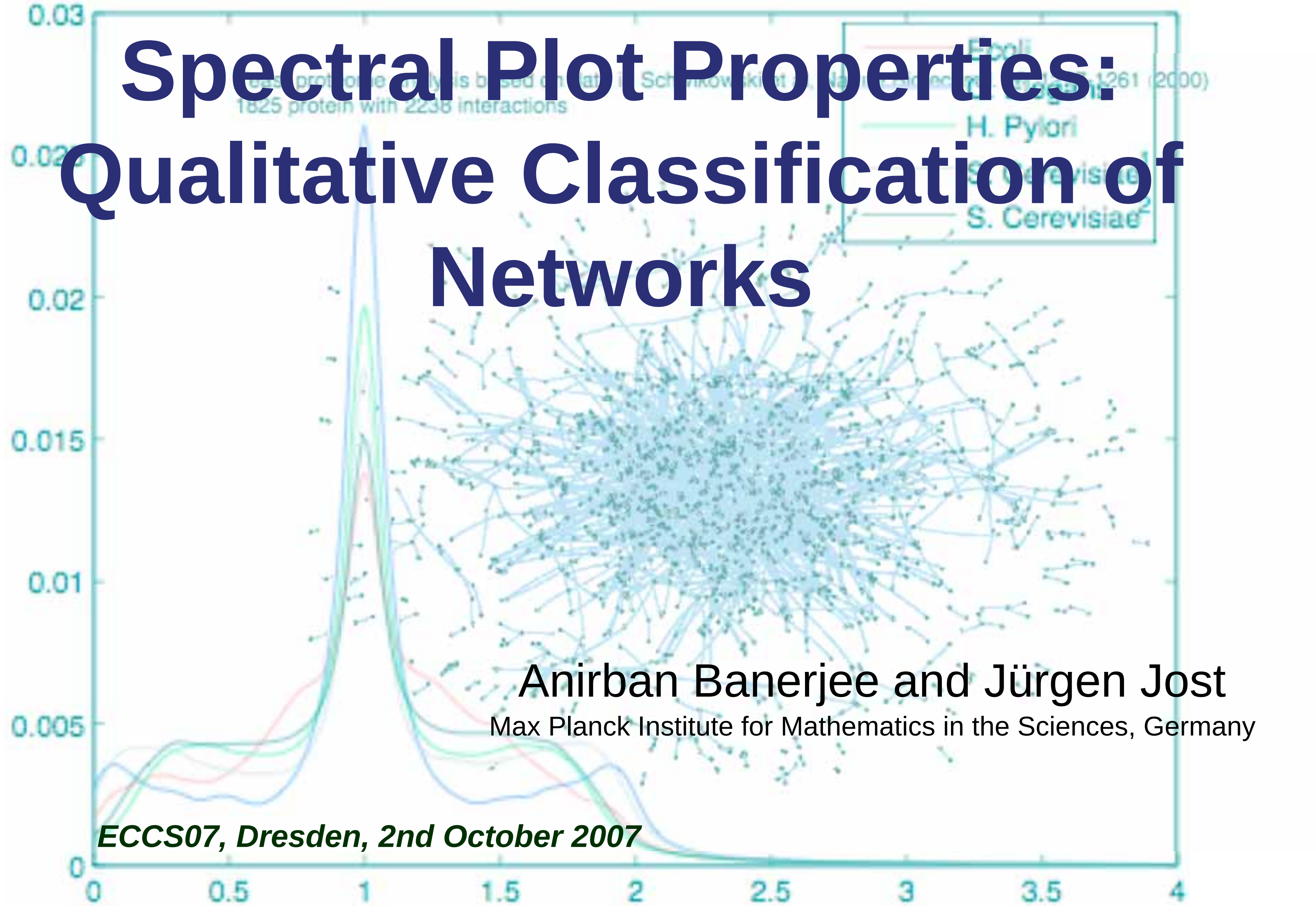
Spectral Plot Properties: Qualitative Classification of Networks



Anirban Banerjee and Jürgen Jost

Max Planck Institute for Mathematics in the Sciences, Germany

ECCS07, Dresden, 2nd October 2007

Overview of the talk

- Introduction
- Graph Laplacian operator
- Eigenfunction and graph structure
- Visualization techniques
- Qualitative classification of networks
- Conclusion and remarks

Overview of the talk

- **Introduction**
- Graph Laplacian operator
- Eigenfunction and graph structure
- Visualization techniques
- Qualitative classification of networks
- Conclusion and remarks

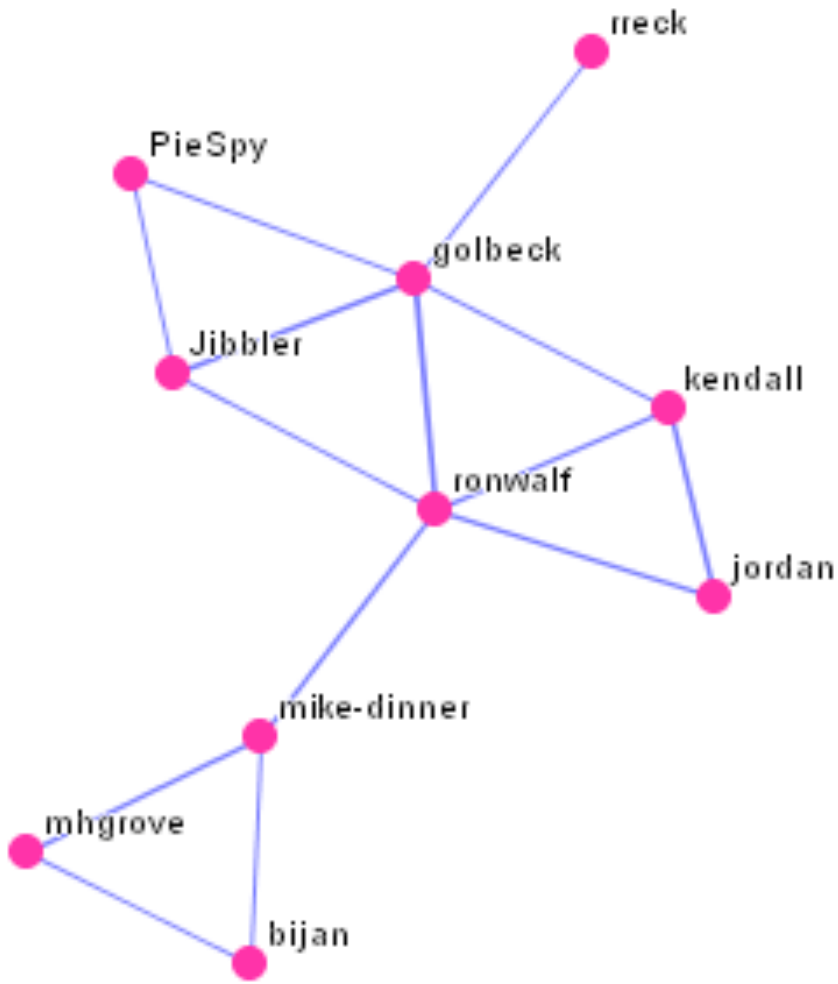
Network construction

- ❑ Component ⇒ Vertex
- ❑ Interaction or relation ⇒ Edge

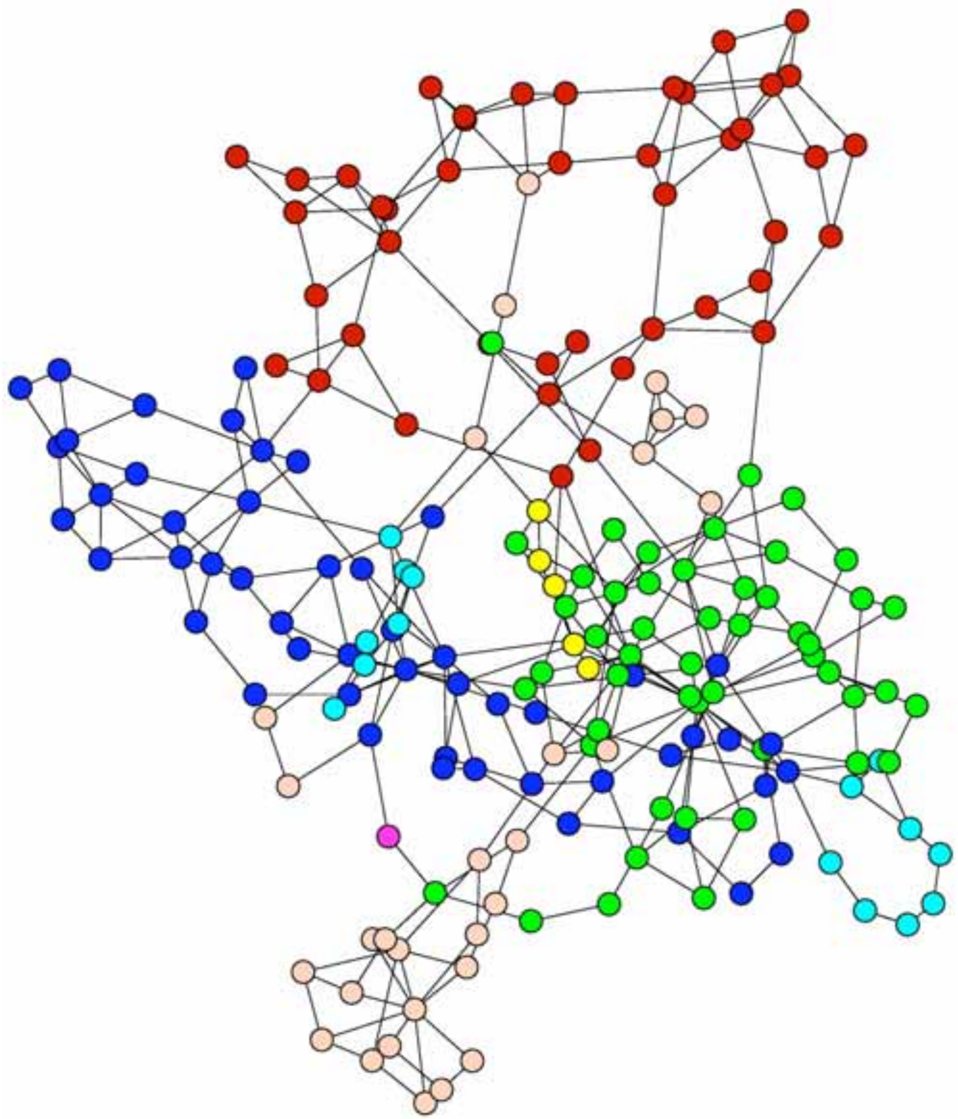
Degree of a vertex = number of its connected neighbors

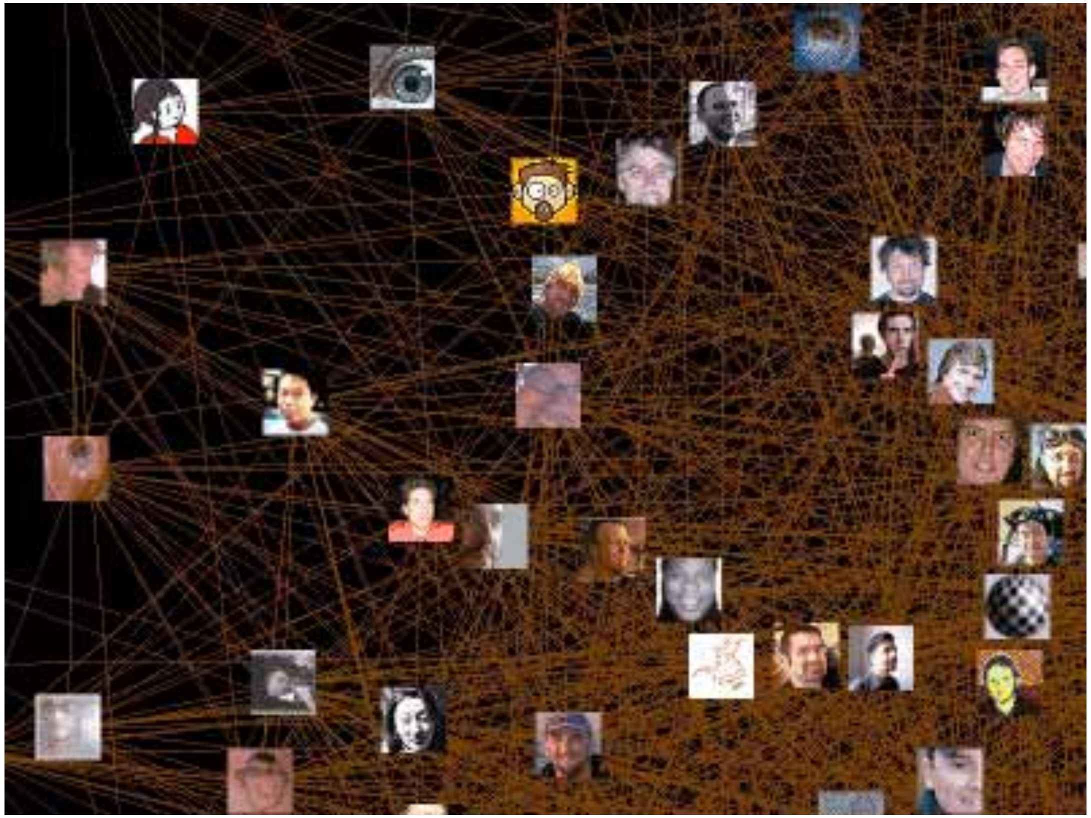
Few examples of real networks

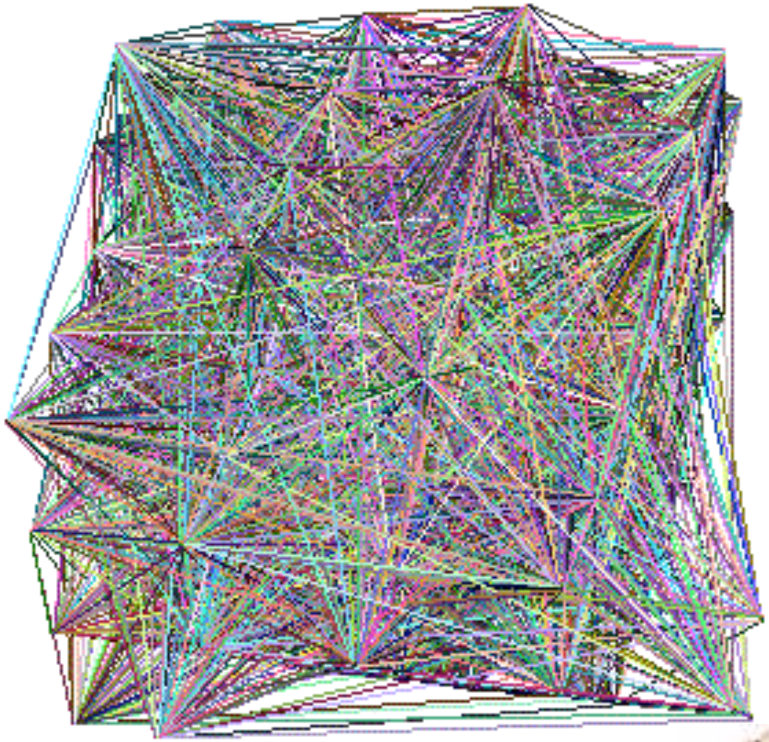
1. Protein-protein interaction network
   - Protein → Vertex
   - Direct physical interactions (binding) → Edges
2. Power grid network
   - Generator, transformers, substations → Vertices
   - High-voltage transmission lines → Edges
3. Neuronal network
   - Neuron → Vertex
   - Synaptic connections → Edges
4. Scientific collaboration network
   - Scientists → Vertices
   - Having joint publication → Edge
5. Food-web network
   - Species → Vertices
   - Predator-prey relation → Edge

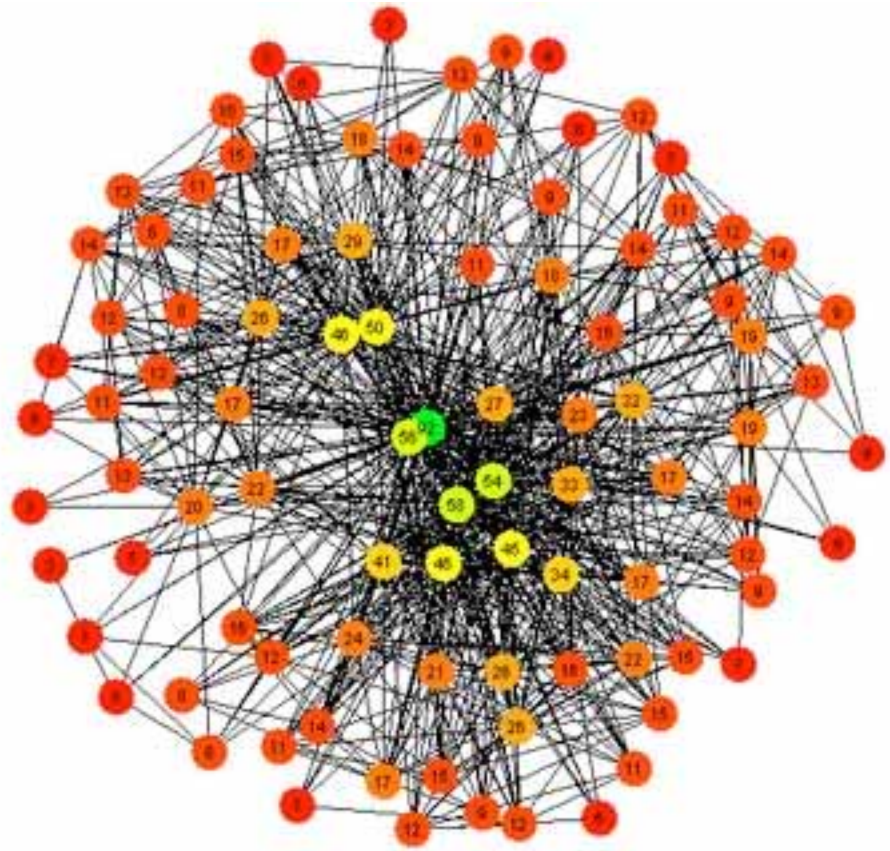


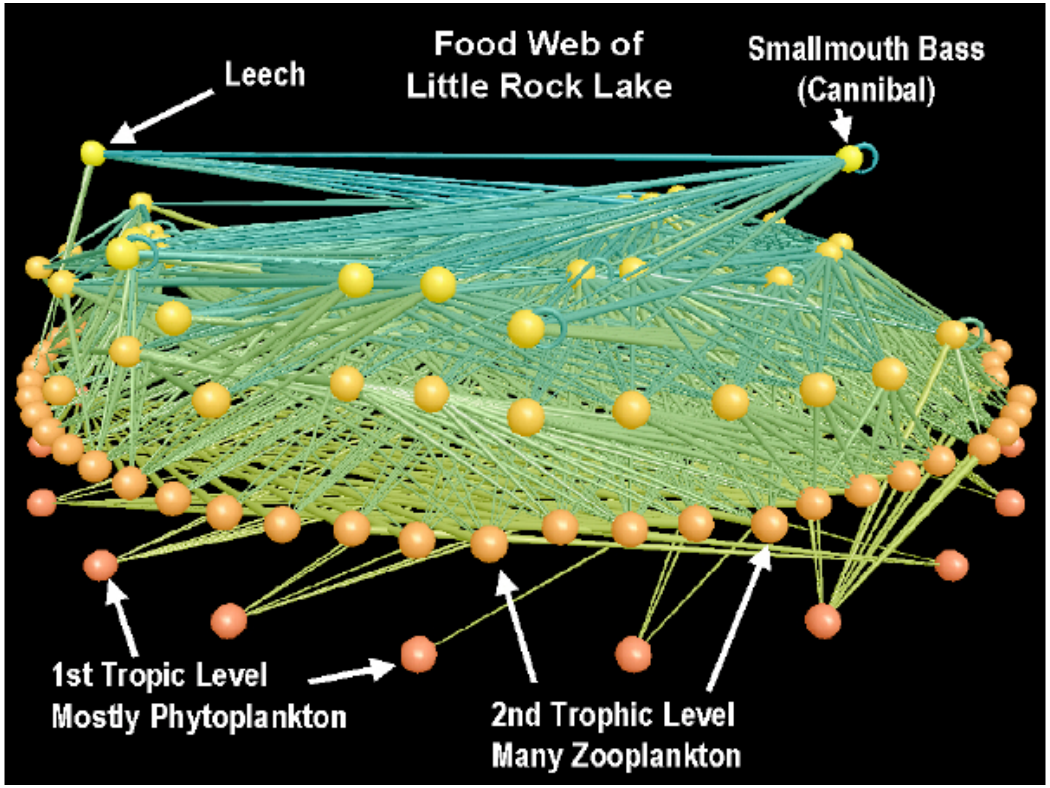
rreck
PieSpy
golbeck
Jibbler
kendall
ronwalf
jordan
mike-dinner
mhgrove
bijan



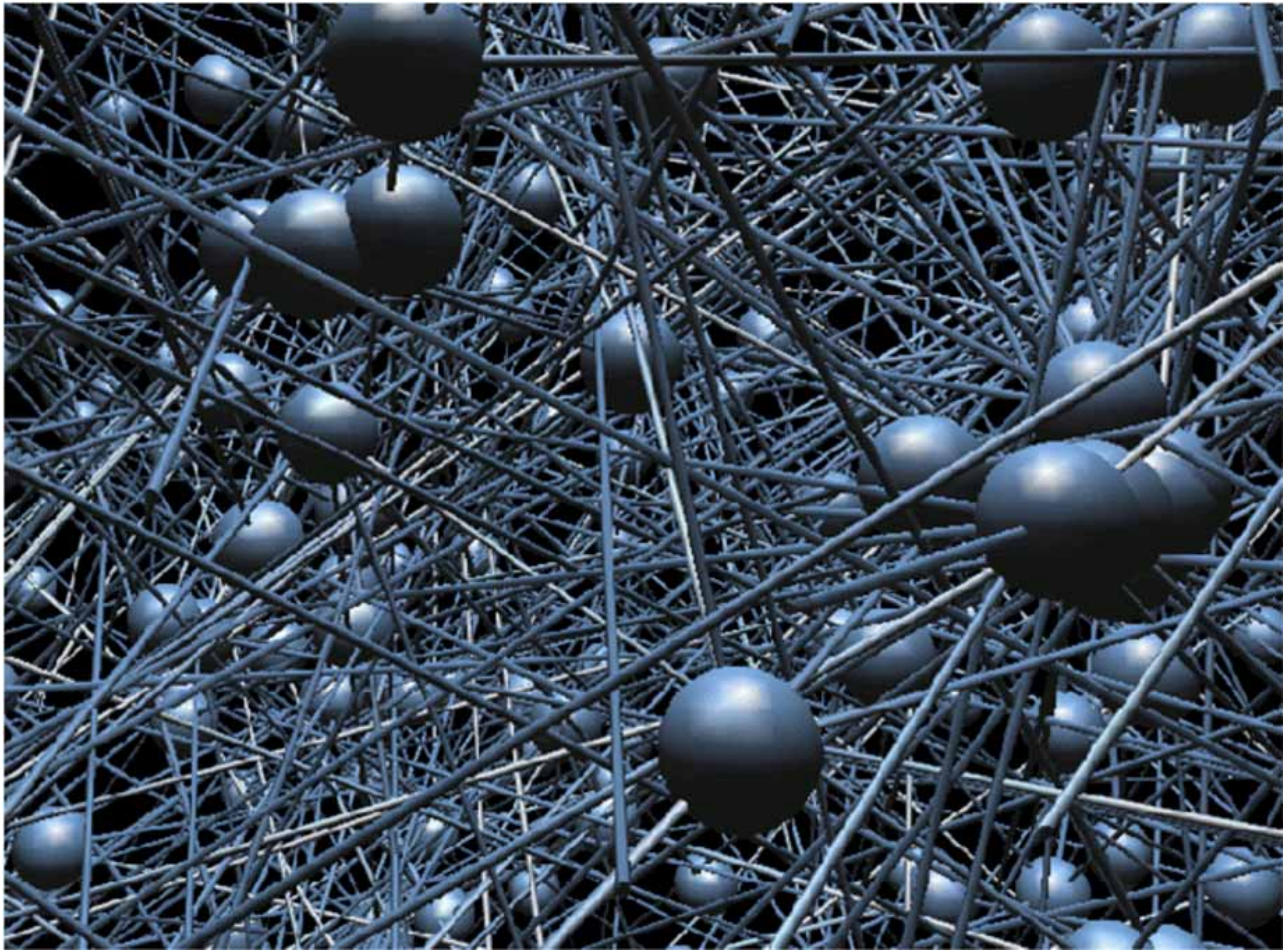


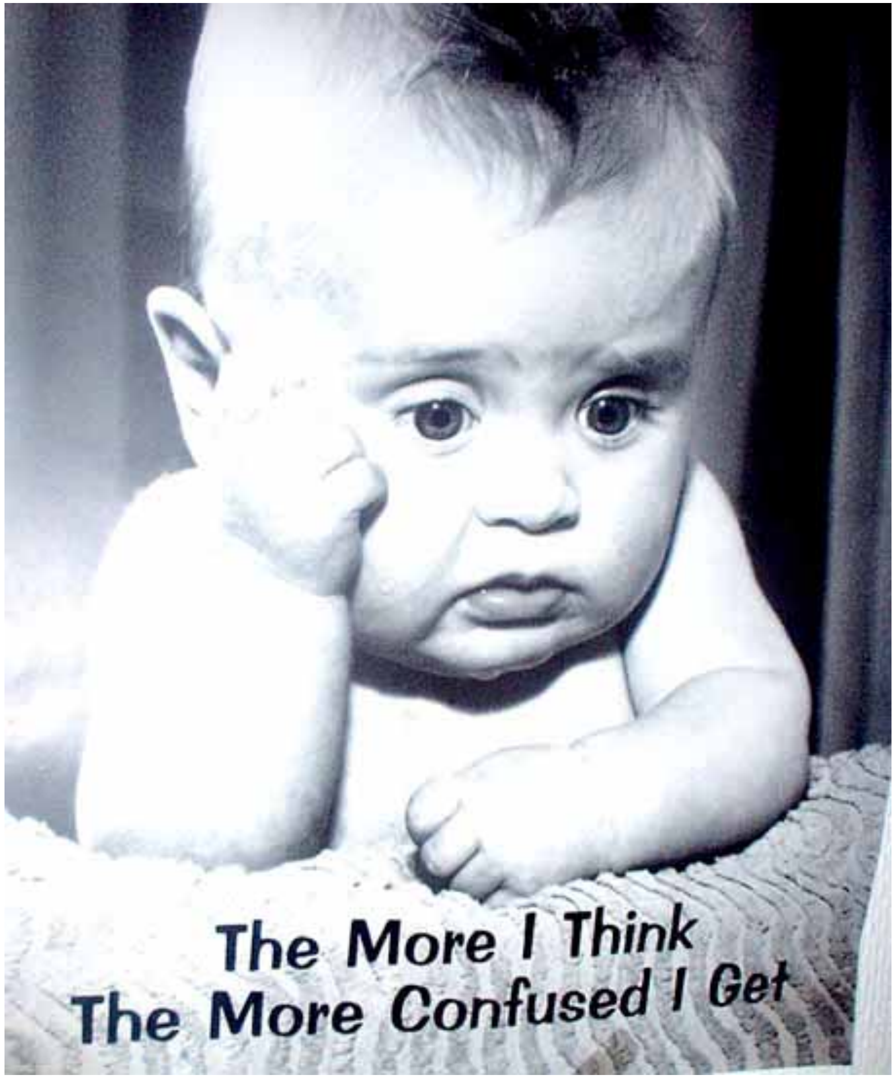






Food Web of
Little Rock Lake
Leech
Smallmouth Bass
(Cannibal)
1st Tropic Level
Mostly Phytoplankton
2nd Trophic Level
Many Zooplankton




The More I Think
The More Confused I Get

Parameters for analyzing network structure

1. Degree distribution
2. Average path length
3. Diameter
4. Clustering coefficient
5. Betweenness centrality

etc.

These invariants can not capture all qualitative aspects of a graph.

e.g. graphs with same degree distribution can have completely different structural (and dynamical) properties.

And two graphs with very different structure can have same clustering coefficient.

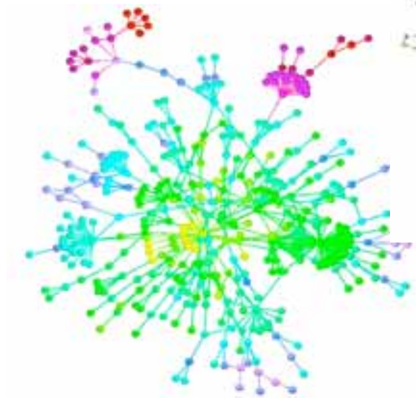
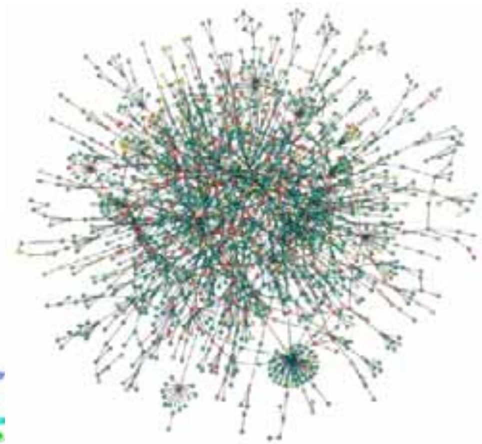
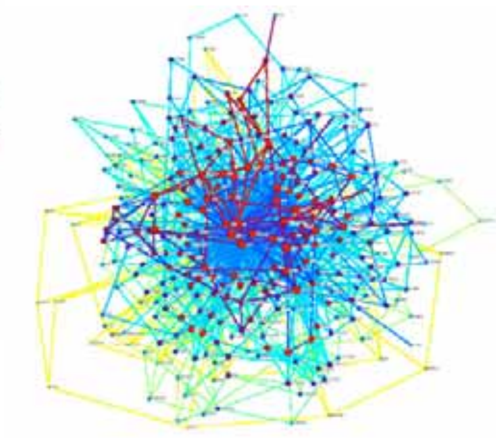
Power-law every where

Network	Size	$\langle k \rangle$	κ	γ_{out}	γ_{in}	ℓ_{real}	ℓ_{rand}	ℓ_{pow}	Reference	Nr.
WWW	325,729	4.51	900	2.45	2.1	11.2	8.32	4.77	Albert, Jeong, Barabási 1999	1
WWW	4×10^7	7		2.38	2.1				Kumar *et al.* 1999	2
WWW	2×10^8	7.5	4,000	2.72	2.1	16	8.85	7.61	Broder *et al.* 2000	3
WWW, site	260,000				1.94				Huberman, Adamic 2000	4
Internet, domain*	3,015 - 4,389	3.42 - 3.76	30 − 40	2.1 - 2.2	2.1 - 2.2	4	6.3	5.2	Faloutsos 1999	5
Internet, router*	3,888	2.57	30	2.48	2.48	12.15	8.75	7.67	Faloutsos 1999	6
Internet, router*	150,000	2.66	60	2.4	2.4	11	12.8	7.47	Govindan 2000	7
Movie actors*	212,250	28.78	900	2.3	2.3	4.54	3.65	4.01	Barabási, Albert 1999	8
Coauthors, SPIRES*	56,627	173	1,100	1.2	1.2	4	2.12	1.95	Newman 2001b,c	9
Coauthors, neuro.*	209,293	11.54	400	2.1	2.1	6	5.01	3.86	Barabási *et al.* 2001	10
Coauthors, math*	70,975	3.9	120	2.5	2.5	9.5	8.2	6.53	Barabási *et al.* 2001	11
Sexual contacts*	2810			3.4	3.4				Liljeros *et al.* 2001	12
Metabolic, E. coli	778	7.4	110	2.2	2.2	3.2	3.32	2.89	Jeong *et al.* 2000	13
Protein, S. cerev.*	1870	2.39		2.4	2.4				Mason *et al.* 2000	14
Ythan estuary*	134	8.7	35	1.05	1.05	2.43	2.26	1.71	Montoya, Solé 2000	14
Silwood park*	154	4.75	27	1.13	1.13	3.4	3.23	2	Montoya, Solé 2000	16
Citation	783,339	8.57			3				Redner 1998	17
Phone-call	53×10^6	3.16		2.1	2.1				Aiello *et al.* 2000	18
Words, cooccurence*	460,902	70.13		2.7	2.7				Cancho, Solé 2001	19
Words, synonyms*	22,311	13.48		2.8	2.8				Yook *et al.* 2001	20

R. Albert and A Barabási, *Rev. of Modern Phys. 74, 47-97, 2002*

How many invariants do we need to consider to investigate the systematic structural difference and similarities between the graphs from different classes?

For a given a particular structure, which of each features or qualities are universal, that is, shared by other structures?


What is unique and special for the structure from a particular class?

Are those invariants qualitatively good enough to identify the domain of a given an empirical graph?

So instead of focusing on particular and specific aspects and quantities in details, of a given a large and complex structure,

we can try to obtain, at least at some rough level, a simultaneous representation of all its qualitative features.

Therefore, we are advocating here a tentative classification scheme for empirical networks based on underlying global qualitative properties detected through the *graph Laplacian spectrum*

that,

on one hand, are complete qualitative characterization of a graph and on other hand, can be easily graphically represented and therefore visually analyzed and compared.

Overview of the talk

- Introduction
- **Graph Laplacian operator**
- Eigenfunction and graph structure
- Visualization techniques
- Qualitative classification of networks
- Conclusion and remarks

Γ : A finite undirected graph with N vertices.

Vertices: $i, j \in \Gamma$ are connected $\Rightarrow$ $i \sim j$

n_i $\Rightarrow$ degree of vertex i

For any function $v : \Gamma \to \mathbb{R}$

$$\Delta v(i) := v(i) - \frac{1}{n_i} \sum_{j, j \sim i} v(j)$$

A.Banerjee, J.Jost, *On the spectrum of the normalized graph Laplacian*, submitted.

$$\mathcal{L}u(i) := u(i) - \frac{1}{n_i} \sum_{j, j \sim i} \frac{1}{\sqrt{n_i n_j}} u(j)$$

F.Chung, *Spectral graph theory*,1997

$$a_{ij} = \begin{cases} 1, & \text{if } i = j \text{ and } n_i \neq 0 \\ -\frac{1}{\sqrt{n_i n_j}}, & \text{if } ij \text{ is an edge} \\ 0, & \text{otherwise.} \end{cases}$$

$$Lu(i) := n_i u(i) - \sum_{j, j \sim i} u(j)$$

A.Bollobás, *Modern graph theory*,1998

$$a_{ij} = \begin{cases} n_i, & \text{if } i = j \\ -1, & \text{if } ij \text{ is an edge} \\ 0, & \text{otherwise} \end{cases}$$

Properties of this operator

- This operator is symmetric about the product

  $(u, v) := \sum_{i \in \Gamma} n_i u(i) v(i)$ ⇒ eigenvalues are real

- $(\Delta u, u) \geq 0$ ⇒ eigenvalues are non-negative

- $\Delta u = 0$ for u = constant ⇒ $\lambda_{min} = 0$

Eigenvalues of this operator

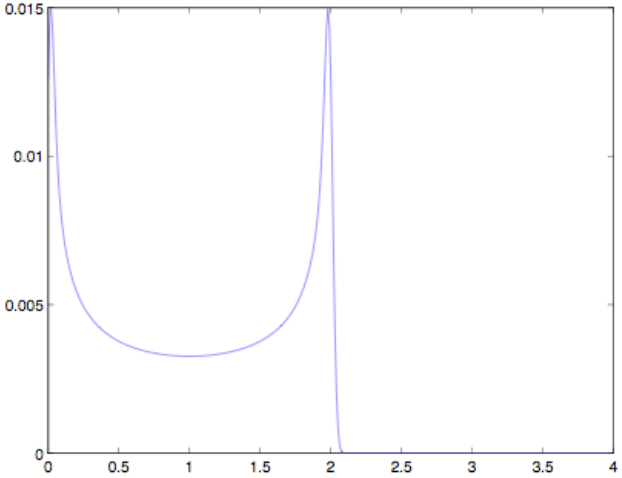
$$0 = \lambda_0 \leq \lambda_1 \leq \cdots \leq \lambda_{N-1} \leq 2$$

- Multiplicity of 0 ⇔ # components in the graph
- Eigenvalues λ_1 and λ_{N-1}
- $\lambda_{N-1} = 2$ ⇔ graphs is bipartite
  - Spectrum is symmetric about 1
- Complete graph with *N* vertices ⇔

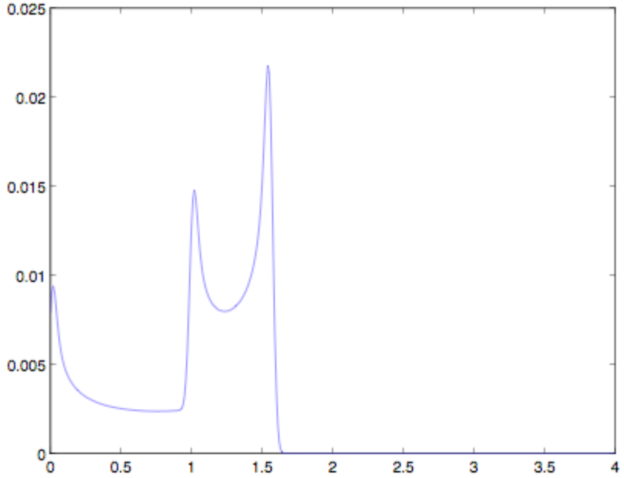
$$\lambda_1 = \lambda_2 = \cdots = \lambda_{N-1} = \frac{N}{N-1}$$

Few example of spectral plots

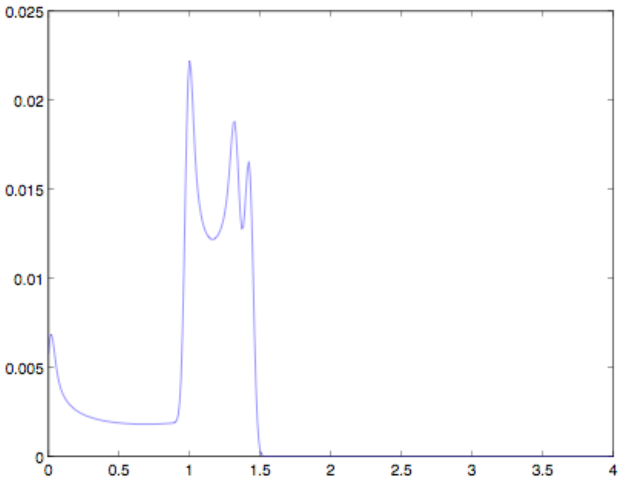
1D regular ring lattices



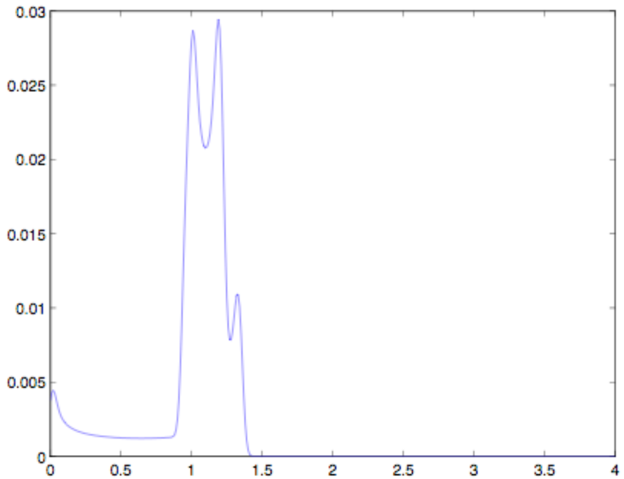
Deg. of each node = 2



Deg. of each node = 4

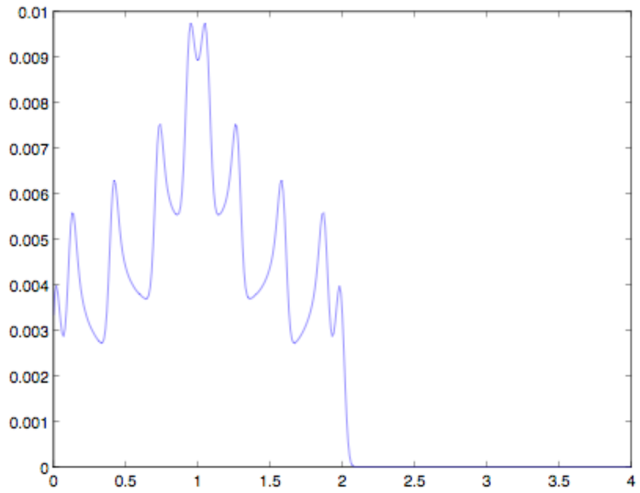


Deg. of each node = 6

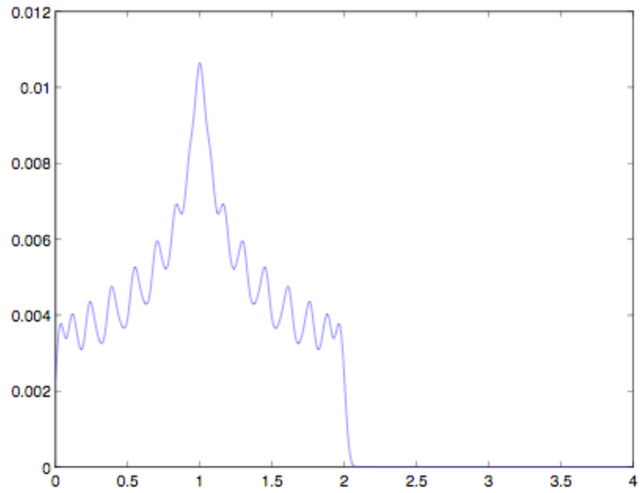


Deg. of each node = 10

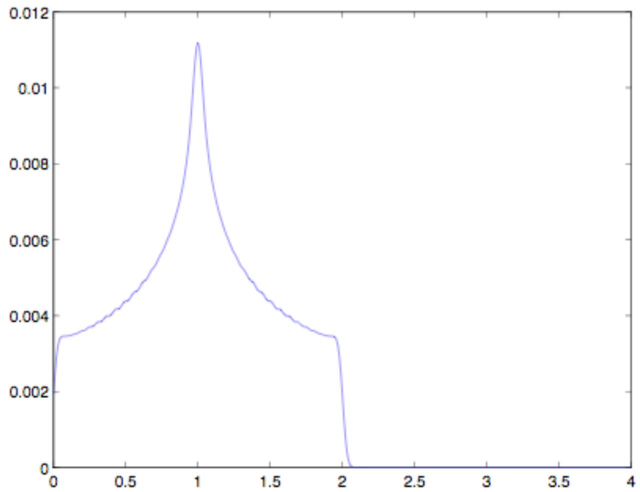
2D square grids



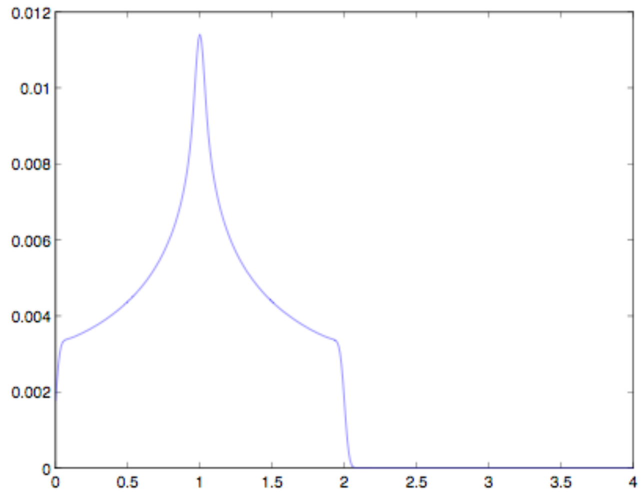
Dimension = 5 X 2000



Dimension = 10 X 1000

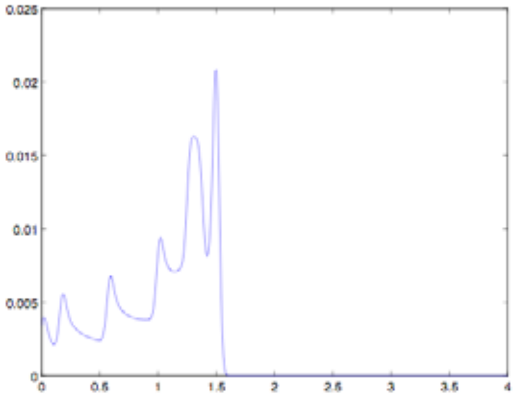


Dimension = 25 X 400

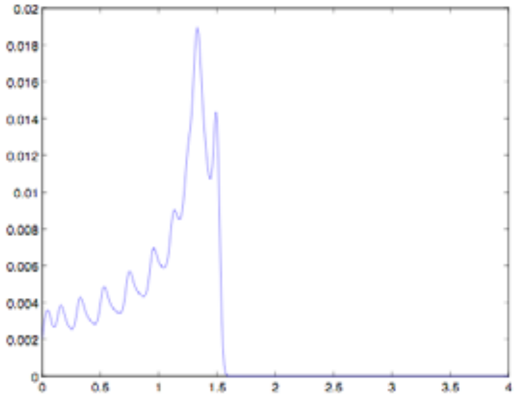


Dimension = 100 X 100

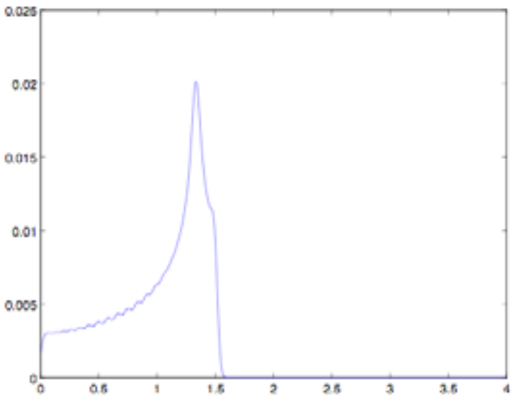
2D square grid with one diagonal



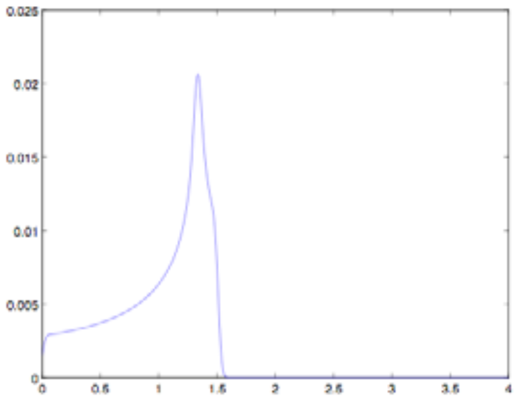
Dimension = 5 X 2000



Dimension = 10 X 1000

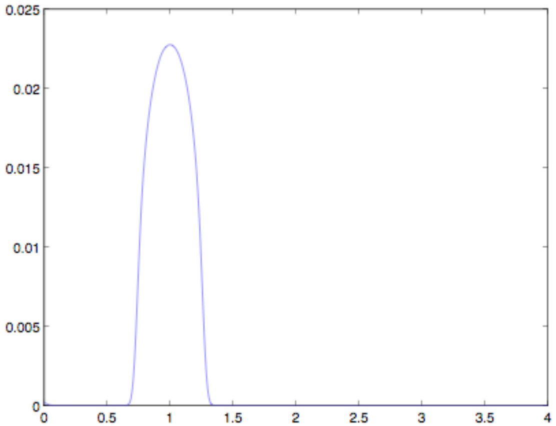


Dimension = 25 X 400

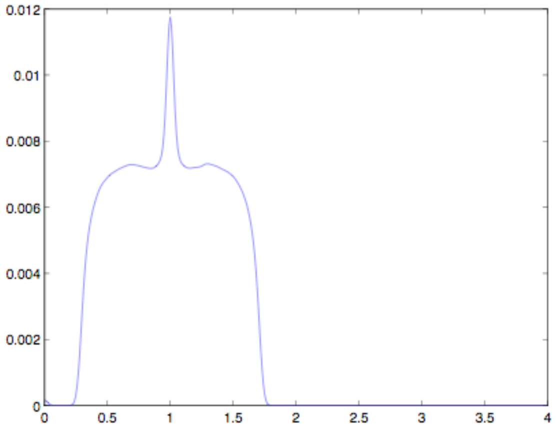


Dimension = 100 X 100

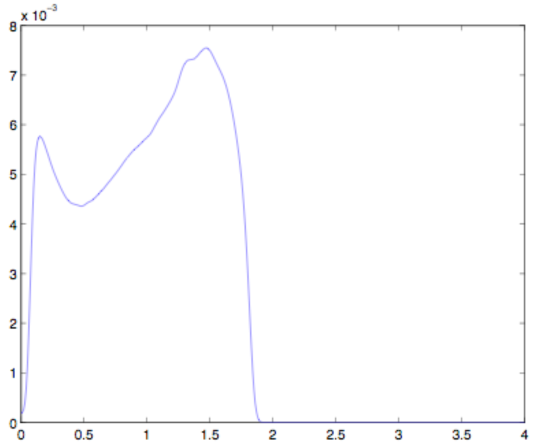
Three generic models



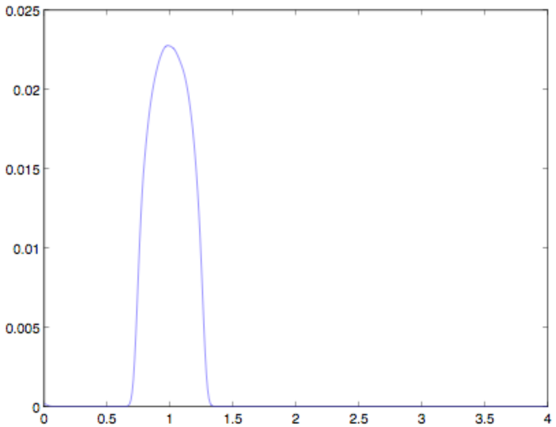
Random network

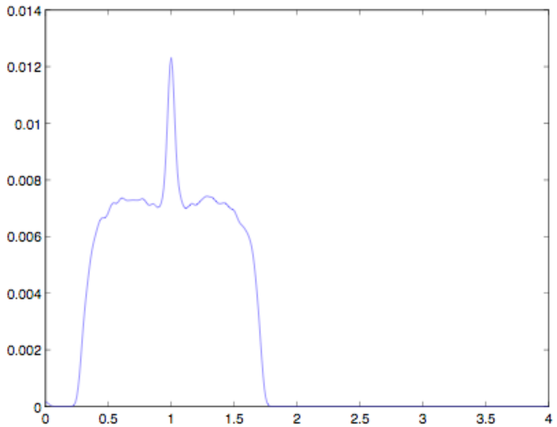


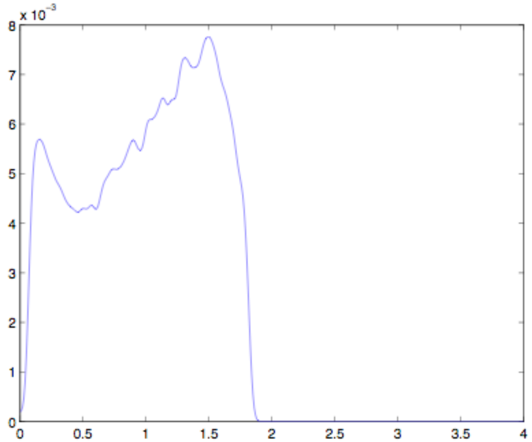
Small-world network



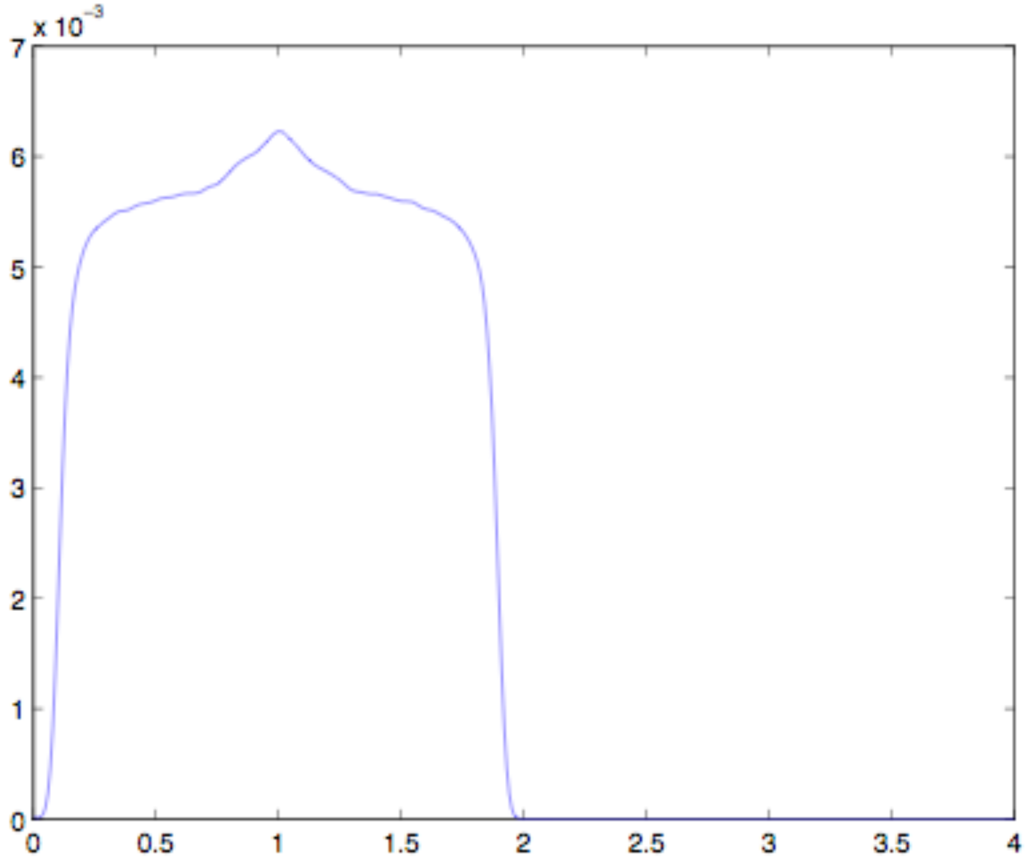
Scale-free network





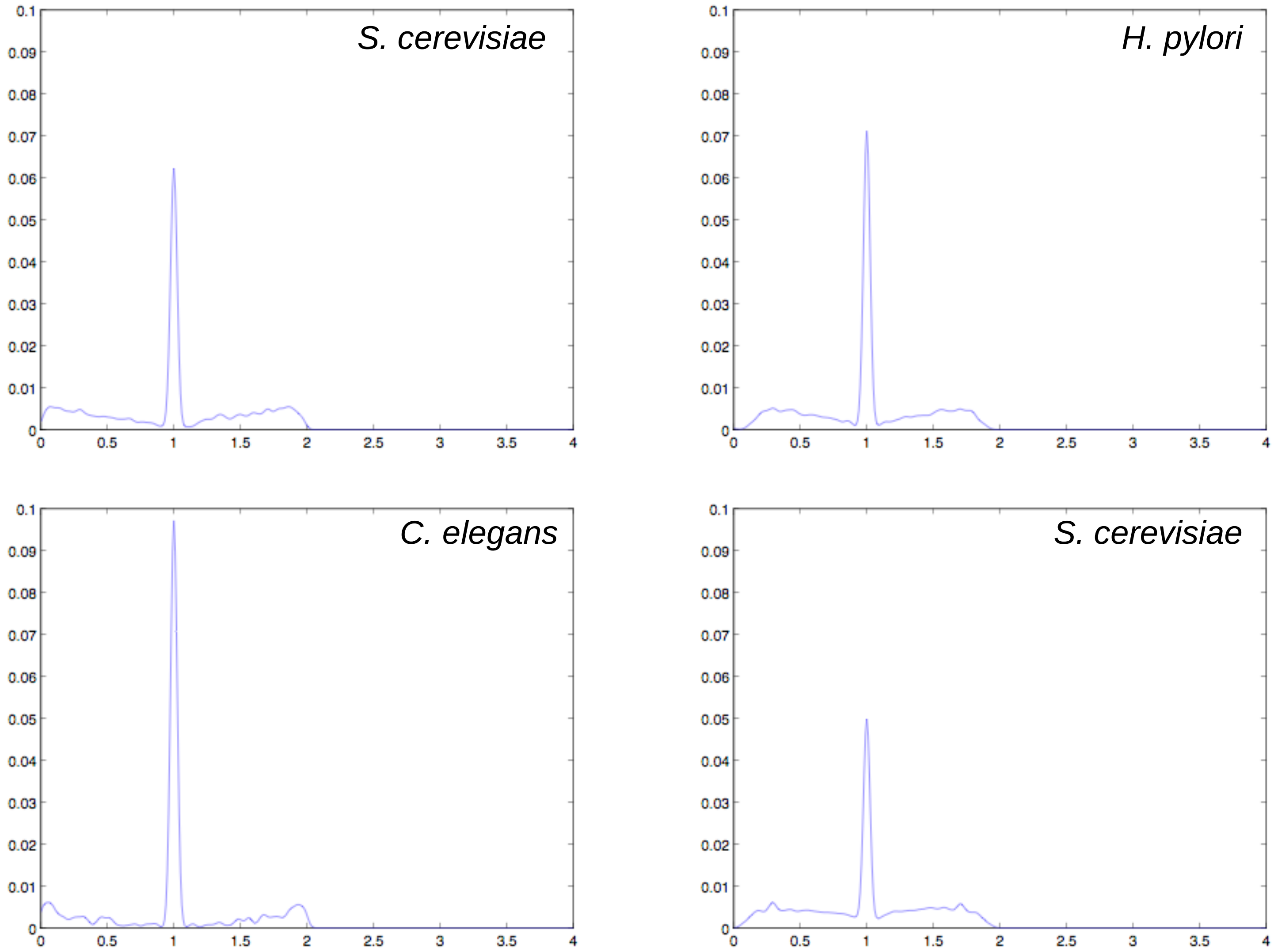


Small world network (created from 2D grid)



Small-world network, created by rewiring 2-D square grid

Protein-protein interaction networks



Overview of the talk

- Introduction
- Graph Laplacian operator
- **Eigenfunction and graph structure**
- Visualization techniques
- Qualitative classification of networks
- Conclusion and remarks

A.Banerjee, J.Jost, *Graph spectra as a systematic tool in computational biology*, submitted.

Let think of a graph, representing real data as a structure that has evolved from some simpler precursors.

Constructions with different graph operation related to the evolution of a network describe certain processes of graph formation that leave characteristic traces in the spectrum.

A solution u_k of the eigenvalue equation

$$\Delta u_k - \lambda_k u_k = 0$$

- can be localize
- can be global

A.Banerjee, J.Jost, *Discrete Applied Mathematics*, to appear.

Vertex doubling

Doubling a single vertex $j_0 \in \Gamma$

i_0 is double of j_0

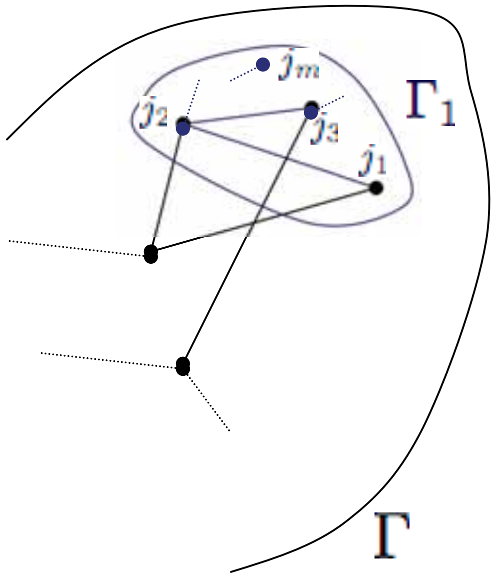
-- generates eigenvalue 1 with an eigenfunction

$$u_1(i) = \begin{cases} 1, & \text{for } i = j_0 \\ -1, & \text{for } i = i_0 \\ 0, & \text{otherwise} \end{cases}$$

High peak at 1: evolve by sequence of vertex duplication

Motif duplication

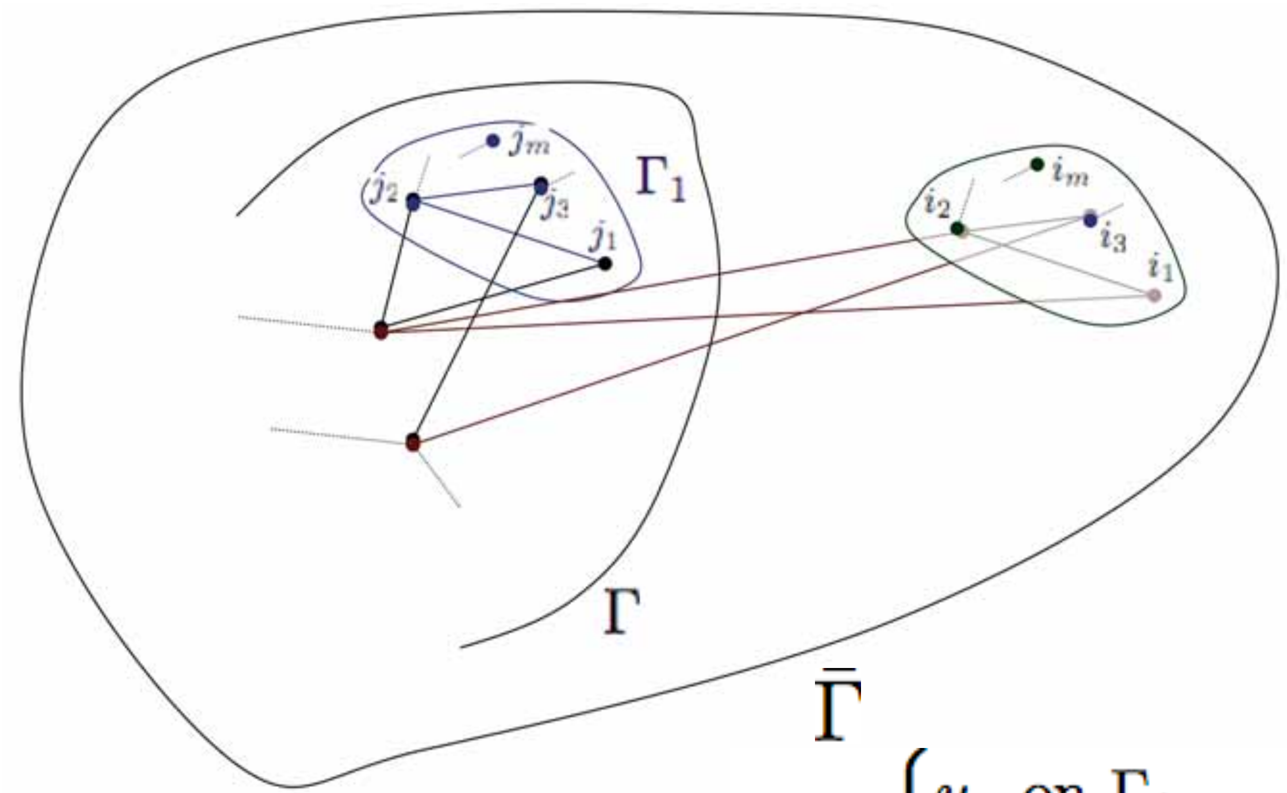
Motif: small sub-graph (where as the graph is supposed to be large) containing all edges of the graph between vertices of that subgraph



$$\frac{1}{n_i} \sum_{j \in \Gamma_1, j \sim i} u(j) = (1 - \lambda) u(i) \text{ for all } i \in \Gamma_1 \text{ and some } \lambda$$

Motif duplication

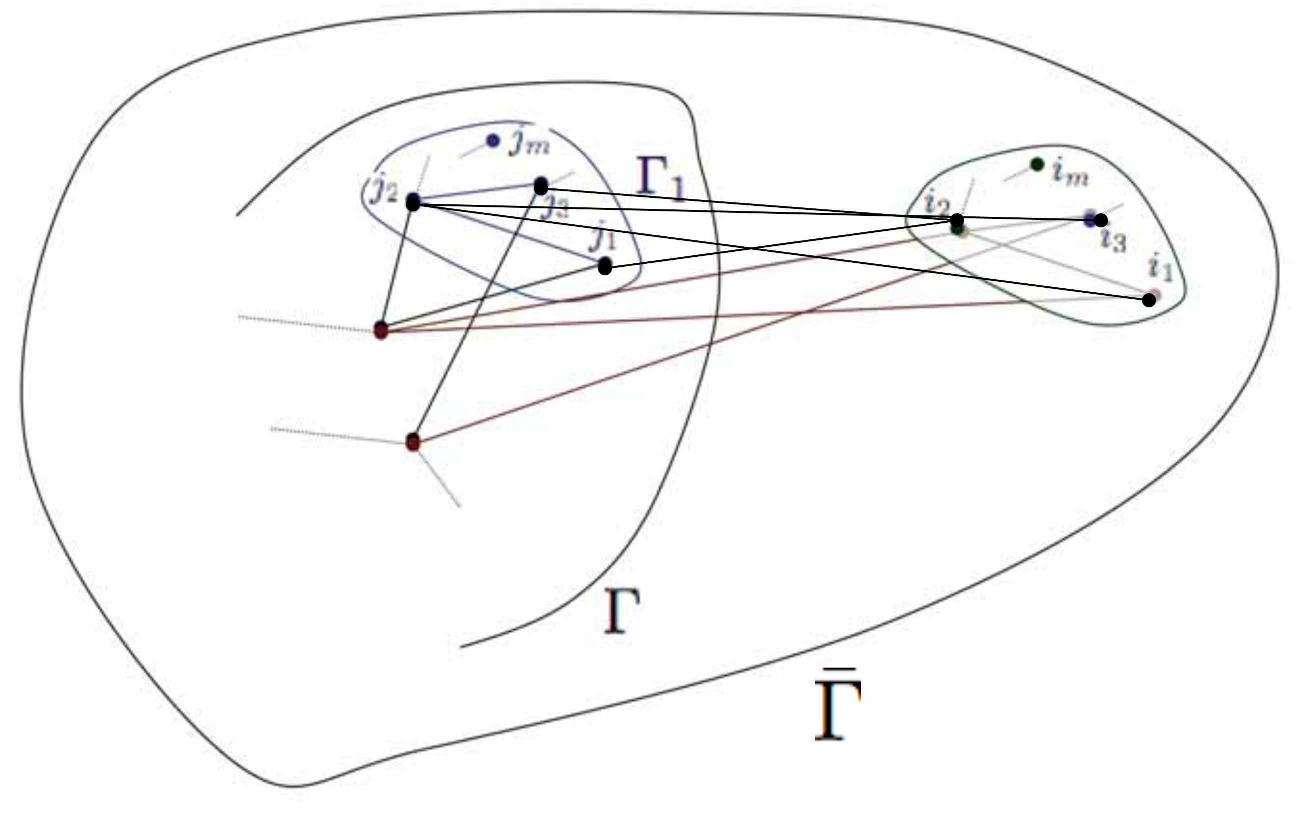
Motif: small sub-graph (where as the graph is supposed to be large) containing all edges of the graph between vertices of that subgraph



$$u^{\lambda}_{\bar{\Gamma}} = \begin{cases} u, & \text{on } \Gamma_1 \\ -u. & \text{on the double of } \Gamma_1 \end{cases}$$

Another motif duplication

Motif: small sub-graph (where as the graph is supposed to be large) containing all edges of the graph between vertices of that subgraph



Increment of multiplicity of the eigenvalue 1 by *m*

Edge doubling

Doubling an edge that connects vertices j_1, j_2

-- produce eigenvalues

$$\lambda_{\pm} = 1 \pm \frac{1}{\sqrt{n_{j_1} n_{j_2}}}$$

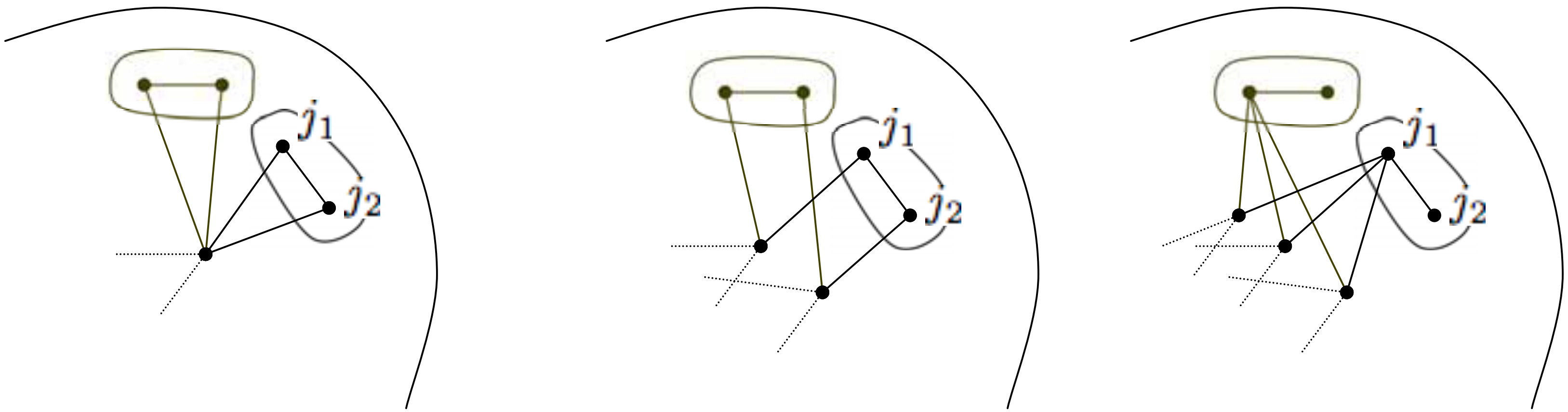
Symmetric about 1 and close to 1 when n_{j_1}, n_{j_2} are sufficiently large.

High peak at 1, but not too sharp

Edge doubling

Doubling an edge that connects vertices j_1, j_2
with $n_{j_1} n_{j_2} = 4$

-- produce eigenvalues 3/2 and 1/2



Entire graph doubling

Double the entire graph Γ with vertices $p_1, \ldots, p_N$

Γ' be copy of Γ with vertices $q_1, \ldots, q_N$ and with same connection pattern.

Connect each q_α to all neighbors of p_α

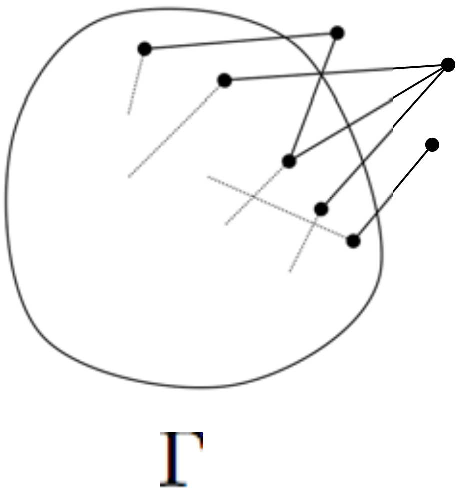
New graph has the same eigenvalues as Γ, plus the eigenvalue 1 with the multiplicity *N.*

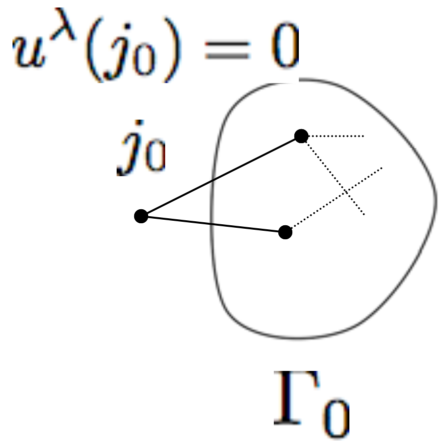
protein-protein interaction network do have high multiplicity, not of the order of half of the system size -- subsequent mutations after the genome duplication.

A.Banerjee, J.Jost, *Laplacian spectrum and protein-protein interaction networks*, preprint.

Motif joining

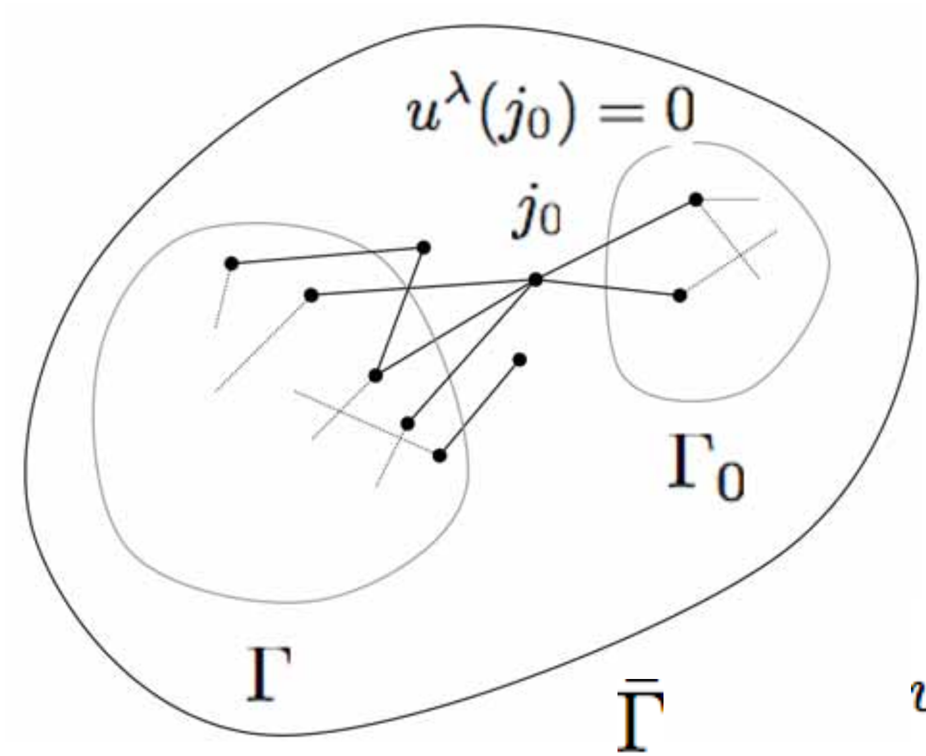
Motif: small sub-graph (where as the graph is supposed to be large) containing all edges of the graph between vertices of that subgraph





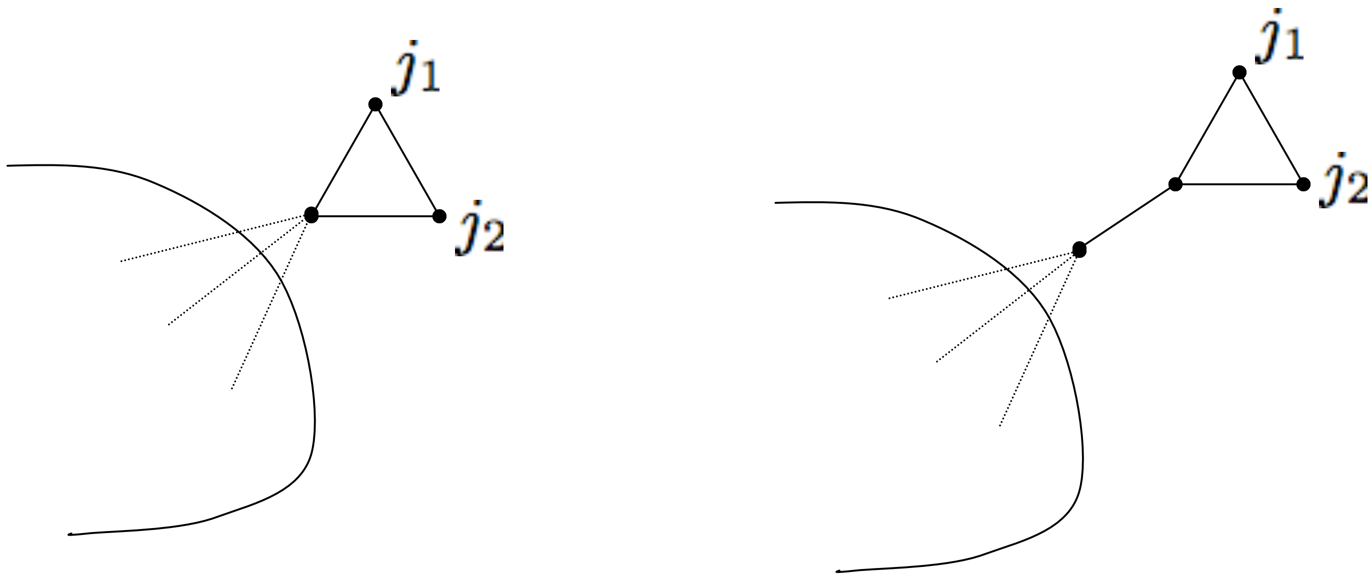
Motif joining

Motif: small sub-graph (where as the graph is supposed to be large) containing all edges of the graph between vertices of that subgraph



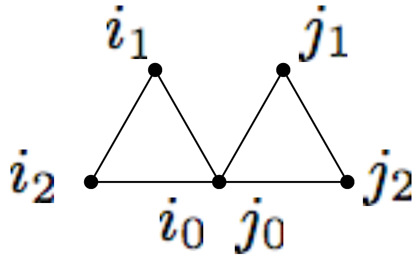
$$u^{\lambda}_{\bar{\Gamma}} = \begin{cases} u^{\lambda}, & \text{on } \Gamma_0 \\ 0. & \text{elsewhere} \end{cases}$$

Triangle joining



$$u_{3/2}(i) = \begin{cases} 1, & \text{for } i = j_1 \\ -1, & \text{for } i = j_2 \\ 0, & \text{otherwise} \end{cases}$$

Triangle joining

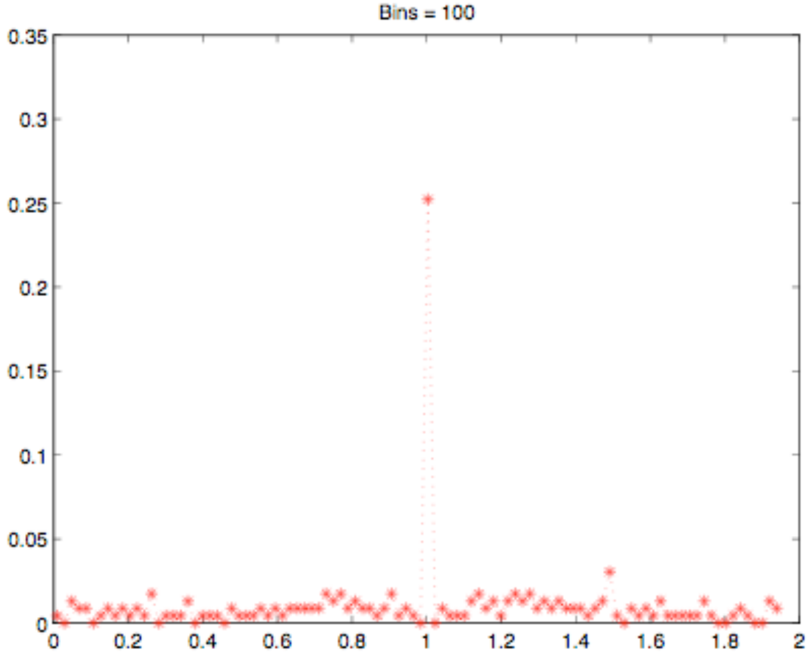


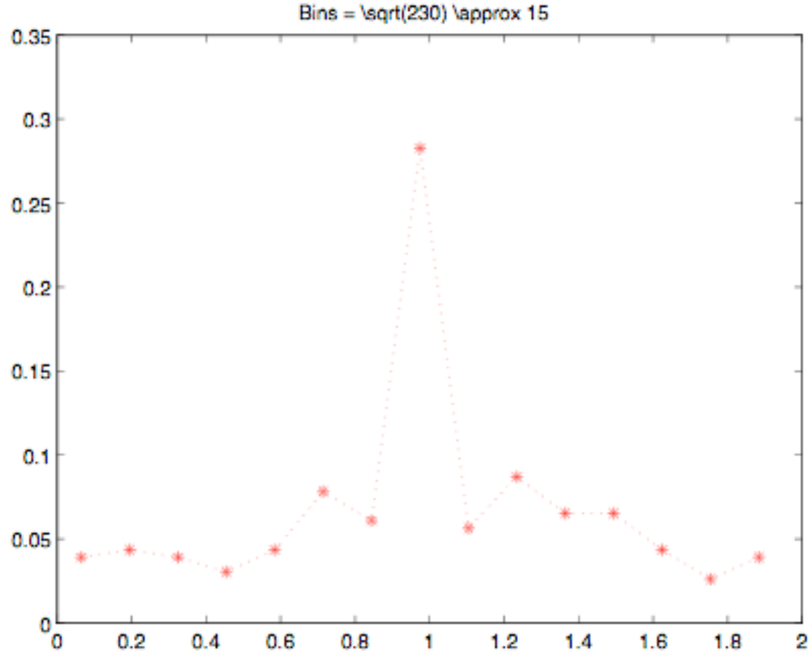
Generate not only eigenvalue 3/2, but also 1/2

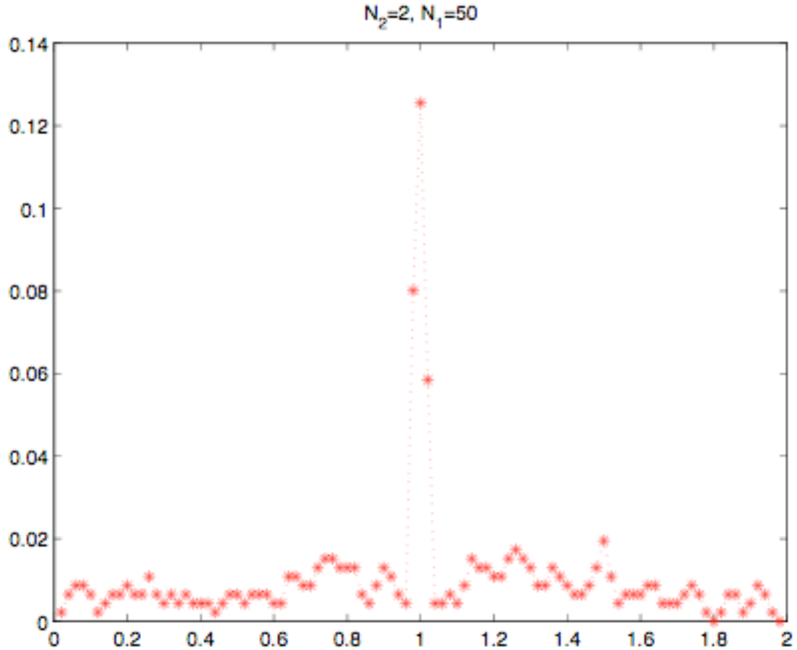
Overview of the talk

- Introduction
- Graph Laplacian operator
- Eigenfunction and graph structure
- **Visualization techniques**
- Qualitative classification of networks
- Conclusion and remarks

A.Banerjee, J.Jost, *Theory in Biosciences*, 126(1) 15-21, 2007.







(Overlapping bins such as, $[0, \frac{N_2}{N_1}], [\frac{1}{N_1}, \frac{N_2+1}{N_1}], \ldots, [\frac{2N_1-N_2}{N_1}, \frac{2N_1}{N_1}]$)

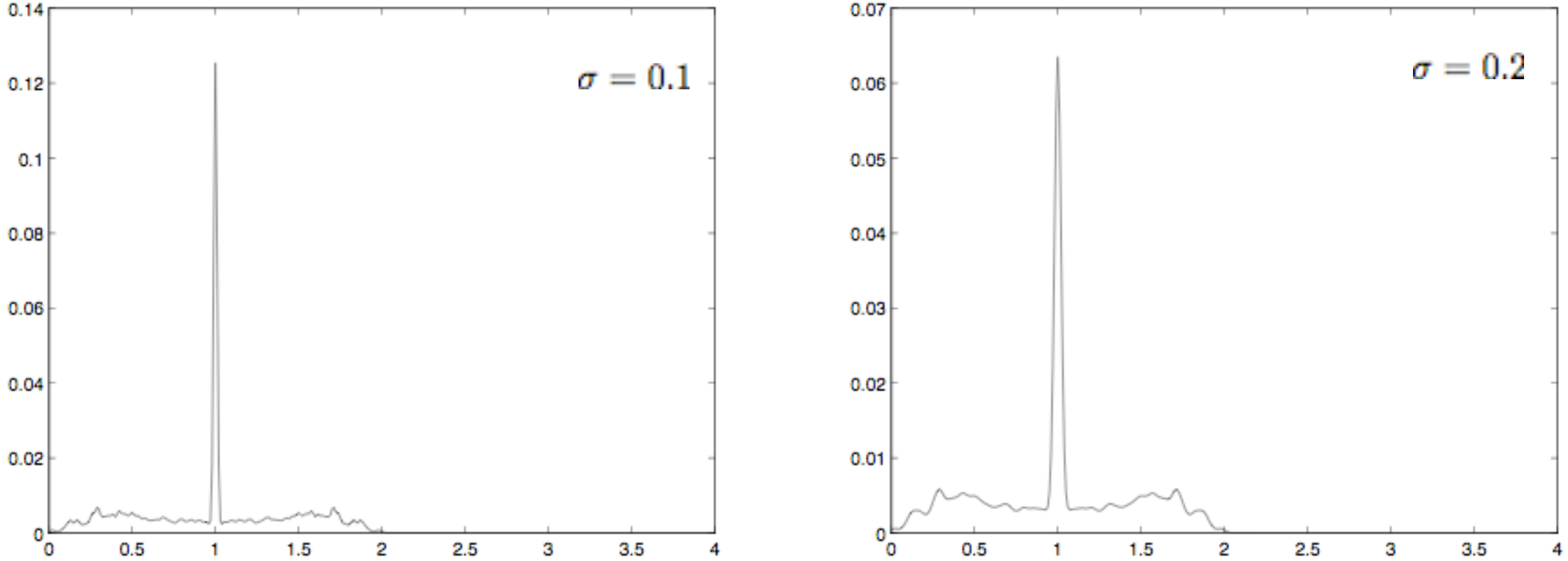
Protein-protein interaction network *E.coli.* Size = 230

Convolve with a kernel

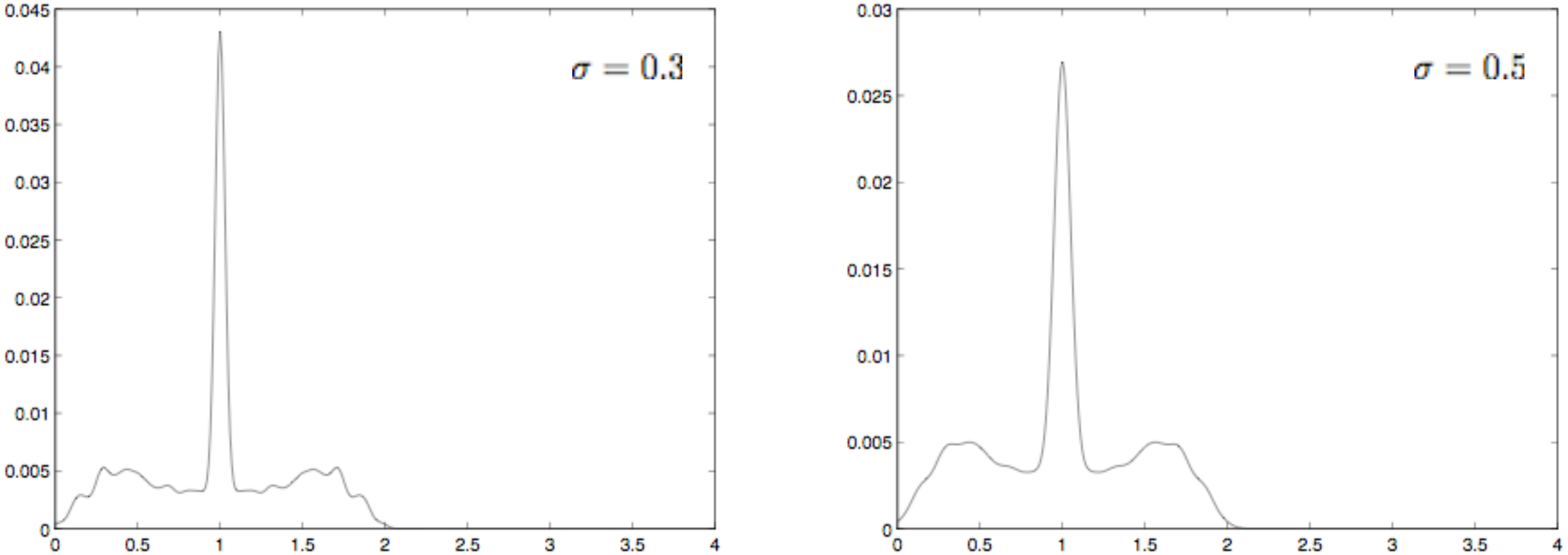
- Gaussian: $\frac{1}{\sqrt{2\pi\sigma}} exp(-\frac{(x-m_x)^2}{2\sigma^2})$
- Cauchy-Lorentz: $\frac{1}{\pi}\frac{\gamma}{(x-m)^2+\gamma^2}$

$$f(x) = \int g(x,\lambda) \sum_k \delta(\lambda, \lambda_k)\, \mathrm{d}\lambda = \sum_k g(x, \lambda_k).$$

Plots with different kernel values



C. elegans

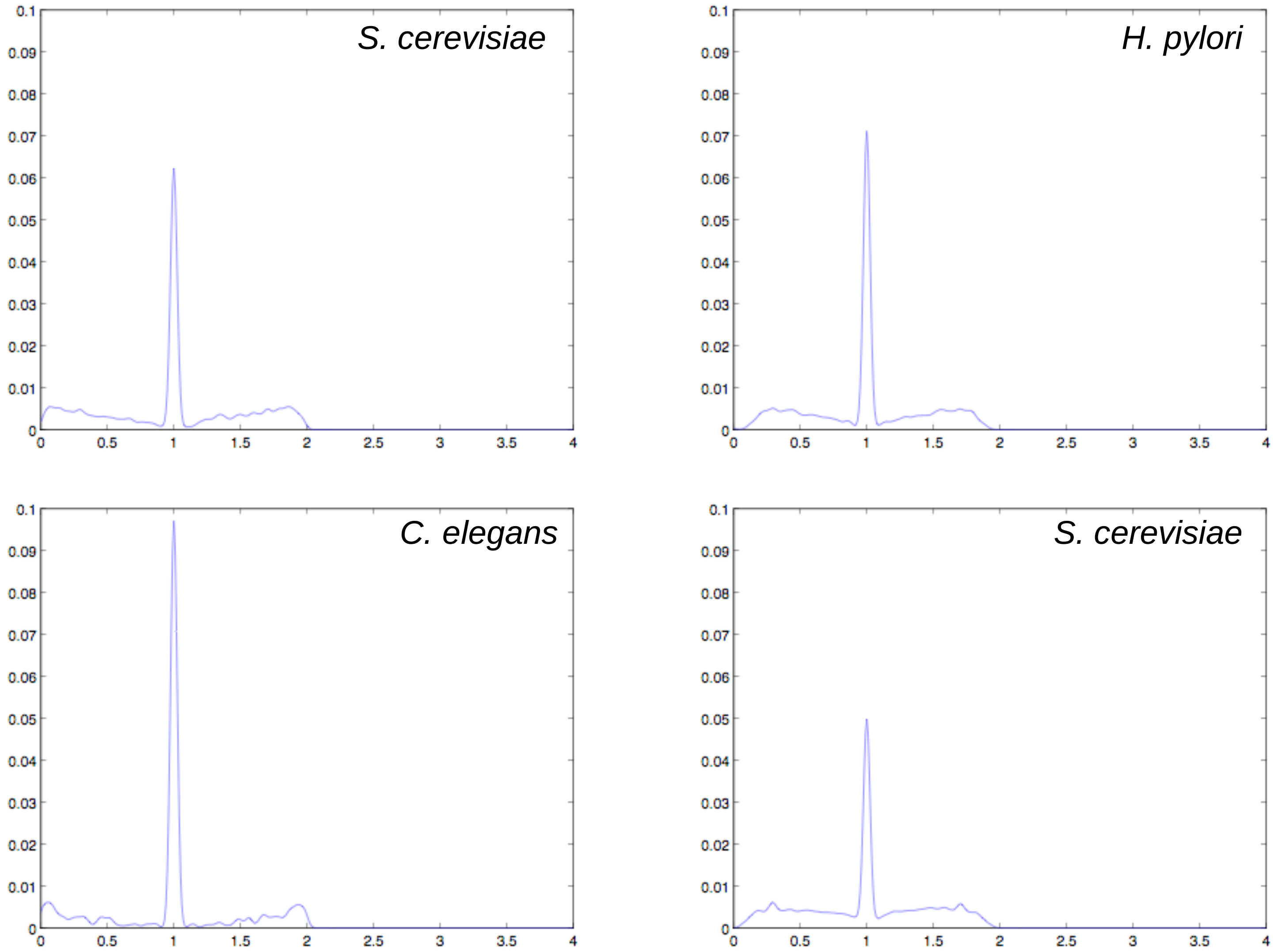


Overview of the talk

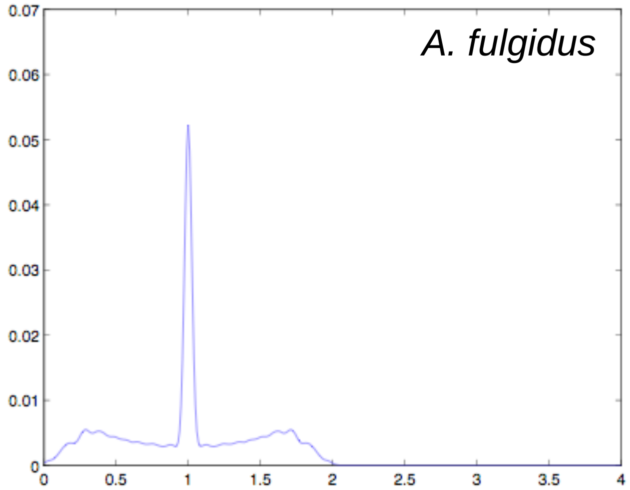
- Introduction
- Graph Laplacian operator
- Eigenfunction and graph structure
- Visualization techniques
- **Qualitative classification of networks**
- Conclusion and remarks

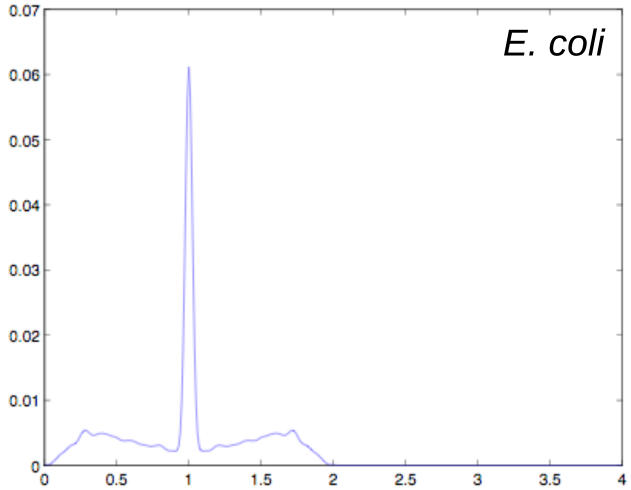
Type I

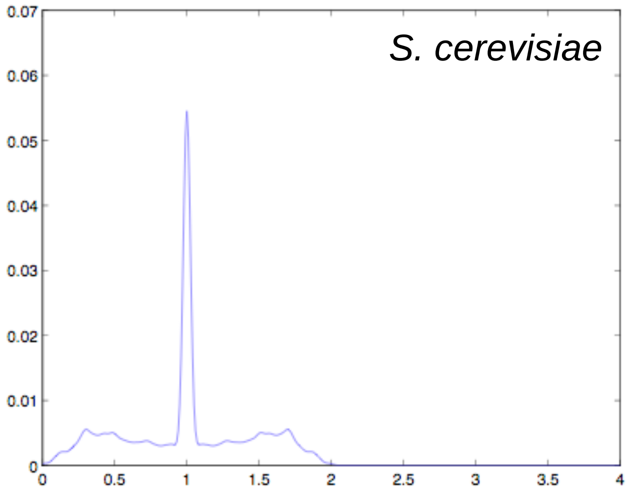
Protein-protein interaction networks



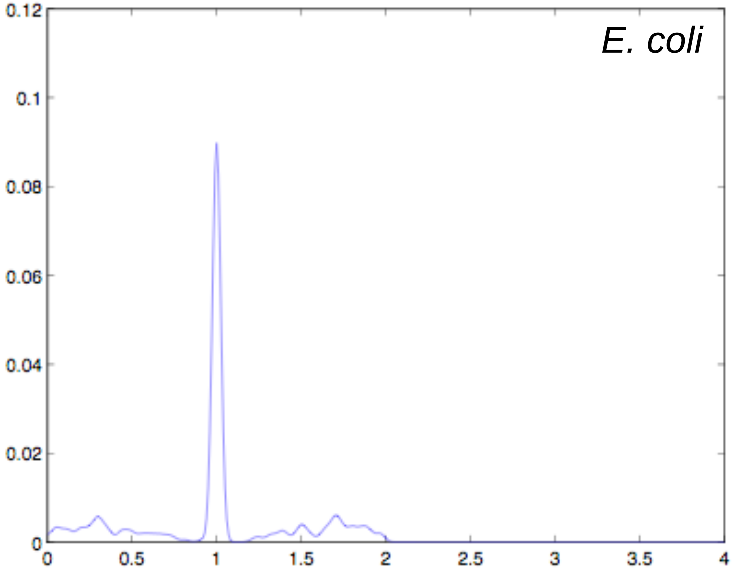
Metabolic networks

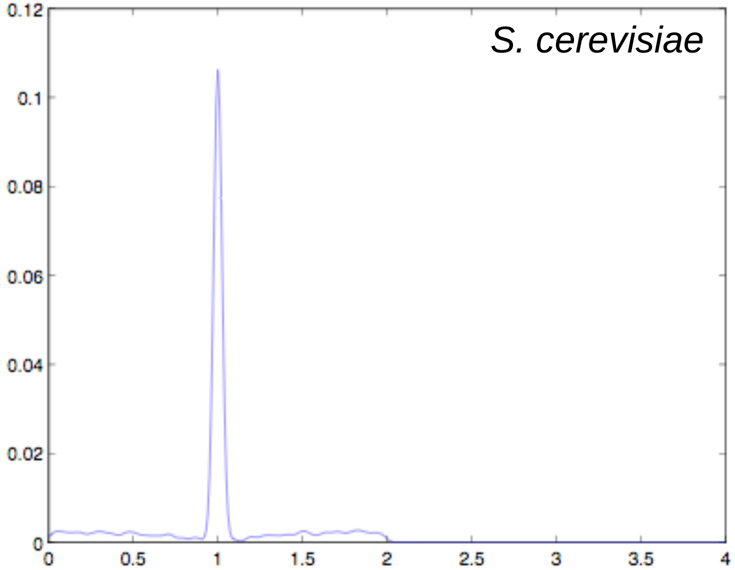




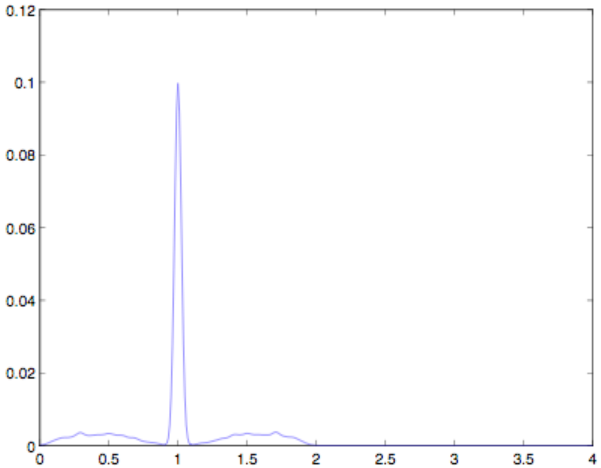


Transcription networks

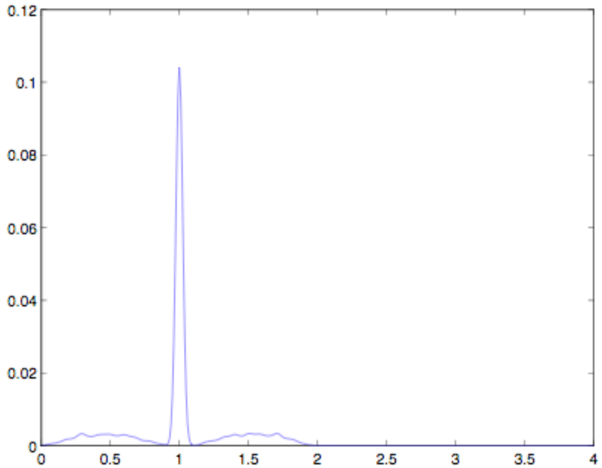




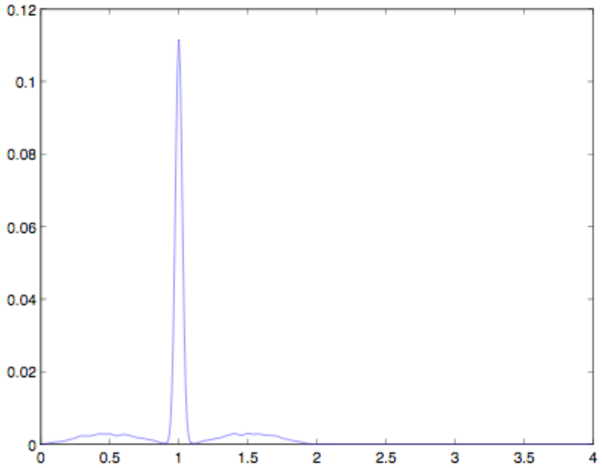
Autonomous Systems topology of the Internet



AS graph of 1997/11/08

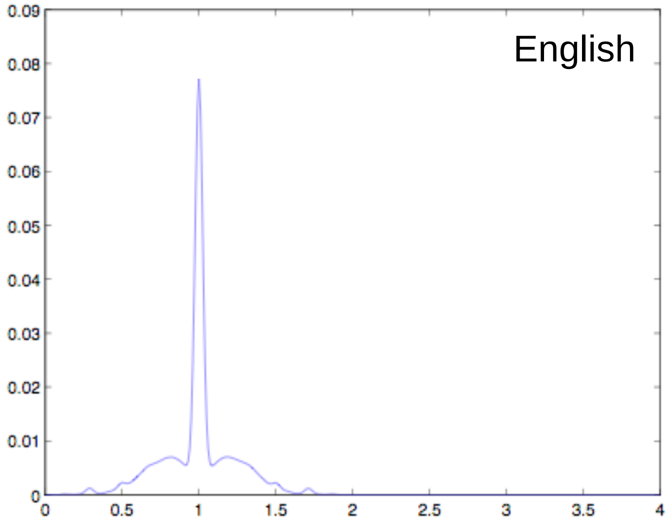


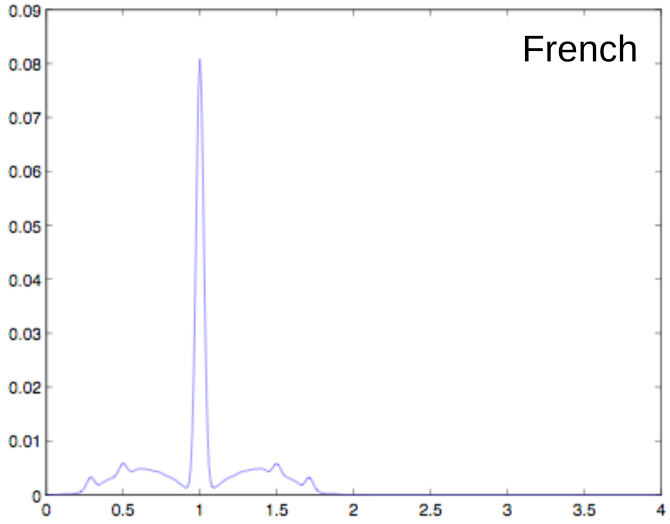
AS graph of 1999/07/02

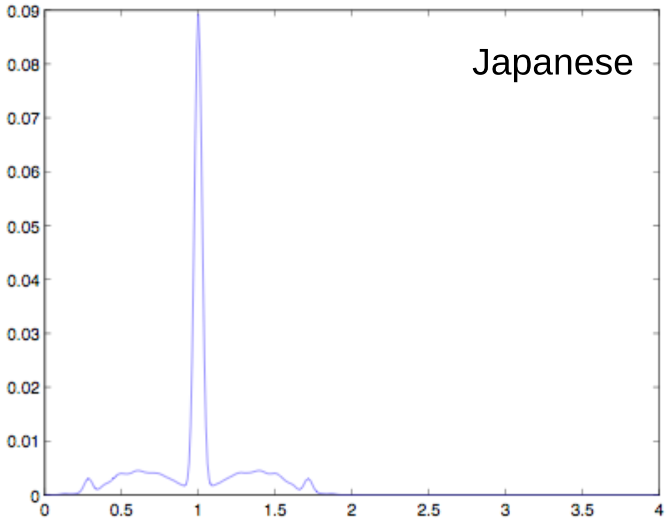


AS graph of 2001/03/16

Word-adjacency networks

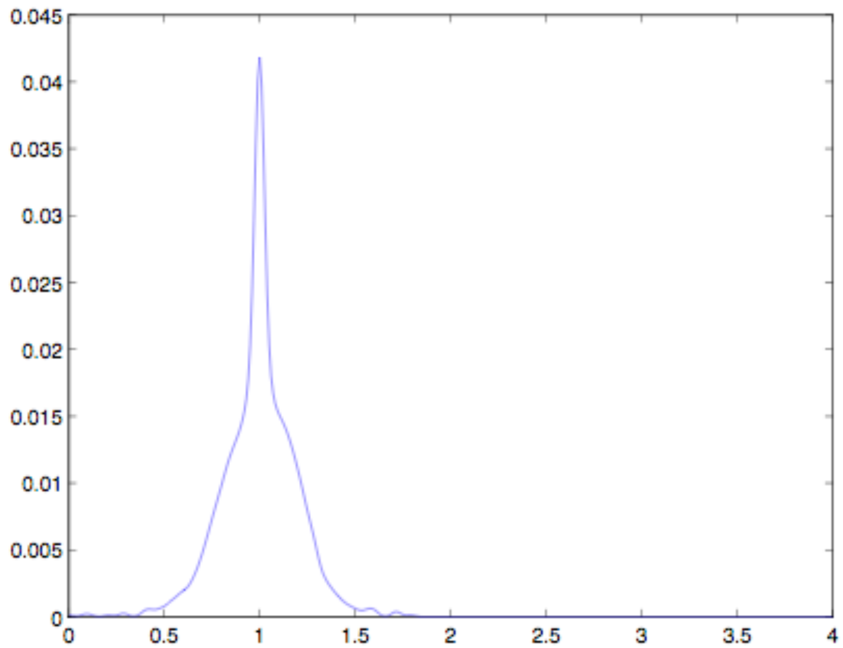




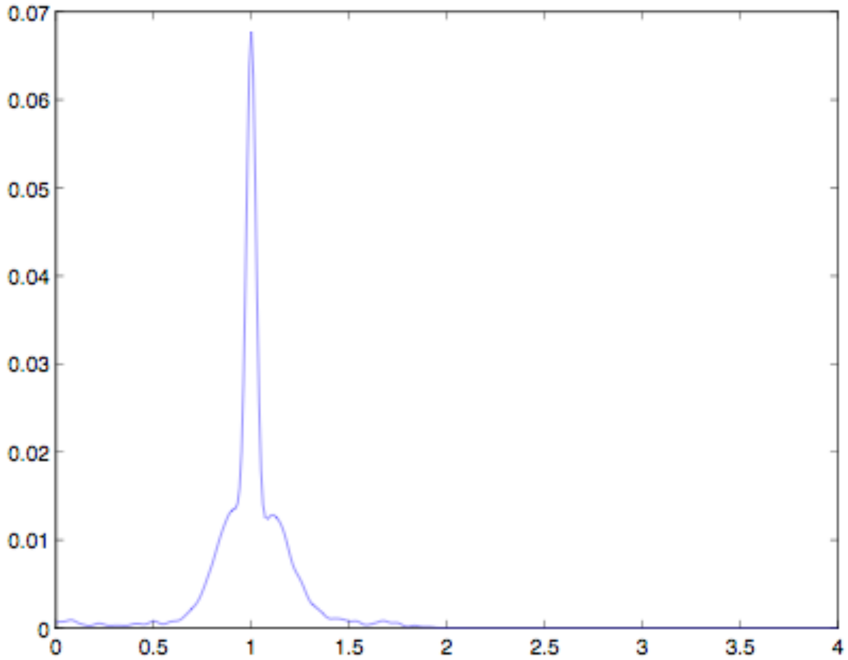


Type II

Network of hyperlinks between weblogs of US politics

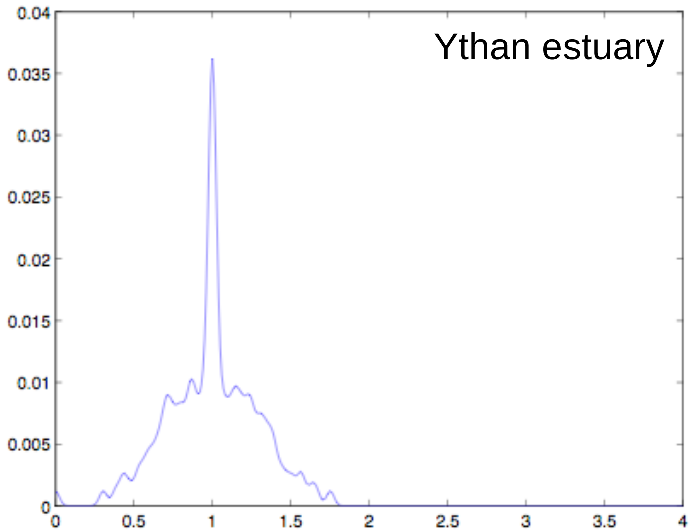


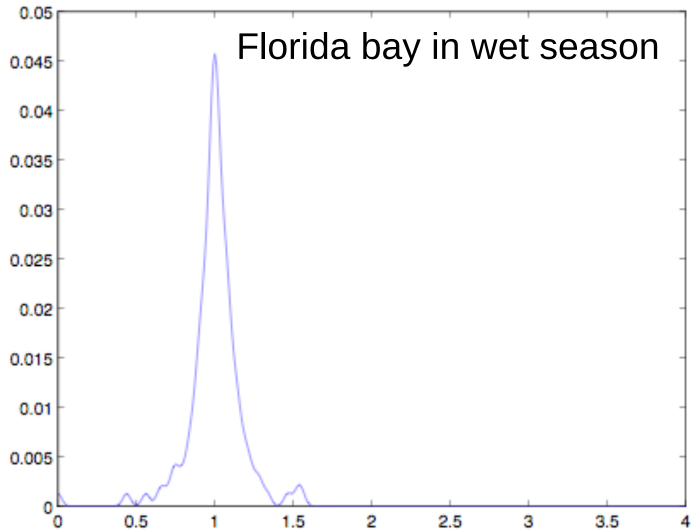
Network of conformation space

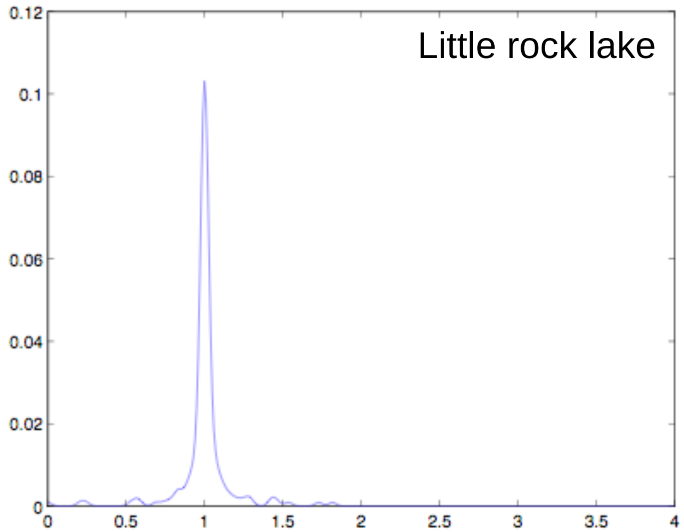


Only conformations that are visited at least 20 times during the simulation are considered in the building of the network.

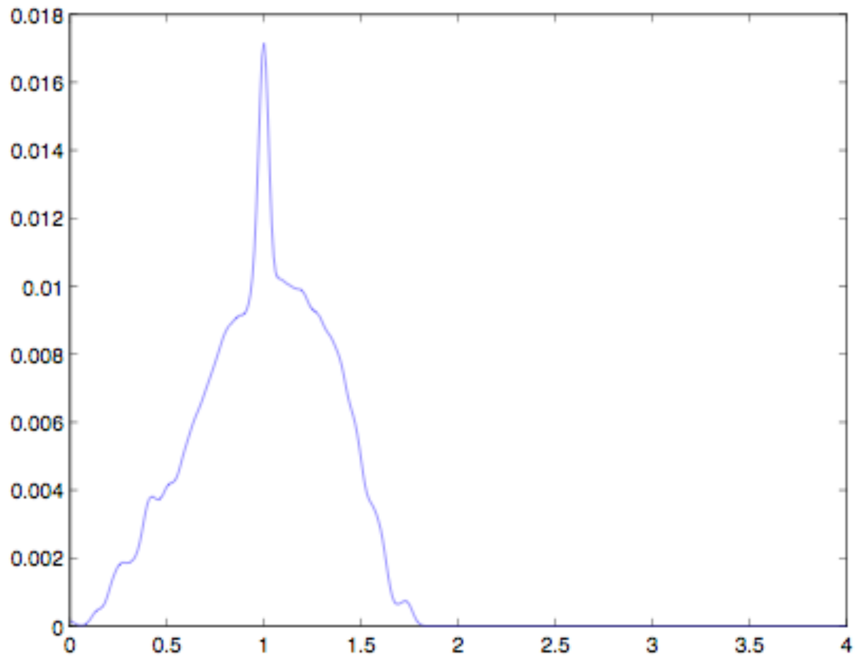
Food-web networks





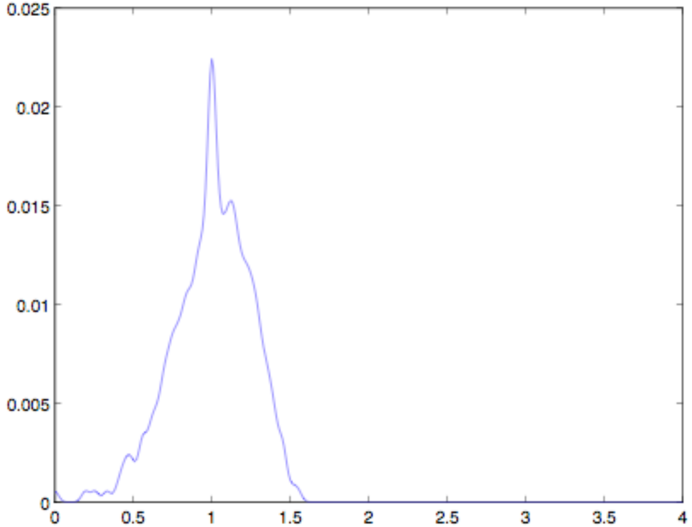


E-mail interchanges network

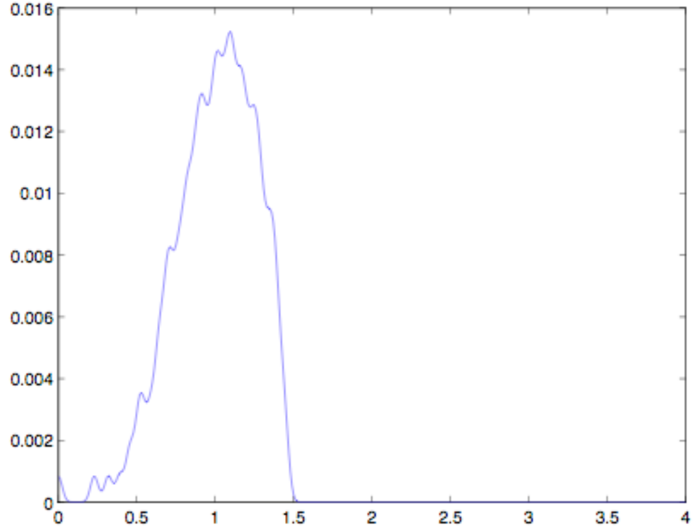


E-mail interchanges between members of the Univeristy Rovira i Virgili (Tarragona)

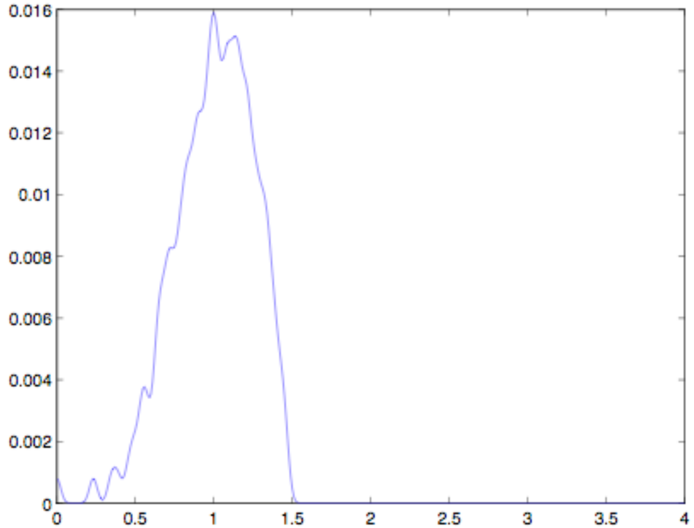
Neural networks



C. elegans



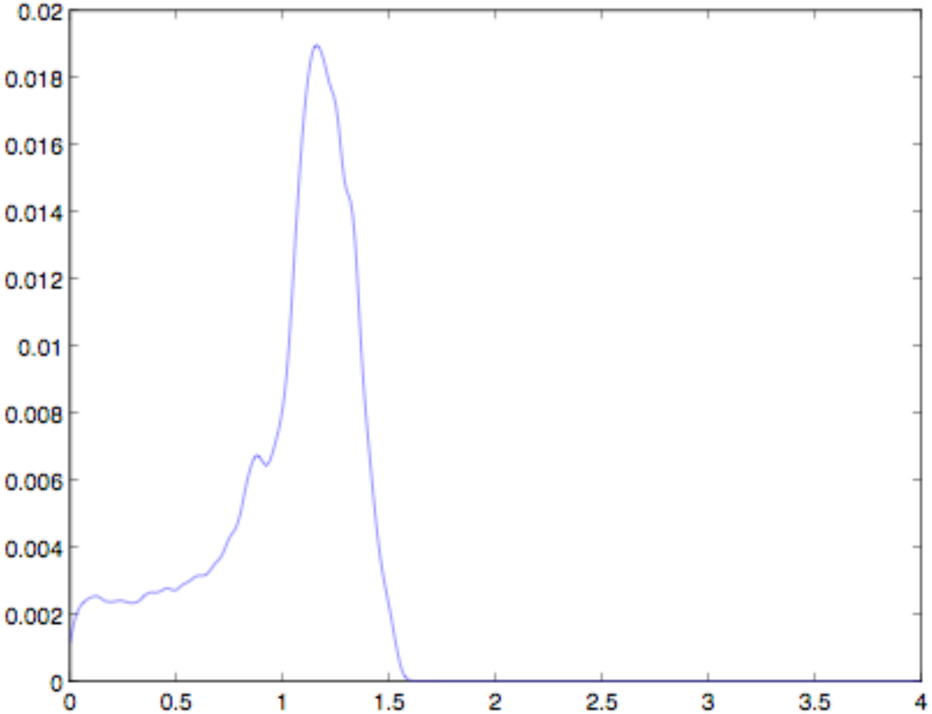
C. elegans (animal JSH, L4 male) in RVG region



C. elegans (animal N2U, adult hermaphrodite) in RVG region

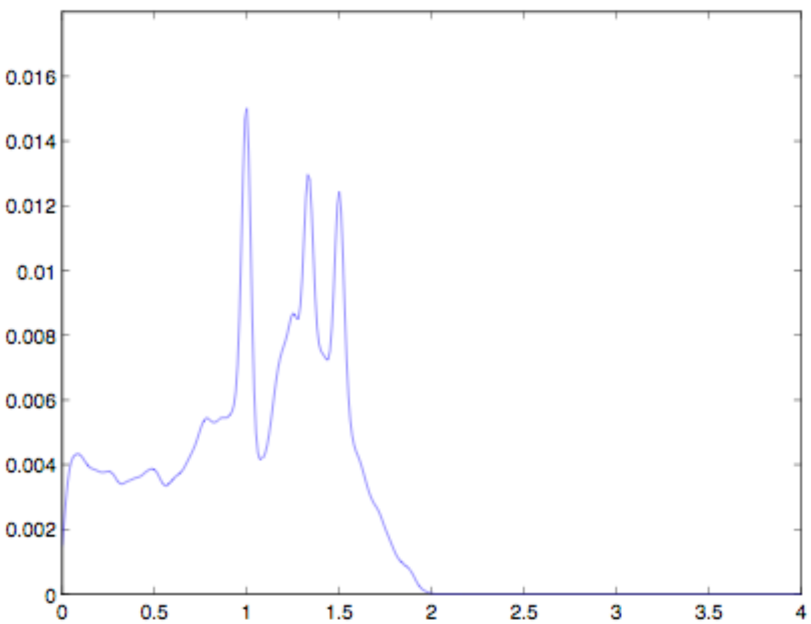
Type III

Power-grid network

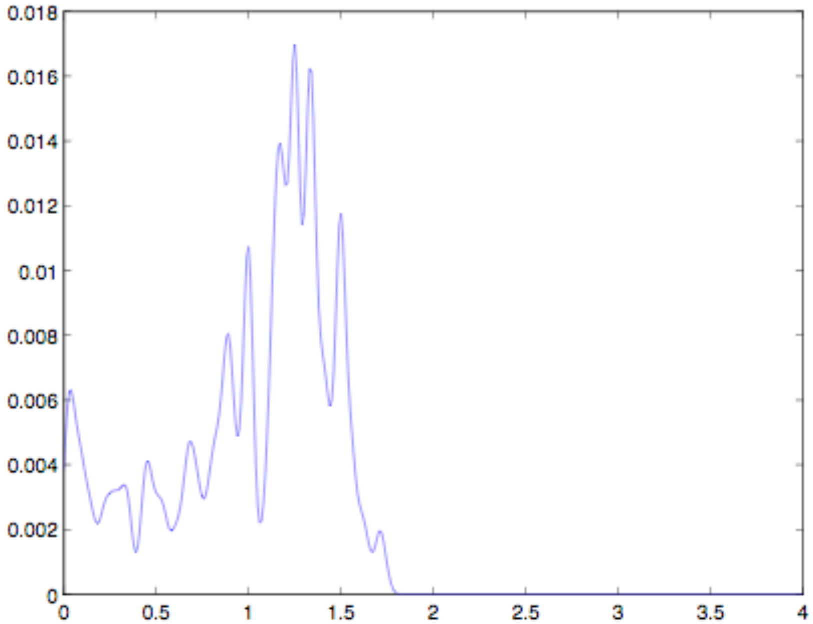


Topology of the Western States Power Grid of the United States

Networks of co-authorships

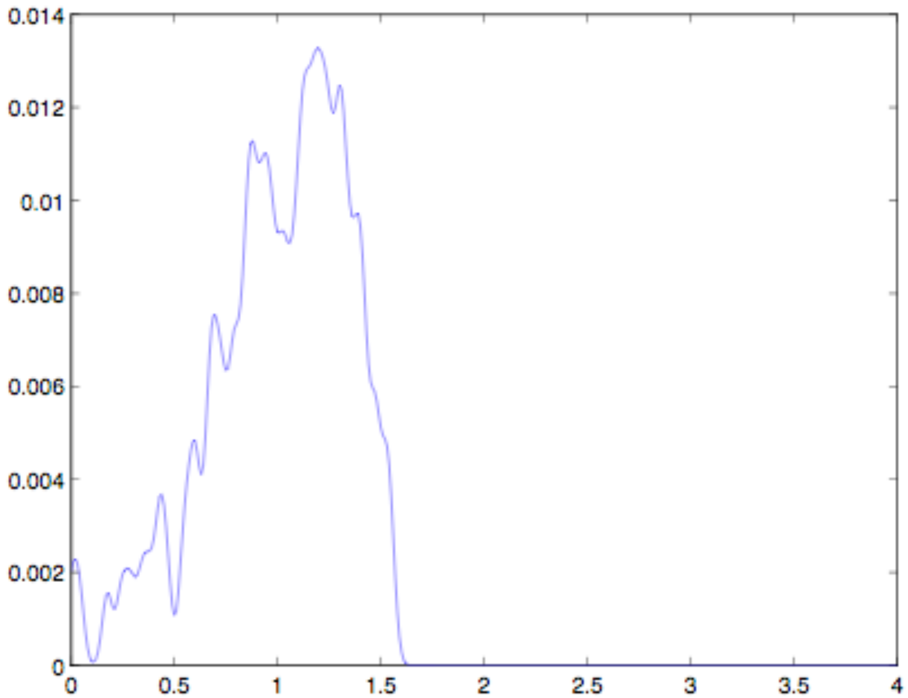


Scientists in high-energy theory



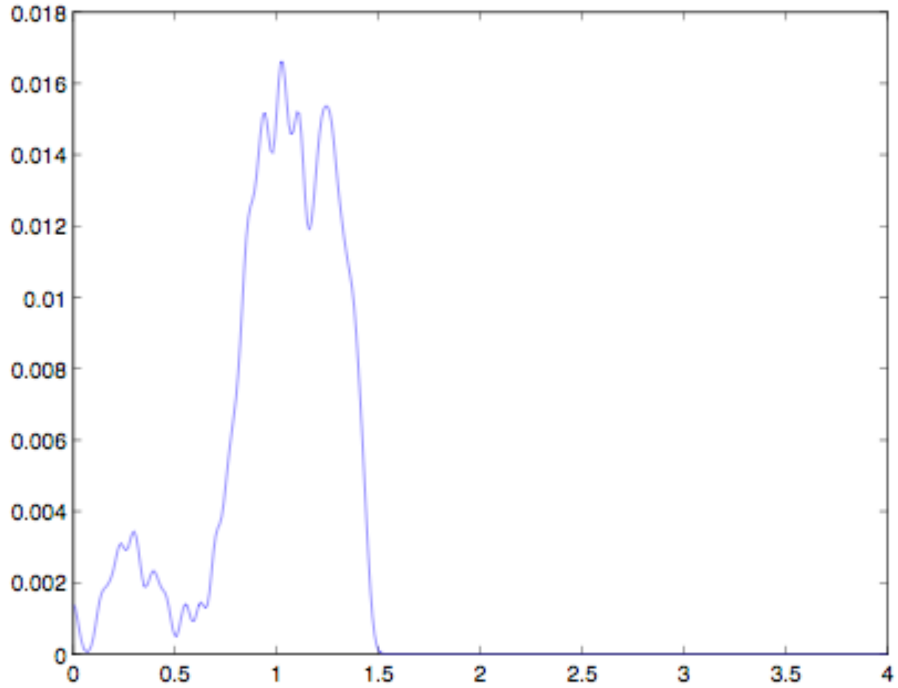
Scientists in network theory and experiment

Network of co-purchasing of books



Books about recent US politics sold by the online bookseller Amazon.com.

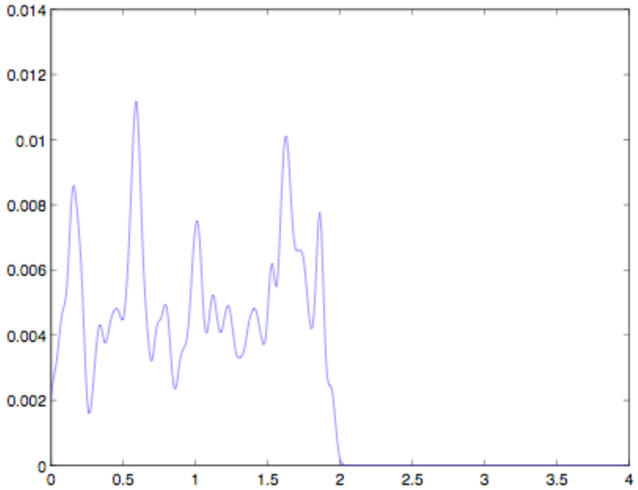
Network of US football games

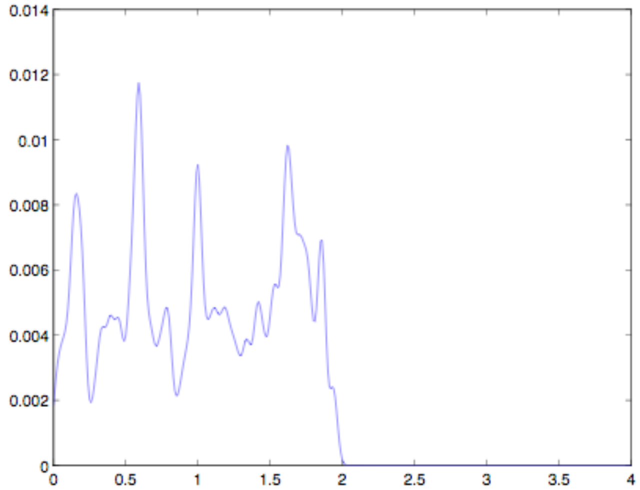


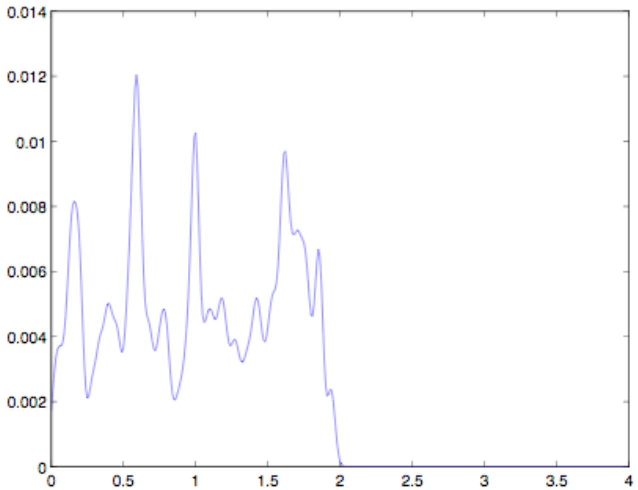
American football games between Division IA colleges during regular season Fall 2000

Type IV

Networks in electronic circuits





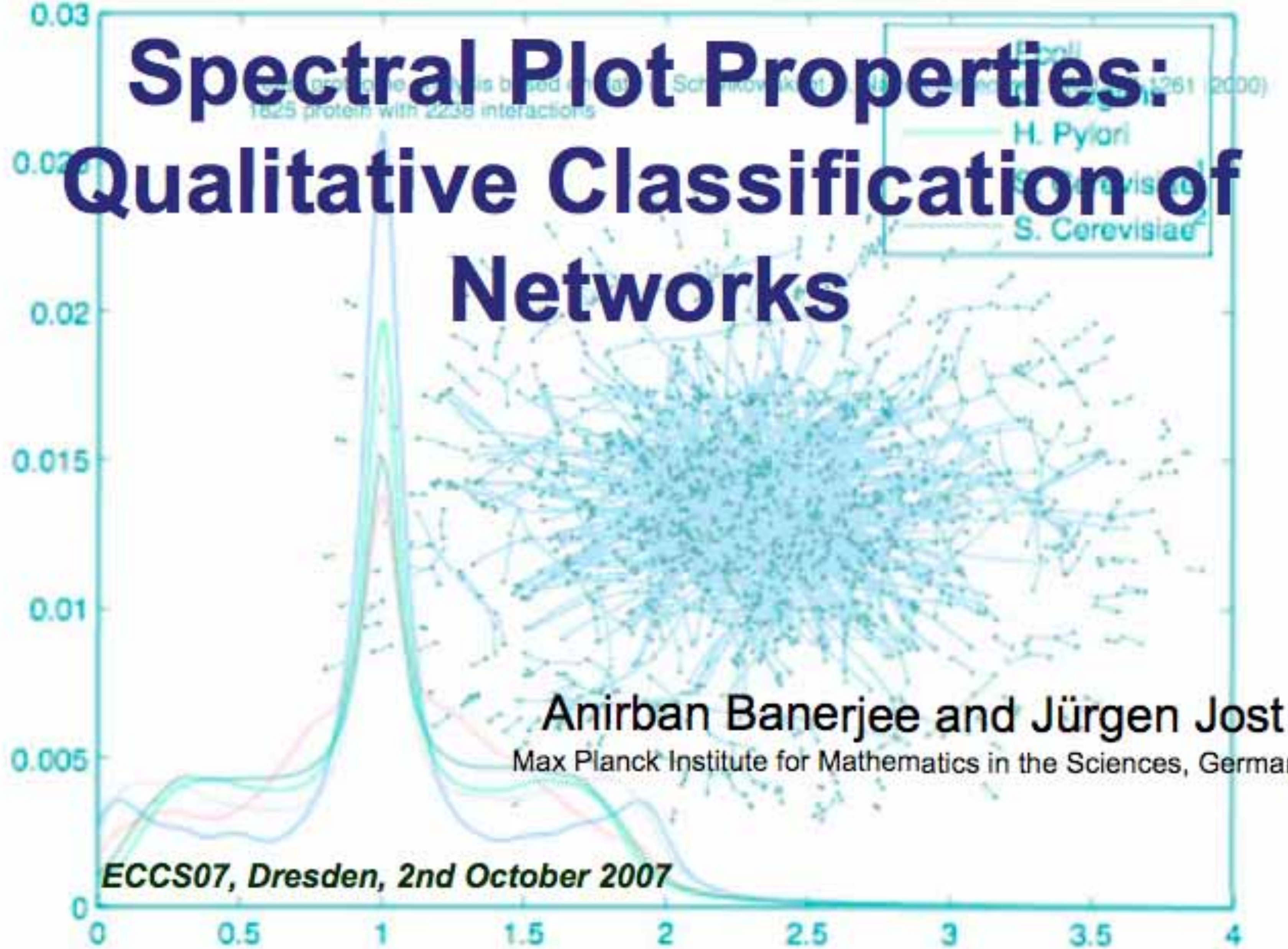


Overview of the talk

- Introduction
- Graph Laplacian operator
- Eigenfunction and graph structure
- Visualization techniques
- Qualitative classification of networks
- Conclusion and remarks

Thank you

Spectral Plot Properties: Qualitative Classification of Networks



Anirban Banerjee and Jürgen Jost
Max Planck Institute for Mathematics in the Sciences, Germany

ECCS07, Dresden, 2nd October 2007