Commun. Math. Phys. 401, 1291–1309 (2023)
Digital Object Identifier (DOI) https://doi.org/10.1007/s00220-023-04709-6

Communications in
Mathematical Physics

Many Nodal Domains in Random Regular Graphs

Shirshendu Ganguly[1], **Theo McKenzie**[2], **Sidhanth Mohanty**[1], **Nikhil Srivastava**[1]

[1] UC Berkeley, Berkeley, USA. E-mail: sganguly@berkeley.edu; sidhanthm@cs.berkeley.edu; nikhil@math.berkeley.edu
[2] Harvard University, Cambridge, USA. E-mail: tmckenzie@math.harvard.edu

Received: 13 December 2021 / Accepted: 25 January 2023
Published online: 24 April 2023 – © The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2023

Abstract: Let G be a random d-regular graph on n vertices. We prove that for every constant $\alpha > 0$, with high probability every eigenvector of the adjacency matrix of G with eigenvalue less than $-2\sqrt{d-2} - \alpha$ has $\Omega(n/\text{polylog}(n))$ nodal domains.

1. Introduction

Courant's nodal domain theorem states that the zero set of the kth smallest Dirichlet eigenfunction of the Laplacian on a smooth bounded domain in $\mathbb{R}^d$ partitions it into at most k connected components [CH53]. These components, known as the *nodal domains* of the eigenfunction, have garnered significant interest over time in spectral geometry and mathematical physics (see e.g. [Zel17]). The analogous definition for a finite discrete graph $G = (V, E)$ is the following.

Definition 1.1 (*Nodal domains*). A (*weak*) *nodal domain* of a function $f : V \to \mathbb{R}$ on G is a maximal connected subgraph S of G such that $f(u) \geqslant 0$ for all $u \in S$ or $f(u) \leqslant 0$ for all $u \in S$. A *strong nodal domain* of $f : V \to \mathbb{R}$ on G is a maximal connected subgraph S of G such that $f(u) > 0$ for all $u \in S$ or $f(u) < 0$ for all $u \in S$.

Fiedler [Fie75] showed that for a tree, the eigenvector of the kth smallest eigenvalue of the discrete Laplacian (defined as $L_G = D_G - A_G$ where D_G is the diagonal matrix of vertex degrees and A_G is the adjacency matrix) has exactly k nodal domains. Davies et al. [DGLS00] showed that for an arbitrary graph that the kth Laplacian eigenvector has at most k nodal domains and at most $k + m - 1$ strong nodal domains, where m is the multiplicity of the kth eigenvalue. Berkolaiko [Ber08] showed that for a connected

Shirshendu Ganguly supported by NSF Grant DMS-1855688, NSF Career Grant DMS-1945172, and a Sloan Fellowship.
Theo McKenzie supported by NSF GRFP Grant DGE-1752814 and NSF Grant DMS-2212881.
Sidhanth Mohanty supported by Google Ph.D. Fellowship.
Nikhil Srivastava supported by NSF Grant CCF-2009011

graph with n vertices and $n + \ell - 1$ edges (such that removing ℓ edges would produce a tree) the kth eigenvector of a Schrödinger operator with arbitrary potential has between $k - \ell$ and k nodal domains. Beyond these results, we are not aware of any lower bounds on the number of nodal domains of eigenvectors of any large class of graphs.

Our main result is the following lower bound on the number of nodal domains of a random regular graph.[1] We refer to a nodal domain with a single vertex as a *singleton* nodal domain.

Theorem 1.2. *Fix $d \geqslant 3$ and $\alpha > 0$ and let G be a random d-regular graph on n vertices. Then with probability $1 - o(1)$ as $n \to \infty$, every eigenvector of A_G with eigenvalue $\lambda \leqslant -2\sqrt{d-2} - \alpha$ has $\Omega\left(\frac{n}{\log^{C_{1.2}}(n)}\right)$ singleton nodal domains, where $C_{1.2} \leqslant 301$ is an absolute constant.*

Note that for large enough n, almost every $d-$regular graph has at least $\Omega(d^{-3/2}n)$ eigenvalues with $\lambda \leqslant -2\sqrt{d-2}$, as the spectrum of A_G converges weakly to the Kesten–McKay measure [McK81]. Since the Laplacian of a $d-$regular graph is equal to $dI - A_G$, the conclusion of the theorem also holds for the "high energy"[2] eigenvectors of the Laplacian with eigenvalues $\lambda \geqslant d + 2\sqrt{d-2} + \alpha$. We will accordingly also refer to highly negative eigenvalues of the adjacency matrix as high energy.

The proof of Theorem 1.2 appears in Sect. 3 to Sect. 6 and employs tools from random matrix theory (ℓ_∞ delocalization of eigenvectors of random regular graphs [HY21]), graph limits (weak convergence of eigenvectors of random regular graphs [BS19]), and combinatorics (expansion and short cycle counts of random regular graphs), and is outlined in Sect. 1.3. The conceptual phenomenon articulated by the proof is that (under certain conditions) high energy eigenvectors of graphs cannot simultaneously have few nodal domains and be delocalized. A simple demonstration of this tension for the easier case of $d = 3, 4$ is presented in Sect. 1.2. Due to the use of a weak convergence argument, there is no effective bound on the $o(1)$ probability in the statement of Theorem 1.2, and the proof requires d to be constant.

We complement Theorem 1.2 by observing in Sect. 7 (Theorem 7.2) that by an application of the expander mixing lemma, *every* non-leading eigenvector f of a d-regular expander graph G with sufficiently large spectral gap has two nodal domains which together contain a $(1 - o_d(1))$ fraction of the vertices of G.

1.1. History and related work. Random Graphs. Dekel et al. [DLL11] initiated the study of nodal domains of eigenvectors of Erdös-Rényi $G(n, p)$ random graphs. They showed that for constant p, with high probability all but $O(1)$ of the vertices are contained in two large nodal domains for every non-leading adjacency eigenvector. Arora and Bhaskara [AB11] improved this by establishing that when $p \geqslant n^{-1/19+o(1)}$ there are typically exactly 2 nodal domains in each non-leading eigenvector. Huang and Rudelson [HR20] proved that these two domains are approximately the same size for eigenvectors of eigenvalues macroscopically away from the edge when $p \in [n^{-c}, 1/2]$ for some fixed c and also for the first and last $e^{c(\log\log n)^2}$ eigenvectors when $p \in (0, 1)$ is constant. Linial

[1] We restrict our attention to weak nodal domains as there are at least as many strong domains as weak domains.

[2] The quantum chaos literature considers eigenfunctions of Laplacians on manifolds in the high energy limit. We analogously consider the highest Laplacian eigenvalues of graphs of fixed degree with the number of vertices going to infinity. High eigenvalue eigenfunctions have high frequency, and therefore, by the Planck formula, high energy.

suggested studying the shape of these nodal domains; for example, how many vertices are on the boundary of a domain, what is the distribution of distances to the boundary, etc. For sufficiently dense graphs sampled from $G(n, p)$, this geometry turned out to be trivial—in particular, Rudelson [Rud17, Section 5.2] showed that with high probability, for $G(n, p)$ with fixed $p \in (n^{-c}, 1)$, every vertex is adjacent to $\Omega(n/\text{polylog}n)$ vertices that have the opposite sign in each eigenvector f. This left open the question of nontrivial structure of the nodal domains for sparse graphs[3]. Theorem 1.2 and Theorem 7.2 show that both the number and the geometry of nodal domains is nontrivial for high energy eigenvalues of sparse random regular graphs.

In contrast to the situation for dense graphs, Dekel, Lee, and Linial observed that in simulations, a randomly selected d-regular graph with d constant has a number of nodal domains that increases as the eigenvalue becomes more negative. Our results confirm their observation that the most negative eigenvalues have many nodal domains.

Random Matrix Theory and Graph Limits. The results for $G(n, p)$ described above rely crucially on delocalization estimates in random matrix theory. There are two relevant notions of delocalization, namely ℓ_∞ norm bounds (the strongest being of order $\log^C n/\sqrt{n}$) and (ℓ_2) *no-gaps delocalization*, which asserts that every subset of tn vertices has at least a $\beta(t)$ fraction of the ℓ_2 mass of an eigenvector. Both types of bounds have been derived for models with independent entries via geometric arguments [RV15,RV16], and ℓ_∞ bounds have been obtained for random regular graphs via estimates on entries of their Green's functions [BHY19]. No-gaps delocalization is so far only known for sufficiently dense random $G(n, p)$ graphs [HR20], and remains open for sparse random graphs.

The proof of Theorem 1.2 relies on both notions of delocalization and combines them in a new way. We first consider no-gaps delocalization at scale $t = 1 - \delta$ for a small constant δ. In the case that this property holds for an eigenvector, we employ a weak convergence result of Backhausz and Szegedy [BS19] to argue that the local distribution of eigenvector entries around a randomly chosen vertex behaves like a Gaussian wave (defined in Sect. 2.4), implying that a random vertex is a singleton nodal domain with constant probability. Note that if the discrete version of Berry's conjecture ([Elo08], see below) is true as is believed, then no-gaps delocalization indeed holds with high probability for *every* bulk eigenvector in $(-2\sqrt{d-1}, 2\sqrt{d-1})$, so we expect this to be the only case.

Although we expect no gaps delocalization to hold, we manage to prove that even if it does not hold, then eigenvectors of sufficiently negative adjacency eigenvalue must have many nodal domains. To do this we apply the ℓ_∞ delocalization estimate of [BHY19, HY21] to the putative subset of δn vertices on which a given eigenvector is ℓ_2-localized; the ℓ_∞ bound allows us to exploit the locally almost-treelike structure of the graph on this subset and deduce many singleton nodal domains via a different argument which hinges on the negativity of the eigenvalue λ. Thus, we sidestep the current lack of no-gaps estimates for random regular graphs, as well as the difficulty of examining individual eigenvector entries solely using the Green's function method.[4]

[3] As a starting point, Eldan, H. Huang, and Rudelson asked in 2020 [Rud20] whether the most negative eigenvector of a sparse $G(n, p)$ graph has more than two nodal domains. Such graphs may have nontrivial nodal domain structure when p is near or below the connectivity threshold $\log n/n$. However, these graphs have a different, irregular, Benjamini-Schramm limit from random $d-$regular graphs. We expect our methods to work in this new setting, but this would require significant modifications, and we have not pursued this in this work.

[4] The Green's function $(A - zI)^{-1}$ of a random regular graph can only approximate that of the infinite tree when $\Im(z) \geqslant \text{polylog}n/n$, meaning that it inherently reflects the aggregate behavior of polylog$n$ eigenvectors.

Mathematical Physics. The field of quantum chaos aims to relate the classical dynamics of the geodesic flow on a manifold to the behavior of its high energy Laplacian eigenfunctions [Rud08], and the number of nodal domains has also been studied in this context [BGS02]. A guiding question in this area is Berry's random wave conjecture [Ber77], which asserts that the high energy eigenfunctions of quantum chaotic billiards behave like "Gaussian random waves" in the limit. Random d-regular graphs have been studied as a discrete model of quantum chaos [KS97,BOS07,Smi13]; in particular, a discrete analogue of Berry's conjecture considered in [Elo08] asserts that the bulk eigenvectors of random d-regular graphs have a (locally) jointly Gaussian distribution with a specific *nonzero* covariance matrix depending on the degree d. This conjecture implies the existence of many nodal domains in all bulk eigenvectors of random regular graphs. Theorem 1.2 proves the implication of the conjecture for sufficiently negative λ, and one branch of its proof (Sect. 3) is directly inspired by the "Gaussian wave" heuristic, which we make rigorous via the weak convergence result of [BS19].

1.2. Low degree case. As a warm-up, we prove a weaker version of Theorem 1.2 which applies to any eigenvector of a regular graph with sufficiently negative eigenvalue and an ℓ_∞ bound.

Proposition 1. *For $\alpha, \eta \geqslant 1$, $d \geqslant 3$, assume f is a unit eigenvector of a d-regular n-vertex graph $G = (V, E)$ with eigenvalue $\lambda \leqslant -(d-1) - \alpha$ and*

$$\|f\|_\infty \leqslant \frac{\eta}{\sqrt{n}}. \tag{1}$$

Then f has at least

$$\frac{n}{(2\eta)^{2+\frac{\log(d-1)}{\log(1+\alpha/(d-1))}}}$$

nodal domains.

Proof. Assume that $u \in V$ is not a singleton nodal domain and $|f(u)| \geqslant \frac{1}{2\sqrt{n}}$. Then u has at most $d-1$ neighbors v such that $f(u)f(v) \leqslant 0$, so as $\sum_{v\sim u} f(v) = \lambda f(u)$, we must have that for some neighbor v of u, $|f(v)| \geqslant (1+\alpha/(d-1))|f(u)|$. Repeating this argument, if there are no singleton nodal domains at distance at most k from u, then there is a path $(u = x_0, \ldots, x_k)$ such that $|f(x_i)| \geqslant (1+\alpha/(d-1))|f(x_{i-1})|$ for each i. By (1), we must have $k \leqslant \tilde{k}$ for

$$\tilde{k} := \frac{\log(2\eta)}{\log(1+\frac{\alpha}{d-1})}.$$

Every u with $|f(u)| \geqslant \frac{1}{2\sqrt{n}}$ must have a vertex w that is a singleton nodal domain and $d(u,w) \leqslant \tilde{k}$. By (1), there are at least $\frac{3}{4}n/\eta^2$ vertices u with $|f(u)| \geqslant 1/2\sqrt{n}$.

Any vertex w has at most $d(d-1)^{\tilde{k}-1}$ vertices at distance at most $\tilde{k}$. Therefore there are at least

$$\frac{\frac{3}{4}\cdot\frac{n}{\eta^2}}{d(d-1)^{\tilde{k}-1}} \geqslant \frac{n}{(2\eta)^2(d-1)^k} \geqslant \frac{n}{(2\eta)^{2+\frac{\log(d-1)}{\log(1+\alpha/(d-1))}}}$$

singleton nodal domains. □

The ℓ_∞ delocalization bound of [HY21] corresponds to $\eta = \text{polylog}n$. Thus if $d \leqslant 4$, $\alpha > 0$ are fixed and $\lambda \leqslant -(d-1)-\alpha$, Proposition 1 yields $\Omega(n/\text{polylog}n)$ nodal domains for an eigenvector of a random d-regular graph, recovering the conclusion of Theorem 1.2 up to polylogarithmic factors in the spectral window $[-d, -(d-1)-\alpha]$. We recall that every nontrivial eigenvalue λ of a random $d-$regular graph satisfies $|\lambda| \leqslant 2\sqrt{d-1} + o(1)$ with high probability [Fri03], so for $d > 5$ there are typically no eigenvectors with $\lambda \leqslant -(d-1)$ and Proposition 1 is vacuous. To improve the required bound on λ from $-(d-1)$ to $-2\sqrt{d-2}$, we shift from a local analysis of the entries of f to a more global one.

1.3. Proof outline and organization. In Sect. 2, we go over notation and some preliminary statements. In Sect. 3, we use the weak convergence result of Backhausz and Szegedy [BS19] to show that with high probability, if the ℓ_2 mass of an eigenvector f is not concentrated on a set of size δn for a small constant δ, then it has many singleton nodal domains. The remainder of the proof focuses on the case where the eigenvector f is ℓ_2-localized on a small set $S \subset G$. In Sect. 4 we give a deterministic upper bound of the spectral radius of "almost treelike" graphs in terms of their maximum degree, average degree, and girth; in particular, the bound implies that certain small subgraphs of G of maximum degree Δ have spectral radius approximately bounded by $2\sqrt{\Delta-1}$, with high probability. In Sect. 5 we show that if f has few singleton nodal domains in S, then we may pass to an edge subgraph $H \subset G[S]$ (of the induced subgraph $G[S]$) of maximum degree at most $d-1$ such that the restriction of f to S, denoted by f_S, satisfies

$$f_S^T A_H f_S \approx f^T A_G f = \lambda. \tag{2}$$

This is the step in which both the ℓ_2-localization assumption and the ℓ_∞ bound of [HY21] are crucially used. This subgraph H satisfies the conditions of the argument in Sect. 4, therefore $\|A_H\|_{op}$ is approximately at most $2\sqrt{d-2}$. Therefore, (2) cannot be satisfied if $\lambda \ll -2\sqrt{d-2}$, so we conclude that there must be many singleton nodal domains of f in S. We combine the above cases to prove Theorem 1.2 in Sect. 6. We conclude by showing that any sparse expander graph contains two nodal domains whose total size is large in Sect. 7.

2. Preliminaries

2.1. Notation and basic definitions. Given a graph G, we shall use $V(G)$ to denote its vertex set, $E(G)$ to denote its edge set, and A_G to denote its adjacency matrix. Throughout, given a graph G, n will be the number of vertices of G. We will order the n eigenvalues of A_G and denote them as:

$$\lambda_{\max}(G) = \lambda_1(G) \geqslant \lambda_2(G) \geqslant \cdots \geqslant \lambda_n(G).$$

For a subset of vertices $S \subseteq V(G)$ we use $G[S]$ to denote the induced subgraph of G on S. We use $N(S)$ to denote the set of vertices that have a neighbor in S. We use $E(S,T)$ to denote the collection of edges with one endpoint in S and one endpoint in T. We use $\overline{S}$ to denote the the set of vertices $V(G)\backslash S$. We use $B_G(S,\ell)$ to denote the induced subgraph on the set of all vertices of distance at most ℓ from S, and we write $B_G(v,\ell) := B_G(\{v\},\ell)$.

Given a vector $f \in \mathbb{R}^{V(G)}$, we use f_S to denote the vector in $\mathbb{R}^S$ obtained by restricting f to coordinates in S. We also will write $\|f\| := \|f\|_2$. For a matrix A, we use $\|A\|$ to denote the spectral norm of A.

We write $\Pr_{x\sim\mu}(E)$ to denote the probability that a random variable x sampled from the distribution μ satisfies E.

2.2. Graph theory. We use the following standard facts about expansion and cycle counts in random regular graphs.

Definition 2.1. The *spectral expansion* of a graph G, denoted $\lambda(G)$, is defined as $\max\{\lambda_2(G), -\lambda_n(G)\}$.

Definition 2.2. The *ε-edge expansion* of a graph G, denoted $\Psi_\varepsilon(G)$, is defined as:

$$\Psi_\varepsilon(G) := \max_{\substack{S\subseteq V(G)\\ |S|\leqslant \varepsilon n}} \frac{|E(S,\overline{S})|}{|S|}.$$

Definition 2.3 (*Bicycle-freeness*). We say G is ℓ-bicycle-free if for every vertex v, $B_G(v,\ell)$ contains at most 1 cycle.

Lemma 2.4 (Expander Mixing Lemma [BMS93, Lemma 9]). *Let G be a d-regular graph, $S, T \subseteq V(G)$, and $e(S,T)$ is the number of tuples (u,v) such that $u \in S$, $v \in T$ and $\{u,v\} \in E(G)$. Then:*

$$e(S,T) \in \frac{d}{n}|S|\cdot|T| \pm \lambda(G)\sqrt{|S|\cdot|T|\cdot\left(1-\frac{|S|}{n}\right)\cdot\left(1-\frac{|T|}{n}\right)}.$$

Lemma 2.5 (Edge expansion in random graphs [HLW18, Theorem 4.16]) . *Let G be a random d-regular graph. For every $\delta > 0$, there is an $\varepsilon > 0$ such that with probability $(1-o(1))$:*

$$\Psi_\varepsilon(G) \geqslant d - 2 - \delta.$$

Lemma 2.6 (Bicycle-freeness in random regular graphs [Bor19, Lemma 9]). *Let G be a random d-regular graph. There exists an absolute constant $c_{2.6} \in (0,1)$ such that with probability $1-o(1)$, G is ℓ-bicycle-free for any $\ell \leqslant c_{2.6}\log_{d-1} n$.*

We write $G\backslash F$ to signify $(V, E\backslash F)$. We use Lemma 2.6 to derive the following:

Lemma 2.7. *Let G be a random d-regular graph. Then with probability $1-o_n(1)$ there exists a collection of edges F with cardinality bounded by $(d-1)n^{1-c_{2.6}/2}$ such that $G\backslash F$ has girth $\ell := \frac{c_{2.6}}{2}\log_{d-1} n$.*

Proof. Let $\mathcal{C}$ be the collection of all cycles in G of length at most ℓ. By Lemma 2.6, G is 2ℓ-bicycle-free with probability 1-$o(1)$. Consequently, the collection of graphs given by $\mathcal{C}' := \{B_G(C,\ell) : C \in \mathcal{C}\}$ must be pairwise vertex-disjoint. Indeed, if there are distinct $C, C' \in \mathcal{C}$ for which $B_G(C,\ell)$ and $B_G(C',\ell)$ share a vertex v, then $B_G(v,2\ell)$ contains both C and C' contradicting bicycle-freeness.

By bicycle-freeness, for any $C \in \mathcal{C}$, the number of vertices in $B_G(C,\ell)$ is at least $(d-1)^{\ell-1} = \frac{n^{c_{2.6}/2}}{d-1}$, and by vertex-disjointness of the balls around cycles, $|\mathcal{C}'| \leqslant (d-1)n^{1-c_{2.6}/2}$. However, since $|\mathcal{C}| = |\mathcal{C}'|$, we have the same bound on $|\mathcal{C}|$. We can then construct F by choosing one edge per $C \in \mathcal{C}$, which completes the proof. □

2.3. Delocalization of eigenvectors of random regular graphs. A key ingredient in our proof is the following result about ℓ_∞-delocalization of eigenvectors in random regular graphs, as stated in [HY21, Theorem 1.4], which built on the previous result [BHY19].

Theorem 2.8. *Let $d \geqslant 3$ be a constant, and let G be a random d-regular graph. With probability $1 - O(n^{-1+o(1)})$ for all eigenvectors v:*

$$\|v\|_\infty \leqslant \frac{\log^{C_{\mathrm{HY}}} n}{\sqrt{n}} \|v\|,$$

where $C_{\mathrm{HY}} \leqslant 150$ is an absolute constant independent of d.

2.4. Gaussian wave. Our results also use results concerning the before-mentioned Gaussian wave.

Definition 2.9. Consider the infinite d-regular tree T_d with vertex set V_d and origin o. An *eigenvector process* with eigenvalue λ is a joint distribution $\{X_v\}_{v \in V_d}$, such that it is invariant under all automorphisms of the tree, $\mathbf{E}(X_o^2) = 1$, and satisfies the eigenvector equation

$$\lambda X_o = \sum_{v \sim o} X_v \tag{3}$$

with probability 1.

Observe that the eigenvector process must satisfy the eigenvector equation at every vertex by automorphism invariance, and that by taking the expectation of (3) and automorphism invariance, if $\mathbf{E}(X_o) \neq 0$, then $\lambda = d$.

Definition 2.10. A *Gaussian wave* is an eigenvector process that is also a Gaussian process.

Theorem 2.11 (Theorem 1.1 of [Elo09]). *For any $-d \leqslant \lambda \leqslant d$, there exists a unique Gaussian wave with parameter λ.*

We call this Gaussian wave Λ_λ.

Definition 2.12. The *Lévy Prokhorov distance* between two Borel probability measures μ_1 and μ_2 on $\mathbb{R}^k$ is given by

$$\tilde{d}(\mu_1, \mu_2) := \inf\{\epsilon > 0 | \forall A \in \mathcal{B}_k, \mu_1(A) \leqslant \mu_2(A_\epsilon) + \epsilon \text{ and } \mu_2(A) \leqslant \mu_1(A_\epsilon) + \epsilon\},$$

where $\mathcal{B}_k$ is the set of Borel measurable sets in $\mathbb{R}^k$ and A_ϵ is the neighborhood of radius ϵ around A.

Define C_ℓ to be the number of vertices in $B_{T^d}(v, \ell)$, where T_d is the infinite d-regular tree, and v is an arbitrary vertex. Namely

$$C_\ell := 1 + \frac{d((d-1)^\ell - 1)}{d-2}.$$

A vector $f \in \mathbb{R}^{V(G)}$ on the vertices of a d-regular graph G on n vertices defines the following distribution $\nu_{G,f,\ell}$ on $\mathbb{R}^{C_\ell}$. Select a vertex $u \in V$ uniformly at random. Order the vertices in $B(u, \ell)$ by starting a breadth first search at u, breaking ties in the order

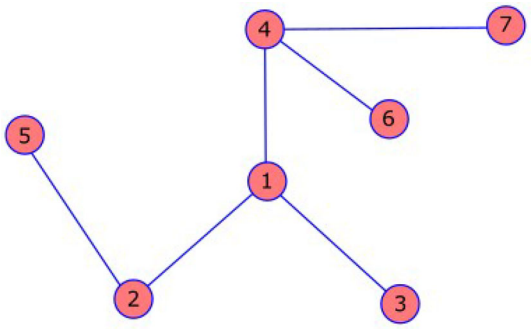

Fig. 1. An example of the ordering created by a breadth first search. On step 1 we label a given vertex 1. Then at step $k \geqslant 2$, we choose the vertex with the lowest label that has an unlabeled neighbor, then give the label k to an unlabeled neighbor of that vertex chosen arbitrarily

of the search uniformly at random 1. Create the vector $x := (x_1, \ldots x_{C_\ell})$ such that $x_k := \sqrt{n} f(u_k)$, where u_k is the kth vertex in this breadth first search. If $B(u, \ell)$ has fewer than C_ℓ vertices, then have $x_k = 0$ for $1 \leqslant k \leqslant C_\ell$. Finally, let $\nu_{G,f,\ell}$ be the distribution of $\overline{x}(u)$.

Theorem 2.13 (Theorem 2 of [BS19]). *For every $d \geqslant 3$, $\epsilon > 0$ and $R \in \mathbb{N}$, there exists N such that for $n > N$, with probability at least $1 - \epsilon$, a random regular graph G of degree d on n vertices has the following property. Any eigenvector f of G is such that $\nu_{G,f,R}$ is at most ϵ in Lévy-Prokhorov distance from the distribution of $\sigma \cdot \Lambda_\lambda$ restricted to the vertices of $B_{T^d}(o, R)$ for some $\sigma \in [0, 1]$, where λ is the eigenvalue of f.*

In fact, [BS19] proves that there is an N and a $\delta > 0$ such that a $G(n, d)$ graph has this property for all normalized vectors f such that there exists a constant λ such that $\|(A - \lambda I) f\| \leqslant \delta$. Namely, this statement is true for all "pseudo-eigenvectors".

3. Either ℓ_2-Localization or Many Nodal Domains

In this section, we show (Lemma 3.2) that if an eigenvector of a random regular graph is appropriately delocalized in ℓ_2, then its proximity to the Gaussian wave implies it has many nodal domains. We begin by showing that the root vertex in a Gaussian wave with negative parameter λ has a constant probability of being a singleton domain.

Lemma 3.1. *For $d \geqslant 3$ and $0 < \alpha \leqslant d$, let*

$$c_{3.1} := \frac{\alpha^d}{3^{d+2} d^{d+1}}.$$

Assume that $\lambda \leqslant -\alpha$. With probability at least $c_{3.1}$, $\{o\}$ is a singleton nodal domain in Λ_λ with all entries in $B(o, 1)$ of modulus at least $\alpha/5d$.

Proof. The proof proceeds by using the covariance of the Gaussian wave to pass to a Gaussian vector with i.i.d. entries, then showing that with probability at least $c_{3.1}$, this vector has a direction and norm that imply Lemma 3.1.

The distribution of Λ_λ restricted to $B(o, 1)$ is given by the multivariate normal distribution $N(\mathbf{0}, \Sigma)$ for a $(d+1) \times (d+1)$ covariance matrix Σ. The distribution according to $N(\mathbf{0}, \Sigma)$ is the same as the distribution of $\Sigma^{1/2} g$, where g is a length $(d+1)$ vector with

i.i.d. Gaussian $N(0,1)$ entries. Denote by $\{v_1, \ldots, v_d\}$ the neighbors of o and denote by e_v the elementary vector on v. Notice that $\langle \Sigma^{1/2} e_o, \Sigma^{1/2} e_o \rangle = \mathbf{E}(X_o^2) = 1$, and by automorphism invariance $\langle \Sigma^{1/2} e_{v_i}, \Sigma^{1/2} e_{v_i} \rangle = 1$. Moreover, by the eigenvector equation and automorphism invariance $\langle \Sigma^{1/2} e_o, \Sigma^{1/2} e_{v_i} \rangle = \mathbf{E}(X_o X_{v_i}) = \lambda/d \leqslant -\alpha/d$.

Let $\tilde{g} := g/\|g\|$. Next, we show that if $\tilde{g}$ is sufficiently close to $\Sigma^{1/2} e_o$, then it must have negative inner product with $\Sigma^{1/2} e_{v_i}$ for each $1 \leqslant i \leqslant d$.

If $\langle \tilde{g}, \Sigma^{1/2} e_o \rangle \geqslant 1 - \frac{\alpha^2}{16d^2}$,

$$\begin{aligned}
\langle \tilde{g}, \Sigma^{1/2} e_{v_i} \rangle &= 1 - \frac{1}{2} \|\tilde{g} - \Sigma^{1/2} e_{v_i}\|^2 \\
&\leqslant 1 - \frac{1}{2} \left(\|\Sigma^{1/2} e_o - \Sigma^{1/2} e_{v_i}\| - \|\tilde{g} - \Sigma^{1/2} e_o\| \right)^2 \\
&\leqslant 1 - \left(\sqrt{1 - \langle \Sigma^{1/2} e_o, \Sigma^{1/2} e_{v_i} \rangle} - \sqrt{1 - \langle \Sigma^{1/2} e_o, \tilde{g} \rangle} \right)^2 \\
&\leqslant 1 - \left(\sqrt{1 + \frac{\alpha}{d}} - \sqrt{\frac{\alpha^2}{16d^2}} \right)^2 \\
&\leqslant -\frac{\alpha}{d} - \frac{\alpha^2}{16d^2} + \frac{\alpha}{2d} \sqrt{1 + \frac{\alpha}{d}} \\
&\leqslant -\frac{\alpha}{5d}
\end{aligned}$$

for each i. The first inequality is the triangle inequality. The second is the parallelogram law. The last inequality is true as $\alpha/d \leqslant 1$.

The probability that $\|g\| \geqslant 1$ is at least the probability that the first coordinate of g has modulus at least 1. As this coordinate is standard normal, this probability is at least 0.3. The probability that $\langle \tilde{g}, e_o \rangle \geqslant 1 - \frac{\alpha^2}{16d^2}$ is the surface area of the spherical cap where this inequality is true divided by the surface area of the sphere. The surface area of the spherical cap is at least the volume of the d dimensional sphere base of the spherical cap.

The radius of the d-dimensional sphere is

$$\sqrt{1 - \left(1 - \frac{\alpha^2}{16d^2}\right)^2} = \sqrt{\frac{\alpha^2}{8d^2} - \frac{\alpha^4}{256d^4}} \geqslant \frac{\alpha}{3d}.$$

The volume of the d-dimensional unit sphere is $\pi^{d/2}/\Gamma(\frac{d}{2}+1))$, and the surface area is $2\pi^{d/2}/\Gamma(n/2)$. This means that the probability that $\langle \tilde{g}, e_o \rangle \geqslant 1 - \frac{\alpha^2}{16d^2}$ is at least

$$\left(\left(\frac{\alpha}{3d} \right)^d \cdot \frac{\pi^{d/2}}{\Gamma(\frac{d}{2}+1)} \right) \Bigg/ \left(\frac{2\pi^{(d+1)/2}}{\Gamma(\frac{d}{2}+\frac{1}{2})} \right) \geqslant \frac{\alpha^d}{3^d d^{d+1} \sqrt{\pi}}.$$

The probability that both $\langle g, e_o \rangle \geqslant 1 - \frac{\alpha^2}{16d^2}$ and $\langle g, \Sigma^{1/2} e_{v_i} \rangle \leqslant -\alpha/2d$ for each i is at least the probability that $\|g\| \geqslant 1$ and $\langle \tilde{g}, e_o \rangle \geqslant 1 - \frac{\alpha^2}{16d^2}$. By rotational invariance of g these are independent, so this probability is at least

$$0.3 \cdot \frac{\alpha^d}{3^d d^{d+1}\sqrt{\pi}} \geqslant \frac{\alpha^d}{3^{d+2} d^{d+1}}.$$

□

Lemma 3.2. *For any $d \geqslant 3$, $\delta > 0$ and $0 < \alpha \leqslant d$, there exists $N = N(d, \delta, \alpha)$ such that if $n > N$, then with probability at least $1 - \delta$ with respect to $G(n, d)$, for any unit eigenvector f with eigenvalue less than $-\alpha$ either*

1. *f has at least $c_{3.1} n/2$ singleton nodal domains, or*
2. *There is a set of vertices $S \subset V$, $|S| \leqslant \delta n$ such that $\sum_{v \in S} f(v)^2 \geqslant 1 - \delta$.*

Proof. Set σ to be the scalar of the Gaussian wave associated with f in Theorem 2.13. Define $\mu = \mu(d, \lambda, \sigma)$ to be the distribution of the Gaussian wave $\sigma \cdot \Lambda_\lambda$ restricted to $B(o, 1)$. Assume that $d(\mu, \nu_{G,f,1}) \leqslant \epsilon$, for $\epsilon \leqslant c_{3.1}/2$ to be fixed later. We consider two cases depending on the relationship between σ and ϵ. If σ is much larger than ϵ, then the eigenvector is delocalized, and we can use Lemma 3.1. Otherwise, the eigenvector is localized.

First, assume $\sigma \geqslant 10\epsilon d \alpha^{-1}$. Define A to be the set of vectors $\overline{x} := (x_o, x_{v_1}, \ldots, x_{v_d}) \in \mathbb{R}^{d+1}$ such that

1. $\min\{|x_o|, |x_{v_1}|, \ldots, |x_{v_d}|\} \geqslant \frac{\sigma\alpha}{5d}$ and
2. $x_o \cdot x_{v_i} < 0$ for each $1 \leqslant i \leqslant d$.

By Lemma 3.1, $\mu(A) \geqslant c_{3.1}$. By the definition of A, a given vector $\overline{x} \in A$ is such that all entries are of modulus at least $\frac{\sigma\alpha}{5d}$. Moreover, by the assumption on σ, we have $\epsilon \leqslant \frac{\sigma\alpha}{10d}$. Therefore, for a vector $\overline{y} := (y_o, y_{v_1}, \ldots, y_{v_d})$ such that $\|\overline{x} - \overline{y}\| \leqslant \epsilon$, the entries of $\overline{y}$ are of the same sign as the entries of $\overline{x}$. Therefore, if $x_o \cdot x_{v_i} < 0$ for each $1 \leqslant i \leqslant d$, then $y_o \cdot y_{v_i} < 0$ for each $1 \leqslant i \leqslant d$, meaning that if $B(o, 1)$ is colored as per $\overline{y}$, then $\{o\}$ is a singleton nodal domain.

As $\tilde{d}(\mu, \nu_{G,f,1}) \leqslant \epsilon$, we have $\nu_{G,f,1}(A_\epsilon) \geqslant \mu(A) - \epsilon \geqslant c_{3.1}/2$. By the previous paragraph, all vectors in A_ϵ correspond to singleton nodal domains. $\nu_{G,f,1}$ measures the empirical distribution of neighborhoods of radius 1 in G. Considering singleton nodal domains are defined by the radius 1 neighborhood, this means there are at least $c_{3.1} n/2$ singleton domains of f in G.

Now assume $\sigma < 10\epsilon d \alpha^{-1}$. In this case, we will show that because $\nu_{G,f,1}$ is close to a Gaussian with low variance, the distribution of entries of f must be concentrated around 0.

Denote by μ_0 the distribution of the value on o in μ, and $\nu_0 := \nu_{G,f,0}$. Note that μ_0 is the distribution $N(0, \sigma^2)$. The Euclidean distance between two points can only decrease when projecting onto a single coordinate, therefore the Lévy Prokhorov distance can only decrease as well. This means that as $\tilde{d}(\mu, \nu_{G,f,1}) \leqslant \epsilon$, then $\tilde{d}(\mu_0, \nu_0) \leqslant \epsilon$. Therefore for each $z \geqslant 0$,

$$\Pr_{x \sim \nu_0}(x \in [-z - \epsilon, z + \epsilon]) \geqslant \Pr_{x \sim \mu_0}(x \in [-z, z]) - \epsilon.$$

Fix $z := \sigma\sqrt{2 \log \frac{1}{\epsilon}}$ and observe that by Gaussian tail bounds

$$\Pr_{x \sim \mu_0}(x \notin [-z, z]) \leqslant 2\epsilon. \tag{4}$$

Also, by examining the endpoints of the interval, we have

$$\mathbf{E}_{x\sim\nu_0}\left(\mathbf{1}\left[x\in[-z-\epsilon,z+\epsilon]\right]\cdot x^2\right)\leqslant\left(\sigma\sqrt{2\log\frac{1}{\epsilon}}+\epsilon\right)^2.$$

By assumption, we have $\sigma<10\epsilon d\alpha^{-1}$. Therefore

$$\left(\sigma\sqrt{2\log\frac{1}{\epsilon}}+\epsilon\right)^2\leqslant\left(10\epsilon d\alpha^{-1}\sqrt{2\log\frac{1}{\epsilon}}+\epsilon\right)^2=\left(1+10d\alpha^{-1}\sqrt{2\log\frac{1}{\epsilon}}\right)^2\epsilon^2$$
$$\leqslant 250d^2\alpha^{-2}\epsilon^2\log\frac{1}{\epsilon}.$$

The last inequality follows as we assumed $d/\alpha\geqslant 1$. As $\frac{1}{\epsilon}>\log\frac{1}{\epsilon}$ and $\mathbf{E}_{x\sim\nu_0}(x^2)=1$, this means that

$$\mathbf{E}_{x\sim\nu_0}\left(\mathbf{1}\left[x\notin[-z-\epsilon,z+\epsilon]\right]\cdot x^2\right)\geqslant 1-250d^2\alpha^{-2}\epsilon.$$

Combining this with (4) and the definition of ν_0, this means that if $S=\{u\in V\,|\,f(u)^2\geqslant 2\sigma^2\log\frac{1}{\epsilon}\}$, then $|S|\leqslant 3\epsilon n$, and

$$\sum_{u\in S}f(u)^2=\frac{1}{n}\sum_{u\in S}nf(u)^2=\mathbf{E}_{x\sim\nu_0}\left(\mathbf{1}\left[x\notin[-z-\epsilon,z+\epsilon]\right]\cdot x^2\right)\geqslant 1-250d^2\alpha^{-2}\epsilon.$$

It is therefore sufficient to choose N as per Theorem 2.13 for

$$\epsilon<\min\left\{\frac{c_{3.1}}{2},\frac{\alpha^2}{250d^2}\delta\right\}.$$

□

4. Spectral Radius Bounds

The main result of this section is Lemma 4.6, where we prove bounds on the spectral radius of high-girth graphs with bounded maximum degree and hereditary degree (defined below) approximately equal to 2.

Definition 4.1. The *hereditary degree* of a graph H is defined as:

$$\max_{H'\subseteq H}\text{AvgDegree}(H')$$

where $\text{AvgDegree}(H')=2|E(H')|/|V(H')|$.

Definition 4.2. Given a collection of edges F, we will use $v(F)$ to denote the number of vertices adjacent to F, and $c(F)$ to denote the number of connected components formed by F.

Definition 4.3. Given a graph H and a collection of edges $F\subseteq E(H)$, we use 1_F to denote its indicator vector in $\mathbb{R}^{E(H)}$. The *forest polytope* of H is defined to be the convex hull of $\{1_F : F \text{ forest}\}$.

We will also need the following two ingredients.

Lemma 4.4. [Kes59] *If T is a forest with maximum degree bounded by Δ, then $\lambda_{\max}(A_T) \leqslant 2\sqrt{\Delta-1}$.*

The following fact about the forest polytope is a consequence of [KV12, Theorem 13.21].

Lemma 4.5. *The forest polytope of a graph H is equal to the feasible region of the following linear program:*

$$\begin{aligned} & y \in \mathbb{R}^{E(H)} \\ & y \geqslant 0 \\ \sum_{e \in F} y_e & \leqslant v(F) - c(F) \qquad\qquad \forall F \subseteq E(H). \end{aligned}$$

Lemma 4.6. *Let H be a graph with hereditary degree $2(1+\delta)$, maximum degree Δ, and girth g. Then:*

$$\lambda_{\max}(A_H) \leqslant 2\frac{1+\delta}{1-\frac{1}{g}}\sqrt{\Delta-1}.$$

Proof. Since A_H is a symmetric matrix with nonnegative entries,

$$\lambda_{\max}(A_H) = \max_{f \in \mathbb{R}^{V(H)} \setminus \{0\}} \frac{f^\top A_H f}{\|f\|^2}.$$

We will bound $f^\top A_H f$ for any f. Observe that:

$$f^\top A_H f = \sum_{(u,v) \in E(H)} f_u f_v.$$

We will prove that there is a forest T for which:

$$\frac{1-\frac{1}{g}}{1+\delta} f^\top A_H f \leqslant f^\top A_T f, \tag{5}$$

which by Lemma 4.4 is bounded by $2\sqrt{\Delta-1}$ hence implying

$$f^\top A_H f \leqslant 2\frac{1+\delta}{1-\frac{1}{g}}\sqrt{\Delta-1}.$$

To prove (5) we exhibit a distribution $\mathcal{D}$ on forests such that:

$$\mathbf{E}_{\boldsymbol{T} \sim \mathcal{D}} \left[f^\top A_{\boldsymbol{T}} f \right] = \frac{1-\frac{1}{g}}{1+\delta} f^\top A_H f.$$

Let $y \in \mathbb{R}^{E(H)}$ be the vector with $\frac{1-\frac{1}{g}}{1+\delta}$ in every entry. We claim that y is inside the forest polytope of H. To verify this, it suffices to check if y satisfies the linear constraints given by the linear program description of the polytope from Lemma 4.5. By construction, each $y_e \geqslant 0$.

For any $F \subseteq E(H)$, write it as $F_1 \cup \cdots \cup F_{c(F)}$ where each F_i is a connected component given by F. Since the girth of H is at least g, for any $|F_i| < g$ we know F_i forms a tree and hence $|F_i| = v(F_i) - 1$. For the remaining components, we know $|F_i| \leqslant v(F_i)(1+\delta)$ by our bound on the hereditary average degree. Now:

$$\begin{aligned}
\sum_{e\in F} y_e &= \sum_{i=1}^{c(F)} \sum_{e\in F_i} y_e \\
&= \sum_{i=1}^{c(F)} \frac{1-\frac{1}{g}}{1+\delta} |F_i| \\
&= \sum_{i\in[c(F)]:|F_i|<g} \frac{1-\frac{1}{g}}{1+\delta} |F_i| + \sum_{i\in[c(F)]:|F_i|\geqslant g} \frac{1-\frac{1}{g}}{1+\delta} |F_i| \\
&\leqslant \sum_{i\in[c(F)]:|F_i|<g} (v(F_i)-1) + \sum_{i\in[c(F)]:|F_i|\geqslant g} \left(1-\frac{1}{g}\right) v(F_i) \\
&\leqslant \sum_{i=1}^{c(F)} (v(F_i)-1) \\
&= v(F) - c(F).
\end{aligned}$$

The second to last inequality follows from the fact that for a graph of girth g, a subgraph with at least g edges has at least g vertices. Since y is in the forest polytope of H it must be expressible as a convex combination $p_1 T_1 + \cdots + p_s T_s$ of indicator vectors of forests in H. Let $\mathcal{D}$ be the distribution given by choosing forest T_i with probability p_i. Notice that for $\boldsymbol{T} \sim \mathcal{D}$ the probability that any given edge e is chosen is $\frac{1-\frac{1}{g}}{1+\delta}$. Now:

$$\begin{aligned}
\mathbf{E}_{\boldsymbol{T}\sim\mathcal{D}}\left[f^\top A_{\boldsymbol{T}} f\right] &= 2\mathbf{E}_{\boldsymbol{T}\sim\mathcal{D}}\left[\sum_{(u,v)\in E(H)} \mathbf{1}[(u,v)\in \boldsymbol{T}] f_u f_v\right] \\
&= 2\sum_{(u,v)\in E(H)} f_u f_v \Pr[(u,v)\in \boldsymbol{T}] \\
&= 2\frac{1-\frac{1}{g}}{1+\delta} \sum_{(u,v)\in E(H)} f_u f_v \\
&= \frac{1-\frac{1}{g}}{1+\delta} f^\top A_H f,
\end{aligned}$$

which completes the proof. □

5. ℓ_2-Localization Implies Many Nodal Domains

In this section G is a d-regular graph and f is a vector in $\mathbb{R}^{V(G)}$. We prove that under some suitable assumptions on G and f, it is not possible for f to simultaneously be

localized and have few nodal domains. Next, we verify that all of these conditions are simultaneously satisfied by random graphs and eigenvectors corresponding to sufficiently negative eigenvalues with high probability.

The conditions we impose on G are:

Almost high-girth: There is $F \subseteq E(G)$ such that $|F| \leqslant O(n^{1-c})$ and the girth of $G \backslash F$ is at least $c \log_{d-1} n$ for some absolute constant $c > 0$.

Lossless edge expansion: $\Psi_\varepsilon(G) \geqslant d - 2 - \delta$ for some constants $\varepsilon > 0$ and $0 < \delta < d - 2$.

The conditions we impose on f are:

ℓ_2-localization: There is a set $S \subseteq V(G)$ of size εn such that $\|f_S\|^2 \geqslant (1 - \eta)\|f\|^2$ for some small constant $\eta > 0$ such that $4d\sqrt{\eta} < \delta\sqrt{d-2}$.

ℓ_∞-delocalization: $\|f\|_\infty \leqslant \dfrac{\log^C n}{\sqrt{n}} \|f\|$ for some constant $C > 0$.

High energy: $f^\top A_G f = \lambda \|f\|^2$ for $\lambda < -2(1 + 2\delta)\sqrt{d-2}$.

We note that the labels for the conditions on G and f are not definitions of those properties, but rather for readability in back-referencing.

The key result of this section is the following. We emphasize that all nodal domains considered are weak nodal domains of f defined with respect to the graph G, and not its subgraphs.

Lemma 5.1. *If G and f satisfy the above conditions then f must have $\Omega\left(\frac{n}{\log^{2C+1} n}\right)$ singleton nodal domains.*

A key lemma in service of proving Lemma 5.1 is:

Lemma 5.2. *Let G, F, f and S satisfy the above conditions, and let c, d, δ and η be the parameters from above. If f has fewer than $\frac{n}{\log^{2C+1} n}$ singleton nodal domains in S, then there is a subgraph H of G on vertex set S such that:*

1. *The girth of H is at least $c \log_{d-1} n$.*
2. *The maximum degree of H is at most $d - 1$.*
3. *The hereditary degree of H is at most $2 + \delta$.*
4. *$f_S^\top A_H f_S \leqslant (\lambda + 4d\sqrt{\eta})\|f_S\|^2$.*

Proof. Let H be the graph obtained by starting with $G[S]$, and then deleting the edge subgraph

$$L := L_+ \cup L_\circ \cup (F \cap E(G[S]))$$

where L_+ is the subgraph obtained by choosing every edge $\{u, v\}$ in $G[S]$ such that $f_S(u) f_S(v) \geqslant 0$, and $L_\circ$ is obtained by choosing one arbitrary incident edge in $G[S]$ to each singleton nodal domain $v \in S$ with degree d in $G[S]$.

Proof of Item 1.. H is a subgraph of $G \backslash F$ and hence has girth at least $c \log_{d-1} n$.

Proof of Item 2.. Every vertex v with degree d in $G[S]$ has an incident edge in L: indeed, if v is a singleton nodal domain with degree d in $G[S]$ then one of its incident edges is added to $L_\circ$; otherwise v has a neighbor $u \in S$ such that $f_S(u) f_S(v) \geqslant 0$, which means $\{u, v\} \in L_+$. Consequently, every vertex in H has degree bounded by $d - 1$.

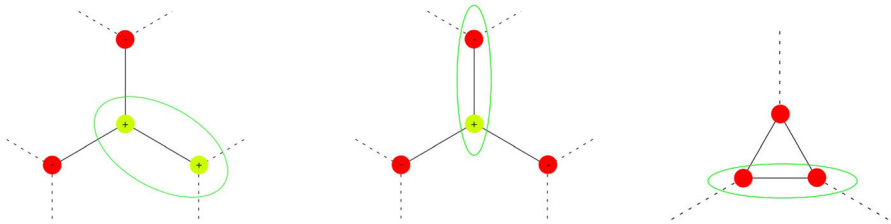

Fig. 2. The subgraph H of $G[S]$ is defined by deleting edges of three different types: (1) edges between vertices of the same sign in f; (2) one arbitrary edge from each singleton nodal domain; (3) a small set of edges to make our graph high girth

Proof of Item 3.. For any $T \subseteq S$, since $|T| \leqslant \varepsilon n$, it must be the case that $|E(T, \overline{T})| \geqslant d - 2 - \delta$ by "lossless edge expansion". Since G is a d-regular graph, the average degree of $G[T]$ must be at most $2 + \delta$. Consequently since $H[T]$ is a subgraph of $G[T]$, the average degree of $H[T]$ is also bounded by $2 + \delta$.

Proof of Item 4.. First observe that:

$$\begin{aligned}
\lambda \|f\|^2 &= f^\top A_G f \\
&= f_S^\top A_G f_S + 2 f_{\overline{S}}^\top A_G f_S + f_{\overline{S}}^\top A_G f_{\overline{S}} \\
&\geqslant f_S^\top A_{G[S]} f_S - 2d\sqrt{\eta}\|f\|^2 - d\eta\|f\|^2 \\
&\geqslant f_S^\top A_{G[S]} f_S - 3d\sqrt{\eta}\|f\|^2
\end{aligned}$$

where the third line follows from "ℓ_2-localization", Cauchy-Schwarz inequality, and $\lambda_{\max}(A_G) \leqslant d$. Consequently $f_S^\top A_{G[S]} f_S \leqslant (\lambda + 3d\sqrt{\eta})\|f\|^2$. Next, observe that:

$$\begin{aligned}
f_S^\top A_{G[S]} f_S &= f_S^\top A_H f_S + f_S^\top A_L f_S \\
&= f_S^\top A_H f_S + f_S^\top A_{L_+} f_S + f_S^\top A_{L_\circ} f_S + f_S^\top A_{F \cap E(G[S])} f_S \\
&\geqslant f_S^\top A_H f_S + f_S^\top A_{L_\circ} f_S + f_S^\top A_{F \cap E(G[S])} f_S && (\text{since } f_S^\top A_{L_+} f_S \geqslant 0) \\
&\geqslant f_S^\top A_H f_S - 2|L_\circ| \cdot \|f_S\|_\infty^2 - 2|F \cap E(G[S])| \cdot \|f_S\|_\infty^2 \\
&\geqslant f_S^\top A_H f_S - \left(\frac{2}{\log n} + O(n^{-c} \log^{2C} n) \right) \|f\|^2,
\end{aligned}$$

where the last inequality is because $|L_\circ| \leqslant \frac{n}{\log^{2C+1} n}$ by assumption, $\|f_S\|_\infty^2 \leqslant \frac{\log^{2C} n}{n}\|f\|^2$ by "ℓ_∞-delocalization", and $|F \cap E(G[S])| = O\left(n^{1-c}\right)$ by "almost-high girth".

Chaining the above two inequalities together gives us:

$$f_S^\top A_H f_S \leqslant \left(\lambda + 3d\sqrt{\eta} + O\left(\frac{1}{\log n} \right) \right) \|f\|^2 \leqslant (\lambda + 4d\sqrt{\eta})\|f\|^2.$$

Since $\lambda + 4d\sqrt{\eta} < 0$ and $\|f\|^2 \geqslant \|f_S\|^2$, the above is bounded by $(\lambda + 4d\sqrt{\eta})\|f_S\|^2$, completing the proof of Item 4.. □

We are now ready to prove Lemma 5.1.

Proof. (*Proof of Lemma* 5.1) We prove the desired statement by contradiction. If f has less than $\frac{n}{\log^{2C+1} n}$ singleton nodal domains then consider the subgraph H that is promised by Lemma 5.2. On one hand by Item 4. of Lemma 5.2:

$$f_S^\top A_H f_S \leqslant (\lambda + 4d\sqrt{\eta})\|f_S\|^2 \leqslant (-2(1+2\delta)\sqrt{d-2} + \delta\sqrt{d-2})\|f_S\|^2$$
$$= -2\left(1 + \frac{3}{2}\delta\right)\sqrt{d-2}\|f_S\|^2.$$

which implies that the spectral radius of A_H is lower bounded by $2\left(1+\frac{3}{2}\delta\right)\sqrt{d-2}$. On the other hand, by Item 1., Item 2. and Item 3. of Lemma 5.2 in conjunction with Lemma 4.6 the spectral radius of A_H is upper bounded by $\frac{(2+\delta)\sqrt{d-2}}{1-\frac{1}{c\log_{d-1} n}}$, which is at most $2(1+\delta)\sqrt{d-2}$, which is a contradiction. □

Remark 5.3. (*Sharpness of Lemma* 5.2) We remark that Item 2. in Lemma 5.2 is the source of the $\lambda \leqslant -2\sqrt{d-2} - \alpha$ hypothesis in Theorem 1.2; reducing the degree of H below $d-1$ would yield a larger spectral window in Theorem 1.2. The entirely local argument of the Lemma is seen to be sharp by taking $G[S] = \cup_{i=1}^{k} T_i$ to be a disjoint union of finite $d-$ary trees T_i of depth $O(\log\log n)$ such that the graph distance between any two trees is at least 2, and f to be an eigenfunction of a $(d-1)$-ary tree $T_i' \subset T_i$ with eigenvalue $\lambda \approx -2\sqrt{d-2}$ in each copy, and zero everywhere else. Then f and G satisfy the hypotheses of Lemma 5.2 locally, f has no singleton nodal domains in $G[S]$ (since each vertex has a path on which $f = 0$ to the leaves of the tree), and there is no subgraph of $G[S]$ of maximum degree strictly less than $d-1$ satisfying Item 4. Thus, improving Lemma 5.2 will require either additional hypotheses or a more global examination of the structure of G and f.

6. Many Nodal Domains in Random Regular Graphs

We are now ready to prove our main result.

Proof. (*Proof of Theorem* 1.2)

Fix $\delta, \eta > 0$, such that $4d\sqrt{\eta} < \delta\sqrt{d-2}$. We then set $\epsilon \leqslant \delta, \eta$ and ϵ small enough that δ, ϵ satisfy the conditions of Lemma 2.5. By Lemma 3.2, if we set the parameter δ from the theorem to be ϵ, then we obtain that with probability $1-\epsilon$, every eigenvector f either has $c_{3.1}n/2$ singleton nodal domains or satisfies "ℓ_2-localization" with parameters δ, η, ϵ.

We define the following events:

- $\mathcal{E}_1$: G satisfies "almost-high girth" with constant $c_{2.6}$ and "lossless edge expansion"; f satisfies "ℓ_∞-delocalization" with constant C_{HY} and "high energy",
- $\mathcal{E}_2$: f has at least $c_{3.1}n/2$ singleton nodal domains,
- $\mathcal{E}_3$: f satisfies "ℓ_2-localization" and has fewer than $c_{3.1}n/2$ singleton nodal domains.

Clearly, when $\mathcal{E}_2$ occurs there are $\Omega(n/\log^{2C+1} n)$ nodal domains. Next, observe that when both $\mathcal{E}_1$ and $\mathcal{E}_3$ occur, the conditions of Lemma 5.1 are satisfied for δ, η, ϵ, and f has $\Omega(n/\log^{2C+1} n)$ singleton nodal domains.

Thus, it suffices to lower bound $\Pr[\mathcal{E}_2 \cup (\mathcal{E}_1 \cap \mathcal{E}_3)]$. Since $\mathcal{E}_2$ and $\mathcal{E}_3$ are mutually exclusive, $\mathcal{E}_2$ and $\mathcal{E}_3 \cap \mathcal{E}_1$ are also mutually exclusive, and hence:

$$\Pr[\mathcal{E}_2 \cup (\mathcal{E}_1 \cap \mathcal{E}_3)] = \Pr[\mathcal{E}_2] + \Pr[\mathcal{E}_1 \cap \mathcal{E}_3] \geqslant \Pr[\mathcal{E}_2] + \Pr[\mathcal{E}_3] - \Pr\left[\overline{\mathcal{E}_1}\right]. \quad (6)$$

Lemma 3.2 implies that $\Pr[\mathcal{E}_2] + \Pr[\mathcal{E}_3] \geqslant 1 - \epsilon$. We further have $\Pr\left[\overline{\mathcal{E}_1}\right] = o(1)$ by a combination of Lemmas 2.7, 2.5 and Theorem 2.8. Thus,

$$\Pr[\mathcal{E}_2 \cup (\mathcal{E}_1 \cap \mathcal{E}_3)] \geqslant 1 - \epsilon.$$

ϵ is arbitrary, and the constant in the $\Omega(\cdot)$ term has no dependence on ϵ. Therefore, we can send $\epsilon \to 0$ while keeping the statement true, meaning there are $\Omega(n/\log^{2C+1} n)$ nodal domains with probability $1 - o(1)$.

□

7. Large Nodal Domains in Expanders

In this section we prove that as a consequence of expansion in random graphs, for any eigenvector of a random d-regular graph, most vertices are part of a macroscopic nodal domain. Key to our result in this section is the following lemma, which proves that by the expander mixing lemma, the only way to have the "correct" number of internal edges in a large subgraph is to have a large connected component.

Lemma 7.1. *Let G be a n-vertex d-regular graph and let $S \subseteq V(G)$ of size cn, where c is arbitrary. Also assume $\lambda(G) < d$. Then $G[S]$ has a connected component of size at least:*

$$\left(c - \frac{2(1-c)\lambda(G)}{d - \lambda(G)}\right) \cdot n.$$

Proof. By the expander mixing lemma (Lemma 2.4), we know that the average degree of $G[S]$ is:

$$\begin{aligned}\text{AvgDegree}(G[S]) &= \frac{|E(S,S)|}{|S|} \\ &\geqslant cd - \lambda(G)(1-c).\end{aligned}$$

Let the size of the connected component C^* in $G[S]$ with maximum average degree be $c'n$. We know that $\text{AvgDegree}(G[C^*])$ is at least $\text{AvgDegree}(G[S])$, and by the expander mixing lemma (Lemma 2.4):

$$\begin{aligned}\text{AvgDegree}(G[C^*]) &= \frac{e(C^*, C^*)}{|C^*|} \\ &\leqslant c'd + \lambda(G)(1-c').\end{aligned}$$

Consequently, we have:

$$\begin{aligned}c'd + \lambda(G)(1-c') &\geqslant cd - \lambda(G)(1-c) \\ c'(d - \lambda(G)) &\geqslant c(d + \lambda(G)) - 2\lambda(G) \\ c' &\geqslant c \cdot \frac{d + \lambda(G)}{d - \lambda(G)} - \frac{2\lambda(G)}{d - \lambda(G)} \\ &= c - \frac{2(1-c)\lambda(G)}{d - \lambda(G)}.\end{aligned}$$

which proves the claim. □

The result about nodal domains (which actually really applies to any signing of the vertices independent of being an eigenvector) in expanders is:

Theorem 7.2. *Let G be a d-regular graph and let f be any eigenvector of A_G. Suppose C_1 and C_2 be the two largest nodal domains in f, then $|C_1| + |C_2| \geqslant \left(1 - \frac{2\lambda(G)}{d-\lambda(G)}\right) n$.*

Proof. Let $S_+ := \{v \in V(G) : f(v) \geqslant 0\}$ and $S_- := \{v \in V(G) : f(v) < 0\}$. Let's denote $|S_+|$ as cn and $|S_-|$ as $(1-c)n$. By Lemma 7.1 we know that the largest component C_+ in S_+ has size at least $\left(c - \frac{2(1-c)\lambda(G)}{d-\lambda(G)}\right) \cdot n$ and the largest component C_- in S_- (which is distinct from C_+) has size at least $\left(1 - c - \frac{2c\lambda(G)}{d-\lambda(G)}\right) \cdot n$. It then follows:

$$\begin{aligned} |C_1| + |C_2| &\geqslant |C_+| + |C_-| \\ &\geqslant \left(1 - \frac{2\lambda(G)}{d - \lambda(G)}\right) \cdot n. \end{aligned}$$

□

Remark 7.3. When G is a random d-regular graph, then $\frac{2\lambda(G)}{d-\lambda(G)} = O\left(\frac{1}{\sqrt{d}}\right)$, and so for large enough d, the statement implies that a large constant fraction of the vertices are part of the two largest nodal domains. For instance, when $d \geqslant 99$, at least half the vertices are part of the two largest nodal domains.

References

[AB11] Arora, S., Bhaskara, A.: Eigenvectors of random graphs: delocalization and nodal domains. http://www.cs.princeton.edu/~bhaskara/files/deloc.pdf (2011)

[Ber77] Berry, M.V.: Regular and irregular semiclassical wavefunctions. J. Phys. A: Math. Gen. **10**(12), 2083 (1977)

[Ber08] Berkolaiko, G.: A lower bound for nodal count on discrete and metric graphs. Commun. Math. Phys. **278**(3), 803–819 (2008)

[BGS02] Blum, G., Gnutzmann, S., Smilansky, U.: Nodal domains statistics: a criterion for quantum chaos. Phys. Rev. Lett. **88**(11), 114101 (2002)

[BHY19] Bauerschmidt, R., Huang, J., Yau, H.-T.: Local Kesten–McKay law for random regular graphs. Commun. Math. Phys. **369**(2), 523–636 (2019)

[BMS93] Beigel, R., Margulis, G., Spielman, D.A.: Fault diagnosis in a small constant number of parallel testing rounds. In: Proceedings of the Fifth Annual ACM Symposium on Parallel Algorithms and Architectures, pp. 21–29 (1993)

[Bor19] Bordenave, C.: A new proof of Friedman's second eigenvalue theorem and its extension to random lifts. In: Annales scientifiques de l'Ecole normale supérieure (2019)

[BOS07] Band, R., Oren, I., Smilansky, U.: Nodal domains on graphs-how to count them and why? arXiv:0711.3416 (2007)

[BS19] Backhausz, Á., Szegedy, B.: On the almost eigenvectors of random regular graphs. Ann. Probab. **47**(3), 1677–1725 (2019)

[CH53] Courant, R., Hilbert, D.: Methods of Mathematical Physics: Partial Differential Equations. Wiley, New York (1953)

[DGLS00] Davies, E.B., Gladwell, G., Leydold, J., Stadler, P.F.: Discrete nodal domain theorems. arXiv:math/0009120 (2000)
[DLL11] Dekel, Y., Lee, J.R., Linial, N.: Eigenvectors of random graphs: nodal domains. Random Struct. Algorithms **39**(1), 39–58 (2011)
[Elo08] Elon, Y.: Eigenvectors of the discrete Laplacian on regular graphs-a statistical approach. J. Phys. A: Math. Theor. **41**(43), 435203 (2008)
[Elo09] Elon, Y.: Gaussian waves on the regular tree. arXiv:0907.5065 (2009)
[Fie75] Fiedler, M.: A property of eigenvectors of nonnegative symmetric matrices and its application to graph theory. Czechoslov. Math. J. **25**(4), 619–633 (1975)
[Fri03] Friedman, J.: A proof of Alon's second eigenvalue conjecture. In: Proceedings of the Thirty-Fifth Annual ACM Symposium on Theory of Computing, pp. 720–724 (2003)
[HLW18] Hoory, S., Linial, N., Wigderson, A.: Expander graphs and their applications. Bull. Am. Math. Soc. **43**(4), 439–561 (2018)
[HR20] Huang, H., Rudelson, M.: Size of nodal domains of the eigenvectors of a graph. Random Struct. Algorithms **57**(2), 393–438 (2020)
[HY21] Huang, J., Yau, H.-T.: Spectrum of random d-regular graphs up to the edge. arXiv:2102.00963 (2021)
[Kes59] Kesten, H.: Symmetric random walks on groups. Trans. Am. Math. Soc. **92**(2), 336–354 (1959)
[KS97] Kottos, T., Smilansky, U.: Quantum chaos on graphs. Phys. Rev. Lett. **79**(24), 4794 (1997)
[KV12] Korte, B., Vygen, J.: Combinatorial Optimization, vol. 2. Springer, Berlin (2012)
[McK81] McKay, B.D.: The expected eigenvalue distribution of a large regular graph. Linear Algebra Appl. **40**, 203–216 (1981)
[Rud08] Rudnick, Z.: Quantum chaos? Notices of the AMS **55**(1), 32–34 (2008)
[Rud17] Rudelson, M.: Delocalization of eigenvectors of random matrices. lecture notes. arXiv:1707.08461 (2017)
[Rud20] Rudelson, M.: Nodal domains of G(n,p) graphs. https://www.youtube.com/watch?v=ItMCuWiDRcI (2020)
[RV15] Rudelson, M., Vershynin, R.: Delocalization of eigenvectors of random matrices with independent entries. Duke Math. J. **164**(13), 2507–2538 (2015)
[RV16] Rudelson, M., Vershynin, R.: No-gaps delocalization for general random matrices. Geom. Funct. Anal. **26**(6), 1716–1776 (2016)
[Smi13] Smilansky, U.: Discrete graphs—a paradigm model for quantum chaos. In: Chaos, pp. 97–124. Springer (2013)
[Zel17] Zelditch, S.: Eigenfunctions of the Laplacian on a Riemannian Manifold, vol. 125. American Mathematical Society, New York (2017)

Communicated by J. Ding