Expansion of Random Graphs:
New Proofs, New Results

Doron Puder†

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

July 31, 2014

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 $\Gamma$ be a random $d$-regular graph on $n$ vertices, and let $\lambda$ 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 $\varepsilon > 0$, a.a.s. $\lambda < 2\sqrt{d-1} + \varepsilon$. 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 $\lambda < 2\sqrt{d-1} + 1$ asymptotically almost surely.

A main advantage of our approach is that it is applicable to a generalized conjecture: 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 $\Omega$, we study the spectrum of $\Gamma$, a random $n$-covering of $\Omega$. Let $\lambda$ be the largest absolute value of a non-trivial eigenvalue of $\Gamma$. Extending Alon’s conjecture to this more general model, Friedman [Fri03] conjectured that for every $\varepsilon > 0$, a.a.s. $\lambda < \rho + \varepsilon$, where $\rho$ is the spectral radius of the universal cover of $\Omega$. When $\Omega$ is regular we get the same bound as before: $\rho + 1$, and for an arbitrary $\Omega$, 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 Introduction
2 Overview of the Proof
3 Preliminaries: Core Graphs and Algebraic Extensions
4 Counting Words and Critical Subgroups
5 Controlling the Error Term of $\mathbb{E} [F_{w,n}]$
6 Completing the Proof for Regular Graphs

†Supported by Adams Fellowship Program of the Israel Academy of Sciences and Humanities, and the ERC.
1 Introduction

Random $d$-regular graphs

Let $\Gamma$ be a finite $d$-regular graph on $n$ vertices ($d \geq 3$) and let $A_\Gamma$ be its adjacency matrix. The spectrum of $\Gamma$ is the spectrum of $A_\Gamma$ 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 $\Gamma$. Let $\lambda(\Gamma)$ be the largest absolute value of a non-trivial eigenvalue of $\Gamma$, i.e. $\lambda(\Gamma) = \max \{ \lambda_2, -\lambda_n \}$. This value measures the spectral expansion of the graph: the smaller $\lambda(\Gamma)$ is, the better expander $\Gamma$ is (see Appendix B for details).

The well-known Alon-Boppana bound states that $\lambda(\Gamma) \geq 2\sqrt{d-1} - o_n(1)$, bounding the spectral expansion of an infinite family of $d$-regular graphs. There is no equivalent deterministic non-trivial upper bound: for example, if $\Gamma$ is disconnected or bipartite then $\lambda(\Gamma) = d$. However, Alon conjectured [Alo86, Conj. 5.1] that if $\Gamma$ 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 $A_\Gamma$ and shows that it is likely to be small on all points of an appropriate $\varepsilon$-net on the unit sphere. This approach yielded an asymptotic bound of $\lambda(\Gamma) < c\sqrt{d}$ for some unspecified constant $c$ [FKSS9]. In the recent work [DJPP11 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 paths in $\Gamma$. These works include [BS87], in which Broder and Shamir show that a.a.s. $\lambda(\Gamma) \leq \sqrt{2d^{3/4} + \varepsilon}$ ($\forall \varepsilon > 0$); [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 [PP14]. We show the following:

**Theorem 1.1.** Fix $d \geq 3$ and let $\Gamma$ 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.[]

[]^1Unless 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.

[]^2In fact, Alon’s original conjecture referred only to $\lambda_2(\Gamma)$, the second largest eigenvalue.

[]^3For small $d$’s a 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 \textit{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 \textit{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 \( \Gamma \) on the set of vertices \([n]\) is obtained by choosing independently and uniformly at random \( \frac{d}{2} \) permutations \( \sigma_1, \ldots, \sigma_d \) in the symmetric group \( S_n \), and introducing an edge \((v, \sigma_j(v))\) for every \( v \in [n] \) and \( j \in \{1, \ldots, \frac{d}{2}\} \). Of course, \( \Gamma \) may be disconnected and can have loops or multiple edges.

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 \textit{Ramanujan}, i.e. that \( \lambda(\Gamma) \leq 2\sqrt{d-1} \)? There are contradicting experimental pieces of evidence (in [MNS08] it is conjectures that this probability tends to 27% as \( n \) grows; simulations in [HJW06 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 are inside \([-2\sqrt{d-1}, 2\sqrt{d-1}]\)). 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 \([-2\sqrt{d-1}, 2\sqrt{d-1}]\). Namely, \( 2\sqrt{d-1} \) is the \textit{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 \( \Gamma \) against the spectral radius of its covering tree. Several authors call graphs whose non-trivial spectrum is bounded by this value \textit{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 \( \Omega \) 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 probability \( 1 - o_n(1) \) 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 \( \sup \{|\lambda| : \lambda \in \text{Spec } M\} \).
All other finite quotients of $T$ are then coverings of $\Omega$, 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 $\Omega$ 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.

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

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 $\Gamma$ 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 $\Omega$, and let $\Gamma$ be a random $n$-covering space of $\Omega$. More specifically, $\Gamma$ is sampled as follows: its set of vertices is $V(\Omega) \times \{1, \ldots, n\}$. A permutation $\sigma_e \in S_n$ is then chosen uniformly and independently at random for every edge $e = (u, v)$ of $\Omega$, and for every $i \in \{1, \ldots, n\}$ the edge $((u, i), (v, \sigma_e(i)))$ is introduced in $\Gamma$.

We denote this model by $\mathcal{C}_{n,\Omega}$ (so that $\mathcal{C}_{n,B^2_d} = \mathcal{P}_{n,d}$, where $B^2_d$ is the bouquet with $\frac{d}{2}$ loops). For example, all bipartite $d$-regular graphs on $2n$ vertices cover the graph 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 $\Omega$” 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 a priori clear which operator best describes in spectral terms the properties of the graph. In this paper we consider two operators: the adjacency operator $A_\Gamma$ defined as above, and the Markov operator $M_\Gamma$ defined by

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

(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 (see Appendix B for the relations of these spectra to expansion properties of $\Gamma$).

For a finite graph $\Omega$ on $m$ vertices, the spectrum of the adjacency matrix $A_\Omega$ is

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

$\text{pf}(\Omega)$ being the Perron-Frobenius eigenvalue of $A_\Gamma$. The spectrum of $M_\Omega$ is

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

the eigenvalue $1$ corresponding to the constant function. Every finite covering $\Gamma$ of $\Omega$ shares the same Perron-Frobenius eigenvalue, and moreover, inherits the entire spectrum of $\Omega$ (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 \(\Gamma, f \circ \pi\), with the same eigenvalue. Thus, every eigenvalue of \(\Omega\) (with multiplicity) is trivially an eigenvalue of \(\Gamma\) as well. We denote by \(\lambda_A(\Gamma)\) the largest absolute value of a new eigenvalue of \(A_T\), namely the largest one not inherited from \(\Omega\). Equivalently, this is the largest absolute eigenvalue of an eigenfunction of \(\Gamma\) which sums to zero on every fiber of \(\pi\). In a familiar way we define \(\lambda_M(\Gamma)\), the largest absolute value of a new eigenvalue of \(M_T\). Note that in the regular case (i.e. when \(\Omega\) is \(d\)-regular), \(A_T = d \cdot M_T\), and so \(\lambda_A(\Gamma) = d \cdot \lambda_M(\Gamma)\). Moreover, when \(\Omega = B_d\) is the bouquet, \(\lambda_A(\Gamma) = \lambda(\Gamma)\).

As in the regular case, the largest non-trivial eigenvalue is closely related to the spectral radius of \(T\), the universal covering tree of \(\Omega\) (which is also the universal covering of every covering \(\Gamma\) of \(\Omega\).

We denote by \(\rho_A(\Omega)\) and \(\rho_M(\Omega)\) the spectral radii of \(A_T\) and \(M_T\), resp. (So when \(\Omega\) 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:

**Proposition 1.2.** Let \(\Gamma\) be an \(n\)-covering of \(\Omega\). Then

1. \(\lambda_A(\Gamma) \geq \rho_A(\Omega) - o_n(1)\) [Gre92, Thm 2.11].
2. \(\lambda_M(\Gamma) \geq \rho_M(\Omega) - o_n(1)\) [Bur87, GZ99, Prop. 6].

When \(\Omega\) is \(d\)-regular (but not necessarily a bouquet), Greenberg’s result was first observed by Serre [Ser90].

As in the \(d\)-regular case, the only deterministic upper bounds are trivial: \(\lambda_A(\Gamma) \leq \rho f(\Omega)\) 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 \(\Omega\) be a finite connected graph. If \(\Gamma\) is a random \(n\)-covering of \(\Omega\), 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 \(\Gamma\) (see Appendix B), Conjecture L3 asserts that if the base graph \(\Omega\) is a good (near-optimal) expander then with high probability so is its random covering \(\Gamma\).

In the same paper ([Fri03]), Friedman generalizes the method of Broder-Shamir mentioned above and shows that \(\lambda_A(\Gamma) < \rho f(\Omega)^{1/2} \rho_A(\Omega)^{1/2} + \varepsilon\) 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 \(\lambda_A(\Gamma) < 3\rho f(\Omega)^{1/3} \rho_A(\Omega)^{2/3} + \varepsilon\) (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 \(\Omega\) 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 L3 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 $\Omega$ and, to a larger extent, in the general case of any finite base-graph.

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

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

asymptotically almost surely, and similarly

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

asymptotically almost surely.

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

**Theorem 1.5.** Let $\Omega$ be a finite connected $d$-regular graph ($d \geq 3$) and let $\Gamma$ be a random $n$-covering of $\Omega$. Then for every $\varepsilon > 0$,

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

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 $\Omega$ consisting of two vertices and $d$ edges connecting them.

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

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

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

This means that alongside the two trivial eigenvalues $\pm d$, all other eigenvalues of the bipartite graph $\Gamma$ are a.a.s. within $[-2\sqrt{d-1} - 0.84, 2\sqrt{d-1} + 0.84]$. 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 $A$.

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 Erdős-Rényi model $G(n, p)$, 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 (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 $\Omega$ & Any graph & $d$-regular & $B_{2d} = a$ bouquet of $\frac{d}{2}$ loops \\
Deterministic lower bound for $\lambda_{A}(\Gamma)$ & $\rho - \alpha_{n}(1)$ \cite{Gre95} & $\rho - \alpha_{n}(1)$ \cite{Ser90} & $\rho - \alpha_{n}(1)$ \cite{Nil91} (Alon-Boppana) \\
Conjectured probabilistic upper bound & $\rho + \varepsilon$ & $\rho + \varepsilon$ \cite{Fri03} \\
Probabilistic upper bounds, ordered by asymptotic strength for growing $\rho$ & $\sqrt{\rho}(\Omega) + \varepsilon$ \cite{Fri03} & $\sqrt{d\rho} + \varepsilon$ & $\sqrt{d\rho} + \varepsilon$ \cite{BS87} \\
& $3 \cdot \sqrt{\rho}(\Omega)^{2/3} + \varepsilon$ \cite{LP10} & $3 \cdot d^{1/3} \rho^{2/3} + \varepsilon$ \\
& $C \cdot \max(\lambda(\Omega), \rho) \log 3$ \cite{LSV11} & $\sqrt{3} \cdot \rho + \varepsilon$ \cite{Thm 1.4} \\
& $265,000 \cdot \rho$ \cite{ABG10} & $6,200 \cdot \rho$ \cite{PKS89} \cite{DJPP11} \\
& $\rho + 0.84$ \cite{Thm 1.5} & $\rho + 0.84$ \cite{Thm 1.1} \\
& $\rho + \varepsilon$ \cite{Fri91} \\

Table 1: Our results compared with former ones. As above, $\Omega$ 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 $\Omega$ be a fixed base graph with $k$ edges and $\Gamma$ a random $n$-covering in the model $C_{n,\Omega}$. In the spirit of the trace method, the spectrum of $\Gamma$ is analyzed by counting closed paths. More concretely, denote by $\mathcal{CP}_{t}(\Gamma)$ the set of closed paths of edge-length $t$ in $\Gamma$. If $\text{Spec}(A_{\Gamma})$ denotes the multiset of eigenvalues of $A_{\Gamma}$, then for every $t \in \mathbb{N}$,

$$
\sum_{\mu \in \text{Spec}(A_{\Gamma})} \mu^{t} = \text{tr}(A_{\Gamma}^{t}) = |\mathcal{CP}_{t}(\Gamma)|.
$$

Orient each of the $k$ edges of $\Omega$ 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 $\Gamma$ is defined: for each edge $x_{i} = (u, v)$ of $\Omega$ and each $i \in [n]$, $\Gamma$ has an edge $((u, i), (v, \sigma_{i}(i)))$. Note that every closed path in $\Gamma$ projects to a closed path in $\Omega$. Thus, instead of counting directly closed paths in $\Gamma$, one can count, for every closed path in $\Omega$, the number of closed paths in $\Gamma$ projecting onto it.

Let $w = x_{j_{1}}^{\varepsilon_{1}} \cdots x_{j_{t}}^{\varepsilon_{t}} \in \mathcal{CP}_{t}(\Omega) \subseteq (X \cup X^{-1})^{t}$ be a closed path in the base graph $\Omega$, beginning (and terminating) at some vertex $v \in V(\Omega)$. (Here $\varepsilon_{i} = \pm 1$ and $x_{j}^{-1}$ means the path traverses the edge $x_{j}$ in the opposite orientation.) For every $i \in [n]$ there is a unique lift of $w$ to some path in $\Gamma$, not necessarily closed, which begins at the vertex $(v, i)$. This lifted path 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 $\sigma_{1}, \ldots, \sigma_{k} \in S_{n}$ according to $w$, namely, $w(\sigma_{1}, \ldots, \sigma_{k}) = \sigma_{j_{1}}^{\varepsilon_{1}} \cdots \sigma_{j_{t}}^{\varepsilon_{t}} \in S_{n}$. Then $j$ is the
image of \( i \) under this permutation: \( j = w(\sigma_1, \ldots, \sigma_k)(i) = \sigma_{j_1}^{\varepsilon_{j_1}} \cdots \sigma_{j_t}^{\varepsilon_{j_t}}(i) \) Thus, the \( i \)-th lift of \( w \) is a closed path if and only if \( i \) is a fixed point of the permutation \( w(\sigma_1, \ldots, \sigma_k) \), and the number of closed paths in \( \Gamma \) projecting onto \( w \) is equal to the number of fixed points of \( w(\sigma_1, \ldots, \sigma_k) \). We denote this number by \( F_{w,n} = F_{w,n}(\sigma_1, \ldots, \sigma_k) \).

Claim 2.1. For every even \( t \in \mathbb{N} \),

\[
\mathbb{E}\left[ \lambda_A(\Gamma)^t \right] \leq \sum_{w \in CP_t(\Omega)} [\mathbb{E}[F_{w,n}] - 1].
\] (2.1)

(The expectation on the l.h.s. is over \( C_{n,\Omega} \), 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,

\[
\lambda_A(\Gamma)^t = \max_{\mu \in \text{Spec}(A_{\Omega})} \left| \text{Spec}(\Gamma) \right| \mu^t \leq \sum_{\mu \in \text{Spec}(A_{\Omega})} \mu^t = \sum_{\mu \in \text{Spec}(A_{\Omega})} \mu^t - \sum_{\mu \in \text{Spec}(A_{\Omega})} \mu^t = [CP_t(\Gamma)] - [CP_t(\Omega)] = \sum_{w \in CP_t(\Omega)} [F_{w,n}(\sigma_1, \ldots, \sigma_k) - 1].
\]

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

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 \( d \) loops, \( \text{Spec}(A_{\Omega}) = \{ d \} \), and \( CP_t(\Omega) = (X \cup X^{-1})^t \), i.e. it consists of all words of length \( t \) in the letters \( X \cup X^{-1} \) (not necessarily reduced), so that \( |CP_t(B_{\frac{d}{2}})| = d^t \).

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 \( \mathbb{E}[F_{w,n}] \). This is where the results from [PP14] come to bear. Let \( F_k = F(X) \) be the free group on \( k \) generators. Every word \( w \in CP_t(\Omega) \subseteq (X \cup X^{-1})^t \) corresponds to an element of \( F_k \) (by abuse of notation we let \( w \) denote an element of \( (X \cup X^{-1})^t \) and of \( F_k = F(X) \) at the same time; it is important to stress that reduction\(^\dagger\) of \( w \) does not affect the associated permutation \( w(\sigma_1, \ldots, \sigma_k) \).)

The main theorem in [PP14] estimates the expected number of fixed points of the 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 primitive rank, which we now describe.

A word \( w \in F_k \) is primitive if it belongs to a basis\(^\dagger\) of \( F_k \). For a given \( w \), one can also ask whether \( w \) is primitive as an element of different subgroups of \( F_k \) (which are free as well by a classical theorem of Nielsen and Schreier [Bog08, Chap. 2.8]). If \( w \) is primitive in \( F_k \), it is also primitive in every subgroup \( J \leq F_k \) (e.g. [Pud14b] Claim 2.5). However, if \( w \) is not primitive in \( F_k \), it is sometimes primitive and sometimes not so in subgroups containing it. Theoretically, one can go over all subgroups of \( F_k \) containing \( w \), ordered by their rank\(^\ddagger\), and look for the first time at which \( w \) is not primitive. First introduced in [Pud14b], the primitivity rank of \( w \in F_k \) captures this notion:

\[\text{rank}(F_k) \leq \text{size}(F_k) \]

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

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

\( \dagger \)A basis of a group is a 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 \( F_k \) this is equivalent to a generating set of size \( \text{size}(F_k) \) [Bog08, Chap. 2.20].

\( \ddagger \)The rank of a free group \( F_k \), denoted \( \text{rk}(F_k) \), is the size of (every) basis of \( F_k \).

8
Definition 2.2. The primitivity rank of $w \in F_k$, denoted $\pi(w)$, is

$$\pi(w) = \min \left\{ \text{rk}(J) \middle| w \in J \leq F_k \text{ s.t. } w \text{ is not primitive in } J \right\}.$$ 

If no such $J$ exists, i.e. if $w$ is primitive in $F_k$, 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 $\text{Crit}(w)$.

For instance, $\pi(w) = 1$ if and only if $w$ is a proper power ($w = v^d$ for some $v \in F_k$ and $d \geq 2$). By Corollary 4.2 and Lemma 6.8 in [Pud14], the set of possible primitivity ranks in $F_k$ 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 $\mathbb{E}[F_{w,n}]$, 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. [PPT4, Thm 1.7] For every $w \in F_k$, the expected number of fixed points in $w(\sigma_1, \ldots, \sigma_k)$ is

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

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 $\mathcal{CP}^m_t(\Omega) = \{w \in \mathcal{CP}_t(\Omega) \mid \pi(w) = m\}$ for the subsets we obtain by splitting $\mathcal{CP}_t(\Omega)$ according to primitivity ranks:

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

$$= \sum_{m=0}^k \sum_{w \in \mathcal{CP}^m_t(\Omega)} \left( \frac{|\text{Crit}(w)|}{m^{m-1}} + O \left( \frac{1}{m^m} \right) \right)$$

(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 $\mathbb{E}[F_{w,n}]$

The error term $O \left( \frac{1}{n^{\pi(w)}} \right)$ in (2.2) depends on $w$. For a given $w \in \mathcal{CP}_t(\Omega)$, 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 paths of length $t$ with a given primitivity rank in $\Omega$. Namely, for every $m$ one needs to control the $O(\cdot)$ term for all $w \in \mathcal{CP}^m_t(\Omega)$ simultaneously. The third stage is therefore the following proposition:

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

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

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

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

$$\mathbb{E} \left[ \lambda_A(\Gamma)^t \right] \leq (1 + o_n(1)) \sum_{m=0}^k \frac{1}{m^{m-1}} \sum_{w \in \mathcal{CP}^m_t(\Omega)} |\text{Crit}(w)|.$$  (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 \( \sum_{w \in \mathcal{CP}_t^\rho(\Omega)} |\text{Crit}(w)| \) 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 \( \sum_{w \in \mathcal{CP}_t^\rho(\Omega)} |\text{Crit}(w)| = |\mathcal{CP}_t^\rho(\Omega)| \). Moreover, words reducing to 1 are precisely the completely back-tracking closed paths, i.e. the paths lifting to closed paths in the covering tree. It follows that the exponential growth rate of \( |\mathcal{CP}_t^\rho(\Omega)| \) 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 \( \Omega \) be a finite, connected graph with \( k \geq 2 \) edges, and let \( m \in \{1, \ldots, k\} \). Then

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

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

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

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

Moreover, for \( \Omega = B_2^d \) the bouquet, there is equality:

\[
\limsup_{t \to \infty} \left[ \sum_{w \in \mathcal{CP}_t^\rho(B_2^d)} |\text{Crit}(w)| \right]^{1/t} = \begin{cases} 2\sqrt{2k - 1} & \text{if } 2m - 1 \leq \sqrt{2k - 1} \\ 2m - 1 + \frac{2k - 1}{2m - 1} & \text{if } 2m - 1 \geq \sqrt{2k - 1} \end{cases}.
\]

**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 \( (X \cup X^{-1})^t \) with primitivity rank \( m \)-see Theorem 8.5.

Table 2 summarizes the content of Theorems 2.3, 4.11 and 8.3.

Whereas in the special case of the bouquet we count words in \( F_k \) 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 \( F_k \). (There is one such subgroup for each vertex \( v \) of \( \Omega \): the one consisting of the words which correspond to closed paths 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 1.4 which are tight up to a small additive constant, and Theorem 1.4 which is tight up to a small multiplicative factor (assuming, of course, that Conjecture 1.3 is true).

\[\text{Table 2}\]
| $\pi (w)$ | Description of $w$ | $\mathbb{E}[F_{w,n}]$ | Growth rate for the bouquet $B_2$ | Bound on growth rate for general $\Omega$ |
|----|-------------------|----------------|------------------|------------------|
| 0  | $w = 1$           | $n$            | $2\sqrt{2k-1}$  | $\rho$          |
| 1  | a power           | $\sim 1 + |\text{Crit}(w)|$ | $2\sqrt{2k-1}$  | $\rho$          |
| 2  | E.g. $x_1^2x_2^2x_2^2$ | $\sim 1 + |\text{Crit}(w)|$ | $2\sqrt{2k-1}$  | $3\rho$        |
| 3  | E.g. $x_1^2x_2^2x_2^2$ | $\sim 1 + |\text{Crit}(w)|$ | $2\sqrt{2k-1}$  | $5\rho$        |
| : | :                | :             | :               | :               |
| $\lfloor \sqrt{2k-1}+1 \rfloor$ | :            | $2\sqrt{2k-1}$ | :               | :               |
| $\lceil \sqrt{2k-1}+1 \rceil$ | :            | $2\pi(w)-1+\frac{2k-1}{\pi(w)-1}$ | :               | :               |
| : | :                |             | :               | :               |
| $k-1$ | :           | $2k-2 + \frac{2}{2k-3}$ | :               | :               |
| $k$ | E.g. $x_1^2x_2^2...x_k^2$ | $\sim 1 + |\text{Crit}(w)|$ | $2k$            | :               |
| $\infty$ | primitive  | 1            | $2k-2 + \frac{2}{2k-3}$ | :               |

Table 2: Primitivity rank, the average number of fixed points, the exponential growth rate of $\sum_{w \in \cal{P}^m} (B_2) |\text{Crit}(w)|$, and bounds on the exponential growth rate of $\sum_{w \in \cal{P}^m} |\text{Crit}(w)|$.

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 [PP14]. Section 6 completes the proof of Theorems 1.4 and 1.5 and addresses the source of the gap between Theorem 1.4 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 $\mathbb{F}_k$ (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 $H \leq_{alg} J$, 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] [PPT14] and especially to [MVW07].

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

**Claim 3.1.** [Pud14b] Cor. 4.4] Every $w$-critical subgroup is an algebraic extension of $\langle w \rangle$ (the subgroup generated by $w$).

More precisely, $\text{Crit} (w)$ 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, $\langle w \rangle \leq H$. If $H$ is not an algebraic extension of $\langle w \rangle$, then there is a proper intermediate free factor $\langle w \rangle \leq L \leq_{ff} H$. Since $w$ is not primitive in $H$, it is also not primitive in $L$ (e.g. [Pud14b] Claim 2.5), but $\text{rk} (L) < \text{rk} (M)$, 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 $\Gamma_X (H)$. 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 $\Gamma_X (H)$ by omitting all the vertices and edges of $\Gamma_X (H)$ which are not traced by any reduced (i.e., non-backtracking) path that starts and ends at the basepoint. Stated informally, we trim all “hanging trees” from $\Gamma_X (H)$. To illustrate, Figure 3.1 shows the graphs $\Gamma_X (H)$ and $\Gamma_X (H)$ for $H = \langle x_1 x_2 x_1^{-1}, x_2^2 x_2 x_1^{-2} \rangle \leq \mathbf{F}_2$.

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

$$
\bullet \rightarrow x_2 \rightarrow \bullet \rightarrow x_2 \rightarrow \bullet \rightarrow x_1 \rightarrow \bullet \rightarrow x_2 \rightarrow \bullet \rightarrow x_3 \rightarrow \bullet \rightarrow x_2 \rightarrow \bullet \rightarrow x_1 \rightarrow \bullet
$$

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 paths around the basepoint in $\Gamma$ is a subgroup of $\mathbf{F} (X)$ which we call the labeled fundamental group of $\Gamma$, and denote by $\pi^X (\Gamma)$. Note that $\pi^X (\Gamma)$ need not be isomorphic to $\pi_1 (\Gamma)$, the standard fundamental group $\pi_1 (\Gamma)$.

---

†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 $\text{Crit} (w) = \{ \emptyset \} = \{ \langle w \rangle \}$.

§Alternatively, $\Gamma_X (H)$ is the quotient $H/T$, where $T$ is the Cayley graph of $\mathbf{F}_k$ with respect to the basis $X$, and $\mathbf{F}_k$ (and thus also $H$) acts on this graph from the left. Moreover, this is the covering-space of $\Gamma_X (\mathbf{F}_k) = \Gamma_X (\mathbf{F}_k)$, 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 $\pi_1 (Y)$. 

---

12
The Schreier coset graph $\Gamma_X(H)$ 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 $Hx_1x_2^{-1}$, etc. The core graph $\Gamma_X(H)$ is the finite graph on the right, which is obtained from $\Gamma_X(H)$ by omitting all vertices and edges that are not traced by reduced closed paths around the basepoint.

However, it is not hard to show that when $\Gamma$ is a core graph, then $\pi_1(\Gamma)$ 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 $F(X)$. It is an easy observation that

$$\pi_1^n(\Gamma_X(H)) = \pi_1^n(\Gamma_X(H)) = H$$

This gives a one-to-one correspondence between subgroups of $F(X) = F_k$ and core graphs labeled by $X$. Namely, $\pi_1^n$ and $\Gamma_X$ are the inverses of each other in a bijection (Galois correspondence)

$$\begin{align*}
\left\{ \begin{array}{c}
\text{Subgroups} \\
\text{of } F(X)
\end{array} \right\} \xrightarrow{\Gamma_X} \left\{ \begin{array}{c}
\text{Core graphs} \\
\text{labeled by } X
\end{array} \right\}.
\end{align*}$$

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, Pud14b]).

**Claim 3.2.** Let $H$ be a subgroup of $F_k$ with an associated core graph $\Gamma = \Gamma_X(H)$.

1. $\text{rk} (H) < \infty \iff \Gamma$ is finite.
2. $\text{rk} (H) = |E(\Gamma)| - |V(\Gamma)| + 1$ (for f.g. subgroup $H$).
3. The correspondence (3.2) restricts to a correspondence between finitely generated subgroups of $F_k$ and finite core graphs.
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, Pud14b or an easy observation:

**Claim 3.3.** Let $H, J, L \leq F_k$ 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 $\eta^X_{H \to J}$.
3. Whenever $H \leq L \leq J$, $\eta^X_{H \to J} = \eta^X_{L \to J} \circ \eta^X_{H \to L}$.
4. If $\Gamma_X (H)$ is a subgraph of $\Gamma_X (J)$, namely if $\eta^X_{H \to J}$ 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

$$\sum_{w \in \mathcal{CP}_t(X)} |\text{Crit} (w)|.$$  

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

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 $w \in F_k$ and let $N \leq_{f.g} F_k$ be a proper algebraic extension of $\langle w \rangle$. Then the closed path 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 $\eta^X_{H \to N}$ is onto: see Definition 5.3 and e.g. [PP14, Claim 4.2]. We repeat the simple argument here.) Otherwise, let $J$ be the subgroup of $N$ corresponding to the subgraph $\Delta$ traced by $w$ (so $\Delta = \text{im} \eta^X_{\langle w \rangle \to N}$, and $J = \pi^X_N (\Delta)$, see Section 3.2 and in particular Claim 3.3). Then $w \in J \leq_{ff} N$ (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 $\Gamma$ into two connected components, then it is obvious that the path of $w$ in $\Gamma$ must traverse $e$ an even number of times, and since this number is $\geq 1$, it is in fact $\geq 2$.

Finally, assume that $e$ is not separating, and $w$ traverses it exactly once, so that the path corresponding to $w$ in $\Gamma_X (N)$ is $w_1 w_2$ (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 $r = \text{rk} (N)$ 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 path 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.

\footnote{Points (1)-(3) can be formulated by saying that 3.2 is in fact an isomorphism of categories, given by the functors $\pi^X_N$ and $\Gamma_X$.}

\footnote{But not vice-versa: for example, consider $\langle x_1 x_2^2 \rangle \leq_{ff} F_2$.}
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 path through \( T \) from \( \otimes \) to the origin of \( e \), and \( p_2 \) the path from the terminus of \( e \) back to \( \otimes \). Then
\[
u_1 = p_1 ep_2 = p_1 w_1^{-1} w_1 e w_2^{-1} p_2 = (p_1 w_1^{-1}) w (w_2^{-1} p_2)
\]
and we are done because \( p_1 w_1^{-1} \) and \( w_2^{-1} p_2 \) avoid \( e \) and thus belong to \( \langle u_2, \ldots, u_r \rangle \).

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 \( J \leq F_k \) of rank \( m \). Then,

1. After omitting the string to \( \otimes \) if the basepoint is a leaf, all vertices of \( \Gamma \) are of degree at most \( 2m \).
2. \( \Gamma \) 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 \( \Gamma \) are of degree \( \geq 2 \). 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 \( \Gamma \) is topologically a bouquet of \( m \) loops (plus, possibly, a string to \( \otimes \)).

(2) Consider \( \Gamma \) 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 \( \leq 2 \) 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 \( e \leq 3m - 1 \).

### 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
\[
\sum_{w \in \mathcal{CP}_1 \left( B_{\frac{d}{2}} \right)} |\text{Crit}(w)| = \sum_{w \in (X \cup X^{-1})^*: \pi(w) = m} |\text{Crit}(w)|.
\]
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
\[
\limsup_{t \to \infty} \left[ \sum_{w \in F_k: |w| = t \land \pi(w) = m} |\text{Crit}(w)| \right]^{1/t} \leq \begin{cases} \sqrt{2k-1} & 2m - 1 \leq \sqrt{2k-1} \\ 2m - 1 & 2m - 1 \geq \sqrt{2k-1} \end{cases}
\]

Put differently, the \( \limsup \) is bounded by \( \max \{ \sqrt{2k-1}, 2m - 1 \} \) (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.
Proof. Note that
\[
\sum_{w \in F_k : \|w\| = t, \pi(w) = m} |\text{Crit}(w)| = \sum_{J \leq F_k : \text{rk}(J) = m} |\{w \in F_k : \|w\| = t, J \in \text{Crit}(w)\}|
\]
\[
\leq \sum_{J \leq F_k : \text{rk}(J) = m} \left|\left\{w \in F_k : \|w\| = t, \langle w \rangle \leq_{\text{alg}} J\right\}\right|
\]
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 \(J \leq F_k\) let \(\nu_t(J)\) denote the corresponding summand:
\[
\nu_t(J) = \left|\left\{w \in J : \|w\| = t, w \text{ traces each edge of } \Gamma_X(J) \text{ at least twice}\right\}\right|.
\]
We classify all \(J\)'s of rank \(m\) by the number of edges in \(\Gamma_X(J)\). Consider all \(X\)-labeled core-graphs \(\Gamma\) of total size \(\delta t\) and rank \(m\) (so that \(\delta t\) 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 \(\delta \in [0, \frac{1}{2}]\).

The counting is performed in several steps:

- First, let us bound the number of unlabeled and unoriented connected pointed graphs with \(\delta t\) 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 paths through the tree from \(\otimes\) to the origins and termini of these edges cover the entire tree. Denote these paths 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 \(\leq \delta t\). Now, \(p_{2,1}\) leaves \(p_{1,1} \cup p_{1,2}\) at one of \(\leq \delta t\) possible vertices and goes on for \(\leq \delta t\) 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 \((\delta t)^2\) possible unlabeled pointed graphs of rank \(m\) with \(\delta t\) edges.

- Next, we bound the number of labelings of each such graph \(\Gamma\) (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 \(\Gamma\) is \(\leq 2k \cdot (2k - 1)^{3t - 1}\).

- For a given labeled core-graph \(\Gamma\), let \(J = \pi_1^X(\Gamma)\) be the corresponding subgroup. We claim that \(\nu_t(J) \leq (4t^2)^{3m - 1} \cdot (2m - 1)^{(1 - 2\delta)t}\). 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 \(\geq 3\)), 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 \(\delta t\) 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 \(\bar{e}\) in \(\Gamma\), it begins in one of \(\leq t\) possible positions (in \(w\)), and from \(\leq 2\) possible directions of \(\bar{e}\). So there \(\leq 4t^2\) possible ways in which \(w\) traces \(\bar{e}\) for the first two times. By Claim 4.2(2) there are at most \(3m - 1\) topological edges, and so at most \((4t^2)^{3m - 1}\) possibilities for how \(w\) traces each topological edge of \(\Gamma\) 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).

\(^1\)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.
Hence,

\[
\sum_{J : f_k: \text{rk}(J) = m} \nu_t(J) \leq (\delta t)^{4m} \cdot 2k(2k - 1)^{\delta t - 1} \cdot (4t^2)^{3m - 1} (2m - 1)^{(1 - 2\delta)t} \\
\leq c \cdot t^{10m - 2} \cdot \left[(2k - 1)^{\delta} (2m - 1)^{1 - 2\delta}\right]^t \\
= c \cdot t^{10m - 2} \cdot \left[\left(\frac{2k - 1}{(2m - 1)^2}\right)^{\delta} (2m - 1)\right]^t. \tag{4.2}
\]

Recall that \(\delta \in [0, \frac{1}{2}]\) and \(\delta t \in \mathbb{N}\). We bound \(\sum_{J : f_k: \text{rk}(J) = m} \nu_t(J)\) by \(\frac{t}{2}\) times the maximal possible value of the r.h.s. of (4.2) (when going over all possible values of \(\delta\)). 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

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

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

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

The proposition follows. \(\square\)

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 \(\Gamma\) be a connected \(d\)-regular graph. Let \(b_{\Gamma,v}(t)\) denote the number of cycles of length \(t\) at some vertex \(v\) in \(\Gamma\), 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 \(A_\Gamma\), denoted \(\text{rad}(\Gamma)\), 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 \(\Gamma\) is defined as \(\text{cogr}(\Gamma) = \text{cogr}(\cdot)\limsup_{t \to \infty} n_{\Gamma,v}(t)^{1/t}\), and is also independent of \(v\).

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

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

Another way to view the parameters \(\text{rad}(\Gamma)\) and \(\text{cogr}(\Gamma)\) 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 paths of length \(t\) in \(T_d\) emanating from \(\otimes\) and terminating inside \(S\). Similarly, \(n_{\Gamma,v}(t)\) is the number of non-backtracking paths 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 \(S_t\) of \(T_d\) belonging to \(S\).

For our needs we introduce (in a separate paper [Pud14a]) an extended formula applying to other types of subsets 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 \(1_S\):)

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

\[\text{1. If } \Gamma \text{ is finite, } \text{rad}(\Gamma) = d. \text{ If } \Gamma \text{ is the } d\text{-regular tree, } \text{rad}(\Gamma) = 2\sqrt{d - 1}.\]

\[\text{2. The } t\text{-th sphere of the pointed } T_d \text{ is the set of vertices at distance } t \text{ from } \otimes.\]

\[\text{3. The results in [Pud14a] include a new proof of the original cogrowth formula.}\]
Remark over all (possibly backtracking) paths 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 paths of length $t$ emanating from $\otimes$. $\nu_f (t)$

**Theorem 4.4.** [Extended Cogrowth Formula] Let $d \geq 3$, $f : V (T_d) \to \mathbb{R}$, $\beta_f (t)$ and $\nu_f (t)$ as above. If $\nu_f (t) \leq \pi \cdot \alpha^t$ 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 $m \in \{1, \ldots, k\}$,

$$\limsup_{t \to \infty} \left| \sum_{w \in \text{Crit} (w)} \right| \leq \left\{ \begin{array}{ll}
2 \sqrt{2k - 1} & \text{if } 2m - 1 \leq \sqrt{2k - 1}, \\
2k - 1 + 2m - 1 & \text{if } 2m - 1 \geq \sqrt{2k - 1}.
\end{array} \right.$$  

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

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 [Pud14a].

**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 $(X \cup X^{-1})^t$ reducing to 1. This is a well-known result:

**Claim 4.7.**

$$\limsup_{t \to \infty} \left| \mathcal{P}^t \left(B_{\frac{1}{2}}\right) \right|^{1/t} = \limsup_{t \to \infty} \left| \{ w \in (X \cup X^{-1})^t \mid w \text{ reduces to 1} \} \right|^{1/t} = 2 \sqrt{2k - 1}.$$  

**Proof.** Denote by $c_{\Gamma} (t, u, v)$ the number of paths of length $t$ from the vertex $u$ to the vertex $v$. $c_{\Gamma} (t, u, v)$ in a connected graph $\Gamma$. If, as above, $A_{\Gamma}$ denotes the adjacency operator on $l^2 (V (\Gamma))$, then

$$c_{\Gamma} (t, u, v) = \langle A_{\Gamma}^t \delta_u, \delta_v \rangle_{1} \quad (\langle \cdot, \cdot \rangle_{1} \text{ marks the standard inner product).}$$

If $\Gamma$ has bounded degrees, then $A_{\Gamma}$ is a bounded self-adjoint operator, hence

$$\text{rad} (\Gamma) = \| A_{\Gamma} \| = \limsup_{t \to \infty} \sqrt{c_{\Gamma} (t, u, v)}$$

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

$$c_{\Gamma} (t, u, v) = \langle A_{\Gamma}^t \delta_u, \delta_v \rangle_{1} \leq \| A_{\Gamma}^t \delta_u \| \cdot \| \delta_v \| \leq \| A_{\Gamma} \|^{t} \cdot \| \delta_u \| \cdot \| \delta_v \| = \text{rad} (\Gamma)^t$$  

(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 paths of length $t$ at the basepoint of the 2k-regular tree $T_{2k}$. So the number we seek is

$$\limsup_{t \to \infty} \sqrt{c_{T_{2k}} (t, v, v)}$$

which therefore equals $\text{rad} (T_{2k}) = 2 \sqrt{2k - 1}$. □
4.2 An arbitrary regular base-graph $\Omega$

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

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

**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, $\Omega$ 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 path in $\Omega$ of length $t$ can be regarded as an element of $(X \cup X^{-1})^t$ and (after reduction) of $F_k = \mathcal{F}(X)$. If a word $w \in \mathcal{CP}_t(\Omega)$ begins (and ends) at $v \in V(\Omega)$, then $w \in J_v$, where $J_v = \pi_X^X(\Omega_v)$ is the subgroup of $F_k$ corresponding to the $X$-labeled graph $\Omega$ pointed at $v$. The rank of $J_v$ is independent of $v$ and equals $\text{rk}(\Omega)$. It is easy to see that $J_v \leq_{ff} F_k$ (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 4.11. This basis can then be extended to a basis of $F_k$ by the $x_i$'s associated with the edges inside the spanning tree. So if $w$ is primitive in $J_v$, it is also primitive in $F_k$ and $\pi(w) = \infty$. Otherwise, $\pi(w) \leq \text{rk}(J_v) = \text{rk}(\Omega)$. \hfill $\Box$

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

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

**Proof.** As $J_v \leq_{ff} F_k$, it follows that $J_v \cap N \leq_{ff} N$ (see e.g. [PP14, 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 $N \leq J_v$. \hfill $\Box$

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

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

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

**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.

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

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:

$$\limsup_{t \to \infty} \left[ \sum_{w \in \mathcal{CP}^m_t(\Omega): w \text{ is reduced}} |\text{Crit}(w)| \right]^{1/t} \leq \begin{cases} \frac{\sqrt{d-1}}{d-1} 2m-1 \in \left[1,\sqrt{d-1}\right] \\ 2m-1 \in \left[d^{-1},d^{-1},d-1\right] \\ d-1 \in \left[d^{-1},d-1,2\text{rk}(\Omega)-1\right] \end{cases}. $$
From Claim 4.9 we deduce that every critical subgroup is necessarily a subgroup of \( J_v = \pi_X^1 (\Omega_v) \) for some vertex \( v \in V (\Omega) \). As in the proof of Proposition 4.3, we denote

\[
\nu_t (J) = \left\{ w \in F_k \left| \begin{array}{c}
|w| = t, \ w \text{ traces each edge of } \Gamma_X (J) \text{ at least twice}
\end{array} \right. \right\}
\]

for every \( J \leq F_k \), and as in (4.1), we obtain the bound:

\[
\sum_{w \in CP^m_t (\Omega); w \text{ is reduced}} |\text{Crit} (w)| \leq \sum_{v \in V (\Omega) \ J \leq J_v; \text{rk} (J) = m} \nu_t (J).
\]

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 \( \Omega_v \). 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 \(|V (\Omega)| \cdot d (d - 1)^{d-1}\).
- 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 \( \nu_t (J) \leq (4t^2)^{3m-1} \cdot (\min \{2m, d\} - 1)^{(1-2s)t} \).

We conclude as in the proof of Proposition 4.3

\[
\square
\]

### 4.3 An arbitrary base-graph \( \Omega \)

We now return to the most general case of an arbitrary connected base graph \( \Omega \). Theorem 4.11 below is needed for proving the bound on the new spectrum of the adjacency operator on \( \Gamma \), the random covering of \( \Omega \) in the \( C_{n, \Omega} \) 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 \( \Omega \) (and of \( \Gamma \)), 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, \( \Omega \) becomes a non-pointed \( X \)-labeled graph, where \( X = \{x_1, \ldots, x_k\} \). Every path in \( \Omega \) of length \( t \) can be regarded as an element of \( (X \cup X^{-1})^t \) and (after reduction) of \( F_k = F (X) \). We also denoted \( \text{rk} (\Omega) = |E (\Omega)| - |V (\Omega)| + 1 \) and showed that \( \pi (w) \in \{0, 1, \ldots, \text{rk} (\Omega), \infty\} \) for every \( w \in CP_t (\Omega) \) (Lemma 4.8). The main theorem of this subsection is the following:

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

\[
\limsup_{t \to \infty} \left[ \sum_{w \in CP^m_t (\Omega)} |\text{Crit} (w)| \right]^{1/t} \leq (2m - 1) \cdot \rho.
\]

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 \( F_k \) has exactly one critical subgroup, so \( \sum_{w \in CP^0_t (\Omega)} |\text{Crit} (w)| \) equals \( |CP^0_t (\Omega)| \).
Claim 4.12.  
\[
\limsup_{t \to \infty} \left| CP_t^0 (\Omega) \right|^{1/t} = \rho.
\]

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 \( \hat{v} \), where \( \hat{v} \in p^{-1} (v) \) 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 \( \left| A_T^{t, \delta_v} \right| \), and
\[
\left[ A_T^{t, \delta_v} \right] = \langle A_T^{t, \delta_v^1} \rangle = \left\langle A_T^{t, \delta_v^2} \delta_v^1 \right\rangle \leq \left\| A_T^{t, \delta_v^1} \right\| \cdot \left\| \delta_v^1 \right\|^2 = \left\| A_T^{t, \delta_v^1} \right\| = \rho^t
\]
(the last equality follows from \( A_T \) being self-adjoint), and thus
\[
\limsup_{t \to \infty} \left| CP_t^0 (\Omega) \right|^{1/t} = \limsup_{t \to \infty} \left[ V (\Omega) \right] \cdot \rho^t = \rho.
\]

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

We return to the proof of Theorem 4.11. By Claim 3.1
\[
\sum_{w \in CP_t^0 (\Omega)} \left| \text{Crit} (w) \right| = \sum_{N \leq F_k : \text{rk}(N) = m} \left| \left\{ w \in CP_t (\Omega) \mid N \in \text{Crit} (w) \right\} \right|
\]
\[
\leq \sum_{N \leq F_k : \text{rk}(N) = m} \left| \left\{ w \in CP_t (\Omega) \mid \langle w \rangle \leq_{\text{alg}} N \right\} \right| \tag{4.7}
\]

and we actually bound the latter summation. For every \( N \leq F_k \), we let \( \beta_t (N) \) denote the corresponding summand, namely
\[
\beta_t (N) = \left| \left\{ w \in CP_t (\Omega) \mid \langle w \rangle \leq_{\text{alg}} N \right\} \right|.
\]

Note that while a non-reduced element \( w \in CP_t (\Omega) \) with \( w \in N \) might not correspond to a close path in \( \Gamma_X (N) \), it always does correspond to a close path at the basepoint of the Schreier coset graph \( \Gamma_X (N) \).

If \( N \leq F_k \) 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 \( N \leq F_k \) 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 \( w' \in CP_t (\Omega) \) satisfies \( \langle w' \rangle \leq_{\text{alg}} N' \), then the path corresponding to \( w' \) in the Schreier graph \( \Gamma_X (N') \) must visit all vertices and edges of the core of \( \Gamma_X (N') \), and in particular the basepoint of \( \Gamma_X (N) \) (by Lemma 4.1). So there is some cyclic rotation \( w' \) of \( w \) satisfying \( \langle w \rangle \leq_{\text{alg}} N \) (clearly, \( w \) also belongs to \( CP_t (\Omega) \)). On the other hand, each such \( w \) has at most \( t \) possible cyclic rotations, each of which corresponds to one \( w' \) and one \( N' \). \( \square \)

Next, we classify the subgroups \( N \leq F_k \) according to their “topological” core graph \( \Lambda \). 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 \( F_k \), we also restrict attention to one CR representative \( \Lambda \) of each “conjugacy class” of topological core graphs. Ignoring the basepoints, any \( \Lambda \) in the “conjugacy class” of \( \Lambda \) retracts to this representative. For example, we need exactly three such representatives in rank 2, as shown in Figure 4.11.

The following proposition is the key step in the proof of Theorem 4.11.
**Proposition 4.14.** Let $\Lambda$ be a pointed finite connected graph without vertices of degree 1 or 2 except for possibly the basepoint, and let $\delta$ denote its maximal degree. Then the sum of $\beta_i(N)$ over all subgroup $N \leq F_k$ whose core graph is topologically $\Lambda$ is at most

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

**Proof.** Denote $r = |E(\Lambda)|$. Order and orient the edges of $\Lambda \{e_1, e_2, \ldots, e_r\}$ 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 $e_1, \ldots, e_{i-1}$. 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: $v_0 \otimes \cdots \otimes v_2$. Also, denote by $\hat{v}_i$ the smallest index $j$ such that $e_i$ begins at $v_j$, so $\hat{v}_i$ is a direct edge from $v_{\text{beg}(i)}$ to $v_i$ and $\hat{v}_i < i$. In our example, $\hat{v}(1) = \hat{v}(3) = 0$ and $\hat{v}(2) = 1$.

Note that each $N$ corresponding to $\Lambda$ is determined by the paths (words in $F_k$) 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 $\hat{v}_0 \in V(T)$ which projects to $v_0$. We claim that every subgroup $N \leq J_{v_0}$ corresponding to $\Lambda$ is completely determined by a set of vertices $\hat{v}_1, \ldots, \hat{v}_r$ in $T$: the topological edge in $\Gamma_X(N)$ associated with $e_i$ corresponds to the path in $T$ from $\hat{v}_\text{beg}(i)$ to $\hat{v}_i$. (There are some constraints on the choices of the $\hat{v}_i$'s. For example, if $v_i = v_j$ then $\hat{v}_i$ and $\hat{v}_j$ 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 $\hat{v}_1, \ldots, \hat{v}_r$ in $T$.

The counting argument that follows resembles the one in Proposition 4.3. Fix a particular $N \leq J_{v_0}$ corresponding to $\Lambda$ and let $\hat{v}_1, \ldots, \hat{v}_r$ be the corresponding vertices in $T$. By Lemma 4.1, if $w \in (X \cup X^{-1})^t$ satisfies $\langle w \rangle \leq_\text{alg} N$, 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 $0 \leq \tau_{i,1} < t - 1$), 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 $\epsilon_{i,1}, \epsilon_{i,2} \in \{\pm 1\}$. 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 $\epsilon_{i,j}$'s: a total of less than $(4t^4)^t$ options. There are $t - \ell_{i,1} - \ell_{i,2} - \ldots - \ell_{r,1} - \ell_{r,2}$ 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 $t$'th segment reduces to some path in $\Gamma_X(N)$, with at most $(\delta - 1)^{q_{4r}}$ possibilities (recall that $\delta$ marks the maximal degree of a vertex in $\Lambda$). Overall, there are at most $(\delta - 1)^{q_1 + \ldots + q_{4r}} \leq (\delta - 1)^t$ options to choose the reduced paths traced by these $4r$ segments in $w$. Given such a reduced path for the $i$'th segment, let $\tilde{x}_i, \tilde{y}_i \in V(T)$ be suitable vertices in the tree such that the reduced path lifts to the unique reduced path from $\tilde{x}_i$ to $\tilde{y}_i$.

Now, we sum over all subgroups $N$ corresponding to $\Lambda$ and all words $w \in CP_t(\Omega)$ with $\langle w \rangle \leq_\text{alg} N$. By adding a factor of $|V(\Omega)| (4t^4)^t \cdot (\delta - 1)^t$ we assume we already know $v_0$ and $\hat{v}_0$, the $\epsilon_{i,j}$'s, the $\ell_{i,j}$'s, the $\hat{v}_i$'s, the $\tilde{x}_i$'s and the reduced $4r$ paths. Moreover, conditioning on knowing $\hat{v}_1, \ldots, \hat{v}_r$, we also know the $\tilde{x}_i$'s and the $\tilde{y}_i$'s. Recall that $C_T(t, u, v)$ denotes the number of paths of length $t$ in a graph $\Gamma$ 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 $C_T(\ell_{i,j}, \hat{v}_{\text{beg}(i)}, \hat{v}_i)$ possible subwords corresponding to the $j$'th traverse of $e_i$ (even if $\epsilon_{i,j} = -1$, because $C_T(\ell_{i,j}, \tilde{x}_i, \tilde{y}_i) = C_T(\ell_{i,j}, \hat{v}_i, \hat{v}_{\text{beg}(i)})$. Similarly, there are at most $C_T(q_i, \tilde{x}_i, \tilde{y}_i)$ subwords corresponding to the the $i$'th intermediate segment. Thus,
if \( \alpha = |V(\Omega)| \cdot (4t^4)^{\gamma} \cdot (\delta - 1)^t \) then

\[
\sum_{N \leq F_k; \quad \Gamma(X(N)) \geq \Lambda} \beta_t(N) \leq \alpha \cdot \sum_{\hat{v}_1, \ldots, \hat{v}_r \in V(T)} \left[ \prod_{i=1}^{r-1} \prod_{j=1}^{2} \text{ct} \left( \ell_{i,j}, \hat{\nu}_{\text{beg}(i)}, \hat{\nu}_i \right) \right] \prod_{i=1}^{4r} \text{ct} \left( q_i, \hat{\nu}_i, \hat{\nu}_i \right)
\]

\[
\leq \alpha \cdot \left[ 4^r \rho^{2\gamma} \right] \sum_{\hat{v}_1, \ldots, \hat{v}_r \in V(T)} \left[ \prod_{i=1}^{r-1} \prod_{j=1}^{2} \text{ct} \left( \ell_{i,j}, \hat{\nu}_{\text{beg}(i)}, \hat{\nu}_i \right) \right]
\]

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

\[
\sum_{N \leq F_k; \quad \Gamma(X(N)) \geq \Lambda} \beta_t(N) \leq \alpha \cdot \rho \sum_{\hat{v}_1 \in V(T)} \left( f(1) \sum_{\hat{v}_2 \in V(T)} f(2) \ldots \right)
\]

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 \( \sum_{\hat{v}_r \in V(T)} f(r) \) as follows:

\[
\sum_{\hat{v}_r \in V(T)} f(r) = \sum_{\hat{v}_r \in V(T)} \text{ct} \left( \ell_{r,1}, \hat{\nu}_{\text{beg}(r)}, \hat{\nu}_r \right) \text{ct} \left( \ell_{r,2}, \hat{\nu}_{\text{beg}(r)}, \hat{\nu}_r \right) = \sum_{\hat{v}_r \in V(T)} \text{ct} \left( \ell_{r,1}, \hat{\nu}_{\text{beg}(r)}, \hat{\nu}_r \right) \text{ct} \left( \ell_{r,2}, \hat{\nu}_r, \hat{\nu}_{\text{beg}(r)} \right) = \text{ct} \left( \ell_{r,1} + \ell_{r,2}, \hat{\nu}_{\text{beg}(r)}, \hat{\nu}_{\text{beg}(r)} \right) \leq \rho^{\ell_{r,1} + \ell_{r,2}}.
\]

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

\[
\sum_{N \leq F_k; \quad \Gamma(X(N)) \geq \Lambda} \beta_t(N) \leq \alpha \cdot \rho \sum_{\hat{v}_1 \in V(T)} \prod_{i=1}^{r} \rho^{\ell_{i-1} + \ell_i} = |V(\Omega)| \cdot (4t^4)^{\gamma} \cdot (\delta - 1)^t \cdot \rho^t.
\]

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

\[
\sum_{w \in G^{\text{m}}(\Omega)} |\text{Crit} \left( w \right)| \leq \sum_{N \leq F_k; \quad \text{rk}(N)=m} \beta_t(N)
\]

\[
\leq \sum_{[N] \in \text{ConjCl}(\text{Cosp}(F_k, m))} t \beta_t(N) \quad (4.8)
\]

where the final summation is over all conjugacy classes of subgroups of rank \( m \) in \( F_k \), 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 \( \Lambda \).
Finally, split the summation of the CR representatives \( N \) by their topological graph \( \Lambda \). By Claim 4.12 each such \( \Lambda \) has maximal degree at most \( 2m \) and at most \( 3m - 1 \) edges, so by Proposition 4.14 the \( N \)'s corresponding to each \( \Lambda \) contribute to the summation in (4.8) at most 
\[
t \cdot |V(\Omega)| \cdot (4t^t)^{3m-1} \cdot (2m-1)^t \cdot \rho^t.
\]
This finishes the proof of Theorem 4.11 as there is a finite number of topological graphs \( \Lambda \) of rank \( m \). □

5 Controlling the Error Term of \( \mathbb{E} [\mathcal{F}_{w,n}] \)

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 \( w \in \mathcal{F}_k \) the following holds:
\[
\mathbb{E} [\mathcal{F}_{w,n}] = 1 + \frac{|\text{Crit}(w)|}{n^{\pi(w)-1}} + O \left( \frac{1}{n^{\pi(w)}} \right).
\]
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 \( w \in (X \cup X^{-1})^t \) satisfy \( \pi(w) \neq 0 \) (so \( w \) does not reduce to 1). If \( n > t^2 \) then
\[
\mathbb{E} [\mathcal{F}_{w,n}] \leq 1 + \frac{1}{n^{\pi(w)-1}} \left( |\text{Crit}(w)| + t^{2+2\pi(w)} \right).
\]

Achieving such a bound requires more elaborated details from the proof of Theorem 2.3, which appears in \cite{PP14}. We therefore begin with recalling relevant concepts and results from \cite{PP14}. 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 \( (X \cup X^{-1})^t \) that reduce to 1:

**Claim 5.2.** Let \( w \in (X \cup X^{-1})^t \) satisfy \( \pi(w) = 0 \) (so \( w \) reduces to 1). If \( n > t^2 \) then
\[
\mathbb{E} [\mathcal{F}_{w,n}] \leq 1 + \frac{1}{n^{\pi(w)-1}} \left( |\text{Crit}(w)| + t^{2+2\pi(w)} \right).
\]

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

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 \( H \leq J \leq \mathcal{F}_k \). 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 \( H \ X\text{-covers} J \) and denote this by \( H \preceq_X J \).

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 $C_{n,H}$ are topological covering maps, and we reserve the term “coverings” for these.

For instance, $H = \langle x_1x_2x_1^{-1}x_1^2x_2x_1^{-2} \rangle \leq \mathbb{F}_k$ $X$-covers the group $J = \langle x_2, x_1^2, x_1x_2x_1 \rangle$, the corresponding core graphs of which are the leftmost and rightmost graphs in Figure 5.1. As another example, a core graph $\Gamma$ $X$-covers $\Gamma_X (\mathbb{F}_k)$ (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 $H \leq X J$ indeed depends on the given basis $X$ of $\mathbb{F}_k$. For example, if $H = \langle x_1x_2 \rangle$ then $H \leq X \mathbb{F}_2$. However, for $Y = \{x_1x_2, x_2 \}$, $H$ does not $Y$-cover $\mathbb{F}_2$, 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 “$\leq X$” indeed constitutes a partial ordering of the set of subgroups of $\mathbb{F}_k$. In fact, restricted to f.g. subgroups it becomes a locally-finite partial order, which means that if $H \leq X J$ then the interval of intermediate subgroups $[H, J]_{\leq X} = \{ M \leq F_k | H \leq X M \leq X J \}$ is finite:

**Claim 5.4.** If $H \leq F_k$ is a f.g. subgroup then it $X$-covers only a finite number of groups. In particular, the partial order “$\leq X$” restricted to f.g. subgroups of $\mathbb{F}_k$ 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). □

### 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 $\eta_H^{\Gamma} \rightarrow J$). However, not every partition $P$ of $V (\Gamma_X (H))$ corresponds to a quotient core-graph. Indeed, $\Delta$, 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. [Pud14b, 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 $\pi_1^X$, this core graph $\Gamma_X (H)/P$ is $\Gamma_X (J)$, where $J = \pi_1^X (\Delta)$. The resulting partition $\tilde{P}$ of $V (\Gamma_X (H))$ (the blocks of which are the fibers of $\eta_H^{\Gamma} \rightarrow J$) 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.

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 $P \leq 2^X$ is a partition of some set $X$, let

$$\|P\| \overset{df}{=} |X| - |P| = \sum_{B \in P} (|B| - 1). \tag{5.1}$$

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 $X$ in order to obtain the equivalence relation $P$. Restricting to pairs of subgroups $H, J$ with $H \leq X J$, we can define the following distance function:

**Definition 5.5.** Let $H, J \leq F_k$ be subgroups such that $H \leq X J$, 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 $\rho_X (H, J)$ or $\rho (\Gamma, \Delta)$ as

$$\rho_X (H, J) = \min \left\{ \|P\| \left| P \text{ a partition of } V (\Gamma_X (H)) \text{ satisfying } \Gamma_X (H)/P = \Gamma_X (J) \right. \right\}. \tag{5.2}$$

1 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 [Pud14b, PPT14].

2 In [PP14], the notation $\Gamma_X (H)/P$ was used to denote something a bit different (the unfolded graph $\Delta$).
Recall that Theorem 2.3 estimates \( \sigma \) where
\[ \text{5.3 From random elements of } S_n \text{ to random subgroups} \]

Recall that Theorem 2.3 estimates \( \mathbb{E} [\mathcal{F}_{w,n}] \), the expected number of fixed points of \( \sigma_1, \ldots, \sigma_k \), 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 \( H \leq J \leq F_k \), let \( \alpha_{J,S_n} : J \to S_n \) be a random homomorphism chosen at uniform distribution (there are exactly \( |S_n|^{rk(J)} \) such homomorphisms). Then \( \alpha_{J,S_n} (H) \) is a

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 [Pud14b, PP14].

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.** [Pud14b, Theorem 1.1 and Lemma 3.3] Let \( H, J \leq F_k \) such that \( H \leq \bar{X} J \). Then
\[
rk (J) - rk (H) \leq \rho_X (H, J) \leq rk (J).
\]

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

This theorem is used, in particular, in the proof in [PP14] 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 \( \Gamma_X (H)/(P_1, \ldots, P_r) \). It is easy to see that one 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:

\[
\rho_X (H, J) = \min \left\{ \left\| P_i \right\| + \ldots + \left\| P_r \right\| \middle| P_i : \text{a partition of } V (\Gamma_X (H)) \text{ or of } E (\Gamma_X (H)) \text{ s.t. } \Gamma_X (H)/(P_1, \ldots, P_r) = \Gamma_X (J) \right\}. \quad (5.3)
\]

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 \( \sigma_1, \ldots, \sigma_k \), 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 \( H \leq J \leq F_k \), let \( \alpha_{J,S_n} : J \to S_n \) be a random homomorphism chosen at uniform distribution (there are exactly \( |S_n|^{rk(J)} \) such homomorphisms). Then \( \alpha_{J,S_n} (H) \) is a

Figure 5.1: The left graph is the core graph \( \Gamma_X (H) \) of \( H = \langle x_1 x_2 x_1^{-1}, x_1^2 x_2 x_1^{-2} \rangle \leq F_2 \). 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 \( P = \{ \{v_1, v_4\}, \{v_2, v_3\} \} \).
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
\[
\Phi_{H,J}(n) \overset{\text{def}}{=} \mathbb{E}[\text{common fixed points} (\alpha_{J,S_n}(H))].
\]
This indeed generalizes $\mathbb{E}[\mathcal{F}_w,n]$ for
\[
\mathbb{E}[\mathcal{F}_w,n] = \Phi_{(w),k}(n). \quad (5.4)
\]

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 $[x, y]_\preceq = \{z \mid x \preceq z \preceq y\}$ 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 [PP14]).

In our case, the poset in consideration is $\text{sub}_g (F_k) = \{ H \leq F_k \mid H \text{ is f.g.} \}$, and the partial order is $\preceq_x$, which is indeed locally-finite (Claim 5.4). We define three derivations of the function $\Phi$ 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 $\Phi$ with the Möbius function of $\text{sub}_g (F_k) \subseteq \mathbb{R}$ (see [PP14]) but here we define them in an equivalent simpler way: these are the functions satisfying, for every $H \leq_x J$,
\[
\Phi_{H,J}(n) = \sum_{M \leq x H, J} L_{M,J}(n) = \sum_{M \leq x N \leq x J} C_{M,N}(n) = \sum_{N \leq x H \leq x J} R_{H,N}(n). \quad (5.5)
\]
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 $H \leq_x J$, $L_{H,J}(n) = \Phi_{H,J}(n) - \sum_{M \leq x H, J} L_{M,J}(n)$ 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 [PP14].

Proposition 5.7 ([PP14], 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 $H \leq_{\text{alg}} J$ then $H \leq_x J$ (e.g. [PP14] Claim 4.2), we obtain that
\[
\Phi_{H,J}(n) = \sum_{N \leq_{\text{alg}} x J} R_{H,N}(n). \quad (5.6)
\]

Next, $\Phi_{H,J}(n)$ is given a geometric interpretation: it turns out it equals the expected number of lifts of $\eta_{H \rightarrow J} : \Gamma_X(H) \rightarrow \Gamma_X(J)$ to a random $n$-covering of $\Gamma_X(J)$ in the model $\mathcal{C}_{n,\Gamma_X(J)}$ [PP14] Lemma 6.2]. Similarly, $L_{H,J}(n)$ counts the average number of injective lifts [PP14] 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 [PP14] 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 $H, J \leq F_k$ be f.g. subgroups such that $H \leq_x J$. Then for $n \geq |E(\Gamma_X(H))|$, 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 $n$. 

27
After some involved combinatorial arguments, one obtains from this the following expression for \( C_{M,N}(n) \): Denote by \( \text{Sym}(S) \) the set of permutations of a given set \( S \). Every permutation \( \sigma \in \text{Sym}(S) \) defines, in particular, a partition on \( S \) whose blocks are the cycles of \( \sigma \). By abuse of notation we denote by \( \sigma \) 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 \( \sigma \). We also use \( V_M \) and \( E_M \) as short for \( V(\Gamma_X(M)) \) and \( E(\Gamma_X(M)) \), \( V_M, E_M \) respectively.

**Proposition 5.9** ([PP14], Section 7.1). Let \( M, N \leq F_k \) be f.g. subgroups with \( M \leq \leq N \). Consider the set 
\[
\mathcal{T}_{M,N} = \left\{ (\sigma_0, \sigma_1, \ldots, \sigma_r) \mid r \in \mathbb{N}, \sigma_0 \in \text{Sym}(V_M), \sigma_1, \ldots, \sigma_r \in \text{Sym}(E_M) \setminus \{\text{id}\} \right\}.
\]
Then
\[
C_{M,N}(n) = \frac{1}{n^{|\text{Crit}(M)| - 1}} \sum_{(\sigma_0, \sigma_1, \ldots, \sigma_r) \in \mathcal{T}_{M,N}} (-1)^r \left( -\frac{1}{n} \right)^{\sum_{i=0}^r \|\sigma_i\|}.
\]

The derivation of the main result of [PP14] (Theorem 2.3) from Theorem 5.6 and Propositions 5.7 and 5.9 is short: see the beginning of Section 7 in [PP14].

### 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 \( 1 \neq w \in F_k \) of length \( t \) and every \( n > t^2 \),
\[
\mathbb{E} [\mathcal{F}_{w,n}] \leq 1 + \frac{1}{n^{\pi(w)-1}} \left( |\text{Crit}(w)| + \frac{t^2 + 2\pi(w)}{n - t^2} \right).
\]
(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 \( \mathbb{E} [\mathcal{F}_{w,n}] = \Phi_{(w),F_k}(n) \) and this quantity is given by some rational expression in \( n \) (for large enough \( n \), say \( n \geq |w| \), 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 [Pud14b] 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 \( a_{\pi(w)-1} = |\text{Crit}(w)| \) (unless \( \pi(w) = 1 \) in which case \( a_0 = 1 + |\text{Crit}(w)| \)). So our goal here is to bound the remaining coefficients \( a_s(w) \) for \( s \geq \pi(w) \).

The discussion in Section 5.4 yields the following equalities:
\[
\mathbb{E} [\mathcal{F}_{w,n}] = \Phi_{(w),F_k}(n) = \sum_{N: (w) \leq u \leq N \leq F_k} R_{(w),N}(n) = \sum_{M,N: (w) \leq M \leq N} C_{M,N}(n) = \sum_{M: (w) \leq M \leq N} \sum_{N: M \leq N} C_{M,N}(n)
\]

28
From Proposition [5.9] we obtain that for a fixed \( M \),
\[
\sum_{N:M \leq N} C_{M,N}(n) = \frac{1}{n^{|\text{rk}(M)|-1}} \sum_{r \in \mathbb{N}} (-1)^r \sum_{\sigma_0 \in \text{Sym}(V_{M}) \atop \sigma_1, \ldots, \sigma_r \in \text{Sym}(E_M) \setminus \{\text{id}\}} (-1)^{\|\sigma_0\| + \cdots + \|\sigma_r\|}.
\]

For every \( q \geq 0 \) define the following set:
\[
P_{M,q} = \left\{ (\sigma_0, \ldots, \sigma_r) \mid r \in \mathbb{N}, \sigma_0 \in \text{Sym}(V_M) \atop \sigma_1, \ldots, \sigma_r \in \text{Sym}(E_M) \setminus \{\text{id}\}, \|\sigma_0\| + \cdots + \|\sigma_r\| = q \right\}, \tag{5.7}
\]
so that
\[
\sum_{N:M \leq N} C_{M,N}(n) = \frac{1}{n^{|\text{rk}(M)|-1}} \sum_{q=0}^{\infty} \frac{(-1)^q}{n^q} \sum_{(\sigma_0, \ldots, \sigma_r) \in P_{M,q}} (-1)^r.
\]
Hence,
\[
a_s(w) = \sum_{i=1}^{s+1} \sum_{\substack{M: (w) \leq M \atop \text{rk}(M) = i}} (-1)^{s-(i-1)} \sum_{(\sigma_0, \ldots, \sigma_r) \in P_{M,s-(i-1)}} (-1)^r. \tag{5.8}
\]
In what follows we ignore the alternating signs of the summands in (5.8) and bound \(|a_s(w)|\) by
\[
|a_s(w)| \leq \sum_{i=1}^{s+1} \sum_{\substack{M: (w) \leq M \atop \text{rk}(M) = i}} |P_{M,s-(i-1)}|. \tag{5.9}
\]

**Claim:** For every \( M \leq f_g \mathbf{F}_k \) with \( \langle w \rangle \leq M \), we have \( |P_{M,q}| \leq t^{2q} \).

**Proof of Claim:** Fix \( M \) and denote \( b_q = |P_{M,q}| \). Clearly, \( b_0 = 1 \), and we proceed by induction on \( q \). Let \( q \geq 1 \). We split the set \( P_{M,q} \) by the value of \( \sigma_r \). For \( r = 0 \) there are at most
\[
|\{ \sigma \in \text{Sym}(V_M) \mid \|\sigma\| = q \}| \leq \left( \frac{|V_M|}{2} \right)^q \leq \left( \frac{t}{2} \right)^q \leq \frac{t^{2q}}{2^q}
\]
elements with \( r = 0 \). (For the middle inequality note that \( |V_M| \leq |V_{(w)}| \leq t \); this is also the case with the edges: \( |E_M| \leq |E_{(w)}| \leq t \). For \( r \geq 1 \), \( \sigma_r \) is a permutation of the set of edges \( E_M \) and given \( \sigma_r \), the number of options for \( \sigma_0, \ldots, \sigma_{r-1} \) is exactly \( b_{q-\|\sigma_r\|} \). By the induction hypothesis we obtain:
\[
b_q \leq \frac{t^{2q}}{2^q} + \sum_{\sigma_r \in \text{Sym}(E_M) \setminus \{\text{id}\}} b_{q-\|\sigma_r\|} = \frac{t^{2q}}{2^q} + \sum_{\alpha=1}^{q} b_{q-\alpha} |\{ \sigma \in \text{Sym}(E_M) \mid \|\sigma\| = \alpha \}|
\]
\[
\leq \frac{t^{2q}}{2^q} + \sum_{\alpha=1}^{q} t^{2q-2\alpha} \frac{t^{2\alpha}}{2^\alpha} = t^{2q}. \quad \square
\]

We proceed with the proof of the proposition. For a given \( w \in (X \cup X^{-1})^t \) there are at most \( \left( \frac{|V_{(w)}|}{2} \right)^{\beta} \leq \left( \frac{t}{2} \right)^{\beta} \) partitions of norm \( \beta \) of \( V_{(w)} \), and so at most \( \left( \frac{t}{2} \right)^{\beta} \) subgroups \( M \) of rank \( \beta \) with \( \langle w \rangle \leq M \) (see Theorem [5.6]). Hence from (5.9) we obtain,
\[
|a_s(w)| \leq \sum_{i=1}^{s+1} \left( \frac{t}{2} \right)^i \frac{t^{2(s-(i-1))}}{2^i} \leq \sum_{i=1}^{s+1} \left( \frac{t}{2} \right)^{2i} t^{2(s-i+1)} \leq t^{2s+2}.
\]
Finally,
\[ |E[\mathcal{F}_{w,n}] - 1 - \left| \text{Crit}(w) \right| |_{\pi(w)-1} \leq \sum_{s=\pi(w)}^{\infty} \frac{a_s(w)}{n^s} \leq \sum_{s=\pi(w)}^{\infty} \frac{|a_s(w)|}{n^s} \}
\[ \leq \sum_{s=\pi(w)}^{\infty} \frac{t^{2s+2}}{n^s} = t^2 \cdot \left( \frac{t^2}{n} \right) \pi(w) \cdot \frac{n}{n-t^2} \]  

This finishes the proof.

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 \( \Gamma \) 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 \( A_{\Gamma} \). As explained in more details in Appendix A, this yields the same result for a uniformly random \( d \)-regular simple graph.

Let \( \varepsilon > 0 \). For \( m \in \{0, 1, \ldots, k\} \), Corollary 4.5 (for \( m \geq 1 \)) and Claim 4.7 (for \( m = 0 \)) yield that for large enough \( t \),
\[ \sum_{w \in (X \cup X^{-1})^t: \pi(w)=m} |\text{Crit}(w)| \leq g(2m-1) + \varepsilon t^t, \]
where \( g(\cdot) \) is defined as in (4.4):
\[ g(2m-1) = \begin{cases} 2\sqrt{d-1} & 2m-1 \leq \sqrt{d-1} \\ 2m-1 + \frac{d-1}{2m-1} & 2m-1 > \sqrt{d-1} \end{cases} \]
Thus

\[
\mathbb{E} \left[ \lambda(\Gamma)^t \right] \leq \left( 1 + \frac{t^2 + 2k}{n - t^2} \right) \sum_{m=0}^{k} \frac{\left[ g(2m - 1) + \varepsilon \right]^t}{n^{m-1}}
\]

\[
\leq \left( 1 + \frac{t^2 + 2k}{n - t^2} \right) \cdot (k + 1) \cdot \left[ \max \left\{ \frac{n^{1/2} \left[ g(-1) + \varepsilon \right], g(1) + \varepsilon, 2(3\varepsilon + \varepsilon, \ldots \right. \right. \\
\left. \left. \left. \left. \ldots \right\} \right\} \right]^t
\]

(6.1)

Recall that \( \Gamma \) 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 \( t = \theta \log n \); 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, \( \left( 1 + \frac{t^2 + 2k}{n - t^2} \right) = 1 + o_n(1) \). We show that for every \( d \) there is some constant \( c = c(d) \), such that if \( t \) is chosen so that \( n^{1/2} \approx c \), 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 \( c \)). Thus, for large enough \( t \), \( \mathbb{E} \left[ \lambda(\Gamma)^t \right] \leq \left[ 2\sqrt{d - 1} + 0.835 \right]^t \). A standard application of Markov’s inequality then shows that

\[
\text{Prob} \left[ \lambda(\Gamma) < 2\sqrt{d - 1} + 0.84 \right] \to 1.
\]

Indeed, for \( d \geq 26 \), one can set \( n^{1/2} = e^{\sqrt{\frac{\pi}{2}} \varepsilon} \). Simple analysis shows that for \( d \geq 26 \), \( e^{\frac{\pi}{2} \sqrt{\varepsilon}} < 1 + \frac{5}{12\sqrt{d - 1}} \), so the element corresponding to \( m = 0 \) is at most \( 2\sqrt{d - 1} \cdot e^{\sqrt{\frac{\pi}{2}} \varepsilon} < 2\sqrt{d - 1} \left( 1 + \frac{5}{12\sqrt{d - 1}} \right) = 2\sqrt{d - 1} + \frac{5}{6} < 2\sqrt{d - 1} + 0.835 \). 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 \( n^{1/2} \) so that \( n^{1/2} \cdot 2\sqrt{d - 1} = 2\sqrt{d - 1} + 0.8 \) works (and see the table in Section 6.2).

**6.2 From even \( d \) to odd \( d \)**

In this subsection we derive the statement of Theorem \ref{thm2} for \( d \) odd from the now established statement for \( d \) even. We showed that for \( d \) even we have a.a.s. \( \lambda(\Gamma) < 2\sqrt{d - 1} + 0.84 \). The idea is that every upper bound applying to some value of \( d \) also applies to \( d - 1 \).

As explained in Appendix \ref{appA} by contiguity results from [GJKW02], it is enough to show the \( 2\sqrt{d - 1} + 1 \) upper bound for random graphs \( \Gamma \) in a random model denoted \( G^*_n,d \) (the result for random simple graphs than follows immediately).

**Claim 6.1.** Let \( \lambda \geq 3 \) be odd. Assume that a random \((d + 1)\)-regular graph \( \Gamma \) in the permutation model satisfies a.a.s. \( \lambda(\Gamma) < C \). Then a random \( d \)-regular graph \( \Gamma \) in \( G^*_n,d \) also satisfies a.a.s. \( \lambda(\Gamma) < C \).

**Proof.** Let \( \Gamma \) be a random \( d \)-regular graph in \( G^*_n,d \). By [GJKW02] Theorem 1.3, the permutation model \( P_{n,d+1} \) is contiguous to the distribution on \((d + 1)\)-regular graphs obtained by considering \( \Gamma \) and adding a uniformly random perfect matching \( m \). (As \( d \) is odd, the number of vertices \( n \) in \( \Gamma \) is necessarily even.) Denote by \( \Gamma \) the random graph obtained this way. It is enough to show that

\[
\lambda(\Gamma) \geq \lambda(\Gamma) - o_n(1) \text{ with probability tending to } 1 \text{ as } n \to \infty.
\]

Indeed, let \( \mu \) be the eigenvalue of \( \Gamma \) whose absolute value is largest (so \( \lambda(\Gamma) = |\mu|) \), and let \( f \in L^2(\hat{\chi}(\Gamma)) \) be a corresponding real eigenfunction with \( ||f|| = 1 \). In particular, \( \sum_{v \in V(\Gamma)} f(v) = 0 \).
and \( \sum_{v \in V(G)} f(v)^2 = 1 \). We have
\[
\lambda(\Gamma) \geq \langle A_f f, f \rangle = \langle A_r f, f \rangle + 2 \sum_{e \in m} f(e^+) f(e^-) = \mu + 2 \sum_{e \in m} f(e^+) f(e^-),
\]
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 \( 2 \sum_{e \in m} f(e^+) f(e^-) \). 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,
\[
\mathbb{E}[R] = n \cdot \frac{1}{\binom{n}{2}} \sum_{\{u, v\} \in \binom{V}{2}} f(u) f(v) = \frac{2}{n-1} s_2(f(v_1), \ldots, f(v_n)),
\]
where \( s_2 \) is the second elementary symmetric function: \( s_2(x_1, \ldots, x_n) = \sum_{i<j} x_i x_j \). Since \( s_2 = \frac{1}{2} (p_1^2 - p_2) = -\frac{1}{2} \), we conclude that \( \mathbb{E}[R] = -\frac{1}{n-1} = o_n(1) \).

Similarly,
\[
\mathbb{E}[R^2] = 4 \cdot \frac{n}{2} \cdot \frac{1}{\binom{n}{2}} \sum_{\{u, v\} \in \binom{V}{2}} f(v)^2 f(u)^2 + 8 \cdot \binom{n}{4} \cdot \frac{1}{\binom{V}{4}} \sum_{\{u, v, w, x\} \in \binom{V}{4}} f(u) f(v) f(w) f(x),
\]
\[
= \frac{4}{n-1} \sum_{\{u, v\} \in \binom{V}{2}} f(v)^2 f(u)^2 + \frac{48}{(n-1)(n-3)} \sum_{\{u, v, w, x\} \in \binom{V}{4}} f(u) f(v) f(w) f(x).
\]

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

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 \( 2\sqrt{d-1} + c \) can be obtained for 4-regular graphs in our methods.

In order to minimize \( \max \{ n^{1/4} \cdot 2\sqrt{d-1}, 2\sqrt{d-1}, \frac{4}{n^{1/4}} \} \) (see (6.1)), we choose \( n^{1/4} = \sqrt{\frac{4}{2\sqrt{d-1}}} \) 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 \( 2\sqrt{d-1} + c \) | \( n^{1/4} \) | \( d \) | Upper 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 | 5 | 4.933 | 0.933 |
| 8 | 5.868 | 0.576 | 1.109 | 7 | 5.868 | 0.969 |
| 10 | 6.646 | 0.646 | 1.108 | 9 | 6.646 | 0.989 |
| 12 | 7.323 | 0.689 | 1.104 | 11 | 7.323 | 0.998 |
| 14 | 7.928 | 0.7169 | 1.099 | 13 | 7.928 | 0.9998 |
| 16 | 8.482 | 0.7352 | 1.095 | 15 | 8.482 | 0.999 |
| 18 | 8.994 | 0.747 | 1.091 | 17 | 8.994 | 0.994 |
| 20 | 9.473 | 0.755 | 1.087 | 19 | 9.473 | 0.988 |

32
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 $\mathbb{F}_k \ast \mathbb{Z}/2\mathbb{Z} = \langle a_1, \ldots, a_k, b \mid b^2 = 1 \rangle$. 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 $\operatorname{rk}(\Omega) = |V(\Omega)| \left(\frac{d}{2} - 1\right) + 1$. However, when $m > k$, it follows from Corollary 4.10 that the corresponding term inside the max operator is $\left(\frac{d}{m} - \frac{d}{m-1}\right)^{m - 1}$, which is strictly less than $\left(\frac{d}{m - 2}\right)^{m - 2}$ (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 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.

The first inequality in our proof (as outlined in Section 2) is in bounding $\left\{ \mathbb{E} \left[ \lambda(\Gamma)^t \right] \right\}^{1/t}$ by $\left\{ \mathbb{E} \left[ \sum_{\mu \in \operatorname{Spec}(A_t) \setminus \{d\}} \mu^t \right] \right\}^{1/t}$. As long as $t = \theta(\log n)$, our loss here is bounded, and if $t \gg \log n$ we lose nothing. This is because

$$\lambda(\Gamma)^t \leq \sum_{\mu \in \operatorname{Spec}(A_t) \setminus \{d\}} \mu^t \leq n \cdot \lambda(\Gamma)^t = \left( n^{1/t} \cdot \lambda(\Gamma) \right)^t.$$ 

On the other hand, if $t \ll \log n$, one cannot get 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 $\Gamma$ 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

$$\left\{ \sum_{\mu \in \operatorname{Spec}(A_t) \setminus \{d\}} \mu^t \right\}^{1/t} \gg \left\{ \varepsilon n \left( \rho - \delta \right)^t \right\}^{1/t} \xrightarrow{n \to \infty} \infty.$$ 

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 3.1 is negligible ($t = o\left( n^{1/(2+2k)} \right)$ suffices), the upper bound our technique yields for $\mathbb{E} \left[ \sum_{\mu \in \operatorname{Spec}(A_t) \setminus \{d\}} \mu^t \right]$ is tight. In particular, for large enough $d$, and $t \approx \varepsilon \log n$ with a suitable constant $c = c(d)$,

$$\left\{ \mathbb{E} \left[ \sum_{\mu \in \operatorname{Spec}(A_t) \setminus \{d\}} \mu^t \right] \right\}^{1/t} \approx 2\sqrt{d - 1} + 0.84.$$
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 $\sum_{w \in \{X \cup X^{-1}\}^t: \pi(w) = m} |\text{Crit}(w)|$ is, in fact, the correct value (see Theorem 5.5). In the final, fifth step we can tighten our calculation 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 $\mathbb{E}\left[\lambda(\Gamma)^t\right]$ substantially. For example, in the permutation model every vertex of $\Gamma$ is isolated with probability $1/\pi^2$, so overall there are on average $1/n^2$ isolated vertices. Each such vertex is responsible to an additional eigenvalue $d$, alongside the trivial one. These rare events alone contribute $1/n^2 \cdot d^t$ 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 is exactly the correct value (see Theorem 8.5). In the final, fifth step we can tighten our calculation to come closer to the real constant (slightly smaller than 0.84), but we cannot improve it considerably.

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 $\Gamma$ contains the subgraph $\circ \longrightarrow \circ \longrightarrow \circ \longrightarrow \circ$. 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) $\Gamma$, 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 $\Gamma$, 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 $[2\sqrt{d-1}]^t \approx 3.464^t$.

Each such small graph corresponds to a few particular subgroups of $F_k$. For example, the subgraph $\circ \longrightarrow \circ \longrightarrow \circ \longrightarrow \circ$ corresponds to one of four subgroups, one of which is $x_1 \circ \longrightarrow x_2 \odot \circ \longrightarrow x_1$. One therefore needs to realize which are all these 'bad' subgroups, show their overall 'probability' is small (the average number of appearances of $H \leq F_k$ in $\Gamma$ is exactly $L_{H,F_k}$), and somehow omit their contribution to $\mathbb{E}\left[\lambda(\Gamma)^t\right]$. This would be relatively easy if our analysis of $\mathbb{E}[F_w]$ were based on $\mathbb{E}[F_w] = \sum_{M \in [w,F_k]} L_{M,F_k}$, but it is based, instead, on $\mathbb{E}[F_w] = \sum_{N \in [w,F_k]} R_{(w),N}$ (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 parallel to the notion of tangles in [FR10].

### 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 $\Gamma$, the random $n$-covering of the fixed base graph $\Omega$. The variations needed in order to establish the statement about the Markov operator are described in Section 7.1.

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

$$\lambda_\Lambda(\Gamma) < \sqrt{3} \cdot \rho + \varepsilon. \quad (7.1)$$

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
\[
E \left[ \lambda_\mu (\Gamma)^t \right] \leq \sum_{w \in CP_t(\Omega)} (E \left[ F_w \right] - 1) = \\
\leq \left( 1 + \frac{t^2 + 2\operatorname{rk}(\Omega)}{n - t^2} \right) \sum_{m=0}^{\operatorname{rk}(\Omega)} \frac{1}{n^{m-1}} \sum_{w \in CP_t^{m}(\Omega)} |\operatorname{Crit} (w)|
\]

Let \(\varepsilon > 0\). From Theorem \([4.11]\) and Lemma \([4.12]\) it follows now that for \(t\) even and large enough,
\[
E \left[ \lambda_\mu (\Gamma)^t \right] \leq \left( 1 + \frac{t^2 + 2\operatorname{rk}(\Omega)}{n - t^2} \right) \left( 1 + \operatorname{rk}(\Omega) \right) \cdot \\
\left[ \max \left\{ n^{1/t} [\rho + \varepsilon], \rho + \varepsilon, \frac{3\rho + \varepsilon}{n^{1/t}}, \frac{5\rho + \varepsilon}{(n^{1/t})^2}, \ldots, \frac{(2\operatorname{rk}(\Omega) - 1) \rho + \varepsilon}{(n^{1/t})^{\operatorname{rk}(\Omega) - 1}} \right\} \right]^t
\]

Again, to obtain a bound we must have \(t \in \theta (\log n)\), and the best bound we can obtain in this general case is obtained by choosing \(n^{1/t} \approx \sqrt{3}\), so \(\left( 1 + \frac{t^2 + 2\operatorname{rk}(\Omega)}{n - t^2} \right)^{1/t} \to 1\), and the maximal value inside the set in \([7.2]\) is then \(\sqrt{3}(\rho + \varepsilon)\). 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 \([4.4]\), we want to explain how the proof should be modified to apply to \(\lambda_{M}(\Gamma)\), the maximal absolute value of a non-trivial eigenvalue of the Markov operator on \(\Gamma\). The goal is to show that for every \(\varepsilon > 0\)
\[
\lambda_{M}(\Gamma) < \sqrt{3} \cdot \rho_{M}(\Omega) + \varepsilon
\]
asymptotically almost surely.

As we note in Appendix \([B]\), the Markov operator is given by \(B_{\Gamma} D_{\Gamma}^{-1}\), where \(B_{\Gamma}\) is the adjacency matrix and \(D_{\Gamma}\) 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 \(Q_{\Gamma}\) equals \(\frac{1}{\sqrt{\deg(u) \deg(v)}}\) times the number of edges between \(u\) and \(v\). For every path \(w\) in \(\Gamma\) we assign a weight function \(f(w)\) as follows: if \(w\) starts at \(v_0\), then visits \(v_1, v_2, \ldots, v_{t-1}\) and ends at \(v_t\), then \(f(w) = \frac{1}{\sqrt{\deg(v_0) \deg(v_1) \cdots \deg(v_{t-1}) \deg(v_t)}}\). It is easy to see that \([Q_{\Gamma}]_{n,v}\) equals the sum of \(f(w)\) over all paths \(w\) of length \(t\) from \(u\) to \(v\), and thus
\[
\sum_{\lambda \in \operatorname{Spec}(M_{\Gamma})} \lambda^t = \operatorname{tr} M_{\Gamma} = \sum_{w \in CP_t(\Gamma)} f(w).
\]

Moreover, note that when a path from the covering \(\Gamma\) projects to the base graph \(\Omega\), its weight does not change. Using this fact, we can imitate step I from Section \([2]\) to obtain, for \(t\) even,
\[
\lambda_{M}(\Gamma)^t \leq \sum_{\mu \in \operatorname{Spec}(M_{\Gamma})} \mu^t - \sum_{\mu \in \operatorname{Spec}(M_{\Omega})} \mu^t = \sum_{w \in CP_t(\Gamma)} f(w) - \sum_{w \in CP_t(\Omega)} f(w) = \\
\sum_{w \in CP_t(\Omega)} f(w) [\mathcal{F}_{\omega,n}(\sigma_1, \ldots, \sigma_k) - 1].
\]
The second and third steps remain the same, obtaining
\[
E \left[ \lambda_M (\Gamma)^t \right] \leq \left( 1 + \frac{t^{2+2rk(\Omega)}}{n-t^2} \right) \sum_{m=0}^{rk(\Omega)} \frac{1}{n^{m-1}} \sum_{w \in \mathcal{CP}_m(\Omega)} f(w) |\text{Crit}(w)|.
\]

The next modification needs take place in the fourth step, where instead of bounding \( \sum_{w \in \mathcal{CP}_1(\Omega): \pi(w)=m} |\text{Crit}(w)| \), one needs to bound \( \sum_{w \in \mathcal{CP}_1(\Omega): \pi(w)=m} f(w) |\text{Crit}(w)| \). But the exact same proofs work if we merely replace \( \rho_A (\Omega) \) with \( \rho_M (\Omega) \). Theorem 4.11 becomes
\[
\limsup_{t \to \infty} \left[ \sum_{w \in \mathcal{CP}_1(\Omega)} f(w) |\text{Crit}(w)| \right]^{1/t} \leq (2m - 1) \cdot \rho_M (\Omega). \tag{7.4}
\]

and likewise, Lemma 4.12 becomes
\[
\limsup_{t \to \infty} \left[ \sum_{w \in \mathcal{CP}_0(\Omega)} f(w) \right]^{1/t} = \rho_M (\Omega).
\]

Similarly, the definition of \( \beta_t (N) \) (preceding Claim 4.13) should be modified to
\[
\beta_t (N) = \sum_{w \in \mathcal{CP}_1(\Omega): \langle w \rangle \preccurlyeq \text{alg} N} f(w)
\]

and in the proof of Claim 4.13 one should use the fact that \( f(w) \) does not change when the closed path \( w \) is being cyclically rotated. Finally, in the proof of Proposition 4.14 we sometimes replace a path with its inverse and use the symmetry of the operator. This is the reason for working with \( Q_\Gamma \) rather than with \( M_\Gamma \). Also, the coefficient \( |V(\Omega)| \) 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 \( \rho_M (\Omega) \) instead of \( \rho_A (\Omega) \), 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 4.11 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 \( F_k \).

**Theorem 8.1.** \cite{PW14} For every \( k \geq 3 \), let \( p_k (t) \) denote the number of primitive words of length \( t \) in \( F_k \). Then,
\[
\lim_{t \to \infty} \sqrt[3]{p_k (t)} = 2k - 3.
\]

For \( F_2 \) it is known that this exponential growth rate equals \( \sqrt[3]{3} \) \cite{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

$$c_{k,m} (t) = |\{w \in F_k \mid |w| = t, \pi(w) = m\}|.$$ 

Then,

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

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 $F_k$ 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 $(\sqrt{2k-1}, 2m-1)$.

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 $H = \langle S \rangle$. 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 $\sim \frac{4}{2}$ 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 $\sim \frac{4}{2}$. Consider the word $w = w(u) = x_1^2 x_2^2 \ldots x_{m-1}^2 w^2$. Obviously, the growth rate of the number of possible $w$'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 (Pud14b, Thm 1.1). (In fact, it follows from [PP14, 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 $w = x_3 x_2^2 x_1^2 x_3$, we have $\pi(w) = 2$ because $w$ is not a proper power yet is not primitive in $\langle x_1, x_2^2 x_3 \rangle$. 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 $w$ 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 $\frac{4}{2}$. (By (4.2), the total number of words of length $t$ with an algebraic extension $N$ of rank $\tilde{m}$ and $\delta t$ edges in $\Gamma_X (N)$, for some $\delta < \frac{4}{2}$, 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 $\frac{4}{2}$ edges. In particular, each such $w = w(u)$ traces twice some topological edge in $\Gamma_X (N)$ of length at least $\frac{1}{2(2m-1)}$. 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

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

\footnote{That primitive words in $F_k$ are negligible in this sense follows from the earlier results [BV02, BMS02, Thm 10.4] and [Shp05], where the exponential growth rate from Theorem 8.1 is shown to be $2k - 2 - o_k (1)$.}
If, in addition, $m \geq 2$, the same argument as above works also for $w = w(u) = x_1^2 x_2^2 \ldots x_m^2 u^2$ which is of arbitrary odd length. Thus, $\lim_{t \to \infty} c_{k,m}(t)^{1/t} = \sqrt{2k-1}$.

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 $\frac{t}{4}$.

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 $m \in \{0, 1, 2, \ldots, k, \infty\}$. Let

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

Then for $m = 0$ we have

$$\lim_{t \to \infty} b_{k,0}(t)^{1/t} = 2\sqrt{2k-1}.$$ 

For $m \in \{1, \ldots, k\}$,

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

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 $(X \cup X^{-1})^t$ 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 $\sum_{w \in (X \cup X^{-1})^t : \pi(w) = m} |\text{Crit}(w)|$.

**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 [Pud14b, Lemma 6.8]), $\pi(x_1^2 x_2^2 \ldots x_m^2) = \pi(x_1^3 x_2^3 \ldots x_m^3) = m$. 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 [Pud14a], (here a bit more elaborated results from [Pud14a], not mentioned in Theorem 4.4 are required).

The result 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 [PP14]) 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 statement of Theorems 4.11 and 1.4? This may require a further 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]).

• 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, Lub13]). 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

We would like to thank Nati Linial and Ori Parzanchevski for their valuable suggestions and useful comments. We would also like to thank Miklós Abért, Noga Alon, Itai Benjamini, Ron Rosental and Nick Wormald for their beneficial comments.

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 $G_{n,d}^*$. It is shown [GJKW02, Theorem 1.3] that $G_{n,d}^*$ is contiguous to the permutation model $P_{n,d}$ (for $d$ even). If $\Gamma$ is a random $d$-regular graph in $G_{n,d}^*$, the event that $\Gamma$ 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\(^\dagger\). 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.

Random $d$-regular bipartite graphs

As an immediate corollary from Theorem 1.9 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 $C_{n,\Omega}$ (for $\Omega$ 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.

\(^\dagger\)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.
(asymptotically it is $e^{-\frac{3}{d}}$). 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 $\Omega$ in the model $C_{n,\Omega}$, where a uniform random permutation is generated for every edge of $\Omega$. An equivalent model is attained if we cover some spanning tree of $\Omega$ 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 $\Omega$ by the action of $\pi_1(\Omega, \otimes)$ on the fiber $\{\otimes\} \times [n]$ above $\otimes$ shows that $C_{n,\Omega}$ 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 $C_{n,\Omega}$ if $\ker(\Omega) \geq 2$ (note that the random covering $\Gamma$ in $C_{n,\Omega}$ is a.a.s. connected provided that $\ker(\Omega) \geq 2$). Indeed, the only difference is that in the new model, the probability of every connected graph $\Gamma$ from $C_{n,\Omega}$ is proportional to $\frac{1}{\text{Aut}(\Gamma)}$. When $\ker(\Omega) \geq 2$, it seems that a.a.s. $|\text{Aut}(\Gamma)| = 1$, 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 $\Gamma$ 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 $p_f(\Gamma)$. (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 $\Omega$. Interestingly, Lubotzky and Nagnibeda [LN98] showed that there exist such $T$’s with a minimal quotient $\Omega$ 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 $\Omega$, 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 base graph $\Omega$ 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 $\Gamma$ 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]),

\[\dagger\] 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.

\[\ddagger\] 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.
all of which are bounded\(^{\dagger}\) self-adjoint operators and so have real spectrum:

1. The **adjacency operator** \( A_\Gamma \) on \((\ell^2 (V(\Gamma)),1)\)

\[
(A_\Gamma f)(v) = \sum_{w \sim v} f(w)
\]

If \( \Gamma \) is finite this operator is represented in the standard basis by the adjacency matrix, and its spectral radius is the *Perron-Frobenius* eigenvalue \( p_f(\Gamma) \). The spectrum in this case is

\[
p_f(\Gamma) = \lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_m \geq -p_f(\Gamma),
\]

and the spectral gap is \( p_f(\Gamma) - \lambda(\Gamma) \), where \( \lambda(\Gamma) = \max \{ \lambda_2, -\lambda_m \} \). The spectrum of \( A_\Gamma \) was studied in various works, for instance [Gre95, LN98, Fri03, LP10].

2. The **averaging Markov operator** \( M_\Gamma \) on \((\ell^2 (V(\Gamma)), \deg \cdot)\)

\[
(M_\Gamma f)(v) = \frac{1}{\deg(v)} \sum_{w \sim v} f(w)
\]

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 \( \Gamma \) 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 \( 1 - \mu(\Gamma) \) here \( \mu(\Gamma) = \max \{ \mu_2, -\mu_m \} \). Up to a possible affine transformation, the spectrum of \( M_\Gamma \) is the same as the spectrum of the simple *random walk operator* \( (A_\Gamma D_\Gamma^{-1}) \) 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** \( \Delta_\Gamma^+ \) on \((\ell^2 (V(\Gamma)),1)\):

\[
(\Delta_\Gamma^+ f)(v) = \deg(v) f(v) - \sum_{w \sim v} f(w)
\]

The Laplacian equals \( D_\Gamma - A_\Gamma \), where \( D_\Gamma \) is the diagonal operator \((D_\Gamma f)(v) = \deg(v) \cdot f(v)\). The entire spectrum is non-negative, with 0 corresponding to locally-constant functions when \( \Gamma \) 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 [AMS85].

For a regular graph \( \Gamma \), 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 \( A_\Gamma \) and of \( M_\Gamma \). At this point we do not know how to extend our results to the Laplacian operator \( \Delta_\Gamma^+ \).

\(^{\dagger}\) All operators considered here are bounded provided that the degree of vertices in \( \Gamma \) is bounded. This is the case in all the graphs considered in this paper.

\(^{\ddagger}\) Here, \((\ell^2 (V(\Gamma)),1)\) stands for \( \ell^2 \)-functions on the set of vertices \( V(\Gamma) \) with the standard inner product:

\[
\langle f, g \rangle = \sum_v f(v) g(v);
\]

In the summation \( \sum_w \), each vertex \( w \) is repeated with multiplicity equal to the number of edges between \( v \) and \( w \).

\(^{\dagger\dagger}\) Occasionally, the spectral gap is taken to be \( p_f(\Gamma) - \lambda_2(\Gamma) \).

\(^{*}\) Here, \((\ell^2 (V(\Gamma)), \deg \cdot)\) stands for \( \ell^2 \)-functions on the set of vertices \( V(\Gamma) \) with the inner product:

\[
\langle f, g \rangle = \sum_v f(v) g(v) \deg(v).
\]
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

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

where $\text{vol}_{pf}(S) = \langle 1_S, pf(\Gamma) \rangle$ and $pf(\Gamma)$ is the (normalized) Perron-Frobenius eigenfunction. This is particularly useful in the $C_n, \Omega$ model since the $pf(\Gamma)$ is easily obtained from the Perron-Frobenius eigenfunction of $\Omega$ by

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

In the $d$-regular case, this amounts to the usual mixing lemma: $|E(S, T) - d \frac{|S| \cdot |T|}{m}| \leq \lambda(\Gamma) \sqrt{|S| \cdot |T|}.$

If one takes $T = V \setminus S$, one can attain a bound on the Cheeger constant of $\Gamma$ (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 $\text{deg}(S)$ to denote the sum of degrees of the vertices in $S$, then

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

Moreover, consider the conductance of $\Gamma$

$$\phi(\Gamma) = \min_{0 \neq S \subseteq V, \text{deg}(S) \leq \frac{2 \text{deg}(V)}{|S|}} \frac{|E(S, V \setminus S)|}{\text{deg}(S)}.$$ 

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 $\Gamma$, defined as

$$h(\Gamma) = \min_{0 \neq S \subseteq V, |S| \leq \frac{|V|}{2}} \frac{|E(S, V \setminus S)|}{|S|}.$$ (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 $\Delta_k^+ \Gamma$ 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] A. Amit, N. Linial, and J. Matoušek, 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, λ1, 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 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, Amer Mathematical Society, 1997.

[DJPP11] I. Dumitriu, T. Johnson, S. Pal, and E. Paquette, Functional limit theorems for random regular graphs, Arxiv preprint arXiv:1109.4094 (2011).

[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 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 symposium 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 (Cortona, 1997), Sympos. Math 39 (1999), 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.

[Lub13] Alexander Lubotzky, *Ramanujan complexes and high dimensional expanders*, arXiv preprint arXiv:1301.1028 (2013).

[Lyons12] 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 eigenvalues 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.

[MS13] Adam Marcus, Daniel A Spielman, and Nikhil Srivastava, Interlacing families i: Bipartite ramanujan graphs of all degrees, arXiv preprint arXiv:1304.4132 (2013).

[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.

[PP14] Doron Puder and Ori Parzanchevski, Measure preserving words are primitive, Journal of American Mathematical Society (2014), in press.

[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.

[Pud14a] Doron Puder, An extension of the cogrowth formula, preprint, 2014+.

[Pud14b] ______, Primitive words, free factors and measure preservation, Israel Journal of Mathematics (2014), in press.

[PW14] Doron Puder and Conan Wu, Growth of the primitives elements in free groups, Journal of London Mathematical Society (2014), in press.

[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.