The spectrum of the random environment and localization of noise
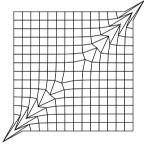
Dimitris Cheliotis · Bálint Virág

Abstract We consider random walk on a mildly random environment on finite transitive d-regular graphs of increasing girth. After scaling and centering, the analytic spectrum of the transition matrix converges in distribution to a Gaussian noise. An interesting phenomenon occurs at $d = 2$: as the limit graph changes from a regular tree to the integers, the noise becomes localized.

The graphs of the noise covariance structure for $d = 4, 3, 2.1$ from above.

Mathematics Subject Classification (2000)   15A52 · 05C50 · 60K37

D. Cheliotis (✉)
EURANDOM, Eindhoven University of Technology, L.G. 1.26, P.O. Box 513,
5600 MB, Eindhoven, The Netherlands
e-mail: dimitrisc@gmail.com

B. Virág
University of Toronto, 40 St George St., Toronto, ON M5S 2E4, Canada
e-mail: balint@math.toronto.edu

1 Introduction

Localization phenomena for eigenvalues of random media have received considerable attention lately. Among the several results, we point out a specific one. If one perturbs the Laplacian of the nearest-neighbor graph on the integers by a very small i.i.d. potential, its spectrum becomes pure point immediately. In contrast, the same change on a higher degree regular tree preserves the continuous spectrum, see [8].

The results of this paper point to an analogous but distinct phenomenon. While the above localization phenomena are due to large-scale behavior of the eigenvectors, we find a similar dichotomy for the local behavior.

For this, we consider sequences of vertex-transitive finite graphs G_n with degree $d \geq 2$ and increasing girth (the length of the shortest simple cycle); these converge locally to either the integers ($d = 2$) or a regular tree of higher degree ($d \geq 3$). On these graphs, we consider a perturbation of the Laplacian, given by the well-studied random walk in random environment model with small noise (see, for example, [14]). The noise in the Laplacian creates a Gaussian noise in the spectrum. Our main discovery is that in the limit this noise is local for the integers and has long-range correlations for higher-degree trees.

We first consider more general graphs G. Let M denote the transition probability matrix for nearest-neighbor simple random walk on G. For the random environment on G, one randomly modifies transition probabilities along edges. The new transition matrix is defined as $M + \varepsilon B$, where B is a random matrix.

We will have the following standing assumptions in effect for the rest of the paper.

Assumptions

(A) G is a vertex-transitive graph of finite degree, and the distribution of the random function $B_{u,v}$ is invariant under automorphisms of G.

(B)   When uv is an edge of G, the random variable $B_{u,v}$ has mean zero and variance 1.

(C)   The $B_{u,v}$'s corresponding to different u are independent.

(D)   For some constant c_1 and all vertices $u, v \in G$, we have

$$|B_{u,v}| \leq c_1 M_{u,v}, \tag{1}$$

$$\sum_{g \in \mathrm{Stab}(u)} B_{u,gv} = 0. \tag{2}$$

Here $\mathrm{Stab}(u)$ is the stabilizer of u in the automorphism group of G. Condition (2) means that the sum of the bias of the random environment over symmetric directions is zero, which is a bit stronger than just requiring $M + \varepsilon B$ to be stochastic. Condition (1) implies that $B_{u,v}$ vanishes when uv is not an edge and that for small ε the entries of $B + \varepsilon M$ are in $[0, 1]$.

For a finite G, let μ_ε denote the empirical probability measure of the eigenvalues $\{\lambda_1(\varepsilon), \ldots, \lambda_{|G|}(\varepsilon)\}$ of $M + \varepsilon B$, i.e., $\mu_\varepsilon := |G|^{-1} \sum_{i=1}^{|G|} \delta_{\lambda_i(\varepsilon)}$. Then as $\varepsilon \to 0$, we will show an expansion

$$\mu_\varepsilon = \mu_0 + \frac{1}{2}\mu'' \varepsilon^2 + o(\varepsilon^2),$$

where μ'' is a random functional. We define the second order difference quotient

$$m_\varepsilon := |G|^{1/2} \frac{\mu_\varepsilon - \mu_0}{\varepsilon^2},$$

where the extra factor upfront makes the scaling consistent as G changes. This is the centered and scaled empirical eigenvalue measure of $M + \varepsilon B$. Let $\mathcal{X}$ denote the space of complex functions analytic in a neighborhood of the closed unit disk. We call f real if it maps $\mathbb{R}$ to $\mathbb{R}$. In Proposition 4, we will show that for $f \in \mathcal{X}$ the limit

$$T_G(f) := \lim_{\varepsilon \to 0} \int f \, dm_\varepsilon \tag{3}$$

exists. Our first theorem identifies the covariance structure of $T_G(f)$ as f ranges over real functions in $\mathcal{X}$; this may depend on B.

For a possibly infinite vertex-transitive graph G and complex $|\lambda| < 1$, let

$$p_\lambda(x) := 1/(1 - \lambda x), \quad \mathcal{G}_\lambda := p_\lambda(M) = (I - \lambda M)^{-1},$$

the Green's function corresponding to G. Let o be a marked vertex of G, and define

$$H_G(p_\lambda, p_\mu) := \frac{\lambda \mu}{2} \, \partial^2_{\lambda,\mu} \left(\lambda^2 \mu^2 \sum_{v \neq o} \mathbf{E}\left[(B\mathcal{G}_\lambda)_{o,v} (B\mathcal{G}_\mu)_{o,v} \right] \mathbf{E}\left[(B\mathcal{G}_\lambda)_{v,o} (B\mathcal{G}_\mu)_{v,o} \right] \right). \tag{4}$$

We will show in Lemma 7 and Proposition 8 that $H_G(p_\lambda, p_\mu)$ is well-defined and extends uniquely to a bilinear form H_G on $\mathcal{X}$. H_G may depend on the covariance structure of the matrix B.

Theorem 1 *For a finite graph G and $f, g \in \mathcal{X}$, we have*

$$\mathbf{E}T_G(f) = 0, \quad \mathbf{E}[T_G(f)T_G(g)] = H_G(f, g).$$

In particular, when f, g are real, $H_G(f, g)$ is the covariance of $T_G(f), T_G(g)$. The proof of Theorem 1 is given in Sect. 4.

We now consider sequences of transitive graphs $G_n \to G$ locally, which means that for every r, the r-neighborhood of a fixed vertex o of G_n eventually agrees with that in G. We say that $(G_n, B_n) \to (G, B)$ locally with covariances if, in addition, the covariances of $(B_n)_{o,v}$ as v ranges over the neighbors of o in G_n also converge to the corresponding quantities in G and the constant c_1 is uniform over the sequence.

Theorem 2 (Normality) *Let G_n be a sequence of finite graphs, and let G be an infinite graph. Assume that $(G_n, B_n) \to (G, B)$ locally with covariances.*

Then as $n \to \infty$, jointly for all real $f \in \mathcal{X}$, the random variables $T_{G_n}(f)$ converge weakly to mean zero normal random variables $T(f)$. Moreover, for all $f, g \in \mathcal{X}$, we have

$$\mathbf{E}\,[T_{G_n}(f)T_{G_n}(g)] \to \mathbf{E}\,[T(f)T(g)] = H_G(f, g).$$

The proof of Theorem 2 is given in Sect. 5.

Finally, we consider the case when $G_n \to \mathbb{T}_d$, the d-regular tree. Since $\mathbb{T}_d$ is bipartite, we expect that the random functionals m_ε have a symmetric limit. Therefore it is natural to consider T_n for even functions of the form $f \circ \mathfrak{s}$ where $\mathfrak{s}(x) = x^2$ and $f \in \mathcal{X}$. Note that for any finite graph G, we have

$$T_G(f \circ \mathfrak{s}) = \tilde{T}_G(f),$$

where $\tilde{T}_G(f)$ is defined analogously to $T_G(f)$ but in terms of $\tilde{\mu}_\varepsilon$, the empirical eigenvalue measure of $(M + \varepsilon B)^2$.

It is known (see [6], Theorem 7.2) that $\tilde{\mu}_n$ corresponding to G_n converges weakly to the measure with density

$$a_d(x) := \frac{2d^2}{\pi} \frac{\sqrt{x(\rho - x)}}{1 - x} \mathbf{1}_{0 < x < \rho}, \tag{5}$$

where $\rho = \rho_d := 4(d - 1)/d^2$, the squared spectral radius of the walk on the infinite d-regular tree $\mathbb{T}_d$. It turns out that for regular trees, $H_{\mathbb{T}_d}$ does not depend on the specific structure of the noise. Our next theorem gives an explicit expression for the limiting covariance structure. Its proof is given in Sect. 7.

Theorem 3 (Tree limits) *Assume that $G_n \to \mathbb{T}_d$, the d-regular tree, and G_n, $\mathbb{T}_d$ satisfy the assumptions above with uniform constant c_1. Then jointly for all real $f \in \mathcal{X}$, we have the convergence in distribution $\tilde{T}_{G_n}(f) \to \tilde{T}(f)$, where the $\tilde{T}(f)$ are jointly normal and have mean 0. Moreover,*

$$\mathbf{E}[\tilde{T}(f)\tilde{T}(g)] = \int_0^\rho \int_0^\rho f'(x)g'(y)\, \beta_d(x, y)\, dx dy, \tag{6}$$

where for $d > 2$ the kernel β_d is given by

$$\beta_d(x, y) := \frac{2d^4}{\pi^2} \frac{(d - 2)\,\kappa(x)\kappa(y)}{16(2d - 3)(x - y)^2 + (d - 2)^2 A(d, x, y)}, \tag{7}$$

with the semicircle function

$$\kappa(x) := 2d\sqrt{x(\rho - x)} \quad \text{and}$$
$$A(d, x, y) := \rho\kappa \left(\frac{x + y}{2}\right)^2 + 4(d + 3)(x - y)^2 + \rho^2(d - 2)^2.$$

For d = 2, the covariance is given by

$$\frac{32}{\pi} \int\limits_0^1 f'(x)g'(x)\sqrt{x(1-x)}dx,$$

which corresponds to $\beta_2(x, y) := \frac{32}{\pi}\sqrt{x(1-x)}\,\delta_x(y).$

The kernel function β_d becomes singular as $d \downarrow 2$. This can be seen in the figure on the first page showing the graphs of $\beta_d : [0, \rho]^2 \to \mathbb{R}$ for $d = 4, 3, 2.1$ from above. In words, in the $d = 2$ case, the centered and scaled empirical measure converges to the so called H_1-noise with a certain density. In contrast with the $H_{1/2}$-noise that arises in the limit of the Gaussian Orthogonal Ensemble and Haar unitary random matrix models (among others; see, for example, [1,4] and the references therein), this noise is local, as the δ-function in the covariance formula shows. H_1 noise appears typically for complex eigenvalues, see [9]. Formula (6) is written in a form to show how the kernel becomes singular (localized) as $d \downarrow 2$. Recall that $\pi^{-1}\varepsilon/(x^2 + \varepsilon^2)$ converges as $\varepsilon \to 0$ to the delta function at $x = 0$.

The presence of f', g' in the formula for the covariance is a priori expected. When the matrix is randomly perturbed, the eigenvalues move in random directions. For every eigenvalue λ, the function $f(\lambda)$ changes by a small amount proportional to f' times the change in λ. However, the picture here is more complicated because the perturbation is of order ε^2. The kernel β reflects the correlation between the random change in eigenvalues as well as their density.

Theorem 3 motivates the following question:

Question 1 *Which natural sequences of graphs give rise to localized noise?*

Our strategy is as follows. To prove normality, we use a central limit theorem for dependent variables based on Stein's method, rather than the usual method of moment computations. For computing the covariance, we consider traces, and so we will have to count certain paths. We use the Green function to treat paths of different lengths jointly and to avoid complicated computations with orthogonal polynomials that normally arise in this context.

The finite graphs that we study arise naturally. The simplest way to pick an interesting transitive graph is to consider a Cayley graph of a sufficiently complicated finite group. As it is discussed in [5], such Cayley graphs will typically have girth tending to infinity; for example, random Cayley graphs of simple groups of increasing order will have this property. Of course, for $d = 2$, the only examples are graphs consisting of cycles.

Random perturbations of eigenvalues have been extensively studied in the literature. The book [12] studies properties of the eigenvalues of a random perturbation of a fixed matrix without the restriction for stochasticity or positivity of the matrices involved (see Sects. 1.3 and 2.2). Another possible approach to our problem is via perturbation expansions for eigenvalues (see, for example, [3], Sects. 6.3, 6.4), but this requires more control on the eigenvectors.

2 The $\varepsilon \to 0$ limit

We first show that for finite graphs G, the limit (3) defining $T_G(f)$ exists and identify it. We will rely on the Assumptions about the perturbation matrix B, in particular, (2). In fact, throughout this paper, we only need a simple consequence of (2). Namely, for all $k \geq 1$ and $v \in G$, we have

$$(BM^k)_{v,v} = 0. \tag{8}$$

Indeed, for any $g \in \text{Stab}(v)$ this equals

$$\sum_{w \in G} B_{v,w} M^k_{w,v} = \sum_{w \in G} B_{gv,gw} M^k_{gw,gv} = \sum_{w \in G} B_{v,gw} M^k_{w,v}.$$

Averaging each term over all $g \in \text{Stab}(v)$ and using (2) we get (8).

Proposition 4 (The limit as $\varepsilon \to 0$) *For any $f = \sum a_j z^j \in \mathcal{X}$, the limit $T_G(f)$ exists, and*

$$T_G(f) = \frac{1}{2} \text{Tr} \left(\partial_{\varepsilon\varepsilon} f(M + \varepsilon B)|_{\varepsilon=0} \right) = \sum_{j=0}^{\infty} a_j T_G(z^j). \tag{9}$$

Moreover, for integers $j \geq 0$ we have

$$T_G(z^j) = |G|^{-1/2} \frac{j}{2} \sum_{k_1+k_2+2=j} \text{Tr}(BM^{k_1} BM^{k_2}).$$

Proof We use the fact (see, for example, Theorem 6.2.8 in [7]) that for f an analytic function with radius of convergence $r > 0$, and for A in the set $\mathcal{M}_r$ of matrices of some fixed dimension and with spectral radius less than r, the power series $f(A)$ is absolutely convergent and is analytic as a function of (the entries of) A. Thus $\text{Tr} \, f(M + \varepsilon B)$ is an analytic function of all of its (at most) $2|G|^2 + 1$ variables near M, B, and $\varepsilon = 0$. We are free to rearrange its absolutely convergent multiple power series expansion in any way we like. So the first derivative with respect to ε at 0 is the coefficient of the ε terms, i.e.,

$$\sum_{j=1}^{\infty} a_j \sum_{k_1+k_2=j-1} \text{Tr}(M^{k_1} BM^{k_2}) = \sum_{j=1}^{\infty} a_j \sum_{k_1+k_2=j-1} \text{Tr}(BM^{k_1+k_2}),$$

which is zero because each of the summands is zero by (8). Consequently,

$$|G|^{1/2} \int f \, dm_\varepsilon = \varepsilon^{-2} (\text{Tr} \, f(M + \varepsilon B) - \text{Tr} \, f(M)) \to \frac{1}{2} \partial_{\varepsilon\varepsilon} \text{Tr} \, f(M + \varepsilon B)|_{\varepsilon=0}$$

when $\varepsilon \to 0$ since the first derivative vanishes at $\varepsilon = 0$. Now the right hand side equals the coefficient of ε^2 in the expansion. This gives (9).

Since the trace of a product does not change when we cyclically permute the factors, we see that

$$\partial_\varepsilon \operatorname{Tr}((M + \varepsilon B)^j) = \operatorname{Tr}(\partial_\varepsilon (M + \varepsilon B)^j) = j \operatorname{Tr}(B(M + \varepsilon B)^{j-1}).$$

Taking another derivative, we get the second claim of the proposition. □

3 Properties of the bilinear form H_G

The goal of this section is to define the bilinear form H_G introduced in (4) and establish its continuity properties.

Recall that $\mathcal{X}$ denotes the set of power series centered at zero, and with radius of convergence more than 1. For a power series f, let $[x^k]f(x)$ denote the coefficient of x^k in the expansion of f. Given a finite or infinite graph G (together with B and M) as in the introduction, we fix a vertex o of G, and re-define the bilinear form on $\mathcal{X}$ as

$$\begin{aligned}
H_G(f, g) &:= \frac{1}{2} \sum_{i,j \geq 1} i j \alpha_{i,j} [x^i]f(x)[x^j]g(x) \\
&= \frac{1}{2} \sum_{i,j \geq 1} \alpha_{i,j} [x^{i-1}]f'(x)[x^{j-1}]g'(x),
\end{aligned} \tag{10}$$

where

$$\alpha_{i,j} := \mathbf{E}\left[\sum_{v \in G \setminus \{o\}} Y_i(G, o, v) Y_j(G, o, v) \right],$$

and for any vertices v, w of G, we define

$$Y_j(G, v, w) := \sum_{j_1 + j_2 + 2 = j} (B M^{j_1})_{v,w} (B M^{j_2})_{w,v}. \tag{11}$$

The last sum is always finite and symmetric in v, w. In Proposition 8, we will show that this definition agrees with (4) for the functions p_λ. Note that the definition of H_G does not depend on o if Assumption A holds. For any graph G, the bilinear form $H_G(f, g)$ clearly makes sense for polynomials f, g. To go beyond polynomials, we need to show that the infinite sum in (10) is well-defined. For $f = \sum a_k z^k \in \mathcal{X}$, set

$$\|f\|_* := \sum_{k=1}^{\infty} |a_k| k^2,$$

which defines a norm on $\mathcal{X}$. (To be precise, it is a norm on $\mathcal{X}$ modulo the constants; however, we note that for the purposes of this paper the constant terms for functions

in $\mathcal{X}$ do not matter, so we may think of $\mathcal{X}$ itself as the space of functions modulo the constants.) Note that on $\mathcal{X}$ this norm is finite.

Remark 5 On the subspace of $\mathcal{X}$ consisting of power series with radius of convergence strictly larger than a fixed $r > 1$, the norm $\| \cdot \|_*$ is dominated by the supremum norm on $D_r := \{z \in \mathbb{C} : |z| \leq r\}$. Indeed, for $f(z) = \sum_{k=0}^{\infty} a_k z^k$ in that subspace, Cauchy's inequalities give $|a_k| \leq r^{-k} \|f\|_{D_r}^{\infty}$. So that

$$\|f\|_* \leq \frac{r(1+r)}{(r-1)^3} \|f\|_{D_r}^{\infty}.$$

As a direct consequence, we get the following.

Lemma 6 *Polynomials are $\| \cdot \|_*$—dense in $\mathcal{X}$.*

We are now ready to show that H_G is well-defined and continuous.

Lemma 7 *For $f, g \in \mathcal{X}$ and any G satisfying the assumptions A–D in the introduction, the sum giving $H_G(f, g)$ is absolutely convergent, and*

$$|H_G(f, g)| \leq c_1^4 \|f\|_* \|g\|_*.$$

Proof We will show that $|a_{i,j}| \leq c_1^4 ij$ for all $i, j \geq 1$. Note that since $|B_{v,w}| \leq c_1 M_{v,w}$ (Assumption D), we have

$$\sum_{w \in G} |Y_j(G, v, w)| \leq c_1^2 \sum_{w \in G} \sum_{j_1+j_2+2=j} M_{v,w}^{j_1+1} M_{w,v}^{j_2+1} = c_1^2 (j-1)(M^j)_{v,v}$$
$$\leq c_1^2 j (M^j)_{v,v}. \tag{12}$$

Therefore

$$\frac{1}{c_1^4 ij} \sum_{v \in G \setminus \{o\}} |Y_i(G, o, v) Y_j(G, o, v)| \leq \frac{1}{c_1^4 ij} \sum_{v \in G} |Y_i(G, o, v)| \sum_{w \in G} |Y_j(G, o, w)|$$
$$\leq M_{o,o}^i M_{o,o}^j \leq 1.$$

The claim now follows by summing over all i, j. □

Proposition 8 *H_G satisfies (4). Moreover, (4) uniquely defines H_G as a $\| \cdot \|_*$ -continuous bilinear form.*

Proof The absolute convergence of the series defining H_G implies

$$H_G(p_\lambda, p_\mu) = \frac{1}{2} \sum_{i,j \geq 1} ij\, \alpha_{i,j} \lambda^i \mu^j = \frac{1}{2} \lambda \mu\, \partial_\lambda \partial_\mu \sum_{i,j \geq 1} \alpha_{i,j} \lambda^i \mu^j.$$

Using the definition of $\alpha_{i,j}$, we write the sum above as

$$\sum_{\substack{v \in G \setminus \{o\} \\ k,\ell,k',\ell' \geq 0}} \lambda^{k+\ell+2} \mu^{k'+\ell'+2} \mathbf{E}\{(BM^k)_{o,v}(BM^\ell)_{v,o}(BM^{k'})_{o,v}(BM^{\ell'})_{v,o}\}$$

$$= \lambda^2 \mu^2 \sum_{v \in G \setminus \{o\}} \mathbf{E}\{(B\mathcal{G}_\lambda)_{o,v}(B\mathcal{G}_\lambda)_{v,o}(B\mathcal{G}_\mu)_{o,v}(B\mathcal{G}_\mu)_{v,o}\}$$

$$= \lambda^2 \mu^2 \sum_{v \in G \setminus \{o\}} \mathbf{E}\left\{(B\mathcal{G}_\lambda)_{o,v}(B\mathcal{G}_\mu)_{o,v}\right\} \mathbf{E}\left\{(B\mathcal{G}_\lambda)_{v,o}(B\mathcal{G}_\mu)_{v,o}\right\}.$$

This proves the first part. For the second, it suffices to show that the linear span of $\{p_\lambda : |\lambda| < 1\}$ is $\|\cdot\|_*$-dense in $\mathcal{X}$; this will be done in the next lemma. $\qquad\square$

Lemma 9 *The vector space generated by* $\{p_\lambda : |\lambda| < 1\}$ *is* $\|\cdot\|_*$—*dense in* $\mathcal{X}$.

Proof Because of the density of polynomials in $\mathcal{X}$ and Remark 5, it suffices to prove that any polynomial P can be approximated in the supremum norm on a disk D_r with $r > 1$ by elements of the linear span of $\{p_\lambda : |\lambda| < 1\}$.
Pick $1 < r < r_1$, and let $C_{r_1} := \{z \in \mathbb{C} : |z| = r_1\}$. Then

$$P(z) = \frac{1}{2\pi i} \int\limits_{C_{r_1}} \frac{P(\zeta)}{\zeta - z} d\zeta = \frac{1}{2\pi i} \int\limits_{C_{r_1}} \frac{P(\zeta)}{\zeta} p_{1/\zeta}(z) d\zeta.$$

Call $R_n(z)$ the Riemann sum corresponding to an equipartition of the circle with n pieces. The sequence of functions R_n is equicontinuous on $\{z : |z| \leq r\}$ and it converges pointwise to $P(z)$, thus the convergence is uniform on that set. So that for a given $\delta > 0$, there is a finite linear combination

$$A_\delta(z) := c_1 p_{1/\zeta_1}(z) + c_2 p_{1/\zeta_2}(z) + \cdots + c_k p_{1/\zeta_k}(z)$$

such that

$$|f(z) - A_\delta(z)| \leq \delta \tag{13}$$

for all $z \in \mathbb{C}$ with $|z| \leq r$. This completes the proof of the lemma. $\qquad\square$

Finally, we check that $H_G(f, g)$ is continuous in G as well. Recall the definition of local convergence of graphs from the introduction. For the following lemma, we use, as usual, the assumptions from the introduction for each G_n and G, and we also assume that the constant c_1 of Assumption D is the same for all G_n.

Lemma 10 *If* $(G_n, B_n) \to (G, B)$ *locally with covariances, then for* $f, g \in \mathcal{X}$ *we have* $H_{G_n}(f, g) \to H_G(f, g)$.

Proof Note that as $n \to \infty$, the neighborhood of radius $i + j$ of o in G_n stabilizes to look like the same neighborhood in the limit graph G.

For $\varepsilon > 0$, polynomials f, g, and for all large n, we have that $|H_{G_n}(f, g) - H_G(f, g)| < \varepsilon$. This is because $H_{G_n}(f, g)$ only depends on a neighborhood of the root o of radius given by the maximal degree of f and g, and this neighborhood eventually agrees with that of G; moreover H only depends on the covariances of the $B_{u,v}$, which also converge by assumption.

Now we have that the sequence of functions $H_{G_n}(f, g)$ is equicontinuous on $\mathcal{X}^2$ (by Lemma 7) and converges on a dense set. Thus, by Lemma (15), they converge on the entire set to a continuous limit. $H_G(\cdot, \cdot)$ is continuous (Lemma 7), and this finishes the proof.   □

4 The covariance structure

After establishing some properties of the bilinear form H_G, we show that for finite graphs G it gives the covariance structure of $T_G(f)$.

Lemma 11 *For any finite vertex-transitive graph G and complex polynomials f, g, we have*

$$\mathbf{E}(T_G(f)) = 0,$$
$$\mathbf{E}(T_G(f)T_G(g)) = H_G(f, g).$$

Proof We will show this for monomials. The extension to polynomials is straightforward from bilinearity. First, by Proposition 4, we have

$$T_G(z^j) = \frac{1}{|G|^{1/2}} \frac{j}{2} \operatorname{Tr} \sum_{j_1 + j_2 + 2 = j} BM^{j_1} BM^{j_2}$$
$$= \frac{1}{|G|^{1/2}} \frac{j}{2} \sum_{\substack{v, w \in G \\ v \neq w}} Y_j(G, v, w). \tag{14}$$

Note that the $v = w$ terms vanish by (8). Each $Y_j(G, v, w)$ with $v \neq w$ has zero mean because different rows of B are independent with entries having zero mean (Assumptions B, C). Thus, $T(z^j)$ has also zero mean. Again because of independence of rows of B, when we compute second moments, the terms in the sum below with $\{v, w\} \neq \{v', w'\}$ vanish. That is,

$$\mathbf{E}(T_G(z^i)T_G(z^j)) = \frac{ij}{4|G|} \sum_{\substack{v, w, v', w' \in G \\ v \neq w, v' \neq w'}} \mathbf{E}\{Y_i(G, v, w)Y_j(G, v', w')\}$$

$$= \frac{ij}{4|G|} \sum_{v,w \in G : v \neq w} \mathbf{E}\left\{ Y_i(G, v, w) \left[Y_j(G, v, w) + Y_j(G, w, v) \right] \right\}$$

$$= \frac{ij}{2} \sum_{v \in G \setminus \{o\}} \mathbf{E}\{ Y_i(G, o, v) Y_j(G, o, v) \}.$$

For the last equality, we fixed a vertex o of G and used Assumption A and the symmetry of Y in its last two arguments. □

Lemma 12 *For G finite, $|T_G(f)| \leq c_1^2 |G|^{1/2} \|f\|_*$.*

Proof This follows directly from (9), (14), and (12). □

We are now ready to prove Theorem 1.

Proof of Theorem 1 Lemmas 7 and 12 show that $\mathbf{E}[T_G(f)T_G(g)]$ and $H_G(f, g)$ are $\| \cdot \|_*$-continuous in f, g. We conclude the proof by approximating f and g with polynomials and using Lemmas 6 and 11. □

5 Asymptotic normality

For this section, we consider a sequence of graphs G_n together with an associated sequence of noise matrices B_n so that $(G_n, B_n) \to (G, B)$ locally with covariances. The assumptions from Sect. 1 hold for all G_n and the limit G with the same constant c_1. We may also assume that all graphs have degree d.

We will prove convergence of certain sequences to normal under a topology which we introduce now.

Let $\mathcal{Y}$ denote the space of probability measures on $\mathbb{R}$ with finite second moment, equipped with the 2-Wasserstein distance, call it d_2 (see [11], Chapter 6). For two measures μ, ν in the space, this is the minimal L^2-distance over all possible couplings of them, i.e.,

$$d_2(\mu, \nu) := \inf_k \left(\int |x - y|^2 \, dk(x, y) \right)^{1/2},$$

where the infimum is over the set of probability measures on $\mathbb{R}^2$ with first marginal μ, and second ν.

The space $(\mathcal{Y}, d_2)$ is complete, and its topology is stronger than weak convergence. It is easy to show that for a sequence X_n of random variables, $X_n \to X$ in this topology (i.e., the corresponding laws converge) if and only if $X_n \to X$ weakly and $\mathbf{E}X_n^2 \to \mathbf{E}X^2$. In particular, in this topology the function variance, Var $: \mathcal{Y} \to \mathbb{R}$, is continuous.

We will need the following case of Lemma 2.4 of [2]. It is a normal approximation theorem for dependent variables. It is proved with the use of Stein's method.

Lemma 13 [2] *Let $\mathcal{I}$ be an index set, $\{X_i : i \in \mathcal{I}\}$ a family of random variables, and for $A \subset \mathcal{I}$, let $X_A := \{X_i : i \in A\}$. Assume that*

(1) *For each $i \in \mathcal{I}$, there exist $A_i \subset B_i \subset C_i \subset \mathcal{I}$ such that X_i is independent of $X_{\mathcal{I} \setminus A_i}$, X_{A_i} is independent of $X_{\mathcal{I} \setminus B_i}$, and X_{B_i} is independent of $X_{\mathcal{I} \setminus C_i}$.*

(2) *There exists a constant γ so that for all $i \in \mathcal{I}$ we have*

$$\max(|N(C_i)|, |\{j : i \in C_j\}|) \leq \gamma,$$

where

$$N(C_i) := \{j \in \mathcal{I} : C_i \cap B_j \neq \emptyset\}.$$

(3) *Each X_i has zero mean and finite variance, and $W := \sum_{i \in \mathcal{I}} X_i$ satisfies $\mathrm{Var}(W) = 1$.*

Then for $2 < p \leq 3$, we have

$$\sup_{z \in \mathbb{R}} |F(z) - \Phi(z)| \leq 75 \gamma^{p-1} \sum_{i \in \mathcal{I}} \mathbf{E}|X_i|^p,$$

where F, Φ are the distribution functions of W and of the standard normal $N(0, 1)$.

An immediate consequence is the following convergence result.

Lemma 14 *For any real polynomial f, the sequence $T_{G_n}(f)$ converges, as $n \to \infty$, in the 2-Wasserstein distance to a normal random variable with zero mean and variance $H_G(f, f)$.*

Proof We will apply Lemma 13. Let $f(x) = \sum_{j=0}^{k} a_j z^j$, and $\mathcal{I}$ the set of vertices of G_n. In the following, we will omit the subscripts for the matrices M, B associated with the graph G_n. By Proposition 4, we have

$$T_{G_n}(f) = \frac{1}{|G_n|^{1/2}} \mathrm{Tr} \left(\sum_{j=0}^{k} \frac{j}{2} a_j \sum_{k_1 + k_2 + 2 = j} B M^{k_1} B M^{k_2} \right) = \sum_{v \in G_n} Y_{n,v},$$

where

$$Y_{n,v} := \frac{1}{|G_n|^{1/2}} \sum_{j=0}^{k} \frac{j}{2} a_j \sum_{k_1 + k_2 + 2 = j} (B M^{k_1} B M^{k_2})_{v,v}.$$

It holds $\lim_{n \to +\infty} \mathrm{Var}\, T_{G_n}(f) = H_G(f, f)$ by Lemma 10. If this limit is zero, then the sequence $T_{G_n}(f)$ will converge to δ_0 in the 2-Wasserstein topology, and the result is proved. We may therefore assume that the limit is positive, and $\mathrm{Var}\, T_{G_n}(f) > 0$ for all n.

For $v \in \mathcal{I}$, define

$$X_v := \frac{Y_{n,v}}{\mathrm{Var}\, T_{G_n}(f)}, \quad W := \sum_{v \in G_n} X_v = \frac{T_{G_n}(f)}{\mathrm{Var}\, T_{G_n}(f)},$$

and the sets

$$A_v := \{w : \text{dist}(v, w) \leq k\},$$
$$B_v := \{w : \text{dist}(v, w) \leq 2k\},$$
$$C_v := \{w : \text{dist}(v, w) \leq 3k\}.$$

These sets satisfy the conditions of Lemma 13 (because the X_v's corresponding to vertices that are distance at least $k + 1$ apart are independent random variables), and for all v, we have $|N(C_v)| \leq |\{w : \text{dist}(v, w) \leq 5k\}| \leq d(d - 1)^{5k-1}$, and $|\{w : v \in C_w\}| = |C_v| \leq d(d - 1)^{3k-1}$. [Here we used the regularity of the graphs, but of course a uniform bound on the degree of the vertices of all graphs would work in the same way.] Also, the X_v's have zero mean and finite variance by Lemma 11. Pick any $p \in (2, 3]$. Lemma 13 gives that the distribution functions F_W, F_Z of W and of the standard Normal $N(0, 1)$ satisfy

$$\sup_{x \in \mathbb{R}} |F_W(x) - F_Z(x)| \leq C|G_n|^{1-\frac{p}{2}} \mathbf{E}|A|^p, \tag{15}$$

with $C := 75(d(d - 1)^{5k-1})^{p-1}$ and

$$A := \sum_{j=0}^{k} \frac{j}{2} a_j \sum_{k_1+k_2+2=j} (BM^{k_1} BM^{k_2})_{o,o}.$$

This is bounded because of the uniform constant c_1 in assumption (1) and the fact that the j-neighborhood of o in G_n is of uniformly bounded size. Thus relation (15), $|G_n| \to \infty$, and $p > 2$ imply that the sequence

$$\frac{T_{G_n}(f)}{\text{Var } T_{G_n}(f)}$$

converges to a standard normal random variable. By Theorem 1 and Lemma 10, we have Var $T_{G_n}(f) \to H_G(f, f)$, and the result follows.   □

Our next goal is to strengthen Lemma 14 by showing that its conclusion is true also for all functions f in $\mathcal{X}$, and thus get Theorem 2. The proof is simply by approximation. We will use the fact that the set of polynomials is $\| \cdot \|_*-$ dense in $\mathcal{X}$ (Lemma 6) and the following simple lemma (Lemma 38, Chapter 7 of [10]).

Lemma 15 *Let X be a metric space, and Y a complete metric space. Assume that $f_n : X \to Y$ is an equicontinuous sequence of functions that converge pointwise on a dense subset of X. Then the sequence f_n converges pointwise on the entire X to a continuous limit.*

We are now ready to prove Theorem 2.

Proof of Theorem 2 Let $\hat{T}_n(f)$ denote the distribution of $T_{G_n}(f)$. The sequence of functions $\hat{T}_n : \mathcal{X} \to \mathcal{Y}$ is uniformly equicontinuous because, by Lemma 7 and Theorem 1, we have

$$\mathbf{E}|T_{G_n}(f) - T_{G_n}(g)|^2 \le c_1^4 \|f - g\|_*^2.$$

By Lemma 14, they converge pointwise at polynomials, which form a dense subset of $\mathcal{X}$ by Lemma 6. By Lemma 15, the limit $\hat{T}(f)$ of $\hat{T}_n(f)$ exists for all functions $f \in \mathcal{X}$ and is continuous. Also, for the 2-Wasserstein distance between $\hat{T}(f)$ and $\hat{T}(g)$, we have

$$d_2(\hat{T}(f), \quad \hat{T}(g)) \le c_1^2 \|f - g\|_*. \tag{16}$$

Since the limit is normal on a dense set of points, and limits of normal random variables are normal, it follows that all limits $\hat{T}(f)$ are normal. Also, the functionals

$$f \mapsto \operatorname{Var} \hat{T}(f), \quad f \mapsto H_G(f, f)$$

are $\| \cdot \|_*$-continuous [the first because of (16) and a property of the 2-Wasserstein distance, the second by Lemma 7] and they agree on a $\| \cdot \|_*$-dense set by Lemma 14. Thus, $\operatorname{Var} \hat{T}(f) = H_G(f, f)$ for all $f \in \mathcal{X}$. □

6 A formula for the covariance

The goal of this section is to reduce the problem of computing the covariance kernel β of Theorem 3 to inverting a Stieltjes transform.

When the limit graph G is bipartite (see the discussion before Theorem 3), we would like to write the limiting covariance in the form

$$H_G(f \circ \mathfrak{s}, g \circ \mathfrak{s}) = \int f'(x)g'(y)d\beta(x, y), \tag{17}$$

where $\mathfrak{s}(x) = x^2$. It is sufficient to check the identity (17) for the functions $f = p_{\lambda^2}$ and $g = p_{\mu^2}$, since their linear span is a dense subset of $\mathcal{X}$ with respect to the norm $\| \cdot \|_*$ (Lemma 9), and the integral on the right is clearly continuous with respect to the product topology based on this norm. For this choice of f, g, the right hand side of (17) equals

$$\int \frac{\lambda^2}{(1 - \lambda^2 x)^2} \frac{\mu^2}{(1 - \mu^2 y)^2} \, d\beta(x, y) = \frac{1}{4} \lambda \mu \, \partial_{\lambda,\mu}^2 \int \frac{1}{(x - \lambda^{-2})} \frac{1}{(y - \mu^{-2})} \, d\beta(x, y).$$

Note that $p_\lambda - p_{\lambda^2} \circ \mathfrak{s}$ is an odd function. For bipartite graphs G, it is clear from the expressions (10) and (11) that H_G vanishes if one its arguments is odd. So we have

$$H_G(p_{\lambda^2} \circ \mathfrak{s}, p_{\mu^2} \circ \mathfrak{s}) = H_G(p_\lambda, p_\mu).$$

In light of expression (4), we arrive at the following:

Proposition 16 *If the Stieltjes transform relation*

$$\lambda^2\mu^2 \sum_{v \neq o} 2\mathbf{E}\Big[(B\mathcal{G}_\lambda)_{o,v}(B\mathcal{G}_\mu)_{o,v}\Big]\mathbf{E}\Big[(B\mathcal{G}_\lambda)_{v,o}(B\mathcal{G}_\mu)_{v,o}\Big]$$

$$= \int \frac{1}{(x - \lambda^{-2})} \frac{1}{(y - \mu^{-2})} \, d\beta(x, y) \tag{18}$$

holds for all $|\lambda|, |\mu| < 1$, *then* (17) *holds for all* $f, g \in \mathcal{X}$.

7 Explicit formulas for trees

In this section, we look at the case where the limiting graph G is the infinite d-regular tree $\mathbb{T}_d$, we compute the measure $d\beta$ introduced in (17), and thus complete the proof of Theorem 3.

From now on, we define square roots for complex numbers with a branch cut at the negative real axis. More precisely, for $z = |z|e^{i\theta}$, with $\theta \in (-\pi, \pi]$, we set $\sqrt{z} := \sqrt{|z|}e^{i\theta/2}$.

Proof of the covariance formula in Theorem 3 We start by computing the left hand side of (18) explicitly. The Green's function for the infinite d-regular tree, $\mathcal{G}_\lambda(v, w) = \mathcal{G}_\lambda(\text{dist}(v, w))$, is given by

$$\mathcal{G}_\lambda(r) = b_\lambda a_\lambda^r,$$

where

$$a_\lambda := d \frac{1 - \sqrt{1 - \rho\lambda^2}}{2(d-1)\lambda}, \quad b_\lambda := (1 - \lambda a_\lambda)^{-1},$$

and $\rho := 4(d-1)/d^2$ as in the introduction. See [13], Lemma 1.24.

First, an observation. For two vertices i, j, let ℓ denote the unique neighbor of i closest to j, r the distance between i and j, and define $B_{i,j}^* = B_{i,\ell}$. For λ with modulus less than 1, by symmetry and since $\sum_k B_{i,k} = 0$ (Assumption D), we get

$$(B\mathcal{G}_\lambda)_{i,j} = B_{i,\ell}\mathcal{G}_\lambda(r-1) + \sum_{\substack{k \sim i \\ k \neq \ell}} B_{i,k}\mathcal{G}_\lambda(r+1) = B_{i,j}^*\mathcal{G}_\lambda^*(r),$$

where

$$\mathcal{G}_\lambda^*(r) := \mathcal{G}_\lambda(r-1) - \mathcal{G}_\lambda(r+1) = b_\lambda(1 - a_\lambda^2)a_\lambda^{r-1}.$$

Thus for any vertex $w \neq o$, with $r := \text{dist}(o, w)$, we have

$$\mathbf{E}(B\mathcal{G}_\lambda)_{o,w}(B\mathcal{G}_\mu)_{o,w} = \mathbf{E}B_{o,w}^{*2}\mathcal{G}_\lambda^*(r)\mathcal{G}_\mu^*(r) = \mathcal{G}_\lambda^*(r)\mathcal{G}_\mu^*(r), \tag{19}$$

because $\mathbf{E}B_{o,w}^{*\,2} = 1$ (Assumption B), and therefore the sum in the left-hand-side of (18) equals

$$\sum_{w \in \mathbb{T}^d \setminus \{o\}} 2\mathcal{G}_\lambda^*(r)^2 \mathcal{G}_\mu^*(r)^2 = 2d \sum_{r=1}^{\infty} (d-1)^{r-1} \, \mathcal{G}_\lambda^*(r)^2 \mathcal{G}_\mu^*(r)^2$$

$$= 2\frac{(b_\lambda b_\mu)^2 (1 - a_\lambda^2)^2 (1 - a_\mu^2)^2 d}{1 - (d-1)(a_\lambda a_\mu)^2}. \tag{20}$$

Now this can be expressed in terms of $s = \sqrt{1 - \rho\lambda^2}$ and $t = \sqrt{1 - \rho\mu^2}$. Indeed, in these variables, we have the simpler expressions

$$a_\lambda^2 = \frac{1-s}{(d-1)(1+s)}, \qquad b_\lambda = \frac{2(d-1)}{(d-2) + sd},$$

and (20) becomes

$$\frac{32\,(d-1)d}{(1+s)(1+t)\,((d-2)(1+s)(1+t) + 2(s+t))}. \tag{21}$$

Introduce new variables u, v by $u = \lambda^{-2}$, $v = \mu^{-2}$, and let $\hat{\beta}(u, v)$ denote expression (21) as a function of them. In terms of these variables, relation (18) becomes

$$\frac{\hat{\beta}(u, v)}{uv} = \int \frac{1}{(x-u)} \frac{1}{(y-v)} \, d\beta(x, y).$$

Due to our convention for square roots (see beginning of this section), the quantity s is an analytic function of u in $\mathbb{C} \setminus [0, \rho]$, and the same holds for t as a function of v.

Assume that $d > 2$. Then the denominator in (21) does not vanish because s, t have nonnegative real parts (if we set the last factor in the denominator equal to zero and solve for t, we get a quantity with negative real part). So $\hat{\beta}$ defined by (21) is a holomorphic function of (u, v) on $(\mathbb{C} \setminus [0, \rho])^2$. In fact, even the limits of the denominator when u or v approaches $[0, \rho)$ are not zero.

Since the function $h(u, v) := (uv)^{-1} \hat{\beta}(u, v)$ is holomorphic in $(\mathbb{C} \setminus [0, \rho])^2$ and it decays as $(uv)^{-1}$ near infinity, Cauchy's formula allows us to express its values in terms of double contour integrals around the segment $[0, \rho]$. Shrinking the contour around $[0, \rho]$, we get a line integral, and we take into account the different limits of h as one of its arguments approaches the segment from the upper or the lower half plane. That is, when u approaches $x \in [0, \rho)$ from the upper half plane, we have $s(u) \to s(x)$. While when u approaches x from the lower half plane, we have $s(u) \to -s(x)$. The difference comes from the branch cut discontinuity of the square root in the definition of s. Thus we have

$$\frac{\hat{\beta}(u, v)}{uv} = -\frac{1}{4\pi^2} \int_0^\rho \int_0^\rho \frac{1}{(x-u)(y-v)} \frac{\tilde{\beta}(x, y)\,dx\,dy}{xy},$$

where for $x, y \in [0, \rho]$ we have

$$\tilde{\beta}(x, y) := \sum_{\sigma, \tau = \pm 1} \sigma \tau \hat{\beta}[\sigma s, \tau t]$$

$$= \frac{512(d-1)^2 d^2 (d-2)\, st}{(d^2(s^2+t^2) - (d-2)^2(1+s^2t^2))^2 - (8(d-1)st)^2},$$

with $\hat{\beta}[\sigma s, \tau t]$ denoting the expression (21) with (s, t) replaced by $(\sigma s, \tau t)$, and $s = \sqrt{1 - \rho x^{-1}}, t = \sqrt{1 - \rho y^{-1}}$.

Consequently, the density of the measure $d\beta$ is

$$\beta_d(x, y) = -\frac{1}{4\pi^2} \frac{\tilde{\beta}(x, y)}{xy}.$$

Substituting the expressions of s, t in terms of x, y, we find

$$\beta_d(x, y) = \frac{128}{\pi^2} d^2 (d-1)^2 (d-2) \frac{\sqrt{x(\rho - x)}\sqrt{y(\rho - y)}}{A},$$

with

$$A := x^2 y^2 \left((d^2(s^2+t^2) - (d-2)^2(1+s^2t^2))^2 - (8(d-1)st)^2 \right)$$

$$= \rho^2 \left[16(2d-3)(x-y)^2 + (d-2)^2 \right.$$

$$\times \left. \left(\rho\kappa \left(\frac{x+y}{2}\right)^2 + 4(d+3)(x-y)^2 + \rho^2(d-2)^2 \right) \right],$$

and $\kappa(x) = 2d\sqrt{x(\rho - x)}$. Then

$$\beta_d(x, y)$$

$$= \frac{2\,\kappa(x)\kappa(y)\,(d-2)\,d^4\,\pi^{-2}}{16(2d-3)(x-y)^2 + (d-2)^2\left(\rho\kappa\left(\frac{x+y}{2}\right)^2 + 4(d+3)(x-y)^2 + \rho^2(d-2)^2\right)}.$$

This proves the $d > 2$ case.

The case $d = 2$ can be easily shown by using continuity in the formulas as $d \downarrow 2$. We get

$$\beta_2(x, y) = \frac{8}{\pi} \kappa(x)\delta_x(y) = \frac{32}{\pi}\sqrt{x(1-x)}\,\delta_x(y).$$

The qualitative difference here is that for $d = 2$ the denominator of (21) does vanish along a line.

Finally, we have established that the covariance structure for $T(f)$ on the regular trees does not depend on the specific covariance structure of the noise B apart from

Assumptions B, C [note (19)]. It suffices to show the distributional convergence of the random variable $\{T_{G_n}(f_i), i = 1, \ldots k\}$ for any k and any real $f_i \in \mathcal{X}$. For any subsequence of (G_n, B_n) with $G_n \to \mathbb{T}_d$, we may pick a further subsequence so that the covariances of B_n converge to the covariance of some noise B for $\mathbb{T}_d$. Then the required joint convergence follows from Theorem 2 for this subsequence. Every subsequence contains a further subsequence that converges to the right limit, so the required joint convergence in fact holds for the original sequence. □

References

1. Anderson, G.W., Zeitouni, O.: A CLT for a band matrix model. Probab. Theory Rel. Fields **134**(2), 283–338 (2006)
2. Chen, L.H.Y., Shao, Q.-M.: Normal approximation under local dependence. Ann. Probab. **32**(3A), 1985–2028 (2004)
3. Deif., A.S.: Advanced Matrix Theory for Scientists and Engineers, 2nd edn. Abacus Press, New York (1991)
4. Diaconis, P., Evans, S.N.: Linear functionals of eigenvalues of random matrices. Trans. Am. Math. Soc. **353**(7), 2615–2633 (2001)
5. Gamburd, A., Hoory, S., Shahshahani, M., Shalev, A., Virág, B.: On the girth of random cayley graphs. Random Struct. Algorithms. arxiv:0707.1833 (2009, to appear)
6. Hoory, S., Linial, N., Wigderson, A.: Expander graphs and their applications. Bull. Am. Math. Soc. (N.S.) **43**(4), 439–561 (2006) (electronic)
7. Horn, R.A., Johnson, C.R.: Topics in Matrix Analysis. Cambridge University Press, Cambridge (1991)
8. Klein, A.: Extended states in the Anderson model on the Bethe lattice. Adv. Math. **133**(1), 163–184 (1998)
9. Rider, B., Virág, B.: The noise in the circular law and the Gaussian free field. Int. Math. Res. Not. IMRN (2), Art. ID rnm006, 33 (2007)
10. Royden, H.L.: Real Analysis, 3rd edn. Macmillan Publishing Company, New York (1988)
11. Villani, C.: Optimal Transport, Old and New, Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], vol. 338. Springer, Berlin (2009)
12. vom Scheidt, J., Purkert, W.: Random Eigenvalue Problems, vol. 63. Akademie-Verlag, Berlin (1983)
13. Woess, W.: Random Walks on Infinite Graphs and Groups, Cambridge Tracts in Mathematics, vol. 138. Cambridge University Press, Cambridge (2000)
14. Zeitouni, O.: Lecture notes on random walks in random environment. In: Lectures on probability theory and statistics, volume 1837 of Lecture Notes in Mathematics. Springer-Verlag, Berlin (2004). Jean Picard (ed.) Lectures from the 31st Summer School on Probability Theory held in Saint-Flour, 8–25 July 2001