Community detection with spectral methods

Marc Lelarge [1]
Charles Bordenave[2] Laurent Massoulié[3] Jiaming Xu[4]

[1]INRIA-ENS

[2]CNRS Université de Toulouse

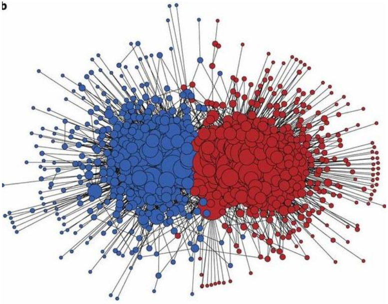
[3]INRIA-Microsoft Research Joint Centre

[4]UIUC

University of Bristol, November 2014

Motivation

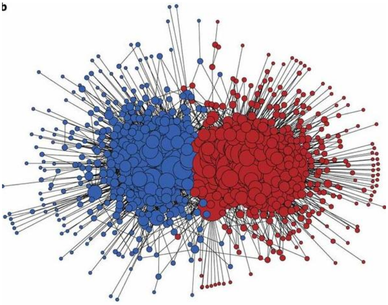
- Community detection in social or biological networks in the sparse regime with a small average degree.



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

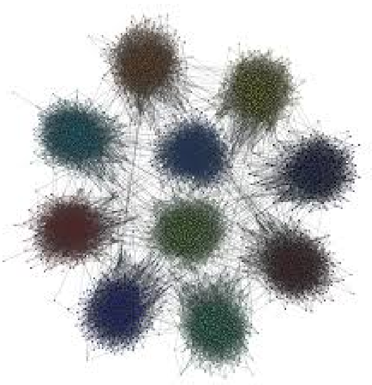
Motivation

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



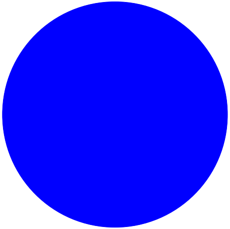
- Performance analysis of spectral algorithms on a toy model (where the ground truth is known!).

A model: the stochastic block model



The sparse stochastic block model

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

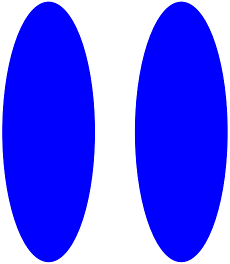


total population

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.

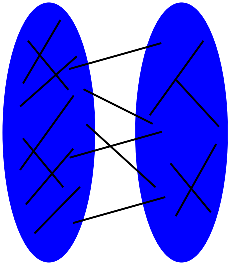


$+1$ and -1

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: as $n \to \infty$, the parameters a, b, c might depend of n and tend to infinity as well.
- Sparse graph: in all cases, $\max(a, b, c)/n \to 0$.
- 2 notions of performance:
  w.h.p. $o(n)$ vertices are misclassified = almost exact partition
  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: as $n \to \infty$, the parameters a, b, c might depend of n and tend to infinity as well.
- Sparse graph: in all cases, $\max(a, b, c)/n \to 0$.
- 2 notions of performance:
  w.h.p. $o(n)$ vertices are misclassified = almost exact partition
  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: as $n \to \infty$, the parameters a, b, c might depend of n and tend to infinity as well.
- Sparse graph: in all cases, $\max(a, b, c)/n \to 0$.
- 2 notions of performance:
  w.h.p. $o(n)$ vertices are misclassified = almost exact partition
  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: as $n \to \infty$, the parameters a, b, c might depend of n and tend to infinity as well.
- Sparse graph: in all cases, $\max(a, b, c)/n \to 0$.
- 2 notions of performance:
  w.h.p. $o(n)$ vertices are misclassified = almost exact partition
  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:
  $D_+ \sim Bin\left(\frac{n}{2} - 1, \frac{a}{n}\right) + Bin\left(\frac{n}{2}, \frac{b}{n}\right)$
- As soon as $\frac{\max(a,b)}{n} \to 0$, 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:
  $D_+ \sim Bin\left(\frac{n}{2} - 1, \frac{a}{n}\right) + Bin\left(\frac{n}{2}, \frac{b}{n}\right)$
- As soon as $\frac{\max(a,b)}{n} \to 0$, 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:
  $D_+ \sim Bin\left(\frac{n}{2} - 1, \frac{a}{n}\right) + Bin\left(\frac{n}{2}, \frac{b}{n}\right)$
- As soon as $\frac{\max(a,b)}{n} \to 0$, 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?

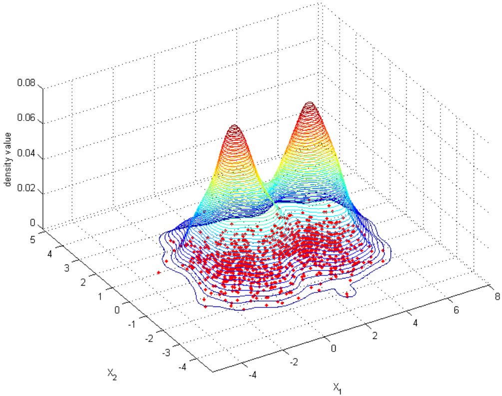
Data: A the adjacency matrix of the graph.
We define the mean column for each community:

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

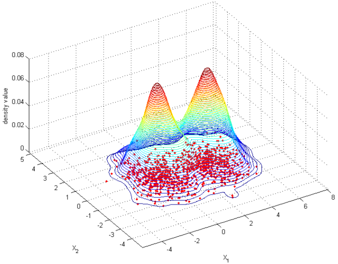
The variance of each entry is $\leq \max(a, b, c)/n$.
Pretend the columns are i.i.d., spherical Gaussian and $k = n$!

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.



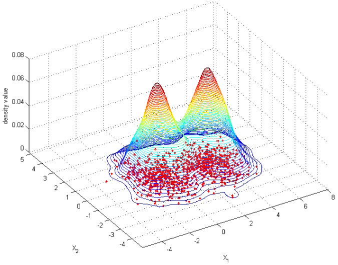
Doing better than naive algorithm



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:

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

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.

Symmetric model: $a = c$

Symmetric model: total population of size n splitted in 2 equal size communities. Probability of an edge intra: a/n and inter b/n.

As a result, the degree in each community is:
$D_+ \sim D_- \sim D \sim Bin\left(\frac{n}{2} - 1, \frac{a}{n}\right) + Bin\left(\frac{n}{2}, \frac{b}{n}\right)$.

Are we close to the Gaussian case?
Degree is a projection so is it Gaussian?

- if $a + b \to \infty$, then $D \approx \frac{a+b}{2} + \sqrt{\frac{a+b}{2}}\mathcal{N}(0,1)$
- if $a + b \prec \infty$, then $D \approx Poi\left(\frac{a+b}{2}\right)$.

Additional difficulties: the matrix A is symmetric, i.e. non i.i.d. columns and the number of samples is equal to the dimension n.

Symmetric model: $a = c$

Symmetric model: total population of size n splitted in 2 equal size communities. Probability of an edge intra: a/n and inter b/n.
As a result, the degree in each community is:
$D_+ \sim D_- \sim D \sim Bin\left(\frac{n}{2} - 1, \frac{a}{n}\right) + Bin\left(\frac{n}{2}, \frac{b}{n}\right)$.

Are we close to the Gaussian case?
Degree is a projection so is it Gaussian?

- if $a + b \to \infty$, then $D \approx \frac{a+b}{2} + \sqrt{\frac{a+b}{2}}\mathcal{N}(0,1)$
- if $a + b \prec \infty$, then $D \approx Poi\left(\frac{a+b}{2}\right)$.

Additional difficulties: the matrix A is symmetric, i.e. non i.i.d. columns and the number of samples is equal to the dimension n.

Symmetric model: $a = c$

Symmetric model: total population of size n splitted in 2 equal size communities. Probability of an edge intra: a/n and inter b/n.
As a result, the degree in each community is:
$D_+ \sim D_- \sim D \sim Bin\left(\frac{n}{2} - 1, \frac{a}{n}\right) + Bin\left(\frac{n}{2}, \frac{b}{n}\right)$.

Are we close to the Gaussian case?
Degree is a projection so is it Gaussian?

- if $a + b \to \infty$, then $D \approx \frac{a+b}{2} + \sqrt{\frac{a+b}{2}}\mathcal{N}(0,1)$
- if $a + b \prec \infty$, then $D \approx Poi\left(\frac{a+b}{2}\right)$.

Additional difficulties: the matrix A is symmetric, i.e. non i.i.d. columns and the number of samples is equal to the dimension n.

Symmetric model: $a = c$

Symmetric model: total population of size n splitted in 2 equal size communities. Probability of an edge intra: a/n and inter b/n.

As a result, the degree in each community is:

$D_+ \sim D_- \sim D \sim Bin\left(\frac{n}{2} - 1, \frac{a}{n}\right) + Bin\left(\frac{n}{2}, \frac{b}{n}\right)$.

Are we close to the Gaussian case?

Degree is a projection so is it Gaussian?

- if $a + b \to \infty$, then $D \approx \frac{a+b}{2} + \sqrt{\frac{a+b}{2}}\mathcal{N}(0, 1)$
- if $a + b \prec \infty$, then $D \approx Poi\left(\frac{a+b}{2}\right)$.

Additional difficulties: the matrix A is symmetric, i.e. non i.i.d. columns and the number of samples is equal to the dimension n.

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:

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

Phase transition

Theorem

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

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

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

Theorem

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

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

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

Theorem

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

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

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.

2 improvements

In the case $a, b \to \infty$, we remove the log factor in Coja-Oghlan's result.

In the case a, b finite, we compute the detectability threshold using the non-backtracking operator .

2 improvements

In the case $a, b \to \infty$, we remove the log factor in Coja-Oghlan's result.

In the case a, b finite, we compute the detectability threshold using the non-backtracking operator .

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 $a \sim b$.

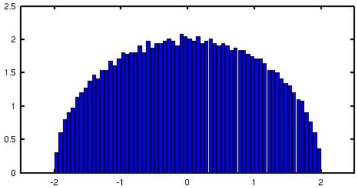
$$A = \frac{a+b}{2} \frac{\mathbf{1}}{\sqrt{n}} \frac{\mathbf{1}^T}{\sqrt{n}} + \frac{a-b}{2} \frac{\sigma}{\sqrt{n}} \frac{\sigma^T}{\sqrt{n}} + A - \mathbb{E}[A]$$

$\frac{a+b}{2}$ 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:

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

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

We should be able to detect signal as soon as

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

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:

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

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

Phase transition for $a \to \infty$

Proposition

Assume $a \succ \ln n$. Then the simple spectral method outputs an almost exact partition, provided $\frac{(a-b)^2}{(a+b)} \succ 1$. Moreover, no algorithm can find an almost exact parition if $\frac{(a-b)^2}{(a+b)} \prec 1$.
If $a \geq \ln^4 n$, then the simple spectral method outputs a positively correlated partition, provided

$$\frac{(a-b)^2}{(a+b)} > 2.$$

Proof: control the spectral norm thanks to Vu '05 and adapt the argument in Benaych-Georges, Nadakuditi '11.

Phase transition for $a \to \infty$

Proposition

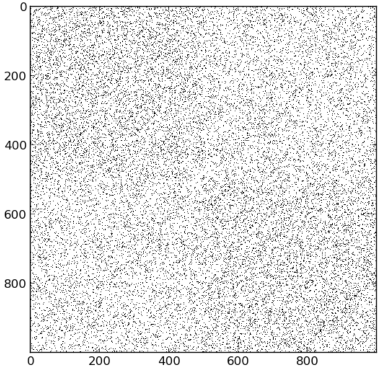
Assume $a \succ \ln n$. Then the simple spectral method outputs an almost exact partition, provided $\frac{(a-b)^2}{(a+b)} \succ 1$. Moreover, no algorithm can find an almost exact parition if $\frac{(a-b)^2}{(a+b)} \prec 1$.
If $a \geq \ln^4 n$, then the simple spectral method outputs a positively correlated partition, provided

$$\frac{(a-b)^2}{(a+b)} > 2.$$

Proof: control the spectral norm thanks to Vu '05 and adapt the argument in Benaych-Georges, Nadakuditi '11.

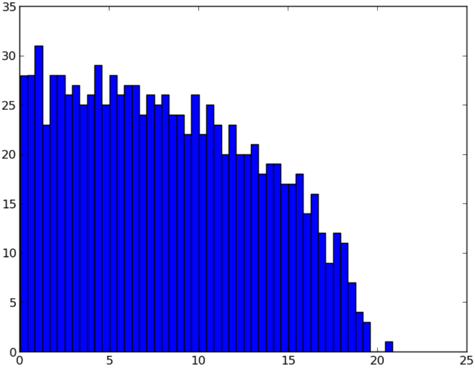
Spectral Algorithm

Original adjacency matrix with 2 communities. $a = 120$, $b = 92$, $\theta = \frac{a-b}{\sqrt{2(a+b)}} = 1.46385...$



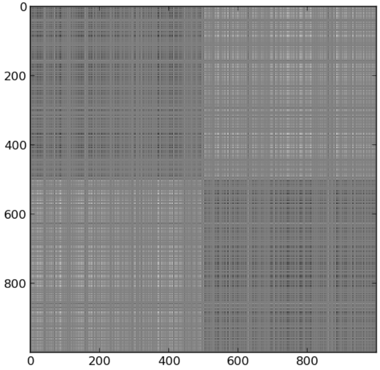
Spectral Algorithm

Spectrum of the original adjacency matrix. $a = 120$, $b = 92$,
$\theta = \frac{a-b}{\sqrt{2(a+b)}} = 1.46385...$



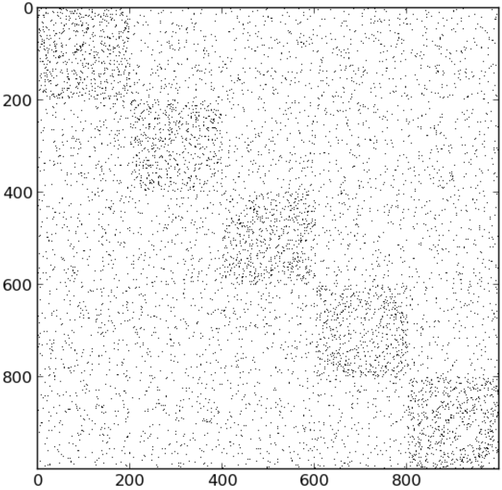
Spectral Algorithm

Rank-1 approximation of the adjacency matrix. $a = 120$, $b = 92$, $\theta = \frac{a-b}{\sqrt{2(a+b)}} = 1.46385...$



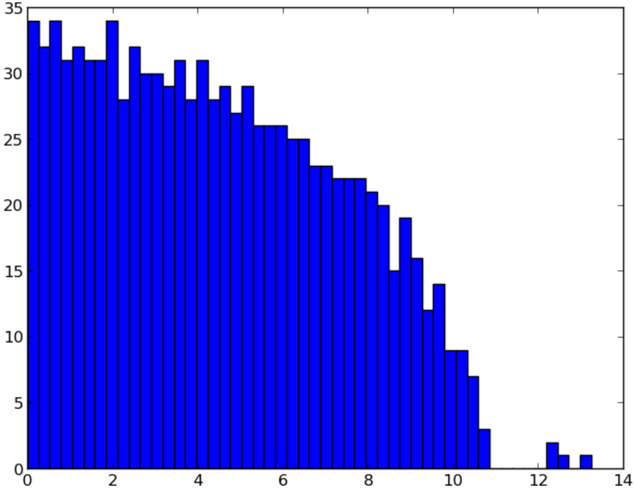
Spectral Algorithm: more communities

Original adjacency matrix with 5 communities.



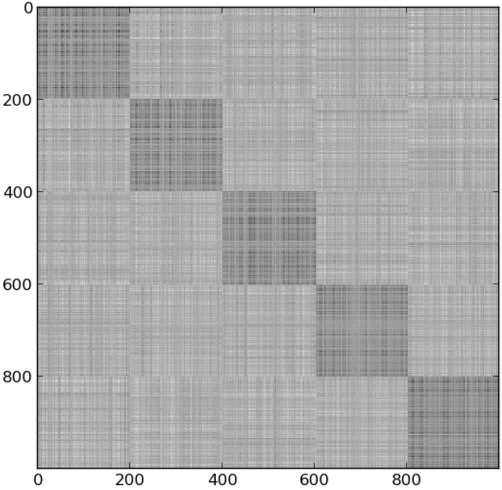
Spectral Algorithm: more communities

Spectrum of the original adjacency matrix.



Spectral Algorithm: more communities

Rank-4 approximation of the adjacency matrix.



Extension: r symmetric communities

Proposition

Assume $a \geq \ln^4 n$ and $r \geq 2$ symmetric communities. Then the clustering problem is solvable by the simple spectral method, provided

$$\frac{(a-b)^2}{r(a+(r-1)b)} > 1.$$

A parenthesis: Ramanujan graph

Spectral method perfoms well on matrices enjoying a spectral separation property.
For a d-regular graph G, the relaxation of the minimum bisection computes the second eigenvalue λ_2:

$$\max \sum_{(u,v)} \sigma_u A_{uv} \sigma_v$$

$$\text{s.t.} \sum_i \sigma_i = 0, \ \|\sigma\|_2 = 1.$$

G is Ramanujan if $\max_{|\lambda_i|<d} |\lambda_i| \leq \sqrt{d-1}$. Ramanujan graphs maximize the spectral gap.
Random d-regular graphs are Ramanujan Friedman '08
Erdős-Rényi graphs with average degree d are such that $\rho(A - dJ) \leq O(\sqrt{d})$ provided $d \succ \log n$ Feige Ofek '05

A parenthesis: Ramanujan graph

Spectral method perfoms well on matrices enjoying a spectral separation property.
For a d-regular graph G, the relaxation of the minimum bisection computes the second eigenvalue λ_2:

$$\max \sum_{(u,v)} \sigma_u A_{uv} \sigma_v$$

$$\text{s.t. } \sum_i \sigma_i = 0, \ \|\sigma\|_2 = 1.$$

G is Ramanujan if $\max_{|\lambda_i|<d} |\lambda_i| \leq \sqrt{d-1}$. Ramanujan graphs maximize the spectral gap.
Random d-regular graphs are Ramanujan Friedman '08
Erdős-Rényi graphs with average degree d are such that $\rho(A - dJ) \leq O(\sqrt{d})$ provided $d \succ \log n$ Feige Ofek '05

A parenthesis: Ramanujan graph

Spectral method perfoms well on matrices enjoying a spectral separation property.
For a d-regular graph G, the relaxation of the minimum bisection computes the second eigenvalue λ_2:

$$\max \sum_{(u,v)} \sigma_u A_{uv} \sigma_v$$
$$\text{s.t. } \sum_i \sigma_i = 0, \ \|\sigma\|_2 = 1.$$

G is Ramanujan if $\max_{|\lambda_i|<d} |\lambda_i| \leq \sqrt{d-1}$. Ramanujan graphs maximize the spectral gap.
Random d-regular graphs are Ramanujan Friedman '08
Erdős-Rényi graphs with average degree d are such that $\rho(A - dJ) \leq O(\sqrt{d})$ provided $d \succ \log n$ Feige Ofek '05

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.

One solution: trimming is working for the SBM. But what if the degree distribution is more skewed?

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.

One solution: trimming is working for the SBM. But what if the degree distribution is more skewed?

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.

One solution: trimming is working for the SBM. But what if the degree distribution is more skewed?

Non-backtracking matrix

Let $\vec{E} = \{(u, v); \{u, v\} \in E\}$ be the set of oriented edges, $m = |\vec{E}|$.
If $e = (u, v) \in \vec{E}$, we denote $e_1 = u$ and $e_2 = v$.

The non-backtracking matrix is an $m \times m$ matrix defined by

$$B_{ef} = 1(e_2 = f_1)1(e_1 \neq f_2)$$

B is NOT symmetric: $B^T \neq B$. We denote its eigenvalues by $\lambda_1, \lambda_2, \ldots$ with $\lambda_1 \geq \cdots \geq |\lambda_m|$.
Proposed by Krzakala et al. '14.

Connection with a multi-type branching process

Idea 1: iterating B counts the number of non-backtracking walks.

Stars (indeed trees) will have only zero as eigenvalues.

Idea 2: couple the local structure of the random graphs with a branching process.

Each individual has a $Poi(a/2)$ number of children of the same type and a $Poi(b/2)$ number of children from the opposite type.

Let $Z_t = (Z_t^+, Z_t^-)$ be the population at generation t.

Connection with a multi-type branching process

Idea 1: iterating B counts the number of non-backtracking walks.
Stars (indeed trees) will have only zero as eigenvalues.
Idea 2: couple the local structure of the random graphs with a branching process.
Each individual has a $Poi(a/2)$ number of children of the same type and a $Poi(b/2)$ number of children from the opposite type.
Let $Z_t = (Z_t^+, Z_t^-)$ be the population at generation t.

Connection with a multi-type branching process

Idea 1: iterating B counts the number of non-backtracking walks.
Stars (indeed trees) will have only zero as eigenvalues.
Idea 2: couple the local structure of the random graphs with a branching process.
Each individual has a $Poi(a/2)$ number of children of the same type and a $Poi(b/2)$ number of children from the opposite type.
Let $Z_t = (Z_t^+, Z_t^-)$ be the population at generation t.

Convergence of martingales

The mean progeny matrix

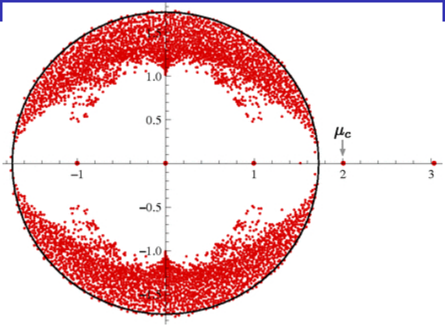
$$\frac{1}{2}\begin{pmatrix} a & b \\ b & a \end{pmatrix}$$

has eigenvalues $\alpha = \frac{a+b}{2}$ with eigenvector $\binom{1}{1}$ and $\beta = \frac{a-b}{2}$ with eigenvector $\binom{1}{-1}$.
The martingales

$$M_t = \frac{Z_t^+ + Z_t^-}{\alpha^t} \quad , \quad N_t = \frac{Z_t^+ - Z_t^-}{\beta^t}$$

converge a.s. and in L^2 as soon as $\beta^2 > \alpha$.
If $\beta^2 < \alpha$, then $\frac{Z_t^+ - Z_t^-}{\alpha^{t/2}}$ converges weakly to a random variable with finite variance.
Kesten Stigum '66

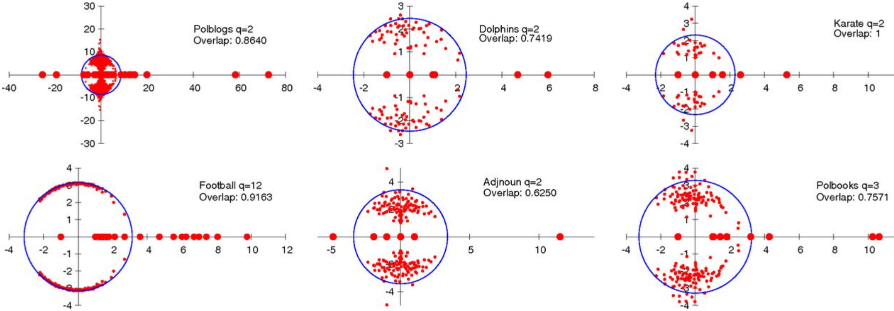
Spectrum of the non-backtracking matrix



If $\beta^2 > \alpha$, then there are two eigenvalues: $\lambda_1 = \alpha$ and $\lambda_2 = \beta$ out of the bulk $|\lambda_3| \leq \sqrt{\alpha} + o(1)$.

$$\beta^2 > \alpha \Leftrightarrow (a - b)^2 > 2(a + b).$$

The non-backtracking matrix on real data



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

Extensions

- For the labeled stochastic block model, we also conjecture a phase transition. We have partial results and an 'optimal' spectral algorithm.
- Some results for models with latent space allowing to relax the low-rank assumption and overlapping communities. If the signal strength is at least $\log n$, then consistent estimation of the edge label distribution is possible.
- Connections with the reconstruction problem on a tree and conjectures about computational complexity phase transition.

THANK YOU!

Extensions

- For the labeled stochastic block model, we also conjecture a phase transition. We have partial results and an 'optimal' spectral algorithm.
- Some results for models with latent space allowing to relax the low-rank assumption and overlapping communities. If the signal strength is at least $\log n$, then consistent estimation of the edge label distribution is possible.
- Connections with the reconstruction problem on a tree and conjectures about computational complexity phase transition.

THANK YOU!

Extensions

- For the labeled stochastic block model, we also conjecture a phase transition. We have partial results and an 'optimal' spectral algorithm.
- Some results for models with latent space allowing to relax the low-rank assumption and overlapping communities. If the signal strength is at least $\log n$, then consistent estimation of the edge label distribution is possible.
- Connections with the reconstruction problem on a tree and conjectures about computational complexity phase transition.

THANK YOU!

Extensions

- For the labeled stochastic block model, we also conjecture a phase transition. We have partial results and an 'optimal' spectral algorithm.
- Some results for models with latent space allowing to relax the low-rank assumption and overlapping communities. If the signal strength is at least $\log n$, then consistent estimation of the edge label distribution is possible.
- Connections with the reconstruction problem on a tree and conjectures about computational complexity phase transition.

THANK YOU!