Community detection with the non-backtracking operator

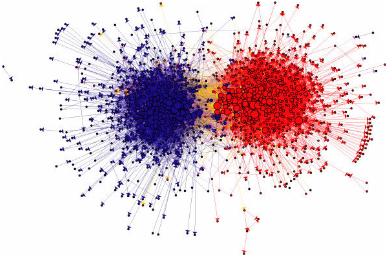
Marc Lelarge

INRIA-ENS

Aalto University, Helsinki, October 2016

Motivation

- Community detection in social or biological networks in the sparse regime with a small average degree.

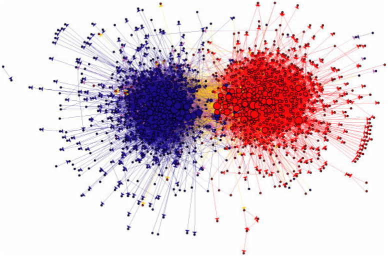


Adamic Glance '05

- Performance analysis of spectral algorithms on a toy model (where the ground truth is known!).

Motivation

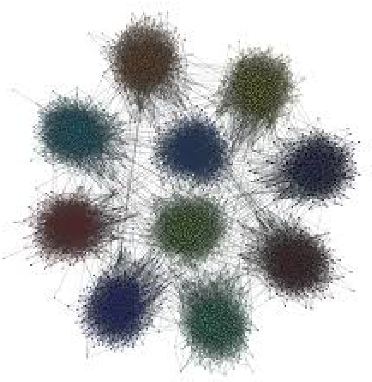
- Community detection in social or biological networks in the sparse regime with a small average degree.



Adamic Glance '05

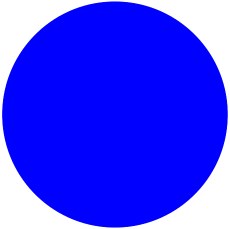
- Performance analysis of spectral algorithms on a toy model (where the ground truth is known!).

A model: the stochastic block model



The sparse stochastic block model

A random graph model on n nodes with three parameters, $a, b, c \geq 0$.

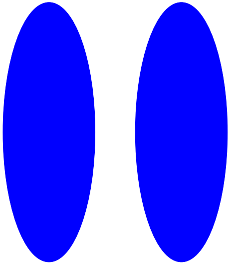


total population

The sparse stochastic block model

A random graph model on n nodes with three parameters, $a, b, c \geq 0$.

- Assign each vertex spin $+1$ or -1 uniformly at random.

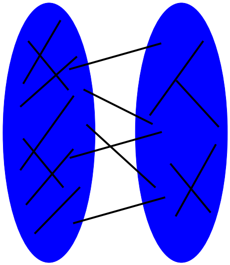


$+1$ and -1

The sparse stochastic block model

A random graph model on n nodes with three parameters, $a, b, c \geq 0$.

- Independently for each pair (u, v):
  - if $\sigma_u = \sigma_v = +1$, draw the edge w.p. a/n.
  - if $\sigma_u \neq \sigma_v$, draw the edge w.p. b/n.
  - if $\sigma_u = \sigma_v = -1$, draw the edge w.p. c/n.



a/n, b/n, c/n.

Community detection problem

- Reconstruct the underlying communities (i.e. spin configuration σ) based on one realization of the graph.
- Asymptotics: $n \to \infty$
- Sparse graph: the parameters a, b, c are fixed.
- notion of performance:
  w.h.p. strictly less than half of the vertices are misclassified
  = positively correlated partition.

Community detection problem

- Reconstruct the underlying communities (i.e. spin configuration σ) based on one realization of the graph.
- Asymptotics: $n \to \infty$
- Sparse graph: the parameters a, b, c are fixed.
- notion of performance:
  w.h.p. strictly less than half of the vertices are misclassified
  = positively correlated partition.

Community detection problem

- Reconstruct the underlying communities (i.e. spin configuration σ) based on one realization of the graph.
- Asymptotics: $n \to \infty$
- Sparse graph: the parameters a, b, c are fixed.
- notion of performance:
  w.h.p. strictly less than half of the vertices are misclassified
  = positively correlated partition.

Community detection problem

- Reconstruct the underlying communities (i.e. spin configuration σ) based on one realization of the graph.
- Asymptotics: $n \to \infty$
- Sparse graph: the parameters a, b, c are fixed.
- notion of performance:
  w.h.p. strictly less than half of the vertices are misclassified
  = positively correlated partition.

A first attempt: looking at degrees

- Degree in community $+1$ is:
  $D_+ \sim Bin\left(\frac{n}{2} - 1, \frac{a}{n}\right) + Bin\left(\frac{n}{2}, \frac{b}{n}\right)$
- We have
  $$\mathbb{E}[D_+] \approx \frac{a+b}{2}, \text{ and } Var(D_+) \approx \frac{a+b}{2}.$$
  and similarly, in community -1:
  $$\mathbb{E}[D_-] \approx \frac{c+b}{2}, \text{ and } Var(D_-) \approx \frac{c+b}{2}.$$
- Clustering based on degrees should 'work' as soon as:
  $$(\mathbb{E}[D_+] - \mathbb{E}[D_-])^2 \succ \max(Var(D_+), Var(D_-))$$
  i.e. (ignoring constant factors)
  $$(a-c)^2 \succ b + \max(a, c).$$

A first attempt: looking at degrees

- Degree in community $+1$ is:
  $D_+ \sim Bin\left(\frac{n}{2} - 1, \frac{a}{n}\right) + Bin\left(\frac{n}{2}, \frac{b}{n}\right)$
- We have

  $$\mathbb{E}[D_+] \approx \frac{a+b}{2}, \text{ and } Var(D_+) \approx \frac{a+b}{2}.$$

  and similarly, in community -1:

  $$\mathbb{E}[D_-] \approx \frac{c+b}{2}, \text{ and } Var(D_-) \approx \frac{c+b}{2}.$$

- Clustering based on degrees should 'work' as soon as:

  $$(\mathbb{E}[D_+] - \mathbb{E}[D_-])^2 \succ \max(Var(D_+), Var(D_-))$$

  i.e. (ignoring constant factors)

  $$(a-c)^2 \succ b + \max(a, c).$$

A first attempt: looking at degrees

- Degree in community $+1$ is:
  $D_+ \sim Bin\left(\frac{n}{2} - 1, \frac{a}{n}\right) + Bin\left(\frac{n}{2}, \frac{b}{n}\right)$
- We have

$$\mathbb{E}[D_+] \approx \frac{a+b}{2}, \text{ and } Var(D_+) \approx \frac{a+b}{2}.$$

  and similarly, in community -1:

$$\mathbb{E}[D_-] \approx \frac{c+b}{2}, \text{ and } Var(D_-) \approx \frac{c+b}{2}.$$

- Clustering based on degrees should 'work' as soon as:

$$(\mathbb{E}[D_+] - \mathbb{E}[D_-])^2 \succ \max(Var(D_+), Var(D_-))$$

  i.e. (ignoring constant factors)

$$(a-c)^2 \succ b + \max(a, c).$$

Is it any good?

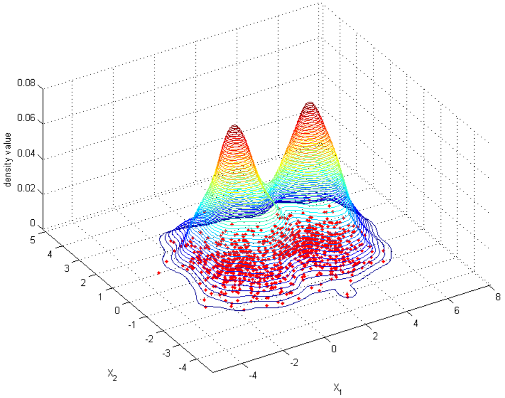
Data: A the adjacency matrix of the graph.
We define the mean column for each community:

$$A_+ = \frac{1}{n}\begin{pmatrix} a \\ \vdots \\ a \\ b \\ \vdots \\ b \end{pmatrix} \text{, and} \quad A_- = \frac{1}{n}\begin{pmatrix} b \\ \vdots \\ b \\ c \\ \vdots \\ c \end{pmatrix}$$

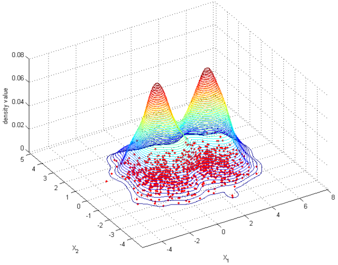
The variance of each entry is $\leq \max(a, b, c)/n$.
Pretend the columns are i.i.d., spherical Gaussian and $k = n$...

Clustering a mixture of Gaussians

Consider a mixture of two spherical Gaussians in $\mathbb{R}^n$ with respective means $\mathbf{m}_1$ and $\mathbf{m}_2$ and variance σ^2.
Pb: given k samples $\sim 1/2\mathcal{N}(\mathbf{m}_1, \sigma^2) + 1/2\mathcal{N}(\mathbf{m}_2, \sigma^2)$, recover the unknown parameters $\mathbf{m}_1$, $\mathbf{m}_2$ and σ^2.



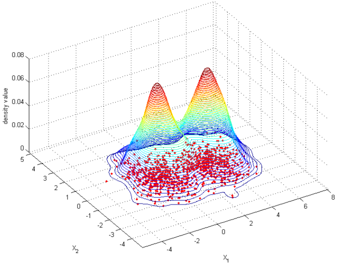
Doing better than naive algorithm



If $\|\mathbf{m}_1 - \mathbf{m}_2\|^2 \succ n\sigma^2$, then the densities 'do not overlap' in $\mathbb{R}^n$.

Projection preserves variance σ^2. So projecting onto the line formed by $\mathbf{m}_1$ and $\mathbf{m}_2$ gives 1-dim. Gaussian variables with no overlap as soon as $\|\mathbf{m}_1 - \mathbf{m}_2\|^2 \succ \sigma^2$. We gain a factor of n.

Doing better than naive algorithm



If $\|\mathbf{m}_1 - \mathbf{m}_2\|^2 \succ n\sigma^2$, then the densities 'do not overlap' in $\mathbb{R}^n$.

Projection preserves variance σ^2. So projecting onto the line formed by $\mathbf{m}_1$ and $\mathbf{m}_2$ gives 1-dim. Gaussian variables with no overlap as soon as $\|\mathbf{m}_1 - \mathbf{m}_2\|^2 \succ \sigma^2$. We gain a factor of n.

Algorithm for clustering a mixture of Gaussians

Each sample is a column of the following matrix:

$$A = (\mathbf{A}_1, \mathbf{A}_2, \ldots, \mathbf{A}_k) \in \mathbb{R}^{n \times k}$$

Consider the SVD of A:

$$A = \sum_{i=1}^{n} \lambda_i \mathbf{u}_i \mathbf{v}_i^T, \quad \mathbf{u}_i \in \mathbb{R}^n,\ \mathbf{v}_i \in \mathbb{R}^k,\ \lambda_1 \geq \lambda_2 \geq \ldots$$

Then the best approximation for the direction $(\mathbf{m}_1, \mathbf{m}_2)$ given by the data is $\mathbf{u}_1$.

Project the points from $\mathbb{R}^n$ onto this line and then do clustering. Provided k is large enough, this 'works' as soon as: $\|\mathbf{m}_1 - \mathbf{m}_2\|^2 \succ \sigma^2$.

Back to our clustering problem

Data: A the adjacency matrix of the graph.
The mean columns for each community are:

$$A_{+} = \frac{1}{n}\begin{pmatrix} a \\ \vdots \\ a \\ b \\ \vdots \\ b \end{pmatrix} \text{, and} \quad A_{-} = \frac{1}{n}\begin{pmatrix} b \\ \vdots \\ b \\ c \\ \vdots \\ c \end{pmatrix}$$

The variance of each entry is $\leq \max(a, b, c)/n$.

Heuristics for community detection

The naive algorithm should work as soon as

$$\|A_+ - A_-\|^2 \succ n \underbrace{\frac{\max(a,b,c)}{n}}_{Var}$$

$$(a-b)^2 + (b-c)^2 \succ n \max(a,b,c)$$

Spectral clustering should allow you a gain of n, i.e.

$$(a-b)^2 + (b-c)^2 \succ \max(a,b,c)$$

Our previous analysis shows that clustering based on degrees works as soon as

$$(a-c)^2 \succ \max(a,b,c).$$

When $a = c$, no information given by the degrees.

Heuristics for community detection

The naive algorithm should work as soon as

$$\|A_+ - A_-\|^2 \succ n \underbrace{\frac{\max(a,b,c)}{n}}_{Var}$$

$$(a-b)^2 + (b-c)^2 \succ n \max(a,b,c)$$

Spectral clustering should allow you a gain of n, i.e.

$$(a-b)^2 + (b-c)^2 \succ \max(a,b,c)$$

Our previous analysis shows that clustering based on degrees works as soon as

$$(a-c)^2 \succ \max(a,b,c).$$

When $a = c$, no information given by the degrees.

Heuristics for community detection

The naive algorithm should work as soon as

$$\|A_+ - A_-\|^2 \succ n \underbrace{\frac{\max(a,b,c)}{n}}_{Var}$$

$$(a-b)^2 + (b-c)^2 \succ n \max(a,b,c)$$

Spectral clustering should allow you a gain of n, i.e.

$$(a-b)^2 + (b-c)^2 \succ \max(a,b,c)$$

Our previous analysis shows that clustering based on degrees works as soon as

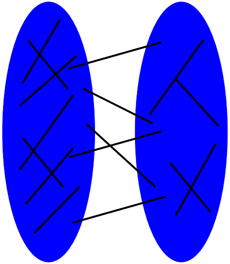
$$(a-c)^2 \succ \max(a,b,c).$$

When $a = c$, no information given by the degrees.

The sparse symmetric stochastic block model

A random graph model on n nodes with two parameters, $a, b \geq 0$.

- Independently for each pair (u, v):
  - if $\sigma_u = \sigma_v$, draw the edge w.p. a/n.
  - if $\sigma_u \neq \sigma_v$, draw the edge w.p. b/n.



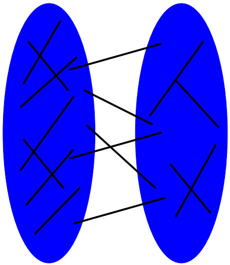
a/n, b/n, a/n.

Heuristic: spectral should work as soon as $(a - b)^2 \succ a + b$

The sparse symmetric stochastic block model

A random graph model on n nodes with two parameters, $a, b \geq 0$.

- Independently for each pair (u, v):
  - if $\sigma_u = \sigma_v$, draw the edge w.p. a/n.
  - if $\sigma_u \neq \sigma_v$, draw the edge w.p. b/n.



a/n, b/n, a/n.

Heuristic: spectral should work as soon as $(a-b)^2 \succ a + b$

Efficiency of Spectral Algorithms

Boppana '87, Condon, Karp '01, Carson, Impagliazzo '01, McSherry '01, Kannan, Vempala, Vetta '04...

Theorem

Suppose that for sufficiently large K and K',

$$\frac{(a-b)^2}{a+b} \geq (\succ) K + K' \ln(a+b),$$

then 'trimming+spectral+greedy improvement' outputs a positively correlated (almost exact) partition w.h.p.

Coja-Oghlan '10

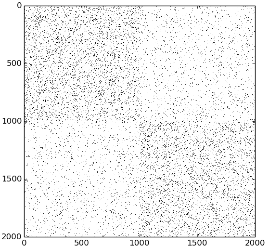
Heuristic based on analogy with mixture of Gaussians:

$$(a-b)^2 \quad \succ \quad a+b$$

Another look at spectral algorithms

Take a finite, simple, non-oriented graph $G = (V, E)$.
Adjacency matrix : symmetric, indexed on vertices, for $u, v \in V$,

$$A_{uv} = 1(\{u, v\} \in E).$$



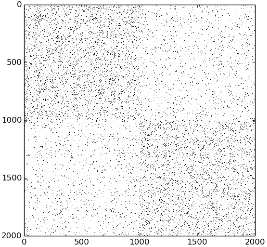
Low rank approximation of the adjacency matrix works as soon as

$$(a - b)^2 \succ a + b$$

Another look at spectral algorithms

Take a finite, simple, non-oriented graph $G = (V, E)$.
Adjacency matrix : symmetric, indexed on vertices, for $u, v \in V$,

$$A_{uv} = 1(\{u, v\} \in E).$$



Low rank approximation of the adjacency matrix works as soon as

$$(a - b)^2 \succ a + b$$

Spectral analysis

Assume that $a \to \infty$, and $a - b \approx \sqrt{a+b}$ so that $a \sim b$.

$$A = \frac{a+b}{2} \frac{\mathbf{1}}{\sqrt{n}} \frac{\mathbf{1}^T}{\sqrt{n}} + \frac{a-b}{2} \frac{\sigma}{\sqrt{n}} \frac{\sigma^T}{\sqrt{n}} + A - \mathbb{E}[A]$$

$\frac{a+b}{2}$ is the mean degree and degrees in the graph are very concentrated if $a \succ \ln n$. We can construct

$$A - \frac{a+b}{2n} J = \frac{a-b}{2} \frac{\sigma}{\sqrt{n}} \frac{\sigma^T}{\sqrt{n}} + A - \mathbb{E}[A]$$

Spectral analysis

Assume that $a \to \infty$, and $a - b \approx \sqrt{a+b}$ so that $a \sim b$.

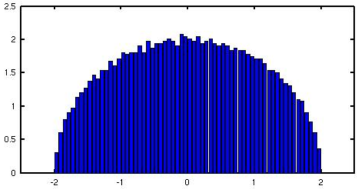
$$A = \frac{a+b}{2} \frac{\mathbf{1}}{\sqrt{n}} \frac{\mathbf{1}^T}{\sqrt{n}} + \frac{a-b}{2} \frac{\sigma}{\sqrt{n}} \frac{\sigma^T}{\sqrt{n}} + A - \mathbb{E}[A]$$

$\frac{a+b}{2}$ is the mean degree and degrees in the graph are very concentrated if $a \succ \ln n$. We can construct

$$A - \frac{a+b}{2n} J = \frac{a-b}{2} \frac{\sigma}{\sqrt{n}} \frac{\sigma^T}{\sqrt{n}} + A - \mathbb{E}[A]$$

Spectrum of the noise matrix

The matrix $A - \mathbb{E}[A]$ is a symmetric random matrix with independent centered entries having variance $\sim \frac{a}{n}$.
To have convergence to the Wigner semicircle law, we need to normalize the variance to $\frac{1}{n}$.



$$ESD\left(\frac{A - \mathbb{E}[A]}{\sqrt{a}}\right) \to \mu_{sc}(x) = \begin{cases} \frac{1}{2\pi}\sqrt{4 - x^2}, & \text{if } |x| \leq 2; \\ 0, & \text{otherwise.} \end{cases}$$

Naive spectral analysis

To sum up, we can construct:

$$M = \frac{1}{\sqrt{a}}\left(A - \frac{a+b}{2n}J\right) = \theta\frac{\sigma}{\sqrt{n}}\frac{\sigma^T}{\sqrt{n}} + \frac{A-\mathbb{E}[A]}{\sqrt{a}},$$

with $\theta = \frac{a-b}{\sqrt{2(a+b)}}$.

We should be able to detect signal as soon as

$$\theta > 2 \Leftrightarrow \frac{(a-b)^2}{2(a+b)} > 4$$

Naive spectral analysis

To sum up, we can construct:

$$\begin{aligned} M &= \frac{1}{\sqrt{a}}\left(A - \frac{a+b}{2n}J\right) \\ &= \theta \frac{\sigma}{\sqrt{n}}\frac{\sigma^T}{\sqrt{n}} + \frac{A - \mathbb{E}[A]}{\sqrt{a}}, \end{aligned}$$

with $\theta = \frac{a-b}{\sqrt{2(a+b)}}$.
We should be able to detect signal as soon as

$$\theta > 2 \Leftrightarrow \frac{(a-b)^2}{2(a+b)} > 4$$

We can do better!

A lower bound on the spectral radius of $M = \theta \frac{\sigma}{\sqrt{n}} \frac{\sigma^T}{\sqrt{n}} + W$:

$$\lambda_1(M) = \sup_{\|x\|=1} \|Mx\| \geq \|M \frac{\sigma}{\sqrt{n}}\|$$

But

$$\begin{aligned} \|M \frac{\sigma}{\sqrt{n}}\|^2 &= \theta^2 + \|W \frac{\sigma}{\sqrt{n}}\|^2 + 2\langle W, \frac{\sigma}{\sqrt{n}} \rangle \\ &\approx \theta^2 + \frac{1}{n} \sum_{i,j} W_{ij}^2 \\ &\approx \theta^2 + 1. \end{aligned}$$

As a result, we get

$$\lambda_1(M) > 2 \Leftrightarrow \theta > 1 \Leftrightarrow (a-b)^2 > 2(a+b).$$

We can do better!

A lower bound on the spectral radius of $M = \theta \frac{\sigma}{\sqrt{n}} \frac{\sigma^T}{\sqrt{n}} + W$:

$$\lambda_1(M) = \sup_{\|x\|=1} \|Mx\| \geq \|M \frac{\sigma}{\sqrt{n}}\|$$

But

$$\begin{aligned} \|M \frac{\sigma}{\sqrt{n}}\|^2 &= \theta^2 + \|W \frac{\sigma}{\sqrt{n}}\|^2 + 2\langle W, \frac{\sigma}{\sqrt{n}} \rangle \\ &\approx \theta^2 + \frac{1}{n} \sum_{i,j} W_{ij}^2 \\ &\approx \theta^2 + 1. \end{aligned}$$

As a result, we get

$$\lambda_1(M) > 2 \Leftrightarrow \theta > 1 \Leftrightarrow (a-b)^2 > 2(a+b).$$

We can do better!

A lower bound on the spectral radius of $M = \theta \frac{\sigma}{\sqrt{n}} \frac{\sigma^T}{\sqrt{n}} + W$:

$$\lambda_1(M) = \sup_{\|x\|=1} \|Mx\| \geq \|M \frac{\sigma}{\sqrt{n}}\|$$

But

$$\begin{aligned} \|M \frac{\sigma}{\sqrt{n}}\|^2 &= \theta^2 + \|W \frac{\sigma}{\sqrt{n}}\|^2 + 2\langle W, \frac{\sigma}{\sqrt{n}} \rangle \\ &\approx \theta^2 + \frac{1}{n} \sum_{i,j} W_{ij}^2 \\ &\approx \theta^2 + 1. \end{aligned}$$

As a result, we get

$$\lambda_1(M) > 2 \Leftrightarrow \theta > 1 \Leftrightarrow (a-b)^2 > 2(a+b).$$

Baik, Ben Arous, Péché phase transition

Rank one perturbation of a Wigner matrix:

$$\lambda_1(\theta\sigma\sigma^T + W) \stackrel{a.s}{\to} \begin{cases} \theta + \frac{1}{\theta} & \text{if } \theta > 1, \\ 2 & \text{otherwise.} \end{cases}$$

Let $\tilde{\sigma}$ be the eigenvector associated with $\lambda_1(\theta uu^T + W)$, then

$$|\langle \tilde{\sigma}, \sigma \rangle|^2 \stackrel{a.s}{\to} \begin{cases} 1 - \frac{1}{\theta^2} & \text{if } \theta > 1, \\ 0 & \text{otherwise.} \end{cases}$$

Watkin Nadal '94, Baik, Ben Arous, Péché '05
Newman, Rao '14
For SBM with $a, b \to \infty$,

$$\theta^2 = \frac{(a-b)^2}{2(a+b)} > 1$$

Baik, Ben Arous, Péché phase transition

Rank one perturbation of a Wigner matrix:

$$\lambda_1(\theta\sigma\sigma^T + W) \overset{a.s}{\to} \begin{cases} \theta + \frac{1}{\theta} & \text{if } \theta > 1, \\ 2 & \text{otherwise.} \end{cases}$$

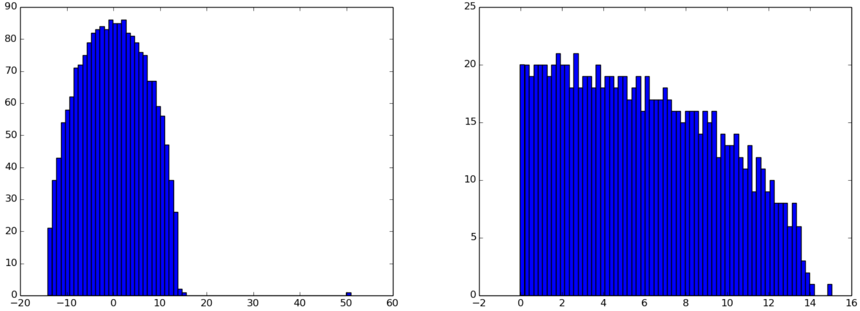
Let $\tilde{\sigma}$ be the eigenvector associated with $\lambda_1(\theta uu^T + W)$, then

$$|\langle \tilde{\sigma}, \sigma \rangle|^2 \overset{a.s}{\to} \begin{cases} 1 - \frac{1}{\theta^2} & \text{if } \theta > 1, \\ 0 & \text{otherwise.} \end{cases}$$

Watkin Nadal '94, Baik, Ben Arous, Péché '05
Newman, Rao '14
For SBM with $a, b \to \infty$,

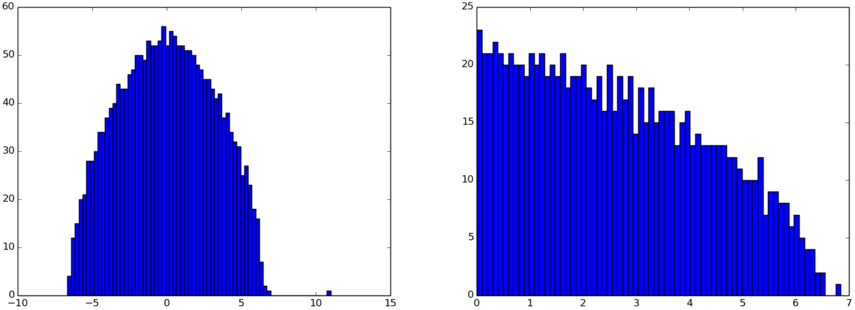
$$\theta^2 = \frac{(a-b)^2}{2(a+b)} > 1$$

When $a, b \to \infty$ spectral is optimal



SBM with $n = 2000$, average degree 50 and $\frac{(a-b)^2}{2(a+b)} = 2$.
Random matrix theory predicts $\lambda_1 = 51$, $\lambda_2 = 15$ and noise at $|\lambda_3| < 14.14$

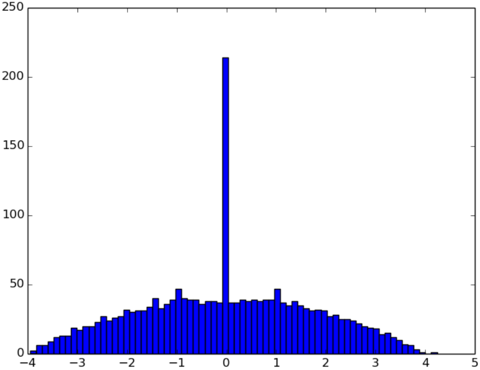
Decreasing the average degree



SBM with $n = 2000$, average degree 10 and $\frac{(a-b)^2}{2(a+b)} = 2$.

Random matrix theory predicts $\lambda_1 = 11$, $\lambda_2 = 6.7$ and noise at $|\lambda_3| < 6.3$

Problems when the average degree is small



SBM with $n = 2000$, average degree 3 and $\frac{(a-b)^2}{2(a+b)} = 2$.
Random matrix theory predicts $\lambda_1 = 4$, $\lambda_2 = 3.67$ and noise at $|\lambda_3| < 3.46$

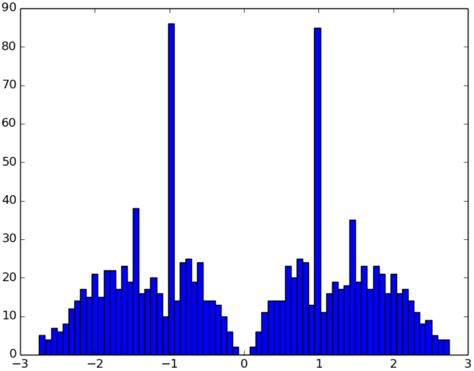
Problems when the average degree is finite

- High degree nodes: a star with degree d has eigenvalues $\{-\sqrt{d}, 0, \sqrt{d}\}$.
  In the regime where a and b are finite, the degrees are asymptotically Poisson with mean $\frac{a+b}{2}$. The adjacency matrix has $\Omega\left(\sqrt{\frac{\ln n}{\ln \ln n}}\right)$ eigenvalues.
- Low degree nodes: instead of the adjacency matrix, take the (normalized) Laplacian but then isolated edges produce spurious eigenvalues.

Problems when the average degree is finite

- High degree nodes: a star with degree d has eigenvalues $\{-\sqrt{d}, 0, \sqrt{d}\}$.
  In the regime where a and b are finite, the degrees are asymptotically Poisson with mean $\frac{a+b}{2}$. The adjacency matrix has $\Omega\left(\sqrt{\frac{\ln n}{\ln \ln n}}\right)$ eigenvalues.
- Low degree nodes: instead of the adjacency matrix, take the (normalized) Laplacian but then isolated edges produce spurious eigenvalues.

Problems when the average degree is small



Same graph after trimming.

Phase transition for $a, b = O(1)$

Theorem

$$\tau = \frac{(a-b)^2}{2(a+b)}$$

If $\tau > 1$, then positively correlated reconstruction is possible.
If $\tau < 1$, then positively correlated reconstruction is impossible.

Conjectured by Decelle, Krzakala, Moore, Zdeborova '11 based on statistical physics arguments.

- Non-reconstruction proved by Mossel, Neeman, Sly '12.
- Reconstruction proved by Massoulié '13 and Mossel, Neeman, Sly '13.

Phase transition for $a, b = O(1)$

Theorem

$$\tau = \frac{(a-b)^2}{2(a+b)}$$

If $\tau > 1$, then positively correlated reconstruction is possible.
If $\tau < 1$, then positively correlated reconstruction is impossible.

Conjectured by Decelle, Krzakala, Moore, Zdeborova '11 based on statistical physics arguments.

- Non-reconstruction proved by Mossel, Neeman, Sly '12.
- Reconstruction proved by Massoulié '13 and Mossel, Neeman, Sly '13.

Phase transition for $a, b = O(1)$

Theorem

$$\tau = \frac{(a-b)^2}{2(a+b)}$$

If $\tau > 1$, then positively correlated reconstruction is possible.
If $\tau < 1$, then positively correlated reconstruction is impossible.

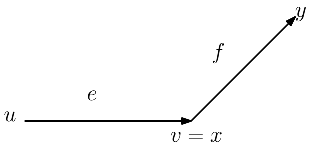
Conjectured by Decelle, Krzakala, Moore, Zdeborova '11 based on statistical physics arguments.

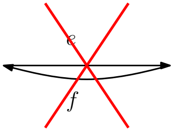
- Non-reconstruction proved by Mossel, Neeman, Sly '12.
- Reconstruction proved by Massoulié '13 and Mossel, Neeman, Sly '13.

Regularization through the non-backtracking matrix

Let $\vec{E} = \{u \to v; \{u,v\} \in E\}$ be the set of oriented edges.
$m = |\vec{E}|$ is twice the number of unoriented edges.
The non-backtracking matrix is an $m \times m$ matrix defined by

$$B_{u \to v, v \to w} = 1(\{u,v\} \in E)1(\{v,w\} \in E)1(u \neq w)$$





B is NOT symmetric: $B^T \neq B$. We denote its eigenvalues by $\lambda_1, \lambda_2, \dots$ with $\lambda_1 \geq \dots \geq |\lambda_m|$.
Proposed by Krzakala et al. '13.

Ihara-Bass' Identity

Let D the diagonal matrix with $D_{vv} = \deg(v)$. We have

$$det(z - B) \;=\; (z^2 - 1)^{|E|-|V|} det(z^2 - Az + D - Id)$$

If G is d-regular, then $D = dId$ and,

$$\sigma(B) = \{\pm 1\} \cup \left\{ \lambda : \lambda^2 - \lambda\mu + (d-1) = 0 \text{ with } \mu \in \sigma(A) \right\}.$$

Ihara-Bass' Identity

Let D the diagonal matrix with $D_{vv} = \deg(v)$. We have

$$det(z - B) = (z^2 - 1)^{|E|-|V|} det(z^2 - Az + D - Id)$$

If G is d-regular, then $D = dId$ and,

$$\sigma(B) = \{\pm 1\} \cup \left\{\lambda : \lambda^2 - \lambda\mu + (d-1) = 0 \text{ with } \mu \in \sigma(A)\right\}.$$

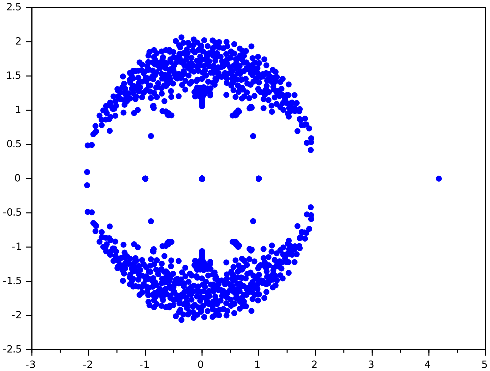
Non-Backtracking matrix of regular graphs

For a d-regular graph, $\lambda_1 = d - 1$,

- ⋆ Alon-Boppana bound : $\max_{k \neq 1} \Re(\lambda_k) \geq \sqrt{\lambda_1} - o(1)$.
- ⋆ Ramanujan (non bipartite) : $|\lambda_2| = \sqrt{\lambda_1}$
- ⋆ Friedman's thm : $|\lambda_2| \leq \sqrt{\lambda_1} + o(1)$ if G random uniform.

Simulation for Erdős-Rényi Graph

Eigenvalues of B for an Erdős-Rényi graph $G(n, \lambda/n)$ with $n = 500$ and $\lambda = 4$.



Erdős-Rényi Graph

Eigenvalues of B: $\lambda_1 \geq |\lambda_2| \geq \ldots$.

Theorem

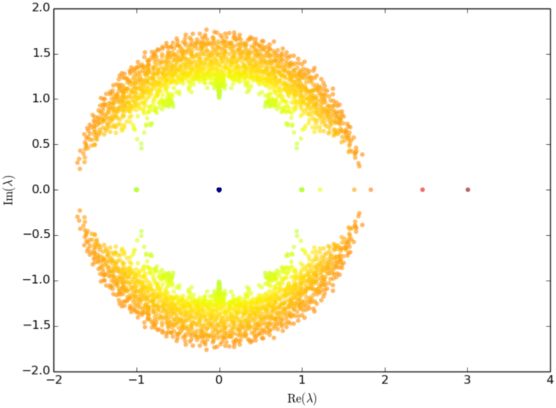
Let $\lambda > 1$ and G with distribution $G(n, \lambda/n)$. With high probability,

$$\begin{aligned} \lambda_1 &= \lambda + o(1) \\ |\lambda_2| &\leq \sqrt{\lambda} + o(1). \end{aligned}$$

Bordenave, Lelarge, Massoulié '15

Simulation for Stochastic Block Model

Eigenvalues of B for a Stochastic Block Model with $n = 2000$, mean degree $\frac{a+b}{2} = 3$ and $\frac{a-b}{2} = 2.45$



Stochastic Block Model

Eigenvalues of B: $\lambda_1 \geq |\lambda_2| \geq \dots$.

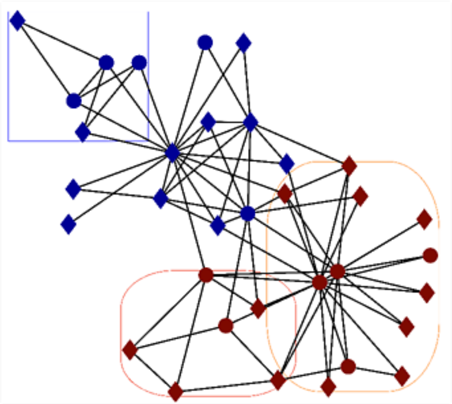
Theorem

Let G be a Stochastic Block Model with parameters a, b. *If* $(a-b)^2 > 2(a+b)$, *then with high probability,*

$$\begin{aligned} \lambda_1 &= \frac{a+b}{2} + o(1) \\ \lambda_2 &= \frac{a-b}{2} + o(1) \\ |\lambda_3| &\leq \sqrt{\frac{a+b}{2}} + o(1). \end{aligned}$$

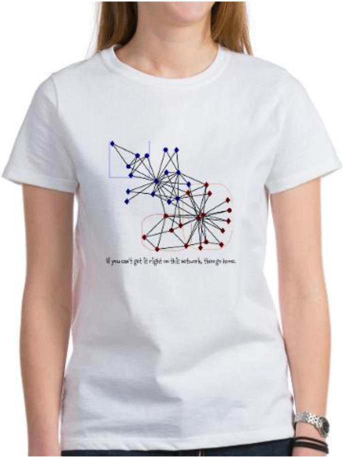
Bordenave, Lelarge, Massoulié ’15

Test with real benchmarks



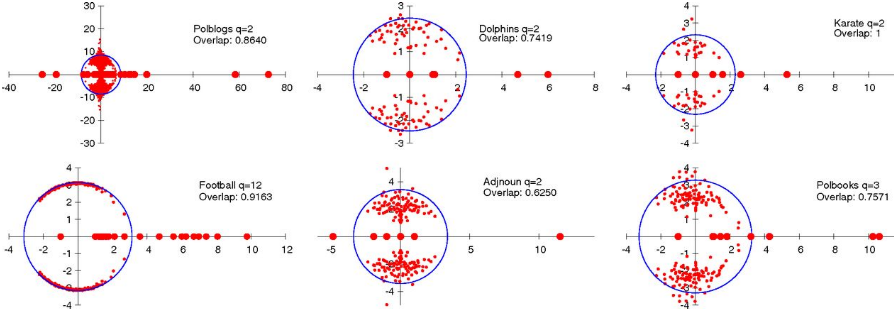
If you can't get it right on this network, then go home.

Test with real benchmarks



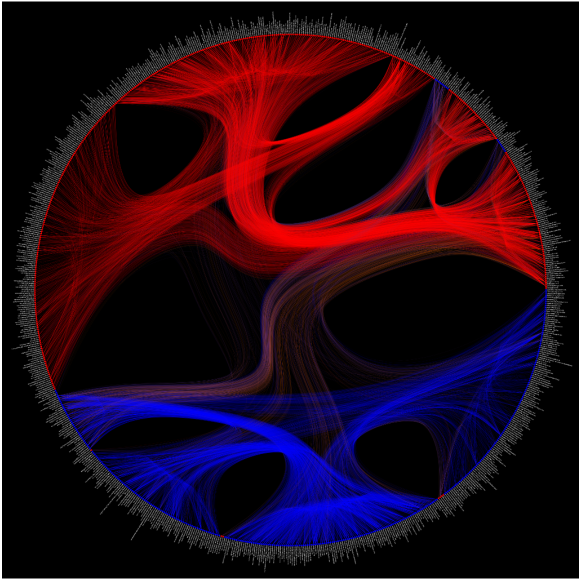
The Power Law Shop

The non-backtracking matrix on real data



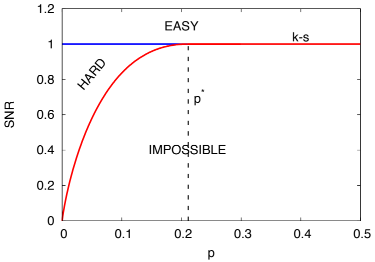
from Krzakala, Moore, Mossel, Neeman, Sly, Zdeborovà ’13

Back to political blogging network data



Non-symmetric Stochastic Block Model

Consider the case where there is a small community of size pn with $p < 1/2$, then the SNR is given by $d(1-b)^2$ where d is the average degree.

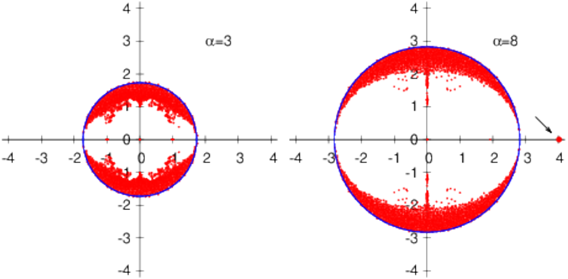


Phase diagram with $p^* = \frac{1}{2} - \frac{1}{2\sqrt{3}}$.

Lelarge, Caltagirone & Miolane, '16

Some extensions

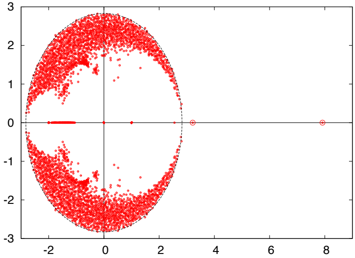
For the labeled stochastic block model, we also conjecture a phase transition. We have partial results and an optimal spectral algorithm.



Saade, Krzakala, Lelarge, Zdeborovà, '15,'16

Some extensions

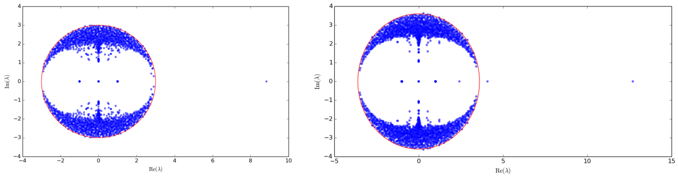
The non-backtracking matrix is also working for the degree-corrected SBM.
ongoing work with Gulikers and Massoulié.
We can adapt the non-backtracking matrix to deal with small cliques.



ongoing work with Caltagirone.

Some extensions

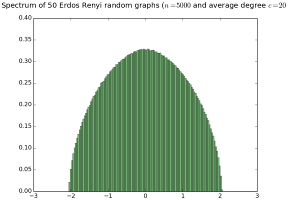
SBM with no noise $b = 0$ but with overlap.
Spectrum of the non-backtracking operator with $n = 1200$, $sn = 400$ and $a = 9$ and 13. The circle has radius $\sqrt{a(2-3s)}$ in each case.

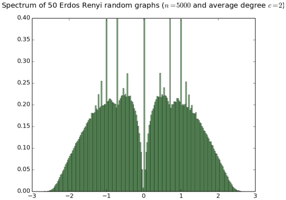


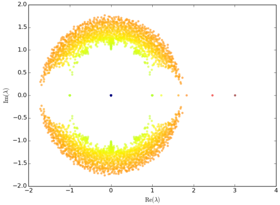
Kaufmann, Bonald, Lelarge ’16

Non-backtracking vs adjacency

On the sparse stochastic block model with probability of intra-edge a/n and inter-edge b/n.





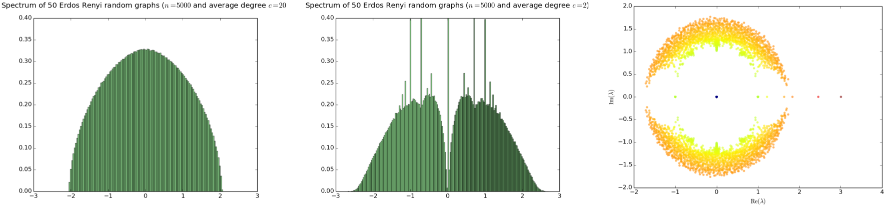


The problem: if $a, b \to \infty$, then Wigner's semi-circle law + BBP phase transition but if $a, b < \infty$ as $n \to \infty$, then Lifshitz tails.
The solution: the non-backtracking matrix on directed edges of the graph: $B_{u \to v, v \to w} = 1(\{u, v\} \in E) 1(\{v, w\} \in E) 1(u \neq w)$ achieves optimal detection on the SBM.

THANK YOU!

Non-backtracking vs adjacency

On the sparse stochastic block model with probability of intra-edge a/n and inter-edge b/n.



The problem: if $a, b \to \infty$, then Wigner's semi-circle law + BBP phase transition but if $a, b < \infty$ as $n \to \infty$, then Lifshitz tails.
The solution: the non-backtracking matrix on directed edges of the graph: $B_{u \to v, v \to w} = 1(\{u, v\} \in E)1(\{v, w\} \in E)1(u \neq w)$ achieves optimal detection on the SBM.

THANK YOU!