Community detection thresholds and the weak Ramanujan property

Laurent Massoulié

Community detection thresholds and the weak Ramanujan property

Laurent Massoulié
Microsoft Research–Inria Joint Centre
laurent.massoulie@inria.fr

ABSTRACT

Decelle et al. [1] conjectured the existence of a sharp threshold on model parameters for community detection in sparse random graphs drawn from the *stochastic block model*. Mossel, Neeman and Sly [2] established the negative part of the conjecture, proving impossibility of non-trivial reconstruction below the threshold. In this work we solve the positive part of the conjecture. To that end we introduce a modified adjacency matrix B which counts *self-avoiding* paths of a given length ℓ between pairs of nodes. We then prove that for logarithmic length ℓ, the leading eigenvectors of this modified matrix provide a non-trivial reconstruction of the underlying structure, thereby settling the conjecture. A key step in the proof consists in establishing a *weak Ramanujan property* of the constructed matrix B. Namely, the spectrum of B consists in two leading eigenvalues $\rho(B)$, λ_2 and $n-2$ eigenvalues of a lower order $O(n^{\epsilon}\sqrt{\rho(B)})$ for all $\epsilon > 0$, $\rho(B)$ denoting B's spectral radius.

Categories and Subject Descriptors

F.2 [**Theory of Computation**]: Analysis of Algorithms and Problem Complexity

Keywords

spectral clustering, community detection, phase transition, spectral separation

1. INTRODUCTION

1.1 Background

Community detection, like clustering, aims to identify groups of similar items from a global population. It is a generic primitive useful e.g. for performing recommendation of contacts to users of online social networks. The stochastic block model has been introduced by Holland et al. [3] to represent interactions between individuals. It consists of a random graph on n nodes, each node $i \in \mathcal{N} = \{1, \ldots, n\}$ being assigned a type σ_i from some fixed set Σ. Conditionally on node types, edge (i, j) is present with probability $p(\sigma_i, \sigma_j)$ independently of other edges, for some matrix of probabilities $\{p(\sigma, \sigma')\}_{\sigma,\sigma' \in \Sigma}$. It constitutes an adequate testbed for community detection. Indeed the performance of candidate detection schemes, captured by the fraction of nodes i for which estimated types $\hat{\sigma}_i$ and true types σ_i coincide, can be compared and analysed on instances of the stochastic block model.

Decelle et al. [1] conjectured the existence of a phase transition in the sparse regime where the graph's average degree is $O(1)$. Specifically, they predicted that for parameters below a certain threshold, no estimates $\hat{\sigma}_i$ of node types existed that would be positively correlated with true types σ_i, while above the threshold, belief propagation algorithms could determine estimates $\hat{\sigma}_i$ achieving such a positive correlation. Their conjecture is formulated on a simple symmetric instance of the stochastic block model featuring two node types $\{+1, -1\}$. The phenomenon appears more general though: Heimlicher et al. [4] extended the conjecture to the more general setup of labeled stochastic block models.

The study of this phenomenon is important for two reasons. First, by localizing precisely the transition point below which no useful signal is present in the observations, one thus characterizes how much subsampling of the original graph can be performed before all information is lost. Second, algorithms leading to estimates $\hat{\sigma}_i$ that achieve positive correlation all the way down to the transition are expected to constitute more robust approaches than alternatives which would fail before the transition. It is therefore important to identify such algorithms.

The negative part of the conjecture has been proven by Mossel, Neeman and Sly [2]. Essentially they established that existence of estimates $\hat{\sigma}_i$ positively correlated with true types σ_i would imply feasibility of a reconstruction problem on a random tree model describing the local statistics of the original random graph. However by results of Evans et al. [5] such reconstruction is infeasible below the conjectured transition point.

Until now, positive results did not apply down to the transition point. The best results to date (see [2]) relied on Coja-Oghlan [6], showing that spectral clustering applied to the adjacency matrix suitably trimmed by removal of high degree nodes yields positively correlated estimates. However this does not apply down to the conjectured threshold.

This limitation stems from the following fact. Spectral methods perform well on matrices enjoying a spectral sepa-

STOC '14, May 31 - June 03 2014, New York, NY, USA

ACM 978-1-4503-2710-7/14/05 ...$15.00.
http://dx.doi.org/10.1145/2591796.2591857.

ration property, namely the spectrum should comprise a few large eigenvalues whose associated eigenvectors reflect the sought structure and all other eigenvalues should be negligible. The prototype of such separation is the Ramanujan property, according to which d-regular graphs have the second eigenvalue λ no larger than $2\sqrt{d-1}$ in absolute value. Friedman [7] established that random d-regular graphs almost satisfy this, in that for them $|\lambda| \leq 2\sqrt{d-1}+o(1)$. Erdős-Rényi graphs with average degree d are such that $|\lambda| \leq O(\sqrt{d})$, *provided* $d=\Omega(\log n)$ (see Feige and Ofek [8]), but such Ramanujan-like separation is lost for smaller d. This lack of separation inherently limits the power of classical spectral methods in the sparse case.

1.2 Main result

We focus on the stochastic block model in Decelle et al. [1]. The graph is denoted $\mathcal{G}$, node types (or spins) σ_i are uniformly and i.i.d. drawn from $\{-1,+1\}$. An edge is present between any two nodes i,j with probability a/n if $\sigma_i=\sigma_j$, and b/n if $\sigma_i=-\sigma_j$, constants a and b being the model parameters. Letting $\tau=(a-b)^2/[2(a+b)]$, it is known that for $\tau<1$ positively correlated detection is impossible. We set out to prove that it is feasible for $\tau>1$.

Introduce the notations $\alpha:=(a+b)/2$, $\beta:=(a-b)/2$. The detectability condition $\tau>1$ can be restated as

$$\beta^2>\alpha. \tag{1}$$

As mentioned, Coja-Oghlan regularizes the adjacency matrix of the random graph by removing high degree nodes before applying spectral clustering. In contrast, we regularize the initial data through *path expansion*. Namely, we consider matrix $B^{(\ell)}$, where $B^{(\ell)}_{ij}$ counts the number of self-avoiding paths of graph edges of length ℓ connecting i to j.

Our main result is then the following

THEOREM 1.1. *Assume Condition (1) holds. Set the path length parameter ℓ to $\ell\sim c\log(n)$ for a constant c such that $c\log(\alpha)<1/4$. Let x be a normed eigenvector corresponding to the second largest eigenvalue of $B^{(\ell)}$. There exists t such that, defining the spin estimates $\hat{\sigma}_i$ as*

$$\hat{\sigma}_i=\begin{cases} +1 & \text{if } x_i\sqrt{n}\geq t,\\ -1 & \text{otherwise,}\end{cases} \tag{2}$$

the empirical overlap *between the true and estimated spins defined as*

$$ov(\sigma,\hat{\sigma}):=\frac{1}{n}\sum_{i\in\mathcal{N}}\sigma_i\hat{\sigma}_i \tag{3}$$

converges in probability to the set $\{-r,+r\}$ for some strictly positive constant $r>0$ as $n\to\infty$.

It proves the positive part of Decelle et al.'s conjecture and identifies a specific spectral method based on the path-expanded matrix $B^{(\ell)}$. An auxiliary result consists in showing that matrix $B^{(\ell)}$ enjoys a spectral separation property that is a weak version of the Ramanujan property. Namely, denoting by $\rho(B^{(\ell)})$ the spectral radius of $B^{(\ell)}$, we show that the third largest eigenvalue λ of matrix $B^{(\ell)}$ satisfies for all positive constant ϵ:

$$|\lambda|\leq n^{\epsilon}\sqrt{\rho(B^{(\ell)})}.$$

We note that computation of $B^{(\ell)}$ and hence of the $\hat{\sigma}_i$ can be done in polynomial time: as shown in Lemma 4.2 the ℓ-neighborhood of any i contains at most one cycle so that each $B^{(\ell)}_{ij}$ is readily evaluated by suitable breadth-first search.

1.3 Related work

Krzakala et al. [9] conjectured that a spectral method based on the "non-backtracking" edge-to-edge matrix achieves positive overlap when above the threshold. Nadakuditi and Newman [10] conjectured that the same holds for modularity maximization. Mossel, Neeman and Sly [11] showed that a modification of belief propagation achieves maximal overlap when initialized with any reconstruction having positive overlap. Shortly after completion of the present paper, Mossel, Neeman and Sly [12] have independently proposed another proof of the positive part of the conjecture. They rely on a reconstruction method markedly distinct from the one introduced here. In particular it is not a spectral method.

1.4 Paper organization

Section 2 describes the structure of Theorem 1.1's proof. Section 3 proves Theorem 2.2, which expresses matrix $B^{(\ell)}$ as an expansion in terms of the matrices $B^{(m)}$, $m<\ell$, together with bounds on the spectral norm of the matrix coefficients involved. Section 4 contains the so-called "local analysis" of node neighborhoods. Specifically it gives controls on the vectors $B^{(m)}e$ and $B^{(m)}\sigma$, where e is the all-ones vector and σ is the vector of spins, establishing a quasi-deterministic growth pattern with respect to m. Section 5 concludes.

2. PROOF STRUCTURE

We characterize the spectral structure of $B^{(\ell)}$ as follows[1]:

THEOREM 2.1. *Assume (1) and $\ell=c\log n$ with $c\log(\alpha)<1/4$. Then with high probability:*

(i) The leading eigenvalue of $B^{(\ell)}$ is $\tilde{\Theta}(\alpha^{\ell})$, with corresponding eigenvector asymptotically parallel to $B^{(\ell)}e$.

(ii) Its second eigenvalue is $\Omega(\beta^{\ell})$, with corresponding eigenvector asymptotically parallel to $B^{(\ell)}\sigma$.

(iii) There is a random variable X with unit mean and variance $1/(\beta^2/\alpha-1)$ such that for all x that is an atom of neither X's nor $-X$'s distribution, the following convergence in probability holds for any normed vector y asymptotically aligned with $B^{(\ell)}\sigma$:

$$\frac{1}{n}\sum_{i=1}^{n}\mathbf{1}_{\sigma_i=\pm 1}\mathbf{1}\left\{y_i\geq\frac{x}{\sqrt{n\mathbf{E}(X^2)}}\right\}\to\frac{1}{2}\mathbf{P}(\pm X\geq x). \tag{4}$$

(iv) For any $\epsilon>0$, all other eigenvalues are $O\left(n^{\epsilon}\sqrt{\alpha^{\ell}}\right)$.

Before we describe the steps used to establish this, let us verify how it implies Theorem 1.1. Note that since $\mathbf{E}(X)=1$, writing

$$\mathbf{E}(X)=\int_0^{\infty}\left(\mathbf{P}(X\geq x)-\mathbf{P}(-X\geq x)\right)dx,$$

inequality $\mathbf{P}(X\geq x)-\mathbf{P}(-X\geq x)>0$ must hold on a set of x's of positive Lebesgue measure. Since the points x at

[1]We note $u=\tilde{O}(v)$ (respectively $\tilde{\Omega}(v)$) if $u=O(v\log^k(n))$ (respectively $\Omega(v\log^k(n))$ for some constant k and $u=\tilde{\Theta}(v)$ if u is both $\tilde{O}(v)$ and $\tilde{\Omega}(v)$.

which the distribution of either X or $-X$ has an atom is at most countable, there thus exists an x at which neither distribution has an atom, and inequality $\mathbf{P}(X \geq x) - \mathbf{P}(-X \geq x) > 0$ holds. Letting $t = x/\sqrt{\mathbf{E}(X^2)}$ and $r = \mathbf{P}(X \geq x) - \mathbf{P}(-X \geq x)$ we conclude by (ii), (iii) and (4) that the empirical overlap in (3) must converge to $\{-r, +r\}$.

Theorem 2.1 will follow from the combination of two analyses. Let $\bar{A}$ denote the expectation of the graph's adjacency matrix conditional on the spin vector σ, that is

$$\bar{A} = \frac{a}{n}\left[\frac{1}{2}(ee' + \sigma\sigma') - I\right] + \frac{b}{2n}(ee' - \sigma\sigma'). \tag{5}$$

The first analysis establishes the following

THEOREM 2.2. *Matrix $B^{(\ell)}$ verifies the identity*

$$B^{(\ell)} = \Delta^{(\ell)} + \sum_{m=1}^{\ell}(\Delta^{(\ell-m)}\bar{A}B^{(m-1)}) - \sum_{m=1}^{\ell}\Gamma^{\ell,m}, \tag{6}$$

for matrices $\Delta^{(\ell)}$, $\Gamma^{\ell,m}$ such that for $\ell = O(\log n)$ and any fixed $\epsilon > 0$, with high probability

$$\rho(\Delta^{(\ell)}) \leq n^{\epsilon}\alpha^{\ell/2}, \tag{7}$$

$$\rho(\Gamma^{\ell,m}) \leq n^{\epsilon-1}\alpha^{(\ell+m)/2},\ m = 1, \dots, \ell. \tag{8}$$

A *local* analysis is then needed to establish properties of the ℓ-neighborhoods of nodes in graph $\mathcal{G}$. Noting $d_{\mathcal{G}}$ the graph distance, the key quantities in this analysis are

$$\begin{aligned} S_t(i) &= |\{j : d_{\mathcal{G}}(i,j) = t\}|, \\ D_t(i) &= \textstyle\sum_j \mathbf{1}_{d_{\mathcal{G}}(i,j)=t}\sigma_j. \end{aligned} \tag{9}$$

They are close (in a sense made precise in Section 4) to the corresponding quantities $(B^{(t)}e)_i$, $(B^{(t)}\sigma)_i$, and are easier to analyze. In particular, they enjoy a *quasi-deterministic growth* property:

THEOREM 2.3. *Assume (1) and $\ell = c\log n$ with $c\log(\alpha) < 1/4$. Then there are constants $C, \epsilon > 0$ such that with probability $1 - O(n^{-\epsilon})$ the following holds for all $i \in \mathcal{N}$ and $t = 1, \dots, \ell$:*

$$\begin{aligned} S_t(i) &\leq C(\log(n))\alpha^t, \\ |D_t(i)| &\leq C(\log(n))\beta^t. \end{aligned} \tag{10}$$

$$\begin{aligned} S_t(i) &= \alpha^{t-\ell}S_\ell(i) + C(\log(n) + \sqrt{\log(n)\alpha^t}), \\ D_t(i) &= \beta^{t-\ell}D_\ell(i) + C(\log(n) + \sqrt{\log(n)\alpha^t}). \end{aligned} \tag{11}$$

This combined with Theorem 2.2 yields the key intermediate step:

THEOREM 2.4. *Assume (1) and $\ell = c\log n$ with $c\log(\alpha) < 1/4$. Then with high probability the matrix $B^{(\ell)}$ satisfies the following weak Ramanujan property*

$$\sup_{|x|=1, x'B^{(\ell)}e = x'B^{(\ell)}\sigma = 0} |B^{(\ell)}x| \leq n^{\epsilon}\alpha^{\ell/2}. \tag{12}$$

Another ingredient consists in coupling the neighborhoods of nodes in graph $\mathcal{G}$ with a random tree process, and performing a martingale analysis of this tree process. This is done in Section 4.2. It establishes (see Theorem 4.2) that the vector $(\beta^{-\ell}D_\ell(i))$ is close in some sense to a vector $(\sigma_i D_i)$ where the D_i are i.i.d., distributed as the limit of a martingale. This limiting martingale distribution is precisely that of variable X in the statement of Theorem 2.1.

3. MATRIX EXPANSION AND SPECTRAL RADII BOUNDS

We now establish Theorem 2.2. Denote ξ_{ij} the indicator of edge (i,j)'s presence in $\mathcal{G}$. Let P_{ij} be the set all so-called *self-avoiding*, or *simple* paths $i_0^\ell := \{i_0, \dots, i_\ell\}$ from i to j, i.e. such that $i_0 = i$, $i_\ell = j$ and $|\{i_0, \dots, i_\ell\}| = \ell + 1$. We have

$$B_{ij}^{(\ell)} = \sum_{i_0^\ell \in P_{ij}} \prod_{t=1}^{\ell} \xi_{i_{t-1}i_t}. \tag{13}$$

Use matrix $\bar{A}$ introduced in (5) to define

$$\Delta_{ij}^{(\ell)} := \sum_{i_0^\ell \in P_{ij}} \prod_{t=1}^{\ell}(A - \bar{A})_{i_{t-1}i_t} \tag{14}$$

We then have the expansion:

$$\begin{aligned} \Delta_{ij}^{(\ell)} = B_{ij}^{(\ell)} - \sum_{m=1}^{\ell}\sum_{i_0^\ell \in P_{ij}} \prod_{t=1}^{\ell-m}(A - \bar{A})_{i_{t-1}i_t} \times \cdots \\ \times \bar{A}_{i_{\ell-m}i_{\ell-m+1}} \prod_{t=\ell-m+2}^{\ell} A_{i_{t-1}i_t}. \end{aligned} \tag{15}$$

Let Q_{ij}^m be the set of paths i_0^ℓ defined by $i_0 = i$, $i_\ell = j$ and

$$|\{i_0, \dots, i_{\ell-m}\}| = \ell - m + 1 \ \&\ |\{i_{\ell-m+1}, \dots, i_\ell\}| = m.$$

Paths in Q_{ij}^m are thus concatenations of simple paths $i_0^{\ell-m}$ and $i_{\ell-m+1}^\ell$. Note that $P_{ij} \subset Q_{ij}^m$. Let R_{ij}^m denote the set difference $Q_{ij}^m \setminus P_{ij}$. It then consists of paths i_0^ℓ such that both $i_0^{\ell-m}$ and $i_{\ell-m+1}^\ell$ are simple, and further verify that the intersection of the corresponding sets is not empty.

Define matrix $\Gamma^{\ell,m}$ as

$$\Gamma_{ij}^{\ell,m} := \sum_{i_0^\ell \in R_{ij}^m} \prod_{t=1}^{\ell-m}(A-\bar{A})_{i_{t-1}i_t}\bar{A}_{i_{\ell-m}i_{\ell-m+1}} \prod_{t=\ell-m+2}^{\ell} A_{i_{t-1}i_t}. \tag{16}$$

Add and subtract contributions of paths in R_{ij}^m to the m-th term in (15) for $m = 1, \dots, \ell$ to obtain a similar identity with summation over paths in Q_{ij}^m instead of P_{ij}, and a term $\sum_{m=1}^{\ell}\Gamma_{ij}^{\ell,m}$ added to the right-hand side. Noting that the summation

$$\sum_{i_0^\ell \in Q_{ij}^m} \prod_{t=1}^{\ell-m}(A - \bar{A})_{i_{t-1}i_t}\bar{A}_{i_{\ell-m}i_{\ell-m+1}} \prod_{t=\ell-m+2}^{\ell} A_{i_{t-1}i_t}$$

coincides with the (i,j) entry of matrix $\Delta^{(\ell-m)}\bar{A}B^{(m-1)}$, this yields expansion (6).

We now show the following

PROPOSITION 3.1. *For all integers $k, \ell \geq 1$, it holds that*

$$\begin{aligned} \mathbf{E}\left[\rho(\Delta^{(\ell)})^{2k}\right] \leq \sum_{v=\ell+1}^{k\ell+1}\sum_{e=v-1}^{k\ell} n^v[(v+1)^2(\ell+1)]^{2k(1+e-v+1)} \\ \times \left(\frac{\alpha}{n}\right)^{v-1}\left[\frac{\max(a,b)}{n}\right]^{e-v+1}. \end{aligned} \tag{17}$$

Inequality (7) readily follows: Indeed for $\ell = O(\log(n))$ and fixed $\epsilon > 0$, choose an integer $k > 0$ such that $\epsilon > 1/(2k)$. By (17), noting $\rho := \rho(\Delta^{(\ell)})$ it holds that

$$\mathbf{E}(\rho^{2k}) \leq (1 + o(1))n\alpha^{k\ell}[(k\ell + 2)^2(\ell + 1)]^{2k}.$$

Thus

$$\begin{aligned}\mathbf{P}(\rho \geq n^{\epsilon}\alpha^{\ell/2}) &\leq \frac{\mathbf{E}(\rho^{2k})}{n^{2k\epsilon}\alpha^{k\ell}} \\ &\leq (1+o(1))\frac{n\alpha^{k\ell}[(k\ell+2)^2(\ell+1)]^{2k}}{n^{2k\epsilon}\alpha^{k\ell}} \\ &\leq (1-o(1))n^{1-2k\epsilon}[(k\ell+2)^2(\ell+1)]^{2k} \\ &= o(1),\end{aligned}$$

since we chose k so that $2k\epsilon > 1$ and the last term is polylogarithmic in n. This establishes (7).

PROOF. (of Proposition 3.1) We use the *trace method*, adapting combinatorial arguments of Füredi and Komlós [13] to the present context. Specifically chose $k > 0$. One has

$$\mathbf{E}(\rho^{2k}) \leq \mathbf{E}\mathrm{Tr}((\Delta^{(\ell)})^{2k}). \quad (18)$$

Note that $\mathrm{Tr}((\Delta^{(\ell)})^{2k})$ is the sum over circuits of length $2k$ of the products of the entries $\Delta_e^{(\ell)}$ over the edges e in the circuit. Moreover, given the definition of $\Delta^{(\ell)}$, these correspond to products of entries $A_e - \bar{A}_e$ over edges e of circuits of length $2k\ell$ such that consecutive length ℓ-paths are simple.

We bound the expectation of the corresponding sum as follows. Let v (respectively, e) be the number of nodes (respectively, edges) traversed by a particular circuit. We represent the corresponding circuit as follows.

We number nodes by the order in which they are met by the circuit, starting with node 1. We break each length ℓ-simple path into consecutive sequences consisting of

- a path using only edges already used in the circuit, and lying on the tree of new node discoveries
- a path of discoveries of new nodes
- a cycle edge connecting the end of the two previous steps to a node already spanned. Such a cycle edge may have already been traversed by the circuit.

Given the tree spanned so far and the current position on it, the first part of the sequence is characterized by the node label of its end: indeed, since on this subsequence we require the path to be simple, there is only one path on the tree going from the origin to the destination. We represent the first part by the number of the destination node if this part is non-empty, by zero otherwise.

The second part of the sequence is simply represented by its length, which is constrained to lie in $\{0, \ldots, \ell\}$. Indeed, it cannot exceed ℓ, as we consider sequences that lie within a length ℓ-simple path.

Finally, the third part of the sequence is simply characterized by the number of its end point, and by zero if this part is not present. We must allow for this case, as when we break up a length ℓ-simple path into constituting such sequences, the last such sequence may not end up by traversal of such a redundant edge.

We now use this representation to bound the number of corresponding sequences. An individual sequence is represented by a triplet (p, q, r) with $p \in \{0, \ldots, v\}$, $q \in \{0, \ldots, \ell\}$, and $r \in \{0, \ldots, v\}$. Note further that each such sequence corresponds to either the end of an individual length ℓ-simple path, or the traversal of a redundant edge. The number of such edges is $e - v + 1$, and each edge can be traversed at most $2k$ times by the constraint that circuits are formed from length ℓ simple paths. Thus the number of valid circuits corresponding to v and e is at most

$$[(v+1)^2(\ell+1)]^{2k(1+e-v+1)}.$$

For a given number of nodes v and edges e, the number of ordered sequences of corresponding nodes in $\{1, \ldots, n\}$ is at most n^v. For a given edge present with multiplicity $m \in \{1, \ldots, 2k\}$, the corresponding expectation is zero if $m = 1$, and for $m \geq 2$ one has

$$\mathbf{E}((\xi_{ij} - \mathbf{E}(\xi_{ij}|\sigma))^m|\sigma) \leq \frac{a(\sigma_i, \sigma_j)}{n},$$

where $a(\sigma_i, \sigma_j)$ equals a if $\sigma_i = \sigma_j$ and b otherwise. For the $e - v + 1$ cyclic edges we use the upper bound $\max(a, b)/n$. We are left with a tree with $v-1$ edges, for which upon averaging over σ we get a contribution $(\alpha/n)^{v-1}$. The number of nodes v on any configuration whose contribution in expectation does not vanish must lie between $\ell+1$ and $k\ell+1$: indeed each node discovery costs one edge, but this edge must be doubled for the contribution not to vanish. Since there are in total $2k\ell$ edges, at most $k\ell$ nodes can be discovered in addition to the original node of the circuit. The number of distinct edges is similarly bounded by $k\ell$ in any configuration with non-vanishing expectation. The right-hand side of (18) is then upper-bounded by the right-hand side of (17). Inequality (17) follows. $\square$

We now establish a bound on the spectral radius of the matrix $\Gamma^{\ell,m}$. Specifically, we have

PROPOSITION 3.2. *For all $k, \ell \geq 1$ and $m \in \{1, \ldots, \ell\}$ we have the following*

$$\mathbf{E}((\rho(\Gamma^{\ell,m})^{2k}) \leq \sum_{v=m\vee(\ell-m+1)}^{1+k(\ell+m)} \sum_{e=v-1}^{k(\ell+m)} \left(\frac{a \vee b}{n}\right)^{2k+e-v+1} \times \\ \times v^{2k}[(v+1)^2(\ell+1)]^{4k(1+e-v+1)} n^v \left(\frac{\alpha}{n}\right)^{v-1}. \quad (19)$$

The proof, postponed to the appendix, follows the same lines as that of Proposition 3.1. It readily implies inequality (8). Indeed for $\ell = O(\log(n))$, and any fixed $\epsilon > 0$, choose $k > 0$ such that $\epsilon > 1/(2k)$. By (19) it holds that

$$\begin{aligned}\mathbf{E}(\rho(\Gamma^{\ell,m})^{2k}) \leq\ & (1+o(1))n\alpha^{k(\ell+m)} \times \cdots \\ & \cdots \times \left[\frac{(a\vee b)(k(\ell+m)+2)^5(\ell+1)^2}{n}\right]^{2k}.\end{aligned}$$

Thus

$$\begin{aligned}\mathbf{P}(\rho \geq n^{\epsilon-1}\alpha^{(\ell+m)/2}) &\leq \frac{\mathbf{E}(\rho(\Gamma^{\ell,m})^{2k})}{n^{2k(\epsilon-1)}\alpha^{k(\ell+m)}} \\ &\leq (1+o(1))n^{1-2k\epsilon}\left[(a\vee b)(k(\ell+m)+2)^5(\ell+1)^2\right]^{2k},\end{aligned}$$

and this bound goes to zero as a power of n since $2k\epsilon > 1$ and the last factor is polylogarithmic in n. This completes the proof of Theorem 2.2.

4. LOCAL ANALYSIS: STRUCTURE OF NODE NEIGHBORHOODS

This section establishes properties of node neighborhoods. We start with general bounds. We then relate vectors of interest $B^{(\ell)}e$ and $B^{(\ell)}\sigma$ to the neighborhood structures. The martingale analysis of neighborhood structures follows. For any $k \geq 0$, the number of nodes with spin $\pm$ at distance k (respectively $\leq k$) of node i is denoted $U_k^{\pm}(i)$ (respectively, $U_{\leq k}^{\pm}(i)$). We thus have

$$S_t(i) = U_t^+(i) + U_t^-(i), \; D_t(i) = U_t^+(i) - U_t^-(i). \quad (20)$$

Index i is omitted when considering a fixed node i. In the remainder of the section we condition on the spins σ of all nodes. We denote $n_\pm$ as the number of nodes with spin $\pm$.

For fixed $i \in \mathcal{N}$ it is readily seen that, conditionally on $\mathcal{F}_{k-1} := \sigma(U_t^+, U_t^-, t \le k-1)$,

$$\begin{aligned} U_k^+ &\sim \mathrm{Bin}\left(n_+ - U_{\le k-1}^+, 1-(1-\tfrac{a}{n})^{U_{k-1}^+}(1-\tfrac{b}{n})^{U_{k-1}^-}\right), \\ U_k^- &\sim \mathrm{Bin}\left(n_- - U_{\le k-1}^-, 1-(1-\tfrac{a}{n})^{U_{k-1}^-}(1-\tfrac{b}{n})^{U_{k-1}^+}\right). \end{aligned} \tag{21}$$

Theorem 2.3 is proven in the Appendix from these characterizations and Chernoff bounds for binomial variables.

The next technical result establishes approximate independence of neighborhoods of distinct nodes. It is instrumental in Section 4.2 e.g. in establishing weak laws of large numbers on the fraction of nodes satisfying a given property.

LEMMA 4.1. *Consider any two fixed nodes i, j with $i \neq j$. Let $\ell = c\log(n)$ where constant c is such that $c\log(\alpha) < 1/2$. Then the variation distance between the joint law of their neighborhood processes $\mathcal{L}((U_k^\pm(i))_{k\le\ell}, (U_k^\pm(j))_{k\le\ell})$ and the law with the same marginals and independence between them, denoted $\mathcal{L}((U_k^\pm(i))_{k\le\ell}) \otimes \mathcal{L}((U_k^\pm(j))_{k\le\ell})$, goes to zero as a negative power of n as $n \to \infty$.*

PROOF. Take two independent realizations of the processes $(U_k^\pm(i))_{k\le\ell}$ and $(U_k^\pm(j))_{k\le\ell}$. Use them to perform a joint construction of the two processes as follows. Having constructed the corresponding sets $\mathcal{U}_t^\pm(i) \subset \mathcal{N}$, $\mathcal{U}_t^\pm(j) \subset \mathcal{N}$ for $t = 1, \ldots, k-1$ and assuming the i-sets and the j-sets have not yet met, we construct them at step k as follows. To construct $\mathcal{U}_k^\pm(i)$ we select a size $U_k^\pm(i)$ subset uniformly at random from $\mathcal{N}^\pm \setminus \mathcal{U}_{\le k-1}^\pm(i)$. We do similarly for j. The construction can proceed based on the independent inputs so long as the resulting i-sets and j-sets do not intersect. However on $\cap_{t\le k}\{S_t(i) \vee S_t(j) \le C\log(n)\alpha^t\}$, the expected size of the intersection will be upper-bounded by $O(\log^2(n)\alpha^{2k}/n) = O(\log^2(n)n^{-2\epsilon})$, where $c\log(\alpha) = 1/2 - \epsilon$. Theorem 2.3 ensures that the probability of $\cap_{t\le k}\{S_t(i) \vee S_t(j) \le C\log(n)\}$ is $1 - O(n^{-\epsilon})$ and the result follows. ☐

We now state a lemma on the presence of cycles in the ℓ-neighborhoods of nodes. It will be instrumental in bounding the discrepancy between vectors $B^{(\ell)}e$ (resp. $B^{(\ell)}\sigma$) and $\{S_\ell(i)\}$ (resp. $\{D_\ell(i)\}$). Its proof, deferred to the Appendix, relies on the previous coupling Lemma 4.1.

LEMMA 4.2. *Assume $\ell = c\log(n)$ with $c\log(\alpha) < 1/2$. Then with high probability the number of nodes i whose ℓ-neighborhood contains one cycle is $O(\log^4(n)\alpha^{2\ell})$. Assume further that $c\log(\alpha) < 1/4$. Then with high probability no node i has more than one cycle in its ℓ-neighborhood.*

4.1 From variables S_t and D_t to matrix $B^{(\ell)}$

We first state how to transport the deterministic growth controls (11) of Theorem 2.3 to vectors $B^{(m-1)}e$ and $B^{(m-1)}\sigma$, a key step in the proof of Theorem 2.4. One has the following

LEMMA 4.3. *Let $\mathcal{B}$ denote the set of nodes i whose ℓ-neighborhood contains a cycle. For $\ell = c\log n$ with $c\log\alpha < 1/4$, with high probability, for $m \le \ell$ and $i \notin \mathcal{B}$:*

$$\begin{cases} (B^{(m-1)}e)_i = \alpha^{m-1-\ell}(B^{(\ell)}e)_i + \tilde{O}(\sqrt{\alpha^{m-1}}), \\ (B^{(m-1)}\sigma)_i = \beta^{m-1-\ell}(B^{(\ell)}\sigma)_i + \tilde{O}(\sqrt{\alpha^{m-1}}), \end{cases} \tag{22}$$

while for $i \in \mathcal{B}$:

$$|(B^{(m)}\sigma)_i| \le |(B^{(m)}e)_i| \le 2\sum_{t=0}^{m} S_i(t) = \tilde{O}(\alpha^m). \tag{23}$$

Proof is in the Appendix, together with that of the following

COROLLARY 4.1. *For all $m \in \{1, \ldots, \ell\}$ it holds with high probability that*

$$\sup_{|x|=1, x'B^{(\ell)}e = x'B^{(\ell)}\sigma = 0} |e'B^{(m-1)}x| = \tilde{O}\left(\sqrt{n}\sqrt{\alpha^{m-1}}\right), \tag{24}$$

$$\sup_{|x|=1, x'B^{(\ell)}e = x'B^{(\ell)}\sigma = 0} |\sigma'B^{(m-1)}x| = \tilde{O}\left(\sqrt{n}\sqrt{\alpha^{m-1}}\right). \tag{25}$$

We are now ready to prove Theorem 2.4:

PROOF. (of Theorem 2.4). Using identity (6), write for unit norm x:

$$|B^{(\ell)}x| \le \rho(\Delta^{(\ell)}) + \sum_{m=1}^{\ell} \rho(\Delta^{(\ell-m)})|\bar{A}B^{(m-1)}x| + \sum_{m=1}^{\ell} \rho(\Gamma^{\ell,m}).$$

The terms $\rho(\Delta^{(\ell)})$ and $\rho(\Gamma^{\ell,m})$ are less than $n^\epsilon\alpha^{\ell/2}$ by (7) and (8). Expression (5) of $\bar{A}$, Cauchy-Schwarz inequality and the fact that $|e| = |\sigma| = \sqrt{n}$ yield

$$|\bar{A}B^{(m-1)}x| \le \frac{a}{n}|B^{(m-1)}x| + O\left(\frac{|\sigma'B^{(m-1)}x| + |e'B^{(m-1)}x|}{\sqrt{n}}\right).$$

Using bounds (24,25), the right-hand side is no larger than

$$\frac{a}{n}|B^{(m-1)}x| + \tilde{O}\left(\sqrt{\alpha^{m-1}}\right).$$

By the previous inequalities (10,22,23) and the row sum bound, we have that

$$\rho(B^{(m-1)}) = \tilde{O}(\alpha^{m-1}).$$

This thus yields

$$\begin{aligned} |\bar{A}B^{(m-1)}x| &\le \tilde{O}\left(\tfrac{\alpha^{m-1}}{n} + \sqrt{\alpha^{m-1}}\right) \\ &= \tilde{O}(\sqrt{\alpha^{m-1}}). \end{aligned}$$

We thus have

$$\begin{aligned} |B^{(\ell)}x| &\le n^\epsilon\alpha^{\ell/2} + \textstyle\sum_{m=1}^{\ell} \rho(\Delta^{(\ell-m)})\tilde{O}(\sqrt{\alpha^{m-1}}) \\ &\le n^\epsilon\alpha^{\ell/2} + \textstyle\sum_{m=1}^{\ell} n^\epsilon\alpha^{(\ell-m)/2}\tilde{O}(\sqrt{\alpha^{m-1}}) \\ &\le n^\epsilon\alpha^{\ell/2}\tilde{O}(1). \end{aligned}$$

The result readily follows. ☐

We now state two Lemmas which will allow to establish Theorem 4.1.

LEMMA 4.4. *Assume (1) and $\ell = c\log n$ with $c\log(\alpha) < 1/4$. Then with high probability one has*

$$\begin{aligned} |B^{(\ell)}e - \{S_\ell(i)\}_{i\in\mathcal{N}}| &= o\left(|B^{(\ell)}e|\right), \\ |B^{(\ell)}\sigma - \{D_\ell(i)\}_{i\in\mathcal{N}}| &= o\left(|B^{(\ell)}\sigma|\right), \\ < B^{(\ell)}e, B^{(\ell)}\sigma > &= o\left(|B^{(\ell)}e| \times |B^{(\ell)}\sigma|\right). \end{aligned} \tag{26}$$

LEMMA 4.5. *Assume (1) and $\ell = c\log n$ with $c\log(\alpha) < 1/4$. Then for some fixed $\gamma > 0$ with high probability one has*

$$\Omega(\alpha^\ell)|B^{(\ell)}e| \le |B^{(\ell)}B^{(\ell)}e| \le O(\log(n)\alpha^\ell)|B^{(\ell)}e|, \tag{27}$$

$$\Omega(\beta^\ell)|B^{(\ell)}\sigma| \le |B^{(\ell)}B^{(\ell)}\sigma| \le O(n^{-\gamma}\alpha^\ell)|B^{(\ell)}\sigma|. \tag{28}$$

Using these, we now establish the following

THEOREM 4.1. *Assume (1) and $\ell = c\log n$ with $c\log(\alpha) < 1/4$. Then with high probability the two leading eigenvectors of $B^{(\ell)}$ are asymptotically aligned with vectors $\{S_\ell(i)\}$, $\{D_\ell(i)\}$, with corresponding eigenvalues of order $\tilde{\Theta}(\alpha^\ell)$ and $\Omega(\beta^\ell)$; all other eigenvalues are $O(n^\epsilon\sqrt{\alpha^\ell})$ for any fixed $\epsilon > 0$.*

PROOF. Estimates (27–28) and the weak Ramanujan property of Theorem 2.4 imply that the leading eigenvector is aligned with $B^{(\ell)}e$ and has eigenvalue $\tilde{\Theta}(\alpha^\ell)$. They also imply that the second eigenvector is asymptotically in the span of $\{B^{(\ell)}e, B^{(\ell)}\sigma\}$ and with eigenvalue $\Omega(\beta^\ell)$. By asymptotic orthonormality of vectors $B^{(\ell)}e$ and $B^{(\ell)}\sigma$ and their asymptotic alignment with $\{S_\ell(i)\}$, $\{D_\ell(i)\}$ respectively, the conclusion regarding the first two eigen-elements follows. The bound on the magnitude of other eigenvalues follows from Theorem 2.4 and the Courant-Fisher theorem. □

4.2 Coupling with Poisson tree growth process

Introduce the stochastic process $\{V_t^\pm\}_{t\geq 0}$ defined by

$$\begin{array}{l} V_0^+ = 1,\ V_0^- = 0,\\ V_t^+,\ V_t^- \text{independent conditionally on } \mathcal{G}_{t-1},\\ \mathcal{L}(V_t^\pm|\mathcal{G}_{t-1}) = \mathrm{Poi}((a/2)V_{t-1}^\pm + (b/2)V_{t-1}^\mp) \end{array} \tag{29}$$

where $\mathcal{G}_{t-1} = \sigma(V^{\pm_k}, k \leq t-1)$. The following is a version of Proposition 4.2 in [2]. The reader is addressed to either [2] or [14] for a proof based on the Stein-Chen method for Poisson approximation.

LEMMA 4.6. *Let $i \in \mathcal{N}$ be fixed with spin $\sigma_i = \sigma$. For a constant $c > 0$ such that $c\log(\alpha) < 1/2$, and $\ell = c\log(n)$, the following holds. The variation distance between $(U_t^\pm(i))_{t\leq\ell}$ and $(V_t^{\pm\sigma})_{t\leq\ell}$ goes to zero as a negative power of n as $n \to \infty$.*

Define now the processes

$$\begin{array}{ll} M_t & = \alpha^{-t}(V_t^+ + V_t^-),\\ \Delta_t & = \beta^{-t}(V_t^+ - V_t^-), \end{array} \tag{30}$$

where $V_t^\pm$ is as in (29). We will need the following results on these processes, which follow from Kesten and Stigum [15] (see also [14] for a direct proof).

LEMMA 4.7. *Processes $\{M_t\}$, $\{\Delta_t\}$ are $\mathcal{G}_t$-martingales. Process $\{M_t\}$ is uniformly integrable when $\alpha > 1$. Under Condition (1) process $\{\Delta_t\}$ is uniformly integrable.*

COROLLARY 4.2. *Under Condition (1) the martingale $\{\Delta_t\}$ converges almost surely to a unit mean random variable Δ_∞. Moreover this random variable has a finite variance $1/(\beta^2/\alpha - 1)$ to which the variance of Δ_t converges. It further holds that $\mathbf{E}|\Delta_t^2 - \Delta_\infty^2| \to 0$ as $t \to \infty$.*

Together these properties allow to establish the following

THEOREM 4.2. *One has the following convergence in probability*

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^n \beta^{-2\ell}D_\ell^2(i) = \mathbf{E}(\Delta_\infty^2). \tag{31}$$

Let $y \in \mathbb{R}^n$ be the normed vector defined as

$$y_i = \frac{D_\ell(i)}{\sqrt{\sum_{j=1}^n D_\ell(j)^2}},\ i = 1,\dots,n. \tag{32}$$

Let x be a vector in $\mathbb{R}^n$ such that we have the convergence in probability

$$\lim_{n\to\infty} ||x - y|| = 0. \tag{33}$$

For all $\tau \in \mathbb{R}$ that is a point of continuity of the distribution of both Δ_∞ and $-\Delta_\infty$, one has the following convergence in probability for both signs $\pm$

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i\in\mathcal{N}:\sigma_i=\pm} \mathbf{1}_{x_i\geq\tau/\sqrt{n\mathbf{E}(\Delta_\infty^2)}} = \frac{1}{2}\mathbf{P}(\pm\Delta_\infty \geq \tau). \tag{34}$$

We now establish (31); the rest of the proof, which relies on similar ideas, is deferred to the Appendix.

PROOF. (of (31)) By the coupling lemma 4.6, with probability $1 - O(n^{-\epsilon})$ for fixed positive ϵ, $\sigma(i)\beta^{-\ell}D_\ell(i)$ coincides with Δ_ℓ, an event denoted $\mathcal{C}$. Define events $\Omega_k(i)$ by

$$\Omega_k(i) = \{S_k(i) \leq C(\log n)\alpha^k\} \tag{35}$$

and Ω by

$$\Omega := \cap_{i\in\mathcal{N}}\Omega(i) \text{ where } \Omega(i) := \cap_{k\leq\ell}\Omega_k(i), \tag{36}$$

where constant C is as in Theorem 2.3. When $\mathcal{C}$ fails, $\beta^{-\ell}D_\ell(i)$ is $O(\log(n))$ on the event Ω. Let $\tilde{\mathbf{P}}$ and $\tilde{\mathbf{E}}$ denote probability and expectation conditional on Ω respectively. The left-hand side of (31) thus verifies

$$\tilde{\mathbf{E}}\left(\frac{1}{n}\sum_{i=1}^n \beta^{-2\ell}D_\ell^2(i)\right) = O(\log^2(n))n^{-\epsilon} + \tilde{\mathbf{E}}(\Delta_\ell^2\mathbf{1}_\mathcal{C}).$$

Write

$$\begin{array}{ll} |\tilde{\mathbf{E}}(\Delta_\ell^2\mathbf{1}_\mathcal{C}) - \mathbf{E}(\Delta_\infty^2)| & \leq (\frac{1}{\mathbf{P}(\Omega)} - 1)\mathbf{E}(\Delta_\infty^2) + \frac{\mathbf{E}|\Delta_\ell^2 - \Delta_\infty^2|}{\mathbf{P}(\Omega)}\\ & + \frac{\mathbf{E}\Delta_\infty^2\mathbf{1}_{\overline{\Omega\cap\mathcal{C}}}}{\mathbf{P}(\Omega)}. \end{array}$$

By Theorem 2.3 and Corollary 4.2, the first and second term in the right-hand side go to zero with n and ℓ respectively; the third term goes to zero as $\mathbf{P}(\Omega\cap\mathcal{C}) \to 1$ (e.g. by Hardy-Littlewood-Polya's rearrangement inequalities). Thus the expectation of the left-hand side of (31) converges to $\mathbf{E}\Delta_\infty^2$.

We now evaluate $\tilde{\mathbf{E}}\left(\frac{1}{n}\sum_{i=1}^n \beta^{-2\ell}D_\ell^2(i)\right)^2$, the second moment of this empirical sum. We break it into two terms, the first being

$$\frac{1}{n^2}\tilde{\mathbf{E}}\sum_{i=1}^n \beta^{-4\ell}D_\ell^4(i).$$

By Theorem 2.3, this is $O(\log(n)^4)/n = o(1)$. The second term reads

$$\frac{2}{n^2}\sum_{i<j}\beta^{-4\ell}\tilde{\mathbf{E}}(D_\ell^2(i)D_\ell^2(j)). \tag{37}$$

Fix $i < j$. As $\mathbf{P}(\Omega) \geq 1 - O(n^{-\epsilon})$, (37) is equivalent to $\mathbf{E}(\mathbf{1}_\Omega\beta^{-4\ell}D_\ell^2(i)D_\ell^2(j))$, itself equivalent to

$$\mathbf{E}(\mathbf{1}_{\Omega(i)\cap\Omega(j)}\beta^{-4\ell}D_\ell^2(i)D_\ell^2(j)) + O(\log^4(n)\mathbf{P}(\overline{\Omega})).$$

The second term in the right-hand side is $o(1)$ by Theorem 2.3, while the first term in the right-hand side is asymptotic to $\left(\mathbf{E}(\mathbf{1}_{\Omega(i)}\beta^{-2\ell}D_\ell^2(i))\right)^2$ by Lemma 4.1. This is in turn equivalent to $[\mathbf{E}(\Delta_\infty^2)]^2$ by the analysis of the first moment of the empirical sum. It readily follows that

$$\lim_{n\to\infty}\tilde{\mathbf{E}}\left[\frac{1}{n}\sum_{i=1}^n \beta^{-2\ell}D_\ell^2 - \mathbf{E}(\Delta_\infty^2)\right]^2 = 0.$$

Convergence (31) then follows by Tchebitchev's inequality. □

Theorems 4.1 and 4.2 readily imply Theorem 2.1.

5. CONCLUSIONS

The methods developed here may find further applications, e.g. to prove the more general conjecture by Heimlicher et al. [4] of a phase transition in the labeled stochastic block model or the "spectral redemption" conjecture of [9]. More generally one might ask what is the range of applicability of path expansion approaches to "fix" spectral methods by recovering Ramanujan-like spectral separation properties. It is likely that a similar regularization would occur by considering matrix $\hat{B}$ defined by $\hat{B}_{ij} = \mathbf{1}_{d_G(i,j)=\ell}$ but we have not been able to prove this yet.

acknowledgements: The author gratefully acknowledges stimulating discussions on the topic with Marc Lelarge and Charles Bordenave.

6. REFERENCES

[1] A. Decelle, F. Krzakala, C. Moore, and L. Zdeborová, "Asymptotic analysis of the stochastic block model for modular networks and its algorithmic applications," *Physics Review E*, vol. 84:066106, 2011.

[2] E. Mossel, J. Neeman, and A. Sly, "Reconstruction and estimation in the planted partition model," Feb. 2012, available at: http://arxiv.org/abs/1202.1499.

[3] P. W. Holland, K. B. Laskey, and S. Leinhardt, "Stochastic blockmodels: First steps," *Social Networks*, vol. 5, no. 2, pp. 109–137, 1983.

[4] S. Heimlicher, M. Lelarge, and L. Massoulié, "Community detection in the labelled stochastic block model," *NIPS Workshop on Algorithmic and Statistical Approaches for Large Social Networks*, 2012.

[5] W. Evans, C. Kenyon, Y. Peres, and L. Schulman, "Broadcasting on trees and the Ising model," *The Annals of Applied Probability*, vol. 10, no. 2, pp. 410–433, 2000.

[6] A. Coja-Oghlan, "Graph partitioning via adaptive spectral techniques," *Comb. Probab. Comput.*, vol. 19, no. 2, pp. 227–284, 2010. [Online]. Available: http://dx.doi.org/10.1017/S0963548309990514

[7] J. Friedman, "A proof of Alon's second eigenvalue conjecture and related problem," *Mem. Amer. Math. Soc.*, no. 910., 2008.

[8] U. Feige and E. Ofek, "Spectral techniques applied to sparse random graphs," *Random Struct. Algorithms*, vol. 27, no. 2, pp. 251–275, Sept. 2005. [Online]. Available: http://dx.doi.org/10.1002/rsa.v27:2

[9] F. Krzakala, C. Moore, E. Mossel, J. Neeman, A. Sly, L. Zdeborová, and P. Zhang, "Spectral redemption: clustering sparse networks," *Proceedings of the National Academy of Sciences*, no. 110(52), pp. 20 935–20 940, 2013.

[10] R. R. Nadakuditi and M. E. J. Newman, "Graph spectra and the detectability of community structure in networks," *Phys. Rev. Lett.*, no. 108, 2012.

[11] E. Mossel, J. Neeman, and A. Sly, "Belief propagation, robust reconstruction and optimal recovery of block models," Sep. 2013, available at: http://arxiv.org/abs/1309.1380.

[12] ——, "A proof of the block model threshold conjecture," Nov. 2013, available at: http://arxiv.org/abs/1311.4115.

[13] Z. Füredi and J. Komlós, "The eigenvalues of random symmetric matrices," *Combinatorica*, no. 1(3), pp. 233–241, 1981.

[14] L. Massoulié, "Community detection thresholds and the weak Ramanujan property," Nov. 2013, available at: http://arxiv.org/abs/1311.3085.

[15] H. Kesten and B. P. Stigum, "Additional limit theorems for indecomposable multidimensional Galton-Watson processes," *Ann. Math. Statist.*, no. 37, pp. 1463–1481, 1966.

APPENDIX

A. PROOF OF PROPOSITION 3.2

PROOF. The trace of $[\Gamma^{\ell,m}(\Gamma^{\ell,m})']^k$ upper-bounds ρ^{2k} where ρ is the spectral radius of $\Gamma^{\ell,m}$. This trace is the sum over circuits of length $2\ell k$ of products of terms that can be either $A_e - \bar{A}_e$, $\bar{A}_e$ or A_e such that a length ℓ chunk of the circuit is the concatenation of two simple paths of length $m-1$ and $\ell - m$, and that the two of them have a non-empty intersection.

We represent such contributions as follows. Denote by v the number of nodes and by e the number of edges traversed by the circuit, while ignoring edges that are weighed by an $\bar{A}$-term. Note that by the constraint that the concatenated simple parts of each length ℓ-chunk intersect, the corresponding graph is necessarily connected. We adopt the following representation of the corresponding circuits.

Nodes are again denoted by the order in which they are first met, starting with node 1. We represent each simple path that constitutes the circuit by sequences of three phases as in the proof of Proposition 3.1. Note that there are now $4k$ such simple paths: each length ℓ chunk of the original circuit is broken into an $m-1$- and an $\ell-m$-path. We adopt the same representation as before, except that we must now also incorporate the label of the starting point after traversal of an $\bar{A}$-edge.

Thus we have the upper bound on the number of valid circuit labels with v nodes and e edges:

$$v^{2k}[(v+1)^2(\ell+1)]^{4k(1+e-v+1)}.$$

Let us bound the values that v and e can take. Necessarily, $v \geq \max(m, \ell-m+1)$: indeed, each length ℓ chunk comprises simple paths of length $m-1$ and $\ell-m$. Moreover, there are overall $2k(\ell-1)$ edges (recall that we discount the $\bar{A}$-edges). Out of these, $2k(\ell-m)$ must be doubled for the expectation not to vanish. There are thus at most $1+k(\ell+m)$ nodes v in total, and at most $k(\ell+m)$ distinct edges in total. Finally, bounding the contribution of each $\bar{A}$ term and each cycle edge by $(a \vee b)/n$, arguing as in the proof of Proposition 3.1 one obtains (19). $\square$

B. PROOF OF THEOREM 2.3

The following inequality is easily seen to hold for any non-negative u, v, a, b, n such that $a/n, b/n \leq 1$:

$$\frac{au+bv}{n} - \frac{1}{2}\left(\frac{au+bv}{n}\right)^2 \leq 1-(1-\frac{a}{n})^u(1-\frac{b}{n})^v \leq \frac{au+bv}{n}. \quad (38)$$

Next lemma is the key ingredient to establish Theorem 2.3.

LEMMA B.1. *Let $\epsilon \in (0,1)$, $\gamma > 0$ and $\ell = c\log(n)$ with $c\log(\alpha) < 1/2$. Then there exists some constant $K > 0$ such* *that with probability $1-O(n^{-\gamma})$ the following properties hold for all $i \in \mathcal{N}$ and all $t \le \ell$.*

(i) For $T := \inf\{t \le \ell : S_t \ge K\log(n)\}$, then $S_T = \Theta(\log(n))$.

(ii) Let $\epsilon_t := \epsilon\alpha^{-(t-T)/2}$. Then for all $t, t' \in \{T, \dots, \ell\}$, $t > t'$, the vector $U_t = (U_t^+, U_t^-)$ verifies the coordinate-wise bounds:

$$U_t \in \left[\prod_{s=t'+1}^{t}(1-\epsilon_s)M^{t-t'+1}U_{t'}, \prod_{s=t'+1}^{t}(1+\epsilon_s)M^{t-t'+1}U_{t'}\right], \tag{39}$$

where M denotes the matrix $(a/2\ b/2, b/2\ a/2)$.

Proof. By definition of T, $U_{T-1} < K\log(n)$. Thus by (21)

$$U_T^{\pm} \le \text{Bin}(n^{\pm}, (a \vee b)\frac{K\log(n)}{n}).$$

The mean of the Binomial distribution in the right-hand side of the above is equivalent to $(a \vee b)(1/2)K\log(n)$ and less than $\kappa\log(n)$ for $\kappa = (a \vee b)K$. Hence by Chernoff's inequality, for $h(x) := x\log(x) - x + 1$ and $K' > 2\kappa$,

$$\mathbf{P}(U_T^{\pm} \le K'/2\log(n)|\mathcal{F}_{T-1}) \ge e^{-\kappa\log(n)h(K'/2\kappa)}.$$

Take $K' > 2\kappa$ so that $\kappa h(K'/2\kappa) > 2 + \gamma$. The right-hand side of the above is then no larger than $n^{-2-\gamma}$.

Thus by the union bound, Property (i) holds with probability $1 - O(n^{-1-\gamma})$ for all $i \in \mathcal{N}$.

Conditional on $\mathcal{F}_T$, the binomial distribution of $U_{T+1}^{\pm}$ has mean

$$[n^{\pm} - U_{<T+1}^{\pm}] \times [1 - (1-a/n)^{U_T^{\pm}}(1-b/n)^{U_T^{\mp}}],$$

which by the inequalities (38) lies in the interval

$$[(a \wedge b)\frac{1}{3}K\log(n), (a \vee b)K'\log(n)].$$

For a given $\epsilon > 0$, we can choose K sufficiently large so that[2]

$$(a \wedge b)\frac{1}{3}Kh(1+\epsilon) > 2 + \gamma.$$

It follows that $U_{T+1}^{\pm}$ admits a relative deviation from its conditional mean by ϵ with probability at most $n^{-2-\gamma}$.

We now define the events $\mathcal{A}_t := \{U_t^{\pm} \in [1-\epsilon_t, 1+\epsilon_t]\frac{aU_{t-1}^{\pm} + bU_{t-1}^{\mp}}{2}\}$ where ϵ_t is as in Statement (ii) of the Lemma. Conditionally on $\mathcal{A}_T, \dots, \mathcal{A}_t$, the vector $U_t = (U_t^+, U_t^-)$ verifies the announced inequality (39). This in turn implies that $U_t^{\pm} \ge (1-O(\epsilon))\alpha^{t-T}K''\log(n)$ where $K'' := (a \wedge b)K/3$. We now check that Chernoff's bound applies to show by induction that (39) holds at step t with high enough probability. It suffices to ensure that

$$U_t^{\pm}\tilde{h}(\epsilon_t) \ge (2+\gamma)\log(n),$$

where $\tilde{h}(u) := \min[(1+u)\log(1+u) - u, (1-u)\log(1-u) + u)$. However as we just saw the left-hand side of this expression is lower-bounded by

$$(1-O(\epsilon))\alpha^{t-T}K''\log(n)\tilde{h}(\epsilon_t) \ge (1-O(\epsilon))\alpha^{t-T}K''\log(n)\frac{\epsilon_t^2}{3},$$

where we took a second-order expansion of $\tilde{h}$ around 0. The condition is therefore met as soon as $(1-O(\epsilon))K''\epsilon^2/3 \ge 2 + \gamma$. For K large enough this holds. □

[2]We assume here that $a \wedge b > 0$, as the result trivially holds if $a \wedge b = 0$.

Proof. (of Theorem 2.3). For $t \le \ell$, if $t \le T$, we necessarily have that $S_t, |D_t| = O(\log n)$. Consider then $t > T$. Note that matrix M is such that

$$M^k = \frac{1}{2}\begin{pmatrix} \alpha^k + \beta^k & \alpha^k - \beta^k \\ \alpha^k - \beta^k & \alpha^k + \beta^k \end{pmatrix}.$$

Using (39), we readily have for $t, t' \le T$, with $t > t'$:

$$\begin{aligned} S_t &\le \textstyle\prod_{s=t'+1}^{t}(1+\epsilon_s)(1,1)M^{t-t'}U_{t'} \\ &= \textstyle\prod_{s=t'+1}^{t}(1+\epsilon_s)\alpha^{t-t'}S_{t'}. \end{aligned}$$

A similar lower bound holds with $-\epsilon_s$ in place of $+\epsilon_s$. Setting $t' = T$ in the upper bound, since $S_T = O(\log(n))$, the upper bound (10) follows for S_t, as $\prod_{s=T+1}^{t}(1+\epsilon_s) = O(1)$.

It readily follows that (11) holds for S_t by noting that

$$\begin{aligned} \max\left(\prod_{s=t'+1}^{t}(1+\epsilon_s) - 1, 1 - \prod_{s=t'+1}^{t}(1-\epsilon_s)\right) &= O(\epsilon_{t'}) \\ &= O(\alpha^{-t'/2}). \end{aligned}$$

Consider now D_t. Using (39) again, we have

$$\beta D_{t-1} - \alpha\epsilon_t S_t \le D_t \le \beta D_{t-1} + \alpha\epsilon_t S_t.$$

Iterating, we obtain

$$|D_t - \beta^{t-t'}D_{t'}| \le \sum_{s=t'+1}^{t}\alpha\beta^{t-s}\epsilon_s S_s. \tag{40}$$

Since $S_s = O(\log(n)\alpha^{s-T})$, $|D_T| = O(\log(N))$ and $\epsilon_s = O(\alpha^{-(s-T)/2})$, we obtain for $t' = T$:

$$\begin{aligned} |D_t| &= O\big(\log(n)\beta^t + \textstyle\sum_{s=T+1}^{t}\beta^{t-s}\log(n)\alpha^{(s-T)/2}\big) \\ &= O(\log(n)\beta^t), \end{aligned}$$

where we used inequality $\beta^2 > \alpha$ to bound $\sum_{u>0}\beta^{-u}\alpha^{u/2}$. Property (10) thus holds for D_t.

Finally, the right-hand side of (40) is of order

$$\sum_{s=t'+1}^{t}\beta^{t-s}\alpha^{(s-T)/2}\log(n) = O(\log(n)\beta^{t-t'}\alpha^{t'/2}).$$

Thus setting $t = \ell$, for $\ell > t' \ge T$ we have

$$D_{t'} = \beta^{t'-\ell}D_\ell + O(\log(n)\alpha^{t'/2}).$$

Since for $t' < T$ we readily have $D_{t'} = O(\log(n)$ by definition of T, property (11) follows for D_t. □

C. PROOF OF LEMMA 4.2

Proof. There are two ways for creating cycles within the distance k-neighborhood of i: an edge may be present between two nodes at distance $k-1$ of i, or two nodes at distance $k-1$ may be connected to the same node at distance k of i. The number of edges of the first type is stochastically dominated by $\text{Bin}(S_{k-1}^2, a \vee b/n)$. Its expected number conditionally on $\Omega_{k-1}(i)$, as defined in (35), is at most $O(\log^2(n)\alpha^{2\ell}/n)$. Thus by the union bound the probability that there is such an edge in the ℓ-neighborhood of i is, by Theorem 2.3 at most:

$$\ell \times O(\log^2(n)\alpha^{2\ell}/n) + \sum_{k=1}^{\ell}(1 - \mathbf{P}(\Omega_k(i))) = O(\log^3(n)\alpha^{2\ell}/n).$$

The number of cycle edges of the second type is stochastically dominated by

$$\mathrm{Bin}(n, (a \vee b/n)^2 S_{k-1}^2).$$

On $\Omega_{k-1}(i)$ its conditional expectation is $O(\log^2(n)\alpha^{2\ell})$.

By the same argument, the probability that there are two cycle-edges within the ℓ-neighborhood of i is upper-bounded by $O(\log^6(n)\alpha^{4\ell}/n^2)$. By the union bound we readily have that with high probability no node has two cycle-edges within its ℓ-neighborhood as soon as $\log^6(n)\alpha^{4\ell} \ll n$, which holds for $\ell = c\log(n)$ with $c\log(\alpha) < 1/4$.

Let Z_i denote the event that the ℓ-neighborhood of i contains a cycle. On the event Ω defined by (36), the ℓ-neighborhoods of an arbitrary pair of distinct nodes i, j are disjoint with probability $1 - O(\log^2(n)\alpha^{2\ell}/n)$, conditionally upon which the probability that they both have a cycle in their neighborhood is upper-bounded by $O(\log^6(n)\alpha^{4\ell}/n^2)$. Conditionally on the event that their neighborhoods meet, the expectation of the product Z_iZ_j is still upper-bounded by $O(\log^3(n)\alpha^{2\ell}/n)$.

Eventually Markov's inequality yields

$$\begin{aligned}
&\mathbf{P}(\textstyle\sum_i Z_i \geq m\log^3(n)\alpha^{2\ell}) \leq \frac{\mathbf{E}(\sum_i Z_i)^2}{m^2\log^6(n)\alpha^{4\ell}}\\
&\leq \frac{n\mathbf{E}(Z_1)+n^2\mathbf{E}(Z_1Z_2)}{m^2\log^6(n)\alpha^{4\ell}}\\
&\leq \frac{O(\log^3(n)\alpha^{2\ell})+n^2[O(\log^6(n)\alpha^{4\ell}/n^2)+(\log^2(n)\alpha^{2\ell}/n)(\log^3(n)\alpha^{2\ell}/n)}{m^2\log^6(n)\alpha^{4\ell}}\\
&= O(\tfrac{1}{m^2}).
\end{aligned}$$

Taking $m = \log(n)$ (say), then with high probability $\sum_i Z_i = O(\log^4(n)\alpha^{2\ell})$. □

D. PROOF OF LEMMA 4.3

Let $i \notin \mathcal{B}$ be a node whose ℓ-neighborhood is a tree. For any $k \in \mathcal{N}$ and any $m \leq \ell$, $B_{ik}^{(m)} = \mathbf{1}_{d_{\mathcal{G}}(i,k)=m}$. For such i, one therefore has the following identities:

$$i \notin \mathcal{B} \Rightarrow \begin{cases} (B^{(m)}e)_i = S_i(m), \\ (B^{(m)}\sigma)_i = D_i(m), \end{cases} \tag{41}$$

Relations (22) readily follow from Theorem 2.3.

Let $i \in \mathcal{B}$. By Lemma 4.2, such nodes i have in their ℓ-neighborhood only one cycle. Clearly only nodes at distance at most ℓ of i can be counted in $(B^{(\ell)}e)_i$, and they can be counted at most twice because the neighborhood contains only one cycle. Control (23) readily follows.

E. PROOF OF COROLLARY 4.1

Proof. Let $\mathcal{B}$ denote the set of nodes i such that their ℓ-neighborhood contains a cycle. Let x be a normed vector such that $x'B^{(\ell)}e = 0$. We then have

$$\begin{aligned}
&|e'B^{(m-1)}x| = |\textstyle\sum_{i\in\mathcal{N}} x_i(B^{(m-1)}e)_i|\\
&\leq |\textstyle\sum_{i\in\mathcal{B}} x_i(B^{(m-1)}e)_i| + |\sum_{i\in\mathcal{B}} x_i\alpha^{m-1-\ell}(B^{(\ell)}e)_i|\\
&\quad+|\textstyle\sum_{i\in\mathcal{N}} x_i[\alpha^{m-1-\ell}(B^{(\ell)}e)_i + O(\log(n)+\sqrt{\log(n)\alpha^{m-1}})|.
\end{aligned} \tag{42}$$

Using the bound (23) for $i \in \mathcal{B}$, we can bound the first summation, using Cauchy-Schwarz's inequality by

$$\begin{aligned}|\textstyle\sum_{i\in\mathcal{B}} x_i[(B^{(m-1)}e)_i| &\leq O(\log(n)\alpha^{m-1})\sqrt{|\mathcal{B}|}\\ &\leq O(\log^3(n)\alpha^{\ell+m-1}),\end{aligned}$$

where we have used the bound on the size of $\mathcal{B}$ derived in Lemma 4.2. The second summation in (42) is similarly bounded. As for the third summation, using the fact that $e'B^{(\ell)}x = 0$, it is upper-bounded by

$$|\sum_{i\in\mathcal{N}} x_iO(\log(n)+\sqrt{\log(n)\alpha^{m-1}})|.$$

By Cauchy-Schwarz again, this is no larger than

$$O\left(\sqrt{n}(\log(n)+\sqrt{\log(n)\alpha^{m-1}})\right).$$

The announced bound (24) on $|e'B^{(m-1)}x|$ follows. Similarly, the bound (25) on $|\sigma'B^{(m-1)}x|$ follows by using property $\sigma'B^{(\ell)}x = 0$ instead of property $e'B^{(\ell)}x = 0$. □

F. PROOF OF THEOREM 4.2

Proof. We now turn to establishing (34). We shall only consider the case of sign +, the other being handled similarly. Fix some arbitrarily small $\delta > 0$. Because τ is a continuity point of the distribution of Δ_∞, we can find two bounded Lipschitz-continuous functions f, g such that

$$f(u) \leq \mathbf{1}_{u\geq\tau} \leq g(u),\ u \in \mathbb{R}$$

and

$$0 \leq \mathbf{E}(g(\Delta_\infty) - f(\Delta_\infty)) \leq \delta.$$

Consider then the empirical sum

$$\frac{1}{n}\sum_{i\in n_+} f(x_i\sqrt{n\mathbf{E}(\Delta_\infty^2)}).$$

Denoting by K the Lipschitz constant of function f, it differs from the simpler one

$$\frac{1}{n}\sum_{i\in n_+} f(\beta^{-\ell}D_\ell(i)) \tag{43}$$

by at most

$$\begin{aligned}&\tfrac{K}{n}\textstyle\sum_{i\in n_+}\left|(x_i-y_i)\sqrt{n\mathbf{E}(\Delta_\infty^2)}\right| + \cdots\\ &\cdots\tfrac{K}{n}\textstyle\sum_{i\in n_+}\beta^{-\ell}|D_\ell(i)| \times \left|1-\sqrt{\frac{\mathbf{E}(\Delta_\infty^2)}{A}}\right|,\end{aligned}$$

where A is the empirical sum in (31). The first sum tends to zero in probability by the assumed convergence in probability $\lim_{n\to\infty}||x-y|| = 0$ and Cauchy-Schwarz inequality. The second sum tends to zero in probability by dominated convergence. Indeed, convergence to zero of $1-\sqrt{\mathbf{E}(\Delta_\infty^2)/A}$ has just been established, and with similar arguments one easily shows that the empirical average of the $|\beta^{-\ell}D_\ell(i)|$ converges in probability to $\mathbf{E}|\Delta_\infty|$ and is hence bounded in probability.

Convergence in probability of (43) to $(1/2)\mathbf{E}(f(\Delta_\infty))$ is similarly established.

The same argument with g instead of f yields convergence in probability

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i\in n_+} g(x_i\sqrt{n\mathbf{E}(\Delta_\infty^2)}) = \frac{1}{2}\mathbf{E}(g(\Delta_\infty)).$$

It readily follows that

$$\limsup_{n\to\infty}\left|\frac{1}{n}\sum_{i\in n_+}\mathbf{1}_{x_i\geq\tau/\sqrt{n\mathbf{E}(\Delta_\infty^2)}} - \frac{1}{2}\mathbf{P}(\Delta_\infty\geq\tau)\right| \leq \delta.$$

As δ is arbitrary, this establishes (34). □

G. PROOF OF LEMMA 4.4

PROOF. The first and second evaluations follow by noting that the vectors whose difference is considered in the left-hand side agree on the set of entries i whose ℓ-neighborhood is cycle-free. Thus

$$\begin{aligned}|B^{(\ell)}e - \{S_\ell(i)\}| &\le \sqrt{|\mathcal{B}|}O(\log(n)\alpha^\ell)\\ &\le O(\log^3(n)\alpha^{2\ell}),\end{aligned}$$

and the same bound holds for $|B^{(\ell)}\sigma - \{D_\ell(i)\}|$. This upper bound is $o(\sqrt{n}\beta^\ell)$ so that the first two assertions follow, by further noticing that $|\{D_\ell(i)\}| = \Theta(\sqrt{n}\beta^\ell)$, as follows from Theorem 4.2, (31).

For the third assertion, consider the scalar product $< \{S_\ell(i)\}, \{D_\ell(i)\} >$. The same arguments as in the proof of convergence (31) in Theorem 4.2 allow to establish (details omitted for brevity) the following convergences in probability:

$$\begin{aligned}\lim_{n\to\infty}\frac{1}{n}\sum_{i\in\mathcal{N}}\alpha^{-2\ell}S_\ell^2(i) &= \mathbf{E}M_\infty^2,\\ \lim_{n\to\infty}\frac{1}{n}\sum_{i\in\mathcal{N}}\alpha^{-\ell}S_\ell(i)\beta^{-\ell}D_\ell(i) &= \tfrac{1}{2}\mathbf{E}[M_\infty\Delta_\infty - M_\infty\Delta_\infty]\\ &= 0,\end{aligned} \tag{44}$$

where M_∞ is the almost sure limit of martingale M_t as in (30). Thus the scalar product $< \{S_\ell(i)\}, \{D_\ell(i)\} >$ is $o(n\alpha^\ell\beta^\ell)$ and is indeed negligible compared to $|\{S_\ell(i)\}| \times |\{D_\ell(i)\}|$, which is precisely of order $\Theta(n\alpha^\ell\beta^\ell)$. $\square$

H. PROOF OF LEMMA 4.5

PROOF. To establish the lower bound of (27), note that by Cauchy-Schwarz,

$$< e, B^{(\ell)}B^{(\ell)}e > \le |e| \times |B^{(\ell)}B^{(\ell)}e|.$$

However the left-hand side reads $|B^{(\ell)}e|^2$. Thus

$$|B^{(\ell)}B^{(\ell)}e| \ge \frac{|B^{(\ell)}e|^2}{|e|}.$$

However $|B^{(\ell)}e| = \Theta(\sqrt{n}\alpha^\ell)$ by (44) and Lemma 4.4. Since $|e| = \sqrt{n}$, the lower bound in (27) follows. For the upper bound, we note that by Lemma 4.3 and Theorem 2.3, the max row sum for matrix $B^{(\ell)}$ is of order $O(\log(n)\alpha^\ell)$.

The lower bound in (28) is established similarly, from the inequality

$$< \sigma, B^{(\ell)}B^{(\ell)}\sigma > \le |\sigma| \times |B^{(\ell)}B^{(\ell)}\sigma|$$

We turn to the upper bound. Write vector $B^{(\ell)}B^{(\ell)}\sigma$ as a sum $z + z' + z''$ where

$$\begin{aligned}z_i &= \mathbf{1}_{\bar{\mathcal{B}}}(i)\textstyle\sum_{j:d_{\mathcal{G}}(i,j)=\ell} D^{(\ell)}(j),\\ z_i' &= \mathbf{1}_{\bar{\mathcal{B}}}(i)\textstyle\sum_{j:d_{\mathcal{G}}(i,j)=\ell}\tilde{O}(\alpha^\ell)\mathbf{1}_{\mathcal{B}}(j),\\ z_i'' &= \mathbf{1}_{\mathcal{B}}(i)\tilde{O}(\alpha^{2\ell}).\end{aligned}$$

Adapting the results of Lemma 4.2 one has that both $|z'|$ and $|z''|$ are $\tilde{O}(\alpha^{3\ell})$. This is $O(n^{-\gamma}\beta^\ell|B^{(\ell)}\sigma|)$ by our choice of ℓ for some $\gamma > 0$.

Write then

$$\tilde{\mathbf{E}}|z|^2 = n^{1-\epsilon}\tilde{O}(\alpha^{2\ell}\beta^{2\ell}) + n\tilde{\mathbf{E}}(X^2\mathbf{1}_{\mathcal{C}}), \tag{45}$$

where $\mathcal{C}$ is the event that coupling between 2ℓ-neighborhood of i with random tree as per Lemma 4.6 has succeeded, $n^{-\epsilon}$ is the coupling failure probability and X is defined as

$$X = \sum_{d=0}^{\ell}\sum_{j:d(j,i)=2d}\sigma_j|\{k : d(j,k) = d(i,k) = \ell\}|.$$

Let $\mathcal{T}$ denote a branching process with offspring Poi(α). The process of spins is constructed by sampling uniformly the root's spin, and then propagating spins in a Markovian fashion with transition matrix $(a/(a+b)b(a+b), b(a+b), a(a+b))$ that is $\alpha^{-1}M$. Its eigenvalues are thus $(1, \beta/\alpha)$. Write

$$\begin{aligned}X^2 = &\textstyle\sum_{d=0}^{\ell}\sum_{d'=0}^{\ell}\sum_{j':d(j',i)=2d'}\sum_{j:d(j,i)=2d}\sigma_j\sigma_{j'} \times \cdots\\ &\cdots|\{k : d(j,k) = d(i,k) = \ell\}| \cdot |\{k' : d(j',k') = d(i,k') = \ell\}|.\end{aligned}$$

Now it holds that

$$\mathbf{E}(\sigma_j\sigma_{j'}|\mathcal{T}) = O\left(\left(\frac{\beta}{\alpha}\right)^{d(j,j')}\right).$$

We will use this formula, and further distinguish nodes j' according to their distance $2(d + d' - \tau)$ from j for $\tau = 0, \ldots, 2(d \wedge d')$. This yields

$$\begin{aligned}\mathbf{E}(X^2|\mathcal{T}) = &\sum_{d,d'=0}^{\ell}\sum_{\tau=0}^{2(d\wedge d')}\sum_{j':d(j',i)=2d'}\sum_{j:d(j,i)=2d}\cdots\\ &\cdots\mathbf{1}_{d(j,j')=2(d+d'-\tau)}O\left(\left(\tfrac{\beta}{\alpha}\right)^{2(d+d'-\tau)}\right) \times \cdots\\ &\cdots|\{k : d(j,k) = d(i,k) = \ell\}| \cdot |\{k' : d(j',k') = d(i,k') = \ell\}|.\end{aligned}$$

Note that on Ω the following evaluations hold uniformly for all nodes involved in the above expression:

$$\begin{aligned}&|\{k : d(j,k) = d(i,k) = \ell\}| = \tilde{O}(\alpha^{\ell-d}),\\ &|\{k' : d(j',k') = d(i,k') = \ell\}| = \tilde{O}(\alpha^{\ell-d'}),\\ &|\{j : d(j,i) = 2d\}| = \tilde{O}(\alpha^{2d}),\\ &|\{j' : d(j',i) = 2d' \& d(j,j') = 2(d+d'-\tau)\}| = \tilde{O}(\alpha^{2d'-\tau}).\end{aligned}$$

Plugging these in, we have

$$\begin{aligned}\mathbf{E}(X^2\mathbf{1}_\Omega|\mathcal{T}) &= \sum_{d,d'=0}^{\ell}\sum_{\tau=0}^{2(d\wedge d')}\tilde{O}\left(\left(\frac{\beta}{\alpha}\right)^{2(d+d'-\tau)}\right)\alpha^{2\ell+d+d'-\tau}\\ &= \textstyle\sum_{d,d'=0}^{\ell}\sum_{\tau=0}^{2(d\wedge d')}\tilde{O}\left(\alpha^{2\ell}\left(\frac{\beta^2}{\alpha}\right)^{d+d'-\tau}\right)\\ &= \tilde{O}(\alpha^{2\ell}(\beta^2/\alpha)^{2\ell})\\ &= \tilde{O}(\beta^{4\ell}).\end{aligned}$$

This combined with (45) and Tchebitchev's inequality entails that $|z|$ is with high probability no larger than

$$(\sqrt{n}\beta^\ell)\tilde{O}((\beta^\ell) \vee (\alpha^\ell n^{-\epsilon/2})).$$

Since $|B^{(\ell)}\sigma| = \Theta(\sqrt{n}\beta^\ell)$ the result follows. $\square$