Expansion of Random Graphs:
New Proofs, New Results

Doron Puder[†]

Einstein Institute of Mathematics
Hebrew University, Jerusalem
doronpuder@gmail.com

February 10, 2015

Abstract

We present a new approach to showing that random graphs are nearly optimal expanders. This approach is based on recent deep results in combinatorial group theory. It applies to both regular and irregular random graphs.

Let Γ be a random d-regular graph on n vertices, and let λ be the largest absolute value of a non-trivial eigenvalue of its adjacency matrix. It was conjectured by Alon [Alo86] that a random d-regular graph is "almost Ramanujan", in the following sense: for every $\varepsilon > 0$, $\lambda < 2\sqrt{d-1} + \varepsilon$ asymptotically almost surely. Friedman famously presented a proof of this conjecture in [Fri08]. Here we suggest a new, substantially simpler proof of a nearly-optimal result: we show that a random d-regular graph satisfies $\lambda < 2\sqrt{d-1} + 1$ a.a.s.

A main advantage of our approach is that it is applicable to a generalized conjecture: For d even, a d-regular graph on n vertices is an n-covering space of a bouquet of $d/2$ loops. More generally, fixing an arbitrary base graph Ω, we study the spectrum of Γ, a random n-covering of Ω. Let λ be the largest absolute value of a non-trivial eigenvalue of Γ. Extending Alon's conjecture to this more general model, Friedman [Fri03] conjectured that for every $\varepsilon > 0$, a.a.s. $\lambda < \rho + \varepsilon$, where ρ is the spectral radius of the universal cover of Ω. When Ω is regular we get a bound of $\rho + 0.84$, and for an arbitrary Ω, we prove a nearly optimal upper bound of $\sqrt{3}\rho$. This is a substantial improvement upon all known results (by Friedman, Linial-Puder, Lubetzky-Sudakov-Vu and Addario-Berry-Griffiths).

Contents

1

1 Introduction

Random d-regular graphs

Let Γ be a finite d-regular graph[†] on n vertices ($d \geq 3$) and let A_Γ be its adjacency matrix. The *spectrum* of Γ is the spectrum of A_Γ and consists of n real eigenvalues,

$$d = \lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n \geq -d.$$

The eigenvalue $\lambda_1 = d$ corresponds to constant functions and is considered the *trivial* eigenvalue of Γ. Let $\lambda(\Gamma)$ be the largest absolute value of a non-trivial eigenvalue of Γ, i.e. $\lambda(\Gamma) = \max\{\lambda_2, -\lambda_n\}$. $\lambda(\Gamma)$
This value measures the spectral expansion of the graph: the smaller $\lambda(\Gamma)$ is, the better expander Γ is (see Appendix B for details).

The well-known Alon-Boppana bound states that $\lambda(\Gamma) \geq 2\sqrt{d-1} - o_n(1)$ ([Nil91]), bounding the spectral expansion of an infinite family of d-regular graphs. There is no equivalent deterministic non-trivial upper bound: for example, if Γ is disconnected or bipartite then $\lambda(\Gamma) = d$. However, Alon conjectured [Alo86, Conj. 5.1] that if Γ is a *random d-regular graph*, then $\lambda(\Gamma) \leq 2\sqrt{d-1} + o_n(1)$ a.a.s. (asymptotically almost surely, i.e. with probability tending to 1 as $n \to \infty$) [‡].

Since then, a series of papers have dealt with this conjecture. One approach, due to Kahn and Szemerédi, studies the Rayleigh quotient of the adjacency matrix A_Γ and shows that it is likely to be small on all points of an appropriate ε-net on the unit sphere. This approach yielded an asymptotic bound of $\lambda(\Gamma) < c\sqrt{d}$ for some unspecified constant c [FKS89]. In the recent work [DJPP13, Thm. 26], it is shown that this bound can be taken to be 10^4. Other works, as well as the current paper, are based on the idea of the *trace method*, which amounts to bounding $\lambda(\Gamma)$ by means of counting closed walks in Γ. These works include [BS87], in which Broder and Shamir show that a.a.s. $\lambda(\Gamma) \leq \sqrt{2}d^{3/4} + \varepsilon$ ($\forall\varepsilon > 0$); [Fri91] where Friedman obtains $\lambda(\Gamma) \leq 2\sqrt{d-1} + 2\log d + c$ a.a.s.; and, most famously, Friedman's 100-page-long proof of Alon's conjecture [Fri08]. Friedman shows that for every $\varepsilon > 0$, $\lambda(\Gamma) \leq 2\sqrt{d-1} + \varepsilon$ a.a.s.

In the current paper we prove a result which is slightly weaker than Friedman's. However, the proof we present is substantially shorter and simpler than the sophisticated proof in [Fri08]. Our proof technique relies on recent deep results in combinatorial group theory [PP15]. We show the following:

Theorem 1.1. *Fix $d \geq 3$ and let Γ be a random d-regular simple graph on n vertices chosen at uniform distribution. Then*

$$\lambda(\Gamma) < 2\sqrt{d-1} + 1$$

asymptotically almost surely[§].

[†]Unless otherwise specified, a graph in this paper is undirected and may contain loops and multiple edges. A graph without loops and without multiple edges is called here *simple*.

[‡]In fact, Alon's original conjecture referred only to $\lambda_2(\Gamma)$, the second largest eigenvalue.

[§]For small d's better bounds are attainable - see the table in Section 6.2.

For d even, or d odd large enough, we obtain a better bound of $2\sqrt{d-1} + 0.84$. The same result, for d even, holds also for random d-regular graphs in the *permutation model* (see below). In fact, we first prove the result stated in Theorem 1.1 for random graphs in this model (with d even). The derivation of Theorem 1.1 for the uniform model and d even is then immediate by results of Wormald [Wor99] and Greenhill et al. [GJKW02] showing the *contiguity*[†] of different models of random regular graphs (see Appendix A). Finally, we derive the case of odd d relying on the even case and a contiguity argument in which we loose some in the constant and get 1 instead of 0.84 (Section 6.2).

The permutation model, which we denote by $\mathcal{P}_{n,d}$, applies only to even values of d. In this model, a random d-regular graph Γ on the set of vertices $[n]$ is obtained by choosing independently and uniformly at random $\frac{d}{2}$ permutations $\sigma_1, \ldots, \sigma_{\frac{d}{2}}$ in the symmetric group S_n, and introducing an edge $(v, \sigma_j(v))$ for every $v \in [n]$ and $j \in \left\{1, \ldots, \frac{d}{2}\right\}$. Of course, Γ may be disconnected and can have loops or multiple edges.

$\mathcal{P}_{n,d}$

We stress that even after Alon's conjecture is established, many open questions remain concerning $\lambda(\Gamma)$. In fact, very little is known about the distribution of $\lambda(\Gamma)$. A major open question is the following: what is the probability that a random d-regular graph is *Ramanujan*, i.e. that $\lambda(\Gamma) \leq 2\sqrt{d-1}$? There are contradicting experimental pieces of evidence (in [MNS08] it is conjectured that this probability tends to 27% as n grows; simulations depicted in [HLW06, Section 7] suggest it may be larger than 50%) . However, even the following, much weaker question is not known: are there infinitely many Ramanujan d-regular graphs for every $d \geq 3$? The only positive results here are by explicit constructions of Ramanujan graphs when $d - 1$ is a prime power by [LPS88, Mar88, Mor94]. In a recent major breakthrough, Marcus, Spielman and Srivastava [MSS13] show the existence of infinitely many d-regular *bipartite-Ramanujan* graphs for every $d \geq 3$ (namely, these graphs have two "trivial" eigenvalues, d and $-d$, while all others lie inside $\left[-2\sqrt{d-1}, 2\sqrt{d-1}\right]$). Still, the original problem remains open. We hope our new approach may eventually contribute to answering these open questions.

Random coverings of a fixed base graph

The hidden reason for the number $2\sqrt{d-1}$ in Alon's conjecture and Alon-Boppana Theorem is the following: All finite d-regular graphs are covered by the d-regular (infinite) tree $T = T_d$. Let $A_T : \ell^2(V(T)) \to \ell^2(V(T))$ be the adjacency operator of the tree, defined by

$$(A_T f)(u) = \sum_{v \sim u} f(v).$$

Then A_T is a self-adjoint operator and, as firstly proven by Kesten [Kes59], the spectrum of A_T is $\left[-2\sqrt{d-1}, 2\sqrt{d-1}\right]$. Namely, $2\sqrt{d-1}$ is the *spectral radius*[‡] of A_T. In this respect, among all possible (finite) quotients of the tree, Ramanujan graphs are "ideal", having their non-trivial spectrum as good as the "ideal object" T.

It is therefore natural to measure the spectrum of any graph Γ against the spectral radius of its covering tree. Several authors call graphs whose non-trivial spectrum is bounded by this value *Ramanujan*, generalizing the regular case. Many of the results and questions regarding the spectrum of d-regular graphs extend to this general case. For example, an analogue of Alon-Boppana's Theorem is given in Proposition 1.2.

Ideally, one would like to extend Alon's conjecture on nearly-Ramanujan graphs to every infinite tree T with finite quotients, and show that most of its quotients are nearly Ramanujan. However, as shown in [LN98], there exist trees T with some minimal finite quotient Ω which is not Ramanujan.

[†]Two models of random graphs are contiguous if the following holds: (*i*) for every (relevant) n they define distributions on the same set of graphs on n vertices, and (*ii*) whenever a sequence of events has a probability of $1 - o_n(1)$ in one distribution, it has a probability of $1 - o_n(1)$ in the other distribution as well.

[‡]The spectral radius of an operator M is defined as $\sup\{|\lambda| \,|\, \lambda \in \operatorname{Spec} M\}$.

All other finite quotients of T are then coverings of Ω, and inherit the "bad" eigenvalues of this quotient (we elaborate a bit more in Appendix A) . Such examples invalidate the obvious analogue of Alon's conjecture.

But what if we ignore this few, fixed, "bad" eigenvalues originated in the minimal quotient Ω and focus only on the remaining, "new" eigenvalues of each larger quotient? In this sense, a generalized version of Alon's conjecture is indeed plausible. Instead of studying the spectrum of a random finite quotient of T, one may consider the spectrum of a random finite covering of a fixed finite graph. This is the content of the generalized conjecture of Friedman appearing here as Conjecture 1.3.

In order to describe this conjecture precisely, let us first describe the random model we consider. This is a generalization of the permutation model for random regular graphs, which generates families of graphs with a common universal covering tree. A random graph Γ in the permutation model $\mathcal{P}_{n,d}$ can be equivalently thought of as a random n-sheeted covering space of the bouquet with $\frac{d}{2}$ loops. In a similar fashion, fix a finite, connected base graph Ω, and let Γ be a random n-covering space of Ω. More specifically, Γ is sampled as follows: its set of vertices is $V(\Omega) \times [n]$. A permutation $\sigma_e \in S_n$ is then chosen uniformly and independently at random for every edge $e = (u, v)$ of Ω, and for every $i \in [n]$ the edge $((u, i), (v, \sigma_e(i)))$ is introduced in $\Gamma^{\dagger}$. We denote this model by $\mathcal{C}_{n,\Omega}$ (so that $\mathcal{C}_{n,B_{\frac{d}{2}}} = \mathcal{P}_{n,d}$,

Figure 1.1: A 5-covering of a base graph using permutations.

$\mathcal{C}_{n,\Omega}$

where $B_{\frac{d}{2}}$ is the bouquet with $\frac{d}{2}$ loops). For example, all bipartite d-regular graphs on $2n$ vertices cover the graph $\bullet \!\!\! \equiv \!\!\! \bullet$ with two vertices and d edges connecting them. Various properties of random graphs in the $\mathcal{C}_{n,\Omega}$ model were thoroughly examined over the last decade (e.g. [AL02, ALM02, Fri03, LR05, AL06, BL06, LP10]). From now on, by a "random n-covering of Ω" we shall mean a random graph in the model $\mathcal{C}_{n,\Omega}$.

A word about the spectrum of a non-regular graph is due. In the case of d-regular graphs we have considered the spectrum of the adjacency operator. In the general case, it is not apriori clear which operator best describes in spectral terms the properties of the graph. In this paper we consider two operators: the *adjacency operator* A_Γ defined as above, and the *Markov operator* M_Γ defined by

A_Γ
M_Γ

$$(M_\Gamma f)(u) = \frac{1}{\deg(u)} \sum_{v \sim u} f(v).$$

(A third possible operator is the *Laplacian* - see Appendix B.) With a suitable inner product, each of these operators is self-adjoint and therefore admits a real spectrum (and see Appendix B for the relations of these spectra to expansion properties of Γ).

For a finite graph Ω on m vertices, the spectrum of the adjacency matrix A_Ω is

$$\mathfrak{pf}(\Omega) = \lambda_1 \geq \ldots \geq \lambda_m \geq -\mathfrak{pf}(\Omega),$$

$\mathfrak{pf}(\Omega)$ being the Perron-Frobenius eigenvalue of A_Γ. The spectrum of M_Ω is

$\mathfrak{pf}(\Omega)$

$$1 = \mu_1 \geq \ldots \geq \mu_m \geq -1,$$

the eigenvalue 1 corresponding to the constant function. Every finite covering Γ of Ω shares the same Perron-Frobenius eigenvalue, and moreover, inherits the entire spectrum of Ω (with multiplicity):

---

†We stress that we consider undirected edges. Although one should first choose an arbitrary orientation for each edge in order to construct the random covering, the orientation does not impact the resulting probability space.
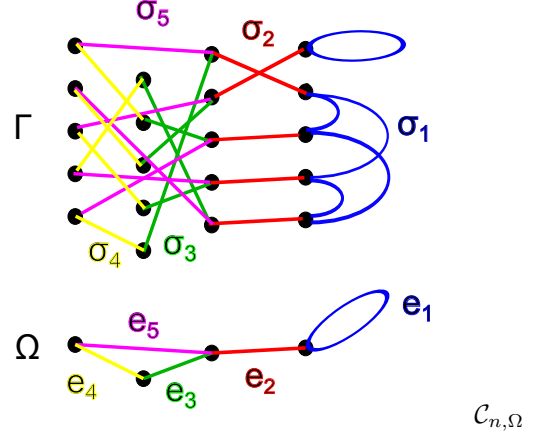
Let $\pi : \Gamma \to \Omega$ be the covering map, sending the vertex (v, i) to v and the edge $((u, i), (v, j))$ to (u, v). Every eigenfunction $f : V(\Omega) \to \mathbb{C}$ of any operator on $l^2(V(\Omega))$ as above, can be pulled back to an eigenfunction of Γ, $f \circ \pi$, with the same eigenvalue. Thus, every eigenvalue of Ω (with multiplicity) is trivially an eigenvalue of Γ as well. We denote by $\lambda_A(\Gamma)$ the largest absolute value of a *new* eigenvalue of A_Γ, namely the largest one not inherited from Ω. Equivalently, this is the largest absolute eigenvalue of an eigenfunction of Γ which sums to zero on every fiber of π. In a similar fashion we define $\lambda_M(\Gamma)$, the largest absolute value of a new eigenvalue of M_Γ. Note that in the regular case (i.e. when Ω is d-regular), $A_\Gamma = d \cdot M_\Gamma$, and so $\lambda_A(\Gamma) = d \cdot \lambda_M(\Gamma)$. Moreover, when $\Omega = B_{\frac{d}{2}}$ is the bouquet, $\lambda_A(\Gamma) = \lambda(\Gamma)$.

$\lambda_A(\Gamma)$

$\lambda_M(\Gamma)$

As in the regular case, the largest non-trivial eigenvalue is closely related to the spectral radius of T, the universal covering tree of Ω (which is also the universal covering of every connected covering Γ of Ω). We denote by $\rho_A(\Omega)$ and $\rho_M(\Omega)$ the spectral radii of A_T and M_T, resp. (So when Ω is d-regular, $\rho_A(\Omega) = d \cdot \rho_M(\Omega) = 2\sqrt{d-1}$.) First, there are parallels of Alon-Boppana's bound in this more general scenario. The first part of the following proposition is due to Greenberg, while the second one is due to Burger:

$\rho_A(\Omega), \rho_M(\Omega)$

Proposition 1.2. *Let Γ be an n-covering of Ω. Then*

(1) $\lambda_A(\Gamma) \geq \rho_A(\Omega) - o_n(1)$ *[Gre95, Thm 2.11].*

(2) $\lambda_M(\Gamma) \geq \rho_M(\Omega) - o_n(1)$ *[Bur87, GZ99, Prop. 6].*

When Ω is d-regular (but not necessarily a bouquet), this proposition was also observed by Serre [Ser90].

As in the d-regular case, the only deterministic upper bounds are trivial: $\lambda_A(\Gamma) \leq \mathfrak{pf}(\Omega)$ and $\lambda_M(\Gamma) \leq 1$. But there are interesting probabilistic phenomena. The following conjecture is the natural extension of Alon's conjecture. The adjacency-operator version is due to Friedman [Fri03]. We extend it to the Markov operator M as well:

Conjecture 1.3 (Friedman, [Fri03]). *Let Ω be a finite connected graph. If Γ is a random n-covering of Ω, then for every $\varepsilon > 0$,*
$$\lambda_A(\Gamma) < \rho_A(\Omega) + \varepsilon$$
asymptotically almost surely, and likewise
$$\lambda_M(\Gamma) < \rho_M(\Omega) + \varepsilon$$
asymptotically almost surely.

Since $\lambda_A(\Gamma)$ and $\lambda_M(\Gamma)$ provide an indication for the quality of expansion of Γ (see Appendix B), Conjecture 1.3 asserts that if the base graph Ω is a good (nearly optimal) expander then with high probability so is its random covering Γ.

In the same paper ([Fri03]), Friedman generalizes the method of Broder-Shamir mentioned above and shows that $\lambda_A(\Gamma) < \mathfrak{pf}(\Omega)^{1/2} \rho_A(\Omega)^{1/2} + \varepsilon$ a.a.s. An easy variation on his proof gives $\lambda_M(\Gamma) < \rho_M(\Omega)^{1/2} + \varepsilon$ a.a.s. In [LP10], Linial and the author improve this to $\lambda_A(\Gamma) < 3\mathfrak{pf}(\Omega)^{1/3} \rho_A(\Omega)^{2/3} + \varepsilon$ (and with the same technique one can show $\lambda_M(\Gamma) < 3\rho_M(\Omega)^{2/3} + \varepsilon$). This is the best known result for the general case prior to the current work.

Several works studied the special case where the base-graph Ω is d-regular (recall that in this case $\lambda_A(\Gamma) = d \cdot \lambda_M(\Gamma)$ and $\rho_A(\Omega) = 2\sqrt{d-1}$). Lubetzky, Sudakov and Vu [LSV11] find a sophisticated improvement of the Kahn-Szemerédi approach and prove that a.a.s. $\lambda_A(\Gamma) \leq C \cdot \max(\lambda(\Omega), \rho_A(\Omega)) \cdot \log \rho_A(\Omega)$ for some unspecified constant C. An asymptotically better bound of $430{,}656\sqrt{d}$ is given by Addario-Berry and Griffiths [ABG10], by further ameliorating the same basic technique (note that this bound becomes meaningful only for $d \geq 430{,}656^2$).

The following theorems differ from Conjecture 1.3 only by a small additive or multiplicative factor, and are nearly optimal by Proposition 1.2. They pose a substantial improvement upon all

former results, both in the special case of a d-regular base-graph Ω and, to a larger extent, in the general case of any finite base-graph.

Theorem 1.4. *Let Ω be an arbitrary finite connected graph, and let Γ be a random n-covering of Ω. Then for every $\varepsilon > 0$,*

$$\lambda_A\left(\Gamma\right) < \sqrt{3} \cdot \rho_A\left(\Omega\right) + \varepsilon$$

asymptotically almost surely, and similarly

$$\lambda_M\left(\Gamma\right) < \sqrt{3} \cdot \rho_M\left(\Omega\right) + \varepsilon$$

asymptotically almost surely.

For the special case where Ω is regular, we obtain the same bound as in the case of the bouquet (Theorem 1.1 for d even):

Theorem 1.5. *Let Ω be a finite connected d-regular graph ($d \geq 3$) and let Γ be a random n-covering of Ω. Then*

$$\lambda_A\left(\Gamma\right) < \rho_A\left(\Omega\right) + 0.84 = 2\sqrt{d-1} + 0.84$$

asymptotically almost surely.

We stress the following special case concerning random *bipartite d-regular* graphs. It follows as all bipartite regular graphs cover the graph Ω consisting of two vertices and d edges connecting them.

Corollary 1.6. *Let Γ be a random bipartite d-regular graph on n vertices ($d \geq 3$). Then*

$$\lambda_A\left(\Gamma\right) < 2\sqrt{d-1} + 0.84$$

asymptotically almost surely (as $n \to \infty$)†.

This means that alongside the two trivial eigenvalues $\pm d$, all other eigenvalues of the bipartite graph Γ are a.a.s. within $\left[-2\sqrt{d-1} - 0.84, 2\sqrt{d-1} + 0.84\right]$. The result applies also to random simple bipartite regular graphs: see appendix A.

To put Theorems 1.1, 1.4 and 1.5 in context, Table 1 summarizes the results mentioned above for the different cases in question, with respect to the adjacency operator A_Γ. (See also the late remark on Page 39 referring to the very recent work [FK14].)

Finally, let us stress that alongside the different models for random d-regular graphs, random coverings of a fixed, good expander, are probably the most natural other source for random, good expanders ("good" expanders are *sparse* graphs with high quality of expansion). Other known models for random graphs do not necessarily have this property. For example, the Erdös-Rényi model $G\left(n, p\right)$, fails to produce good expander graphs: when p is small ($O\left(\frac{1}{n}\right)$) the generic graph is not an expander (due, e.g., to lack of connectivity), whereas for larger values of p, the average degree grows unboundedly.

2 Overview of the Proof

In this section we present the outline of the proof of Theorems 1.1, 1.4 and 1.5. For simplicity, only the spectrum of the adjacency operator is considered in this section. We assume the reader has some familiarity with free groups, although we recall the basic definitions and classical relevant results throughout the text. For a good exposition of free groups and combinatorial group theory we refer the reader to [Bog08].

---

†Again, for small values of d a better bound is reachable - see Sections 6.2 and 6.3.

The base-graph Ω	Any graph	d-regular	$B_{\frac{d}{2}}$ = a bouquet of $\frac{d}{2}$ loops
		$\rho = 2\sqrt{d-1}$	$\rho = 2\sqrt{d-1}$
Deterministic lower bound for $\lambda_A(\Gamma)$	$\rho - o_n(1)$ [Gre95]	$\rho - o_n(1)$ [Ser90]	$\rho - o_n(1)$ (Alon-Boppana) [Nil91]
Conjectured probabilistic upper bound	$\rho + \varepsilon$ [Fri03]		$\rho + \varepsilon$ [Alo86]
Probabilistic upper bounds, ordered by asymptotic strength for growing ρ	$\sqrt{\mathfrak{pf}(\Omega)}\,\rho + \varepsilon$ [Fri03] $\implies$	$\sqrt{d\rho} + \varepsilon$	$\sqrt{d\rho} + \varepsilon$ [BS87]
	$3 \cdot \mathfrak{pf}(\Omega)^{1/3}\rho^{2/3} + \varepsilon$ [LP10] $\implies$	$3 \cdot d^{1/3}\rho^{2/3} + \varepsilon$	
		$C \cdot \max(\lambda(\Omega), \rho)\log\rho$ [LSV11]	
		$265{,}000 \cdot \rho$ [ABG10]	$6{,}200 \cdot \rho$ [FKS89, DJPP13]
	$\sqrt{3} \cdot \rho + \varepsilon$ **(Thm 1.4)**		
			$\rho + 2\log d + c$ [Fri91]
		$\rho + 0.84$ **(Thm 1.5)**	$\rho + 0.84$ **(Thm 1.1)**
			$\rho + \varepsilon$ [Fri08]

Table 1: Our results compared with former ones. As above, Ω is the connected base-graph and $\rho = \rho_A(\Omega)$ is the spectral radius of its universal covering tree. The results are ordered by their asymptotic strength.

Step I: The trace method

Let Ω be a fixed base graph with k edges and Γ a random n-covering in the model $\mathcal{C}_{n,\Omega}$. In the spirit of the trace method, the spectrum of Γ is analyzed by counting closed walks. More concretely, denote by $\mathcal{CW}_t(\Gamma)$ the set of closed walks of edge-length t in Γ. If $\mathrm{Spec}(A_\Gamma)$ denotes the multiset of eigenvalues of A_Γ, then for every $t \in \mathbb{N}$,

$$\sum_{\mu \in \mathrm{Spec}(A_\Gamma)} \mu^t = \mathrm{tr}\left(A_\Gamma^t\right) = |\mathcal{CW}_t(\Gamma)|.$$

Orient each of the k edges of Ω arbitrarily, label them by $x_1, \ldots, x_k$ and let $X = \{x_1, \ldots, x_k\}$. Let $\sigma_1, \ldots, \sigma_k \in S_n$ denote the random permutations by which Γ is defined: for each edge $x_j = (u, v)$ of Ω and each $i \in [n]$, Γ has an edge $((u, i), (v, \sigma_j(i)))$. Note that every closed walk in Γ projects to a closed walk in Ω. Thus, instead of counting directly closed walks in Γ, one can count, for every closed walk in Ω, the number of closed walks in Γ projecting onto it.

Let $w = x_{j_1}^{\varepsilon_1} \ldots x_{j_t}^{\varepsilon_t} \in \mathcal{CW}_t(\Omega) \subseteq \left(X \cup X^{-1}\right)^t$ be a closed walk in the base graph Ω, beginning (and terminating) at some vertex $v \in V(\Omega)$. (Here $\varepsilon_i = \pm 1$ and x_j^{-1} means the walk traverses the edge x_j in the opposite orientation.) For every $i \in [n]$ there is a unique lift of w to some walk in Γ, not necessarily closed, which begins at the vertex (v, i). This lifted walk terminates at the vertex (v, j), where j is obtained as follows: let $w(\sigma_1, \ldots, \sigma_k)$ denote the permutation obtained by composing $\sigma_1, \ldots, \sigma_k$ according to w, namely, $w(\sigma_1, \ldots, \sigma_k) = \sigma_{j_1}^{\varepsilon_1} \ldots \sigma_{j_t}^{\varepsilon_t} \in S_n$. Then j is the

$\mathcal{CW}_t(\Gamma)$

$\mathrm{Spec}(A_\Gamma)$

image of i under this permutation: $j = w(\sigma_1, \ldots, \sigma_k)(i) = \sigma_{j_1}^{\varepsilon_1} \ldots \sigma_{j_t}^{\varepsilon_t}(i)^\dagger$. Thus, the i-th lift of w is a closed walk if and only if i is a fixed point of the permutation $w(\sigma_1, \ldots, \sigma_k)$, and the number of closed walks in Γ projecting onto w is equal to the number of fixed points of $w(\sigma_1, \ldots, \sigma_k)$. We denote this number by $\mathcal{F}_{w,n} = \mathcal{F}_{w,n}(\sigma_1, \ldots, \sigma_k)$.

$\mathcal{F}_{w,n}$

Claim 2.1. *For every even $t \in \mathbb{N}$,*

$$\mathbb{E}\left[\lambda_A(\Gamma)^t\right] \leq \sum_{w \in \mathcal{CW}_t(\Omega)} \left[\mathbb{E}[\mathcal{F}_{w,n}] - 1\right]. \tag{2.1}$$

(The expectation on the l.h.s. is over $\mathcal{C}_{n,\Omega}$, which amounts to the i.i.d. uniform permutations $\sigma_1, \ldots, \sigma_k \in S_n$. The expectation on the r.h.s. is over the same k-tuple of permutations.)

Proof. Since t is even,

$$
\begin{aligned}
\lambda_A(\Gamma)^t &= \left(\max_{\mu \in \mathrm{Spec}(A_\Gamma) \backslash \mathrm{Spec}(A_\Omega)} |\mu|\right)^t \leq \sum_{\mu \in \mathrm{Spec}(A_\Gamma) \backslash \mathrm{Spec}(A_\Omega)} \mu^t = \sum_{\mu \in \mathrm{Spec}(A_\Gamma)} \mu^t - \sum_{\mu \in \mathrm{Spec}(A_\Omega)} \mu^t = \\
&= |\mathcal{CW}_t(\Gamma)| - |\mathcal{CW}_t(\Omega)| = \sum_{w \in \mathcal{CW}_t(\Omega)} \left[\mathcal{F}_{w,n}(\sigma_1, \ldots, \sigma_k) - 1\right].
\end{aligned}
$$

(Recall that we regard the spectrum of an operator as a multiset.) The claim is established by taking expectations. $\square$

We shall assume henceforth that t is an even integer. Note that in the special case where $\Omega = B_{\frac{d}{2}}$ is a bouquet of $\frac{d}{2}$ loops, $\mathrm{Spec}(A_\Omega) = \{d\}$, and $\mathcal{CW}_t(\Omega) = (X \cup X^{-1})^t$, i.e. it consists of all words of length t in the letters $X \cup X^{-1}$ (not necessarily reduced), so that $\left|\mathcal{CW}_t\left(B_{\frac{d}{2}}\right)\right| = d^t$.

Step II: The expected number of fixed points in $w(\sigma_1, \ldots, \sigma_k)$

The next stage in the proof of the main results is an analysis of $\mathbb{E}[\mathcal{F}_{w,n}]$. This is where the results from [PP15] come to bear. Let $\mathbf{F}_k = \mathbf{F}(X)$ be the free group on k generators. Every word $w \in \mathcal{CW}_t(\Omega) \subseteq (X \cup X^{-1})^t$ corresponds to an element of $\mathbf{F}_k$ (by abuse of notation we let w denote an element of $(X \cup X^{-1})^t$ and of $\mathbf{F}_k = F(X)$ at the same time; it is important to stress that reduction‡ of w does not affect the associated permutation $w(\sigma_1, \ldots, \sigma_k)$.) The main theorem in [PP15] estimates the expected number of fixed points of the random permutation $w(\sigma_1, \ldots, \sigma_k) \in S_n$, where $\sigma_1, \ldots, \sigma_k \in S_n$ are random permutations chosen independently with uniform distribution. This theorem shows that this expectation is related to an algebraic invariant of w called its *primitivity rank*, which we now describe.

A word $w \in \mathbf{F}_k$ is *primitive* if it belongs to a *basis*§ of $\mathbf{F}_k$. For a given w, one can also ask whether w is primitive as an element of different subgroups of $\mathbf{F}_k$ (which are free as well by a classical theorem of Nielsen and Schreier [Bog08, Chap. 2.8]). If w is primitive in $\mathbf{F}_k$, it is also primitive in every subgroup $J \leq \mathbf{F}_k$ (e.g. [Pud14, Claim 2.5]). However, if w is *not* primitive in $\mathbf{F}_k$, it is sometimes primitive and sometimes not so in subgroups containing it. Theoretically, one can go over all subgroups of $\mathbf{F}_k$ containing w, ordered by their rank¶, and look for the first time at which w is not primitive. First introduced in [Pud14], the primitivity rank of $w \in \mathbf{F}_k$ captures this notion:

primitive, basis

---

†For convenience, we use in this paper the non-standard convention that permutations are composed from left to right.

‡By reduction of a word we mean the (repeated) deletion of subwords of the form $x_i x_i^{-1}$ or $x_i^{-1} x_i$ for some $x_i \in X$.

§A basis of a free group is a free generating set. Namely, this is a generating set such that every element of the group can be expressed in a unique way as a reduced word in the elements of the set and their inverses. For $\mathbf{F}_k$ this is equivalent to a generating set of size k [Bog08, Chap. 2.29].

¶The *rank* of a free group $\mathbf{F}$, denoted $\mathrm{rk}(\mathbf{F})$, is the size of (every) basis of $\mathbf{F}$.

Definition 2.2. The *primitivity rank* of $w \in \mathbf{F}_k$, denoted $\pi(w)$, is $\hfill \pi(w)$

$$\pi(w) = \min \left\{ \text{rk}(J) \,\middle|\, \begin{array}{c} w \in J \leq \mathbf{F}_k \ s.t. \\ w \text{ is \textbf{not} primitive in } J \end{array} \right\}.$$

If no such J exists, i.e. if w is primitive in $\mathbf{F}_k$, then $\pi(w) = \infty$.
A subgroup J for which the minimum is obtained is called *w-critical*, and the set of w-critical subgroups is denoted $\text{Crit}(w)$. $\hfill \text{Crit}(w)$

For instance, $\pi(w) = 1$ if and only if w is a proper power ($w = v^d$ for some $v \in \mathbf{F}_k$ and $d \geq 2$). By Corollary 4.2 and Lemma 6.8 in [Pud14], the set of possible primitivity ranks in $\mathbf{F}_k$ is $\{0, 1, 2, \ldots, k\} \cup \{\infty\}$ (the only word w with $\pi(w) = 0$ being $w = 1$). Moreover, $\pi(w) = \infty$ iff w is primitive. The same paper also describes an algorithm to compute $\pi(w)$.

The following theorem estimates $\mathbb{E}[\mathcal{F}_{w,n}]$, the expected number of fixed points of $w(\sigma_1, \ldots, \sigma_k)$, where $\sigma_1, \ldots, \sigma_k \in S_n$ are chosen independently at random with uniform distribution:

Theorem 2.3. *[PP15, Thm 1.8] For every* $w \in \mathbf{F}_k$, *the expected number of fixed points in* $w(\sigma_1, \ldots, \sigma_k)$ *is*

$$\mathbb{E}[\mathcal{F}_{w,n}] = 1 + \frac{|\text{Crit}(w)|}{n^{\pi(w)-1}} + O\left(\frac{1}{n^{\pi(w)}}\right).$$

In particular, it is also shown that $\text{Crit}(w)$ is always finite. The three leftmost columns in Table 2 summarize the connection implied by Theorem 2.3 between the primitivity rank of w and the average number of fixed points in the random permutation $w(\sigma_1, \ldots, \sigma_k)$.

With Theorem 2.3 at hand, we can use the primitivity rank to split the summation in (2.1). We shall use the notation $\mathcal{CW}_t^m(\Omega) = \{w \in \mathcal{CW}_t(\Omega) \mid \pi(w) = m\}$ for the subsets we obtain by splitting $\hfill \mathcal{CW}_t^m(\Omega)$
$\mathcal{CW}_t(\Omega)$ according to primitivity ranks:

$$
\begin{aligned}
\mathbb{E}\left[\lambda_A(\Gamma)^t\right] &\leq \sum_{w \in \mathcal{CW}_t(\Omega)} (\mathbb{E}[\mathcal{F}_{w,n}] - 1) = \\
&= \sum_{m=0}^{k} \sum_{w \in \mathcal{CW}_t^m(\Omega)} \left(\frac{|\text{Crit}(w)|}{n^{m-1}} + O\left(\frac{1}{n^m}\right) \right)
\end{aligned}
\tag{2.2}
$$

(note that for primitive words, i.e. words with $\pi(w) = \infty$, the expected number of fixed points is exactly 1, so their contribution to the summation vanishes.)

Step III: A uniform bound for $\mathbb{E}[\mathcal{F}_{w,n}]$

The error term $O\left(\frac{1}{n^m}\right)$ in (2.2) depends on w. For a given $w \in \mathcal{CW}_t(\Omega)$, this error term becomes negligible as $n \to \infty$. However, in order to bound the r.h.s. of (2.2), one needs a uniform bound for all closed walks of length t with a given primitivity rank in Ω. Namely, for every m one needs to control the $O(\cdot)$ term for all $w \in \mathcal{CW}_t^m(\Omega)$ simultaneously. The third stage is therefore the following proposition:

Proposition (Follows from Prop. 5.1 and Claim 5.2)**.** *Let* $t = t(n)$ *and* $w \in \left(X \cup X^{-1}\right)^t$. *If* $t^{2k+2} = o(n)$ *then*

$$\mathbb{E}[\mathcal{F}_{w,n}] \leq 1 + \frac{|\text{Crit}(w)|}{n^{\pi(w)-1}} (1 + o_n(1)),$$

where the $o_n(1)$ *does not depend on* w.

Hence, as long as we keep $t^{2k+2} = o(n)$, we obtain:

$$\mathbb{E}\left[\lambda_A(\Gamma)^t\right] \leq (1 + o_n(1)) \sum_{m=0}^{k} \frac{1}{n^{m-1}} \sum_{w \in \mathcal{CW}_t^m(\Omega)} |\text{Crit}(w)|. \tag{2.3}$$

Step IV: Counting words and critical subgroups

The fourth step of the proof consists of estimating the exponential growth rate (as $t \to \infty$) of the summation $\sum_{w \in \mathcal{CW}_t^m(\Omega)} |\mathrm{Crit}(w)|$ for every $m \in \{0, 1, \ldots, k\}$. For $m = 0$, the only reduced word with $\pi(w) = 0$ is $w = 1$, and its sole critical subgroup is the trivial subgroup $\{1\}$, so $\sum_{w \in \mathcal{CW}_t^0(\Omega)} |\mathrm{Crit}(w)| = |\mathcal{CW}_t^0(\Omega)|$. Moreover, words reducing to 1 are precisely the completely back-tracking closed walks, i.e. the walks lifting to closed walks in the covering tree. It follows that the exponential growth rate of $|\mathcal{CW}_t^0(\Omega)|$ is exactly $\rho = \rho_A(\Omega)$, the spectral radius of the covering tree (see Claim 4.12). For larger m we obtain the following upper bound:

Theorem (Theorem 4.11). *Let Ω be a finite, connected graph with $k \geq 2$ edges, and let $m \in \{1, \ldots, k\}$. Then*

$$\limsup_{t \to \infty} \left[\sum_{w \in \mathcal{CW}_t^m(\Omega)} |\mathrm{Crit}(w)| \right]^{1/t} \leq (2m - 1) \cdot \rho.$$

This upper bound is not tight in general. However, in the special case where Ω is d-regular, we give better bounds:

Theorem (Follows from Corollaries 4.5 and 4.10 and from Theorem 8.5). *Let Ω be a finite, connected d-regular graph $(d \geq 3)$ with k edges, and let $m \in \{0, 1, \ldots, k\}$. Then*

$$\limsup_{t \to \infty} \left[\sum_{w \in \mathcal{CW}_t^m(\Omega)} |\mathrm{Crit}(w)| \right]^{1/t} \leq \begin{cases} 2\sqrt{2k - 1} & 2m - 1 \leq \sqrt{2k - 1} \\ 2m - 1 + \frac{2k - 1}{2m - 1} & 2m - 1 \geq \sqrt{2k - 1} \end{cases}.$$

Moreover, for $\Omega = B_{\frac{d}{2}}$ the bouquet, there is equality:

$$\limsup_{t \to \infty} \left[\sum_{w \in \mathcal{CW}_t^m\left(B_{\frac{d}{2}}\right)} |\mathrm{Crit}(w)| \right]^{1/t} = \begin{cases} 2\sqrt{2k - 1} & 2m - 1 \leq \sqrt{2k - 1} \\ 2m - 1 + \frac{2k - 1}{2m - 1} & 2m - 1 \geq \sqrt{2k - 1} \end{cases}.$$

Remark 2.4. In fact, in the case of the bouquet, the growth rates in the statement of the last theorem remain the same if we assume every word has only a single critical subgroup. That is, the r.h.s. gives also the growth rate of the *number* of words in $\left(X \cup X^{-1}\right)^t$ with primitivity rank m - see Theorem 8.5.

Table 2 summarizes the content of Theorems 2.3, 4.11 and 8.5[†].

Whereas in the special case of the bouquet we count words in $\mathbf{F}_k$ of a given length and a given primitivity rank, the case of a general graph concerns the equivalent question for words which in addition belong to some fixed subgroups of $\mathbf{F}_k$. (There is one such subgroup for each vertex v of Ω: the one consisting of the words which correspond to closed walks at v.) The fact that the bounds in Corollary 4.5 are better than those in Theorem 4.11 explains the gap between Theorems 1.1 and 1.5 which are tight up to a small *additive* constant, and Theorem 1.4 which is tight up to a small *multiplicative* factor.

[†]The number $2k - 2 + \frac{2}{2k-3}$ in the last row of the table is the exponential growth rate of the set of primitives in $\mathbf{F}_k$, namely of $\left| \mathcal{CW}_t^\infty \left(B_{\frac{d}{2}} \right) \right|$. (Primitive words have no critical subgroups.) This result is not necessary for the current work, and is established in a separate paper [PW14], using completely different techniques. We use it here only to show that our bounds for $\sum_{w \in \mathcal{CW}_t^m\left(B_{\frac{d}{2}}\right)} |\mathrm{Crit}(w)|$ are tight - see Section 8.

$\pi(w)$	Description of w	$\mathbb{E}\left[\mathcal{F}_{w,n}\right]$	Growth rate for the bouquet $B_{\frac{d}{2}}$	Bound on growth rate for general Ω
0	$w=1$	n	$2\sqrt{2k-1}$	ρ
1	a power	$\sim 1+\lvert\operatorname{Crit}(w)\rvert$	$2\sqrt{2k-1}$	ρ
2	E.g. $[x_1,x_2], x_1^2 x_2^2$	$\sim 1+\frac{\lvert\operatorname{Crit}(w)\rvert}{n}$	$2\sqrt{2k-1}$	3ρ
3	E.g. $x_1^2 x_2^2 x_3^2$	$\sim 1+\frac{\lvert\operatorname{Crit}(w)\rvert}{n^2}$	$2\sqrt{2k-1}$	5ρ
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$\left\lfloor\frac{\sqrt{2k-1}+1}{2}\right\rfloor$			$2\sqrt{2k-1}$	
$\left\lceil\frac{\sqrt{2k-1}+1}{2}\right\rceil$			$2\pi(w)-1+\frac{2k-1}{2\pi(w)-1}$	
$\vdots$			$\vdots$	
$k-1$	$\vdots$	$\vdots$	$2k-2+\frac{2}{2k-3}$	
k	E.g. $x_1^2\ldots x_k^2$	$\sim 1+\frac{\lvert\operatorname{Crit}(w)\rvert}{n^{k-1}}$	$2k$	
∞	primitive	1	$2k-2+\frac{2}{2k-3}$	

Table 2: Primitivity rank, the average number of fixed points, the exponential growth rate of $\sum_{w\in\mathcal{CW}_t^m\left(B_{\frac{d}{2}}\right)}\lvert\operatorname{Crit}(w)\rvert$, and bounds on the exponential growth rate of $\sum_{w\in\mathcal{CW}_t^m(\Omega)}\lvert\operatorname{Crit}(w)\rvert$.

Step V: Some analysis

The final step is fairly simple and technical: it consists of analyzing the upper bounds we obtain from (2.3) together with Theorem 4.11 and Corollary 4.5. We seek the value of t (as a function of n) which yields the best bounds.

The paper is arranged as follows. Section 3 provides some basic facts about the concepts of core graphs and algebraic extensions which are used throughout this paper. In Section 4 we bound the number of words and critical subgroups and establish the fourth step of the proof (first for the special case of the bouquet, in Section 4.1, then for the intermediate case of an arbitrary regular base graph in Section 4.2, and finally for the most general case in Section 4.3). The third step of the proof, where the error term from Theorem 2.3 is dealt with, is carried out in Section 5, where we have to recall some more details from [PP15]. Section 6 completes the proof of Theorems 1.1 and 1.5 and addresses the source of the gap between Theorem 1.1 and Friedman's result. In Section 7 we complete the proof of Theorem 1.4. We end with results on the accurate exponential growth rate of words with a given primitivity rank in $\mathbf{F}_k$ (Section 8), and then list a few open questions. The appendices provide some background on the relation between different models of random d-regular graphs and between different models of random coverings (Appendix A), and on the theory of spectral expansion of non-regular graphs (Appendix B).

3 Preliminaries: Core Graphs and Algebraic Extensions

This section describes some notions and ideas which are used throughout the current paper.

3.1 Algebraic extensions

Let $H \leq J$ be subgroups of $\mathbf{F}_k$. We say that J is an *algebraic extension* of H and denote $H \leq_{\mathrm{alg}} J$, if there is *no* intermediate subgroup $H \leq L \leq J$ which is a proper free factor[†] of J. The name originated in [KM02], but the notion goes back at least to [Tak51], and was formulated independently by several authors. It is central in the understanding of the lattice of subgroups of $\mathbf{F}$. For example, it can be shown that every extension $H \leq J$ of free groups admits a unique intermediate subgroup $H \leq_{alg} M \leq_{ff} J$ (where $\leq_{ff}$ denotes a free factor). Moreover, if $H \leq \mathbf{F}$ is a finitely generated subgroup, it has only finitely many algebraic extensions in $\mathbf{F}$. Thus, every group containing H is a free extension of one of finitely many extensions of H, which is a well known theorem of Takahasi [Tak51]. For more information we refer the interested reader to [KM02, PP15] and especially to [MVW07].

The importance of algebraic extensions in the current paper stems from the following easy observation:

Claim 3.1. *[Pud14, Cor. 4.4] Every w-critical subgroup is an algebraic extension of $\langle w \rangle$ (the subgroup generated by w).*
More precisely, $\mathrm{Crit}\,(w)$ *consists precisely of the algebraic extensions of $\langle w \rangle$ of minimal rank besides $\langle w \rangle$ itself*[‡]*.*

To see the claim, assume that H is a w-critical subgroup of $\mathbf{F}_k$. Obviously, $\langle w \rangle \lneq H$. If H is not an algebraic extension of $\langle w \rangle$, then there is a proper intermediate free factor $\langle w \rangle \leq L \lneq_{ff} H$. Since w is not primitive in H, it is also not primitive in L (as a basis containing w for L would extend to a basis for H), but $\mathrm{rk}\,(L) < \mathrm{rk}\,(M)$, which is a contradiction. Below, we use properties of w-critical subgroups which are actually shared by all proper algebraic extensions of $\langle w \rangle$.

3.2 Core graphs

Fix a basis $X = \{x_1, \ldots, x_k\}$ of $\mathbf{F}_k$. Associated with every subgroup $H \leq \mathbf{F}_k$ is a directed, pointed graph whose edges are labeled by X. This graph is called *the core-graph associated with H* and is denoted by $\Gamma_X\,(H)$. We illustrate the notion in Figure 3.1.
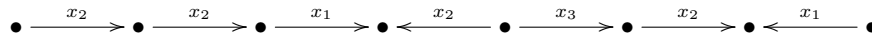
To understand how $\Gamma_X\,(H)$ is constructed, recall first the notion of the Schreier (right) coset graph of H with respect to the basis X, denoted by $\overline{\Gamma}_X\,(H)$. This is a directed, pointed and edge-labeled graph. Its vertex set is the set of all right cosets of H in $\mathbf{F}_k$, where the basepoint corresponds to the trivial coset H. For every coset Hw and every basis-element x_j there is a directed j-edge (short for x_j-edge) going from the vertex Hw to the vertex Hwx_j.[§]

The core graph $\Gamma_X\,(H)$ is obtained from $\overline{\Gamma}_X\,(H)$ by omitting all the vertices and edges of $\overline{\Gamma}_X\,(H)$ which are not traced by any reduced (i.e., non-backtracking) walk that starts and ends at the basepoint. Stated informally, we trim all "hanging trees" from $\overline{\Gamma}_X\,(H)$. To illustrate, Figure 3.1 shows the graphs $\overline{\Gamma}_X\,(H)$ and $\Gamma_X\,(H)$ for $H = \langle x_1 x_2 x_1^{-3}, x_1^2 x_2 x_1^{-2} \rangle \leq \mathbf{F}_2$.

If Γ is a directed pointed graph labeled by some set X, walks in Γ correspond to words in $\mathbf{F}\,(X)$ (the free group generated by X). For instance, the walk (from left to right)

$$\bullet \xrightarrow{x_2} \bullet \xrightarrow{x_2} \bullet \xrightarrow{x_1} \bullet \xleftarrow{x_2} \bullet \xrightarrow{x_3} \bullet \xrightarrow{x_2} \bullet \xleftarrow{x_1} \bullet$$

corresponds to the word $x_2^2 x_1 x_2^{-1} x_3 x_2 x_1^{-1}$. The set of all words obtained from closed walks around the basepoint in Γ is a subgroup of $\mathbf{F}\,(X)$ which we call the *labeled fundamental group* of Γ, and denote by $\pi_1^X\,(\Gamma)$. Note that $\pi_1^X\,(\Gamma)$ need not be isomorphic to $\pi_1\,(\Gamma)$, the standard fundamental

---

[†]If $H \leq J$ are free groups then H is said to be a *free factor* of J if a (every) basis of H can be extended to a basis of J.

[‡]Unless $w = 1$ in which case $\mathrm{Crit}\,(w) = \{\langle\rangle\} = \{\langle w \rangle\}$.

[§]Alternatively, $\overline{\Gamma}_X\,(H)$ is the quotient $H \backslash T$, where T is the Cayley graph of $\mathbf{F}_k$ with respect to the basis X, and F_k (and thus also H) acts on this graph from the left. Moreover, this is the covering-space of $\overline{\Gamma}_X\,(F_k) = \Gamma_X\,(F_k)$, the bouquet of k loops, corresponding to H, via the correspondence between pointed covering spaces of a space Y and subgroups of its fundamental group $\pi_1\,(Y)$.
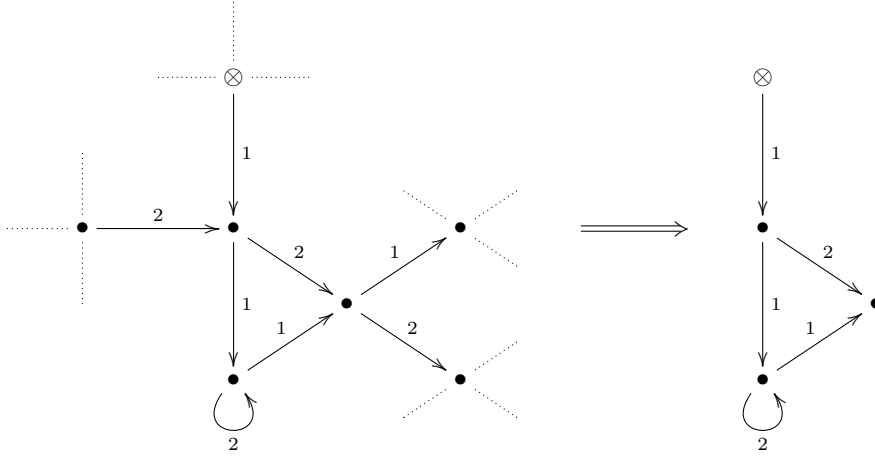
Figure 3.1: $\overline{\Gamma}_X(H)$ and $\Gamma_X(H)$ for $H = \langle x_1 x_2 x_1^{-3}, x_1^2 x_2 x_1^{-2} \rangle \leq \mathbf{F}_2$. The Schreier coset graph $\overline{\Gamma}_X(H)$ is the infinite graph on the left (the dotted lines represent infinite 4-regular trees). The basepoint "⊗" corresponds to the trivial coset H, the vertex below it corresponds to the coset Hx_1, the one further down corresponds to $Hx_1^2 = Hx_1 x_2 x_1^{-1}$, etc. The core graph $\Gamma_X(H)$ is the finite graph on the right, which is obtained from $\overline{\Gamma}_X(H)$ by omitting all vertices and edges that are not traced by reduced closed walks around the basepoint.

group of Γ viewed as a topological space - for example, take $\Gamma = {}^{x_1}\!\!\overset{\curvearrowright}{\bigcirc}\!\!\otimes\!\!\overset{\curvearrowleft}{\bigcirc}\!\!{}^{x_1}$.

However, it is not hard to show that when Γ is a core graph, then $\pi_1^X(\Gamma)$ *is* isomorphic to $\pi_1(\Gamma)$ (e.g. [KM02]). In this case the labeling gives a canonical identification of $\pi_1(\Gamma)$ as a subgroup of $\mathbf{F}(X)$. It is an easy observation that

$$\pi_1^X\left(\overline{\Gamma}_X(H)\right) = \pi_1^X\left(\Gamma_X(H)\right) = H \tag{3.1}$$

This gives a one-to-one correspondence between subgroups of $\mathbf{F}(X) = \mathbf{F}_k$ and core graphs labeled by X. Namely, π_1^X and Γ_X are the inverses of each other in a bijection (Galois correspondence)

$$\left\{ \begin{array}{c} \text{Subgroups} \\ \text{of } \mathbf{F}(X) \end{array} \right\} \xrightleftharpoons[\pi_1^X]{\Gamma_X} \left\{ \begin{array}{c} \text{Core graphs} \\ \text{labeled by } X \end{array} \right\}. \tag{3.2}$$

Core graphs were introduced by Stallings [Sta83]. Our definition is slightly different, and closer to the one in [KM02, MVW07] in that we allow the basepoint to be of degree one, and in that our graphs are directed and edge-labeled.

We now list some basic properties of core graphs which are used in the sequel of this paper (proofs can be found in [Sta83, KM02, MVW07, Pud14]).

Claim 3.2. *Let H be a subgroup of $\mathbf{F}_k$ with an associated core graph $\Gamma = \Gamma_X(H)$.*

(1) $\text{rk}(H) < \infty \Longleftrightarrow \Gamma$ is finite.

(2) $\text{rk}(H) = |E(\Gamma)| - |V(\Gamma)| + 1$ for finitely generated (f.g.) subgroup H.

(3) The correspondence (3.2) restricts to a correspondence between finitely generated subgroups of $\mathbf{F}_k$ and finite core graphs.

13

A *morphism* between two core-graphs is a map that sends vertices to vertices and edges to edges, and preserves the structure of the core graphs. Namely, it preserves the incidence relations, sends the basepoint to the basepoint, and preserves the directions and labels of the edges. As in Claim 3.2, each of the following properties is either proven in (some of) [Sta83, KM02, MVW07, Pud14] or is an easy observation:

Claim 3.3. *Let* $H, J, L \leq \mathbf{F}_k$ *be subgroups. Then*

(1) *A morphism* $\Gamma_X(H) \to \Gamma_X(J)$ *exists if and only if* $H \leq J$.

(2) *If a morphism* $\Gamma_X(H) \to \Gamma_X(J)$ *exists, it is unique. We denote it by* $\eta_{H \to J}^X$. $\qquad\qquad\qquad \eta_{H \to J}^X$

(3) *Whenever* $H \leq L \leq J$, $\eta_{H \to J}^X = \eta_{L \to J}^X \circ \eta_{H \to L}^X$.[†]

(4) *If* $\Gamma_X(H)$ *is a subgraph of* $\Gamma_X(J)$, *namely if* $\eta_{H \to J}^X$ *is injective, then* $H \leq_{ff} J$.[‡]

(5) *Every morphism in an immersion (locally injective at the vertices).*

4 Counting Words and Critical Subgroups

In this section we bound the exponential growth rate (as $t \to \infty$) of

$$\sum_{w \in \mathcal{CW}_t^m(\Omega)} |\mathrm{Crit}(w)|.$$

For the special case of the bouquet with $k = \frac{d}{2}$ loops, where $\mathcal{CW}_t\left(B_{\frac{d}{2}}\right) = \left(X \cup X^{-1}\right)^t$, we find the accurate exponential growth rate. The bound for a general graph Ω is given in terms of the spectral radius $\rho = \rho_A(\Omega)$ of the universal covering tree of Ω.

We begin with a key lemma to be used in the proofs of all cases (a bouquet, a d-regular base graph and an arbitrary base graph):

Lemma 4.1. *Let* $w \in \mathbf{F}_k$ *and let* $N \leq_{f.g.} \mathbf{F}_k$ *be a proper algebraic extension of* $\langle w \rangle$. *Then the closed walk in* $\Gamma_X(N)$ *corresponding to* w *traces every edge at least twice.*

Proof. First, we claim that every edge is traced at least once (in fact, even more generally, if $H \leq_{alg} N$ then $\eta_{H \to N}^X$ is onto: see Definition 5.3 and e.g. [PP15, Claim 4.2]. We repeat the simple argument here.) Otherwise, let J be the subgroup of N corresponding to the subgraph Δ traced by w (so $\Delta = \mathrm{im}\, \eta_{\langle w \rangle \to N}^X$), and $J = \pi_1^X(\Delta)$, see Section 3.2 and in particular Claim 3.3). Then $w \in J \lneq_{ff} N$ (Claim 3.3), contradicting the fact that N is an algebraic extension of $\langle w \rangle$.

Next, we distinguish between *separating* edges and *non-separating* edges in $\Gamma = \Gamma_X(N)$. If e is a separating edge, namely if removing e separates Γ into two connected components, then it is obvious that the walk of w in Γ must traverse e an even number of times, and since this number is ≥ 1, it is in fact ≥ 2.

Finally, assume that e is not separating, and w traverses it exactly once, so that the walk corresponding to w in $\Gamma_X(N)$ is $w_1 e w_2$ (with w_1, w_2 avoiding e; we think of e as oriented according to the direction of w). Choose a spanning tree T of $\Gamma_X(N)$ which avoids e to obtain a basis for N as follows. There are $r = \mathrm{rk}(N)$ excessive edges $e = e_1, e_2, \ldots, e_r$ outside the tree, and they should be oriented arbitrarily. For each $1 \leq i \leq r$ let u_i be the word corresponding to the walk that goes from $\otimes$ to the origin of e_i via T, then traverses e_i and returns to $\otimes$ via T. It is easy to see that $\{u_1, \ldots, u_r\}$ is a basis of N. We claim that so is $\{w, u_2, \ldots, u_r\}$, so that w is primitive in N and therefore $\langle w \rangle \leq_{ff} N$, a contradiction.

[†] Points (1)-(3) can be formulated by saying that (3.2) is in fact an isomorphism of categories, given by the functors π_1^X and Γ_X.

[‡] But not vice-versa: for example, consider $\langle x_1 x_2^2 \rangle \leq_{ff} \mathbf{F}_2$.

It is enough to show that $u_1 \in \langle w, u_2, \ldots, u_r \rangle$ (see footnote on Page 8). Let p_1 be the walk through T from $\otimes$ to the origin of e, and p_2 the walk from the terminus of e back to $\otimes$. Then

$$u_1 = p_1 e p_2 = p_1 w_1^{-1} w_1 e w_2 w_2^{-1} p_2 = \left(p_1 w_1^{-1}\right) w \left(w_2^{-1} p_2\right)$$

and we are done because $p_1 w_1^{-1}$ and $w_2^{-1} p_2$ avoid e and thus belong to $\langle u_2, \ldots, u_r \rangle$. $\qquad \square$

We will also use the following simple properties of the core graph of a subgroup of rank m. A "topological edge" of a graph is an edge of the graph obtained after ignoring all vertices of degree 2, except for (possibly) the basepoint $\otimes$.

Claim 4.2. *Let $\Gamma = \Gamma_X (J)$ be the core graph of a subgroup $J \leq \mathbf{F}_k$ of rank m. Then,*

(1) *After omitting the string to $\otimes$ if the basepoint is a leaf, all vertices of Γ are of degree at most $2m$.*

(2) *Γ has at most $3m - 1$ topological edges.*

Proof. *(1)* After ignoring $\otimes$ and the string leading to $\otimes$ in case it is a leaf, all vertices of Γ are of degree ≥ 2 . Thus all summands in the l.h.s. of

$$\sum_{v \in V(\Gamma)} [\deg(v) - 2] = 2|E(\Gamma)| - 2|V(\Gamma)| = 2m - 2$$

are non-negative. So the degree of every vertex is bounded by $2 + (2m - 2) = 2m$. In fact, there is a vertex of degree $2m$ if and only if Γ is topologically a bouquet of m loops (plus, possibly, a string to $\otimes$).

(2) Consider Γ as a "topological graph" as explained above. Let e and v denote the number of topological edges and vertices. It is still true that $e - v + 1 = m$, but now there are no vertices of degree ≤ 2 except for, possibly, the basepoint. Therefore, the sum of degrees, which equals $2e$, is at least $3(v-1) + 1$. So

$$2e \geq 3(v-1) + 1 = 3(e - m) + 1$$

so $e \leq 3m - 1$. $\qquad \square$

4.1 The special case of the bouquet

For the special case where $\Omega = B_{\frac{d}{2}}$ is the bouquet of $k = \frac{d}{2}$ loops, our goal is to bound the exponential growth rate of

$$\sum_{w \in \mathcal{CW}_t^m \left(B_{\frac{d}{2}}\right)} |\mathrm{Crit}(w)| = \sum_{w \in (X \cup X^{-1})^t : \pi(w) = m} |\mathrm{Crit}(w)|.$$

In order to estimate this number we first estimate the exponential growth rate of the parallel quantity for *reduced* words:

Proposition 4.3. *Let $k \geq 2$ and $m \in \{1, 2, \ldots, k\}$. Then*

$$\limsup_{t \to \infty} \left[\sum_{\substack{w \in \mathbf{F}_k : \\ |w| = t \ \& \ \pi(w) = m}} |\mathrm{Crit}(w)|\right]^{1/t} \leq \begin{cases} \sqrt{2k-1} & 2m - 1 \leq \sqrt{2k-1} \\ 2m - 1 & 2m - 1 \geq \sqrt{2k-1} \end{cases}.$$

Put differently, the $\limsup$ is bounded by $\max\left\{\sqrt{2k-1}, 2m-1\right\}$ (we present it in a lengthier way to stress the threshold phenomenon). In fact, this is not only an upper bound but the actual exponential growth rate - see Theorem 8.2.

15

Proof. Note that

$$\sum_{\substack{w \in \mathbf{F}_k: \\ |w|=t \,\&\, \pi(w)=m}} |\mathrm{Crit}\,(w)| \;=\; \sum_{J \leq \mathbf{F}_k:\, \mathrm{rk}(J)=m} |\{w \in \mathbf{F}_k \,|\, |w|=t,\, J \in \mathrm{Crit}\,(w)\}|$$

$$\leq \sum_{J \leq \mathbf{F}_k:\, \mathrm{rk}(J)=m} |\{w \in \mathbf{F}_k \,|\, |w|=t,\, \langle w \rangle \lesssim_{alg} J\}| \qquad (4.1)$$

$$\leq \sum_{J \leq \mathbf{F}_k:\, \mathrm{rk}(J)=m} \left| \left\{ w \in J \;\middle|\; \begin{array}{l} |w|=t,\, w \text{ traces each edge} \\ \text{of } \Gamma_X\,(J) \text{ at least twice} \end{array} \right\} \right|,$$

where the first inequality stems from Claim 3.1 and the second from Lemma 4.1. We continue to bound the latter sum. For each $J \leq \mathbf{F}_k$ let $\nu_t\,(J)$ denote the corresponding summand:

$$\nu_t\,(J) = \left| \left\{ w \in J \;\middle|\; \begin{array}{l} |w|=t,\, w \text{ traces each edge} \\ \text{of } \Gamma_X\,(J) \text{ at least twice} \end{array} \right\} \right|.$$

We classify all J's of rank m by the number of edges in $\Gamma_X\,(J)$. Consider all X-labeled core-graphs Γ of total size δt and rank m (so that δt is an integer, of course). Since we count words of length t tracing every edge at least twice, $\nu_t\,(J) = 0$ if $\delta > \frac{1}{2}$. So we restrict to the case $\delta \in \left[0, \frac{1}{2}\right]$. The counting is performed in several steps:

- First, let us bound the number of unlabeled and unoriented connected pointed graphs with δt edges and rank m (here the rank of a connected graph is $e - v + 1$). As in the proof of Lemma 4.1, each such graph has some spanning tree and m excessive edges. The walks through the tree from $\otimes$ to the origins and termini of these edges cover *the entire* tree. Denote these walks by $p_{1,1}, p_{1,2}, p_{2,1}, p_{2,2}, \ldots, p_{m,1}, p_{m,2}$. We "unveil" the spanning tree step by step: first we unveil $p_{1,1}$. The only unknown is its length $\in \{0, 1, \ldots, \delta t - 1\}$. Then $p_{1,2}$ leaves $p_{1,1}$ at one of $\leq \delta t$ possible vertices and goes on for some length $< \delta t$. Now, $p_{2,1}$ leaves $p_{1,1} \cup p_{1,2}$ at one of $\leq \delta t$ possible vertices and goes on for $< \delta t$ new edges. This goes on $2m$ times in total (afterward, the ends of $p_{i,1}$ and $p_{i,2}$ are connected by an edge). In total, there are at most $\left((\delta t)^2\right)^{2m} = (\delta t)^{4m}$ possible unlabeled pointed graphs of rank m with δt edges[†].

- Next, we bound the number of labelings of each such graph Γ (here, the labeling includes also the orientation of each edge). Label some edge (there are $2k$ options) and then gradually label edges adjacent to at least one edge which is already labeled (at most $2k - 1$ possible labels for each edge). Over all the number of possible labelings of Γ is $\leq 2k \cdot (2k - 1)^{\delta t - 1}$.

- For a given labeled core-graph Γ, let $J = \pi_1^X\,(\Gamma)$ be the corresponding subgroup. We claim that $\nu_t\,(J) \leq \left(4t^2\right)^{3m-1} \cdot (2m - 1)^{(1-2\delta)t}$. Indeed, note first that if the basepoint $\otimes$ is a leaf, then every reduced w must first follow the string from $\otimes$ to the first "topological" vertex (vertex of degree ≥ 3), and then return to the string only in its final steps back to $\otimes$. So we can assume w traces a leaf-free graph of rank m and at most δt edges. A reduced word $w \in J$ which traces every edge at least twice, also traverses any topological edge at least twice, each time in one shot (without backtracking). Each time w traces some topological edge $\widetilde{e}$ in Γ, it begins in one of $\leq t$ possible positions (in w), and from ≤ 2 possible directions of $\widetilde{e}$. So there $\leq 4t^2$ possible ways in which w traces $\widetilde{e}$ for the first two times. By Claim 4.2(2) there are at most $3m - 1$ topological edges, and so at most $\left(4t^2\right)^{3m-1}$ possibilities for how w traces each topological edge of Γ for the first two times. The rest of w is of length (at most) $(1 - 2\delta)\,t$, and in every step there are at most $2m - 1$ ways to proceed, by Claim 4.2(1).

[†]A tighter bound of $(\delta t)^{3m}$ can also be obtained quite easily. We do not bother to introduce it because this expression is anyway negligible when exponential growth rate is considered.

16

Hence,

$$\sum_{\substack{J \leq \mathbf{F}_k : rk(J)=m \\ |\Gamma_X(J)|=\delta t}} \nu_t(J) \leq (\delta t)^{4m} \cdot 2k\,(2k-1)^{\delta t-1} \cdot \left(4t^2\right)^{3m-1}(2m-1)^{(1-2\delta)t}$$

$$\leq c \cdot t^{10m-2} \cdot \left[(2k-1)^{\delta}(2m-1)^{1-2\delta}\right]^t$$

$$= c \cdot t^{10m-2} \cdot \left[\left(\frac{2k-1}{(2m-1)^2}\right)^{\delta}(2m-1)\right]^t. \tag{4.2}$$

Recall that $\delta \in \left[0, \frac{1}{2}\right]$ and $\delta t \in \mathbb{N}$. We bound $\sum_{J \leq \mathbf{F}_k : rk(J)=m} \nu_t(J)$ by $\frac{t}{2}$ times the maximal possible value of the r.h.s. of (4.2) (when going over all possible values of δ). When $2m - 1 \leq \sqrt{2k-1}$, the r.h.s. of (4.2) is largest when $\delta = \frac{1}{2}$, so we get overall

$$\sum_{J \leq \mathbf{F}_k : rk(J)=m} \nu_t(J) \leq c \cdot t^{10m-1} \cdot \left[\sqrt{2k-1}\right]^t. \tag{4.3}$$

For $2m - 1 \geq \sqrt{2k-1}$, the r.h.s. of (4.2) is largest when $\delta = 0$, so we get overall

$$\sum_{J \leq \mathbf{F}_k : rk(J)=m} \nu_t(J) \leq c \cdot t^{10m-1} \cdot [2m-1]^t.$$

The proposition follows. $\qquad\square$

The next step is to deduce an analogue result for non-reduced words. To this goal, we use an extended version of the well known *cogrowth formula* due to Grigorchuk [Gri77] and Northshield [Nor92]. Let Γ be a connected d-regular graph. Let $b_{\Gamma,v}(t)$ denote the number of *cycles* of length t at some vertex v in Γ, and let $n_{\Gamma,v}(t)$ denote the size of the smaller set of *non-backtracking cycles* of length t at v. The spectral radius of A_Γ, denoted $\mathrm{rad}(\Gamma)^\dagger$, is equal to $\limsup_{t \to \infty} b_{\Gamma,v}(t)^{1/t}$ (in particular, this limit does not depend on v). The *cogrowth* of Γ is defined as $\mathrm{cogr}(\Gamma) = \quad \mathrm{cogr}(\cdot)$ $\limsup_{t \to \infty} n_{\Gamma,v}(t)^{1/t}$, and is also independent of v.

The cogrowth formula expresses $\mathrm{rad}(\Gamma)$ in terms of $\mathrm{cogr}(\Gamma)$: it determines that $\mathrm{rad}(\Gamma) = g(\mathrm{cogr}(\Gamma))$, where $g : [1, d-1] \to \left[2\sqrt{d-1}, d\right]$ is defined by

$$g(\alpha) = \begin{cases} 2\sqrt{d-1} & \alpha \leq \sqrt{d-1} \\ \frac{d-1}{\alpha} + \alpha & \alpha \geq \sqrt{d-1} \end{cases}. \tag{4.4}$$

Another way to view the parameters $\mathrm{rad}(\Gamma)$ and $\mathrm{cogr}(\Gamma)$ is the following: let T_d be the d-regular tree with basepoint $\otimes$, let $p : T_d \to \Gamma$ be a covering map such that $p(\otimes) = v$, and let $S = p^{-1}(v) \subseteq V(T_d)$ be the fiber above v. Then $b_{\Gamma,v}(t)$ is the number of walks of length t in T_d emanating from $\otimes$ and terminating inside S. Similarly, $n_{\Gamma,v}(t)$ is the number of non-backtracking walks of length t in T_d emanating from $\otimes$ and terminating in S. This is also equal to the number of vertices in the t-th sphere‡ of T_d belonging to S.

For our needs we introduce (in a separate paper - [?])§ an extended formula applying to other types of subsets S of $V(T_d)$, which do not necessarily correspond to a fiber of a covering map of a graph. Even more generally, we extend the formula to a class of functions on $V(T_d)$ (this extends the previous case if S is identified with its characteristic function $\mathbb{1}_S$):

For $f : V(T_d) \to \mathbb{R}$, denote by $\beta_f(t)$ the sum $\qquad \beta_f(t)$

---

†If Γ is finite, $\mathrm{rad}(\Gamma) = d$. If Γ is the d-regular tree, $\mathrm{rad}(\Gamma) = 2\sqrt{d-1}$.

‡The t-th sphere of the pointed T_d is the set of vertices at distance t from $\otimes$.

§The results in [?] include a new proof of the original cogrowth formula.

$$\beta_f(t) = \sum_{\substack{p:\text{ a path from } \otimes \\ \text{of length } t}} f(\text{end}(p))$$

over all (possibly backtracking) walks of length t in T_d emanating from $\otimes$. Similarly, denote by $\nu_f(t)$ the same sum over the smaller set of *non-backtracking* walks of length t emanating from $\otimes$. $\quad \nu_f(t)$

Theorem 4.4. *[Extended Cogrwoth Formula [?]] Let $d \geq 3$, $f : V(T_d) \to \mathbb{R}$, $\beta_f(t)$ and $\nu_f(t)$ as above. If $\nu_f(t) \leq c \cdot \alpha^t$ for every t (and some $c > 0$) then*

$$\limsup_{t \to \infty} \beta_f(t)^{1/t} \leq g(\alpha).$$

With this theorem at hand, one can obtain the sought-after bound on the number of non-reduced words from the one on reduced words:

Corollary 4.5. *For every $k \geq 2$ and $m \in \{1, \ldots, k\}$,*

$$\limsup_{t \to \infty} \left[\sum_{w \in \mathcal{CW}_t^m \left(B_{\frac{d}{2}} \right)} |\text{Crit}(w)| \right]^{1/t} \leq \begin{cases} 2\sqrt{2k-1} & 2m-1 \leq \sqrt{2k-1} \\ \frac{2k-1}{2m-1} + 2m - 1 & 2m-1 \geq \sqrt{2k-1} \end{cases}.$$

Proof. Consider the the Cayley graph of $\mathbf{F}_k$ which is a $2k$-regular tree. Every vertex corresponds to a word in $\mathbf{F}_k$, and we let $f_m(w) = \mathbb{1}_{\pi(w)=m} |\text{Crit}(w)|$. The corollary then follows by applying Theorem 4.4 on f_m, using Proposition 4.3. $\qquad \square$

In Section 8 it is shown (Theorem 8.5) that the bound in Corollary 4.5 represents the accurate exponential growth rate of the sum, and even merely of the number of not-necessarily-reduced words with primitivity rank m. This result uses further results from [?].

Remark 4.6. Interestingly, the threshold of $\sqrt{2k-1}$ shows up twice, apparently independently, both in Proposition 4.3 and in the (extended) cogrowth formula.

Finally, for $m = 0$ there is exactly one relevant reduced word: $w = 1$, and this word has exactly one critical subgroup: the trivial subgroup. Thus, it suffices to bound the number of words in $\left(X \cup X^{-1} \right)^t$ reducing to 1. This is a well-known result:

Claim 4.7.

$$\limsup_{t \to \infty} \left| \mathcal{CW}_t^0 \left(B_{\frac{d}{2}} \right) \right|^{1/t} = \limsup_{t \to \infty} \left| \left\{ w \in \left(X \cup X^{-1} \right)^t \,\middle|\, w \text{ reduces to } 1 \right\} \right|^{1/t} = 2\sqrt{2k-1}.$$

Proof. Denote by $c_\Gamma(t, u, v)$ the number of walks of length t from the vertex u to the vertex v $\quad c_\Gamma(t, u, v)$ in a connected graph Γ. If, as above, A_Γ denotes the adjacency operator on $l^2(V(\Gamma))$, then $c_\Gamma(t, u, v) = \langle A_\Gamma^t \delta_u, \delta_v \rangle_1$ ($\langle \cdot, \cdot \rangle_1$ marks the standard inner product). If Γ has bounded degrees, then A_Γ is a bounded self-adjoint operator, hence

$$\text{rad}(\Gamma) = \|A_\Gamma\| = \limsup_{t \to \infty} c_\Gamma(t, u, v)^{1/t} \tag{4.5}$$

for every $u, v \in V(\Gamma)$. Moreover,

$$c_\Gamma(t, u, v) = \left\langle A_\Gamma^t \delta_u, \delta_v \right\rangle_1 \leq \left\| A_\Gamma^t \delta_u \right\| \cdot \|\delta_v\| \leq \|A_\Gamma\|^t \cdot \|\delta_u\| \cdot \|\delta_v\| = \text{rad}(\Gamma)^t \tag{4.6}$$

(For these facts and other related ones we refer the reader to [Lyo12, §6]).

The words of length t reducing to 1 are exactly the closed walks of length t at the basepoint of the $2k$-regular tree T_{2k}. So the number we seek is $\limsup_{t \to \infty} c_{T_{2k}}(t, v, v)^{1/t}$, which therefore equals $\text{rad}(T_{2k}) = 2\sqrt{2k-1}$. $\qquad \square$

18

4.2 An arbitrary regular base-graph Ω

We proceed with the observation that when Ω is d-regular (but not necessarily the bouquet), the bounds from Corollary 4.5 generally apply. We begin with a few claims that will be useful also in the next subsection dealing with irregular base graphs.

Let rk (Ω) denote the rank of the fundamental group of a finite graph Ω, so rk $(\Omega) = |E(\Omega)| - |V(\Omega)| + 1$. We claim there are no words in $\mathcal{CW}_t(\Omega)$ admitting finite primitivity rank which is greater than rk (Ω):

Lemma 4.8. *Let Ω be a finite, connected graph. Then $\pi(w) \in \{0, 1, \ldots, \text{rk}(\Omega), \infty\}$ for every $w \in \mathcal{CW}_t(\Omega)$.*

Proof. Recall from Section 2 that we denote $k = |E(\Omega)|$ and orient each of the k edges arbitrarily and label them by $x_1, \ldots, x_k$. With the orientation and labeling of its edges, Ω becomes a non-pointed X-labeled graph, where $X = \{x_1, \ldots, x_k\}$. (This is not a core-graph, for it has no basepoint and may have leaves.) So every walk in Ω of length t can be regarded as an element of $(X \cup X^{-1})^t$ and (after reduction) of $\mathbf{F}_k = \mathbf{F}(X)$. If a word $w \in \mathcal{CW}_t(\Omega)$ begins (and ends) at $v \in V(\Omega)$, then $w \in J_v$, where $J_v = \pi_1^X(\Omega_v)$ is the subgroup of $\mathbf{F}_k$ corresponding to the X-labeled graph Ω pointed at v. The rank of J_v is independent of v and equals rk (Ω). It is easy to see that $J_v \leq_{ff} \mathbf{F}_k$ (recall that '$\leq_{ff}$' denotes a free factor): obtain a basis for J_v by choosing an arbitrary spanning tree and orienting the edges outside the tree, as in the proof of Lemma 4.1. This basis can then be extended to a basis of $\mathbf{F}_k$ by the x_i's associated with the edges inside the spanning tree. So if w is primitive in J_v, is it also primitive in $\mathbf{F}_k$ and $\pi(w) = \infty$. Otherwise, $\pi(w) \leq \text{rk}(J_v) = \text{rk}(\Omega)$. $\square$

Moreover, proper algebraic extensions of words in $\mathcal{CW}_t(\Omega)$ are necessarily subgroups of J_v for some $v \in V(\Omega)$:

Claim 4.9. *In $w \in \mathcal{CW}_t(\Omega)$ is a cycle around the vertex v and $\langle w \rangle \leq_{alg} N$, then $N \leq J_v$.*

Proof. As $J_v \leq_{ff} \mathbf{F}_k$, it follows that $J_v \cap N \leq_{ff} N$ (see e.g. [PP15, Claim 3.9]). So if w belongs to N, it belongs to the free factor $J_v \cap N$ of N, which is proper, unless $N \leq J_v$. $\square$

If Ω is d-regular, $|E(\Omega)| = \frac{d}{2}|V(\Omega)|$ so that $\text{rk}(\Omega) = \left(\frac{d}{2} - 1\right)|V(\Omega)| + 1 \geq \frac{d}{2}$ (with equality only for the bouquet). The following Corollary distinguishes between three classes of primitivity rank: the interval $0, 1 \ldots, \left\lfloor \frac{\sqrt{d-1}+1}{2} \right\rfloor$, the interval $\left\lceil \frac{\sqrt{d-1}+1}{2} \right\rceil, \ldots, \lfloor \frac{d}{2} \rfloor$ and $\lceil \frac{d}{2} \rceil, \ldots, \text{rk}(\Omega)$.

Corollary 4.10. *Let Ω be a finite, connected d-regular graph, and let $m \in \{0, 1, \ldots, \text{rk}(\Omega)\}$. Then*

$$\limsup_{t \to \infty} \left[\sum_{w \in \mathcal{CW}_t^m(\Omega)} |\text{Crit}(w)| \right]^{1/t} \leq \begin{cases} 2\sqrt{d-1} & 2m-1 \in [-1, \sqrt{d-1}] \\ \frac{d-1}{2m-1} + 2m - 1 & 2m-1 \in [\sqrt{d-1}, d-1] \\ d & 2m-1 \in [d-1, 2\text{rk}(\Omega) - 1] \end{cases}.$$

Proof. First, for words with $\pi(w) = 0$, that is, words reducing to 1, their number is $|V(\Omega)|$ times the number of cycles of length t at a fixed vertex in the d-regular tree. Thus, as in the proof of Claim 4.7,

$$\limsup_{t \to \infty} \left[\sum_{w \in \mathcal{CW}_t^0(\Omega)} |\text{Crit}(w)| \right]^{1/t} = \limsup_{t \to \infty} \left| \mathcal{CW}_t^0(\Omega) \right|^{1/t} = 2\sqrt{d-1} \cdot \limsup_{t \to \infty} |V(\Omega)|^{1/t} = 2\sqrt{d-1}.$$

For $m \geq 1$, since the extended cogrowth formula (Theorem 4.4) applies here too, it is enough to prove that for *reduced* words we have:

$$\limsup_{t \to \infty} \left[\sum_{\substack{w \in \mathcal{CW}_t^m(\Omega): \\ w \text{ is reduced}}} |\text{Crit}(w)| \right]^{1/t} \leq \begin{cases} \sqrt{d-1} & 2m-1 \in [1, \sqrt{d-1}] \\ 2m-1 & 2m-1 \in [\sqrt{d-1}, d-1] \\ d-1 & 2m-1 \in [d-1, 2\text{rk}(\Omega) - 1] \end{cases}$$

19

From Claim 4.9 we deduce that every critical subgroup is necessarily a subgroup of $J_v = \pi_1^X(\Omega_v)$ for some vertex $v \in V(\Omega)$. As in the proof of Proposition 4.3, we denote

$$\nu_t(J) = \left| \left\{ w \in \mathbf{F}_k \; \middle| \; \begin{array}{c} |w| = t, \; w \text{ traces each edge} \\ \text{of } \Gamma_X(J) \text{ at least twice} \end{array} \right\} \right|$$

for every $J \leq \mathbf{F}_k$, and as in (4.1), we obtain the bound:

$$\sum_{\substack{w \in \mathcal{CW}_t^m(\Omega): \\ w \text{ is reduced}}} |\mathrm{Crit}(w)| \leq \sum_{v \in V(\Omega)} \sum_{J \leq J_v: \, \mathrm{rk}(J) = m} \nu_t(J).$$

We carry the same counting argument as in the proof of Proposition 4.3:

- The first stage, where we count unlabeled and unoriented pointed graphs of a certain size and rank remains unchanged.

- For the second stage of labeling and orienting the graph, we first choose v ($|V(\Omega)|$ options), and then we use the fact that whenever $J \leq J_v$, there is a core-graph morphism $\eta : \Gamma_X(J) \to \Omega_v$, which is, as always, an immersion (i.e. locally injective). So we first label an arbitrary edge incident to the basepoint $\otimes$, and this one has to be labeled like one of the d edges incident with $\otimes$ at Ω_v. We then label gradually edges adjacent to at least one already-labeled edge. Thus, the image of one of the endpoints of the current edge under the core-graph morphism is already known, and there are at most $d - 1$ options to label the current edge. Overall, the number of possible labelings is bounded by $|V(\Omega)| \cdot d(d-1)^{\delta t - 1}$.

- The third and last stage, where we estimate $\nu_t(J)$ for a particular J, is almost identical. The only difference is that every vertex in $\Gamma_X(J)$ is of degree at most $\min\{2m, d\}$, so overall we obtain $\nu_t(J) \leq (4t^2)^{3m-1} \cdot (\min\{2m, d\} - 1)^{(1-2\delta)t}$.

We conclude as in the proof of Proposition 4.3.

$\square$

4.3 An arbitrary base-graph Ω

We now return to the most general case of an arbitrary connected base graph Ω. Theorem 4.11 below is needed for proving the bound on the new spectrum of the adjacency operator on Γ, the random covering of Ω in the $\mathcal{C}_{n,\Omega}$ model (the first part of Theorem 1.4). The small variation needed for the second part of this theorem, dealing with the Markov operator, is discussed in Section 7.1.

Recall that T denotes the universal covering of Ω (and of Γ), and $\rho = \rho_A(\Omega)$ denotes the spectral radius of its adjacency operator. Recall also that we denote $k = |E(\Omega)|$ and orient each of the k edges arbitrarily and label them by $x_1, \ldots, x_k$. With the orientation and labeling of its edges, Ω becomes a non-pointed X-labeled graph, where $X = \{x_1, \ldots, x_k\}$. Every walk in Ω of length t can be regarded as an element of $(X \cup X^{-1})^t$ and (after reduction) of $\mathbf{F}_k = \mathbf{F}(X)$. We also denoted $\mathrm{rk}(\Omega) = |E(\Omega)| - |V(\Omega)| + 1$ and showed that $\pi(w) \in \{0, 1, \ldots, \mathrm{rk}(\Omega), \infty\}$ for every $w \in \mathcal{CW}_t(\Omega)$ (Lemma 4.8). The main theorem of this subsection is the following:

Theorem 4.11. *Let Ω be a finite, connected graph, and let $m \in \{1, \ldots, \mathrm{rk}(\Omega)\}$. Then*

$$\limsup_{t \to \infty} \left[\sum_{w \in \mathcal{CW}_t^m(\Omega)} |\mathrm{Crit}(w)| \right]^{1/t} \leq (2m - 1) \cdot \rho.$$

Before proceeding to the proof of this theorem, let us refer to the case $m = 0$ which is left out. These are words reducing to 1, and the trivial element of $\mathbf{F}_k$ has exactly one critical subgroup, so $\sum_{w \in \mathcal{CW}_t^m(\Omega)} |\mathrm{Crit}(w)|$ equals $|\mathcal{CW}_t^0(\Omega)|$.

Claim 4.12.
$$\limsup_{t \to \infty} \left| \mathcal{CW}_t^0 (\Omega) \right|^{1/t} = \rho.$$

Proof. For a given vertex $v \in V(\Omega)$, each cycle at v of length t reducing to 1 lifts to a cycle in T at $\widehat{v}$, where $\widehat{v} \in p^{-1}(v)$ is some vertex at the fiber above v of the covering map $p : T \to \Omega$. The number of cycles of length t reducing to 1 at v is thus $[A_T^t \delta_{\widehat{v}}]_{\widehat{v}}$, and

$$\left[A_T^t \delta_{\widehat{v}} \right]_{\widehat{v}} = \left\langle A_T^t \delta_{\widehat{v}}, \delta_{\widehat{v}} \right\rangle_1 \leq \left\| A_T^t \right\| \cdot \left\| \delta_{\widehat{v}} \right\|^2 = \left\| A_T^t \right\| = \rho^t$$

(the last equality follows from A_T being self-adjoint), and thus

$$\limsup_{t \to \infty} \left| \mathcal{CW}_t^0 (\Omega) \right|^{1/t} \leq \limsup_{t \to \infty} \left[|V(\Omega)| \cdot \rho^t \right]^{1/t} = \rho.$$

To show there is actual equality, repeat the argument from Claim 4.7. $\qquad\square$

We return to the proof of Theorem 4.11. By Claim 3.1,

$$\sum_{w \in \mathcal{CW}_t^m (\Omega)} |\mathrm{Crit}(w)| \quad = \quad \sum_{\substack{N \leq \mathbf{F}_k: \\ \mathrm{rk}(N) = m}} |\{ w \in \mathcal{CW}_t (\Omega) \,|\, N \in \mathrm{Crit}(w) \}|$$

$$\leq \quad \sum_{\substack{N \leq \mathbf{F}_k: \\ \mathrm{rk}(N) = m}} |\{ w \in \mathcal{CW}_t (\Omega) \,|\, \langle w \rangle \lesssim_{alg} N \}| \qquad (4.7)$$

and we actually bound the latter summation. For every $N \leq \mathbf{F}_k$, we let $\beta_t(N)$ denote the corre- $\beta_t(N)$
sponding summand, namely

$$\beta_t(N) = |\{ w \in \mathcal{CW}_t (\Omega) \,|\, \langle w \rangle \lesssim_{alg} N \}|.$$

Note that while a non-reduced element $w \in \mathcal{CW}_t (\Omega)$ with $w \in N$ might not correspond to a close walk in $\Gamma_X(N)$, it always does correspond to a close walk at the basepoint of the Schreier coset graph $\overline{\Gamma}_X(N)$.

If $N \leq \mathbf{F}_k$ satisfies that the basepoint $\otimes$ of $\Gamma_X(N)$ is not a leaf, call N and its core-graph CR   CR
(cyclically reduced). The following claim shows it is enough to consider CR subgroups.

Claim 4.13. *If $N \leq \mathbf{F}_k$ is CR then*

$$\sum_{N' \text{ is conjugate to N}} \beta_t(N') \leq t \beta_t(N).$$

Proof. The Schreier graphs of N and of any conjugate of it differ only by the basepoint. If N' is some conjugate of N and $w' \in \mathcal{CW}_t (\Omega)$ satisfies $\langle w' \rangle \lesssim_{alg} N'$, then the walk corresponding to w' in the Schreier graph $\overline{\Gamma}_X(N')$ must visit all vertices and edges of the core of $\overline{\Gamma}_X(N')$, and in particular the basepoint of $\overline{\Gamma}_X(N)$ (by Lemma 4.1). So there is some cyclic rotation w of w' satisfying $\langle w \rangle \lesssim_{alg} N$ (clearly, w also belongs to $\mathcal{CW}_t (\Omega)$). On the other hand, each such w has at most t possible cyclic rotations, each of which corresponds to one w' and one N'. $\qquad\square$

Next, we classify the subgroups $N \leq \mathbf{F}_k$ according to their "topological" core graph Λ. As implied in the short discussion preceding Claim 4.2, this is the homeomorphism class of the pointed $\Gamma_X(N)$. Namely, this is the graph obtained from $\Gamma_X(N)$ by ignoring vertices of degree two, except for (possibly) the basepoint. As Claim 4.13 allows us to restrict to one CR representative from each conjugacy class of subgroups in $\mathbf{F}_k$, we also restrict attention to one CR representative Λ of each "conjugacy class" of topological core graphs. Ignoring the basepoints, any Λ' in the "conjugacy class" of Λ retracts to this representative. For example, we need exactly three such representatives in rank 2, as shown in Figure 4.1.

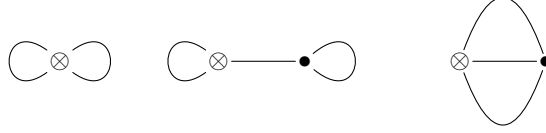The following proposition is the key step in the proof of Theorem 4.11.

Figure 4.1: The three CR representatives of topological graphs of rank 2: Figure-Eight, Barbell and Theta.

Proposition 4.14. *Let Λ be a pointed finite connected graph without vertices of degree 1 or 2 except for possibly the basepoint, and let δ denote its maximal degree. Then the sum of $\beta_t(N)$ over all subgroup $N \leq \mathbf{F}_k$ whose core graph is topologically Λ is at most*

$$|V(\Omega)| \cdot \left(4t^4\right)^{|E(\Lambda)|} \cdot (\delta - 1)^t \cdot \rho^t.$$

Proof. Denote $r = |E(\Lambda)|$. Order and orient the edges of Λ $\{e_1, e_2, \ldots, e_r\}$ so that e_1 emanates from $\otimes$, and for every $i \geq 2$, e_i emanates either from $\otimes$ or from a vertex which is the beginning or endpoint of one of $e_1, \ldots, e_{i-1}$. (This labeling and orientation is usually not unique, but we fix one throughout this proof.) In addition, let v_0 denote $\otimes$ and v_i denote the endpoint of e_i for $1 \leq i \leq r$. For example, one can label the barbell-shaped graph as follows: $\left(\overbrace{e_3}\right) \otimes \xrightarrow{e_1} \bullet \left(\overbrace{e_2}\right)$, where $v_0 = v_3$ are $\otimes$ and $v_1 = v_2$ are $\bullet$. Also, denote by $\mathrm{beg}(i)$ the smallest index j such that e_i begins at v_j, so e_i is a directed edge from $v_{\mathrm{beg}(i)}$ to v_i and $\mathrm{beg}(i) < i$. In our example, $\mathrm{beg}(1) = \mathrm{beg}(3) = 0$ and $\mathrm{beg}(2) = 1$.

Note that each N corresponding to Λ is determined by the walks (words in $\mathbf{F}_k$) associated with $e_1, \ldots, e_r$. From Claim 4.9 it follows one can restrict to subgroups N which are subgroups of J_v for some $v \in V(\Omega)$. So fix some $v_0 \in V(\Omega)$ and also some $\widehat{v}_0 \in V(T)$ which projects to v_0. We claim that every subgroup $N \leq J_{v_0}$ corresponding to Λ is completely determined by a set of vertices $\widehat{v}_1, \ldots, \widehat{v}_r$ in T: the topological edge in $\Gamma_X(N)$ associated with e_i corresponds to the walk in T from $\widehat{v}_{\mathrm{beg}(i)}$ to $\widehat{v}_i$. (There are some constraints on the choices of the $\widehat{v}_i$'s. For example, if $v_i = v_j$ then $\widehat{v}_i$ and $\widehat{v}_j$ must belong to the same fiber of the projection map $p: T \to \Omega$. However, as we only bound from above, we ignore these constraints.) So instead of summing over all possible N's, we go through all possible choices of vertices $\widehat{v}_1, \ldots, \widehat{v}_r$ in T.

The counting argument that follows resembles the one in Proposition 4.3. Fix a particular $N \leq J_{v_0}$ corresponding to Λ and let $\widehat{v}_1, \ldots, \widehat{v}_r$ be the corresponding vertices in T. By Lemma 4.1, if $w \in \left(X \cup X^{-1}\right)^t$ satisfies $\langle w \rangle \lesssim_{alg} N$, then its reduced form traverses every topological edge of $\Gamma_X(N)$ at least twice. For each i, assume that w first traverses the topological edge associated with e_i starting at position $\tau_{i,1}$ (the position is in w, namely $0 \leq \tau_{i,1} \leq t-1$), and in $\ell_{i,1}$ steps, and then from position $\tau_{i,2}$ in $\ell_{i,2}$ steps (recall that w is not reduced so $\ell_{i,2}$ may be different from $\ell_{i,1}$). The directions of these traverses are $\varepsilon_{i,1}, \varepsilon_{i,2} \in \{\pm 1\}$. In total, there are less than t^{2r} options for the $\tau_{i,j}$'s, less than t^{2r} options for the $\ell_{i,j}$'s and less than 2^{2r} options for the $\varepsilon_{i,j}$'s: a total of less than $\left(4t^4\right)^r$ options. There are $t - \ell_{1,1} - \ell_{1,2} - \ldots - \ell_{r,1} - \ell_{r,2}$ remaining steps, and these are divided to at most $4r$ segments (we can always assume one of the $\tau_{i,1}$'s equals 0). Denote the lengths of these segments by $q_1, \ldots, q_{4r}$ (some may be 0). The i'th segment reduces to some walk in $\Gamma_X(N)$, with at most $(\delta - 1)^{q_i}$ possibilities (recall that δ marks the maximal degree of a vertex in Λ). Overall, there are at most $(\delta - 1)^{q_1 + \ldots + q_{4r}} \leq (\delta - 1)^t$ options to choose the reduced walks traced by these $4r$ segments in w. Given such a reduced walk for the i'th segment, let $\widehat{x}_i, \widehat{y}_i \in V(T)$ be suitable vertices in the tree such that the reduced walk lifts to the unique reduced walk from $\widehat{x}_i$ to $\widehat{y}_i$.

Now, we sum over all subgroups N corresponding to Λ and all words $w \in \mathcal{CW}_t(\Omega)$ with $\langle w \rangle \lesssim_{alg} N$. By adding a factor of $|V(\Omega)| \left(4t^4\right)^r \cdot (\delta - 1)^t$ we assume we already know v_0 and $\widehat{v}_0$, the $\tau_{i,j}$'s, $\ell_{i,j}$'s, $\varepsilon_{i,j}$'s, the q_i's and the reduced $4r$ walks. Moreover, conditioning on knowing $\widehat{v}_1, \ldots, \widehat{v}_r$, we also know the $\widehat{x}_i$'s and the $\widehat{y}_i$'s. Recall that $c_\Gamma(t, u, v)$ denotes the number of walks of length t in a graph Γ from the vertex u to the vertex v, and that by (4.6), $c_T(t, u, v) \leq \rho^t$ for every $u, v \in V(T)$. For each $i = 1, \ldots, r$ and $j = 1, 2$, there are $c_T\left(\ell_{i,j}, \widehat{v}_{\mathrm{beg}(i)}, \widehat{v}_i\right)$ possible subwords corresponding to the j'th traverse of e_i (even if $\varepsilon_{i,j} = -1$, because $c_T\left(\ell_{i,j}, \widehat{v}_{\mathrm{beg}(i)}, \widehat{v}_i\right) = c_T\left(\ell_{i,j}, \widehat{v}_i, \widehat{v}_{\mathrm{beg}(i)}\right)$). Similarly, there are at most $c_T(q_i, \widehat{x}_i, \widehat{y}_i)$ subwords corresponding to the the i'th intermediate segment. Thus,

if $\alpha = |V(\Omega)| \cdot \left(4t^4\right)^r \cdot (\delta - 1)^t$ then

$$
\sum_{\substack{N \leq \mathbf{F}_k: \\ \Gamma_X(N) \cong \Lambda}} \beta_t(N) \quad \leq \quad \alpha \cdot \sum_{\widehat{v}_1, \ldots, \widehat{v}_r \in V(T)} \left[\prod_{i=1}^{r} \prod_{j=1}^{2} c_T\left(\ell_{i,j}, \widehat{v}_{\text{beg}(i)}, \widehat{v}_i\right) \right] \prod_{i=1}^{4r} c_T\left(q_i, \widehat{x}_i, \widehat{y}_i\right)
$$

$$
\leq \quad \alpha \cdot \left[\prod_{i=1}^{4r} \rho^{q_i} \right] \sum_{\widehat{v}_1, \ldots, \widehat{v}_r \in V(T)} \left[\prod_{i=1}^{r} \prod_{j=1}^{2} c_T\left(\ell_{i,j}, \widehat{v}_{\text{beg}(i)}, \widehat{v}_i\right) \right]
$$

Note that $\text{beg}(i) < i$, so $c_T\left(\ell_{i,j}, \widehat{v}_{\text{beg}(i)}, \widehat{v}_i\right)$ only depends on $\ell_{i,j}$ and $\widehat{v}_0, \ldots, \widehat{v}_i$ (and not on $\widehat{v}_{i+1}, \ldots, \widehat{v}_r$). Therefore, if we write $f(i) = \prod_{j=1}^{2} c_T\left(\ell_{i,j}, \widehat{v}_{\text{beg}(i)}, \widehat{v}_i\right)$, we can split the sum to obtain:

$$
\sum_{\substack{N \leq \mathbf{F}_k: \\ \Gamma_X(N) \cong \Lambda}} \beta_t(N) \quad \leq \quad \alpha \cdot \rho^{\sum q_i} \sum_{\widehat{v}_1 \in V(T)} f(1) \left[\sum_{\widehat{v}_2 \in V(T)} f(2) [\ldots] \right]
$$

The following step is the crux of the matter. We use the fact that each topological edge is traversed twice to get rid of the summation over vertices in T. We begin with the last edge e_r, where we replace the expression $\sum_{\widehat{v}_r \in V(T)} f(r)$ as follows:

$$
\sum_{\widehat{v}_r \in V(T)} f(r) \quad = \quad \sum_{\widehat{v}_r \in V(T)} c_T\left(\ell_{r,1}, \widehat{v}_{\text{beg}(r)}, \widehat{v}_r\right) c_T\left(\ell_{r,2}, \widehat{v}_{\text{beg}(r)}, \widehat{v}_r\right)
$$

$$
= \quad \sum_{\widehat{v}_r \in V(T)} c_T\left(\ell_{r,1}, \widehat{v}_{\text{beg}(r)}, \widehat{v}_r\right) c_T\left(\ell_{r,2}, \widehat{v}_r, \widehat{v}_{\text{beg}(r)}\right)
$$

$$
\overset{(*)}{=} \quad c_T\left(\ell_{r,1} + \ell_{r,2}, \widehat{v}_{\text{beg}(r)}, \widehat{v}_{\text{beg}(r)}\right) \leq \rho^{\ell_{r,1} + \ell_{r,2}}.
$$

The crucial step here is the equality $\overset{(*)}{=}$. It follows from the fact that $\widehat{v}_r$ can be recovered as the vertex of T visited by the walk of length $\ell_{r,1} + \ell_{r,2}$ after $\ell_{r,1}$ steps. After "peeling" the expression $\sum_{\widehat{v}_r \in V(T)} f(r)$, we can go on and bound $\sum_{\widehat{v}_{r-1} \in V(T)} f(r-1)$ by $\rho^{\ell_{r-1,1} + \ell_{r-1,2}}$ and so on. Eventually, we obtain

$$
\sum_{\substack{N \leq \mathbf{F}_k: \\ \Gamma_X(N) \cong \Lambda}} \beta_t(N) \quad \leq \quad \alpha \cdot \rho^{\sum q_i} \prod_{i=1}^{r} \rho^{\ell_{i,1} + \ell_{i,2}} = |V(\Omega)| \cdot \left(4t^4\right)^r \cdot (\delta - 1)^t \cdot \rho^t.
$$

$\square$

Finally, we are in position to establish the upper bounds stated in Theorem 4.11. Fix $m \in \{1, 2, \ldots, \text{rk}(\Omega)\}$. Then by (4.7) and Claim 4.13,

$$
\sum_{w \in \mathcal{CW}_t^m(\Omega)} |\text{Crit}(w)| \quad \leq \quad \sum_{\substack{N \leq \mathbf{F}_k: \\ \text{rk}(\bar{N}) = m}} \beta_t(N)
$$

$$
\leq \quad \sum_{\substack{[N] \in \text{ConjCls}(\mathbf{F}_k, m) \\ N \text{ is CR}}} t\beta_t(N) \tag{4.8}
$$

where the final summation is over all conjugacy classes of subgroups of rank m in $\mathbf{F}_k$, and for each class N is a CR representative. Moreover, we choose these representatives N so that if $[N_1]$ and

$[N_2]$ correspond the same non-pointed topological graph, the representatives N_1 and N_2 correspond to the same *pointed* topological graph Λ.

Finally, split the summation of the CR representatives N by their topological graph Λ. By Claim 4.2, each such Λ has maximal degree at most $2m$ and at most $3m-1$ edges, so by Proposition 4.14, the N's corresponding to each Λ contribute to the summation in (4.8) at most

$$t \cdot |V(\Omega)| \cdot \left(4t^4\right)^{3m-1} \cdot (2m-1)^t \cdot \rho^t.$$

This finishes the proof of Theorem 4.11 as there is a finite number of topological graphs Λ of rank m. $\quad\square$

5 Controlling the Error Term of $\mathbb{E}\left[\mathcal{F}_{w,n}\right]$

In this section we establish the third step of the proofs of Theorems 1.1, 1.4, and 1.5, as introduced in the overview of the proof (Section 2). Recall that according to Theorem 2.3, for every $w \in \mathbf{F}_k$ the following holds:

$$\mathbb{E}\left[\mathcal{F}_{w,n}\right] = 1 + \frac{|\operatorname{Crit}(w)|}{n^{\pi(w)-1}} + O\left(\frac{1}{n^{\pi(w)}}\right).$$

But the $O(\cdot)$ term depends on w. Our goal here is to obtain a bound on the $O(\cdot)$ term, which depends solely on the length of w and $\pi(w)$, namely a bound which is uniform on all words of a certain length and primitivity rank. This is done in the following proposition:

Proposition 5.1. *Let $w \in \left(X \cup X^{-1}\right)^t$ satisfy $\pi(w) \neq 0$ (so w does not reduce to 1). If $n > t^2$ then*

$$\mathbb{E}\left[\mathcal{F}_{w,n}\right] \leq 1 + \frac{1}{n^{\pi(w)-1}}\left(|\operatorname{Crit}(w)| + \frac{t^{2+2\pi(w)}}{n-t^2}\right).$$

Achieving such a bound requires more elaborated details from the proof of Theorem 2.3, which appears in [PP15]. We therefore begin with recalling relevant concepts and results from [PP15]. We then present the proof of Proposition 5.1 in Section 5.5.

Before that, let us mention that the same statement holds for words in $\left(X \cup X^{-1}\right)^t$ that reduce to 1:

Claim 5.2. *Let $w \in \left(X \cup X^{-1}\right)^t$ satisfy $\pi(w) = 0$ (so w reduces to 1). If $n > t^2$ then*

$$\mathbb{E}\left[\mathcal{F}_{w,n}\right] \leq 1 + \frac{1}{n^{\pi(w)-1}}\left(|\operatorname{Crit}(w)| + \frac{t^{2+2\pi(w)}}{n-t^2}\right).$$

Proof. Recall that $\pi(w) = 0$ if and only if $w = 1$ as an element of $\mathbf{F}_k$. But then the only w-critical subgroup is the trivial one, and so $\mathbb{E}\left[\mathcal{F}_{w,n}\right] = n = 1 + \frac{1}{n^{-1}}\left(|\operatorname{Crit}(w)| - \frac{1}{n}\right)$ which is indeed less than the bound in the statement. $\quad\square$

5.1 The partial order "covers"

In Section 3.2 morphisms of core graphs were discussed. Recall that a morphism $\Gamma_X(H) \to \Gamma_X(J)$ exists (and is unique) if and only if $H \leq J$ (Claim 3.3). A special role is played by *surjective* morphisms of core graphs:

Definition 5.3. Let $H \leq J \leq \mathbf{F}_k$. Whenever the morphism $\eta_{H \to J}^X : \Gamma_X(H) \to \Gamma_X(J)$ is surjective, we say that $\Gamma_X(H)$ *covers* $\Gamma_X(J)$ or that $\Gamma_X(J)$ *is a quotient of* $\Gamma_X(H)$. As for the groups, we say that H *X-covers* J and denote this by $H \leq_{\vec{x}} J$. $\hfill H \leq_{\vec{x}} J$

By "surjective" we mean surjective on both vertices and edges. Note that we use the term "covers" even though in general this is *not* a topological covering map (a morphism between core graphs is always locally injective at the vertices, but it need not be locally bijective). In contrast, the random graphs in $\mathcal{C}_{n,H}$ are topological covering maps, and we reserve the term "coverings" for these.

For instance, $H = \langle x_1 x_2 x_1^{-3}, x_1^2 x_2 x_1^{-2} \rangle \leq \mathbf{F}_k$ X-covers the group $J = \langle x_2, x_1^2, x_1 x_2 x_1 \rangle$, the corresponding core graphs of which are the leftmost and rightmost graphs in Figure 5.1. As another example, a core graph Γ X-covers $\Gamma_X(\mathbf{F}_k)$ (which is merely a wedge of k loops) if and only if it contains edges of all k labels.

As implied by the notation, the relation $H \leq_{\vec{x}} J$ indeed depends on the given basis X of $\mathbf{F}_k$. For example, if $H = \langle x_1 x_2 \rangle$ then $H \leq_{\vec{x}} \mathbf{F}_2$. However, for $Y = \{x_1 x_2, x_2\}$, H does not Y-cover $\mathbf{F}_2$, as $\Gamma_Y(H)$ consists of a single vertex and a single loop and has no quotients apart from itself.

It is easy to see that the relation "$\leq_{\vec{x}}$" indeed constitutes a partial ordering of the set of subgroups of $\mathbf{F}_k$. In fact, restricted to f.g. subgroups it becomes a locally-finite partial order, which means that if $H \leq_{\vec{x}} J$ then the interval of intermediate subgroups $[H, J]_{\vec{x}} = \{ M \leq \mathbf{F}_k \mid H \leq_{\vec{x}} M \leq_{\vec{x}} J \}$ is finite:

Claim 5.4. *If $H \leq \mathbf{F}_k$ is a f.g. subgroup then it X-covers only a finite number of groups. In particular, the partial order "$\leq_{\vec{x}}$" restricted to f.g. subgroups of $\mathbf{F}_k$ is locally finite.*

Proof. The claim follows from the fact that $\Gamma_X(H)$ is finite (Claim 3.2(1)) and thus has only finitely many quotients. Each quotient corresponds to a single group, by (3.2). $\qquad \square$

5.2 Partitions and quotients

It is easy to see that a quotient $\Gamma_X(J)$ of $\Gamma_X(H)$ is determined by the partition it induces on the vertex set $V(\Gamma_X(H))$ (the vertex-fibers of the morphism $\eta_{H \to J}^X$). However, not every partition P of $V(\Gamma_X(H))$ corresponds to a quotient core-graph. Indeed, Δ, the graph we obtain after merging the vertices grouped together in P, might not be a core-graph: two distinct j-edges may have the same origin or the same terminus. (For a combinatorial description of core-graphs see e.g. [Pud14, Claim 2.1].) Then again, when a partition P of $V(\Gamma_X(H))$ yields a quotient which is not a core-graph, we can perform Stallings foldings[†] until we obtain a core graph. We denote the resulting core-graph by[‡] $\Gamma_X(H)/P$. Since Stallings foldings do not affect π_1^X, this core graph $\Gamma_X(H)/P$ is $\Gamma_X(J)$, where $J = \pi_1^X(\Delta)$. The resulting partition $\bar{P}$ of $V(\Gamma_X(H))$ (the blocks of which are the fibers of $\eta_{H \to J}^X$) is the finest partition of $V(\Gamma_X(H))$ which gives a quotient core-graph and which is still coarser than P. We illustrate this in Figure 5.1. $\qquad \Gamma_X(H)/P$

One can think of $\Gamma_X(J) = \Gamma_X(H)/P$ as the core graph "generated" from $\Gamma_X(H)$ by the partition P. It is now natural to look for the "simplest" partition generating $\Gamma_X(J)$. Formally, we introduce a measure for the complexity of partitions: if $P \subseteq 2^{\mathcal{X}}$ is a partition of some set $\mathcal{X}$, let

$$\|P\| \overset{def}{=} |\mathcal{X}| - |P| = \sum_{B \in P} (|B| - 1). \qquad (5.1)$$

Namely, $\|P\|$ is the number of elements in the set minus the number of blocks in the partition. For example, $\|P\| = 1$ iff P identifies only a single pair of elements. It is not hard to see that $\|P\|$ is also the minimal number of identifications one needs to make in $\mathcal{X}$ in order to obtain the equivalence relation P. Restricting to pairs of subgroups H, J with $H \leq_{\vec{x}} J$, we can define the following distance function:

Definition 5.5. Let $H, J \leq_{fg} \mathbf{F}_k$ be subgroups such that $H \leq_{\vec{x}} J$, and let $\Gamma = \Gamma_X(H)$, $\Delta = \Gamma_X(J)$ be the corresponding core graphs. We define the *X-distance* between H and J, denoted $\rho_X(H, J)$ $\qquad \rho_X(H, J)$

[†]A folding means merging two equally-labeled edges with the same origin or with the same terminus. See also Figure 5.1. For a fuller description of Stallings foldings we refer the reader to [Pud14, PP15].

[‡]In [PP15], the notation $\Gamma_X(H)/P$ was used to denote something a bit different (the unfolded graph Δ).
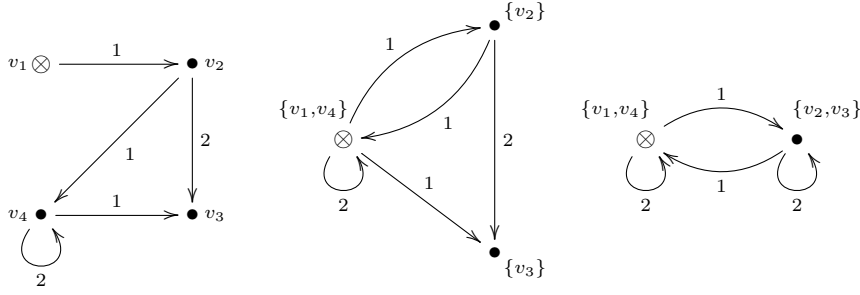
Figure 5.1: The left graph is the core graph $\Gamma_X(H)$ of $H = \langle x_1 x_2 x_1^{-3}, x_1^2 x_2 x_1^{-2} \rangle \leq \mathbf{F}_2$. Its vertices are denoted by $v_1, \ldots, v_4$. The graph in the middle is the quotient corresponding to the partition $P = \{\{v_1, v_4\}, \{v_2\}, \{v_3\}\}$. This is not a core graph as there are two 1-edges originating at $\{v_1, v_4\}$. In order to obtain a core quotient-graph, we use the Stallings folding process and identify these two 1-edges and their termini. The resulting core graph, $\Gamma_X(H)/P$, is shown on the right and corresponds to the partition $\bar{P} = \{\{v_1, v_4\}, \{v_2, v_3\}\}$.

or $\rho(\Gamma, \Delta)$ as

$$\rho_X(H, J) = \min \left\{ \|P\| \, \middle| \, \begin{array}{l} P \text{ is a partition of } V(\Gamma_X(H)) \\ \text{s.t. } \Gamma_X(H)/P = \Gamma_X(J) \end{array} \right\}. \tag{5.2}$$

For example, the rightmost core graph in Figure 5.1 is a quotient of the leftmost one, and the distance between them is 1. For a more geometric description of this distance function, as well as more details and further examples, we refer the readers to [Pud14, PP15].

Of course, the distance function $\rho_X(H, J)$ is computable. It turns out that it can also be used to determine whether H is a free factor of J:

Theorem 5.6. *[[Pud14],Theorem 1.1 and Lemma 3.3] Let $H, J \leq_{fg} \mathbf{F}_k$ such that $H \leq_{\vec{x}} J$. Then*

$$rk(J) - rk(H) \quad \leq \quad \rho_X(H, J) \quad \leq \quad rk(J).$$

Most importantly, the minimum is obtained (namely, $\mathrm{rk}(J) - \mathrm{rk}(H) = \rho_X(H, J)$) if and only if H is a free factor of J.

This theorem is used, in particular, in the proof in [PP15] of Theorem 2.3.

So far the partitions considered here were partitions of the vertex set $V(\Gamma_X(H))$. However, it is also possible to identify (merge) different *edges* in $\Gamma_X(H)$, as long as they share the same label, and then, as before, perform the folding process to obtain a valid core graph. Moreover, it is possible to consider several partitions $P_1, \ldots, P_r$, each one *either* of the vertices *or* of the edges of $\Gamma_X(H)$, identify vertices and edges according to these partitions and then fold. We denote the resulting core graph by $\Gamma_X(H)/\langle P_1, \ldots, P_r \rangle$. It is easy to see that one $\Gamma_X(H)/\langle P_1, \ldots, P_r \rangle$ can incorporate this more involved definition into the definition of the distance function $\rho_X(H, J)$, because, for instance, identifying two edges has the same effect as identifying their origins (or termini). In fact, the following holds:

$$\rho_X(H, J) = \min \left\{ \|P_1\| + \ldots + \|P_r\| \, \middle| \, \begin{array}{l} P_i : \text{ a partition of } V(\Gamma_X(H)) \text{ or of } E(\Gamma_X(H)) \\ \text{s.t. } \Gamma_X(H)/\langle P_1, \ldots, P_r \rangle = \Gamma_X(J) \end{array} \right\}. \tag{5.3}$$

5.3   From random elements of S_n to random subgroups

Recall that Theorem 2.3 estimates $\mathbb{E}[\mathcal{F}_{w,n}]$, the expected number of fixed points of $w(\sigma_1, \ldots, \sigma_k)$, where $\sigma_1, \ldots, \sigma_k \in S_n$ are chosen independently at random in uniform distribution. The first step in its proof consists of a *generalization of the problem to subgroups*:

For every f.g. subgroups $H \leq J \leq \mathbf{F}_k$, let $\alpha_{J,S_n} : J \to S_n$ be a random homomorphism chosen at uniform distribution (there are exactly $|S_n|^{\mathrm{rk}(J)}$ such homomorphisms). Then $\alpha_{J,S_n}(H)$ is a random subgroup of S_n, and we count the number of common fixed points of this subgroup, namely the number of elements in $\{1, \ldots, n\}$ fixed by all permutations in $\alpha_{J,S_n}(H)$. Formally, we define $\quad \Phi_{H,J}$

$$\Phi_{H,J}(n) \stackrel{\text{def}}{=} \mathbb{E}\left|\substack{\text{common} \\ \text{fixed-points}}(\alpha_{J,S_n}(H))\right|.$$

This indeed generalizes $\mathbb{E}[\mathcal{F}_{w,n}]$ for

$$\mathbb{E}[\mathcal{F}_{w,n}] = \Phi_{\langle w \rangle, \mathbf{F}_k}(n). \tag{5.4}$$

5.4 Möbius inversions

The theory of Möbius inversions applies to every poset (partially ordered set) with a *locally-finite* order (recall that an order $\preceq$ is locally-finite if for every x, y with $x \preceq y$, the interval $[x,y]_{\preceq} \stackrel{def}{=} \{z \mid x \preceq z \preceq y\}$ is finite). Here we skip the general definition and define these inversions directly in the special case of interest (for a more general point of view see [PP15]).

In our case, the poset in consideration is $\mathfrak{sub}_{fg}(\mathbf{F}_k) = \{H \leq \mathbf{F}_k \mid H \text{ is f.g.}\}$, and the partial order is $\leq_{\vec{x}}$, which is indeed locally-finite (Claim 5.4). We define three derivations of the function Φ defined in Section 5.3: the left one (L), the right one (R) and the two-sided one (C). These are usually formally defined by convolution of Φ with the Möbius function of $\mathfrak{sub}_{fg}(\mathbf{F}_k)_{\leq_{\vec{x}}}$ (see [PP15]) but here we define them in an equivalent simpler way: these are the functions satisfying, for every $H \leq_{\vec{x}} J$,
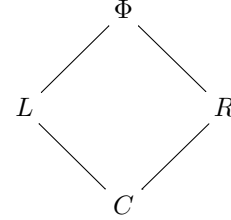
$$\Phi_{H,J}(n) = \sum_{M \in [H,J]_{\vec{x}}} L_{M,J}(n) = \sum_{M,N:\, H \leq_{\vec{x}} M \leq_{\vec{x}} N \leq_{\vec{x}} J} C_{M,N}(n) = \sum_{N \in [H,J]_{\vec{x}}} R_{H,N}(n). \tag{5.5}$$

Note that the summations in (5.5) are well defined because the order is locally finite. To see that (5.5) can indeed serve as the definition for the three new functions, use induction on $|[H,J]|$: for example, for any $H \leq_{\vec{x}} J$, $L_{H,J}(n) = \Phi_{H,J}(n) - \sum_{M \in [H,J]_{\vec{x}}} L_{M,J}(n)$ and all pairs (M, J) on the r.h.s. satisfy $|[M,J]| < |[H,J]|$.

With all this defined, we can state the main propositions along the proof of the main result in [PP15].

Proposition 5.7 ([PP15], Proposition 5.1). *The function R is supported on algebraic extensions.*

Namely, if J is not an algebraic extension of H, then $R_{H,J}(n) = 0$ for every n. Since, if $H \leq_{\mathrm{alg}} J$ then $H \leq_{\vec{x}} J$ (e.g. [PP15, Claim 4.2]), we obtain that

$$\Phi_{H,J}(n) = \sum_{N:\, H \leq_{alg} N \leq J} R_{H,N}(n). \tag{5.6}$$

Next, $\Phi_{H,J}(n)$ is given a geometric interpretation: it turns out it equals the expected number of lifts of $\eta_{H \to J} : \Gamma_X(H) \to \Gamma_X(J)$ to a random n-covering of $\Gamma_X(J)$ in the model $\mathcal{C}_{n,\Gamma_X(J)}$ [PP15, Lemma 6.2]. Similarly, $L_{H,J}(n)$ counts the average number of *injective* lifts [PP15, Lemma 6.3]. For given H and J, it is not hard to come up with an exact rational expression in n for the expected number of injective lifts, i.e. of $L_{H,J}(n)$, for large enough n (in fact, $n \geq |E(\Gamma_X(H))|$ suffices, see [PP15, Lemma 6.4]) . As the other three functions $(\Phi, R$ and $C)$ are obtained via addition and subtraction of a finite number of $L_{M,J}(n)$'s, we obtain

Claim 5.8. *Let $H, J \le \mathbf{F}_k$ be f.g. subgroups such that $H \le_{\vec{x}} J$. Then for $n \ge |E(\Gamma_X(H))|$, the functions $\Phi_{H,J}(n)$, $L_{H,J}(n)$, $R_{H,J}(n)$ and $C_{H,J}(n)$ can all be expressed as rational expressions in n.*

After some involved combinatorial arguments, one obtains from this the following expression for $C_{M,N}(n)$: Denote by $\mathrm{Sym}(S)$ the set of permutations of a given set S. Every permutation $\sigma \in \mathrm{Sym}(S)$ defines, in particular, a partition on S whose blocks are the cycles of σ. By abuse of notation we denote by σ both the permutation and the corresponding partition. For instance, one can consider its "norm" $\|\sigma\|$ (see (5.1); this is also the minimal length of a product of transpositions that gives the permutation σ). We also use V_M and E_M as short for $V(\Gamma_X(M))$ and $E(\Gamma_X(M))$, $\quad V_M, E_M$ respectively.

Proposition 5.9 ([PP15], Section 7.1). *Let $M, N \le \mathbf{F}_k$ be f.g. subgroups with $M \le_{\vec{x}} N$. Consider the set*

$$\mathcal{T}_{M,N} = \left\{ (\sigma_0, \sigma_1, \ldots, \sigma_r) \,\middle|\, \begin{array}{l} r \in \mathbb{N}, \ \sigma_0 \in \mathrm{Sym}(V_M) \\ \sigma_1, \ldots, \sigma_r \in \mathrm{Sym}(E_M) \setminus \{\mathrm{id}\} \\ \Gamma_X(M)/\langle \sigma_0, \sigma_1, \ldots, \sigma_r \rangle = \Gamma_X(N) \end{array} \right\}.$$

Then

$$C_{M,N}(n) = \frac{1}{n^{\mathrm{rk}(M)-1}} \sum_{(\sigma_0, \sigma_1, \ldots, \sigma_r) \in \mathcal{T}_{M,N}} (-1)^r \cdot \left(\frac{-1}{n}\right)^{\sum\limits_{i=0}^{r} \|\sigma_i\|}.$$

The derivation of the main result of [PP15] (Theorem 2.3) from Theorem 5.6 and Propositions 5.7 and 5.9 is short: see the beginning of Section 7 in [PP15].

5.5 Proving the uniform bound for the error term

We now have all the tools required for proving Proposition 5.1. Namely, we now prove that every $1 \ne w \in \mathbf{F}_k$ of length t and every $n > t^2$,

$$\mathbb{E}[\mathcal{F}_{w,n}] \le 1 + \frac{1}{n^{\pi(w)-1}}\left(|\mathrm{Crit}(w)| + \frac{t^{2+2\pi(w)}}{n - t^2}\right).$$

(Note that we pass here to reduced words. Reducing an element of $(X \cup X^{-1})^t$ does not affect $\mathbb{E}[\mathcal{F}_{w,n}]$, and only tightens the upper bound.)

Proof. [of Proposition 5.1] Recall (Section 5.3) that $\mathbb{E}[\mathcal{F}_{w,n}] = \Phi_{\langle w \rangle, \mathbf{F}_k}(n)$ and this quantity is given by some rational expression in n (for large enough n, say $n \ge |w|$, see Claim 5.8). This expression can be expressed as a Taylor series in $\frac{1}{n}$, so write

$$\mathbb{E}[\mathcal{F}_{w,n}] = \sum_{s=0}^{\infty} \frac{a_s(w)}{n^s}$$

where $a_s(w) \in \mathbb{R}$ (in fact these are integers: see [Pud14, Claim 5.1] and also the sequel of the current proof). By Theorem 2.3, $a_0 = 1$, $a_1 = a_2 = \ldots = a_{\pi(w)-2} = 0$ and $\alpha_{\pi(w)-1} = |\mathrm{Crit}(w)|$ (unless $\pi(w) = 1$ in which case $a_0 = 1 + |\mathrm{Crit}(w)|$). So our goal here is to bound the remaining coefficients $a_s(w)$ for $s \ge \pi(w)$.

The discussion in Section 5.4 yields the following equalities:

$$\begin{aligned} \mathbb{E}[\mathcal{F}_{w,n}] &= \Phi_{\langle w \rangle, \mathbf{F}_k}(n) = \sum_{N:\, \langle w \rangle \le_{alg} N \le \mathbf{F}_k} R_{\langle w \rangle, N}(n) = \\ &= \sum_{M,N:\, \langle w \rangle \le_{\vec{x}} M \le_{\vec{x}} N} C_{M,N}(n) = \sum_{M:\, \langle w \rangle \le_{\vec{x}} M} \sum_{N:\, M \le_{\vec{x}} N} C_{M,N}(n) \end{aligned}$$

28

From Proposition 5.9 we obtain that for a fixed M,

$$\sum_{N:\, M \leq_{\vec{x}} N} C_{M,N}(n) = \frac{1}{n^{\mathrm{rk}(M)-1}} \sum_{r \in \mathbb{N}} (-1)^r \sum_{\substack{\sigma_0 \in \mathrm{Sym}(V_M) \\ \sigma_1,\ldots,\sigma_r \in \mathrm{Sym}(E_M) \setminus \{id\}}} \left(\frac{-1}{n}\right)^{\|\sigma_0\| + \ldots + \|\sigma_r\|}.$$

For every $q \geq 0$ define the following set:

$$\mathcal{P}_{M,q} = \left\{ (\sigma_0,\ldots,\sigma_r) \,\middle|\, \begin{array}{c} r \in \mathbb{N},\ \sigma_0 \in \mathrm{Sym}(V_M) \\ \sigma_1,\ldots,\sigma_r \in \mathrm{Sym}(E_M) \setminus \{id\} \\ \|\sigma_0\| + \ldots + \|\sigma_r\| = q \end{array} \right\}, \tag{5.7}$$

so that

$$\sum_{N:\, M \leq_{\vec{x}} N} C_{M,N}(n) = \frac{1}{n^{\mathrm{rk}(M)-1}} \sum_{q=0}^{\infty} \frac{(-1)^q}{n^q} \sum_{(\sigma_0,\ldots,\sigma_r) \in \mathcal{P}_{M,q}} (-1)^r.$$

Hence,

$$a_s(w) = \sum_{i=1}^{s+1} \sum_{\substack{M:\, \langle w \rangle \leq_{\vec{x}} M \\ \mathrm{rk}(M)=i}} (-1)^{s-(i-1)} \sum_{(\sigma_0,\ldots,\sigma_r) \in \mathcal{P}_{M,s-(i-1)}} (-1)^r. \tag{5.8}$$

In what follows we ignore the alternating signs of the summands in (5.8) and bound $|a_s(w)|$ by

$$|a_s(w)| \leq \sum_{i=1}^{s+1} \sum_{\substack{M:\, \langle w \rangle \leq_{\vec{x}} M \\ \mathrm{rk}(M)=i}} \left| \mathcal{P}_{M,s-(i-1)} \right|. \tag{5.9}$$

Claim: For every $M \leq_{fg} \mathbf{F}_k$ with $\langle w \rangle \leq_{\vec{x}} M$, we have $|\mathcal{P}_{M,q}| \leq t^{2q}$.
Proof of Claim: Fix M and denote $b_q = |\mathcal{P}_{M,q}|$. Clearly, $b_0 = 1$, and we proceed by induction on q. Let $q \geq 1$. We split the set $\mathcal{P}_{M,q}$ by the value of σ_r. For $r = 0$ there are at most

$$|\{\sigma \in \mathrm{Sym}(V_M) \mid \|\sigma\| = q\}| \leq \binom{|V_m|}{2}^q \leq \binom{t}{2}^q \leq \frac{t^{2q}}{2^q}$$

elements with $r = 0$. (For the middle inequality note that $|V_M| \leq |V_{\langle w \rangle}| \leq t$; this is also the case with the edges: $|E_M| \leq |E_{\langle w \rangle}| \leq t$.) For $r \geq 1$, σ_r is a permutation of the set of edges E_M and given σ_r, the number of options for $\sigma_0,\ldots,\sigma_{r-1}$ is exactly $b_{q-\|\sigma_r\|}$. By the induction hypothesis we obtain:

$$
\begin{aligned}
b_q &\leq \frac{t^{2q}}{2^q} + \sum_{\sigma_r \in \mathrm{Sym}(E_M) \setminus \{id\}} b_{q-\|\sigma_r\|} = \frac{t^{2q}}{2^q} + \sum_{\alpha=1}^{q} b_{q-\alpha} |\{\sigma \in \mathrm{Sym}(E_M) \mid \|\sigma\| = \alpha\}| \\
&\leq \frac{t^{2q}}{2^q} + \sum_{\alpha=1}^{q} t^{2q-2\alpha} \frac{t^{2\alpha}}{2^\alpha} = t^{2q}. \quad \square
\end{aligned}
$$

We proceed with the proof of the proposition. For a given $w \in (X \cup X^{-1})^t$ there are at most $\binom{|V_{\langle w \rangle}|}{2}^\beta \leq \binom{t}{2}^\beta$ partitions of norm β of $V_{\langle w \rangle}$, and so at most $\binom{t}{2}^\beta$ subgroups M of rank β with $\langle w \rangle \leq_{\vec{x}} M$ (see Theorem 5.6). Hence from (5.9) we obtain,

$$|a_s(w)| \leq \sum_{i=1}^{s+1} \binom{t}{2}^i t^{2(s-(i-1))} \leq \sum_{i=1}^{s+1} \frac{t^{2i}}{2^i} \cdot t^{2(s-i+1)} \leq t^{2s+2}.$$

Finally,

$$\left| \mathbb{E}\left[\mathcal{F}_{w,n}\right] - 1 - \frac{|\mathrm{Crit}\,(w)|}{n^{\pi(w)-1}} \right| = \left| \sum_{s=\pi(w)}^{\infty} \frac{a_s\,(w)}{n^s} \right| \le \sum_{s=\pi(w)}^{\infty} \frac{|a_s\,(w)|}{n^s}$$

$$\le \sum_{s=\pi(w)}^{\infty} \frac{t^{2s+2}}{n^s} = t^2 \cdot \left(\frac{t^2}{n}\right)^{\pi(w)} \cdot \frac{n}{n-t^2}.$$

This finishes the proof. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\quad$ □

6 Completing the Proof for Regular Graphs

In this section we complete the proofs of Theorems 1.1 and 1.5. In addition, we explain (in Section 6.4) the source of the gap between these results on the one hand and Friedman's result and Conjecture 1.3 on the other.

6.1 Proof of Theorem 1.1 for d even

We begin with the case of even d in Theorem 1.1. We show that a random d-regular graph Γ on n vertices in the permutation model (a random n-covering of the bouquet with $\frac{d}{2}$ loops) satisfies a.a.s. $\lambda\,(\Gamma) < 2\sqrt{d-1} + 0.84$, where $\lambda\,(\Gamma)$ is the largest non-trivial eigenvalue of A_Γ. As explained in more details in Appendix A, this yields the same result for a uniformly random d-regular simple graph.

So let $d = 2k$ and $n, t = t\,(n)$ be such that $n > t^2$ and t is even. The base graph Ω is the bouquet with k loops, so $\mathcal{CW}_t\,(\Omega) = \left(X \cup X^{-1}\right)^t$. By (2.2), Proposition 5.1 and Claim 5.2,

$$\mathbb{E}\left[\lambda\,(\Gamma)^t\right] \le \sum_{w \in (X \cup X^{-1})^t} \left(\mathbb{E}\left[\mathcal{F}_{w,n}\right] - 1\right) =$$

$$= \sum_{m=0}^{k} \sum_{\substack{w \in (X \cup X^{-1})^t: \\ \pi(w)=m}} \left(\frac{|\mathrm{Crit}\,(w)|}{n^{m-1}} + O\left(\frac{1}{n^m}\right)\right)$$

$$\le \sum_{m=0}^{k} \frac{1}{n^{m-1}} \sum_{\substack{w \in (X \cup X^{-1})^t: \\ \pi(w)=m}} \left(|\mathrm{Crit}\,(w)| + \frac{t^{2+2m}}{n-t^2}\right)$$

$$\le \left(1 + \frac{t^{2+2k}}{n-t^2}\right) \sum_{m=0}^{k} \frac{1}{n^{m-1}} \sum_{\substack{w \in (X \cup X^{-1})^t: \\ \pi(w)=m}} |\mathrm{Crit}\,(w)|$$

Let $\varepsilon > 0$. For $m \in \{0, 1, \dots, k\}$, Corollary 4.5 (for $m \ge 1$) and Claim 4.7 (for $m = 0$) yield that for large enough t,

$$\sum_{\substack{w \in (X \cup X^{-1})^t: \\ \pi(w)=m}} |\mathrm{Crit}\,(w)| \le \left[g\,(2m-1) + \varepsilon\right]^t,$$

where $g\,(\cdot)$ is defined as in (4.4) with an extended domain:

$$g\,(2m-1) = \begin{cases} 2\sqrt{d-1} & 2m-1 \in \left[-1, \sqrt{d-1}\right] \\ 2m-1 + \frac{d-1}{2m-1} & 2m-1 \in \left[\sqrt{d-1}, d-1\right] \end{cases}.$$

30

Thus

$$
\begin{aligned}
\mathbb{E}\left[\lambda\left(\Gamma\right)^{t}\right] &\leq \left(1+\frac{t^{2+2k}}{n-t^{2}}\right)\sum_{m=0}^{k}\frac{\left[g\left(2m-1\right)+\varepsilon\right]^{t}}{n^{m-1}} \\
&\leq \left(1+\frac{t^{2+2k}}{n-t^{2}}\right)\cdot\left(k+1\right)\cdot \\
&\qquad \cdot\left[\max\left\{\begin{array}{c} n^{1/t}\left[g\left(-1\right)+\varepsilon\right], g\left(1\right)+\varepsilon, \frac{g(3)+\varepsilon}{n^{1/t}}\cdots \\ \cdots, \frac{g(2k-3)+\varepsilon}{\left(n^{1/t}\right)^{k-2}}, \frac{2k+\varepsilon}{\left(n^{1/t}\right)^{k-1}} \end{array}\right\}\right]^{t}
\end{aligned}
\tag{6.1}
$$

Recall that Γ is a random graph on n vertices. In order to obtain the best bound, t needs to be chosen to minimize the maximal summand in the r.h.s. of (6.1). This requires $t = \theta\left(\log n\right)$: if t is larger than that, the last elements are unbounded, and if t is smaller than that, the first element is unbounded. Thus, in particular, $\left(1+\frac{t^{2+2k}}{n-t^{2}}\right) = 1+o_{n}\left(1\right)$. We show that for every d there is some constant $c = c\left(d\right)$, such that if t is chosen so that $n^{1/t} \approx c$, then all $k+1$ elements in the set in the r.h.s. of (6.1) are strictly less than $2\sqrt{d-1}+0.835$ (for small enough ε). Thus, for large enough t, $\mathbb{E}\left[\lambda\left(\Gamma\right)^{t}\right] \leq \left[2\sqrt{d-1}+0.835\right]^{t}$. A standard application of Markov's inequality then shows that $\mathrm{Prob}\left[\lambda\left(\Gamma\right) < 2\sqrt{d-1}+0.84\right] \underset{n\to\infty}{\to} 1$.

Indeed, for $d \geq 26$, one can set $n^{1/t} = e^{\frac{2}{5\sqrt{d-1}}}$. Simple analysis shows that for $d \geq 26$, $e^{\frac{2}{5\sqrt{d-1}}} < 1+\frac{5}{12\sqrt{d-1}}$, so the element corresponding to $m=0$ is at most $2\sqrt{d-1}\cdot e^{\frac{2}{5\sqrt{d-1}}} < 2\sqrt{d-1}\left(1+\frac{5}{12\sqrt{d-1}}\right) = 2\sqrt{d-1}+\frac{5}{6} < 2\sqrt{d-1}+0.835$. This first element is clearly larger than all other elements corresponding to m such that $2m-1 \leq \sqrt{d-1}$. Among all other values of m, the maximal element is obtained when $2m-1 \approx 4.55\sqrt{d-1}$, but its value is bounded from above by $1.94\sqrt{d-1}+0.4$ (again, by simple analysis). For all remaining $d's$ $(4, 6, \ldots, 24)$, it can be checked case by case that choosing $n^{1/t}$ so that $n^{1/t}\cdot 2\sqrt{d-1} = 2\sqrt{d-1}+0.8$ works (and see the table in Section 6.2). $\quad\square$

6.2   From even d to odd d

In this subsection we derive the statement of Theorem 1.1 for d odd from the now established statement for d even. We showed that for d even we have a.a.s. $\lambda\left(\Gamma\right) < 2\sqrt{d-1}+0.84$. The idea is that every upper bound applying to some value of d also applies to $d-1$.

As explained in Appendix A, by contiguity results from [GJKW02], it is enough to show the $2\sqrt{d-1}+1$ upper bound for random graphs Γ in a random model denoted $\mathcal{G}^{*}_{n,d}$ (the result for random simple graphs then follows immediately).

Claim 6.1. *Let $d \geq 3$ be odd. Assume that a random $(d+1)$-regular graph Γ in the permutation model satisfies a.a.s. $\lambda\left(\Gamma\right) < C$. Then a random d-regular graph Γ in $\mathcal{G}^{*}_{n,d}$ also satisfies a.a.s. $\lambda\left(\Gamma\right) < C$.*

Proof. Let Γ be a random d-regular graph in $\mathcal{G}^{*}_{n,d}$. By ([GJKW02, Theorem 1.3], the permutation model $\mathcal{P}_{n,d+1}$ is contiguous to the distribution on $(d+1)$-regular graphs obtained by considering Γ and adding a uniformly random perfect matching m. (As d is odd, the number of vertices n in Γ is necessarily even.) Denote by $\hat{\Gamma}$ the random graph obtained this way. It is enough to show that $\lambda\left(\hat{\Gamma}\right) \geq \lambda\left(\Gamma\right) - o_{n}\left(1\right)$ with probability tending to 1 as $n \to \infty$.

Indeed, let μ be the eigenvalue of Γ whose absolute value is largest (so $\lambda\left(\Gamma\right) = |\mu|$), and let $f \in \ell^{2}\left(V\left(\Gamma\right)\right)$ be a corresponding real eigenfunction with $\|f\| = 1$. In particular, $\sum_{v\in V(\Gamma)} f\left(v\right) = 0$

and $\sum_{v\in V(\Gamma)} f(v)^2 = 1$. We have

$$\lambda\left(\hat{\Gamma}\right) \geq \langle A_{\hat{\Gamma}} f, f\rangle = \langle A_\Gamma f, f\rangle + 2\sum_{e\in m} f\left(e^+\right) f\left(e^-\right) = \mu + 2\sum_{e\in m} f\left(e^+\right) f\left(e^-\right),$$

where the summation is over all edges e in the random perfect matching m, and e^+ and e^- mark the two endpoints of e. Let R denote the random summation $2\sum_{e\in m} f\left(e^+\right) f\left(e^-\right)$. We finish by showing that R is generally very small.

To accomplish that we use standard identities involving symmetric polynomials over $f(v_1),\ldots,f(v_n)$. Let $p_k = \sum_v f(v)^k$ be the k'th symmetric Newton polynomial, so $p_1 = 0$ and $p_2 = 1$. Moreover, since $|f(v)| < 1$ for every v, $|p_k| < p_2 = 1$. We use the fact that every symmetric polynomial is a polynomial in the p_k's and is thus bounded.

To begin with,

$$\mathbb{E}[R] = n\cdot\frac{1}{\binom{n}{2}} \sum_{\{u,v\}\in\binom{V}{2}} f(v) f(u) = \frac{2}{n-1} s_2\left(f(v_1),\ldots,f(v_n)\right),$$

where s_2 is the second elementary symmetric function: $s_2(x_1,\ldots,x_n) = \sum_{i<j} x_i x_j$. Since $s_2 = \frac{1}{2}\left(p_1^2 - p_2\right) = -\frac{1}{2}$, we conclude that $\mathbb{E}[R] = -\frac{1}{n-1} = o_n(1)$.

Similarly,

$$\begin{aligned}
\mathbb{E}\left[R^2\right] &= 4\cdot\frac{n}{2}\cdot\frac{1}{\binom{n}{2}} \sum_{\{u,v\}\in\binom{V}{2}} f(v)^2 f(u)^2 + 8\cdot\binom{n/2}{2}\cdot\frac{1}{\binom{n}{4}} \sum_{\{u,v,w,x\}\in\binom{V}{4}} f(u) f(v) f(w) f(x), \\
&= \frac{4}{n-1} \sum_{\{u,v\}\in\binom{V}{2}} f(v)^2 f(u)^2 + \frac{48}{(n-1)(n-3)} \sum_{\{u,v,w,x\}\in\binom{V}{4}} f(u) f(v) f(w) f(x).
\end{aligned}$$

Since the two summations here are symmetric polynomials, they are bounded, and thus $\mathbb{E}\left[R^2\right] = o_n(1)$ and so is the variance of R. Thus $R = o_n(1)$ with probability tending to 1 as $n\to\infty$. $\qquad\square$

If $d\geq 3$ is odd, we can thus use our bound for $d+1$ to obtain that a.a.s.

$$\lambda(\Gamma) < 2\sqrt{(d+1)-1} + 0.84 = 2\sqrt{d} + 0.84 \approx 2\sqrt{d-1} + \frac{1}{\sqrt{d}} + 0.84.$$

This proves our result for large enough d. Indeed, for $d\geq 41$, $2\sqrt{d} + 0.84 < 2\sqrt{d-1} + 1$.

For smaller values of odd d we use tighter results for $d+1$. For example, we seek the smallest constant c for which a bound of $2\sqrt{4-1} + c$ can be obtained for 4-regular graphs in our methods. In order to minimize $\max\left\{n^{1/t}\cdot 2\sqrt{d-1}, 2\sqrt{d-1}, \frac{4}{n^{1/t}}\right\}$ (see (6.1)), we choose $n^{1/t} = \sqrt{\frac{4}{2\sqrt{d-1}}}$ to get an upper bound of 3.723 (compared with $2\sqrt{d-1} = 3.464$, so here $c\approx 0.259$). For $d = 3$ this bound is useless (it is larger than the trivial bound of 3).

The following table summarizes the bounds we obtain for $d\leq 20$ in the scenario of Theorem 1. This can be carried on to establish Theorem 1.1 for $d\leq 40$.

d	Upper Bound	c in $2\sqrt{d-1}+c$	$n^{1/t}$	d	Uppder Bound	c in $2\sqrt{d-1}+c$
4	3.723	0.259	1.075	3	3	0.172
6	4.933	0.460	1.103 $\Longrightarrow$ 5	4.933	0.933
8	5.868	0.576	1.109 $\Longrightarrow$ 7	5.868	0.969
10	6.646	0.646	1.108 $\Longrightarrow$ 9	6.646	0.989
12	7.323	0.689	1.104 $\Longrightarrow$ 11	7.323	0.998
14	7.928	0.7169	1.099 $\Longrightarrow$ 13	7.928	0.9998
16	8.482	0.7352	1.095 $\Longrightarrow$ 15	8.482	0.999
18	8.994	0.747	1.091 $\Longrightarrow$ 17	8.994	0.994
20	9.473	0.755	1.087 $\Longrightarrow$ 19	9.473	0.988

Remark 6.2. Of course, the method presented here to derive the statement of Theorem 1.1 for odd d's from the statement for even d's works only because of the small additive constant we have in the result. To obtain a tight result (Friedman's Theorem) in our approach, we will need another method to work with odd d's.

One plausible direction is as follows. We may construct a random d-regular graph with d odd using $k = \frac{d-1}{2}$ random permutations plus one random perfect matching. If we label the edges corresponding to the perfect matching by b, and orient the edges corresponding to the permutations and label them by $a_1, \ldots, a_k$, the graphs become Schreier graphs of subgroups of $\mathbf{F}_k * \mathbb{Z}/2\mathbb{Z} = \langle a_1, \ldots, a_k, b \,|\, b^2 = 1 \rangle$. It is conceivable that the machinery we developed for the free group (and especially, Theorem 2.3) can be also developed for this kind of free products.

6.3 Proof of Theorem 1.5

The only change upon the previous case (Theorem 1.1 with d even) is that the summation in (6.1) over the primitivity rank m does not stop at $k = \frac{d}{2}$ but continues until $\mathrm{rk}\,(\Omega) = |V(\Omega)| \left(\frac{d}{2} - 1\right) + 1$. However, when $m > k$, it follows from Corollary 4.10 that the corresponding term inside the max operator is $\frac{d}{\left(n^{1/t}\right)^{m-1}}$ which is strictly less than $\frac{d}{\left(n^{1/t}\right)^{d/2-1}}$ (for every choice of t and n), but this latter term is already there in (6.1). Thus, the maximal term is remained unchanged, and we obtain the same bound overall as in the even case of Theorem 1.1, namely $2\sqrt{d-1} + 0.84$.

Let us stress that in this case the proof as is works for all $d \geq 3$ (odd and even alike). As before, for small d's we can obtain better bounds, even if d is odd. For example, for $d = 3$ one can obtain an upper bound of $\sqrt{3 \cdot 2\sqrt{d-1}} \approx 2.913$.

6.4 The source of the gap

It could be desirable to use the approach presented in this paper and replace the constant 1 in Theorem 1.1 with an arbitrary $\varepsilon > 0$, to obtain Friedman's tight result. Unfortunately, this is still beyond our reach. It is possible, however, to point out the source of the gap and how it may be potentially overcome.

In the first inequality in our proof (as outlined in Section 2), we bound $\left\{\mathbb{E}\left[\lambda\,(\Gamma)^t\right]\right\}^{1/t}$ by $\left\{\mathbb{E}\left[\sum_{\mu \in \mathrm{Spec}(A_\Gamma) \setminus \{d\}} \mu^t\right]\right\}^{1/t}$. Since

$$\lambda\,(\Gamma)^t \leq \sum_{\mu \in \mathrm{Spec}(A_\Gamma) \setminus \{d\}} \mu^t \leq n \cdot \lambda\,(\Gamma)^t = \left[n^{1/t} \cdot \lambda\,(\Gamma)\right]^t,$$

as long as $t = \theta\,(\log n)$ the loss here is bounded, and if $t \gg \log n$ we lose nothing.

On the other hand, if $t \ll \log n$, one cannot obtain anything: It is known (e.g. [GZ99, Corollary 1]) that for every $\delta > 0$ there exists $0 < \varepsilon < 1$ such that at least $\varepsilon \cdot n$ of the eigenvalues of Γ satisfy $|\mu| \geq \rho - \delta$ (here $\rho = 2\sqrt{d-1}$). If $t \in o\,(\log n)$ then $n^{1/t}$ tends to infinity, and thus
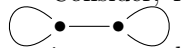
$$\left\{\sum_{\mu \in \mathrm{Spec}(A_\Gamma) \setminus \{d\}} \mu^t\right\}^{1/t} > \left\{\varepsilon n\,(\rho - \delta)^t\right\}^{1/t} \underset{n \to \infty}{\to} \infty.$$
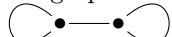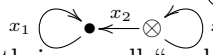
Our proof proceeds by bounding this t-th moment of the non-trivial spectrum. Let us stress that as long as $t = t\,(n)$ is small enough in terms of n so that the error term in Proposition 5.1 is negligible ($t = o\left(n^{1/(2+2k)}\right)$ suffices), the upper bound our technique yields for $\mathbb{E}\left[\sum_{\mu \in \mathrm{Spec}(A_\Gamma) \setminus \{d\}} \mu^t\right]$ is tight. In particular, for large enough d, and $t \approx c \log n$ with a suitable constant $c = c\,(d)$,

$$\left\{\mathbb{E}\left[\sum_{\mu \in \mathrm{Spec}(A_\Gamma) \setminus \{d\}} \mu^t\right]\right\}^{1/t} \approx 2\sqrt{d-1} + 0.84.$$

To see why, note that all relevant steps of the proof yield equalities or tight bounds: the second step, which relies on Theorem 2.3, has only equalities so it is surely tight. In the third step, we prove that the error term is $o_n(1)$ for every w of length t (note the proof bounds the absolute value of the error term). As mentioned above, the bound we have in the fourth step for the exponential growth rate of $\sum_{w \in (X \cup X^{-1})^t:\, \pi(w)=m} |\mathrm{Crit}(w)|$ is, in fact, the correct value (see Theorem 8.5). In the final, fifth step we may tighten our calculation in order to come closer to the real constant (slightly smaller than 0.84), but we cannot improve it considerably.

What is, then, the source of this gap? It seems, therefore, that the reason the bound we get for $\lambda(\Gamma)$ is not tight lies in rare events that enlarge $\mathbb{E}\left[\lambda(\Gamma)^t\right]$ substaintially. For example, in the permutation model every vertex of Γ is isolated with probability $\frac{1}{n^k}$, so overall there are on average $\frac{1}{n^{k-1}}$ isolated vertices. Each such vertex is responsible to an additional eigenvalue d, alongside the trivial one. These rare events alone contribute $\frac{1}{n^{k-1}} \cdot d^t$ to $\mathbb{E}\left[\lambda(\Gamma)^t\right]$. For example, for $d=4$ ($k=2$) and $n^{1/t} \approx 1.075$ as in the table in Section 6.2, isolated vertices contribute about $\left[\frac{4}{(1.075)}\right]^t \approx 3.721^t$ to $\mathbb{E}\left[\lambda(\Gamma)^t\right]$, which is roughly the bound we obtain in this case.

There are other, slightly more complicated, rare events that contribute much to $\mathbb{E}\left[\lambda(\Gamma)^t\right]$. Consider, for instance, the event that when $d=4$ the random graph Γ contains the subgraph . If this subgraph is completed to a 4-regular graph by attaching a tree to each vertex, its spectral radius becomes 3.5. Since this resulting graph topologically covers (the connected component of the subgraph in) Γ, we get a non-trivial eigenvalue which is at least 3.5 (but normally very close to 3.5). On average, there are $\frac{2}{n}$ such subgraphs in Γ, so they contribute about $\frac{2}{n} \cdot 3.5^t$ to $\mathbb{E}\left[\lambda(\Gamma)^t\right]$. When $n^{1/t}$ is small enough, this is strictly larger than $\left[2\sqrt{d-1}\right]^t \approx 3.464^t$.

Each such small graph corresponds to a few particular subgroups of $\mathbf{F}_k$. For example, the subgraph corresponds to one of four subgroups, one of which is $x_1 \bigcirc \bullet \xleftarrow{x_2} \otimes \bigcirc x_1$. One therefore needs to realize which all these "bad" subgroups are, show their overall "probability" is small (the average number of appearances of $H \le \mathbf{F}_k$ in Γ is exactly $L_{H,\mathbf{F}_k}$), and somehow omit their contribution to $\mathbb{E}\left[\lambda(\Gamma)^t\right]$. This would be relatively easy were our analysis of $\mathbb{E}[\mathcal{F}_w]$ based on $\mathbb{E}[\mathcal{F}_w] = \sum_{M \in [\langle w \rangle, \mathbf{F}_k]_{\overrightarrow{X}}} L_{M,\mathbf{F}_k}$. However, it is based, instead, on $\mathbb{E}[\mathcal{F}_w] = \sum_{N \in [\langle w \rangle, \mathbf{F}_k]_{\overrightarrow{X}}} R_{\langle w \rangle, N}$ (see Section 5). It seems that overcoming this difficulty requires a better control over the error term: this might enable us to omit the contribution of these "bad" subgroups from our bounds.

Remark 6.3. These "bad", rare events are somewhat parallel to the notion of *tangles* in [Fri08].

7 Completing the Proof for Arbitrary Graphs

The completion of the proof of Theorem 1.4 is presented in this Section. We begin with the proof of the first statement of the theorem which concerns the spectrum of the adjacency operator of Γ, the random n-covering of the fixed base graph Ω. The variations needed in order to establish the statement about the Markov operator are described in Section 7.1.

Recall that $\rho = \rho_A(\Omega)$ denotes the spectral radius of the adjacency operator of the covering tree. Our goal now is to prove that for every $\varepsilon > 0$, $\lambda_A(\Gamma)$, the largest absolute value of a non-trivial eigenvalue of the adjacency operator A_Γ, satisfies asymptotically almost surely

$$\lambda_A(\Gamma) < \sqrt{3} \cdot \rho + \varepsilon. \tag{7.1}$$

As in the proof of Theorem 1.1 (the beginning of Section 6), let $n, t = t(n)$ be so that $n > t^2$ and t

is even. Using (2.2), Proposition 5.1, Claim 5.2 and Lemma 4.8, one obtains

$$
\mathbb{E}\left[\lambda_A\left(\Gamma\right)^t\right] \leq \sum_{w\in\mathcal{CW}_t(\Omega)}\left(\mathbb{E}\left[\mathcal{F}_w\right]-1\right) =
$$

$$
\leq \left(1+\frac{t^{2+2\mathrm{rk}(\Omega)}}{n-t^2}\right)\sum_{m=0}^{\mathrm{rk}(\Omega)}\frac{1}{n^{m-1}}\sum_{w\in\mathcal{CW}_t^m(\Omega)}\left|\mathrm{Crit}\left(w\right)\right|
$$

Let $\varepsilon > 0$. From Theorem 4.11 and Lemma 4.12 it follows now that for t even and large enough,

$$
\mathbb{E}\left[\lambda_A\left(\Gamma\right)^t\right] \leq \left(1+\frac{t^{2+2\mathrm{rk}(\Omega)}}{n-t^2}\right)\left[n\cdot[\rho+\varepsilon]^t+\sum_{m=1}^{\mathrm{rk}(\Omega)}\frac{[(2m-1)\cdot\rho+\varepsilon]^t}{n^{m-1}}\right].
$$

$$
\leq \left(1+\frac{t^{2+2\mathrm{rk}(\Omega)}}{n-t^2}\right)\left(1+\mathrm{rk}\left(\Omega\right)\right)\cdot
$$

$$
\cdot\left[\max\left\{n^{1/t}\left[\rho+\varepsilon\right],\rho+\varepsilon,\frac{3\rho+\varepsilon}{n^{1/t}},\frac{5\rho+\varepsilon}{\left(n^{1/t}\right)^2},\ldots,\frac{\left(2\mathrm{rk}\left(\Omega\right)-1\right)\rho+\varepsilon}{\left(n^{1/t}\right)^{\mathrm{rk}(\Omega)-1}}\right\}\right]^t \quad (7.2)
$$

Again, to obtain a bound we must have $t\in\theta\left(\log n\right)$, and the best bound we can obtain in this general case is obtained by choosing $n^{1/t}\approx\sqrt{3}$, so $\left(1+\frac{t^{2+2\mathrm{rk}(\Omega)}}{n-t^2}\right)^{1/t}\underset{n\to\infty}{\to}1$, and the maximal value inside the set in (7.2) is then $\sqrt{3}\left(\rho+\epsilon\right)$. Again, a standard application of Markov inequality finishes the proof. $\square$

7.1 The spectrum of the Markov operator

After establishing the first statement of Theorem 1.4, we want to explain how the proof should be modified to apply to $\lambda_M\left(\Gamma\right)$, the maximal absolute value of a non-trivial eigenvalue of the *Markov* operator on Γ. The goal is to show that for every $\varepsilon > 0$

$$
\lambda_M\left(\Gamma\right)<\sqrt{3}\cdot\rho_M\left(\Omega\right)+\varepsilon \quad (7.3)
$$

asymptotically almost surely.

As we note in Appendix B, the Markov operator is given by $B_\Gamma D_\Gamma^{-1}$, where B_Γ is the adjacency matrix and D_Γ the diagonal matrix with the degrees of vertices in the diagonal. This is conjugate to and thus share the same spectrum with $Q_\Gamma=D_\Gamma^{-1/2}B_\Gamma D_\Gamma^{-1/2}$, but the latter has the advantage of being symmetric, so we work with it.

The (u,v) entry of Q_Γ equals $\frac{1}{\sqrt{\deg(u)\deg(v)}}$ times the number of edges between u and v. For every walk w in Γ we assign a weight function $f\left(w\right)$ as follows: if w starts at v_0, then visits $v_1,v_2,\ldots,v_{t-1}$ and ends at v_t, then $f\left(w\right)=\frac{1}{\sqrt{\deg v_0}\cdot\deg v_1\cdot\ldots\cdot\deg v_{t-1}\cdot\sqrt{\deg v_t}}$. It is easy to see that $[Q_\Gamma^t]_{u,v}$ equals the sum of $f\left(w\right)$ over all walks w of length t from u to v, and thus

$$
\sum_{\lambda\in\mathrm{Spec}(M_\Gamma)}\lambda^t=\mathrm{tr}M_\Gamma=\sum_{w\in\mathcal{CW}_t(\Gamma)}f\left(w\right).
$$

Moreover, note that when a walk from the covering Γ projects to the base graph Ω, its weight does not change. Using this fact, we can imitate step I from Section 2 to obtain, for t even,

$$
\lambda_M\left(\Gamma\right)^t \leq \sum_{\mu\in\mathrm{Spec}(M_\Gamma)}\mu^t-\sum_{\mu\in\mathrm{Spec}(M_\Omega)}\mu^t=\sum_{w\in\mathcal{CW}_t(\Gamma)}f\left(w\right)-\sum_{w\in\mathcal{CW}_t(\Omega)}f\left(w\right)=
$$

$$
=\sum_{w\in\mathcal{CW}_t(\Omega)}f\left(w\right)\left[\mathcal{F}_{w,n}\left(\sigma_1,\ldots,\sigma_k\right)-1\right].
$$

35

The second and third steps remain the same, obtaining

$$\mathbb{E}\left[\lambda_M\left(\Gamma\right)^t\right] \leq \left(1 + \frac{t^{2+2\mathrm{rk}(\Omega)}}{n - t^2}\right)^{\mathrm{rk}(\Omega)} \sum_{m=0}^{} \frac{1}{n^{m-1}} \sum_{w\in\mathcal{CW}_t^m(\Omega)} f\left(w\right)\left|\mathrm{Crit}\left(w\right)\right|.$$

The next modification needs take place in the fourth step, where instead of bounding $\sum_{w\in\mathcal{CW}_t(\Omega):\,\pi(w)=m}\left|\mathrm{Crit}\left(w\right)\right|$, one needs to bound $\sum_{w\in\mathcal{CW}_t(\Omega):\,\pi(w)=m} f\left(w\right)\left|\mathrm{Crit}\left(w\right)\right|$. But the exact same proofs work if we merely replace $\rho_A\left(\Omega\right)$ with $\rho_M\left(\Omega\right)$. Theorem 4.11 becomes

$$\limsup_{t\to\infty}\left[\sum_{w\in\mathcal{CW}_t^m(\Omega)} f\left(w\right)\left|\mathrm{Crit}\left(w\right)\right|\right]^{1/t} \leq \left(2m - 1\right)\cdot\rho_M\left(\Omega\right). \tag{7.4}$$

and likewise, Lemma 4.12 becomes

$$\limsup_{t\to\infty}\left[\sum_{w\in\mathcal{CW}_t^0(\Omega)} f\left(w\right)\right]^{1/t} = \rho_M\left(\Omega\right).$$

Similarly, the definition of $\beta_t\left(N\right)$ (preceding Claim 4.13) should be modified to

$$\beta_t\left(N\right) = \sum_{w\in\mathcal{CW}_t(\Omega):\,\langle w\rangle\lesssim_{alg}N} f\left(w\right)$$

and in the proof of Claim 4.13 one should use the fact that $f\left(w\right)$ does not change when the closed walk w is being cyclically rotated. Finally, in the proof of Proposition 4.14 we sometimes replace a walk with its inverse and use the symmetry of the operator. This is the reason for working with Q_Γ rather than with M_Γ. Also, the coefficient $\left|V\left(\Omega\right)\right|$ from the statement of the proposition needs be replaced with some constant function of the degrees of all vertices.

Because the bounds in (7.4) are exactly those in Theorem 4.11 only with $\rho_M\left(\Omega\right)$ instead of $\rho_A\left(\Omega\right)$, the final step of the proof (which appears in Section 7) also remains unchanged.

8   The Distribution of Primitivity Ranks

In this subsection we show that the upper bounds from Proposition 4.3 and Corollary 4.5 are the accurate exponential growth rates of the number of words (reduced or not) and critical subgroups with a given primitivity rank. This is not needed for the proof of the main results of this paper. However, it does show that in the proof of Theorem 1.1, the fourth step of the proof, where words and critical subgroups are counted, yields a tight bound. Thus, the origin of the gap between our result and Friedman's lies elsewhere (see Section 6.4).

First, let us recall a theorem due to the author and Wu which counts primitive words in $\mathbf{F}_k$.

Theorem 8.1. *[PW14] For every $k \geq 3$, let $p_k\left(t\right)$ denote the number of primitive words of length t in $\mathbf{F}_k$. Then,*

$$\lim_{t\to\infty} p_k\left(t\right)^{1/t} = 2k - 3.$$

For $\mathbf{F}_2$ it is known that this exponential growth rate equals $\sqrt{3}$ ([Riv04]). These results show that the portion of primitive words among all words of length t decays exponentially fast[†]. They are used in the following theorem, which states that the upper bounds from Proposition 4.3 are accurate.

Theorem 8.2. *Let $k \geq 2$ and $m \in \{1, 2, \ldots, k\}$. Let*

$$c_{k,m}(t) = |\{w \in \mathbf{F}_k \,|\, |w| = t,\, \pi(w) = m\}|.$$

Then,

$$\limsup_{t \to \infty} c_{k,m}(t)^{1/t} = \begin{cases} \sqrt{2k-1} & 2m-1 \leq \sqrt{2k-1} \\ 2m-1 & 2m-1 \geq \sqrt{2k-1} \end{cases}. \tag{8.1}$$

In fact, as the proof shows, for $m \geq 2$, we can replace the $\limsup$ with regular $\lim$, and for $m = 1$ we can replace $\limsup_{t \to \infty} c_{k,1}(t)^{1/t}$ with $\lim_{t \to \infty} c_{k,1}(2t)^{1/2t}$.

Corollary 8.3. *A generic word in $\mathbf{F}_k$ has primitivity rank k.*

Proof. [of Theorem 8.2] The r.h.s. of (8.1) is an upper bound for the $\limsup$ by Proposition 4.3. It remains to show that for every $m \in \{1, \ldots, k\}$, there is some subset of words with primitivity rank m and growth rate $\max\left(\sqrt{2k-1}, 2m-1\right)$.

Consider first the case $2m-1 > \sqrt{2k-1}$. Take any subset of the generators $S \subseteq X$ of size m and consider the subgroup $H = \mathbf{F}(S)$. Its core graph is a bouquet of m loops. The number of words of length t in H is $2m \cdot (2m-1)^{t-1}$. By Theorem 8.1, a random word in H of length t is a.a.s. non-primitive in H, so its primitivity rank is at most m. On the other hand, the exponential growth rate of all words with $\pi(w) < m$ combined is smaller than $(2m-1)$ (by Proposition 4.3). Thus, a word $w \in H$ of length t satisfies $\pi(w) = m$ a.a.s., and we are done. In particular, we proved that for such values of m,

$$\limsup_{t \to \infty} c_{k,m}(t)^{1/t} = \lim_{t \to \infty} c_{k,m}(t)^{1/t} = 2m-1.$$

Now assume that $2m-1 \leq \sqrt{2k-1}$. Consider subgroups of the form $H = \langle x_1, \ldots, x_{m-1}, u \rangle$ where u is a cyclically reduced word of length $\sim \frac{t}{2}$ such that its first and last letters are *not* one of $\{x_1^{\pm 1}, \ldots, x_{m-1}^{\pm 1}\}$. Then, $\Gamma_X(H)$ has the form of a bouquet of $m-1$ small loops of size 1 and one large loop of size $\sim \frac{t}{2}$. Now consider the word $w = w(u) = x_1^2 x_2^2 \ldots x_{m-1}^2 u^2$. Obviously, the growth rate of the number of possible u's (as a function of t) is $\sqrt{2k-1}$, hence also the growth rate of the number of different w's. It can be shown that w is *not* primitive in H, using the primitivity criterion from Theorem 5.6 ([Pud14, Thm 1.1]). (In fact, it follows from [Pud14, Lemma 6.8] that as an element of the free group H, w has primitivity rank m with H being the sole w-critical subgroup.) Thus, $\pi(w) \leq m$. In general, the primitivity rank might be strictly smaller. For example, for $m = 3$ and $u = x_3 x_1^2 x_2^2 x_3$, we have $\pi(w) = 2$ because w is not a proper power yet is not primitive in $\langle x_3, x_1^2 x_2^2 \rangle$. However, we claim that for a generic u, the primitivity rank of w is exactly m.

Indeed, if this is not the case, then there is some $\tilde{m} < m$ such that the growth rate of words $w = w(u)$ as above with $\pi(w) = \tilde{m}$ is $\sqrt{2k-1}$. By the proof of Proposition 4.3 and especially (4.2), it follows that most of these words ($w = w(u)$ with $\pi(w) = \tilde{m}$) have an algebraic extension N of rank $\tilde{m}$ such that the number of edges in $\Gamma_X(N)$ is close to $\frac{t}{2}$. (By (4.2), the total number of words of length t with an algebraic extension N of rank $\tilde{m}$ and δt edges in $\Gamma_X(N)$, for some $\delta < \frac{1}{2}$, grows strictly slower than $\sqrt{2k-1}$.) So almost all these words $w = w(u)$ trace twice every edge of some $\Gamma_X(N)$ of rank $\tilde{m}$ with roughly $\frac{t}{2}$ edges. In particular, each such $w = w(u)$ traces twice some topological edge in $\Gamma_X(N)$ of length at least $\frac{1}{2(3m-1)}t$. This implies that there is some linear-size two overlapping subwords of u or of u^{-1}. But for a generic u, the longest subword appearing twice in u or in u^{-1} has length of order $\log t$.

Since the w's we obtained are of arbitrary even length, this shows that if $2m-1 \leq \sqrt{2k-1}$, then

$$\limsup_{t \to \infty} c_{k,m}(t)^{1/t} = \lim_{t \to \infty} c_{k,m}(2t)^{1/2t} = \sqrt{2k-1}.$$

[†]That primitive words in $\mathbf{F}_k$ are negligible in this sense follows also from the earlier results [BV02], [BMS02, Thm 10.4] and [Shp05], where the exponential growth rate from Theorem 8.1 is shown to be $\leq 2k - 2 - o_k(1)$.

If, in addition, $m \geq 2$, the same argument as above works also for $w = w(u) = x_1^3 x_2^2 \dots x_{m-1}^2 u^2$ which is of arbitrary odd length. Thus, $\lim_{t \to \infty} c_{k,m}(t)^{1/t} = \sqrt{2k-1}$. $\qquad\square$

Remark 8.4. It follows from the proofs of Proposition 4.3 and Theorem 8.2 that while for $2m-1 > \sqrt{2k-1}$ the main source for words with $\pi(w) = m$ is in subgroups with core graphs of minimal size (and their conjugates), the main source for $2m-1 < \sqrt{2k-1}$ is in subgroups with core graphs of maximal size, namely of size roughly $\frac{t}{2}$.

Recall that in the proof of Theorem 1.1 we used bounds on the number of *not-necessarily-reduced* words (and their critical subgroups). Here, too, the bounds from Corollary 4.5 are accurate for every value of m:

Theorem 8.5. *Let $k \geq 2$ and $m \in \{0, 1, 2, \dots, k, \infty\}$. Let* $\qquad\qquad\qquad b_{k,m}(t)$

$$b_{k,m}(t) = \left| \left\{ w \in \left(X \cup X^{-1}\right)^t \,\middle|\, \pi(w) = m \right\} \right|.$$

Then for $m = 0$ we have
$$\lim_{\substack{t \to \infty \\ t \text{ even}}} b_{k,0}(t)^{1/t} = 2\sqrt{2k-1}.$$

For $m \in \{1, \dots, k\}$,
$$\lim_{t \to \infty} b_{k,m}(t)^{1/t} = \begin{cases} 2\sqrt{2k-1} & 2m-1 \leq \sqrt{2k-1} \\ 2m-1 + \frac{2k-1}{2m-1} & 2m-1 \geq \sqrt{2k-1} \end{cases}.$$

Finally, for $m = \infty$ we have
$$\lim_{t \to \infty} b_{k,\infty}(t)^{1/t} = 2k - 2 + \frac{2}{2k-3}.$$

This shows, in particular, that as in the case of reduced words, a generic word in $\left(X \cup X^{-1}\right)^t$ is of primitivity rank k, namely, the share of words with this property tends to 1 as $t \to \infty$. It also shows that for every m, the growth rate of the number of words with primitivity rank m is equal to the growth rate of the larger quantity of $\sum_{w \in (X \cup X^{-1})^t : \pi(w) = m} |\mathrm{Crit}(w)|$.

Proof. For $m = 0$ this is (the proof of) Claim 4.7 (evidently, there are no odd-length words reducing to 1). For $1 \leq m$ with $2m - 1 \leq \sqrt{2k-1}$ the same proof (as in Claim 4.7) can be followed as long as we present at least one even-length and one odd-length words with primitivity rank m. And indeed, as mentioned above (and see [Pud14, Lemma 6.8]), $\pi\left(x_1^2 x_2^2 \dots x_m^2\right) = \pi\left(x_1^3 x_2^2 \dots x_m^2\right) = m$. If $2m - 1 > \sqrt{2k-1}$, the statement follows from the statements on reduced words (Theorems 8.2 and 8.1) and an application of the extended cogrowth formula [?] (here a bit more elaborated results from [?], not mentioned in Theorem 4.4, are required). $\qquad\square$

The statements of the last theorem are summarized in Table 2.

9 Open Questions

We end with some open problems that suggest themselves from this paper:

- Can one obtain a better control over the error term in Theorem 2.3? This would probably require not ignoring the alternating signs in (5.8). As explained in Section 6.4, this may be the seed to closing the gap in the result of Theorem 1.1.

- Is it possible to generalize the techniques in this paper (and even more so the ones from [PP15]) to odd values of d? (See Remark 6.2).

- Can one obtain the accurate exponential growth rate of the number of not-necessarily-reduced words with a given primitivity rank in a general base graph Ω, thus improving the statements of Theorems 4.11 and 1.4? This may require some sort of clever extension of the cogrowth formula that applies to non-regular graphs (there have been a few attempts in this aim, see e.g. [Bar99, Nor04, AFH07], but see limitations in [?]).

- Several classic results from the theory of expansion in graphs were generalized lately to simplicial complexes of dimension greater than one (see e.g. [GW12, PRT12, Lub14]). In particular, a parallel of Alon-Boppana Theorem is presented in [PR12]. Is there a parallel to Alon's conjecture in this case? Can the methods of the current paper be extended to higher dimensions?

Acknowledgments

Late Remark

Slightly over a year after this manuscript was written and submitted, Friedman and Kohler wrote [FK14], where they prove an asymptotic probabilistic upper bound of $2\sqrt{d-1} + \varepsilon$ for $\lambda(\Gamma)$, where Γ is a random covering of an arbitrary d-regular base graph Ω. They improve Friedman's former techniques from [Fri08] to apply to this more general case. This bound is tight and improves on the statements from Theorem 1.5 and Corollary 1.6 in the current paper. It is claimed in [FK14] that at present, they are unable to make their techniques apply to the most general case of an arbitrary (not necessarily regular) base graph Ω.

Appendices

A   Contiguity and Related Models of Random Graphs

Random d-regular graphs

In this paper, the statement of Theorem 1.1 is first proved for the permutation model of random d-regular graphs with d even. We then derive Theorem 1.1, stated for the uniform distribution on all d-regular *simple* graphs on n vertices with d even or odd, using results of Wormald [Wor99] and Greenhill et al. [GJKW02]. These works show the *contiguity* (see footnote on page 3) of different models of random regular graphs.

In particular, they describe the following model: consider dn labeled points, with d points in each of n buckets, and take a random perfect matching of the points. Letting the buckets be vertices and each pair represent an edge, one obtains a random d-regular graph. This model is denoted $\mathcal{G}^*_{n,d}$. It is shown [GJKW02, Theorem 1.3] that $\mathcal{G}^*_{n,d}$ is contiguous to the permutation model $\mathcal{P}_{n,d}$ (for d even). If Γ is a random d-regular graph in $\mathcal{G}^*_{n,d}$, the event that Γ is a simple graph (with no loops nor multiple edges) has positive probability, bounded away from 0. Moreover, within this event, simple graphs are distributed uniformly[†]. Thus, for even values of d, Theorem 1.1 follows from the corresponding result for the permutation model. The derivation of the odd case also uses contiguity results, as explained in Section 6.2.

[†]To be precise, *vertex-labeled* simple graphs are distributed uniformly in this event. Unlabeled simple graphs have probability proportional to the order of their automorphism group. Then again, for $d \geq 3$, this group is a.a.s. trivial, so the result of Theorem 1.1 applies both to the uniform model of labeled graph and to the uniform model of unlabeled graphs.

Random d-regular bipartite graphs

As an immediate corollary from Theorem 1.5 we deduced that a random d-regular bipartite graph is "nearly Ramanujan" in the sense that besides its two trivial eigenvalues $\pm d$, all other eigenvalues are at most $2\sqrt{d-1} + 0.84$ in absolute value a.a.s. (Corollary 1.6). Our proof works in the model $\mathcal{C}_{n,\Omega}$ (here Ω is the graph with 2 vertices and d parallel edges connecting them). However, by the results of [Ben74], the probability that our graph has no multiple edges is bounded away from zero (asymptotically it is $e^{-\binom{d}{2}}$). Thus, our result applies also to the model of d random *disjoint* perfect matchings between two sets of n vertices. This model, in turn, is contiguous to the uniform model of bipartite (vertex-labeled) d-regular simple graphs (for $d \geq 3$: see [MRRW97, Section 4][†]), so our result applies in the latter model as well.

Random coverings of a fixed graph

In Theorem 1.4 we consider random n-coverings of a fixed graph Ω in the model $\mathcal{C}_{n,\Omega}$, where a uniform random permutation is generated for every edge of Ω. An equivalent model is attained if we cover some spanning tree of Ω by n disjoint copies and then choose a random permutation for every edge outside the tree (that is, the same automorphism-types of non-labeled graphs are obtained with the same distribution). In fact, picking a basepoint $\otimes \in V(\Omega)$, there is yet another description for this model: The classification of n-sheeted coverings of Ω by the action of $\pi_1(\Omega, \otimes)$ on the fiber $\{\otimes\} \times [n]$ above $\otimes$ shows that $\mathcal{C}_{n,\Omega}$ is equivalent to choosing uniformly at random an action of the free group $\pi_1(\Omega, \otimes)$ on $\{\otimes\} \times [n]$.

A different but related model uses the classification of *connected*, pointed coverings of $(\Omega, \otimes)$ by the corresponding subgroups of $\pi_1(\Omega, \otimes)$. A random n-covering is thus generated by choosing a random subgroup of index n. However, it seems that this model is contiguous to $\mathcal{C}_{n,\Omega}$ if $\mathrm{rk}(\Omega) \geq 2$ (note that the random covering Γ in $\mathcal{C}_{n,\Omega}$ is a.a.s. connected provided that $\mathrm{rk}(\Omega) \geq 2$). Indeed, the only difference is that in the new model, the probability of every connected graph Γ from $\mathcal{C}_{n,\Omega}$ is proportional to $\frac{1}{|\mathrm{Aut}(\Gamma)|}$. When $\mathrm{rk}(\Omega) \geq 2$, it seems that a.a.s. $|\mathrm{Aut}(\Gamma)| = 1$, which would show that our result applies to this model as well.

Finally, there is another natural model that comes to mind: given a periodic infinite tree, namely a tree that covers some finite graph, one can consider a random (simple) graph Γ with n vertices covered by this tree (with uniform distribution among all such graphs with n vertices, for suitable n's only). One can then analyze $\lambda(\Gamma)$, the largest absolute value of an eigenvalue besides[‡] $\mathfrak{pf}(\Gamma)$. (This generalizes the uniform model on d-regular graphs.) Occasionally, all the quotients of some given periodic tree T cover the same finite "minimal" graph Ω. Interestingly, Lubotzky and Nagnibeda [LN98] showed that there exist such T's with a minimal quotient Ω which is *not* Ramanujan (in the sense that $\lambda(\Omega)$ is strictly larger than $\rho(T)$, the spectral radius of T). Since all the quotients of T inherit the eigenvalues of Ω, their $\lambda(\cdot)$ is also bounded away from $\rho(T)$ (from above). Hence, the corresponding version of Conjecture 1.3 is false in this general setting.

B   Spectral Expansion of Non-Regular Graphs

In this section we provide some background on the theory of expansion of irregular graphs, describing how spectral expansion is related to other measurements of expansion (combinatorial expansion, random walks and mixing). This further motivates the claim that Theorem 1.4 shows that if the

---

[†]In fact, there is an explicit proof there only for $d = 3$. To derive the general case, one can show that a random $(d+1)$-regular graph is contiguous to a random d-regular bipartite graph plus one edge-disjoint random matching (following, e.g., the computations in [BM86]). We would like to thank Nick Wormald for helpful private communications surrounding this point.

[‡]Leighton showed that two finite graphs with a common covering share also some common finite covering [Lei82]. It follows that all finite quotients of the same tree share the same Perron-Frobenius eigenvalue.

base graph Ω is a good (nearly optimal) expander, then a.a.s. so are its random coverings. We would like to thank Ori Parzanchevski for his valuable assistance in writing this appendix.

The spectral expansion of a (non-regular) graph Γ on m vertices is measured by some function on its spectrum, and most commonly by the *spectral gap*: the difference between the largest eigenvalue and the second largest. As mentioned above, it is not apriori clear which operator best describes in spectral terms the properties of the graph. There are three main candidates (see, e.g. [GW12]), all of which are bounded[†], self-adjoint operators and so have real spectrum:

(1) The *adjacency operator* A_Γ on $\left(\ell^2\left(V\left(\Gamma\right)\right),1\right)$[‡]:

$$(A_\Gamma f)(v) = \sum_{w \sim v} f\left(w\right)$$

If Γ is finite this operator is represented in the standard basis by the adjacency matrix, and its spectral radius is the *Perron-Frobenius* eigenvalue $\mathfrak{pf}\left(\Gamma\right)$. The spectrum in this case is

$$\mathfrak{pf}\left(\Gamma\right) = \lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_m \geq -\mathfrak{pf}\left(\Gamma\right),$$

and the spectral gap is $\mathfrak{pf}\left(\Gamma\right) - \lambda\left(\Gamma\right)$, where $\lambda\left(\Gamma\right) = \max\left\{\lambda_2, -\lambda_n\right\}$[§]. The spectrum of A_Γ was studied in various works, for instance [Gre95, LN98, Fri03, LP10].

(2) The *averaging Markov operator* M_Γ on $\left(\ell^2\left(V\left(\Gamma\right)\right),\deg\left(\cdot\right)\right)$[¶]:

$$(M_\Gamma f)(v) = \frac{1}{\deg\left(v\right)} \sum_{w \sim v} f\left(w\right)$$

This operator is given by $D_\Gamma^{-1} A_\Gamma$, and its spectrum is contained in $[-1, 1]$. The eigenvalue 1 corresponds to locally-constant functions when Γ is finite, and in this case the spectrum is

$$1 = \mu_1 \geq \mu_2 \geq \ldots \geq \mu_m \geq -1.$$

The spectral gap is then $1 - \mu\left(\Gamma\right)$ here $\mu\left(\Gamma\right) = \max\left\{\mu_2, -\mu_m\right\}$. Up to a possible affine transformation, the spectrum of M_Γ is the same as the spectrum of the simple *random walk operator* $\left(A_\Gamma D_\Gamma^{-1}\right)$ or of one of the *normalized Laplacian* operators $(I - A_\Gamma D_\Gamma^{-1}$ or $I - D_\Gamma^{-1/2} A_\Gamma D_\Gamma^{-1/2})$. This spectrum is considered for example in [Sin93, Chu97, GZ99].

(3) The *Laplacian operator* Δ_Γ^+ on $\left(\ell^2\left(V\left(\Gamma\right)\right),1\right)$:

$$\left(\Delta_\Gamma^+ f\right)(v) = \deg\left(v\right) f\left(v\right) - \sum_{w \sim v} f\left(w\right)$$

The Laplacian equals $D_\Gamma - A_\Gamma$, where D_Γ is the diagonal operator $\left(D_\Gamma f\right)\left(v\right) = \deg\left(v\right) \cdot f\left(v\right)$. The entire spectrum is non-negative, with 0 corresponding to locally-constant functions when Γ is finite. In the finite case, the spectrum is

$$0 = \nu_1 \leq \nu_2 \leq \ldots \leq \nu_m,$$

the spectral gap being $\nu_2 - \nu_1 = \nu_2$. The Laplacian operator is studied e.g. in [AM85].

[†]All operators considered here are bounded provided that the degree of vertices in Γ is bounded. This is the case in all the graphs considered in this paper.

[‡]Here, $\left(\ell^2\left(V\left(\Gamma\right)\right),1\right)$ stands for ℓ^2-functions on the set of vertices $V\left(\Gamma\right)$ with the standard inner product: $\langle f, g \rangle = \sum_v f\left(v\right) \overline{g\left(v\right)}$; In the summation $\sum_{w \sim v}$, each vertex w is repeated with multiplicity equal to the number of edges between v and w.

[§]Occasionally, the spectral gap is taken to be $\mathfrak{pf}\left(\Gamma\right) - \lambda_2\left(\Gamma\right)$.

[¶]Here, $\left(\ell^2\left(V\left(\Gamma\right)\right),\deg\left(\cdot\right)\right)$ stands for l^2-functions on the set of vertices $V\left(\Gamma\right)$ with the inner product: $\langle f, g \rangle = \sum_v f\left(v\right) \overline{g\left(v\right)} \deg\left(v\right)$.

For a regular graph Γ, all different operators are identical up to an affine shift. However, in the general case there is no direct connection between the three different spectra. In this paper we consider the spectra of A_Γ and of M_Γ. At this point we do not know how to extend our results to the Laplacian operator Δ_Γ^+.

The spectrum of all three operators is closely related to different notions of expansion in graphs. The adjacency operator, for example, has the following version of *the expander mixing lemma*: for every two subsets $S, T \subseteq V(\Gamma)$ (not necessarily disjoint), one has

$$\left| E(S,T) - \mathfrak{pf}(\Gamma) \operatorname{vol}_{\mathfrak{pf}}(S) \operatorname{vol}_{\mathfrak{pf}}(T) \right| \leq \lambda(\Gamma) \frac{\sqrt{|S| \cdot |T|}}{m},$$

where $\operatorname{vol}_{\mathfrak{pf}}(S) = \langle \mathbf{1}_S, f_{\mathfrak{pf}}(\Gamma) \rangle$ and $f_{\mathfrak{pf}}(\Gamma)$ is the (normalized) Perron-Frobenius eigenfunction. This is particularly useful in the $\mathcal{C}_{n,\Omega}$ model since the $f_{\mathfrak{pf}}(\Gamma)$ is easily obtained from the Perron-Frobenius eigenfunction of Ω by

$$f_{\mathfrak{pf}}(\Gamma) = \frac{1}{\sqrt{n}} f_{\mathfrak{pf}}(\Omega) \circ \pi.$$

In the d-regular case, this amounts to the usual mixing lemma: $\left| E(S,T) - d\frac{|S| \cdot |T|}{m} \right| \leq \lambda(\Gamma) \sqrt{|S| \cdot |T|}$. If one takes $T = V \setminus S$, one can attain a bound on the Cheeger constant of Γ (see (B.1)).

As for the averaging Markov operator, it is standard that $\mu(\Gamma)$ controls the speed in which a random walk converges to the stationary distribution. In addition, if one defines $\deg(S)$ to denote the sum of degrees of the vertices in S, then

$$\left| E(S,T) - \frac{\deg(S)\deg(T)}{2|E(\Gamma)|} \right| \leq \mu(\Gamma) \sqrt{\deg(S)\deg(T)}.$$

Moreover, consider the *conductance* of Γ

$$\phi(\Gamma) = \min_{\substack{\emptyset \neq S \subseteq V \\ \deg(S) \leq \frac{\deg(V)}{2}}} \frac{|E(S, V \setminus S)|}{\deg(S)}.$$

Then the following version of the Cheeger inequality holds [Sin93, Lemmas 2.4, 2.6]:

$$\frac{\phi^2(\Gamma)}{2} \leq 1 - \mu_2 \leq 2\phi(\Gamma).$$

Finally, the spectrum of the Laplacian operator is related to the standard *Cheeger Constant* of Γ, defined as

$$h(\Gamma) = \min_{\substack{\emptyset \neq S \subseteq V \\ |S| \leq \frac{|V|}{2}}} \frac{|E(S, V \setminus S)|}{|S|}. \tag{B.1}$$

By the so-called "discrete Cheeger inequality" [AM85]:

$$\frac{h^2(\Gamma)}{2k} \leq \nu_2 \leq 2h(\Gamma)$$

with k being the largest degree of a vertex. In addition, one has a variation on the mixing lemma for Δ_Γ^+ as well [PRT12, Thm 1.4].

References

[ABG10]   L. Addario-Berry and S. Griffiths, *The spectrum of random lifts*, Arxiv preprint arXiv:1012.4097, 2010.

[AFH07]   O. Angel, J. Friedman, and S. Hoory, *The non-backtracking spectrum of the universal cover of a graph*, arXiv preprint arXiv:0712.0192, 2007.

[AL02]   A. Amit and N. Linial, *Random graph coverings I: General theory and graph connectivity*, Combinatorica **22** (2002), no. 1, 1–18.

[AL06]   ———, *Random lifts of graphs: edge expansion*, Combinatorics, Probability and Computing **15** (2006), no. 03, 317–332.

[ALM02]   A. Amit, N. Linial, and J. Matoušek, *Random lifts of graphs: independence and chromatic number*, Random Structures & Algorithms **20** (2002), no. 1, 1–22.

[Alo86]   N. Alon, *Eigenvalues and expanders*, Combinatorica **6** (1986), no. 2, 83–96.

[AM85]   N. Alon and V.D. Milman, *$\lambda 1$, isoperimetric inequalities for graphs, and superconcentrators*, Journal of Combinatorial Theory, Series B **38** (1985), no. 1, 73–88.

[Bar99]   L. Bartholdi, *Counting paths in graphs*, Enseign. Math., II. Sér. **45** (1999), no. 1-2, 83–131.

[Ben74]   Edward A Bender, *The asymptotic number of non-negative integer matrices with given row and column sums*, Discrete Mathematics **10** (1974), no. 2, 217–223.

[BL06]   Y. Bilu and N. Linial, *Lifts, discrepancy and nearly optimal spectral gap*, Combinatorica **26** (2006), no. 5, 495–519.

[BM86]   B. Bollobás and B.D. McKay, *The number of matchings in random regular graphs and bipartite graphs*, Journal of Combinatorial Theory, Series B **41** (1986), no. 1, 80–91.

[BMS02]   A.V. Borovik, A.G. Myasnikov, and V. Shpilrain, *Measuring sets in infinite groups*, Computational and Statistical Group Theory (Las Vegas, NV/Hoboken, NJ), Contemp. Math., American Mathematical Society, 2002, pp. 21–42.

[Bog08]   Oleg Bogopolski, *Introduction to Group Theory*, EMS Textbooks in Mathematics, European Mathematical Society, Zurich, 2008.

[BS87]   A. Broder and E. Shamir, *On the second eigenvalue of random regular graphs*, Foundations of Computer Science, 1987., 28th Annual Symposium on, IEEE, 1987, pp. 286–294.

[Bur87]   M. Burger, *Chengs inequality for graphs*, preprint, 1987.

[BV02]   J. Burillo and E. Ventura, *Counting primitive elements in free groups*, Geometriae Dedicata **93** (2002), no. 1, 143–162.

[Chu97]   F.R.K. Chung, *Spectral Graph Theory*, no. 92, American Mathematical Society, 1997.

[DJPP13]   I. Dumitriu, T. Johnson, S. Pal, and E. Paquette, *Functional limit theorems for random regular graphs*, Probability Theory and Related Fields **156** (2013), no. 3-4, 921–975.

[FK14]   Joel Friedman and David-Emmanuel Kohler, *The relativized second eigenvalue conjecture of Alon*, arXiv preprint arXiv:1403.3462v1[cs.DM] (2014).

[FKS89]   J. Friedman, J. Kahn, and E. Szemeredi, *On the second eigenvalue of random regular graphs*, Proceedings of the twenty-first annual ACM symposium on Theory of computing, ACM, 1989, pp. 587–598.

[Fri91]   J. Friedman, *On the second eigenvalue and random walks in random d-regular graphs*, Combinatorica **11** (1991), no. 4, 331–362.

[Fri03]   _____, *Relative expanders or weakly relatively Ramanujan graphs*, Duke Mathematical Journal **118** (2003), no. 1, 19–35.

[Fri08]   _____, *A proof of Alon's second eigenvalue conjecture and related problems*, vol. 195, Memoirs of the AMS, no. 910, AMS, september 2008.

[GJKW02]   C. Greenhill, S. Janson, J.H. Kim, and N.C. Wormald, *Permutation pseudographs and contiguity*, Combinatorics, Probability and Computing **11** (2002), no. 03, 273–298.

[Gre95]   Y. Greenberg, *On the spectrum of graphs and their universal coverings, (in Hebrew)*, Ph.D. thesis, Hebrew University, 1995.

[Gri77]   R.I. Grigorchuk, *Symmetric random walks on discrete groups*, Uspekhi Matematicheskikh Nauk **32** (1977), no. 6, 217–218.

[GW12]   A. Gundert and U. Wagner, *On laplacians of random complexes*, Proceedings of the 2012 symposuim on Computational Geometry, ACM, 2012, pp. 151–160.

[GZ99]   R.I. Grigorchuk and A. Zuk, *On the asymptotic spectrum of random walks on infinite families of graphs*, Random walks and discrete potential theory, Sympos. Math, vol. 39, Cambridge, 1999, pp. 188–204.

[HLW06]   S. Hoory, N. Linial, and A. Wigderson, *Expander graphs and their applications*, Bulletin of the American Mathematical Society **43** (2006), no. 4, 439–562.

[Kes59]   Harry Kesten, *Symmetric random walks on groups*, Transactions of the American Mathematical Society (1959), 336–354.

[KM02]   I. Kapovich and A. Myasnikov, *Stallings foldings and subgroups of free groups*, Journal of Algebra **248** (2002), no. 2, 608–668.

[Lei82]   Frank Thomson Leighton, *Finite common coverings of graphs*, Journal of Combinatorial Theory, Series B **33** (1982), no. 3, 231–238.

[LN98]   A. Lubotzky and T. Nagnibeda, *Not every uniform tree covers Ramanujan graphs*, Journal of Combinatorial Theory, Series B **74** (1998), no. 2, 202–212.

[LP10]   Nati Linial and Doron Puder, *Words maps and spectra of random graph lifts*, Random Structures and Algorithms **37** (2010), no. 1, 100–135.

[LPS88]   A. Lubotzky, R. Phillips, and P. Sarnak, *Ramanujan graphs*, Combinatorica **8** (1988), no. 3, 261–277.

[LR05]   N. Linial and E. Rozenman, *Random lifts of graphs: perfect matchings*, Combinatorica **25** (2005), no. 4, 407–424.

[LSV11]   E. Lubetzky, B. Sudakov, and V. Vu, *Spectra of lifted Ramanujan graphs*, Advances in Mathematics **227** (2011), no. 4, 1612–1645.

[Lub14]   Alexander Lubotzky, *Ramanujan complexes and high dimensional expanders*, Japanese Journal of Mathematics **9** (2014), no. 2, 137–169.

[Lyo12]   Y. Lyons, R. with Peres, *Probability on trees and networks*, Cambridge Univ Press, 2012, In preparation. Current version available at
http://mypage.iu.edu/~rdlyons/.

[Mar88]   G.A Margulis, *Explicit group-theoretic constructions of combinatorial schemes and their applications in the construction of expanders and concentrators*, Problemy Peredachi Informatsii **24** (1988), no. 1, 51–60.

[MNS08]   S.J. Miller, T. Novikoff, and A. Sabelli, *The distribution of the largest nontrivial eigen-values in families of random regular graphs*, Experimental Mathematics **17** (2008), no. 2, 231–244.

[Mor94]   M. Morgenstern, *Existence and explicit constructions of q+ 1 regular Ramanujan graphs for every prime power q*, Journal of Combinatorial Theory, Series B **62** (1994), no. 1, 44–62.

[MRRW97] Michael S. O. Molloy, Hanna Robalewska, Robert W. Robinson, and Nicholas C. Wormald, *1-factorizations of random regular graphs*, Random Structures and Algorithms **10** (1997), no. 3, 305–321.

[MSS13]   Adam Marcus, Daniel A Spielman, and Nikhil Srivastava, *Interlacing families I: Bipartite Ramanujan graphs of all degrees*, Foundations of Computer Science (FOCS), 2013 IEEE 54th Annual Symposium on, IEEE, 2013, pp. 529–537.

[MVW07]   A. Miasnikov, E. Ventura, and P. Weil, *Algebraic extensions in free groups*, Geometric group theory (G.N. Arzhantseva, L. Bartholdi, J. Burillo, and E. Ventura, eds.), Trends Math., Birkhauser, 2007, pp. 225–253.

[Nil91]   A. Nilli, *On the second eigenvalue of a graph*, Discrete Mathematics **91** (1991), no. 2, 207–210.

[Nor92]   S. Northshield, *Cogrowth of regular graphs*, Proc. Amer. Math. Soc, vol. 116, 1992, pp. 203–205.

[Nor04]   ———, *Cogrowth of arbitrary graphs*, Random walks and geometry (Kaimanovich V., ed.), de Gruyter, 2004, pp. 501–513.

[PP15]   Doron Puder and Ori Parzanchevski, *Measure preserving words are primitive*, Journal of the American Mathematical Society **28** (2015), no. 1, 63–97.

[PR12]   Ori Parzanchevski and Ron Rosenthal, *Simplicial complexes: spectrum, homology and random walks*, arXiv preprint arXiv:1211.6775 (2012).

[PRT12]   O. Parzanchevski, R. Rosenthal, and R.J. Tessler, *Isoperimetric inequalities in simplicial complexes*, Combinatorica (2012), To appear. Arxiv preprint arXiv:1207.0638.

[Pud14]   Doron Puder, *Primitive words, free factors and measure preservation*, Israel Journal of Mathematics **201** (2014), no. 1, 25–73.

[PW14]   Doron Puder and Conan Wu, *Growth of the primitives elements in free groups*, Journal of London Mathematical Society **90** (2014), no. 1, 89–104.

[Riv04]   I. Rivin, *A remark on 'Counting primitive elements in free groups' (by J. Burillo and E. Ventura)*, Geometriae Dedicata **107** (2004), no. 1, 99–100.

[Ser90]   J.P. Serre, *Lettre à Winnie Li, 8 Octobre*, 1990, see [GZ99].

[Shp05]   V. Shpilrain, *Counting primitive elements of a free group*, Contemporary Mathematics **372** (2005), 91–98.

[Sin93]   A. Sinclair, *Algorithms for random generation and counting: a Markov chain approach*, vol. 7, Birkhauser, 1993.

[Sta83]   John R. Stallings, *Topology of finite graphs*, Inventiones mathematicae **71** (1983), no. 3, 551–565.

[Tak51]   M. Takahasi, *Note on chain conditions in free groups*, Osaka Math. J **3** (1951), no. 2, 221–225.

[Wor99]   N.C. Wormald, *Models of random regular graphs*, London Mathematical Society Lecture Note Series (1999), 239–298.