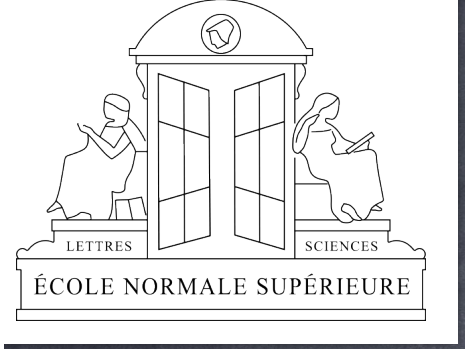


UNIVERSITE PIERRE & MARIE CURIE
LA SCIENCE A PARIS

The non-backtracking operator

Florent Krzakala
LPS, Ecole Normale Supérieure

in collaboration with
Paris: L. Zdeborova, A. Saade
Rome: A. Decelle
Würzburg: J. Reichardt
Santa Fe: C. Moore, P. Zhang
Berkeley: E. Mossel, A. Sly, J. Neeman

http://krzakala.org

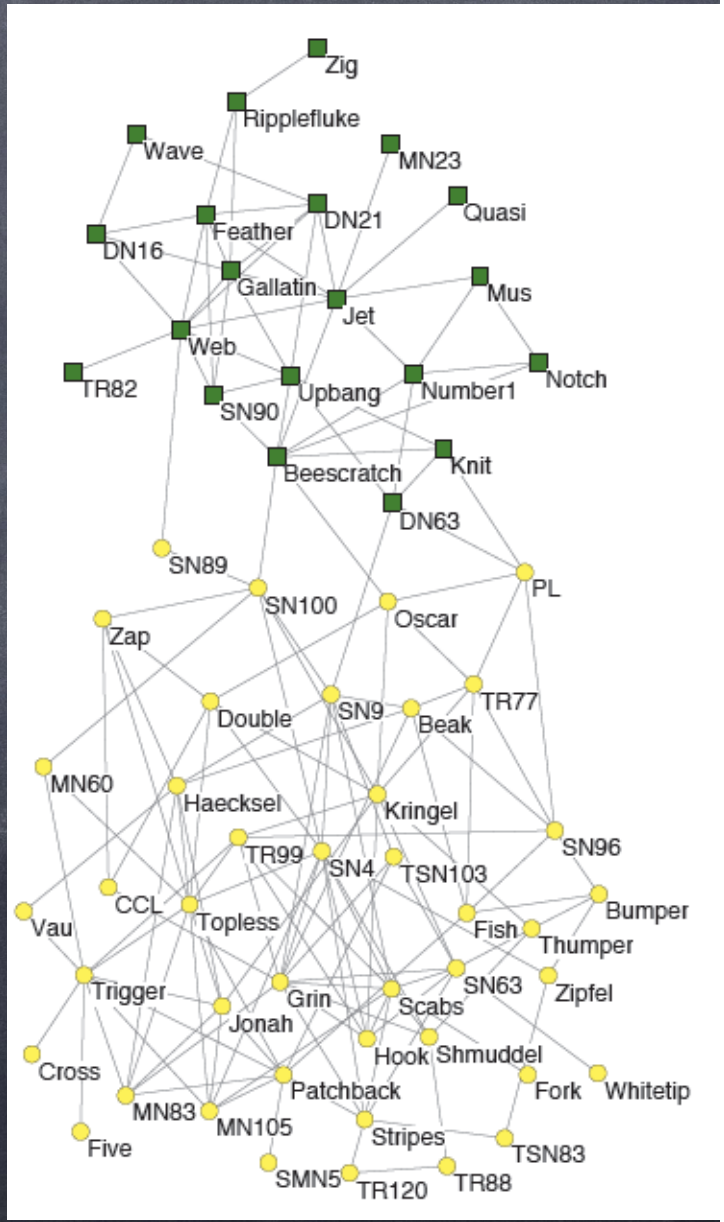
European Research Council

Community detection in network

- Stochastic block modeling (statistics)
- Data Clustering (machine learning)
- Graph-based machine learning (inference)
- Planted constraint satisfaction models

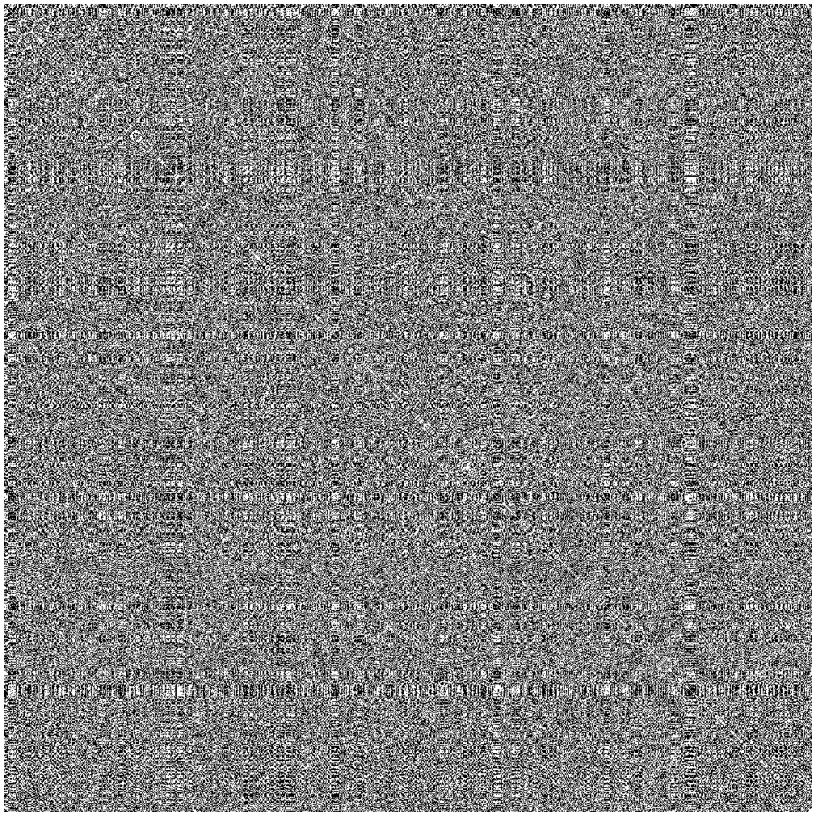
Given the graph,
find the labeling!

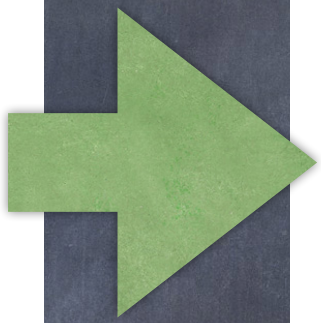
- Online communities
- Word adjacency networks
- Food webs
- Metabolic networks
- Protein-protein interaction networks
- Financial market (sectors)
- ...

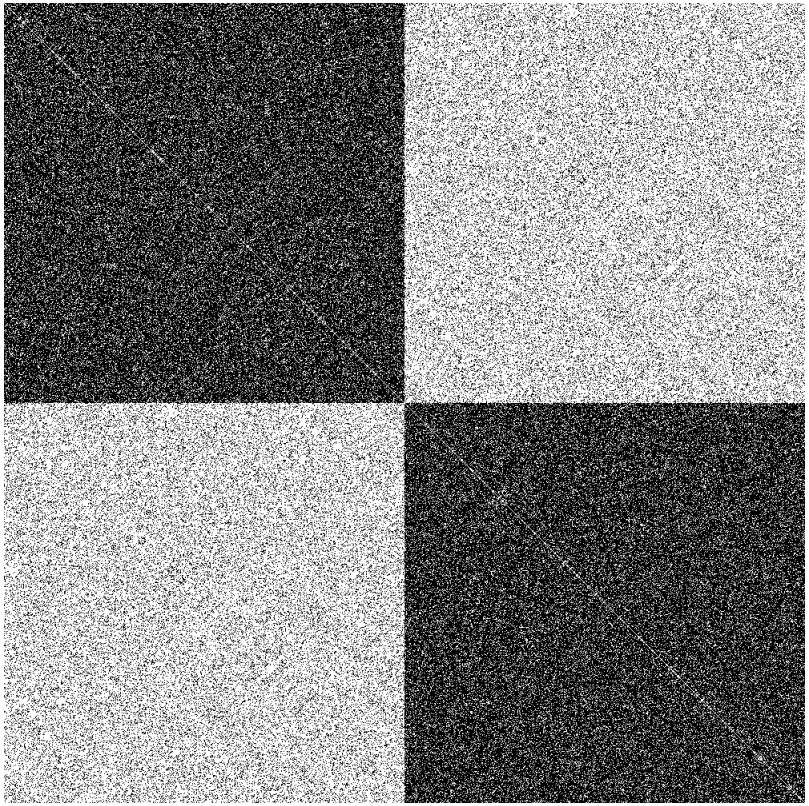


This is a hard problem...

Adjacency matrix



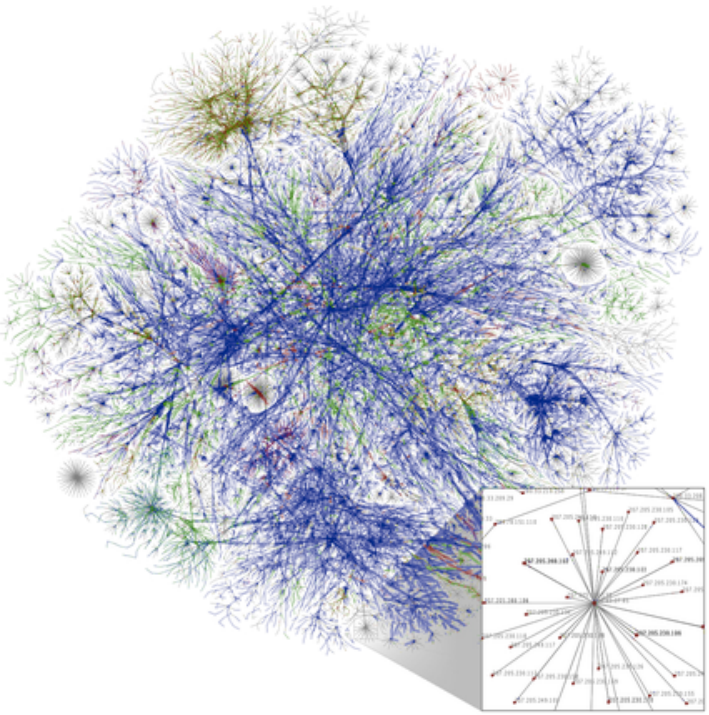




What you have:
The data

What you want:
The structure

This is a hard problem...



Karate club '77 The internet Protein network

... and it is getting harder!

Algorithms for clustering

- Hundreds of different methods - for a review e.g.

  S. Fortunato, Physics Reports, 2010.

- Spectral clustering - the state of art clustering method in machine learning. Associate a matrix to the network (adjacency, random walk, Laplacian, etc.), compute its top eigenvalues. The corresponding eigenvectors encode the clusters in a "visible" way.

This talk

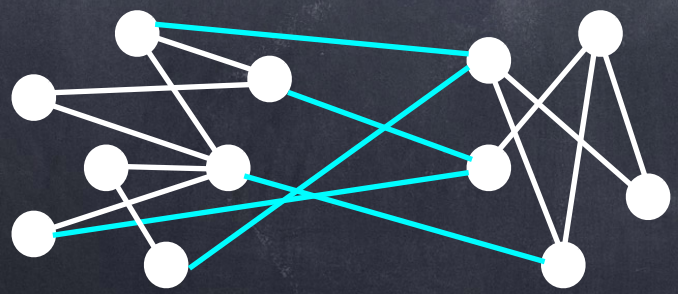
- Inference using the stochastic block model
- Spectral algorithm(s)

Stochastic block model

Generate a random network as follows:

- q groups, N nodes
- n_a proportion of nodes in group $a = 1, \ldots, q$
- $p_{ab} = \dfrac{c_{ab}}{N}$ probability that an edge present between node from group a and another from group b

$$n_1 = 7/12 \qquad n_2 = 5/12$$



$$p_{11} = p_{22} = 0.39$$

$$p_{12} = p_{21} = 0.14$$

Optimal inference in SBM

(when parameters of the model known)

$$P(\{q_i\}) = \prod_{i=1}^{N} n_{q_i} \qquad P(A_{ij}|\{q_i\}) = \prod_{i \neq j} p_{q_i q_j}^{A_{ij}} (1 - p_{q_i q_j})^{1 - A_{ij}}$$

$$P(\{q_i\}|A_{ij}) = \frac{1}{Z} P(\{q_i\}) P(A_{ij}|\{q_i\})$$

n_a proportion of nodes in group a

p_{ab} proportion of edge between groups a and b

$q_i \in \{1, \ldots, q\}$ an assignment of "colors"

Includes ALL available information about the signal

Optimal algorithm in SBM

(for a statistician)

Posterior probability distribution

$$P(\{s_i\}|A_{ij}) = \frac{1}{Z} P(\{s_i\}) P(A_{ij}|\{s_i\})$$

Marginal probabilities

$$\mu(s_i) \equiv \sum_{\{s_j\}_{j\neq i}} P(\{s_j\}_{j\neq i}, s_i|A_{ij})$$

Maximize number of correctly assigned nodes

$$s_i^* = \mathrm{argmax}_{s_i} \mu(s_i)$$

Optimal algorithm in SBM

(for a statistical physicist)

Potts spin variables $s_i \in \{1, \dots, q\}$

Boltzmann measure $P(\{s_i\}) = \frac{1}{Z} e^{-\mathcal{H}(\{s_i\})}$

Hamiltonian: Potts glass in field on the Nishimori line

$$\mathcal{H}(\{s_i\}) = -\sum_{i=1}^{N} \log n_{s_i} - \sum_{i \neq j} \left[A_{ij} \log p_{s_i s_j} + (1 - A_{ij}) \log \left(1 - p_{s_i s_j}\right) \right]$$

magnetic field

pair-wise interactions

Local magnetization $s_i^* = \mathrm{argmax}_{s_i} \mu(s_i)$

How to compute marginals and Z?

- A #P-hard problem in general
- MCMC: (Monte Carlo Markov chain, Gibbs sampling): generic, but potentially a (very) large equilibration time
- Variational mean field: convergence problem and (sometimes) a crude approximation.
- Belief propagation/cavity method: fast, hard to control, generally better than mean field, exact for large networks generated by the model.

Asymptotic analysis of the stochastic block model for modular networks and its algorithmic applications

Phys. Rev. E **84**, 066106 – Published 12 December 2011

Aurelien Decelle, Florent Krzakala, Cristopher Moore, and Lenka Zdeborová

Inference and Phase Transitions in the Detection of Modules in Sparse Networks

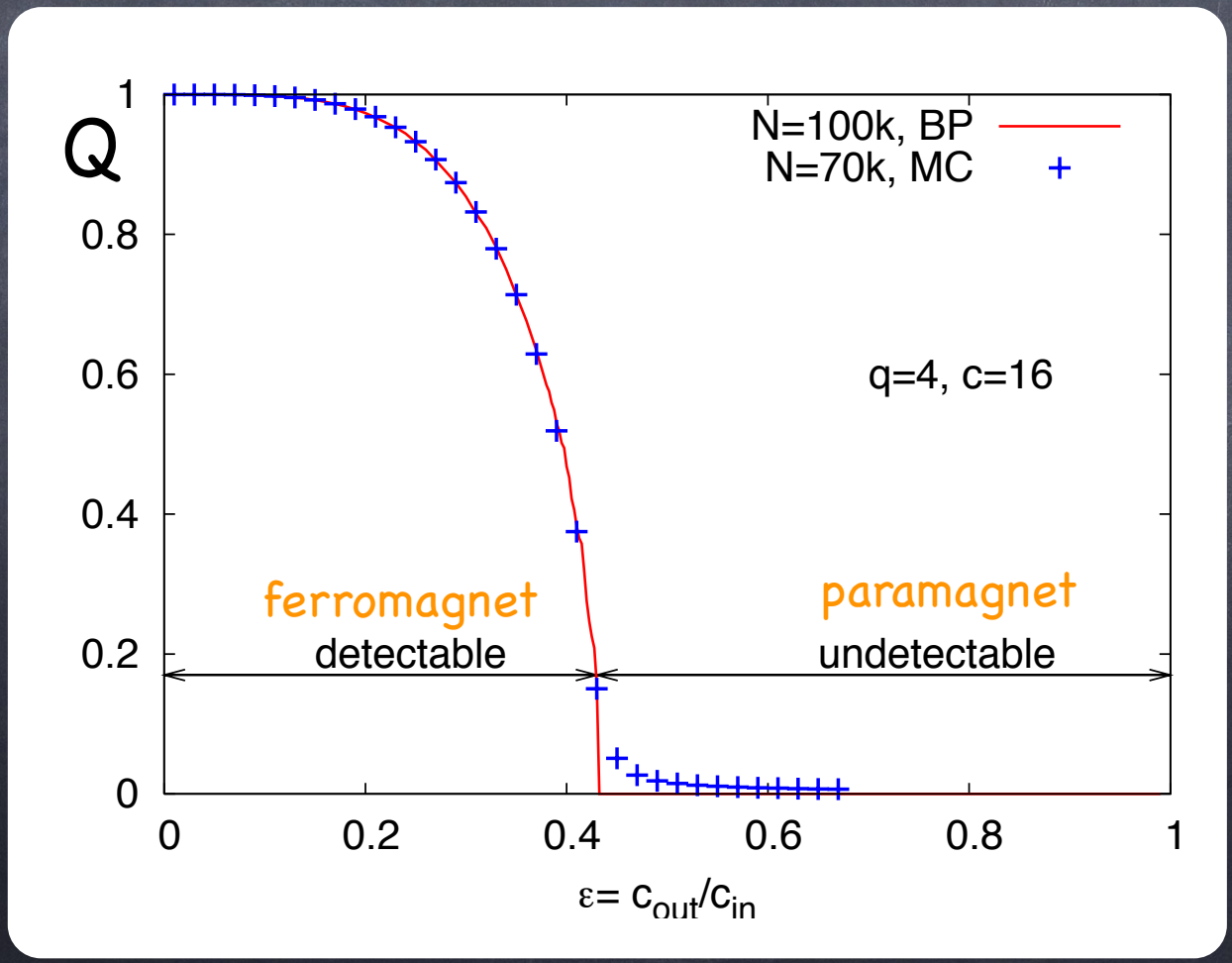
Phys. Rev. Lett. **107**, 065701 – Published 2 August 2011

Aurelien Decelle, Florent Krzakala, Cristopher Moore, and Lenka Zdeborová

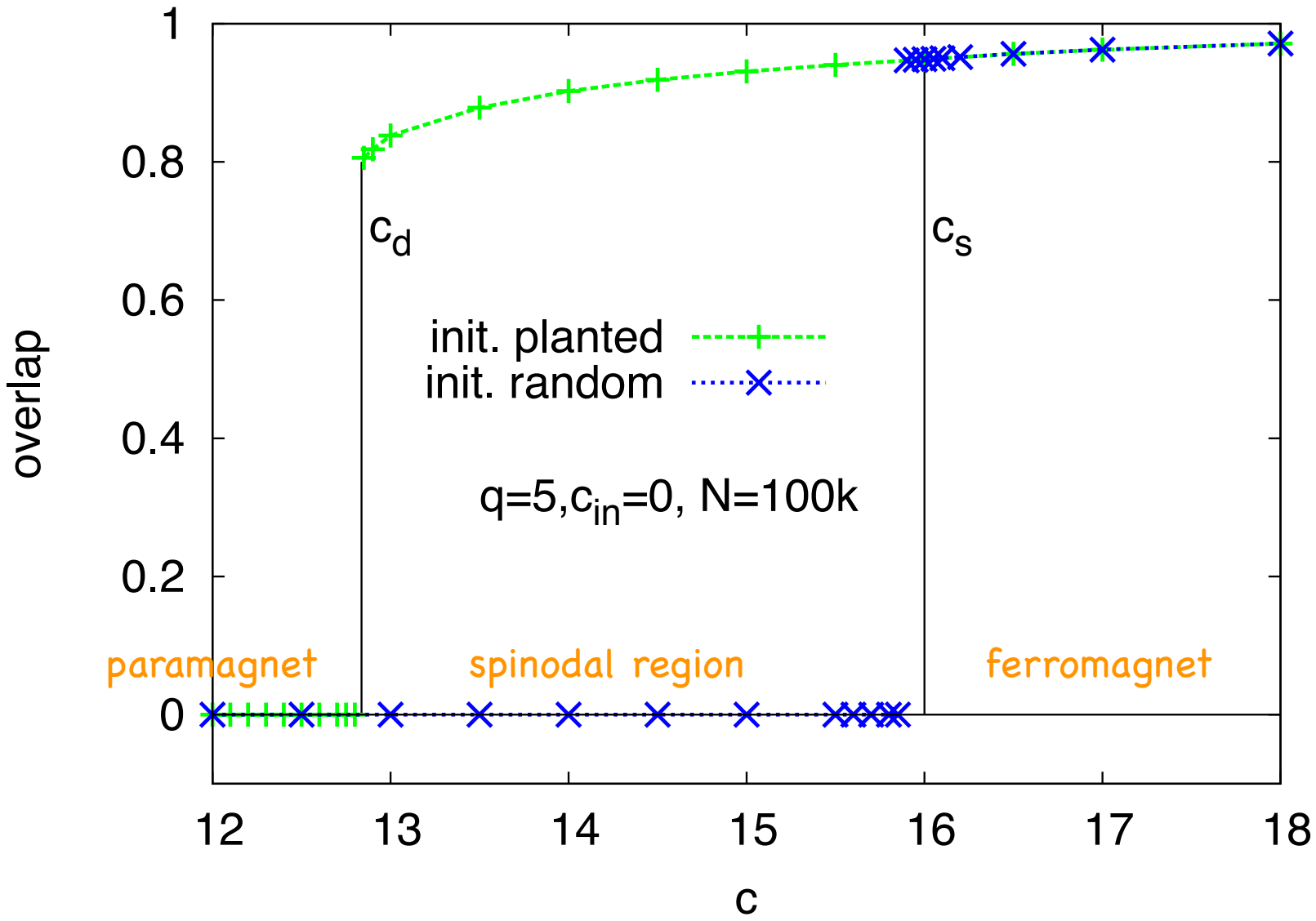
Results (in a nutshell)

$$Q = \max_{\pi} \frac{\sum_{i=1}^{N} \delta_{\pi(t_i),s_i^*}/N - 1/q}{1 - 1/q}$$

Overlap of the optimal estimation with the true labeling



Results (in a nutshell)

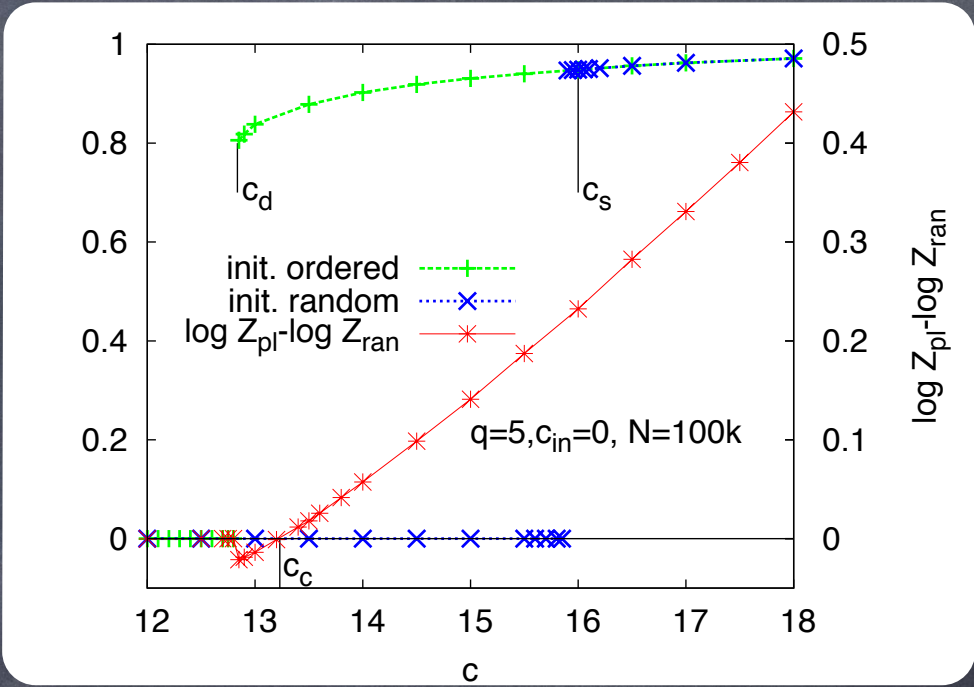


dis-assortative network 5 colors

Algorithmic consequences

$c < c_c$

impossible



$|c_{\rm in} - c_{\rm out}| = q\sqrt{c_{\rm s}}$

$c_c < c < c_s$

possible but hard

$c_s < c$

easy

Rigorous results for q=2

For q=2 we find that the information theoretic thresholds equals the algorithmic one

$$|c_{\text{in}} - c_{\text{out}}| = q\sqrt{c_{\text{s}}}$$

Proven!

Communutities are undetectable bellow c_s and easy to find beyond c_s

- The undetectable regime (Mossel, Neeman, Sly'12) : Planted graphs are essentially Erdos-Renyi random graphs in the undetectable regime.
- The detectable regime (Massoulie'13)

This talk

- Inference using the stochastic block model
- Spectral algorithm(s)

Spectral clustering

- Compute k largest eigenvalues and their eigenvectors for a matrix associated to the graph.
- Cluster these eigenvectors, e.g. using k-means.
  For 2 groups - signs of the 2nd eigenvector.

Matrices:

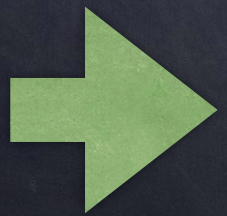
Adjacency	A_{ij}
Laplacian	$L_{ij} = d_i \delta_{i,j} - A_{ij}$
Random walk	$Q_{ij} = \frac{A_{ij}}{d_i}$
Modularity	$M_{ij} = A_{ij} - \frac{d_i d_j}{2M}$

Adjacency matrix for dense graphs

- Largest eigenvalue is the average degree, and the eigenvector is trivial.
- It exists an eigenvector correlated with the communities, with eigenvalue

$$\lambda_c = \frac{c_{\text{in}} - c_{\text{out}}}{2} + \frac{c_{\text{in}} + c_{\text{out}}}{c_{\text{in}} - c_{\text{out}}}$$

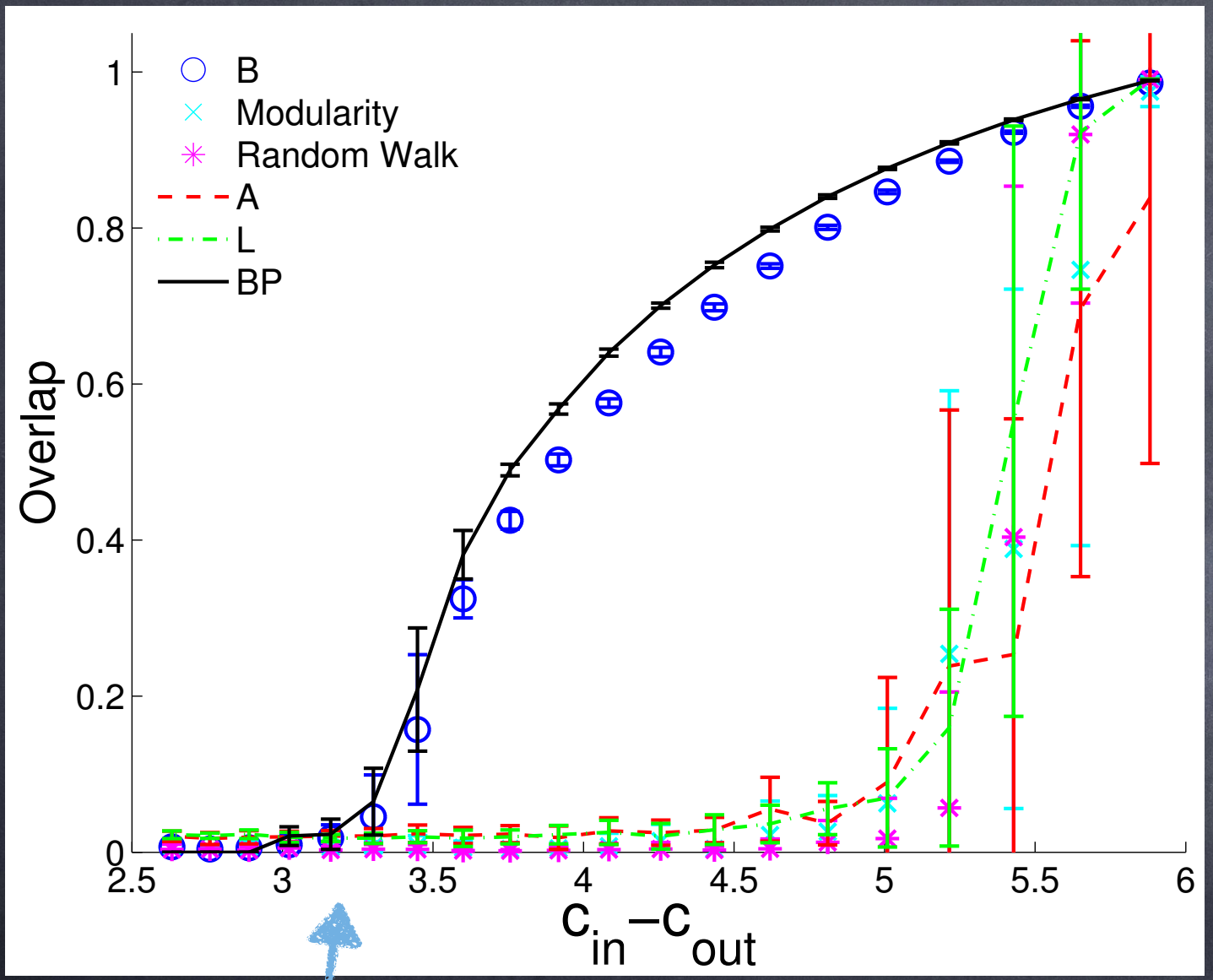
- The bulk follows Wigner's semi circle law: $P(\lambda) = \frac{1}{2\pi c}\sqrt{4c - \lambda^2}$
- If $|c_{\text{in}} - c_{\text{out}}| > 2\sqrt{c}$ the informative eigenvalue is out of the bulk



- Adjacency is optimal for dense graphs!
  If it fails, any thing else will...

Nadakuditi & Newman '12

Performance of spectral clustering



$q = 2$

$n_a = 1/2$

$c = 3$

$N = 10^5$

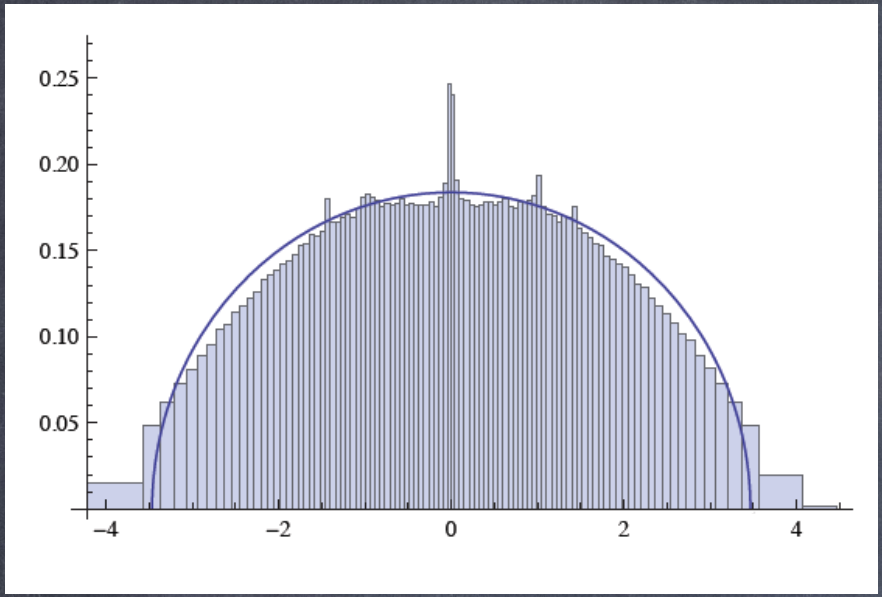
$c_{\rm in} = c_{aa}$

$c_{\rm out} = c_{a\neq b}$

$|c_{\rm in} - c_{\rm out}| = q\sqrt{c}$

$$Q = \max_{\pi} \frac{\sum_{i=1}^{N} \delta_{\pi(t_i),q_i^*}/N - 1/q}{1 - 1/q}$$

Why the suboptimality?

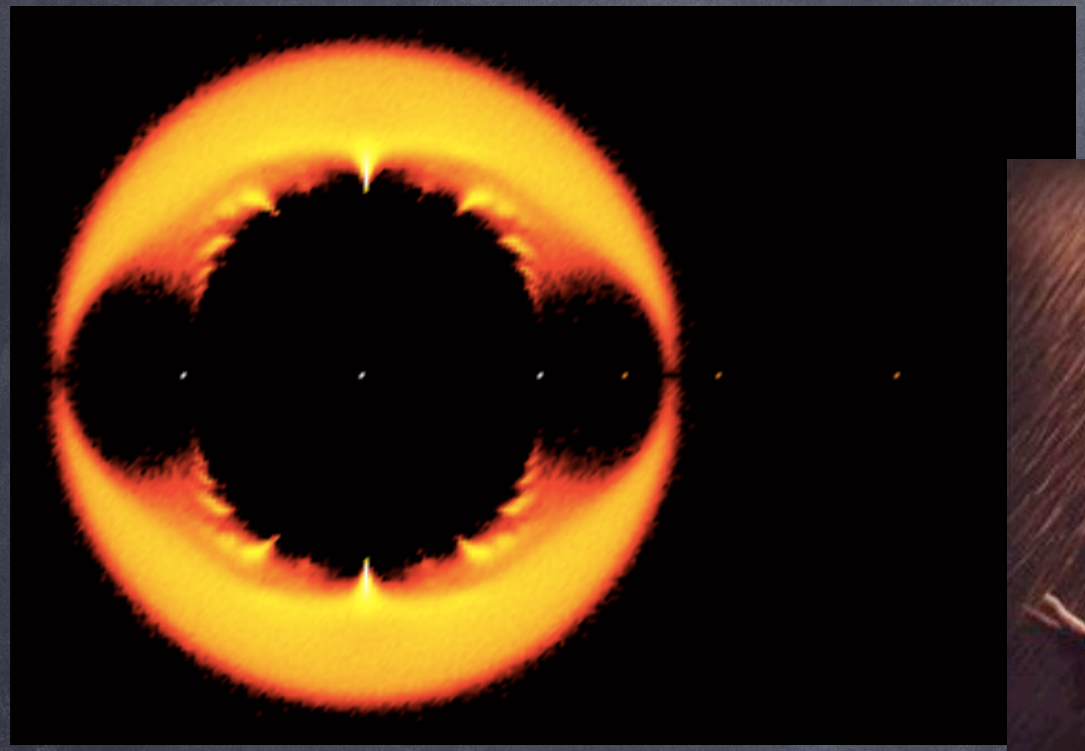


Wigner not good for sparse G

$$P(\lambda) = \frac{1}{2\pi c}\sqrt{4c - \lambda^2}$$

- The edge of the spectrum of sparse graphs is spoiled by nodes of large degree.
- ER graphs, largest degree ~ log(N)/log(log(N)).
- How to correct this? Remove largest degrees? Not good enough - loosing information.

Spectral Redemption



Can the optimal performance (phase transition) be matched by a spectral method?

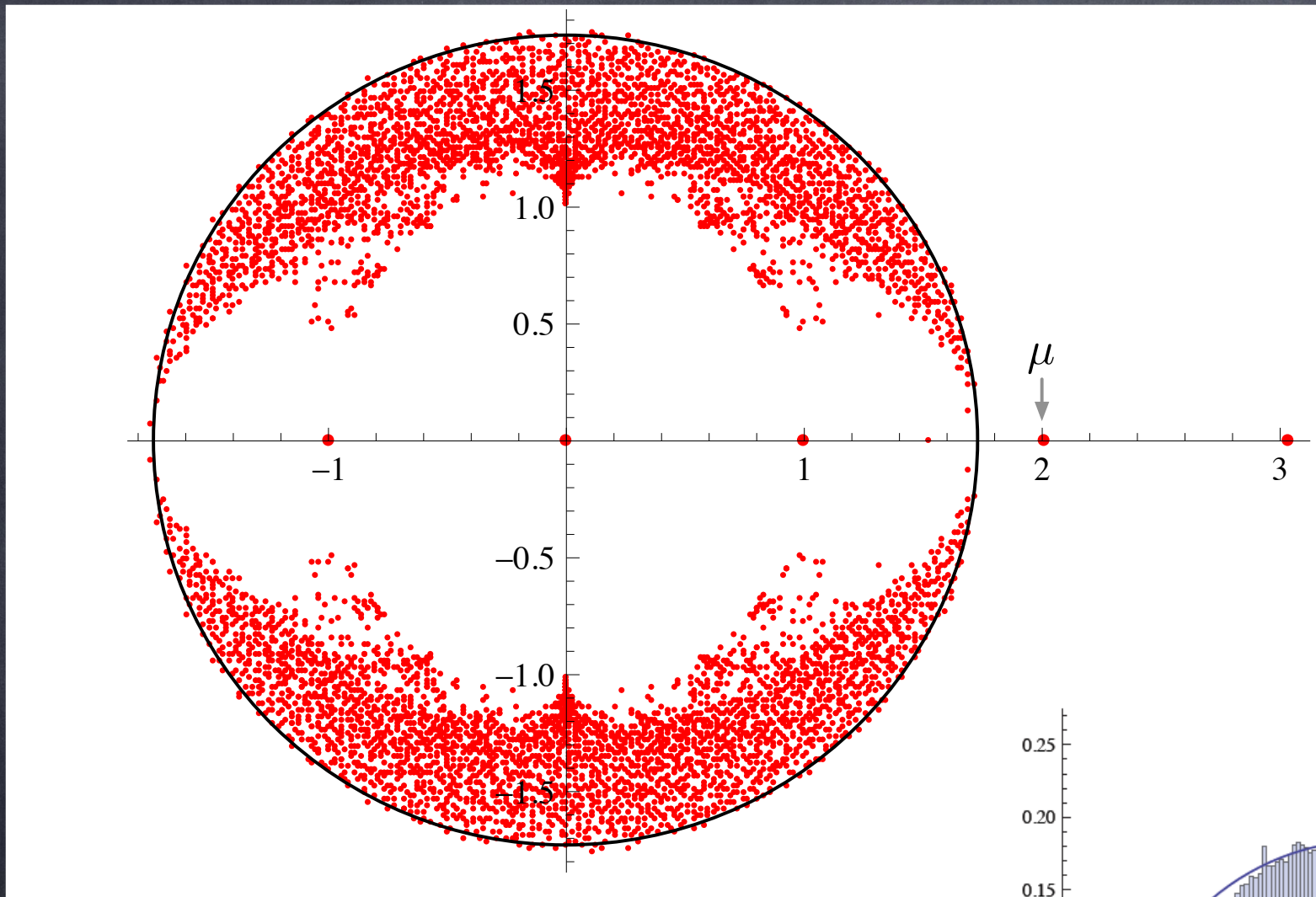
The non-backtracking matrix

- M edges, define matrix B on directed edges, i.e. 2M x 2M matrix as follows

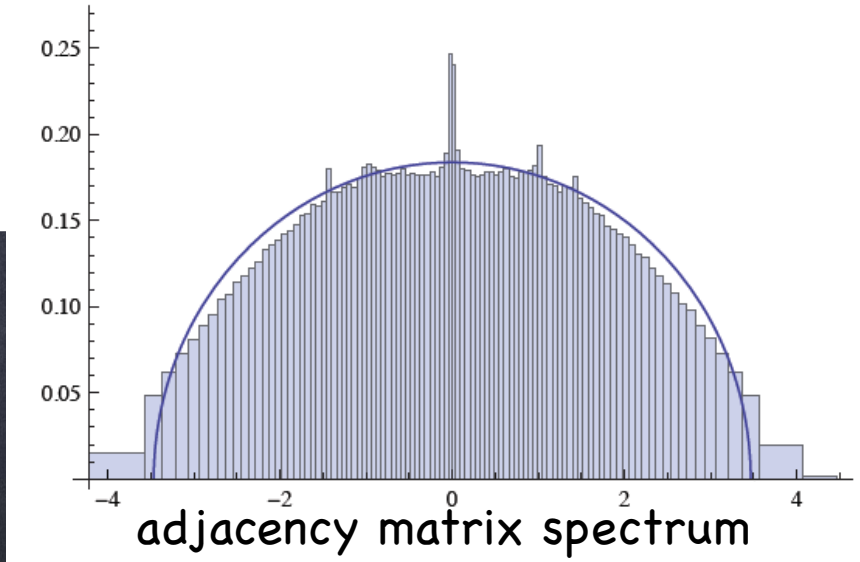
$$B_{i\to j,k\to l} = 1 \quad \text{if} \quad j = k, i \neq l$$
$$B_{i\to j,k\to l} = 0 \quad \text{otherwise}$$

- The Hashimoto ('89) edge adjacency operator used to evaluate the Ihara zeta function
- Directed walk similar to a message passing algorithm (i.e. Belief propagation)

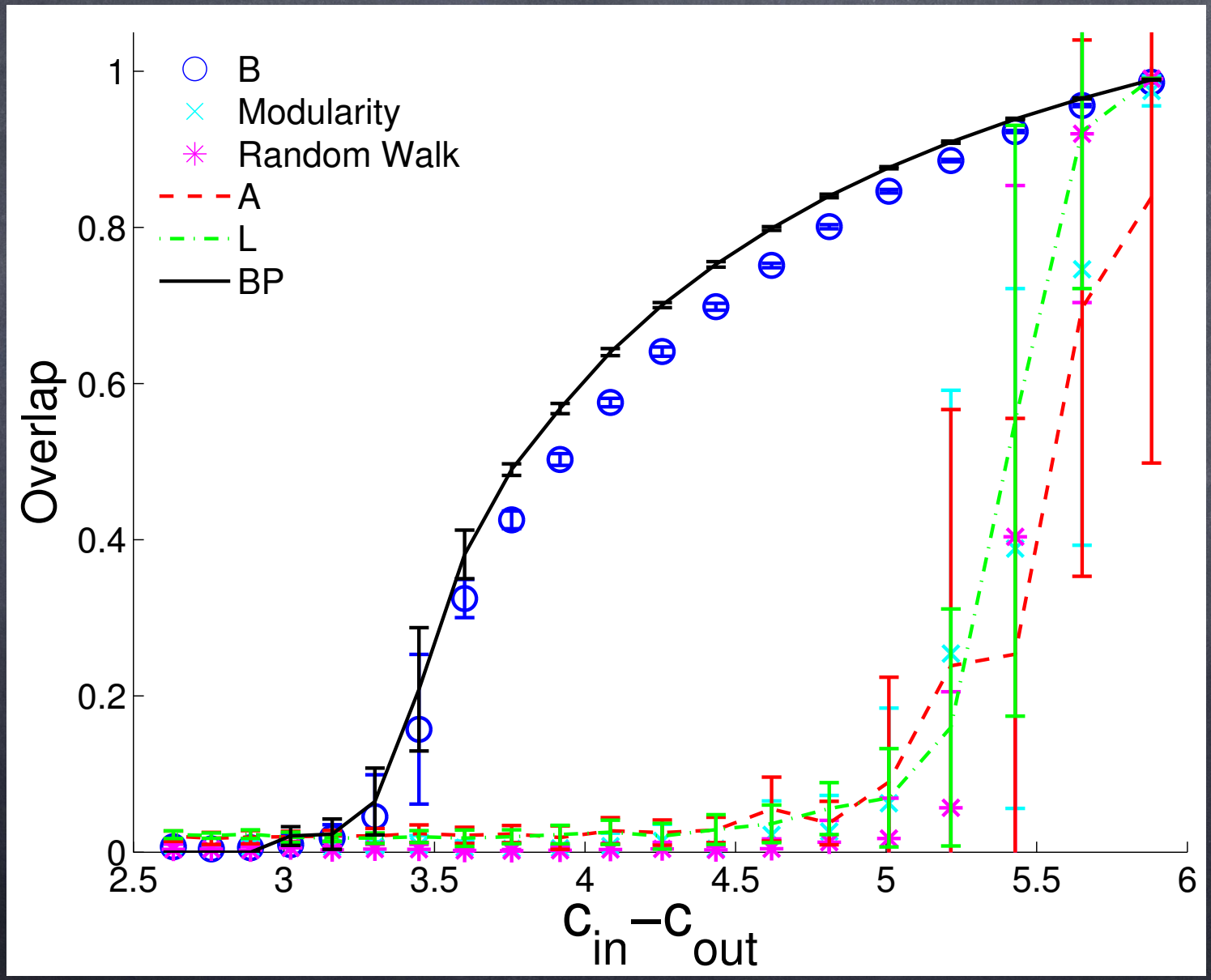
The non-backtracking matrix



Not a random sparse real non-symmetric matrix, elements correlated!



Performance of spectral clustering



Properties of the non-backtracking matrix

- Eigenvalues of B, and sums of eigenvector elements are given by a 2N x 2N matrix.

$$B' = \begin{pmatrix} 0 & D-1 \\ -1 & A \end{pmatrix} \qquad v_{\text{in}}^{j} = \sum_{i \in \partial j} v^{i \to j}$$

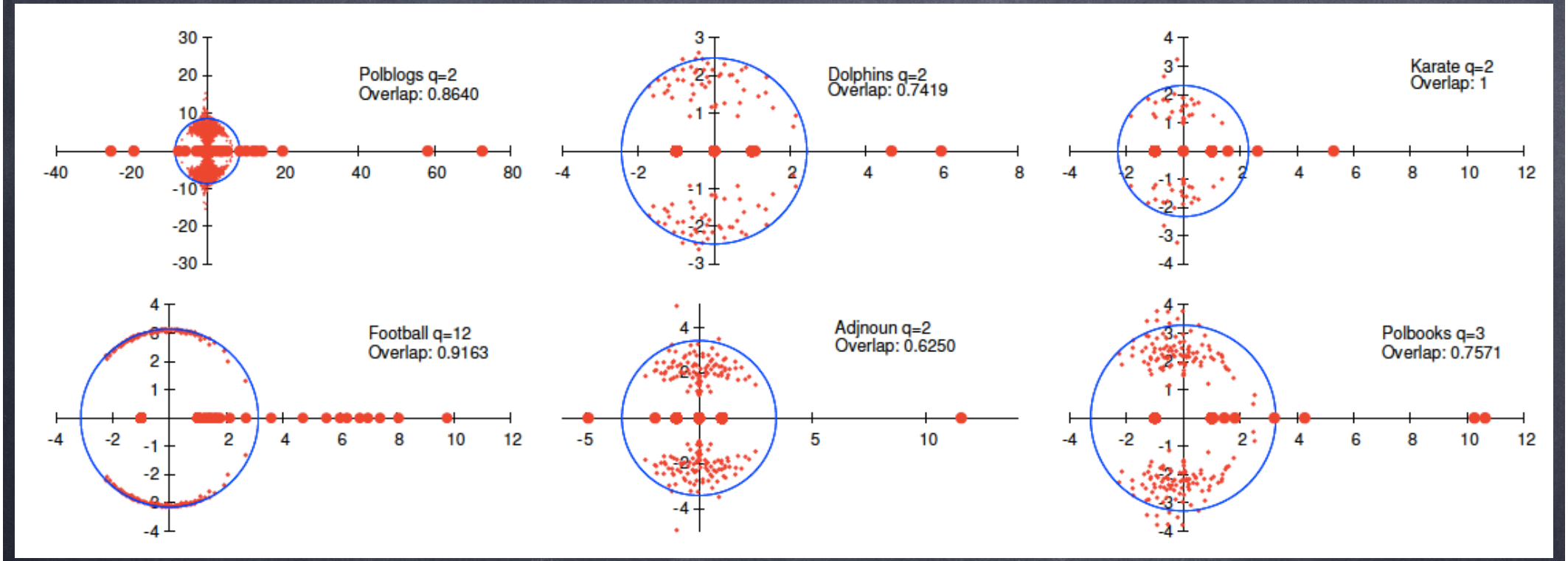
- Characteristic polynomial (Bass'92 formula)

$$\det\left[\mu^2 1 - \mu A + (D-1)\right] = 0$$

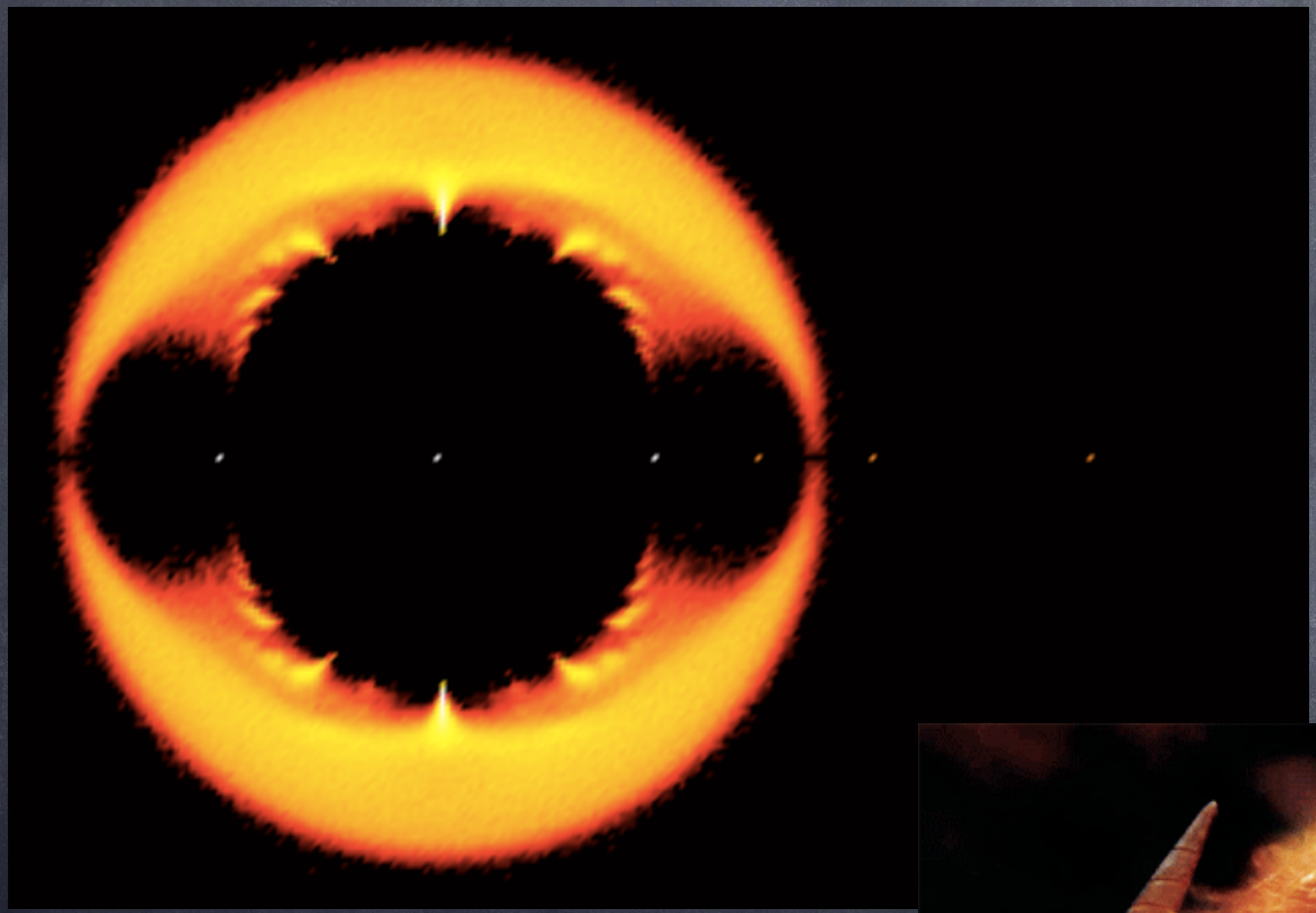
- Largest eigenvalue c.
- If $|c_{\text{in}} - c_{\text{out}}| = q\sqrt{c}$ the 2nd eigenvalue is informative and $> \sqrt{c}$
- All others are inside circle $\sqrt{c}$ (we believed, but proved only in density)
- Number of eigenvalues outside the circle = number of communities!

Complete proof now available
Massoulie, Lelarge and Bordenave

Spectra of some real networks



Spectrum of the non-backtracking matrix



Spectrum of the non-backtracking matrix

(with some physics hocus-pocus...)

$$\det\left[D - zA - (1-z^2)\mathbf{1}\right] = \prod_i^{2N}(z-\lambda_i)\,.$$

⬅Use the characteristic polynomial···

$$\nu(z) = \frac{1}{2N}\sum_{i=1}^{2N}\delta(z-\lambda_i)$$

⬅···to compute the spectrum density···

$$\delta(z-\mu) = \frac{1}{\pi}\partial_{\bar{z}}(z-\mu)^{-1}$$

⬅··· with the complex representation of δ

$$\nu(z) = \lim_{\epsilon\to 0}\frac{1}{2\pi N}\partial_{\bar{z}}\partial_z \log\det\mathcal{M}_\epsilon \qquad \mathcal{M}_\epsilon(z,A) = \begin{pmatrix} \epsilon\mathbf{1} & i(D-zA-(1-z^2)\mathbf{1}) \\ i(D-zA-(1-z^2)\mathbf{1})^\dagger & \epsilon\mathbf{1}\end{pmatrix}$$

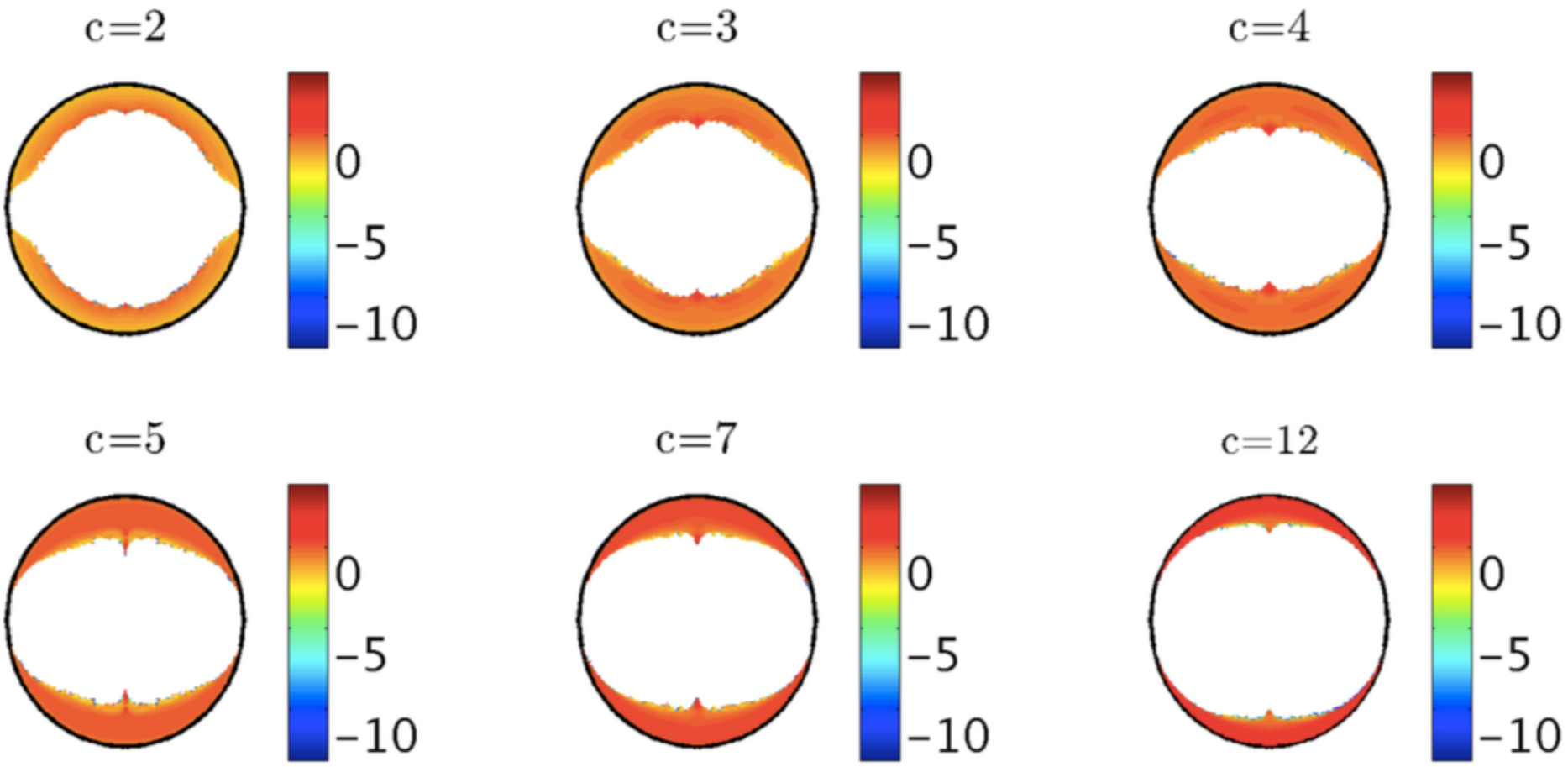
Using the integral representation of a determinant
(complex gaussian integral)
we map the problem to a statistical physics one
(on a random graph)

$$\nu(z) = -\lim_{\epsilon\to 0}\frac{1}{2\pi N}\partial_{\bar{z}}\partial_z \log\mathcal{Z}_\epsilon$$

A. Saade, FK & L. Zdeborova '14

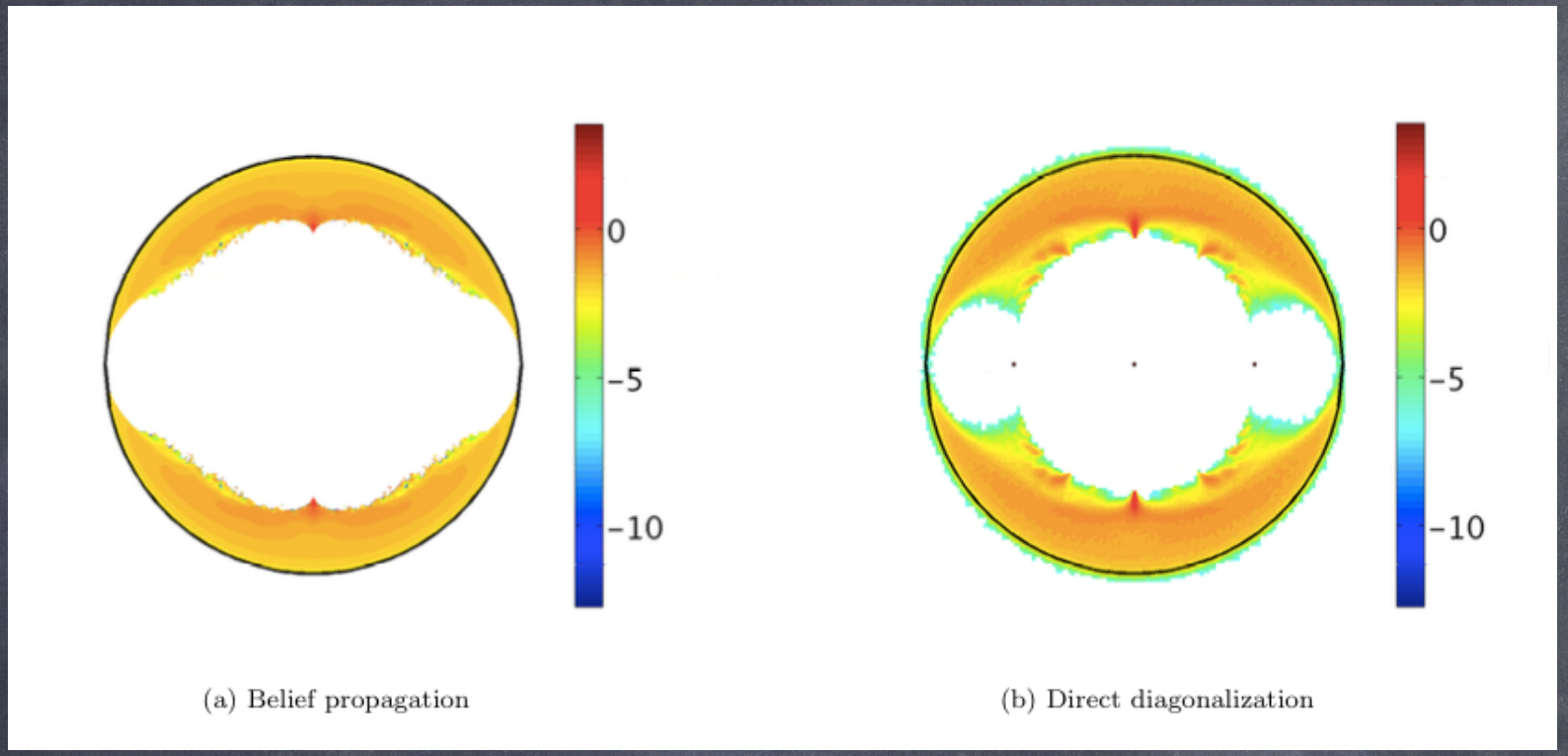
Spectrum of the non-backtracking matrix

(with some physics hocus-pocus...)



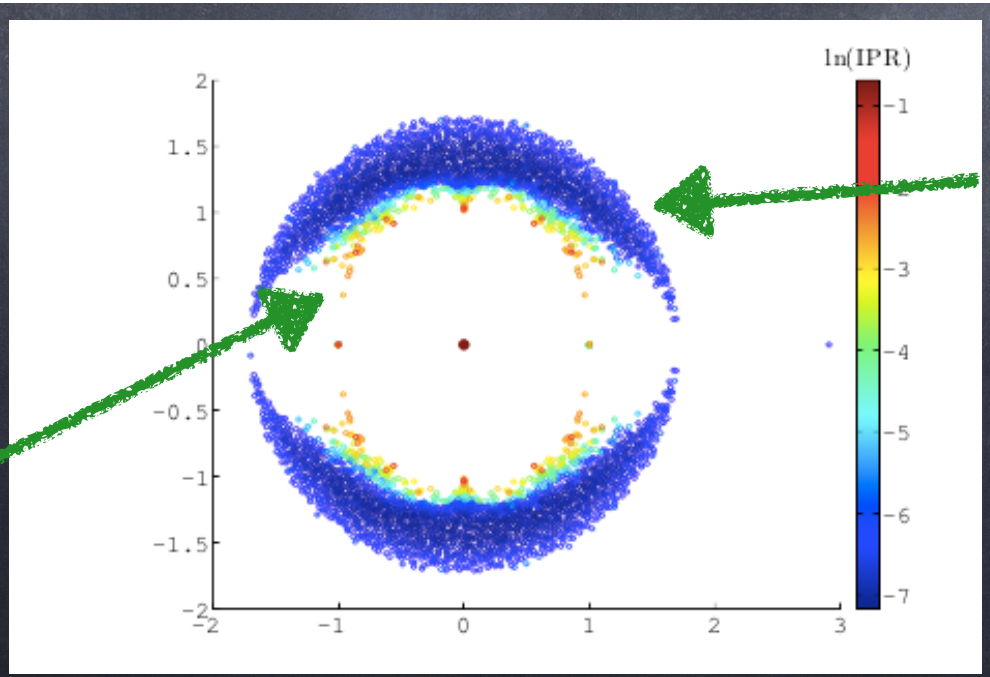
A. Saade, FK & L. Zdeborova '14

Spectrum of the non-backtracking matrix



(a) Belief propagation

(b) Direct diagonalization



Non-localized

localized

The non-backtracking operator

- Allows spectral clustering to match information theoretic performances

- Trivial to implement: spectrum of 2N x 2N matrix.

$$B' = \begin{pmatrix} 0 & D-1 \\ -1 & A \end{pmatrix}$$

- Parameter free: No need for learning

- Avoid the usual BP convergence problems on non-tree like graph.

The non-backtracking operator

Perspectives:

- Can we compute Full spectrum rigorously?
- I believe we have only scratched the surface of the potential of this approach
    - Spectral method for matrix factorization?
    - Censored block model (Abbe et al) ?
    - Percolation ?

Percolation

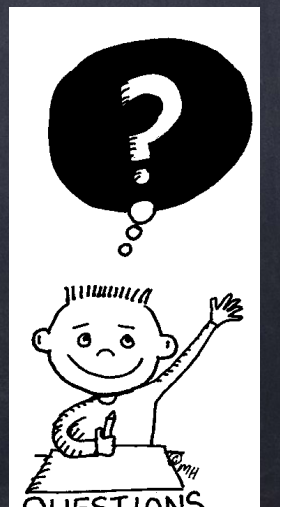
Dense network: The critical occupation probability for percolation on a dense network is equal to the reciprocal of the leading eigenvalue of the adjacency matrix (Bollobas *et al '10*)

ANY network: The critical occupation probability for percolation on a dense network is bounded to the reciprocal of the leading eigenvalue of the non-backtracking matrix (Zdeborova *et al '14*)

(and exact for sparse tree-like graph and dense graphes as well..)

This talk is based on:

- A. Decelle, **FK**, C. Moore, L. Zdeborová, Phase transition in the detection of modules in sparse networks, Phys. Rev. Lett. 107, 065701 (2011).
- A. Decelle, **FK**, C. Moore, L. Zdeborová, Asymptotic analysis of the stochastic block model for modular networks and its algorithmic applications, Phys. Rev. E 84, 066106 (2011).
- P. Zhang, **FK**, J. Reichardt, L. Zdeborová, Comparative Study for Inference of Hidden Classes in Stochastic Block Models, J. Stat. Mech. (2012) P12021
- **FK**, C. Moore, E. Mossel, J. Neeman, A. Sly, L. Zdeborová, P. Zhang, Spectral clustering of Sparse Networks, Proc. Nat. Acad. Sci. (2013)
- A. Saade, **FK** & L. Zdeborová, The spectrum of the non-backtracking matrix EPL (2014)
- A. Saade, **FK** & L. Zdeborová, Clustering with the Bethe Hessian NIPS 2014
- Implementations available at: http://mode_net.krzakala.org/



Thank you for your attention!