The Non-Backtracking Matrix

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

(Note: Any fixed vertex has $\text{Binom}(\frac{n}{2}-1, \frac{a}{n}) + \text{Binom}(\frac{n}{2}, \frac{b}{n}) \approx \text{Poisson}(d)$ neighbors, where $d = \frac{a+b}{2}$. Each neighbor has conditionally $\approx \text{Poisson}(d)$ new neighbors, etc. Thus $G(n, \frac{a}{n}, \frac{b}{n})$ has the same local weak limit, and the same limit ESD for its adjacency matrix, as Erdos-Renyi $G(n, \frac{d}{n})$. 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 $[-2\sqrt{d-1}, 2\sqrt{d-1}]$. For $d \geq 3$, the largest eigenvalue λ_1 does not converge to this support:

Let $x = (1, \dots, 1)$. Then

$$Ax = dx$$

b/c the graph is d-regular. So d is always an eigenvalue of A.

A d-regular graph is called <u>Ramanujan</u> if

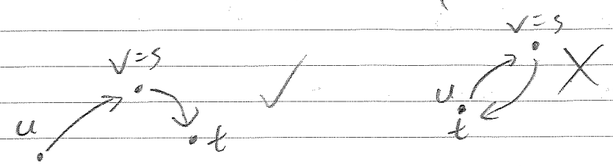
$$\lambda_2 \leq 2\sqrt{d-1} \text{ and } \lambda_n \geq -2\sqrt{d-1}.$$

Is it true that for a random d-regular graph, $\lambda_2 \to 2\sqrt{d-1}$ and

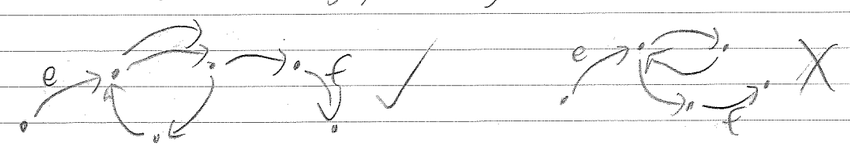
$\lambda_n \to -2\sqrt{d-1}$ as $n\to\infty$, i.e. a random regular graph is "almost Ramanujan"?

Def: The <u>non-backtracking matrix</u> of a graph $G=(V,E)$ is the matrix $B \in \{0,1\}^{2|E|\times 2|E|}$, indexed by directed edges $u\to v$, $v\to u$ for each (undirected) edge $\{u,v\}\in E$, such that for $e=u\to v$, $f=s\to t$,

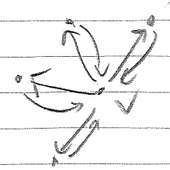
$$B_{ef} \equiv B_{u\to v,\, s\to t} = \begin{cases} 1 & \text{if } v=s \text{ and } t\neq u \\ 0 & \text{otherwise.} \end{cases}$$



Motivation: In the same way that $(A^\ell)_{uv}$ counts paths of ℓ+1 vertices starting from u, ending at v, we have $(B^\ell)_{ef}$ counts non-backtracking paths of ℓ+1 (directed) edges, starting at e and ending at f.

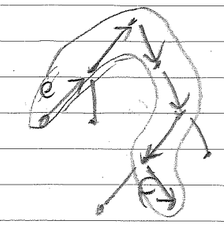


- If G has a vertex v of high degree, $\operatorname{Tr} A^\ell$ may be dominated by paths alternating between v and its neighbors. The number of such paths is $(\deg v)^{\ell/2}$ for even ℓ.



However, non-backtracking paths 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 e to f, which is the shortest path. For locally tree-like graphs these are easier to count than all paths.

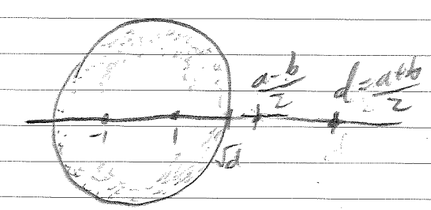


Note that B is <u>not</u> symmetric: $B_{ef} \neq B_{fe}$. The eigenvalues of B are complex-valued.

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

(a) For Erdős-Rényi $G(n, \frac{d}{n})$, the non-backtracking matrix B has one eigenvalue converging to d, and remaining eigenvalues in the disk $\{z\in\mathbb{C} : |z| \le \sqrt{d}+\varepsilon\}$, w.h.p. for any fixed $\varepsilon>0$ and large n.

(b) For the SBM $G(n, \frac{a}{n}, \frac{b}{n})$, if $\frac{a-b}{2} > \sqrt{d}$ where $d=\frac{a+b}{2}$, then B has <u>two</u> eigenvalues converging to d and $\frac{a-b}{2}$, respectively, and remaining eigenvalues in the above disk, w.h.p. for large n.

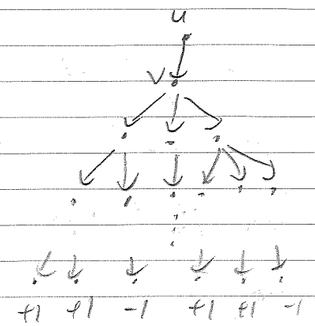


The corresponding eigenvector ξ is such that $\xi^{\text{in}} \in \mathbb{R}^n$ defined by $\xi^{\text{in}}_v = \sum_{u:\{u,v\}\in E} \xi_{u\to v}$ has asymptotically positive correlation with the true community membership vector $\sigma = (\underbrace{+1,\dots,+1}_{n/2}, \underbrace{-1,\dots,-1}_{n/2})$ as $n\to\infty$.

Remark: • For two communities of equal size, the threshold $\frac{a-b}{2} > \sqrt{d}$ is the optimal one for detection and weak recovery (Mossel, Neeman, 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 $\xi^{(r)}_{u\to v} = \left(\frac{a-b}{2}\right)^{-r} \cdot \sum_{s\to t:\, d(u\to v, s\to t)=r} \sigma_t$. (Doesn't depend on u.)



When this a Poisson(d)-GW tree, with each child having label equal to parent w/ probi $\frac{a}{a+b}$, it is known that this is correlated with σ_v iff $\frac{a-b}{2} > \sqrt{d}$.

This is approximately an eigenvector of B w/ eigenvalue $\frac{a-b}{2}$:

$$(B\xi^{(r)})_{u\to v} = \sum_{s\to t} B_{u\to v, s\to t}\, \xi^{(r)}_{s\to t} = \left(\frac{a-b}{2}\right)^{-r} \sum_{s\to t:\, d(u\to v, s\to t)=r+1} \sigma_t = \frac{a-b}{2} \cdot \xi^{(r+1)}_{u\to v}.$$

For $\frac{a-b}{2} > \sqrt{d}$ and large r, can show $\xi^{(r)}_{u\to v} \approx \xi^{(r+1)}_{u\to v}$. Similarly, $\xi^{(r)}_{u\to v} = d^r \sum_{s\to t:\, d(u\to v, s\to t)} 1$ is an approximate eigenvector w/ eigenvalue d, for large r.

If $M \in \mathbb{R}^{m\times n}$ is any (non-symmetric) matrix, w/ eigenvalues $\lambda_1, \dots, \lambda_n$ and and singular values $\sigma_1 \geq \dots \geq \sigma_n \geq 0$, then

$$\sum_{i=1}^n |\lambda_i|^2 \leq \sum_{i=1}^n \sigma_i^2 = \operatorname{Tr} MM^T.$$

Letting $\lambda_1, \dots, \lambda_n$ be the eigenvalues of B, for any $k \geq 0$,

$$\frac{1}{n}\sum_{i=1}^n |\lambda_i|^{2k} = \frac{1}{n}\sum_{i=1}^n |\lambda_i^k|^2 \leq \frac{1}{n} \operatorname{Tr} B^k (B^k)^T$$

We have $(B^k(B^k)^T)_{u\to v, u\to v} = \sum_{s\to t} (B^k)_{u\to v, s\to t} (B^k)_{u\to v, s\to t}$

$= \#\,(s\to t)$ which are k steps from $(u\to v)$.

On expectation, this is d^k. So for large n and fixed k, we expect $\frac{1}{n}\operatorname{Tr} B^k (B^k)^T \approx d^k$, and $\frac{1}{n}\sum_{i=1}^n |\lambda_i|^{2k} \leq d^k$ w.h.p. Thus all but a $o(1)$ fraction of eigenvalues of B should be in the disk $\{z \in \mathbb{C} : |z| \leq \sqrt{d} + \varepsilon\}$, for large n.

Bordenave, Lelarge, Massoulié show that all but two outliers are in this disk.

Thm (Friedman '08): Fix $d \geq 3$. A uniformly random d-regular graph has $\lambda_2 \leq 2\sqrt{d-1}+\varepsilon$ and $\lambda_n \geq -2\sqrt{d-1}-\varepsilon$ for any fixed $\varepsilon > 0$, w.h.p. for large n, where $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$ are the eigenvalues of its adjacency matrix.

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

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

$$\det(I - zB) = (1-z^2)^{m-n} \det(I - zA + z^2(D-I)).$$

Hence: Eigenvalues $\lambda = \frac{1}{z}$ of B are (assuming $m \geq n$)

- $+1, -1$ each with multiplicity $m-n$
- The $2n$ roots of $0 = \det(I - zA + z^2(D-I)) \iff 0 = \det(\lambda^2 I - \lambda A + D - I)$.

For a d-regular graph, $D = dI$. Thus

λ is an eigenvalue of $B \iff \lambda = 1, \lambda = -1$, or $\frac{\lambda^2 + d - 1}{\lambda}$ is an eigenvalue of A.

If A has an eigenvalue μ w/ $|\mu| > 2\sqrt{d-1}$, then B has two real eigenvalues at $\frac{\mu \pm \sqrt{\mu^2 - 4(d-1)}}{2}$, multiplying to $d-1$.

In particular, the eigenvalue d of A corresponds to $\{1, d-1\}$ for B, and Friedman's Theorem may be restated as:

Thm: Let B be the non-backtracking matrix of a random d-regular graph. Then B has a single eigenvalue equal to $d-1$, and remaining eigenvalues in the disc $\{z \in \mathbb{C} : |z| \le \sqrt{d-1} + \varepsilon\}$ for any fixed $\varepsilon > 0$, w.h.p. for large n.

Proof of Ihara-Bass: Let $S \in \mathbb{R}^{n \times 2m}$, $T \in \mathbb{R}^{2m \times n}$ be defined by

$$S_{u, v \to w} = \begin{cases} 1 & u = v \\ 0 & \text{otherwise} \end{cases} \qquad T_{u \to v, w} = \begin{cases} 1 & v = w \\ 0 & \text{otherwise.} \end{cases}$$

So $\xi_u^{out} \equiv (S\xi)_u = \sum_{w \in nbrs(u)} \xi_{u \to w}$, $\xi_w^{in} \equiv (\xi^T T)_w = \sum_{u \in nbrs(w)} \xi_{u \to w}$.

Define the permutation $P \in \mathbb{R}^{2m \times 2m}$ by

$$P_{u \to v, s \to t} = \begin{cases} 1 & \text{if } u = t,\ v = s \\ 0 & \text{otherwise.} \end{cases}$$

So $(P\xi)_{u \to v} = \xi_{v \to u}$, and $P^2 = I$. We have:

$$(ST)_{uv} = \sum_{s \to t} S_{u, s \to t} T_{s \to t, v} = \mathbb{I}\{u \sim v\} = A_{uv}, \text{ so } ST = A.$$

$$(TS)_{u \to v, s \to t} = \sum_w T_{u \to v, w} S_{w, s \to t} = \mathbb{I}\{v = s\}, \text{ so } TS - P = B.$$

$$(SPT)_{uv} = \sum_{s \to t} S_{u, s \to t} T_{t \to s, v} = \mathbb{I}\{u = v\} \cdot \deg(u), \text{ so } SPT = D.$$

Note that P consists of blocks $\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ along the diagonal, so has eigs. ± 1.

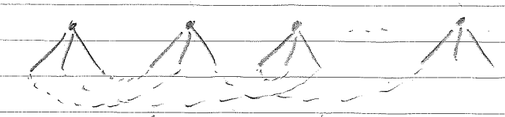
For $z \notin \{+1, -1\}$:

$$\begin{aligned}
\det(I - zB) &= \det(I + zP - zTS) \\
&= \det(I + zP) \cdot \det\Big(I - \underbrace{(I + zP)^{-1}}_{= \frac{1}{1-z^2} I - \frac{z}{1-z^2} P} \cdot zTS\Big) \\
&= (1 - z^2)^m \cdot \det\left(I - zS\left(\tfrac{1}{1-z^2} I - \tfrac{z}{1-z^2} P\right) T\right) \\
&= (1 - z^2)^m \cdot \det\left(I - \tfrac{z}{1-z^2} A + \tfrac{z^2}{1-z^2} D\right) \\
&= (1 - z^2)^{m-n} \det(I - zA + z^2(D - I)).
\end{aligned}$$

This must also hold at $z \in \{+1, -1\}$ by continuity.

<u>Proof ideas of Friedman's Thm</u> (following Bordenave '15).

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



Construct G by randomly pairing the half-edges. This yields a d-regular multi-graph, possibly with

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

Conditional on the event Ω_0 that G is a simple graph with no self-loops, the distribution of G is uniformly random over all d-regular graphs.

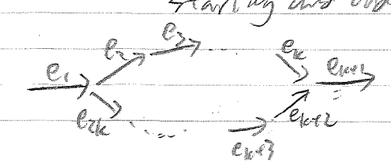
Prop (Bollobas, '01): $\mathbb{P}[\Omega_0] \to e^{-(d^2-1)/4}$ as $n \to \infty$.

If result holds w.h.p. for this multi-graph G, then it also holds w.h.p. conditional on Ω_0. So, suffices to analyze this model.

② Path-counting and centering: Recall $\sum_{i=1}^n |\lambda_i|^{2k} \le \mathrm{Tr}(B^k)(B^T)^k$. Here,

$$\mathrm{Tr}(B^k)(B^T)^k = \sum_{\substack{e_1, \ldots, e_k \\ e_{k+1}, \ldots, e_{2k}}} B_{e_1 e_2} \cdots B_{e_k e_{k+1}} \cdot B_{e_{k+2} e_{k+1}} B_{e_{k+3} e_{k+2}} \cdots B_{e_1 e_{2k}}$$

= # pairs of non-backtracking paths of length k edges, starting and ending at shared directed edges.



This includes the contribution from $\lambda_1 = d-1$, which corresponds to the (both left and right) eigenvector $\chi = (1,1,\dots,1) \in \mathbb{R}^{dn}$: $B\chi = (d-1)\chi$, $\chi^T B = (d-1)\chi^T$.

To remove this contribution, we center B:

- Index both the half-edges and directed edges of G by (u,i): $u \in \{1,\dots,n\}$, $i \in \{1,\dots,d\}$.

  Then $B_{(u,i),(v,j)} = \begin{cases} 1 & \text{if } (u,i) \text{ was paired with } (v,k) \text{ for some } k \neq j \\ 0 & \text{otherwise.} \end{cases}$

  $$= \sum_{k: k \neq j} M_{(u,i),(v,k)} \quad \text{where } M_{(u,i),(v,k)} = \mathbb{I}\{(u,i) \text{ paired w/ } (v,k)\}.$$

  (We take this as the definition of B when G has repeated edges/self loops.)

- Define $\underline{M}_{(u,i),(v,k)} = M_{(u,i),(v,k)} - \frac{1}{dn}$ as the centered version of M.
- If we set $\underline{B}_{(u,i),(v,j)} = \sum_{k: k \neq j} \underline{M}_{(u,i),(v,k)} = B_{(u,i),(v,k)} - \frac{d-1}{dn}$, then for any $\xi \perp \chi$, we have $\underline{B}\xi = B\xi$. If (λ, ξ) is an eigenpair different from $(d-1, \chi)$, then $\xi \perp \chi$ so

$$|\lambda|^{2k} = |\lambda^{2k}| = \|B^{2k}\xi\| = \|\underline{B}^{2k}\xi\| \leq \|\underline{B}^k\|^2 \leq \operatorname{Tr}(\underline{B}^k)(\underline{B}^k)^T.$$

Would suffice to show $\operatorname{Tr} \underline{B}^k (\underline{B}^k)^T \leq (\sqrt{d-1}+\varepsilon)^{2k}$ for some $k = k(n,\varepsilon)$ w.h.p.

③ In the Furedi-Komlos argument, we used Markov's inequality and bounded $\mathbb{E}[\operatorname{Tr} W^{2k}]$. Recall we needed to take $k = C(\varepsilon)\log n$, where $C(\varepsilon) \to \infty$ as $\varepsilon \searrow 0$, to compensate for the factor of n coming from Tr.

This doesn't work here:

Prop: For a constant $c > 0$, w/ prob. at least n^{-c}, G contains a clique of $d+1$ vertices.

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

$$c_0 \cdot (d-1)^{2k} \cdot \left(1 - \frac{d-1}{dn}\right)^{2k}$$ from pairs of non-backtracking paths on this clique.

We have $n^{-c} \cdot (d-1)^{2k}\left(1 - \frac{d-1}{dn}\right)^{2k} \gg (\sqrt{d-1}+2\varepsilon)^{2k}$ when $k \geq C(\varepsilon)\log n$, for small enough $\varepsilon > 0$ and large enough $C(\varepsilon)$. So in fact,

$$\mathbb{E}[\operatorname{Tr} \underline{B}^k (\underline{B}^k)^T] \gg (\sqrt{d-1}+\varepsilon)^{2k} \text{ for such } \varepsilon \text{ and } C(\varepsilon).$$

Idea: Isolate a low-probability event $\mathcal{E}$, such that $\mathbb{P}[\mathcal{E}] \to 0$ and

"$\mathbb{E}[\operatorname{Tr} \underline{B}^k (\underline{B}^k)^T \mathbb{I}\{\mathcal{E}\}] \leq (\sqrt{d-1}+\varepsilon)^{2k}$."

Def: G is ℓ-tangle-free if, for every vertex v, the ball $B_\ell(v)$ of radius ℓ has at most one cycle.

Prop: If $\ell \leq c\log n$ for $c \cdot \log(d-1) < \frac{1}{4}$, then $\mathbb{P}[G \text{ is } \ell\text{-tangle-free}] \to 1$.

Proof sketch: Fix v. Pair the d half-edges of v one at a time, then the (at most) $d-1$ unpaired half-edges of each neighbor one at a time, etc., until we've paired all half-edges from level $\ell-1$. The number of times we do a pairing is at most $|B_\ell(v)|$, which is at most

$$T \equiv \sum_{k=0}^{\ell-1} d(d-1)^k \leq 3(d-1)^\ell.$$

Each pairing, we create a new cycle with probability at most $\frac{d|B_\ell(v)|}{nd - d|B_\ell(v)|} \leq 0.99\frac{T}{n}$. So the number of cycles is $S \leq \operatorname{Binom}(T, 0.99\frac{T}{n})$.

Then $\mathbb{P}[S \geq 2] \approx \binom{T}{2}\left(0.99\frac{T}{n}\right)^2 = C \cdot \frac{(d-1)^{4\ell}}{n^2}$. Taking a union bound over v,

$$\mathbb{P}[G \text{ has an } \ell\text{-tangle}] \leq C \cdot \frac{(d-1)^{4\ell}}{n} \to 0 \text{ if } \ell \leq c\log n,$$

for $c \cdot \log(d-1) < \frac{1}{4}$.

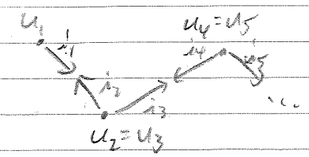
④ Refined argument: Let $\ell = \lfloor c \log n \rfloor$, $m = \lfloor C' \frac{\log n}{\log\log n} \rfloor$, $k = \ell m$.

To bound $|\lambda|^{2k}$, write first $|\lambda|^\ell \le \|B^\ell \xi\|$. Note that

$$(B^\ell)_{(u,i),(v,j)} = \sum_\gamma M_{\gamma_1\gamma_2} M_{\gamma_3\gamma_4} \cdots M_{\gamma_{2\ell-1}\gamma_{2\ell}}$$

where the sum is over all $(2\ell+1)$-tuples of half-edges $(\gamma_t = (u_t, i_t) : t = 1, \dots, 2\ell+1)$ such that

- $u_2 = u_3,\ u_4 = u_5, \dots, u_{2\ell} = u_{2\ell+1}$
- $i_2 \ne i_3,\ i_4 \ne i_5, \dots, i_{2\ell} \ne i_{2\ell+1}$
- $\gamma_1 = (u,i)$ and $\gamma_{2\ell+1} = (v,j)$.



Let G_γ be the sub-graph containing edges $\{u_1, u_2\}, \{u_3, u_4\}, \dots, \{u_{2\ell-1}, u_{2\ell}\}$.

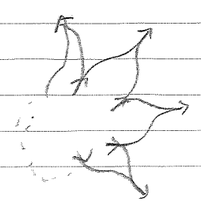
Let $\underline{B}^{(\ell)} = \sum_{\gamma : G_\gamma \text{ has } \le 1 \text{ cycle}} \underline{M}_{\gamma_1\gamma_2} \underline{M}_{\gamma_3\gamma_4} \cdots \underline{M}_{\gamma_{2\ell-1}\gamma_{2\ell}}$.

Lemma: There exists $R^{(\ell)}$ s.t.

- If $\xi^T \chi = 0$ and G is ℓ-tangle-free, then
$$B^\ell \xi = \underline{B}^\ell \xi = \underline{B}^{(\ell)} \xi + R^{(\ell)} \xi, \text{ and}$$
- $\|R^{(\ell)}\| \le \frac{1}{d_n} \cdot (\log n)^{30} \cdot (d-1)^\ell$ whp.

For $\ell = \lfloor c \log n \rfloor$ and c small enough, this remainder is negligible. Then

$$|\lambda|^{2k} = |\lambda^\ell|^{2m} = \|B^\ell \xi\|^{2m} \approx \|\underline{B}^{(\ell)} \xi\|^{2m} \le \|\underline{B}^{(\ell)}\|^{2m} = \|\underline{B}^{(\ell)} \underline{B}^{(\ell)T}\|^m$$
$$\le \mathrm{Tr}\left(\underline{B}^{(\ell)} (\underline{B}^{(\ell)})^T\right)^m.$$



This is a sum over $2m$ non-backtracking paths, each starting with the reversal of the last edge of the previous path, and each traversing a graph with ≤ 1 cycle.

⑤ Counting canonical paths: Call $(\gamma^{(1)}, \gamma^{(2)}, \dots, \gamma^{(2m)})$ in this sum canonical if the distinct vertices are $1, 2, 3, \dots$ in visited order, and the distinct half-edges per vertex are $1, 2, \dots, d$ in visited order.

Lemma: The number of canonical tuples $(\gamma^{(1)}, \dots, \gamma^{(2m)})$ visiting s vertices and a edges is at most
$$(C\ell m)^{Cm + Cm(a+1-s)}.$$

Note: Without the assumption that each $G_{\gamma^{(i)}}$ has ≤ 1 cycle, a naive bound for $s, a = O(1)$ is $C^{\ell m}$, which is much larger.

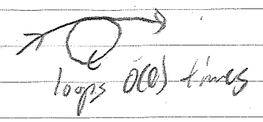
Proof sketch: Call a step $u \to v$ of $\gamma^{(i)}$ a first step if v is a previously unvisited vertex (for any previous step of $\gamma^{(1)}, \dots, \gamma^{(i)}$). The set of edges $\{u, v\}$ for such steps form a spanning tree of the visited graph.

If $\gamma^{(i)}$ has no 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-tree edges is $a+1-s$. Each is traversed at most once by $\gamma^{(i)}$, because $\gamma^{(i)}$ doesn't cycle. Thus, the number of jumps is at most $a+1-s$, and the number of specifications is
$$\le (C\ell m)^{C(a+1-s+1)}.$$

If $\gamma^{(j)}$ has one cycle, it is still specified by the above, but the number of jumps may be too large ($O(\ell)$):



However, let $u \to v$ be the last jump before $\gamma^{(j)}$ begins to cycle, and let t be the number of steps from $u \to v$ until $\gamma^{(j)}$ leaves the cycle. The cycle is unique, so $\gamma^{(j)}$ is still specified by discarding the info for all jump steps between $u \to v$ and when $\gamma^{(j)}$ leaves the cycle. The number of non-discarded jump steps is still at most $a+1-s$, and $t \in \{1, \dots, \ell\}$. Thus the number of specifications is still

$$\leq (C\ell m)^{C(a+1-s+1)}.$$

Multiplying this bound for $j = 1, \dots, m$ concludes the proof. $\boxtimes$

(6) From this lemma and a bound on $\mathbb{E}\left[\prod_{j=1}^{m} \underline{M}_{\gamma_1^{(j)}\gamma_2^{(j)}} \cdots \underline{M}_{\gamma_{2k-1}^{(j)}\gamma_{2k}^{(j)}}\right]$, we get $\mathbb{E}\left[\mathrm{Tr}\left(\underline{B}^{(\ell)}(\underline{B}^{(\ell)})^T\right)^m\right] \leq \left(\sqrt{d-1}+\varepsilon\right)^{2k}$, where $k = \ell m$, $m = C(\varepsilon)\left\lfloor \frac{\log n}{\log\log n}\right\rfloor$.

Then $\mathbb{P}\left[\max(|\lambda_2|, \dots, |\lambda_{nd}|) > \sqrt{d-1}+\varepsilon\right]$

$$\leq \mathbb{P}[G \text{ not } \ell\text{-tangle-free}] + \mathbb{P}\left[\max(|\lambda_2|^{2k}, \dots, |\lambda_{nd}|^{2k}) > \left(\sqrt{\lambda_1}+\varepsilon\right)^{2k}, G \text{ is } \ell\text{-tangle-free}\right]$$

$$\leq \mathbb{P}[G \text{ not } \ell\text{-tangle-free}] + \underbrace{\mathbb{P}\left[\mathrm{Tr}\left(\underline{B}^{(\ell)}(\underline{B}^{(\ell)})^T\right)^m > \left(\sqrt{d-1}+\varepsilon\right)^{2k}\right]}_{\leq \left(\frac{1}{\sqrt{d-1}+\varepsilon'}\right)^{2k} \cdot \mathbb{E}\left[\mathrm{Tr}\left(\underline{B}^{(\ell)}(\underline{B}^{(\ell)})^T\right)^m\right]}$$

$\to 0$, choosing $m = C(\varepsilon')\left\lfloor \frac{\log n}{\log\log n}\right\rfloor$.