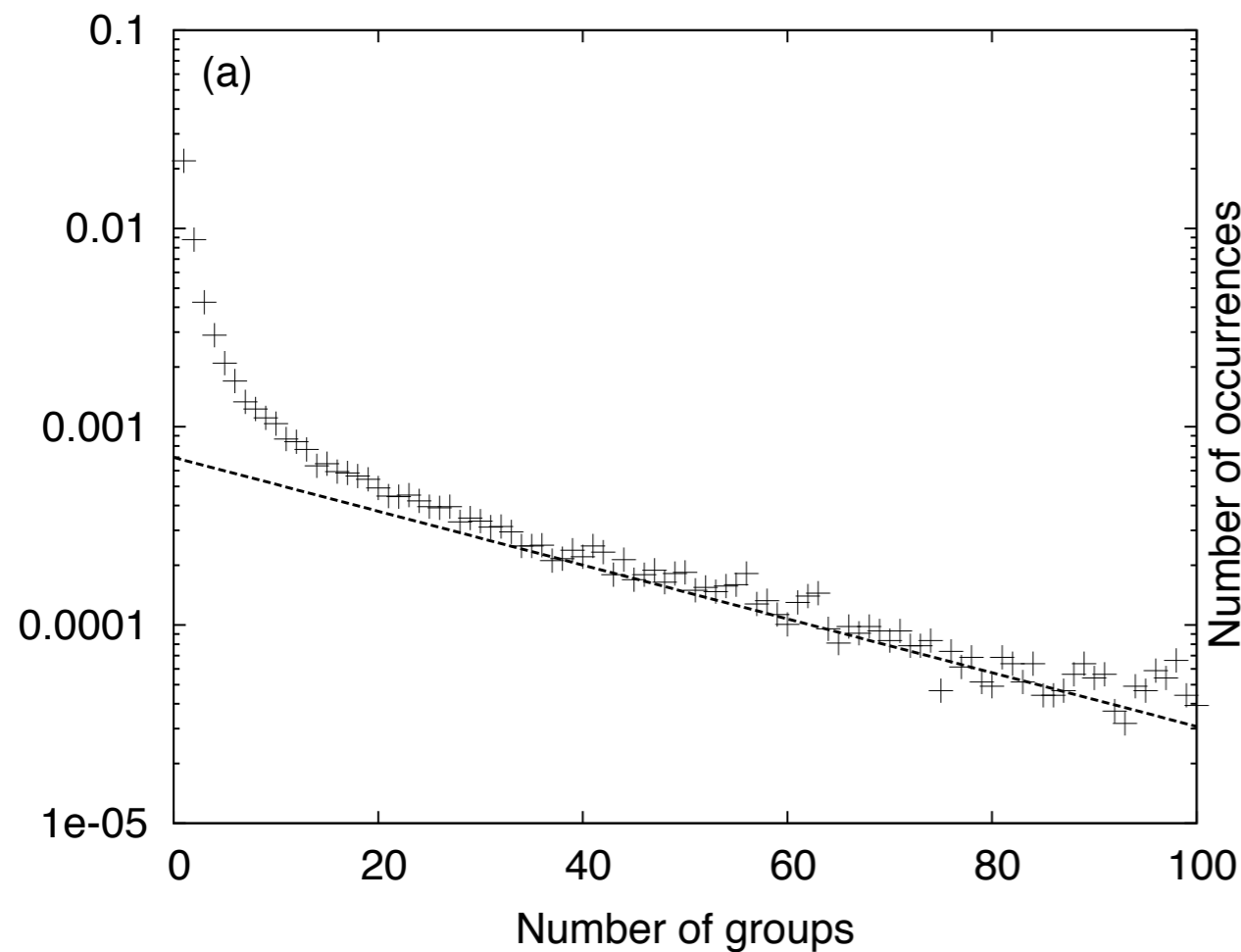
(i) A new user may enter the system, with probability  $p_I$ . His/her library contains one music group, chosen randomly in the set of previous groups with preferential attachment.

(ii) A randomly picked user adds a new music group to the library, with probability  $p_N$ . This new group appends to the collection of available music in the system.

(iii) Two randomly chosen users exchange their music knowledge, with probability  $p_E$ . The pair is selected with a probability proportional to  $e^{\frac{(\cos \theta_{\mu\lambda} - 1)}{T}}$ , where  $\theta_{\mu\lambda}$  is the angle between the vectors of their music libraries (Eq.1), defined by their cosine (Eq.2); the *temperature*  $T$  is a parameter that represents the ability of qualitatively different communities to mix together. If the pair is selected, we compare the two music libraries, and give to each user a fraction of the unknown groups of his/her partner. Let us stress that this rule ensures preferential attachment for the music groups.

## Results for the asymptotic scaling solution:

- recovers the observed pdf
- recovers the high clustering coefficients
- lowering  $T$ , structures disappear



**BUT:** Many parameters, non-stationary solutions...

# Conclusion

- Statistical methods to reveal community structures/  
music genres
- Opinion formation-like model
  
- Applicability for collaborative websites: online  
suggestions
- Musicology/Sociology
  
- More detailed exploration of the network structure
- Detailed study of the dynamical model
- Stationary version of the model => phase diagram
- Evolving structures, trends, avalanches

What kind of music do  
you listen to?

Which music community  
do you belong to?