The Non-Backtracking Matrix

Q1: Recall that for the stochastic block model $G(n, \frac{\eta}{n}, \frac{\zeta}{n})$ and fixed $\alpha > \beta$, the "signal eigenvalue" $\frac{\alpha - \beta}{2}$ is lost in the bulk of the spectrum as $n \to \infty$.

(Note: Any fixed vertex has Binom($n-1, \frac{\alpha}{n}$) + Binom($n-1, \frac{\beta}{n}$) $\sim$ Poisson($d$) neighbors, where $d = \frac{\alpha + \beta}{2}$. Each neighbor has conditionally $\sim$ Pois($d$) new neighbors, etc. Thus $G(n, \frac{\eta}{n}, \frac{\zeta}{n})$ has the same local weak limit, and the same limit ESD for its adjacency matrix as Erdős-Rényi $G(n, \frac{\zeta}{n})$. This limit ESD has unbounded support.)

Is there a spectral method that can identify the communities?

Q2: Recall that the random $d$-regular graph has limit ESD for its adjacency matrix given by the Kesten-McKay law. This is a density supported on $[-2\sqrt{d}, 2\sqrt{d}]$. For $d > 3,$ the largest eigenvalue $\lambda_1$ does not converge to this support. Let $X = (1, \ldots, 1)$. Then

$$AX = \lambda X$$

$A$ is the graph is $d$-regular. So $\lambda_1$ is always an eigenvalue of $A$.

A $d$-regular graph is called Ramanujan if

$$\lambda^2 \leq 2\sqrt{d-1} \quad \text{and} \quad \lambda_n \geq -2\sqrt{d-1}.$$ 

Is it true that for a random $d$-regular graph, $\lambda_2 \to 2\sqrt{d-1}$ and
Definition: The non-backtracking matrix $B_{uv}$ of a graph $G = (V, E)$ is the matrix indexed by directed edges $u \rightarrow v$.

$$B_{uv} = \begin{cases} 1 & \text{if } u \rightarrow v \in E \setminus E' \\ 0 & \text{otherwise} \end{cases}$$

Note that $B$ is not symmetric: $B_{uv} \neq B_{vu}$. The eigenvalues $\lambda$ are complex-valued.

Then (Bordenave, Lelarge, Massoulié '18): Assume $d > 1$.

(a) For Erdős-Rényi $G(n, d/n)$, the non-backtracking matrix $B$ has one eigenvalue converging to $0$, and remaining eigenvalues in the disk $|\lambda| < 1$; why? For any fixed $\epsilon > 0$ and large $n$.

(b) For the SBM $G(n, d/n)$ if $d < \frac{\lambda}{2}$ where $\lambda = \frac{\epsilon}{2}$, then $B$ has two eigenvalues converging to $1$ and $\lambda$, respectively, and remaining eigenvalues in the above disk, why? For large $n$.

If $G$ has a vertex $v$ of high degree, $v \rightarrow A^k v$ may be dominated by paths alternating between $v$ and its neighbors. The number of such paths is $(\deg(v))^k$ for any $k$.

Hence, non-backtracking graphs cannot do this, and must visit many distinct vertices unless $G$ has short cycles.

If $G$ is a tree, there is at most one non-backtracking path from $v$ to $v'$, which is the shortest path. For locally tree-like graphs, the corresponding eigenvector $\xi$ is such that $\xi \in \mathbb{R}^m$ defined by

$$\xi_v = \begin{cases} 1 & \text{if } v \text{ is the root} \\ 0 & \text{otherwise} \end{cases}$$
Remark: For two communities of equal size, the threshold $\frac{\alpha}{\kappa^2} > \frac{1}{2}$ is the optimal one for detection and rank recovery. (Merely Necessary, Sly ’12)

Bordenave et al establish a general version of the preceding theorem for any fixed number k of communities as $n \to \infty$.

Intuition: Let $\mathbf{S}_{uv} = (\alpha \kappa)^{-1} \sum_{i=1}^{\alpha \kappa} \mathbf{O}_{ii}$. (Don’t depend on $u$)

This is an eigenvector of $B$ with eigenvalue $\frac{\alpha}{\kappa}$.

\[
\mathbf{S}^{(k)}_{uv} = \sum_{i=1}^{k \alpha \kappa} \mathbf{B}^{(k)}_{uv} = (\alpha \kappa)^{-1} \sum_{i=1}^{\alpha \kappa} \sum_{t=1}^{k} \mathbf{O}_{ii} = (\alpha \kappa)^{-1} \sum_{i=1}^{\alpha \kappa} \mathbf{O}_{ii} = \frac{\alpha}{\kappa} \mathbf{S}^{(0)}_{uv}
\]

For $\frac{\alpha}{\kappa} > \frac{1}{2}$ and large $k$, we can show $\mathbf{S}^{(k)}_{uv} \approx \mathbf{S}^{(0)}_{uv}$. Similarly, $\mathbf{S}^{(k)}_{ii} = (\alpha \kappa)^{-1} \sum_{i=1}^{\alpha \kappa} \mathbf{O}_{ii}$ is an approximate eigenvector of $\mathbf{O}$, with eigenvalue $\frac{\alpha}{\kappa}$.

If $M_{\mathbf{O}, \mathbf{M}^{-1}}$ is any (non-symmetric) matrix with eigenvalues $\lambda$, $\lambda^*$ and singular values $\sigma_1 > \sigma_2 > \cdots > \sigma_m$, then

\[
\sum_{t=1}^{m} \frac{1}{\sigma_t^2} \leq \frac{\alpha}{\kappa} = \operatorname{Tr} \mathbf{M} \mathbf{M}^T.
\]

Letting $\lambda, \lambda^*$ be the eigenvalues of $B$, for any $k > q$,

\[
\frac{1}{\lambda} \left( \frac{1}{\kappa} \right)^k \left( \frac{1}{\alpha \kappa} \right)^{k^2} \leq \frac{1}{\lambda^*} \left( \frac{1}{\kappa} \right)^k \left( \frac{1}{\alpha \kappa} \right)^{k^2} \leq \frac{1}{\lambda} \operatorname{Tr} B^k \mathbf{B}^k.
\]

We have $\left( \mathbf{B}^k \mathbf{B}^k \right)_{uv} = \sum_{s,t=1}^{k \alpha \kappa} \mathbf{B}^{(k)}_{uv} = \sum_{s,t=1}^{\alpha \kappa} \mathbf{O}_{ss} \mathbf{O}_{tt} = \#(s=t)$ which $\kappa$ of them from $(uv)$.

On expectation this is $\frac{\alpha^2}{\kappa}$. So for large $k$ and fixed $k$, we expect

\[
\frac{1}{\alpha \kappa} B^k \mathbf{B}^k \approx \frac{\alpha^2}{\kappa} \mathbf{I} \mathbf{I} = \alpha \kappa \mathbf{I}.
\]

Bordenave, Lelarge, Masek show that all but two entries are in this link.

Theorem (Friedman ’08): Fix $d \geq 2$. A uniformly random $d$-regular graph has $\lambda \approx 2d^{-1/2}$ and $\lambda \approx 2d^{-1}$. For any fixed $d / 2$, taking $\lambda^*$ when $\lambda \approx 2d^{-1}$, are the eigenvalues of its adjacency matrix.

The proof uses the spectral method and a Furedi-Komlos-type argument. It is easier to analyze $\operatorname{Tr} B^k$, then to analyze $\operatorname{Tr} A^k$.

Prop (Johannsson Identity): For any graph $G = (V, E)$ with $n$ vertices on (undirected) edges, letting $A$, $D$, and $B$ be the adjacency, diagonal degree, and non-backtracking matrices,

\[
\det (I - \lambda \mathbf{B}) = (1 - \lambda)^{n-1} \det (I - \lambda \mathbf{A} + \lambda^2 \mathbf{D} - 2).
\]

Hence: Eigenvalues $\lambda = \frac{\kappa}{\alpha}$ of $B$ are (asymptotically)

\[
\pm 1, -1 \text{ each with multiplicity } m n.
\]

The $2m$ roots of $0 \equiv \det (I - \lambda \mathbf{A} + \lambda^2 \mathbf{D} - 2) \equiv \det (X^2 - A + D - 2)$.

For a regular graph, $D = \mathbf{I}$. Thus

\[
\lambda \text{ is an eigenvalue } \Rightarrow 1 \pm \sqrt{\lambda} \text{ is an eigenvalue } \lambda \text{ of } \mathbf{A}.
\]

If $A$ has an eigenvalue $\mu$ with $|\mu| > 2 \sqrt{d}$, then $B$ has two real eigenvalues at $\pm \frac{1}{2} \frac{\mu - 2 \sqrt{d}}{2} = \pm \sqrt{d}$.
In particular, the eigenvalue \( \lambda \) corresponds to \( \{1, d, \ldots, d\} \) for \( B \) and Friedman's Theorem may be restated as:

Then let \( B \) be the non-backtracking matrix of a random \( d \)-regular graph.
Then \( B \) has a single eigenvalue equal to \( d+1 \), and removing eigenvalues in the disc \( \|z\| < 1 \) for any fixed \( \epsilon > 0 \) when \( d \to \infty \).

### Proof of Ihara-Bass

Let \( S \in \mathbb{R}^{n \times n} \), \( T \in \mathbb{R}^{n \times n} \) be defined by:

\[
S_{uvw} = \begin{cases} 1 & \text{if } \{u, v, w\} \in E \\ 0 & \text{otherwise} \end{cases} \quad T_{uvw} = \begin{cases} 1 & \text{if } \{u, v, w\} \in E \\ 0 & \text{otherwise} \end{cases}
\]

So \( S^{2n+1} = (S^3) = \sum_{w 	ext{ mediate}} S_{uwv} \), \( S^{n+1} = (S^3T) = \sum_{w \text{ mediate}} S_{uwv} \).

Define the permutation \( \pi \in \mathbb{R}^{n \times n} \) by:

\[
P_{u,v,w} = \begin{cases} 1 & \text{if } u, v, w \end{cases} 0 \text{ otherwise.}
\]

So \( S^{2n+1} = S^{n+1} = S^{2} \) and \( P^2 = I \). We have:

\[
(\pi \pi)_{uvw} = \sum_{z \text{ mediate}} S_{uwv}, \quad (\pi \pi)_{uvw} = \sum_{z \text{ mediate}} S_{uwv}.
\]

So \( \pi \pi = \pi \pi \).

### Proof Idea of Friedman's Theorem (Following Bedovran '95)

1. Simplification to configuration model: Link all half-edges to each vertex.

Construct \( G \) by randomly pairing the half-edges. This yields a \( d \)-regular multigraph, possibly with:

- self-loops
- repeated edges between a vertex pair \( (u, v) \).

Conditional on \( S \), the distribution \( G \) is uniformly random over all \( d \)-regular graphs.

Prop (Bulldog '01): \( M_{S} \to e^{-\lambda^2/4} \pi_{n \to \infty} \).

If result holds when \( G \) is this multigraph \( G \), then it also holds only conditional on \( S \to \).

2. Path-counting and covering: Recall \( \Xi(k) \) \& \( \Xi_k(k) \).

Here:

\[
\Xi_k(k) = \sum_{x} \text{Be}_x \cdots \text{Be}_x \cdot \text{Be}_k \cdot \text{Be}_k \cdot \text{Be}_k \cdot \text{Be}_k
\]

\[
\Xi(k) = \sum_{x} \text{Be}_x \cdots \text{Be}_x
\]

\[
\Xi(k)_k = \sum \text{Be}_x \cdots \text{Be}_x \cdot \text{Be}_k \cdot \text{Be}_k \cdot \text{Be}_k \cdot \text{Be}_k
\]

\[
\Xi(k) = \# \text{ pairs of non-backtracking paths of half edges}
\]

starting and ending at shared directed edges.
This includes the contribution from $X_i = d/4$ which corresponds to the left
left and right) eigenvalue $X: (-1, 0) \in \mathbb{R}^d$; \( B_i = \{X: \mathbb{R}^d \}, \ X_i = \{X: \mathbb{R}^d \} \).

To remove this contribution, we center $B_i$:

- Index both the half-edges and directed edges of $G$ by $\text{idx}(i)$: $\{X_{i1}, X_{i2}, X_{i3}, \ldots, X_{id}\}$

Then $B_{i}(\text{idx}(i)) = \{X: \text{idx}(i) \}$ if $X$ is joined with $(X, k)$ for some $k$.

\[
\sum_{k \in \mathbb{Z}} M_{X}(X, k) \text{ when } M_{X}(X, k) \text{ is an edge between $X$ and $(X, k)$}.
\]

(We take this as the definition of $B_i$ when $G$ has repeated edges and loops.)

Define $M_{X}(X, k) = M_{X}(X, k) - \frac{1}{d}$ as the centered version of $M_{X}$.

- If we set $B_{i}(\text{idx}(i)) = \sum_{k \in \mathbb{Z}} M_{X}(X, k) = B_{X} - \frac{1}{d}$, then for every vertex $X$ in the ball $B_{i}(X)$ of radius $i$.

Prop: If $\epsilon = \lambda(G) \geq \epsilon \log(d-1)$, then $\text{IP}[G] = \epsilon$-tangle-free $\Rightarrow$.

Proof sketch: Fix $v$. Pair the $d$ half-edges at $v$ one at a time, then the last one. $d-1$ half-edges of each neighbor at a time, etc., until we've paired all half-edges from level $d-1$. The number of times we do a pairing is at most $1/\epsilon^2$, which is at most $T = \sum_{k \in \mathbb{Z}} d(\epsilon-1)^k \leq 3(d-1)^k$.

Each pairing, we create a new cycle with probability at most
\[
\text{IP}[G] \leq \frac{1/\epsilon^2}{d(\epsilon-1)^k} \leq 0.99 \frac{n}{1/d}.
\]

So the number of cycles is $\leq \text{Bin}(T, (\epsilon-1)^k)$.

Then $\text{IP}[\epsilon \geq \epsilon \log(d-1)] \geq \epsilon \log(d-1) - \epsilon \log d$, taking a union

1. In the Furedi-Kalai argument, we used Markov's inequality and bounded $\text{IP}[W^{(k)}]$. Recall we needed to take $\epsilon = \lambda(G) \geq \epsilon \log(d-1)$, which $\lambda(G)$.

2. For a constant $\gamma$, we prob. at least $\gamma \omega_i$ G contains a clique of $d+1$ vertices.

Then on this event, this is a contribution of at least

\[
\sum_{k \in \mathbb{Z}} (d-1)^k \leq 3(d-1)^k.
\]
Refined argument: Let $k = \log n$, $m = \left\lfloor \frac{n}{2^k} \right\rfloor$, $k \leq \log n$.

To bound $\lambda^{2k}$, write $\lambda^{2k} = \|B^{(k)}\|_2^2$.

Note that

$$B^{(k)(u_2 \cdots u_{2k})} = \sum_{t} M_{u_2 \cdots u_{2k}}^t \cdots M_{u_{2k} \cdots u_{2k}}^t$$

where the sum is over all $(2k)$-tuples of half-edges $(y_k = (u_i, v_i) : i = 1, 2k)$ such that

- $u_2 \neq v_2, u_3 \neq v_3, \ldots, u_{2k} \neq v_{2k}$
- $\delta_{2k} = (u_2, v_2)$ and $\delta_{2k} = (u_2, v_2)$.

Let $G$ be the subgraph containing edges $(u_2, v_2), (u_3, v_3), \ldots, (u_{2k}, v_{2k})$.

Let $B^{(k)} = \sum_{\delta} M_{\delta}$ be his $(2k)$ cycle.

Lemma: There exists $x^{(k)} \neq 0$ such that

- If $x^{(k)} = 0$ and $G$ is $2k$-regular, then $B^{(k)} x^{(k)} = B x^{(k)} = B^{(k)} x^{(k)} = x^{(k)}$, and
- $\|x^{(k)}\|_2 \leq \frac{1}{\sqrt{n}} (\log n)^{2k} (k!)^{k}$

For $\delta = (\alpha \beta)$ and $c$ small enough, this remainder is negligible. Then

$$\lambda^{2k} = \lambda^{2k-2} \leq \|B^{(k)}\|_2^2 \leq \|B^{(k)}\|_2 \|B^{(k)}\|_2 \leq \|B^{(k)}\|_2 \|B^{(k)}\|_2 ^{\frac{1}{2}} = \frac{1}{2} \|B^{(k)}\|_2 ^{\frac{1}{2}} \|B^{(k)}\|_2 ^{\frac{1}{2}}$$

$$\leq \frac{1}{2} \left( \sum_{\delta} \|M_{\delta}\|_2^2 \right)^{\frac{1}{2}} \left( \sum_{\delta} \|M_{\delta}\|_2^2 \right)^{\frac{1}{2}}$$

This is a sum over $2k$ non-backtracking paths $\delta$, each starting with the second edge of the previous path and each traversing a graph with $\leq 1$ cycle.

Counting canonical paths: Call $(y^{(1)}, y^{(2)}, y^{(3)})$ to this sum canonical.

Note: Without the assumption that each $G^{(k)}$ has $\geq 1$ cycle, a naive bound for $\lambda$ is $C_{2k}$, which is much larger.

Proof sketch: Call a step $(u, v)$ of $y^{(k)}$. A $k$-step of $y^{(k)}$ is a previously unvisited vertex (for any previous step $y^{(1)}, y^{(2)}, \ldots$).

The set of edges $\delta_{2k}$ for such steps form a spanning tree of the visited graph.

If $y^{(k)}$ has $\geq 1$ cycle, it is specified (since it is non-backtracking) by:

- All jump steps not belonging to this tree.
- The directed edge corresponding to each jump, immediately after each jump, and corresponding to the next step not on the already-traversed tree (which may be another jump or a first step).
- The first step not on the already-traversed tree, and its directed edge.

The total number of non-back edges is $c + 1 + x$. Each is traversed at most once by $y^{(k)}$, because $y^{(k)}$ doesn't cycle. Thus, the number of jumps is at most $c + 1 + x$ and the number of specific transitions is $\leq (C_{2k})^C_1 (x+1+1)$. 
If \( y(t) \) has one cycle, it is still specified by the above, but the number of jumps may be too large (O(k)).

\[
\sum_{i=0}^{\infty} (\text{loops } 000) \text{ times}
\]

However, let \( u \) be the last jump before \( y(t) \) begins to cycle, and let \( t \) be the number of steps from \( u \) until \( y(t) \) leaves the cycle.

The cycle is unique, so \( y(t) \) is still specified by discarding the info. For all jumps steps between \( u \) and \( v \), \( y(t) \) leaves the cycle.

The number of non-discarded jump steps is still at most \( n + \frac{1}{2} \), and \( t \leq n^2 + n^3 \). Thus, the number of specifications is still

\[
\leq (Cm)^{2(n + \frac{1}{2})^{n^2 + n^3}}
\]

Multiplying this bound for \( j = 1, \ldots, m \) concludes the proof.

From this lemma and a bound on \( I E \left[ \frac{1}{n} \sum_{i=1}^{n} M_{i+1}^{(1)} M_{i+1}^{(2)} \right] \), we get

\[
I E \left[ \left( \mathbf{R}^{(1)} \mathbf{R}^{(2)} \right)^{m} \right] \leq \left( \frac{2n}{k^2} \right)^{2} m = C(6) \left( \frac{1}{k^2} \right)^{2}.
\]

Then

\[
I P \left[ \max \left( \mathbb{A}_{n}, \mathbb{D}_{n} \right) > \sqrt{k} \right]
\]

\[
= I P \left[ G \text{ and } \mathbf{k} - \text{tangle-free} \right] + I P \left[ \max \left( \mathbb{A}_{n}, \mathbb{D}_{n} \right) > \sqrt{k} \right]
\]

\[
= I P \left[ G \text{ and } \mathbf{k} - \text{tangle-free} \right] + I P \left[ \max \left( \mathbb{A}_{n}, \mathbb{D}_{n} \right) > \sqrt{k} \right]
\]

\[
\leq I P \left[ G \text{ and } \mathbf{k} - \text{tangle-free} \right] + I P \left[ \max \left( \mathbb{A}_{n}, \mathbb{D}_{n} \right) > \sqrt{k} \right]
\]

\[
\leq I P \left[ G \text{ and } \mathbf{k} - \text{tangle-free} \right] + I P \left[ \left( \mathbf{R}^{(1)} \mathbf{R}^{(2)} \right)^{m} > \left( \frac{2n}{k^2} \right)^{2} \right]
\]

\[
\leq \left( \frac{2n}{k^2} \right)^{2} I E \left[ \left( \mathbf{R}^{(1)} \mathbf{R}^{(2)} \right)^{m} \right]
\]

\[
\rightarrow 0, \quad \text{choosing } m = C(6) \left( \frac{1}{k^2} \right)^{2}.
\]