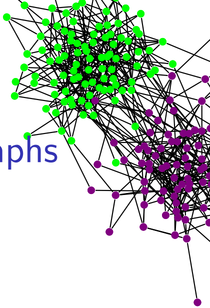
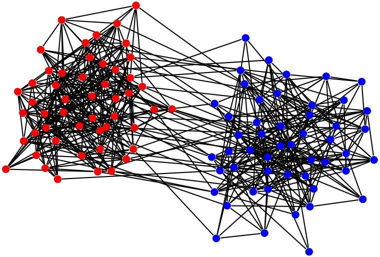


Matrix Concentration and Random Graphs

Sawyer Robertson

November 22, 2024

DOGSPOT Seminar



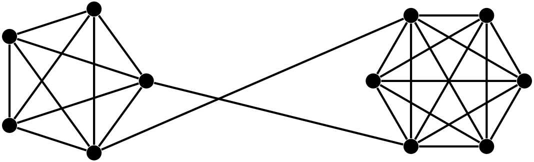
Our journey begins in 1973[1][2].

[1] William E Donath and Alan J Hoffman. "Lower bounds for the partitioning of graphs". In: *IBM Journal of Research and Development* 17.5 (1973), pp. 420–425.

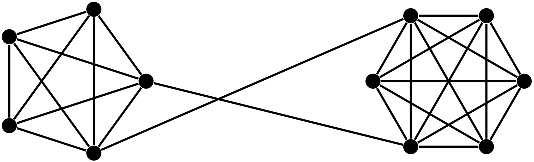
[2] Miroslav Fiedler. "Algebraic connectivity of graphs". In: *Czechoslovak mathematical journal* 23.2 (1973), pp. 298–305.

We are all pretty familiar with spectral clustering at this point.... . .

We are all pretty familiar with spectral clustering at this point.... . .



We are all pretty familiar with spectral clustering at this point.... . .



If G has nice community structure, we can recover it using the eigenvector(s) of the graph Laplacian.

Let $L = I - D^{-1/2}AD^{-1/2}$ be the normalized Laplacian.

Let $L = I - D^{-1/2}AD^{-1/2}$ be the normalized Laplacian.

$$\begin{aligned}\lambda_2 &= \min_{f \perp \mathbf{1}} \frac{f^T L f}{f^T f} \\ &= \min_{f \perp \mathbf{1}} \frac{\sum_{\{i,j\}} |f_i - f_j|^2}{\sum_i |f_i|^2 d_i}.\end{aligned}$$

Let $L = I - D^{-1/2} A D^{-1/2}$ be the normalized Laplacian.

$$\begin{aligned} \lambda_2 &= \min_{f \perp \mathbf{1}} \frac{f^T L f}{f^T f} \\ &= \min_{f \perp \mathbf{1}} \frac{\sum_{\{i,j\}} |f_i - f_j|^2}{\sum_i |f_i|^2 d_i}. \end{aligned}$$

The argmin which we write u_2 solves

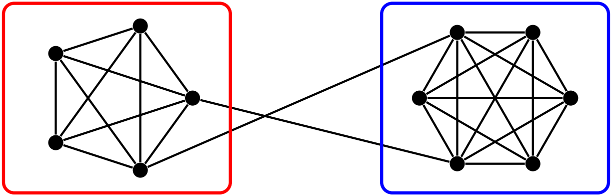
$$L u_2 = \lambda_2 u_2.$$

$$\mathrm{vol}(A, B) = \#\{\{i, j\} : i \in A, j \in B\}$$

$$\mathrm{vol}(A, B) = \#\{\{i, j\} : i \in A, j \in B\}$$

We have the normalized cut problem:

$$\min_{A,B \subseteq V} \frac{\mathrm{vol}(A, B)}{\mathrm{vol}(A, V)} + \frac{\mathrm{vol}(A, B)}{\mathrm{vol}(B, V)}$$



$\{u_2 < 0\}$ $\{u_2 > 0\}$

The sets $\{u_2 > 0\}$ and $\{u_2 < 0\}$ solve a relaxation of this problem.[3]

[3] Jianbo Shi and Jitendra Malik. “Normalized cuts and image segmentation”. In: *IEEE Transactions on pattern analysis and machine intelligence* 22.8 (2000), pp. 888–905.

But is this *consistent* in a structural sense?

But is this *consistent* in a structural sense?

Suppose we have two (latent, planted) communities $A, B \subseteq V$ obtained from some model, and we put:

$$\widehat{A} = \{u_2 < 0\}$$
$$\widehat{B} = \{u_2 > 0\}$$

Can we guarantee $\widehat{A}, A$ are close w.h.p.? Similarly for $\widehat{B}$.

This is a very difficult question because it requires pretty precise knowledge of the **entries** of u_2 in a random graph model[4]...

[4] Ulrike von Luxburg, Mikhail Belkin, and Olivier Bousquet. "Consistency of Spectral Clustering". In: *The Annals of Statistics* 36.2 (Apr. 2008).

This is a very difficult question because it requires pretty precise knowledge of the **entries** of u_2 in a random graph model[4]...

To get to this destination, we start from scratch.

[4] Luxburg, Belkin, and Bousquet, "Consistency of Spectral Clustering".

This is a very difficult question because it requires pretty precise knowledge of the **entries** of u_2 in a random graph model[4]...

To get to this destination, we start from scratch.

What about A?

[4] Luxburg, Belkin, and Bousquet, "Consistency of Spectral Clustering".

Definition (Inhomogeneous ER random graph)

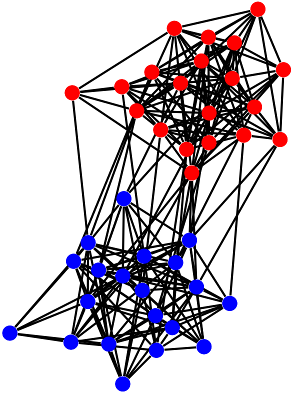
Let $n \geq 2$ and $P \in \mathbb{R}^{n\times n}$ be an $n \times n$ symmetric matrix of probabilities $p_{ij} \in [0,1]$. We assume $p_{ii} = 0$. We construct a random graph G as follows. Let G have vertex set $[n] = \{1, 2, \ldots, n\}$ and, for each $e = \{i, j\} =\in \binom{[n]}{2}$, we add the e to G with probability p_{ij}. We say $G \sim \mathrm{G}(n, P)$.

Example (Planted Community Model)

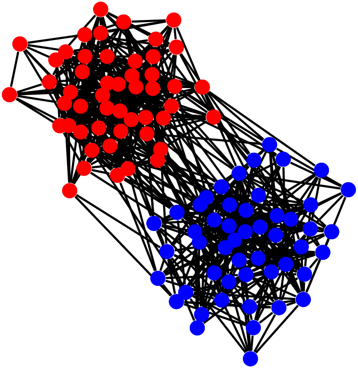
Let $A = \{1, \ldots, n\}$ and $B = \{n+1, \ldots, 2n\}$ for some $n \geq 1$. Let $p, q \in [0,1]$ and choose by convention $p \geq q$. Set:

$$p_{ij} = \begin{cases} p & \text{if } i,j \in A \text{ or } i,j \in B \\ q & \text{otherwise} \end{cases}$$

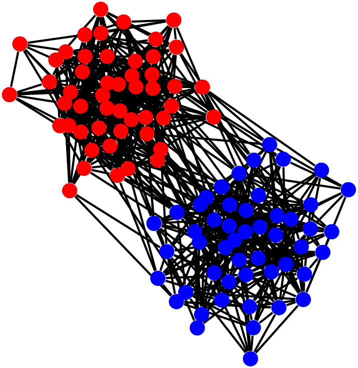
Also called the stochastic block model.



$n = 20,\ p = 0.5,\ q = 0.05$



$n = 50,\ p = 0.25,\ q = 0.025$



$n = 50, \; p = 0.25, \; q = 0.025$

We can see this is a natural petri dish for community detection

Other examples of inhomogeneous ER graphs include:

- Chung-Lu expected degree graphs
- Random dot product graphs
- ...

For $A \sim \mathrm{G}(n, P)$, set $\overline{A}$:

$$\overline{A}_{ij} = \mathbb{E}\left(A_{ij}\right) = p_{ij}$$

$$\overline{D}_{ii} = \sum_{j} \overline{A}_{ij} =: \overline{d}_i$$

For $A \sim \mathrm{G}(n, P)$, set $\overline{A}$:

$$\overline{A}_{ij} = \mathbb{E}\left(A_{ij}\right) = p_{ij}$$
$$\overline{D}_{ii} = \sum_j \overline{A}_{ij} =: \overline{d}_i$$

and $\overline{L} = I - \overline{D}^{-1/2}\overline{AD}^{-1/2}$.

For $A \sim \mathrm{G}(n, P)$, set $\overline{A}$:

$$\overline{A}_{ij} = \mathbb{E}\left(A_{ij}\right) = p_{ij}$$
$$\overline{D}_{ii} = \sum_{j} \overline{A}_{ij} =: \overline{d}_i$$

and $\overline{L} = I - \overline{D}^{-1/2}\overline{AD}^{-1/2}$.

These are adjacency and Laplacian matrices of weighted graphs.

We ask:

[5] Always operator norm.

We ask:

When can we guarantee that $\|A - \overline{A}\|$ is small with high probability?[5]

[5] Always operator norm.

We ask:

When can we guarantee that $\|A - \overline{A}\|$ is small with high probability?[5]

The answer: roughly $O(\sqrt{\log n})$ provided the model is "not too sparse."

[5] Always operator norm.

Our journey picks up in 2009...

[6] Roberto Imbuzeiro Oliveira. "Concentration of the adjacency matrix and of the Laplacian in random graphs with independent edges". In: *arXiv preprint arXiv:0911.0600* (2009).

Our journey picks up in 2009...

Let $d = \min_i \overline{d}_i$, $\Delta = \max_i \overline{d}_i$.

[6] Oliveira, "Concentration of the adjacency matrix and of the Laplacian in random graphs with independent edges".

Our journey picks up in 2009...

Let $d = \min_i \overline{d}_i$, $\Delta = \max_i \overline{d}_i$.

Theorem (Oliveira (2009)[6])

For any $c > 0$ there exist $C = C(c) > 0$, independent of n, P, such that the following holds. If $\Delta > C \ln n$, then for all $n^{-c} \leq \delta \leq 1/2$,

$$\|A - \overline{A}\| \leq 4 \sqrt{\Delta \ln(n/\delta)}$$

w.p. at least $1 - \delta$. If $d \geq C \ln n$, then for the same range of δ:

$$\|L - \overline{L}\| \leq 14 \sqrt{\frac{\ln(4n/\delta)}{d}}$$

w.p. at least $1 - \delta$.

[6] Oliveira, "Concentration of the adjacency matrix and of the Laplacian in random graphs with independent edges".

Chung and Radcliffe improve this two years later...

[7] Fan Chung and Mary Radcliffe. "On the spectra of general random graphs". In: *the electronic journal of combinatorics* (2011), P215–P215.

Chung and Radcliffe improve this two years later...

Theorem (Chung and Radcliffe (2011)[7])

Let $\epsilon > 0$ *be fixed and suppose that for* n *sufficiently large, it holds* $\Delta > \frac{4}{9}\log(2n/\epsilon)$. *Then:*

$$\|A - \overline{A}\| \leq 4\sqrt{\Delta \log(2n/\epsilon)}$$

w.p. at least $1 - \epsilon$. *Moreover there exists* $k = k(\epsilon)$ *such that if* $d \geq k \log n$, *then*

$$\|L - \overline{L}\| \leq 3\sqrt{\frac{3\log(4n/\epsilon)}{d}}$$

w.p. at least $1 - \epsilon$.

[7] Chung and Radcliffe, "On the spectra of general random graphs".

The proof of Chung and Radcliffe's version uses the following matrix Bernstein inequality.

[8] Joel A Tropp. "User-friendly tail bounds for sums of random matrices". In: *Foundations of computational mathematics* 12 (2012), pp. 389–434.

The proof of Chung and Radcliffe's version uses the following matrix Bernstein inequality.

Theorem (Matrix Bernstein inequality[8])

Let $B_1, B_2, \ldots, B_k$ *be independent* $n \times n$ *random Hermitian matrices. Assume that* $\|B_i - \mathbb{E}(B_i)\| \leq M$ *for all* i *and set* $\nu = \|\sum_{i=1}^{k} \mathrm{Var}(B_i)\|$. *Writing* $B = \sum_{i=1}^{k} B_i$, *we have that for any* $a > 0$,

$$\mathbb{P}(\|B - \mathbb{E}(B)\| > a) \leq 2n \exp\left\{-\frac{a^2}{2\nu + 2Ma/3}\right\}. \qquad (1.1)$$

[8] Tropp, "User-friendly tail bounds for sums of random matrices".

The proof of Chung and Radcliffe's version uses the following matrix Bernstein inequality.

Theorem (Matrix Bernstein inequality[8])

Let $B_1, B_2, \ldots, B_k$ *be independent* $n \times n$ *random Hermitian matrices. Assume that* $\|B_i - \mathbb{E}(B_i)\| \leq M$ *for all* i *and set* $\nu = \|\sum_{i=1}^{k} \mathrm{Var}(B_i)\|$. *Writing* $B = \sum_{i=1}^{k} B_i$, *we have that for any* $a > 0$,

$$\mathbb{P}(\|B - \mathbb{E}(B)\| > a) \leq 2n \exp\left\{-\frac{a^2}{2\nu + 2Ma/3}\right\}. \quad (1.1)$$

[8] Tropp, "User-friendly tail bounds for sums of random matrices".

A quick detour. What does $\|A - \overline{A}\|$ small get us?

A quick detour. What does $\|A - \overline{A}\|$ small get us?

Theorem (Weyl's inequality)

Let $B, C \in \mathbb{C}^{n \times n}$ be Hermitian matrices with eigenvalues $\lambda_k(\cdot)$ ordered ascending. Then

$$|\lambda_k(A) - \lambda_k(B)| \leq \|A - B\|.$$

Usually $\lambda_k(\overline{A})$ and $\lambda_k(\overline{L})$ are easy to obtain.

For the planted community model, if $d = d_i = (n-1)p + nq$,

$$\overline{L} = \frac{1}{d}\begin{bmatrix} d & -p & \cdots & -p & -q & & \cdots & -q \\ -p & d & \cdots & -p & -q & & \cdots & -q \\ \vdots & \vdots & \ddots & \vdots & \vdots & & \ddots & \vdots \\ -p & -p & \cdots & d & -q & & \cdots & -q \\ -q & -q & \cdots & -q & d & -p & \cdots & -p \\ -q & -q & \cdots & -q & -p & d & \cdots & -p \\ \vdots & \vdots & \ddots & \vdots & \vdots & & \ddots & \vdots \\ -q & -q & \cdots & -q & -p & -p & \cdots & d \end{bmatrix}$$

$$\lambda_k(\overline{L}) = 0, \frac{2nq}{d}, \frac{d+p}{d} \quad \times (2n-1)$$

$$u_2 \propto [\mathbf{1}_n^T - \mathbf{1}_n^T]^T$$

Theorem (Davis-Kahan Theorem[9])

Let $B, \hat{B} \in \mathbb{R}^{n \times n}$ *be symmetric, with eigenvalues* $\lambda_1 \leq \ldots \leq \lambda_n$ *and* $\hat{\lambda}_1 \leq \ldots \leq \hat{\lambda}_n$ *respectively. Fix* $j \in \{1, \ldots, n\}$, *and assume that* $\min(\lambda_j - \lambda_{j-1}, \lambda_{j+1} - \lambda_j) > 0$, *where* $\lambda_0 := -\infty$ *and* $\lambda_{n+1} := \infty$. *If* $v, \hat{v} \in \mathbb{R}^n$ *satisfy* $Bv = \lambda_j v$ *and* $\hat{B}\hat{v} = \hat{\lambda}_j \hat{v}$, *then*

$$\sin \Theta(\hat{v}, v) \leq \frac{2\|\hat{B} - B\|}{\min(\lambda_{j-1} - \lambda_j, \lambda_j - \lambda_{j+1})}.$$

Moreover, if $\hat{v}^T v \geq 0$, *then*

$$\|\hat{v} - v\| \leq \frac{2^{3/2}\|\hat{B} - B\|}{\min(\lambda_{j-1} - \lambda_j, \lambda_j - \lambda_{j+1})}.$$

---

[9] Yi Yu, Tengyao Wang, and Richard J Samworth. "A useful variant of the Davis–Kahan theorem for statisticians". In: *Biometrika* 102.2 (2015), pp. 315–323.

Theorem (Davis-Kahan Theorem[9])

Let $B, \hat{B} \in \mathbb{R}^{n\times n}$ be symmetric, with eigenvalues $\lambda_1 \leq \ldots \leq \lambda_n$ and $\hat{\lambda}_1 \leq \ldots \leq \hat{\lambda}_n$ respectively. Fix $j \in \{1, \ldots, n\}$, and assume that $\min(\lambda_j - \lambda_{j-1}, \lambda_{j+1} - \lambda_j) > 0$, where $\lambda_0 := -\infty$ and $\lambda_{n+1} := \infty$. If $v, \hat{v} \in \mathbb{R}^n$ satisfy $Bv = \lambda_j v$ and $\hat{B}\hat{v} = \hat{\lambda}_j \hat{v}$, then

$$\sin \Theta(\hat{v}, v) \leq \frac{2\|\hat{B} - B\|}{\min(\lambda_{j-1} - \lambda_j, \lambda_j - \lambda_{j+1})}.$$

Moreover, if $\hat{v}^T v \geq 0$, then

$$\|\hat{v} - v\| \leq \frac{2^{3/2}\|\hat{B} - B\|}{\min(\lambda_{j-1} - \lambda_j, \lambda_j - \lambda_{j+1})}.$$

So we can begin to get a glimpse of our destination...

[9] Yu, Wang, and Samworth, "A useful variant of the Davis–Kahan theorem for statisticians".

In 2014, Lei and Rinaldo provide what is probably the strongest matrix concentration inequality for this model.

[10] Jing Lei and Alessandro Rinaldo. "Consistency of spectral clustering in stochastic block models". In: *The Annals of Statistics* (2015), pp. 215–237.

In 2014, Lei and Rinaldo provide what is probably the strongest matrix concentration inequality for this model.

Theorem (Lei and Rinaldo[10])

Let $r > 0$ be fixed. Assume that $n \max_{ij} p_{ij} \leq s$ for $s \geq c \log n$ for some $c > 0$. Then there exists a constant $C = C(r, c)$ such that

$$\|A - \overline{A}\| \leq C\sqrt{s}$$

with probability at least $1 - n^{-r}$.

[10] Lei and Rinaldo, "Consistency of spectral clustering in stochastic block models".

In 2014, Lei and Rinaldo provide what is probably the strongest matrix concentration inequality for this model.

Theorem (Lei and Rinaldo[10])

Let $r > 0$ be fixed. Assume that $n \max_{ij} p_{ij} \leq s$ for $s \geq c \log n$ for some $c > 0$. Then there exists a constant $C = C(r, c)$ such that

$$\|A - \overline{A}\| \leq C\sqrt{s}$$

with probability at least $1 - n^{-r}$.

Wonderfully clean theorem. The proof? Another story (all things considered not too bad). The catch?

[10] Lei and Rinaldo, "Consistency of spectral clustering in stochastic block models".

In 2014, Lei and Rinaldo provide what is probably the strongest matrix concentration inequality for this model.

Theorem (Lei and Rinaldo[10])

Let $r > 0$ be fixed. Assume that $n \max_{ij} p_{ij} \leq s$ for $s \geq c \log n$ for some $c > 0$. Then there exists a constant $C = C(r, c)$ such that

$$\|A - \overline{A}\| \leq C\sqrt{s}$$

with probability at least $1 - n^{-r}$.

Wonderfully clean theorem. The proof? Another story (all things considered not too bad). The catch?

No L.

[10] Lei and Rinaldo, "Consistency of spectral clustering in stochastic block models".

Finally, we reach 2020 with a paper of Deng, Ling, amd Strohmer which puts the matter to rest.

[11] Shaofeng Deng, Shuyang Ling, and Thomas Strohmer. "Strong consistency, graph laplacians, and the stochastic block model". In: *Journal of Machine Learning Research* 22.117 (2021), pp. 1–44.

Finally, we reach 2020 with a paper of Deng, Ling, amd Strohmer which puts the matter to rest.

Theorem (Deng, Ling, amd Strohmer[11])

Assume that $n \max_{ij} p_{ij} \geq c \log n$ *for some* $c \geq 1$. *Then for any* $r > 0$, *there exists* $C = C(c, r)$ *such that*

$$\|L - \overline{L}\| \leq \frac{C(n \max_{ij} p_{ij})^{5/2}}{\min\{d_{\min}, d\}^3}$$

with probability at least $1 - n^{-r}$. *Here* $d_{\min}$ *is the minimum degree of* A.

[11] Deng, Ling, and Strohmer, "Strong consistency, graph laplacians, and the stochastic block model".

Finally, we reach 2020 with a paper of Deng, Ling, amd Strohmer which puts the matter to rest.

Theorem (Deng, Ling, amd Strohmer[11])

Assume that $n \max_{ij} p_{ij} \geq c \log n$ *for some* $c \geq 1$*. Then for any* $r > 0$*, there exists* $C = C(c, r)$ *such that*

$$\|L - \overline{L}\| \leq \frac{C(n \max_{ij} p_{ij})^{5/2}}{\min\{d_{\min}, d\}^3}$$

with probability at least $1 - n^{-r}$*. Here* $d_{\min}$ *is the minimum degree of* A*.*

This bound needs to be de-randomized to be useful but is strong otherwise.

[11] Deng, Ling, and Strohmer, "Strong consistency, graph laplacians, and the stochastic block model".

In the same paper, we reach our destination...

[12] Deng, Ling, and Strohmer, "Strong consistency, graph laplacians, and the stochastic block model".

In the same paper, we reach our destination...

Theorem (Deng, Ling, amd Strohmer[12])

Let $p = \alpha \frac{\log n}{n}$, $q = \beta \frac{\log n}{n}$ *and assume* $\sqrt{\alpha} - \sqrt{\beta} > \sqrt{2}$. *Then there exists* $\eta = \eta(\alpha, \beta) > 0$ *and* $\sigma \in \{\pm 1\}$ *such that with probability* $1 - o(1)$,

$$\sqrt{2n}(\sigma u_2)_i \geq \eta \text{ for } i \leq n$$

and

$$\sqrt{2n}(\sigma u_2)_i \leq -\eta \text{ for } i \geq n+1.$$

[12] Deng, Ling, and Strohmer, "Strong consistency, graph laplacians, and the stochastic block model".