SHALOM

December 11, 2014

Abstract

This thesis contains most of the fruits of my PhD mathematical research. Two main lines of research show up throughout the five manuscripts composing this thesis: free groups and expansion of random graphs. The two are closely related via the notion of word maps.

In my perspective, the most valuable contributions of my research so far have been in two fields of study: measure-theoretic characterization of words, and expansion of random graphs. In the former, I have studied the measure induced on finite groups by words. Namely, fix some word w in F_k, the free group on k generators $x_1, \ldots, x_k$. This word induces a measure on every finite group via the word map $w : G^k \to G$ (here G^k is the Cartesian product of G) and a push forward of the uniform measure on G^k. (Put differently, for each $1 \leq i \leq k$, substitute x_i with an independent, uniformly distributed random element of G and evaluate the product defined by w to obtain a random element in G.) It is an easy observation that primitive words, namely words belonging to some basis of F_k, induce the uniform measure on every finite group G. Several mathematicians have conjectured that this property actually characterizes primitive elements, i.e. that a word which induces the uniform measure on every finite group is primitive. In Chapter 1 I prove the conjecture for F_2, and in Chapter 2 we prove the conjecture in full.

The second field of study, that of expansion of random graphs, culminated in the work presented in Chapter 3. Here, I use, inter alia, the results from the first two chapters and present a new approach to showing that random graphs are nearly optimal expanders. This new approach applies to both regular and irregular random graphs. This work proves a slightly weakened version of the generalized second eigenvalue conjecture by Alon and Friedman. In the most general setting of this conjecture, it is the best result to date.

Using my new understanding of primitive words, I also studied what a common primitive word looks like. In Chapter 4 we describe the structure of generic primitive words. This also solves a question about the growth of the set of primitive elements, which was open for more than a decade.

Finally, as a side to our work in Chapter 2, we studied a natural conjecture of Miasnikov, Ventura and Weil about algebraic extensions in free groups. We managed to refute it, and this is the main result of Chapter 5.

All five manuscripts composing this thesis were published (at least in electronic format). Here are the bibliographical details:

- **Chapter 1:** Doron Puder, *Primitive words, free factors and measure preservation*, Israel Journal of Mathematics, 201 (1), 2014, pp 25-73. DOI: 10.1007/s11856-013-0055-2
- **Chapter 2:** Doron Puder and Ori Parzanchevski, *Measure preserving words are primitive*, Journal of the American Mathematical Society, 28 (1), 2015, pp 63-97. DOI: 10.1090/S0894-0347-2014-00796-7
- **Chapter 3:** Doron Puder, *Expansion of random graphs: new proofs, new results*, Inventiones Mathematicae, in press. DOI: 10.1007/s00222-014-0560-x
- **Chapter 4:** Doron Puder and Conan Wu, *Growth of primitive elements in free groups*, Journal of the London Mathematical Society, 90 (1), 2014, pp 89-104. DOI: 10.1112/jlms/jdu009

- **Chapter 5:** Ori Parzanchevski and Doron Puder, *Stallings graphs, algebraic extensions and primitives in F_2,* Mathematical Proceedings of the Cambridge Philosophical Society, 157 (1), 2014, pp 1-11. DOI: 10.1017/S0305004114000097

A Letter of Contribution

This thesis consists of five different manuscripts:

(1) "Primitive words, free factors and measure preservation" was published in the Israel Journal of Mathematics in 2014. I am the only author of this work.

(2) "Measure preserving words are primitive" is to be published in the Journal of the American Mathematical Society in 2015. This work is mainly mine, and is the culmination of a long research project (one of the milestones of which is the first manuscript in this thesis). However, at the final stages I did get some small help from a (then) fellow PhD student in the Hebrew University, Ori Parzanchevski. After giving it some consideration, Ori and I decided to publish it as a joint paper. We also decided to balance this in some ways: we put my name first (although non-alphabetical order in authors names is not common in Mathematics), all talks on the paper were given by me, etc.

(3) "Expansion of random graphs: new proofs, new results" was accepted for publication in Inventiones Mathematicae and already appeared online. I am the only author of this work.

(4) "Growth of primitive elements in free groups" was published in the Journal of the London Mathematical Society in 2014. This work is joint with Conan Wu, a (then) Phd student from Princeton University who visited the Hebrew University for one semester. We both contributed to this work roughly equally.

(5) "Stallings graphs, algebraic extensions and primitives in F_2" was published in Mathematical Proceedings of the Cambridge Philosophical Society in 2014. This is joint work with Ori Parzanchevski (again, a then fellow PhD student). We consider our contributions to this work as roughly equal.

Contents

Prologue

Most of my PhD research was devoted to one line of research involving several major questions. These questions belong to the fields of free groups and of expansion of random graphs. My thesis contains five manuscripts describing the main achievements of this study. Chapters 2 and 3 are the culmination of this research, the other chapters describing either partial and supportive results (Chapter 1), or related results that do not belong to the core of the research (Chapters 4 and 5).

The seeds of this main line of research can be traced in my M.Sc. thesis. Together with my advisor, Nati Linial, we studied expansion properties of random graph coverings. We realized that this question can be approached via a question regarding fixed points in random permutations whose distribution is induced by some fixed formal word (see details below). The fruits of this research appear in [LP10]. In the beginning of my PhD studies, I came to realize the right algebraic interpretation of the fixed-points question. Although the original goal was the study of expansion of random graphs, it turned out that the means, i.e. the study of the distribution of random permutations according to fixed formal words, is of no lesser importance. Let me now give a detailed mathematical description of the questions and results.

Contents

0.1 Primitive words and measure preservation

Let $\mathbf{F}_k$ be the free group on k generators, and let $w \in \mathbf{F}_k$. Associated with w and any group G is the word map $w : G^k \to G$, where G^k is the Cartesian product of G. The word map is defined by substitutions, e.g. the word map associated with $w = x_1 x_2 x_1^{-1} \in F_2$ is $w(g_1, g_2) = g_1 g_2 g_1^{-1}$. Via this word map and the push forward of the uniform (Haar) measure on G^k, w induces a measure on every finite (compact) group G. If the push-forward of the uniform/Haar measure on G^k via the word map w yields again the uniform/Haar measure on G, we say that w is *measure preserving.*

Primitive words, namely, words belonging to some basis (free generating set) of $\mathbf{F}_k$, play a special role here. It is an easy observation that if w is primitive, it induces the uniform measure on every finite or compact group, namely

Observation (Observation 2.1.2)**.** *A primitive word is measure preserving.*

But are there any other words with this property? Several mathematicians, most notably from Jerusalem, came independently to the conjecture that the answer is negative, namely, that a word which induces the uniform measure on every finite group is necessarily primitive. From private conversations we know that this has occurred to Tsachik Gelander, Michael Larsen, Alex Lubotzky and Aner Shalev. It also occurred to Nati Linial and myself during our joint research [LP10], and to Alon Amit and Uzi Vishne [AV11]. Each among these researchers was led to the conjecture

by a different motivation: the dynamics of $\mathrm{Aut}(\mathbf{F}_k)$, profinite topology and decidability problems, expansion in random graphs, and word maps in finite simple groups.

In the beginning of my PhD I realized that I may have a road-map to cope with this challenging question. In order to prove the conjecture, one needs to find, for every non-primitive word $w \in \mathbf{F}_k$, a finite or compact group where w does not induce the uniform (Haar) measure. I suspected that this can be resolved using the symmetric groups S_n, namely that every non-primitive word $w \in \mathbf{F}_k$ induces non-uniform distribution on S_n for some n.

Moreover, I suspected it was enough to consider the *expected number of fixed points* in a random permutation σ distributed according to w. Let us denote by G_w a random element in the group G whose distribution is induced by w. Figure 0.1.1 describes the following circle of implications: A uniformly random permutation in S_n has exactly one fixed point on average. Clearly, then, if w is measure preserving, then the expected number of fixed points in $(S_n)_w$ is one. It seemed to me that the converse also holds. Namely, if w is non-primitive, then $\mathbb{E}\left[\mathrm{FixedPoints}\left((S_n)_w\right)\right] \neq 1$ for some n.

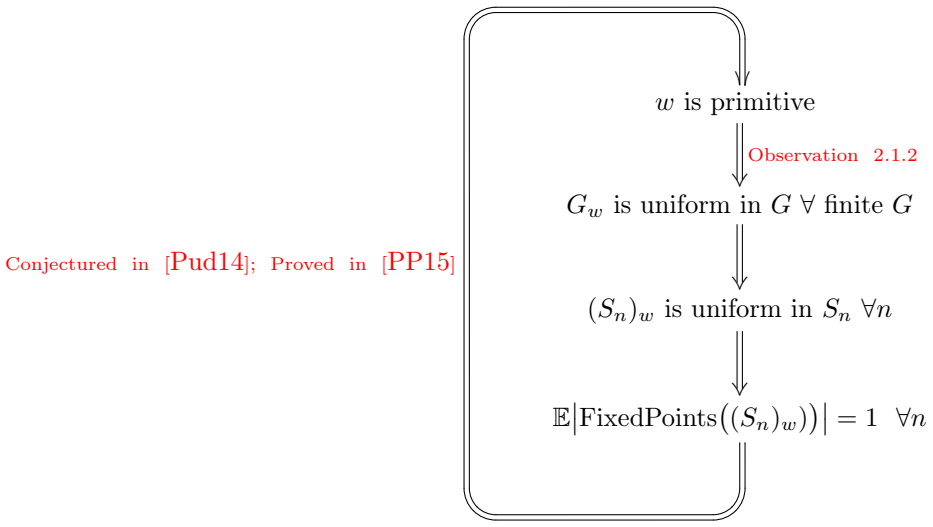


Figure 0.1.1: The relation between primitivity and the expected number of fixed points in random permutations.

In fact, the conjectural picture I described in [Pud14] (Chapter 1) is more elaborated. It says that the expected number of fixed points of $(S_n)_w$ is related to an algebraic property of w, which I called its *primitivity rank*. If w is primitive in $\mathbf{F}_k$, then it is also primitive in every subgroup $H \leq \mathbf{F}_k$ containing it (Claim 1.2.5; For example, the single-letter word x_1 belongs to some basis of the subgroup H whenever $x_1 \in H$). However, if w is not primitive in $\mathbf{F}_k$, it may be either primitive or non-primitive in a given subgroup containing it. For example, every $w \neq 1$ is primitive in the subgroup $\langle w \rangle$. But what is the 'smallest' or 'simplest' subgroup manifesting the non-primitivity of w? Concretely,

Definition (Definition 1.1.7)**.** The **primitivity rank** of $w \in \mathbf{F}_k$, denoted $\pi(w)$, is

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

If no such J exists, $\pi(w) = \infty$. A subgroup J for which the minimum is obtained is called w**-critical**.

(For examples and more details see the discussion following Definition 1.1.7.) Indeed, I noticed there was a close relation between this algebraic feature of w and the expected number of fixed points in $(S_n)_w$. The following statement was conjectured and partially proved in [Pud14], and later fully proved in [PP15]:

Theorem (Conjecture 1.1.7, Theorem 2.1.8)**.** *The average number of fixed points in* $(S_n)_w$ *is*

$$1 + \frac{|\mathrm{Crit}(w)|}{n^{\pi(w)-1}} + O\left(\frac{1}{n^{\pi(w)}}\right),$$

where $\mathrm{Crit}(w)$ *is the set of* w*-critical subgroups.*

In [Pud14] this statement was shown to hold for words in $\mathbf{F}_2$. The techniques there are specialized for this case and could not be generalized to free groups of higher rank. New machinery was later developed in [PP15] to show the general case. This machinery included Möbius analysis on posets of subgroups of $\mathbf{F}_k$, geometric interpretation of the function describing the expected number of fixed points and its Möbius derivations, understanding the role of algebraic extensions in free groups and using the combinatorics of Stirling numbers (see [PP15], here in Chapter 2). The following table summarizes the new categorization of the elements in a free group implied by the primitivity rank and Theorem 2.1.8:

$\pi(w)$	Description of w	$\mathbb{E}[\#\mathrm{FixedPoints}((S_n)_w)]$
0	$w = 1$	n
1	w is a power	$1 + \lvert\mathrm{Crit}(w)\rvert + O\left(\frac{1}{n}\right)$
2	E.g. $[x_1, x_2], x_1^2 x_2^2$	$1 + \frac{\lvert\mathrm{Crit}(w)\rvert}{n} + O\left(\frac{1}{n^2}\right)$
3		$1 + \frac{\lvert\mathrm{Crit}(w)\rvert}{n^2} + O\left(\frac{1}{n^3}\right)$
$\vdots$		$\vdots$
k	E.g. $x_1^2 \dots x_k^2$	$1 + \frac{\lvert\mathrm{Crit}(w)\rvert}{n^{k-1}} + O\left(\frac{1}{n^k}\right)$
∞	w is primitive	1

Table 1: Primitivity rank and the average number of fixed points.

In particular, if $w \in \mathbf{F}_k$ is not primitive, then $\pi(w) < \infty$. Hence, for large enough n, the expected number of fixed points in $(S_n)_w$ is more than one, and so $(S_n)_w$ is not uniformly distributed. This proves that the main conjecture is true:

Theorem (Theorem 2.1.1)**.** *Measure preserving words are primitive.*

An important step in proving these statements was to generalize them to subgroups of F_k: is it true that a subgroup $H \leq F_k$, freely generated by $h_1, \dots, h_r$, is a free factor of F_k if and only if the tuple of maps $(h_1, \dots, h_r) : G^k \to G^r$ induces uniform measure on G^r for every finite group G? In [Pud14] I established this conjecture for subgroups of rank $\geq k - 1$, which in particular yields the original conjecture for F_2, and in [PP15] this generalized conjecture was proven in full. This is shown using results on the expected number of fixed points induced not only by single words but, more generally, by subgroups (see the full statement of Theorem 2.1.8 and the discussion preceding it).

Note that these results provide a new criterion (and a straightforward algorithm) to detect primitive words in $\mathbf{F}_k$. Moreover, we show (Proposition 2.1.6), that a word w of length $\ell > 0$ is primitive if and only if $\mathbb{E}[\#\mathrm{FixedPoints}((S_n)_w)] = 1$ for $n \leq \ell$, which yields a more effective criterion. Detecting primitives is a non-trivial task. The first algorithm was given by Whitehead

[Whi36a] (see Section 4.2), and it is still the most efficient algorithm to identify primitive words, complexity-wise. However, an important ingredient in our proof of the above results is a new algorithm to this goal of identifying primitives. This algorithm, based on Stallings graphs, is interesting for its own sake, and is one of the main results of [Pud14] (Theorem 1.1.1 here).

In working on this project we made a considerable effort to understand the combinatorics of Stallings graphs. This has led us to find new, self-contained proofs for classic theorems regarding primitives and bases of F_2, and also to construct a counterexample for a conjecture of Alexei Miasnikov, Enric Ventura and Pascal Weil, which concerns algebraic extensions in free groups [MVW07]. This is the content of [PP14] (Chapter 5).

0.2 Expansion of Random Graphs

As mentioned, my original interest in the expected number of fixed points in $(S_n)_w$ arose from the study of expansion of random graphs. Let Γ be a d-regular graph on n vertices, and let

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

be the spectrum of its adjacency matrix. We denote by $\lambda = \lambda(\Gamma)$ the largest absolute value of a non-trivial eigenvalue, namely $\lambda = \max(\lambda_2, -\lambda_n)$. It is well known that the smaller λ is, the better expander Γ is (in many aspects: its Cheeger constant, its pseudo-randomness properties expressed by the expander mixing lemma, the mixing time of a random walk, etc.). But λ cannot be too small: by the Alon-Boppana bound, it is at least $2\sqrt{d-1} - o_n(1)$ [Nil91]. For this reason, graphs for which $\lambda \leq 2\sqrt{d-1}$ are considered optimal expanders. They are named *Ramanujan* graphs.

On the other extreme, λ can be as large as d (e.g. for bipartite graphs, where $\lambda_n = -d$). However, in [Alo86], Alon conjectured that a *random* d-regular graph Γ should be almost Ramanujan, in the following sense: for every $\varepsilon > 0$, $\lambda < 2\sqrt{d-1} + \varepsilon$ asymptotically almost surely (a.a.s.) as $n \to \infty$. Friedman famously presented a proof of this conjecture [Fri08].

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_d. The number $2\sqrt{d-1}$ is nothing but the spectral radius of the adjacency operator on $\ell^2(V(T_d))$. It is therefore natural to measure the spectrum of any graph against the spectral radius of its covering tree. Several authors call graphs whose non-trivial spectrum is bounded by this value Ramanujan, thus 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 [Gre95].

Ideally, one would like to extend Alon's conjecture on almost-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 quotient Ω which is not Ramanujan. All other finite quotients of T are then coverings of Ω, and inherit the "bad" eigenvalues of this quotient. Such examples invalidate the obvious analogue of Alon's conjecture.

But what if we ignore this few, fixed, "bad" eigenvalues originated in the minimal quotient Ω and focus only on the remaining, "new" eigenvalues of each larger quotient? In this sense, a generalized version of Alon's conjecture is indeed plausible. Friedman [Fri03] put it as follows: Fix an arbitrary finite base graph Ω, and consider the spectrum of Γ, a random n-sheeted covering of Ω. Let λ be the largest absolute value of a non-trivial (namely, new) eigenvalue of Γ. Friedman conjectured that for every $\varepsilon > 0$ a.a.s. $\lambda < \rho(\Omega) + \varepsilon$, where $\rho(\Omega)$ is the spectral radius of the universal covering (tree) of Ω. (For d even, the d-regular case falls into this framework: all d-regular graphs are coverings of the bouquet with $\frac{d}{2}$ loops.)

In the same paper, Friedman also showed that $\lambda \leq O\left(\rho(\Omega)^{1/2}\, \mathfrak{pf}(\Omega)^{1/2}\right)$ a.a.s., where $\mathfrak{pf}(\Omega)$ is the Perron-Frobenius eigenvalue of Ω (and the analogue of d in the regular case). During my

M.Sc. studies, together with my supervisor, Nati Linial, we improved this result and proved an upper bound of $3\rho(\Omega)^{2/3}\,\mathfrak{pf}(\Omega)^{1/3}$ [LP10].

The seeds of this work grew later to much stronger results that appear in [Pud15a] (Chapter 3). In this work, I rely, among other things, on the abovementioned quantitative results from [PP15] (Theorem 2.1.8), and prove a slightly weakened version of Friedman's generalized conjecture:

Theorem (Theorem 3.1.4)**.** *Let Ω be an arbitrary finite connected graph, and let Γ be a random n-sheeted covering of Ω. Then for every $\varepsilon > 0$,*

$$\lambda(\Gamma) < \sqrt{3}\cdot\rho(\Omega) + \varepsilon$$

asymptotically almost surely.

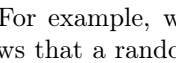
This proves the conjecture up to a (small) multiplicative constant, and is the best known result in the irregular case.

In the special case where Ω is d-regular (but not necessarily a bouquet of $\frac{d}{2}$ loops), there were a few more attempts to attack the problem. Lubetzky, Sudakov and Vu showed in [LSV11] that a.a.s. $\lambda \le C\cdot\rho(\Omega)\cdot\log\rho(\Omega)$ (with an unspecified constant C), and later Addario-Berry and Griffith showed that a.a.s. $\lambda < 265{,}000\cdot\rho(\Omega)$ [ABG10] (here, of course, $\rho(\Omega) = 2\sqrt{d-1}$). These bounds are substantially improved in [Pud15a], where I obtain a nearly optimal bound and show:

Theorem (Theorem 3.1.5)**.** *Let Ω be a finite connected d-regular graph ($d \ge 3$) and let Γ be a random n-sheeted covering of Ω. Then*

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

asymptotically almost surely.

For example, when Ω is  (a 2-vertex graph with d edges between them), my result shows that a random d-regular bipartite graph is a.a.s. nearly Ramanujan in the sense that all its eigenvalues except for $\pm d$ fall inside $\left[-2\sqrt{d-1}-0.84, 2\sqrt{d-1}+0.84\right]$. (For a table summarizing these results with comparison to former ones see the table on Page 84.)

In particular, my work provides a new proof to a slightly weakened version of Alon's original conjecture (Friedman's Theorem): in a random d-regular graph, the second largest absolute value of an eigenvalue is a.a.s. at most $2\sqrt{d-1}+1$. It is important to stress that the proof in [Pud15a] is very different from Friedman's 100-page long proof of the more accurate bound, and to my judgment it is significantly simpler. The new proof is composed of five well-defined steps, whose outline is explained in Section 3.2. This might be meaningful for many questions that remain open even for the d-regular case (see the Epilogue beginning on page 152). Of course, our approach also has the advantage of applying to a more general model of random graphs (the generalized conjecture).

To complete the picture let me also briefly explain the connection between expansion of random graphs and the study of fixed points of random permutations. I demonstrate this connection via the permutation model for random d-regular graphs with d even, but as shown in Chapter 3, this connection extends to the more general model of random n-sheeted coverings of a fixed based graph.

Indeed, fix $d \ge 4$ even, and let Γ be a d-regular random graph on n vertices in the permutation model. Namely, the vertices of Γ are labeled $1,\ldots,n$. choose $\frac{d}{2}$ uniformly random permutations $\sigma_1,\ldots,\sigma_{d/2} \in S_n$ and connect the vertex i with the vertex $\sigma_j(i)$ with a directed edge labeled by the symbol x_j, for every $1 \le i \le n$ and $1 \le j \le \frac{d}{2}$ (loops and multiple edges are allowed). Ignoring the orientation of the edges, this yields a d-regular graph.

The spectrum of Γ may be analyzed by counting closed paths. More concretely, denote by $\mathcal{CP}_t(\Gamma)$ the set of closed paths of edge-length t in Γ. If $\mathrm{Spec}(\mathrm{A}_\Gamma)$ denotes the multiset of eigenvalues of A_Γ, the adjacency matrix of Γ, then for every $t \in \mathbb{N}$,

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

For each $p \in \mathcal{CP}_t(\Gamma)$ one can trace the labels of the edges it traverses. If it goes through an edge corresponding to the permutation σ_j in the right orientation use the label x_j, and if the orientation is reversed, use the label x_j^{-1}. This yields a formal word, not necessarily reduced, in $\left(X \cup X^{-1}\right)^t$ where $X = \left\{x_1, \ldots, x_{d/2}\right\}$. This word can also be thought of as an element of $\mathbf{F}(X) \cong \mathbf{F}_{d/2}$. Now, instead of directly counting closed paths in Γ, one can go over all words $w \in \left(X \cup X^{-1}\right)^t$ and count how many closed paths in $\mathcal{CP}_t(\Gamma)$ correspond to this w. It is easy to see that for each word $w \in \mathbf{F}_{d/2}$, the expected number of closed paths corresponding to w is exactly the expected number of fixed points in $(S_n)_w$ (see Section 3.2 for details).

0.3 The growth of primitivity-rank categories

The proof of the main results about expansion of random graphs in [Pud15a] (Chapter 3), required understanding the size of the set of all words of a given primitivity rank. Namely, for every $m \in \{0, 1, \ldots, k\}$, one needs to obtain the exponential growth rate of the number of words of length N in $\mathbf{F}_k$ whose primitivity rank is m. This is done in Sections 3.4 and 3.8, and is based on analysis of Stallings graphs.

The remaining case which is not required for the results in [Pud15a], is the case of infinite primitivity rank, namely, the case of Primitive words. In fact, this case has attracted attention for several decades and is the content of a well known open question attributed to M. Wicks, which appeared formally in a list of open questions in Combinatorial and Geometric group theory [BMS02a, Problem F17].

It was previously known that in $\mathbf{F}_2$ the exponential growth rate of the primitive is $\sqrt{3}$, but the exact value was not known for $k \geq 3$. In [PW14] (Chapter 4) we answer this question completely. Let $P_{k,N}$ denote the set of primitive words of length N in $\mathbf{F}_k$.

Theorem (Theorem 4.1.1)**.** *For all $k \geq 3$,*

$$\lim_{N\to\infty} \sqrt[N]{|P_{k,N}|} = 2k - 3.$$

Moreover, there are positive constants c_k and C_k such that

$$c_k \cdot N \cdot (2k-3)^N \leq |P_{k,N}| \leq C_k \cdot N \cdot (2k-3)^N .$$

In fact, we showed what a generic primitive word looks like. We established the somewhat surprising fact that generic primitive words are words that are 'obviously' primitive, i.e. contain one of the letters exactly once. More accurately, up to conjugation, a random primitive word of length N contains one of the letters exactly once asymptotically almost surely (as $N \to \infty$). Here we used completely different techniques than those used for other primitive ranks in [Pud15a]. Alongside the results from [PP15], our main tool was a meticulous analysis of Whitehead's algorithm to identify primitive words [Whi36a].

References

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

[Alo86] N. Alon. Eigenvalues and expanders. *Combinatorica*, 6(2):83–96, 1986.

[AV11] A. Amit and U. Vishne. Characters and solutions to equations in finite groups. *Journal of Algebra and Its Applications*, 10(4):675–686, 2011.

[BMS02a] G. Baumslag, A. Myasnikov, and V. Shpilrain. Open problems in combinatorial group theory. *Contemporary Mathematics*, 296:1–38, 2002.

[Fri03] J. Friedman. Relative expanders or weakly relatively ramanujan graphs. *Duke Mathematical Journal*, 118(1):19–35, 2003.

[Fri08] J. Friedman. *A proof of Alon's second eigenvalue conjecture and related problems*, volume 195 of *Memoirs of the AMS*. AMS, september 2008.

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

[LN98] A. Lubotzky and T. Nagnibeda. Not every uniform tree covers ramanujan graphs. *Journal of Combinatorial Theory, Series B*, 74(2):202–212, 1998.

[LP10] N. Linial and D. Puder. Words maps and spectra of random graph lifts. *Random Structures and Algorithms*, 37(1):100–135, 2010.

[LSV11] E. Lubetzky, B. Sudakov, and V. Vu. Spectra of lifted ramanujan graphs. *Advances in Mathematics*, 227(4):1612–1645, 2011.

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

[Nil91] A. Nilli. On the second eigenvalue of a graph. *Discrete Mathematics*, 91(2):207–210, 1991.

[PP14] O. Parzanchevski and D. Puder. Stallings graphs, algebraic extensions and primitive elements in F_2. *Mathematical Proceedings of the Cambridge Philosophical Society*, 157(1):1–11, 2014.

[PP15] D. Puder and O. Parzanchevski. Measure preserving words are primitive. *Journal of the American Mathematical Society*, 28(1):63–97, 2015.

[Pud14] D. Puder. Primitive words, free factors and measure preservation. *Israel Journal of Mathematics*, 201(1):25–73, 2014.

[Pud15a] D. Puder. Expansion of random graphs: New proofs, new results. *Inventiones Mathematicae*, 2015. to appear. arXiv:1212.5216.

[PW14] D. Puder and C. Wu. Growth of the primitives elements in free groups. *Journal of London Mathematical Society*, 90(1):89–104, 2014.

[Whi36a] J.H.C. Whitehead. On certain sets of elements in a free group. *Proc. London Math. Soc.*, 41:48–56, 1936.

Chapter 1

Primitive Words, Free Factors and Measure Preservation

Doron Puder[†]
Einstein Institute of Mathematics
Hebrew University, Jerusalem
doronpuder@gmail.com

Published: Israel Journal of Mathematics, 201 (1), 2014, pp 25-73. DOI: 10.1007/s11856-013-0055-2

Abstract

Let $\mathbf{F}_k$ be the free group on k generators. A word $w \in \mathbf{F}_k$ is called primitive if it belongs to some basis of $\mathbf{F}_k$. We investigate two criteria for primitivity, and consider more generally, subgroups of $\mathbf{F}_k$ which are free factors.

The first criterion is graph-theoretic and uses Stallings core graphs: given subgroups of finite rank $H \leq J \leq \mathbf{F}_k$ we present a simple procedure to determine whether H is a free factor of J. This yields, in particular, a procedure to determine whether a given element in $\mathbf{F}_k$ is primitive.

Again let $w \in \mathbf{F}_k$ and consider the word map $w : G \times \ldots \times G \to G$ (from the direct product of k copies of G to G), where G is an arbitrary finite group. We call w *measure preserving* if given uniform measure on $G \times \ldots \times G$, w induces uniform measure on G (for every finite G). This is the second criterion we investigate: it is not hard to see that primitivity implies measure preservation and it was conjectured that the two properties are equivalent. Our combinatorial approach to primitivity allows us to make progress on this problem and in particular prove the conjecture for $k = 2$.

It was asked whether the primitive elements of $\mathbf{F}_k$ form a closed set in the profinite topology of free groups. Our results provide a positive answer for $\mathbf{F}_2$.
Keywords: word maps, primitive elements of free groups, primitivity rank

Contents

[†]Supported by Advanced ERC Grant 247034 of Aner Shalev, and by Adams Fellowship Program of the Israel Academy of Sciences and Humanities.

1.1 Introduction

An element w of a free group J is called *primitive* if it belongs to some basis (free generating set) of J. When J is given with a basis X, this is equivalent to the existence of an automorphism of J which sends w to a given element of X.

The notion of primitivity has a natural extension to subgroups in the form of free factors. Let H be a subgroup of the free group J (in particular, H is free as well). We say that H is a *free factor* of J and denote $H \overset{*}{\leq} J$, if there is another subgroup $H' \leq J$ such that $H * H' = J$. Equivalently, $H \overset{*}{\leq} J$ if every basis of H can be extended to a basis of J. (This in turn is easily seen to be equivalent to the condition that *some* basis of H extends to a basis of J).

Let $\mathbf{F}_k$ be the free group on k generators with a fixed basis $X = \{x_1, \ldots, x_k\}$. We study finitely generated subgroups of $\mathbf{F}_k$ (denoted $H \leq_{fg} \mathbf{F}_k$) and relations among them using core graphs, also known as Stallings' graphs (See [Sta83]. Actually our definition is a bit different than Stalling's, see below). Associated with every subgroup $H \leq \mathbf{F}_k$ is a directed, pointed, edge-labeled graph denoted $\Gamma_X(H)$. Edges are labeled by the elements of the given basis $X = \{x_1, \ldots, x_k\}$ of $\mathbf{F}_k$. A full definition appears in Section 1.2, but we illustrate the concept in Figure 1.1.1. It shows the core-graph of the subgroup of $\mathbf{F}_2$ generated by $x_1 x_2 {x_1}^{-1} {x_2}^{-1}$ and $x_2 {x_1}^2$.

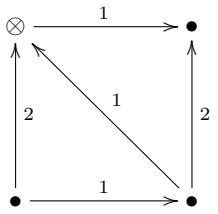


Figure 1.1.1: The core graph $\Gamma_X(H)$ where $H = \langle x_1 x_2^{-1} x_1, x_1^{-2} x_2 \rangle \leq \mathbf{F}_2$.

Core graphs are a key tool in the research of free groups, and are both used for proving new results and for introducing simple proofs to known results (see, for instance, [KM02, MVW07], for a survey of many such results and for further references).

A central new ingredient of our work is a new perspective on core graphs. There is a naturally defined notion of quotient on such graphs (see Section 1.3). In particular, we introduce in Section 1.3 the notion of *immediate quotients*. This in turn yields a directed graph whose vertices are all core graphs of finitely generated subgroups of $\mathbf{F}_k$ (w.r.t. the fixed basis X). A directed edge in this graph stands for the relation of an immediate quotient. This is a directed acyclic graph (DAG) i.e., it contains no directed cycles. As always, reachability in a DAG induces a distance function between vertices. Namely $\rho_X(x, y)$ is the shortest length of a directed path from x to y. We mention that the transitive closure of the immediate quotient relation is the relation "being a quotient of" which is a partial order (a lattice, in fact) on all core graphs of f.g. subgroups of $\mathbf{F}_k$. The following theorem gives a simple criterion for free factorness in terms of this distance:

Theorem 1.1.1. *Let $H, J \leq_{fg} \mathbf{F}_k$, and assume $\Gamma_X(J)$ is a quotient of $\Gamma_X(H)$. Then $H \overset{*}{\leq} J$ if and only if*

$$\rho_X(H, J) = rk(J) - rk(H)$$

We note that $\rho_X(\cdot, \cdot)$ can be explicitly computed, and this theorem thus yields automatically an algorithm to determine, for two given $H, J \leq_{fg} \mathbf{F}_k$ whether H is a free factor of J. In particular, it can serve to detect primitive words (see Appendix 1.A). More generally, for any f.g. free groups $H \leq J$, this theorem can serve to detect the minimal number of complementary generators needed to obtain J from H (Corollary 1.3.6).

In fact, the core graph of every $H \leq_{fg} \mathbf{F}_k$ has finitely many quotients (or reachable vertices). This set is also known in the literature as the *fringe* of H (see, e.g. [MVW07]). For example, Figure 1.3.1 shows the fringe of the subgroup $H = \langle [x_1, x_2] \rangle$. The difference in ranks between H and $\mathbf{F}_2$ is 1. However, the distance between the corresponding core graphs in the fringe is 2. This proves that H is not a free factor of $\mathbf{F}_2$, or equivalently that $[x_1, x_2]$ is not primitive. We elaborate more in Appendix 1.A.1.

Remark 1.1.2. We stress that there are other graph-theoretic algorithms to detect free factors and primitive words, including simplifications of the seminal Whitehead algorithm (the algorithm first appeared in [Whi36a, Whi36b], for its graph-theoretic simplifications see [Ger84, Sta99]). Our approach, however, is very different and does not rely on Whitehead automorphisms. We elaborate more on this in Appendix 1.A.

Theorem 1.1.1 is also used for the other concept we study here, that of measure preservation of word maps. Associated with every $w \in \mathbf{F}_k$ is a *word map*. We view w as a word in the letters of the basis X. For every group G, this mapping which we also denote by w maps $\underbrace{G \times G \times \cdots \times G}_{k} \longrightarrow G$ as follows: It maps the k-tuple $(g_1, \ldots, g_k)$ to the element $w(g_1, \ldots, g_k) \in G$, where $w(g_1, \ldots, g_k)$ is the element obtained by replacing $x_1, \ldots, x_k$ with $g_1, \ldots, g_k$ (respectively) in the expression for w, and then evaluating this expression as a group element in G.

During the last years there has been a great interest in word maps in groups, and extensive research was conducted (see, for instance, [Sha09], [LS09]; for a recent book on the topic see [Seg09]). Our focus here is on the property of measure preservation: We say that the word w preserves measure with respect to a finite group G if when k-tuples of elements from G are sampled uniformly, the image of the word map w induces the uniform distribution on G. (In other words, all fibers of the word map have the same size). We say that w is *measure preserving* if it preserves measure with respect to *every* finite group G.

This concept was investigated in several recent works. See for example [LS08] and [GS09], where certain word maps are shown to be almost measure preserving, in the sense that the distribution induced by w on finite simple groups G tends to uniform, say, in L_1 distance, when $|G| \to \infty$.

Measure preservation can be equivalently defined as follows: fix some finite group G, and select a homomorphism $\alpha_G \in Hom(\mathbf{F}_k, G)$ uniformly at random. A homomorphism from a free group is uniquely determined by choosing the images of the elements of a basis, so that every homomorphism is chosen with probability $1/|G|^k$. We then say that $w \in \mathbf{F}_k$ is measure preserving if for every finite group G and a random homomorphism α_G as above, $\alpha_G(w)$ is uniformly distributed over G.

We note that there is a stronger condition of measure preservation on a word w that is discussed in the literature. In this stronger condition we consider the image of w over the broader class of compact groups G w.r.t. their Haar measure. Our results make use only of the weaker condition that involves only finite groups.

Measure preservation can also be defined for f.g. subgroups.

Definition 1.1.3. For $H \leq_{fg} \mathbf{F}_k$ we say that H is *measure preserving* if and only if for any finite group G and $\alpha_G \in Hom(\mathbf{F}_k, G)$ a randomly chosen homomorphism as before, $\alpha_G|_H$ is uniformly distributed in $Hom(H, G)$.

In particular, $1 \neq w \in \mathbf{F}_k$ is measure preserving if and only if $\langle w \rangle$ is measure preserving.

It is easily seen that primitivity or free factorness yield measure preservation. The reason is that as mentioned, a homomorphism in $Hom(\mathbf{F}_k, G)$ is completely determined by the images of the elements of a basis of $\mathbf{F}_k$, which can be chosen completely arbitrarily and independently.

Several authors have conjectured that the converse is also true:

Conjecture 1.1.4. *For every $w \in \mathbf{F}_k$,*

$$w \text{ is primitive} \iff w \text{ is measure preserving}$$

More generally, for $H \leq_{fg} \mathbf{F}_k$,

$$H \overset{*}{\leq} \mathbf{F}_k \iff H \text{ is measure preserving}$$

From private conversations we know that this has occurred to the following mathematicians and discussed among themselves: T. Gelander, A. Shalev, M. Larsen and A. Lubotzky. The question was mentioned several times in the Einstein Institute Algebra Seminar. This conjecture was independently raised in [LP10][†].

Here we prove a partial result:

Theorem 1.1.5. *Let $H \leq_{fg} \mathbf{F}_k$ have rank $\geq k-1$. Then,*

$$H \overset{*}{\leq} \mathbf{F}_k \iff H \text{ is measure preserving}$$

In particular, for every $w \in \mathbf{F}_2$:

$$w \text{ is primitive} \iff w \text{ is measure preserving}$$

The proof of this result relies, inter alia, on Theorem 1.1.1. Note that a set of $k-1$ elements $w_1, \ldots, w_{k-1} \in \mathbf{F}_k$ can be extended to a basis if and only if it is a free set that generates a free factor. Thus, the result for subgroups can also be stated for finite subsets as follows: Let $r \geq k-1$. A set $\{w_1, \ldots, w_r\} \subset \mathbf{F}_k$ can be extended to a basis if and only if for every finite group G and random homomorphism α_G as above, the r-tuple $(\alpha_G(w_1), \ldots, \alpha_G(w_r))$ is uniformly distributed in G^r, the direct product of r copies of G.

There is an interesting connection between this circle of ideas and the study of profinite groups. For example, an immediate corollary of Theorem 1.1.5 is that

Corollary 1.1.6. *The set of primitive elements in $\mathbf{F}_2$ is closed in the profinite topology.*

We discuss this corollary and other related results in Section 1.7.

In order to prove Conjecture 1.1.4, one needs to find for every non-primitive word $w \in \mathbf{F}_k$, some witness finite group G with respect to which w is not measure preserving. Our witnesses are always the symmetric groups S_n.

It is conceivable that our method of proof for Theorem 1.1.5 is powerful enough to establish Conjecture 1.1.4. We define two categorizations of elements (and of f.g. subgroups) of free groups $\pi(\cdot)$ and $\phi(\cdot)$. They map every free word and free subgroup into $\{0, 1, 2, 3, \ldots\} \cup \{\infty\}$. We believe these two maps are in fact identical. This, if true, yields the general conjecture. Presently we can show that they are equivalent under certain conditions, and this yields our partial result.

[†]It is interesting to note that there is an easy abelian parallel to Conjecture 1.1.4: A word $w \in \mathbf{F}_k$ is primitive, i.e. belongs to a basis, in $\mathbb{Z}^k \cong \mathbf{F}_k/\mathbf{F}_k'$ if and only if for any group G the associated word map is surjective. See [Seg09], Lemma 3.1.1.

The first categorization is called *the primitivity rank*. It is a simple fact that if $w \in \mathbf{F}_k$ is primitive, then it is also primitive in every subgroup of $\mathbf{F}_k$ containing it (see Claim 1.2.5). However, if w is not primitive in $\mathbf{F}_k$, it may be either primitive or non-primitive in subgroups containing it. But what is the smallest rank of a subgroup in which we can realize w is not primitive? Informally, how far does one have to search in order to establish that w is *not* primitive in $\mathbf{F}_k$? Concretely:

Definition 1.1.7. The **primitivity rank** of $w \in \mathbf{F}_k$, denoted $\pi(w)$, is

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

If no such J exists, $\pi(w) = \infty$. A subgroup J for which the minimum is obtained is called w**-critical**.

This extends naturally to subgroups. Namely,

Definition 1.1.8. For $H \leq_{fg} \mathbf{F}_k$, the primitivity rank of H is

$$\pi(H) = min\left\{ rk(J) \;\middle|\; \begin{array}{c} H \leq J \leq \mathbf{F}_k \ s.t. \\ H \text{ is } \mathbf{not} \text{ a free factor of } J. \end{array} \right\}$$

Again, if no such J exists, $\pi(H) = \infty$. A subgroup J for which the minimum is obtained is called H**-critical**.

For instance, $\pi(w) = 1$ if and only if w is a proper power of another word (i.e. $w = v^d$ for some $v \in \mathbf{F}_k$ and $d \geq 2$). In Section 1.4 we show (Corollary 1.4.2) that in $\mathbf{F}_k$ the primitivity rank takes values only in $\{0, 1, 2, \ldots, k\} \cup \{\infty\}$ (the only word w with $\pi(w) = 0$ is $w = 1$). Lemma 1.4.1 shows, moreover, that $\pi(w) = \infty$ ($\pi(H) = \infty$, resp.) if and only if w is primitive ($H \overset{*}{\leq} \mathbf{F}_k$). Finally Lemma 1.6.8 yields that π can take on every value in $\{0, \ldots, k\}$. For example, if $\mathbf{F}_k$ is given with some basis $X = \{x_1, \ldots, x_k\}$ then for every $1 \leq d \leq k$, $\pi(x_1^2 \ldots x_d^2) = d$. It is interesting to mention that $\pi(H)$ also generalizes the notion of *compressed* subgroups, as appears, e.g., in [MVW07]: a subgroup $H \leq_{fg} \mathbf{F}_k$ is compressed if and only if $\pi(H) \geq rk(H)$.

The second categorization of sets of formal words has its roots in [Nic94] and more explicitly in [LP10]. It concerns homomorphisms from $\mathbf{F}_k$ to the symmetric groups S_n, and more concretely the probability that 1 is a fixed point of the permutation $w(\sigma_1, \ldots, \sigma_k)$ for some $w \in \mathbf{F}_k$ when $\sigma_1, \ldots, \sigma_k \in S_n$ are chosen randomly with uniform distribution. More generally, for a subgroup $H \leq_{fg} \mathbf{F}_k$ we study the probability that 1 is a common fixed point of (the permutations corresponding to) all elements in H. We ask how much this probability deviates from the corresponding probability in the case of measure preserving subgroups, i.e. from $\frac{1}{n^{rk(H)}}$. (We continue the presentation for subgroups only. This clearly generalizes the case of a word: for every word $w \neq 1$ consider the subgroup $\langle w \rangle$.)

Formally, for $H \leq_{fg} \mathbf{F}_k$ we define the following function whose domain is all integers $n \geq 1$ where $\alpha_n \in Hom(\mathbf{F}_k, S_n)$ is a random homomorphism with uniform distribution:

$$\Phi_H(n) = Prob\big[\forall w \in H \quad \alpha_n(w)(1) = 1\big] - \frac{1}{n^{rk(H)}} \tag{1.1.1}$$

Clearly, if H is measure preserving, then Φ_H vanishes for every $n \geq 1$.

Nica [Nic94] showed that for a fixed word $w \neq 1$ and large enough n, it is possible to express $\Phi_w(n)$ ($=\Phi_{\langle w \rangle}(n)$) as a rational function in n. We show below that this is easily extended to apply to $\Phi_H(n)$ for arbitrary $H \leq_{fg} \mathbf{F}_k$. Nica's clever observation was used in [LP10] to introduce a new categorization of free words, denoted $\phi(\cdot)$, which, like $\pi(\cdot)$, associates a non-negative integer or ∞ to every formal word (note that in [LP10] the notion of primitive words has a different meaning

than in the current paper). This categorization can also be extended to arbitrary finitely generated subgroups of $\mathbf{F}_k$. More specifically, it is shown in Section 1.5 that for every $H \leq_{fg} \mathbf{F}_k$ and n large enough (say, at least the number of vertices in the core graph of H), we have

$$\Phi_H(n) = \sum_{i=0}^{\infty} a_i(H) \frac{1}{n^i} \tag{1.1.2}$$

where the coefficients $a_i(H)$ are integers depending only on H. We define $\phi(H)$ as follows:

$$\phi(H) := \begin{cases} \text{the smallest integer } i \text{with } a_i(H) \neq 0 & \text{if } \Phi_H(n) \not\equiv 0 \\ \infty & \text{if } \Phi_H(n) \equiv 0 \end{cases} \tag{1.1.3}$$

Thus, $\phi(H)$ measures to what extent the probability that 1 is a common fixed point of H differs from $\frac{1}{n^{rk(H)}}$, the corresponding probability if H were measure preserving. The higher $\phi(H)$ is, the closer the probability is asymptotically to $\frac{1}{n^{rk(H)}}$. If H is a measure preserving subgroup, then $\phi(H) = \infty$.

As it turns out there is a strong connection between $\pi(H)$ and $\phi(H)$. Already Nica's result can be interpreted in the language of $\phi(\cdot)$ to say that $\phi(w) = 1$ if and only if w is a power, that is if and only if $\pi(w) = 1$. But the connection goes deeper. In proving this, we calculate these functions using the core graph of H and its quotients. It turns out that both $\pi(H)$ and $\phi(H)$ can be computed explicitly via the subgraph of the DAG induced by all descendants of $\Gamma_X(H)$.

In the calculation of $\phi(H)$ we use the core graph $\Gamma_X(H)$ and its quotients to partition the event that 1 is a common fixed point of $\alpha_n(w)$ of each $w \in H$ (see Section 1.5).

Fortunately, the same core graph and quotients can also be used to find the primitivity rank $\pi(H)$, as shown in Section 1.4. Lemma 1.4.3 shows that all H-critical subgroups (see Definition 1.1.8) are always represented in the fringe (set of quotients) of H. Theorem 1.1.1 then shows directly how to calculate $\pi(H)$ using the fringe.

We show that under certain conditions, the two categorizations $\pi(\cdot)$ and $\phi(\cdot)$ indeed coincide.

Proposition 1.1.9. *Let $H \leq_{fg} \mathbf{F}_k$. Then for every $i \leq rk(H) + 1$,*

(1) $\pi(H) = i \iff \phi(H) = i$

(2) Moreover, if $\pi(H) = \phi(H) = i$ then $a_i(H)$ equals the number of H-critical subgroups of $\mathbf{F}_k$.

The second part of this proposition is in fact a generalization of a result of Nica. For a single element $w \in \mathbf{F}_k$ which is a proper power, namely $\pi(w) = \phi(w) = 1$, let $w = u^d$ with d maximal (so u is not a proper power). Let M denote the number of divisors of d. It is not hard to see that the number of w-critical subgroups of $\mathbf{F}_k$ equals $M - 1$: these subgroups are exactly $\langle u^m \rangle$ for every $1 \leq m < d$ such that $m|d$. This shows that the average number of fixed points in the permutation $\alpha_n(w)$ goes to M as $n \to \infty$. This corresponds to Corollary 1.3 in [Nic94] (for the case $L = 1$)[†].

The connection between $\pi(\cdot)$ and $\phi(\cdot)$ goes beyond the cases stated in Proposition 1.1.9. To start off, if $\pi(H) = \infty$, then $H \overset{*}{\leq} \mathbf{F}_k$ and therefore H is measure preserving, and thus $\phi(H) = \infty$. In addition, Lemma 1.6.8 states that both $\pi(\cdot)$ and $\phi(\cdot)$ are additive with respect to concatenation of words on disjoint letter sets. Namely, if the words $w_1, w_2 \in \mathbf{F}_k$ have no letters in common then $\pi(w_1 w_2) = \pi(w_1) + \pi(w_2)$ and $\phi(w_1 w_2) = \phi(w_1) + \phi(w_2)$. Moreover, if the disjoint w_1 and w_2 satisfy both parts of Proposition 1.1.9 then so does their concatenation $w_1 w_2$.

In view of this discussion, the following conjecture suggests itself quite naturally:

[†]Nica's result was more general in a different manner: it involved the distribution of the number of L-cycles in the random permutation $\alpha_n(w)$, for any fixed L. He showed that as $n \to \infty$, the limit distribution depends only on d, where $w = u^d$ as above.

Conjecture 1.1.10.

(1) For every $H \leq_{fg} \mathbf{F}_k$
$$\pi(H) \quad = \quad \phi(H)$$

(2) Moreover, $a_{\phi(H)}(H)$ equals the number of H-critical subgroups of $\mathbf{F}_k$.

Specifically, for a single word w, Proposition 1.1.9 states that for $i = 0, 1, 2$, $\pi(w) = i \Leftrightarrow \phi(w) = i$. As mentioned, the possible values of $\pi(H)$ are $\{0, 1, 2, \ldots, k\} \cup \{\infty\}$, and $\pi(H) = \infty$ if and only if $H \stackrel{*}{\leq} \mathbf{F}_k$. We also have $\pi(H) = \infty \Rightarrow \phi(H) = \infty$ (a free factor subgroup is measure preserving). Thus, when $rk(H) \geq k-1$, the value of $\pi(H)$ uniquely determines $\phi(H)$ and the two values coincide. In other words, when $rk(H) \geq k - 1$
$$\pi(H) = \phi(H).$$

This shows, in turn, that when H is measure preserving, we have $\pi(H) = \phi(H) = \infty$, and so H is a free factor. This yields Theorem 1.1.5. The same argument shows that Conjecture 1.1.4 follows from part (1) of Conjecture 1.1.10 and suggests, in particular, a general strategy towards proving Conjecture 1.1.4.

As an aside, the second parts of Proposition 1.1.9 and Conjecture 1.1.10 say something interesting on the average number of fixed points in the random permutation $\alpha_n(w)$. We conjecture that for every w and for large enough n, this average is at least 1. In other words, among the family of distributions of S_n induced by free words, a random uniformly chosen permutation has the least average number of fixed points. This point is further elaborated in Section 1.8.

At this point we should clarify the relation of these results and some of what we did in [LP10]. There we introduced $\beta(\cdot)$ - yet another categorization of formal words. Just like $\phi(\cdot)$ and $\pi(\cdot)$ it maps every formal word to a non-negative integer or ∞. As it turns out, $\pi(\cdot)$ and $\beta(\cdot)$ coincide. This follows from Theorem 1.1.1 and from Section 1.4. The definition of $\pi(\cdot)$ is simpler and more elegant than the original definition of $\beta(\cdot)$. As shown in [LP10] for $i = 0, 1$, $\phi(w) = i \iff \beta(w) = i$. A partial proof was given there as well for the case $i = 2$. In Section 1.6 we complete the argument for $i = 2$ and generalize it to prove Proposition 1.1.9.

The paper is arranged as follows. In section 1.2 we introduce the notions of core graphs, their morphisms and their quotients. In Section 1.3 we introduce our new perspective on core graphs, including the notion of immediate quotients and the mentioned DAG, and then prove Theorem 1.1.1. In Section 1.4 we analyze the primitivity rank of any $H \leq_{fg} \mathbf{F}_k$ and show how it can be computed from the quotients of $\Gamma_X(H)$ in the DAG of finite rank subgroups of $\mathbf{F}_k$. Section 1.5 is devoted to proving that $\phi(H)$ is well defined and can be indeed computed from the same descendants of $\Gamma_X(H)$. In Section 1.6 we establish the results connecting $\phi(\cdot)$ and $\pi(\cdot)$, culminating in the proof of Theorem 1.1.5. The concluding sections are devoted to two different consequences of the main results: the characterization of elements of $\mathbf{F}_k$ which are primitive in its profinite completion (Section 1.7) and the possible values of the average number of fixed points in the image of a word map on S_n (Section 1.8). The discussion in the three appendices is not necessary for the main results of this paper, but it does, in our view, complete the picture. In particular, we illustrate in Appendix 1.A the algorithm to detect free factor subgroups.

1.2 Core Graphs and their Quotients

All groups that appear here are subgroups of $\mathbf{F}_k$, the free group with a given basis $X = \{x_1, \ldots, x_k\}$. Some of the relations we consider depend on the choice of the basis. We first describe core-graphs, which play a crucial role in this paper.

1.2.1 Core Graphs

Associated with every subgroup $H \leq \mathbf{F}_k$ is a directed, pointed, edge-labeled graph. This graph is called *the core-graph associated with* H and is denoted by $\Gamma_X(H)$. We recall the notion of $\overline{\Gamma}_X(H)$ the Schreier (right) coset graph of H with respect to the basis X. 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 letter x_i there is a directed i-edge (short for x_i-edge) going from the vertex Hw to the vertex Hwx_i.

The core graph $\Gamma_X(H)$ is obtained from $\overline{\Gamma}_X(H)$ by omitting all the vertices and edges of $\overline{\Gamma}_X(H)$ which are never traced by a reduced (i.e., non-backtracking) path that starts and ends at the basepoint. Stated informally, we omit all (infinite) "hanging trees" from $\overline{\Gamma}_X(H)$. To illustrate, Figure 1.2.1 shows the graphs $\overline{\Gamma}_X(H)$ and $\Gamma_X(H)$ for $H = \langle x_1 x_2 x_1^{-3}, x_1^2 x_2 x_1^{-2} \rangle \leq \mathbf{F}_2$.

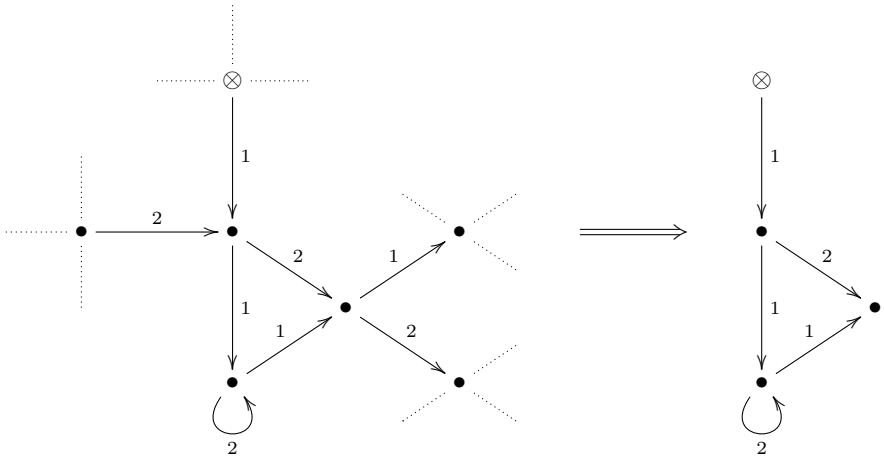


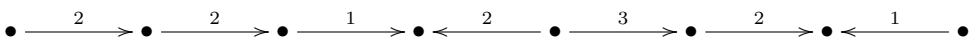
Figure 1.2.1: $\overline{\Gamma}_X(H)$ and $\Gamma_X(H)$ for $H = \langle x_1 x_2 x_1^{-3}, x_1^2 x_2 x_1^{-2} \rangle \leq \mathbf{F}_2$. The Schreier coset graph $\overline{\Gamma}_X(H)$ is the infinite graph on the left (the dotted lines represent infinite 4-regular trees). The basepoint "$\otimes$" corresponds to the trivial coset H, the vertex below it corresponds to the coset Hx_1, the one further down corresponds to $Hx_1^2 = Hx_1x_2x_1^{-1}$, etc. The core graph $\Gamma_X(H)$ is the finite graph on the right, which is obtained from $\overline{\Gamma}_X(H)$ by omitting all vertices and edges that are not traced by reduced closed paths around the basepoint.

Note that the graph $\overline{\Gamma}_X(H)$ is $2k$-regular: Every vertex has exactly one outgoing j-edge and one incoming j-edge for every $1 \leq j \leq k$. Every vertex of $\Gamma_X(H)$ has *at most* one outgoing j-edge, and *at most* one incoming j-edge for every $1 \leq j \leq k$.

It is an easy observation that

$$\pi_1(\overline{\Gamma}_X(H)) = \pi_1(\Gamma_X(H)) \overset{canonically}{\cong} H$$

where the canonical isomorphism is given by associating words in $\mathbf{F}_k$ to paths in the coset graph and in the core graph: We traverse the path by following the labels of outgoing edges. For instance, the path (from left to right)



corresponds to the word $x_2^2 x_1 x_2^{-1} x_3 x_2 x_1^{-1}$. (See also [MVW07], where this fact appears in a slightly different language).

Core graphs were introduced by Stallings [Sta83]. Our definition is slightly different, in that we allow the basepoint to have degree one.

In fact, a "tail" in $\Gamma_X(H)$, i.e., a path to the basepoint can be eliminated by replacing H by an appropriate conjugate. However, we find it unnecessary and less elegant for our needs.

We now list some properties of core graph, most of which are proved in at least one of [Sta83, KM02, MVW07]. The remaining ones are easy observations.

Claim 1.2.1. *Let H be a subgroup of $\mathbf{F}_k$ with an associated core graph $\Gamma = \Gamma_X(H)$. The Euler Characteristic of a graph, denoted $\chi(\cdot)$ is the number of vertices minus the number of edges. Finally, $rk(H)$ denotes the rank of the group H.*

(1) $rk(H) < \infty \Leftrightarrow \Gamma$ is finite

(2) $rk(H) = 1 - \chi(\Gamma)$

(3) Let Λ be a finite, pointed, directed graph with edges labeled by $\{x_1, \ldots, x_k\}$. Then Λ is a core graph (corresponding to some $J \leq \mathbf{F}_k$) if and only if Λ satisfies the following three properties:

   (a) Λ is connected

   (b) With the possible exception of the basepoint, every vertex has degree at least 2.

   (c) For every $1 \leq j \leq k$, no two j-edges share the same origin nor the same terminus.

(4) There is a one-to-one correspondence between subgroups of $\mathbf{F}_k$ and core graphs.

(5) There is a one-to-one correspondence between subgroups of $\mathbf{F}_k$ of finite rank and finite core graphs.

In Appendix 1.C we present a well known algorithm, based on Stallings foldings, to obtain the core graph of every $H \leq_{fg} \mathbf{F}_K$ given some finite generating set for H.

1.2.2 Morphisms of Core Graphs

In our framework, a morphism between two core-graphs Γ_1 and Γ_2 is a map that sends vertices to vertices and edges to edges, and preserves the structure of the 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 1.2.1, the proofs of the following properties are either easy variations on proofs in [Sta83, KM02, MVW07] or just easy observations:

Claim 1.2.2. *Let $H_1, H_2 \leq \mathbf{F}_k$ be subgroups, and Γ_1, Γ_2 be the corresponding core graphs. Then*

(1) A morphism $\eta : \Gamma_1 \to \Gamma_2$ exists $\Leftrightarrow H_1 \leq H_2$,
and in this case, $\eta_* : \pi_1(\Gamma_1) \to \pi_1(\Gamma_2)$ is injective.

(2) If a morphism exists, it is unique.

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

1.2.3 Quotients of Core Graphs

With core-graph morphisms at hand, we can define the following rather natural relation between core-graphs.

Definition 1.2.3. Let Γ_1, Γ_2 be core graphs and $H_1, H_2 \leq \mathbf{F}_k$ the corresponding subgroups. We say that Γ_1 **covers** Γ_2 or that Γ_2 is a **quotient** of Γ_1 if there is a surjective morphism $\eta : \Gamma_1 \twoheadrightarrow \Gamma_2$. We also say in this case that H_1 **covers** H_2, and denote $\Gamma_1 \twoheadrightarrow \Gamma_2$ or $H_1 \leq_{\vec{x}} H_2$.

By "surjective" we mean surjective on both the vertices and the edges. Note that we use the term "covers" even though this is *not* a covering map in general (the morphism from Γ_1 to Γ_2 is always locally injective at the vertices, but not necessarily locally bijective).

For instance, $H = \langle x_1 x_2 x_1^{-3}, x_1^2 x_2 x_1^{-2} \rangle \leq \mathbf{F}_k$ covers the group
$J = \langle x_2, x_1^2, x_1 x_2 x_1 \rangle$, the corresponding core graphs of which are the leftmost and rightmost graphs in Figure 1.2.2. As another example, every core graph Γ that contains edges of all labels covers the wedge graph Δ_k.

We already know (Claim 1.2.2) that if $H_1 \leq_{\vec{x}} H_2$ then, in particular, $H_1 \leq H_2$. However, the converse is incorrect. For example, the group
$K = \langle x_1 x_2 x_1^{-3}, x_1^2 x_2 x_1^{-2}, x_2 \rangle$ contains $H = \langle x_1 x_2 x_1^{-3}, x_1^2 x_2 x_1^{-2} \rangle$ (we simply added x_2 as a third generator), yet K is not a quotient of H: the morphism $\eta : \Gamma_X(H) \to \Gamma_X(K)$ does not contain the 2-loop at the basepoint of $\Gamma_X(K)$ in its image.

Note also that the relation $H_1 \leq_{\vec{x}} H_2$ *depends on the given generating set* X *of* $\mathbf{F}_k$. For example, if $H = \langle x_1 x_2 \rangle$ then $H \leq_{\vec{x}} \langle x_1, x_2 \rangle = F_2$. However, $x_1 x_2$ is primitive and could be taken as part of the original basis of $\mathbf{F}_2$. In that case, the core graph of H would consist of a single vertex and single loop and would have no quotients except for itself.

It is also interesting to note that every quotient of the core-graph Γ corresponds to some partition of $V(\Gamma)$ (the partition determined by the fibers of the morphism). We can simply draw a new graph with a vertex for each block in the partition, and a j-edge from block b_1 to block b_2 whenever there is some j-edge (v_1, v_2) in Γ_1 with $v_1 \in b_1, v_2 \in b_2$. However, not every partition of $V(\Gamma)$ corresponds to a quotient core-graph: In the resulting graph two distinct j-edges may have the same origin or the same terminus. Note that even if a partition P of $V(\Gamma)$ yields a quotient which is not a core-graph, this can be remedied. We can activate the folding process exemplified in Appendix 1.C and obtain a core graph. The resulting partition P' of $V(\Gamma)$ is the finest partition which yields a quotient core-graph and which is still coarser than P. We illustrate this in Figure 1.2.2.

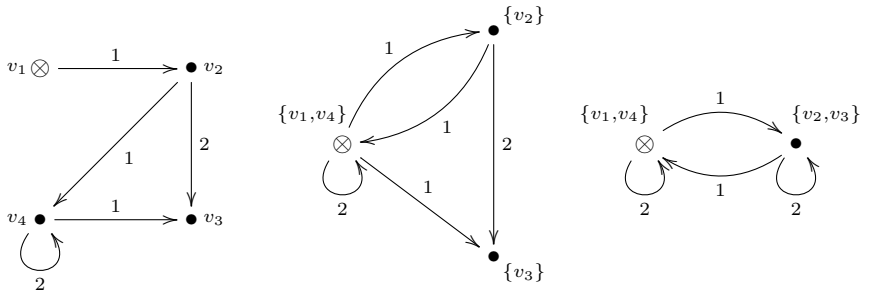


Figure 1.2.2: The left graph is the core graph $\Gamma_X(H)$ of $H = \left\langle x_1 x_2 x_1^{-3}, x_1^2 x_2 x_1^{-2} \right\rangle \leq \mathbf{F}_2$. Its vertices are denoted $v_1, \ldots, v_4$. The graph in the middle is the quotient corresponding to the partition $P = \{\{v_1, v_4\}, \{v_2\}, \{v_3\}\}$. This is not a core graph as there are two different 1-edges originating at $\{v_1, v_4\}$. In order to obtain a core quotient-graph, we use the folding process illustrated in Appendix 1.C. The resulting core graph is on the right, corresponding to the partition $P' = \{\{v_1, v_4\}, \{v_2, v_3\}\}$.

Lemma 1.2.4. *Every finite core-graph has a finite number of quotients.*
Equivalently, every $H \leq_{fg} \mathbf{F}_k$ *covers a finite number of other subgroups.*

Proof. The number of quotients of Γ is bounded from above by the number of partitions of $V(\Gamma)$. □

Following the notations in [MVW07], we call the set of X-quotients of H the X-*fringe* of H and denote ${}_X(H)$. Namely,

$$ {}_X(H) := \{\Gamma_X(J) \mid H \leq_{\vec{x}} J\} \tag{1.2.1}$$

Lemma 1.2.4 states in this terminology that for every $H \leq_{fg} \mathbf{F}_k$ (and every basis X), $|_X(H)| < \infty$.

Before describing our new perspective on core graphs, we remind some useful facts about free factors in free groups:

Claim 1.2.5. *Let $H, J, K \leq \mathbf{F}_k$. Then,*

(1) Free factorness is transitive: If $H \overset{*}{\leq} J \overset{*}{\leq} K$ then $H \overset{*}{\leq} K$.

(2) If $\eta : \Gamma_X(H) \hookrightarrow \Gamma_X(J)$ is an embedding then $H \overset{*}{\leq} J$.

(3) If $H \overset{*}{\leq} J$ then H is a free factor in any subgroup $H \leq M \leq J$ in between.

Proof. The first and second claims are immediate. We give a "graph-theoretic" proof for the third one. Assume that $H \overset{*}{\leq} J$, and let Y be a basis of J extending some basis of H. In particular, $\Gamma_Y(H)$ and $\Gamma_Y(J)$ are both bouquets, consisting of a single vertex and $rk(H)$ (resp. $rk(J)$) loops. Now, for every $H \leq M \leq J$, consider the morphism $\eta : \Gamma_Y(H) \to \Gamma_Y(M)$. It is easy to see that a core-graph-morphism of a bouquet must be an embedding. Thus, by the second claim, $H \overset{*}{\leq} M$. □

1.3 Immediate Quotients and the DAG of Core Graphs

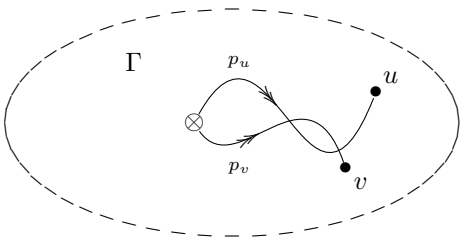
The quotient relation yields a partial order on the set of core graphs. But we are interested in a special case which we call *immediate quotients*. This relation allows us to build the aforementioned DAG (directed acyclic graph) of all (core graphs corresponding to) finite rank subgroups of $\mathbf{F}_k$.

Let Γ be a core graph, and let P be a partition of $V(\Gamma)$. Let Δ be the quotient core graph we obtain from P by the folding process described in Figures 1.C.1 and 1.2.2. We say that Δ is *generated* from Γ by P. We are interested in the case where P identifies only a single pair of vertices:

Definition 1.3.1. Let Γ be a core graph and let P be a partition of $V(\Gamma)$ in which all parts consist of a single vertex with a single exceptional part that contains two vertices. Let Δ be the core graph generated by P. We then say that Δ is an **immediate quotient** of Γ.

Alternatively we say that Δ is *generated by merging a single pair* of vertices of Γ. For instance, the rightmost core graph in Figure 1.2.2 is an immediate quotient of the leftmost core graph.

The relation of immediate quotients has an interesting interpretation for the associated free groups. Let $H, J \leq \mathbf{F}_k$ be free groups and $\Gamma = \Gamma_X(H), \Delta = \Gamma_X(J)$ their core graph, and assume Δ is an immediate quotient of Γ obtained by identifying the vertices $u, v \in V(\Gamma)$. Now let $p_u, p_v \in \mathbf{F}_k$ be words corresponding to some paths in Γ from the basepoint to u and v respectively. It is not hard to see that identifying u and v has the same effect as adding the word $w = p_u p_v^{-1}$ to H and considering the generated group. Namely, $J = \langle H, w \rangle$.



Based on the relation of immediate quotients we consider the DAG $\mathcal{D}_k$. The set of vertices of this graph consists of all finite core graphs with edges labeled by $1, \ldots, k$, and its directed edges connect every core graph to its immediate quotients. Every fixed ordered basis of $\mathbf{F}_k$ $X =$

$\{x_1, \ldots, x_k\}$, determines a one-to-one correspondence between the vertices of this graph and all finite rank subgroups of $\mathbf{F}_k$.

As before, we fix an ordered basis X. For any $H \leq_{fg} \mathbf{F}_k$, the subgraph of $\mathcal{D}_k$ of the descendants of $\Gamma_X(H)$ consists of all quotients of $\Gamma_X(H)$, that is of all elements of the X-fringe ${}_X(H)$. By Lemma 1.2.4, this subgraph is finite. In Figure 1.3.1 we draw the subgraph of $\mathcal{D}_k$ consisting of all quotients of $\Gamma_X(H)$ when $H = \langle x_1 x_2 x_1^{-1} x_2^{-1} \rangle$. The edges of this subgraph (i.e. immediate quotients) are denoted by the broken arrows in the figure.

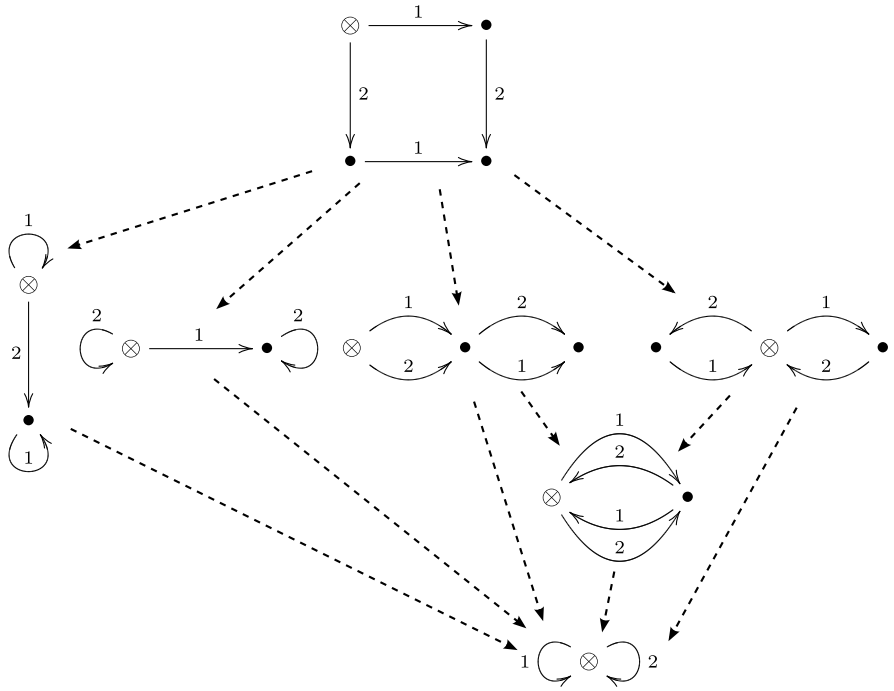


Figure 1.3.1: The subgraph of $\mathcal{D}_k$ induced by ${}_X(H)$, that is, all quotients of the core graph $\Gamma = \Gamma_X(H)$, for $H = \langle x_1 x_2 x_1^{-1} x_2^{-1} \rangle$. The dashed arrows denote immediate quotients, i.e. quotients generated by merging a single pair of vertices. Γ has exactly seven quotients: itself, four immediate quotients, and two quotients at distance 2.

This yields the aforementioned distance function between a finite core graph and a quotient of it:

Definition 1.3.2. Let $H_1, H_2 \leq_{fg} \mathbf{F}_k$ be finite rank subgroups such that $H_1 \leq_{\vec{x}} H_2$, and let $\Gamma_1 = \Gamma_X(H_1), \Gamma_2 = \Gamma_X(H_2)$ be the corresponding core graphs. We define the distance between H_1 and H_2, denoted $\rho_X(H_1, H_2)$ or $\rho(\Gamma_1, \Gamma_2)$, to be the shortest length of a directed path from Γ_1 to Γ_2 in $\mathcal{D}_k$.

In other words, $\rho_X(H_1, H_2)$ is the length of the shortest series of immediate quotients that yields Γ_2 from Γ_1. Equivalently, it is the minimal number of pairs of vertices that need to be identified in Γ_1 in order to obtain Γ_2 (via the folding process). For example, if Γ_2 is an immediate quotient of Γ_1 then $\rho_X(H_1, H_2) = \rho(\Gamma_1, \Gamma_2) = 1$. For $H = \langle x_1 x_2 x_1^{-1} x_2^{-1} \rangle$, $\Gamma_X(H)$ has four quotients at distance 1 and two at distance 2 (see Figure 1.3.1).

As aforementioned, by merging a single pair of vertices of $\Gamma_X(H)$ (and then folding) we obtain the core graph of a subgroup J obtained from H by adding some single generator. Thus, by taking an immediate quotient, the rank of the associated subgroup increases at most by 1 (in fact, it may also stay unchanged or even decrease). This implies that whenever $H \leq_{\vec{x}} J$:

$$rk(J) - rk(H) \quad \leq \quad \rho(H, J) \tag{1.3.1}$$

It is not hard to bound the distance from above as well:

Lemma 1.3.3. *Let $H, J \leq_{fg} \mathbf{F}_k$ such that $H \leq_{\vec{x}} J$. Then*

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

We postpone the proof of the upper bound to Appendix 1.B. (In fact, this upper bound in not needed for the main results of this paper. We give it anyway in order to have the full picture in mind.) Theorem 1.1.1 then states that in the same setting, the lower bound is attained iff H is a free factor of J. In fact one can visualize these results in the following way. Consider a two dimensional table which contains all the elements of the fringe $_X(H)$ (each quotient of $\Gamma_X(H)$ contained in some, not necessarily distinct, cell). The rows correspond to the rank and are indexed $0, 1, 2, 3, ...$ The columns correspond to the distance from H and are also indexed $0, 1, 2, 3, \ldots$. We then put every quotient of H in the suitable cell in the table. Let $r = rk(H)$ denote the rank of H. Lemma 1.3.3 then says that the (finitely many) elements of $_X(H)$ are spread across $r + 1$ diagonals in the table: the main one and the r diagonals below it. Theorem 1.1.1 implies that within $_X(H)$, H is a free factor of exactly those J-s found in the lowest of these $r + 1$ diagonals. (In fact, Lemma 1.4.3 shows that $\pi(H)$ can also be read from this table: it equals the rank of the upmost occupied cell in this table outside the free-factor-diagonal.)

1.3.1 Proof of Theorem 1.1.1

The main result of this section states that if $H \leq_{fg} J \leq_{fg} \mathbf{F}_k$ and $H \leq_{\vec{x}} J$, then

$$\rho_X(H, J) = rk(J) - rk(H) \quad \iff \quad H \overset{*}{\leq} J. \tag{1.3.2}$$

In fact, one of the implications is trivial. As mentioned above, merging two vertices in $\Gamma_X(H)$ is equivalent to adding some generator to H. If we manage to obtain $\Gamma_X(J)$ from $\Gamma_X(H)$ by $\mathrm{rk}(J) - \mathrm{rk}(H)$ merging steps, this means we can obtain J from H by adding $rk(J) - rk(H)$ extra generators to H, hence clearly $H \overset{*}{\leq} J$ (recall that by hopfianity of the free group, every generating set of size $\mathrm{rk}(J)$ is a basis of J, see e.g. [Bog08, Chapter 2.29]). Thus,

$$\rho_X(H, J) = rk(J) - rk(H) \quad \Longrightarrow \quad H \overset{*}{\leq} J \tag{1.3.3}$$

The other implication is not trivial. Assume that $H \overset{*}{\leq} J$. Our goal is to obtain $rk(J) - rk(H)$ complementary generators of J from H, so that each of them can be realized by merging a pair of vertices in $\Gamma_X(H)$.

To this goal we introduce the notion of a "*handle number*" associated with a subgroup M and a word $w \in \mathbf{F}_k$. (It also depends on the fixed basis X of $\mathbf{F}_k$). This number is defined as follows. Let $\Gamma = \Gamma_X(M)$. Denote by p_w the longest prefix of w that corresponds to some path from the basepoint of Γ (we trace the letters of w along Γ until we get stuck). Likewise, denote by s_w the longest suffix of w that ends at the basepoint (here we trace w^{-1} from the basepoint until we get stuck). If $|p_w| + |s_w| < |w|$, then $w = p_w m_w s_w$ as a reduced expression for some $1 \neq m_w \in \mathbf{F}_k$. The handle number of (M, w) is then

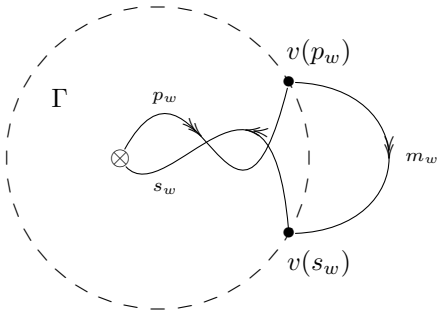
$$h_X(M, w) = h(\Gamma, w) = \begin{cases} |m_w| & |p_w| + |s_w| < |w| \\ 0 & \text{otherwise} \end{cases}.$$

Claim 1.3.4. *Assume that $w \notin M$ and let $N = \langle M, w \rangle$. Then*

(1) $h_X(M, w) > 0$ if and only if $\Gamma_X(M)$ is a (proper) subgraph of $\Gamma_X(N)$, and

(2) $h_X(M,w)=0$ if and only if $\Gamma_X(N)$ is an immediate quotient of $\Gamma_X(M)$.

Proof. Assume first that $h_X(M,w)>0$. In the notations of the previous paragraph, let $v(p_w), v(s_w)$ be the end point of the path corresponding to p_w and the starting point of the path corresponding to s_w. We can then add a "handle" to Γ in the form of a path corresponding to m_w which starts at $v(p_w)$ and ends at $v(s_w)$. (If $v(p_w)=v(s_w)$ this handle looks like a balloon, possibly with a string.)



The resulting graph is a core-graph (the edge conditions at $v(p_w)$ and $v(s_w)$ hold, by the maximality of p_w and s_w), and it corresponds to N. So we actually obtained $\Gamma_X(N)$. It follows that $\Gamma_X(M)$ is a proper subgraph of $\Gamma_X(N)$. On the other hand, if $h_X(M,w)=0$, i.e. if $|p_w|+|s_w|\geq|w|$, we can find a pair of vertices in Γ whose merging adds w to H as a complementary generator for J. (We may take $v(p_w)$ together with the vertex on the path of s_w at distance $|p_w|+|s_w|-|w|$ from $v(s_w)$.)

□

The last claim shows, in particular, that if N is obtained from M by adding a single complementary generator, then either $\Gamma_X(N)$ is an immediate quotient or it contains $\Gamma_X(M)$ as a proper subgraph. This already proves Theorem 1.1.1 for the case $\mathrm{rk}(J)-\mathrm{rk}(H)=1$: if $H\leq_{\vec{x}} J$, we are clearly in the second case of Claim 1.3.4, i.e. J is an immediate quotient of H.

We proceed by defining handle numbers for a subgroup $M\leq_{fg}\mathbf{F}_k$ and an ordered set of words $w_1,\ldots,w_t\in\mathbf{F}_k$. Let $N=\langle M,w_1,\ldots,w_t\rangle$ and $\Delta=\Gamma_X(N)$. Let in addition $N_i=\langle M,w_1,\ldots,w_i\rangle$ and $\Gamma_i=\Gamma_X(N_i)$. We obtain a series of subgroups

$$M=N_0\leq N_1\leq\ldots\leq N_t=N,$$

and a series of graphs $\Gamma=\Gamma_0,\Gamma_1,\ldots,\Gamma_t=\Delta$. We denote by $h_X(M,w_1,\ldots,w_t)$ the t-tuple of the following handle numbers:

$$h_X(M,w_1,\ldots,w_t)\stackrel{\mathrm{def}}{=}\left(h(\Gamma_0,w_1),\,h(\Gamma_1,w_2),\,\ldots,h_X(\Gamma_{t-1},w_t)\right).$$

Let us focus now on the case where t is the cardinality of the smallest set $S\subseteq\mathbf{F}_k$ such that $N=\langle M,S\rangle$. The following lemma characterizes t-tuples of words for which the t-tuple of handle-numbers is lexicographically minimal. It is the crux of the proof of Theorem 1.1.1.

Lemma 1.3.5. *In the above notations, let $(w_1,\ldots,w_t)$ be an ordered set of complementary generators such that the tuple of handle numbers $h_X(M,w_1,\ldots,w_t)$ is lexicographically minimal. Then the zeros in $h_X(M,w_1,\ldots,w_t)$ form a prefix of the tuple.*

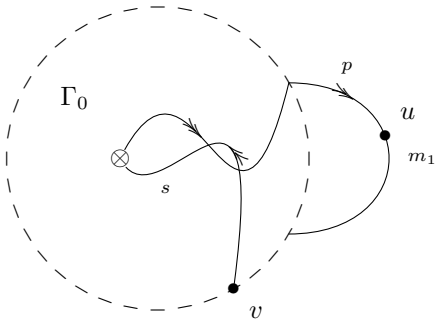
Namely, there is no zero handle-number that follows a positive handle-number.

Proof. It is enough to prove the claim for pairs of words (i.e. for $t=2$), the general case following immediately. Assume then that $N=\langle M,w_1,w_2\rangle$, that 2 is the minimal number of complementary generators of N given M, and that $h_X(M,w_1,w_2)$ is lexicographically minimal. In the above

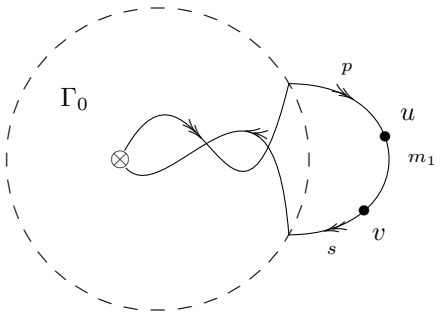
notation, assume to the contrary that $h(\Gamma_0, w_1) > 0$ and $h(\Gamma_1, w_2) = 0$. Let $m_1 = m_{w_1}$ denote the handle of w_1 in Γ_0. Thus Γ_1 was obtained from Γ_0 by adding a handle (or a balloon) representing m_1. The word w_2 can be expressed as $w_2 = ps$ so that there is a path corresponding to p in Γ_1, emanating from the basepoint and ending at some vertex u, and there is a path s to the basepoint from a vertex v. (Clearly, $u \neq v$ for otherwise $w_2 \in N_1$ contradicting the minimality of $t = 2$.) Thus Γ_2 is attained from Γ_1 by identifying the vertices u and v. By possibly multiplying w_2 from the left by a suitable element of N_1, we can assume that p does not traverse the handle m_1 "more than necessary". Namely, if u does not lie on m_1, then p avoids m_1, and if it does lie on m_1, then only the final segment of p traverses m_1 till u. The same holds for s and v (with right multiplication).

The argument splits into three possible cases.

- If both u, v belong already to Γ_0, then $h(\Gamma_0, w_2) = 0$. In this case we can switch between w_2 and w_1 to lexicographically reduce the sequence of handle numbers, contrary to our assumption.
- Consider next the case where, say, $v \in V(\Gamma_0)$ but $u \in V(\Gamma_1) \setminus V(\Gamma_0)$, i.e, u resides on the handle m_1. Then, the handle needed in order to add w_2 to Γ_0 is strictly shorter than $h(\Gamma_0, w_1) = |m_1|$. Again, by switching w_2 with w_1 the sequence of handle numbers goes down lexicographically - a contradiction.



- In the final case that should be considered both u and v are on the handle m_1. I.e. $u, v \in V(\Gamma_1) \setminus V(\Gamma_0)$. Assume w.l.o.g. that when tracing the path of m_1, u precedes v. As before we can premultiply and postmultiply w_2 by suitable elements of N_1 to guarantee the following: The path p, from the basepoint of Γ_1 to u, goes through Γ_0 and then traverses a prefix of m_1 until reaching u, and the path s from v to the basepoint traces a suffix of m_1 and then goes only through Γ_0. Again $h(\Gamma_0, w_2) < h(\Gamma_0, w_1)$, so that switching w_2 with w_1 lexicographically reduces the sequence of handle numbers. (A similar argument works in the case m_1 constitutes a balloon.)



□

Theorem 1.1.1 follows easily from this lemma. Indeed, assume that $H \overset{*}{\leq} J$ and that $H \leq_{\vec{x}} J$. Let $t = \mathrm{rk}(J) - \mathrm{rk}(H)$ denote the difference in ranks, so that t is the smallest number of complementary generators needed to obtain J given H. Let $(w_1, \ldots, w_t)$ be an ordered set of complementary generators so that $h_X(H, w_1, \ldots, w_t)$ is lexicographically minimal. Similarly to the notations above, let $J_i = \langle H, w_1, \ldots, w_i \rangle$ and $\Gamma_i = \Gamma_X(J_i)$.

By the lemma, there is some $0 \leq q \leq t$ so that $h(\Gamma_0, w_1) = \ldots = h(\Gamma_{q-1}, w_q) = 0$ whereas $h(\Gamma_q, w_{q+1}), \ldots, h(\Gamma_{t-1}, w_t)$ are all positive. By Claim 1.3.4 it follows that Γ_i is an immediate quotient of Γ_{i-1} for $1 \leq i \leq q$, and therefore $J_q \in {}_X(H)$ and $\rho_X(H, J_q) = q$. (This in fact shows that $\rho_X(H, J_q) \leq rk(J_q) - rk(H)$, and the equality follows from Lemma 1.3.3).

Using Claim 1.3.4 again, we see that Γ_i is a proper subgraph of Γ_{i+1} for $q \leq i \leq t-1$. So that Γ_q is a subgraph of $\Gamma_t = \Gamma_X(J)$. But then the image of the graph morphism $\eta : \Gamma_X(H) \to \Gamma_X(J)$ is clearly the subgraph Γ_q. If $q < t$ this is a proper subgraph, which contradicts the assumption $H \leq_{\vec{x}} J$. Hence $q = t$ and $\rho_X(H, J) = t$, as required. Together with (1.3.3) this completes the proof of Theorem 1.1.1. □

In fact, the same argument yields a more general result:

Corollary 1.3.6. *Let $H \leq J \leq \mathbf{F}_k$ be f.g. groups, and let t be the minimal number of complementary generators needed to obtain J from H. Then t is computable as follows. Let $\eta : \Gamma_X(H) \to \Gamma_X(J)$ be the unique core-graph morphism, and let M be the intermediate subgroup corresponding to the image $\eta(\Gamma_X(H))$. Then,*

$$t = \rho_X(H, M) + \mathrm{rk}(J) - \mathrm{rk}(M).$$

Proof. In the notation of the last part of the proof of Theorem 1.1.1, we see that $M = J_q \overset{*}{\leq} J$, and so $\rho_X(H, M) + \mathrm{rk}(\mathrm{J}) - \mathrm{rk}(M) = \rho_X(H, J_q) + (\mathrm{t} - \mathrm{q}) = t$. □

Remark 1.3.7. Note that in the crucial arguments of the proof of Theorem 1.1.1 we did not use the fact that the groups where of finite rank. Indeed, this result can be carefully generalized to subgroups of $\mathbf{F}_k$ of infinite rank.

Remark 1.3.8. Another way to interpret Theorem 1.1.1 is by saying that if $H \overset{*}{\leq} J$ and $H \leq_{\vec{x}} J$ with $t = \mathrm{rk}(\mathrm{J}) - \mathrm{rk}(\mathrm{H})$, then there exists some set $\{w'_1, \ldots, w'_t\}$ of complementary generators such that each w_i can be realized by merging a pair of vertices in $\Gamma_X(H)$. To see this, let $w_1, \ldots, w_t$ be as in the proof above, so w_i can be realized by merging a pair of vertices u and v in Γ_{i-1}. Let $\eta_{i-1} : \Gamma_X(H) \to \Gamma_{i-1}$ be the surjective morphism, and pick any vertices in the fibers $u' \in \eta^{-1}(u)$, $v' \in \eta^{-1}(v)$. Let w'_i be some word corresponding to the merging of u' and v' in $\Gamma_X(H)$. It is not hard to see that for each i, $\langle H, w_1, \ldots, w_i \rangle = \langle H, w'_1, \ldots, w'_i \rangle$.

1.4 More on the Primitivity Rank

Recall Definitions 1.1.7 and 1.1.8 where we defined $\pi(w)$, the primitivity rank of a word $w \in \mathbf{F}_k$, and $\pi(H)$, the primitivity rank of $H \leq_{fg} \mathbf{F}_k$. In this subsection we prove some characteristics of this categorization of formal words, and show it actually depends only on the quotients of the core graph $\Gamma_X(H)$ (or $\Gamma_X(\langle w \rangle)$). The claims are stated for subgroups, and can be easily interpreted for elements with the usual correspondence between the element $w \neq 1$ and the subgroup it generates $\langle w \rangle$. We begin by characterizing the possible values of $\pi(H)$.

Lemma 1.4.1. *Let $H \leq_{fg} \mathbf{F}_k$. Then*

$$H \overset{*}{\leq} \mathbf{F}_k \Leftrightarrow \pi(H) = \infty.$$

Proof. Recall that $\pi(H)$ is defined by the smallest rank of subgroups of $\mathbf{F}_k$ where H is contained but not as a free factor. If H is not a free factor of $\mathbf{F}_k$, then $\mathbf{F}_k$ itself is one such subgroup so that $\pi(H) \leq k < \infty$. If $H \overset{*}{\leq} \mathbf{F}_k$, Claim 1.2.5 shows it is a free factor in every other subgroup containing it. Thus, in this case $\pi(H) = \infty$. □

Corollary 1.4.2. *For every $H \leq_{fg} \mathbf{F}_k$*

$$\pi(H) \in \{0, 1, \ldots, k\} \cup \{\infty\}$$

In the definition of the primitivity rank of a subgroup H, we consider all subgroups of $\mathbf{F}_k$ containing H but not as a free factor. It turns out it is enough to consider only subgroups of $\mathbf{F}_k$ that are covered by H, that is, groups whose associated core graphs are in the X-fringe ${}_X(H)$.

Lemma 1.4.3. *For every $H \leq_{fg} \mathbf{F}_k$*

$$\pi(H) = min \left\{ rk(J) \;\middle|\; \begin{array}{c} H \leq_{\vec{x}} J \;\; and \\ H \text{ is } \textbf{\textit{not}} \text{ a free factor of } J \end{array} \right\} \tag{1.4.1}$$

Moreover, all H-critical subgroups of $\mathbf{F}_k$ are covered by H.

Proof. Recall that H-critical subgroups of $\mathbf{F}_k$ are the subgroups of smallest rank in which H is not a free factor (so in particular their rank is exactly $\pi(H)$). It is enough to show that every H-critical subgroup has its associated core graph in the fringe ${}_X(H)$.

Consider an H-critical subgroup $J \leq \mathbf{F}_k$. This J contains H but not as a free factor. By Claim 1.2.2 there exists a morphism $\eta : \Gamma_X(H) \to \Gamma_X(J)$. If η is surjective then $H \leq_{\vec{x}} J$ and $\Gamma_X(J) \in {}_X(H)$. Otherwise, consider J', the group corresponding to the core graph $\eta(\Gamma_X(H))$. This graph, $\Gamma_X(J')$, is a strict subgraph of $\Gamma_X(J)$, and so $J' \overset{*}{\lneq} J$ (see Claim 1.2.5). In particular $H \leq_{\vec{x}} J'$ and $rk(J') < rk(J)$. It is impossible that $H \overset{*}{\leq} J'$, because by transitivity this would yield that $H \overset{*}{\leq} J$. Thus, J' is a subgroup in which H is a not free factor, and of smaller rank than J. This contradicts the fact that J is H-critical. □

We note that in the terminology of [KM02, MVW07], H-critical subgroups are merely a special kind of "algebraic extensions" of H. (An algebraic extension of H is a group J such that for every M with $H \leq M \lneqq J$, M is not a free factor of J.) Specifically, H-critical subgroups are algebraic extensions of H of minimal rank, excluding H itself. Our proof actually shows the more general fact that all algebraic extensions of H can be found in the fringe (this fact appears in [KM02, MVW07]).

At this stage we can describe exactly how the primitivity rank of a subgroup $H \leq_{fg} \mathbf{F}_k$ can be computed. In fact, all algebraic extensions and critical subgroups of H can be immediately identified:

Corollary 1.4.4. *Consider the induced subgraph of $\mathcal{D}_k$ consisting of all core graphs in ${}_X(H)$. Then,*

- The algebraic extensions of H are precisely the core graphs which are not an immediate quotient of any other core graph of smaller rank.
- The H-critical subgroups are the algebraic extensions of smallest rank, excluding H itself, and $\pi(H)$ is their rank.

Proof. The second statement follows from the discussion above and from definition 1.1.8. The first statement holds trivially for H itself. If J is a proper algebraic extension of H, then by the proof of Lemma 1.4.3, $J \in {}_X(H)$. If $\Gamma_X(J)$ is an immediate quotient of some $\Gamma_X(M)$ of smaller rank, where $M \in {}_X(H)$, then $H \leq M \lneq J$ and by (the easier implication of) Theorem 1.1.1 we conclude $M \overset{*}{\leq} J$, a contradiction.

On the other hand, if $J \in {}_X(H)$ is not an algebraic extension of H, then there is some intermediate subgroup L such that $H \leq L \overset{*}{\lneqq} J$. We can assume $L \in {}_X(H)$ for otherwise it can be replaced with L' corresponding to the image of the morphism $\eta : \Gamma_X(H) \to \Gamma_X(L)$ (whence $L' \in {}_X(H)$ and $H \leq L' \overset{*}{\leq} L \overset{*}{\lneqq} J$). From (the harder implication of) Theorem 1.1.1 it follows that $\rho_X(L, J) = \mathrm{rk}(J) - \mathrm{rk}(L)$. The prior-to-last element in a shortest path in $\mathcal{D}_k$ from $\Gamma_X(L)$ to $\Gamma_X(J)$ is then a proper free factor of J at distance 1 that belongs to ${}_X(H)$. □

As an example, consider $H = \langle x_1 x_2 x_1^{-1} x_2^{-1} \rangle$. The full lattice of groups in $_X(H)$ is given in Figure 1.3.1. There is one group of rank 1 (H itself), 5 of rank 2 and one of rank 3. The only group in the lattice where H in not a free factor is the group $\langle x_1 x_2 \rangle = \mathbf{F}_2$, of rank 2, so $\pi(H) = 2$. (And the set of algebraic extensions of H is precisely $\{H, \mathbf{F}_2\}$.)

1.5 The Calculation of ϕ

The proof of Proposition 1.1.9, as well as the reasoning that underlies Conjecture 1.1.10, are based on the fact that both $\phi(H)$ and $\pi(H)$ can be calculated by analyzing $_X(H)$, the set of quotients of $\Gamma_X(H)$. In the previous section it was shown how $\pi(H)$ is determined from $_X(H)$. In this section we show how $\phi(H)$ can be calculated by a simple analysis of the very same set. The origins of the algorithm we present here can be traced to [Nic94] with further development in [LP10]. We present it here from a more general perspective.

Let the group G act on a set Y and let $y_0 \in Y$ be a fixed element. Consider a random homomorphism $\alpha_G \in Hom(F_k, G)$. The core graphs in $_X(H)$ can be used to calculate the probability that $\alpha_G(H) \subset G_{y_0}$ (where G_{y_0} is the stabilizer of the element y_0). The quotients of the core graph $\Gamma_X(H)$ correspond to all the different "coincidence patterns" of the paths of y_0 through the action of the images of all $w \in H$, thereby describing disjoint events whose union is the event that $\alpha_G(H) \subset G_{y_0}$.

The idea is that in order to determine whether $\alpha_G(w)$ stabilizes y_0 for some $w \in \mathbf{F}_k$, we do not need to know all the values $\alpha_G(x_i)$ over $x_i \in X$ (the given basis of $\mathbf{F}_k$). Rather, we only need to know how $\alpha_G(x_i)$ acts on certain elements in Y, those in the path of y_0 through $\alpha_G(w)$. Namely, if $w = x_{j_1}^{\epsilon_1} \dots x_{j_{|w|}}^{\epsilon_{|w|}}$, $j_i \in \{1, \dots, k\}, \epsilon_i \in \{\pm 1\}$, we need to uncover the elements $y_1, \dots, y_{|w|}$ in the following diagram:

$$y_0 \xrightarrow{\alpha_G\left(x_{j_1}^{\epsilon_1}\right)} y_1 \xrightarrow{\alpha_G\left(x_{j_2}^{\epsilon_2}\right)} y_2 \xrightarrow{\alpha_G\left(x_{j_3}^{\epsilon_3}\right)} \cdots \xrightarrow{\alpha_G\left(x_{j_{|w|-1}}^{\epsilon_{|w|-1}}\right)} y_{|w|-1} \xrightarrow{\alpha_G\left(x_{j_{|w|}}^{\epsilon_{|w|}}\right)} y_{|w|}$$

That is, the image of $x_{j_1}^{\epsilon_1}$ acts on y_0, and we denote the resulting element by $y_1 \in Y$. The image of y_1 under the action of $\alpha_G\left(x_{j_2}^{\epsilon_2}\right)$ is denoted by y_2, etc. Then, y_0 is a fixed point of $\alpha_G(w)$ iff $y_{|w|} = y_0$.

There are normally many possible series of elements $y_1, \dots, y_{|w|-1} \in Y$ that can constitute the path of y_0 through $\alpha_G(w)$ such that y_0 is a fixed point. We divide these different series to a *finite* number of categories based on the *coincidence pattern* of this series. Namely, two realizations of this series, $y_1, \dots, y_{|w|-1}$, and $y'_1, \dots, y'_{|w|-1}$ are in the same category iff for every $i, j \in \{0, \dots, |w|-1\}$, $y_i = y_j \Leftrightarrow y'_i = y'_j$ (note that the elements of the series are also compared to y_0). In other words, every coincidence pattern corresponds to some partition of $\{0, 1, \dots, |w|-1\}$.

However, because the elements $\alpha_G(x_j) \in G$ act as permutations on Y, not every partition yields a realizable coincidence pattern: if, for example, $x_{j_2}^{\epsilon_2} = x_{j_7}^{-\epsilon_7}$, and $y_1 = y_7$, we must also have $y_2 = y_6$. This condition should sound familiar. Indeed, for each coincidence pattern we can draw a pointed, directed, edge-labeled graph describing it. The vertices of this graph correspond to blocks in the partition of $\{0, 1, \dots, |w|-1\}$, the basepoint corresponding to the block containing 0. Then, for each $i \in \{1, \dots, |w|\}$ there is a j_i-edge, directed according to ϵ_i, between the block of $i-1$ to the block of i. The constraints that coincidence patterns must satisfy then becomes the very same ones we had encountered in our discussion of core graphs. Namely, no two j-edges share the same origin or the same terminus.

Thus, the different realizable coincidence patterns of the series $y_0, y_1, \dots, y_{|w|-1}$ are exactly those described by core graphs that are quotients of $\Gamma_X(\langle w \rangle)$. For instance, there are exactly seven realizable coincidence patterns that correspond to the event in which y_0 is a fixed point of $\alpha_G(w)$ when $w = [x_1, x_2]$. The seven core graphs in Figure 1.3.1 correspond to these seven coincidence patterns.

Finally, the same phenomenon generalizes to any $H \leq_{fg} \mathbf{F}_k$. Instead of uncovering the path of y_0 through the image of a single word, we uncover the paths trough all words in H. The union of these paths in which y_0 is stabilized by all elements of H is depicted exactly by the core graph $\Gamma_X(H)$. The realizable coincidence patterns correspond then to the quotients of $\Gamma_X(H)$, namely to ${}_X(H)$. To summarize:

$$Prob\big[\alpha_G(H) \subset G_{y_0}\big] = \sum_{\Gamma \in {}_X(H)} Prob\big[\begin{smallmatrix}\Gamma \text{ describes the coincidence pattern}\\ \text{of } y_0 \text{ through the action of } \alpha_G(H)\end{smallmatrix}\big] \tag{1.5.1}$$

The advantage of the symmetric group S_n with its action on $\{1,\dots,n\}$ is that the probabilities in the r.h.s. of (1.5.1) are very easy to formulate. Let $\alpha_n = \alpha_{S_n} \in Hom(\mathbf{F}_k, S_n)$ be a uniformly distributed random homomorphism, and let $\Gamma \in {}_X(H)$ be one of the quotients of $\Gamma_X(H)$. Denote by $P_\Gamma(n)$ the probability that $\alpha_n(H) \subset (S_n)_1$ and that the coincidence pattern of the paths of 1 through the elements $\alpha_G(H)$ are described by Γ. Then we can give an exact expression for $P_\Gamma(n)$ in terms of v_Γ, e_Γ and e^j_Γ, the number of vertices, edges and j-edges in Γ:

There are $(n-1)(n-2)\dots(n-v_\Gamma+1)$ possible assignments of different elements from $\{2,3,\dots,n\}$ to the vertices of Γ (excluding the basepoint which always corresponds to the element 1). Then, for a given assignment, there are exactly e^j_Γ constraints on the permutation $\alpha_n(x_j)$. So the probability that the permutation $\alpha_n(x_j)$ agrees with the given assignment is

$$\frac{(n-e^j_\Gamma)!}{n!} = \frac{1}{n(n-1)\dots(n-e^j_\Gamma+1)}$$

(for $n \geq e^j_\Gamma$). Thus

$$P_\Gamma(n) = \frac{(n-1)(n-2)\dots(n-v_\Gamma+1)}{\prod_{j=1}^k n(n-1)\dots(n-e^j_\Gamma+1)}$$

Recall the definition of $\Phi_H(n)$ in (1.1.1). Since for every j and every $\Gamma \in {}_X(H)$ we have $e^j_\Gamma \leq e^j_{\Gamma_X(H)}$ we can summarize and say that for every $n \geq \max_j e^j_{\Gamma_X(H)}$ (in particular for every $n \geq v_{\Gamma_X(H)}$), we have:

$$\begin{aligned}
\Phi_H(n) &= Prob\left[\forall w \in H \quad \alpha_n(w)(1) = 1\right] - \frac{1}{n^{rk(H)}} \\
&= Prob\left[\alpha_n(H) \subset (S_n)_1\right] - \frac{1}{n^{rk(H)}} \\
&= -\frac{1}{n^{rk(H)}} + \sum_{\Gamma \in {}_X(H)} \frac{(n-1)(n-2)\dots(n-v_\Gamma+1)}{\prod_{j=1}^k n(n-1)\dots(n-e^j_\Gamma+1)} \\
&= -\frac{1}{n^{rk(H)}} + \sum_{\Gamma \in {}_X(H)} \frac{1}{n^{e_\Gamma - v_\Gamma + 1}} \frac{(1-\frac{1}{n})(1-\frac{2}{n})\dots(1-\frac{v_\Gamma-1}{n})}{\prod_{j=1}^k (1-\frac{1}{n})\dots(1-\frac{e^j_\Gamma-1}{n})}
\end{aligned} \tag{1.5.2}$$

For instance, for $H = \langle [x_1, x_2] \rangle$ there are seven summands in the r.h.s. of (1.5.2), corresponding to the seven core graphs in Figure 1.3.1. If we go over these core graphs from top to bottom and left to right, we obtain that for every $n \geq 2$:

$$\begin{aligned}
\Phi_{\langle [x_1,x_2] \rangle}(n) = &-\frac{1}{n} + \frac{(n-1)(n-2)(n-3)}{n(n-1)\cdot n(n-1)} + \\
&+ \frac{n-1}{n(n-1)\cdot n} + \frac{n-1}{n\cdot n(n-1)} + \frac{(n-1)(n-2)}{n(n-1)\cdot n(n-1)} + \\
&+ \frac{(n-1)(n-2)}{n(n-1)\cdot n(n-1)} + \frac{n-1}{n(n-1)\cdot n(n-1)} + \frac{1}{n\cdot n} \\
= &-\frac{1}{n} + \frac{1}{n-1} = \frac{1}{n(n-1)}
\end{aligned}$$

Recall the definition of the second categorization of sets of free words, $\phi(H)$, in (1.1.3). Indeed, we can rewrite (1.5.2) as a power series in $\frac{1}{n}$, and obtain that (for large enough n)

$$\Phi_H(n) = \sum_{i=0}^{\infty} \frac{a_i(H)}{n^i}$$

where the coefficients $a_i(H)$ depend only on H. We need not consider negative values of i because the leading term of every summand in (1.5.2) is $\frac{1}{n^{e_\Gamma - v_\Gamma + 1}}$, and $e_\Gamma - v_\Gamma + 1$ is non-negative for connected graphs. In fact, this number also equals the rank of the free subgroup corresponding to Γ.

The value of $\phi(H)$ equals the smallest i for which $a_i(H)$ does not vanish. For instance, for $H = \langle [x_1, x_2] \rangle$ we have

$$\Phi_{\langle [x_1,x_2] \rangle}(n) = \frac{1}{n(n-1)} = \sum_{i=2}^{\infty} \frac{1}{n^i}$$

so that $a_0(H) = a_1(H) = 0$ and $a_i(H) = 1$ for $i \geq 2$. Hence $\phi(H) = 2$.

In fact, we can write a power series for each $\Gamma \in {}_X(H)$ separately, and obtain:

$$\begin{aligned} P_\Gamma(n) &= \frac{1}{n^{e_\Gamma - v_\Gamma + 1}} \frac{(1-\frac{1}{n})(1-\frac{2}{n}) \dots (1-\frac{v_\Gamma - 1}{n})}{\prod_{j=1}^{k} (1-\frac{1}{n}) \dots (1-\frac{e_\Gamma^j - 1}{n})} \\ &= \frac{1}{n^{e_\Gamma - v_\Gamma + 1}} \big(1 - \frac{\binom{v_\Gamma}{2} - \sum_{j=1}^{k} \binom{e_\Gamma^j}{2}}{n} + O\big(\frac{1}{n^2}\big)\big) \end{aligned} \tag{1.5.3}$$

This shows that if $\Gamma = \Gamma_X(J)$ ($J \leq_{fg} \mathbf{F}_k$), then $P_\Gamma(n)$ never affects $a_i(H)$-s with $i < rk(J)$. It is also easy to see that all the coefficients of the power series expressing $P_\Gamma(n)$ are integers. We summarize:

Claim 1.5.1. *For every $H \leq_{fg} \mathbf{F}_k$, all the coefficients $a_i(H)$ are integers.*
Moreover, $a_i(H)$ is completely determined by core graphs in ${}_X(H)$ corresponding to groups of rank $\leq i$.

1.6 Relations between $\pi(\cdot)$ and $\phi(\cdot)$

We now have all the background needed for the proof of Proposition 1.1.9 and consequently of Theorem 1.1.5. We need to show that for every $H \leq_{fg} \mathbf{F}_k$ and every $i \leq rk(H) + 1$, we have

$$\pi(H) = i \iff \phi(H) = i.$$

The proof is divided into three steps. First we deal with the case $i < rk(H)$, then with $i = rk(H)$. The last case $i = rk(H) + 1$ is by far the hardest.

Lemma 1.6.1. *Let $H \leq_{fg} \mathbf{F}_k$ and $i < rk(H)$. Then*

(1) $\pi(H) = i \Leftrightarrow \phi(H) = i$

(2) If $\pi(H) = \phi(H) = i$ then $a_i(H)$ equals the number of H-critical subgroups of $\mathbf{F}_k$.

Proof. Let m denote the smallest rank of a group $J \leq \mathbf{F}_k$ such that $H \leq_{\vec{x}} J$ (so $m \leq \mathrm{rk}\,(\mathrm{H})$). The first part of the result is derived from the observation that both $\pi(H) = i$ and $\phi(H) = i$ iff $m = i$. Let us note first that $\pi(H) = i \Leftrightarrow m = i$. This follows from Lemma 1.4.3 and the fact that H cannot be a free factor in a subgroup of smaller rank.

We next observe that $\phi(H) = i \Leftrightarrow m = i$: If $m < rk(H)$ then by (1.5.2) and (1.5.3), m is indeed the smallest index for which $a_m(H)$ does not vanish (this does not work for $m = rk(H)$ because

of the term $\left(-\frac{1}{n^{rk(H)}}\right)$ in the definition of $\Phi_H(n)$). Conversely, if $m = rk(H)$ then obviously $\phi(H) \geq rk(H)$.

For the second part of the lemma, recall that H is not a free factor in any subgroup of smaller rank containing it. Thus, by (1.5.3) and Lemma 1.4.3, both $a_i(H)$ and the number of subgroups of rank i containing H equal the number of subgroups of rank i in ${}_X(H)$. □

The case $i = rk(H)$ is slightly different, but almost as easy.

Lemma 1.6.2. *Let $H \leq_{fg} \mathbf{F}_k$. Then,*

(1) $\pi(H) = rk(H) \Leftrightarrow \phi(H) = rk(H)$

(2) If $\pi(H) = \phi(H) = rk(H)$ then $a_{rk(H)}(H)$ equals the number of H-critical subgroups of $\mathbf{F}_k$.

Proof. From Lemma 1.6.1 we infer that $\pi(H) \geq rk(H) \Leftrightarrow \phi(H) \geq rk(H)$. So we assume that indeed $\pi(H), \phi(H) \geq rk(H)$, or, equivalently, that there are no subgroups covered by H of rank smaller than $rk(H)$.

We show that both sides of part (1) are equivalent to the existence of a quotient (corresponding to a subgroup) of rank $rk(H)$ in ${}_X(H)$ *other* than $\Gamma_X(H)$ itself. Indeed, this is true for $\pi(H)$ because the only free product of H of rank $rk(H)$ is H itself.

As for $\phi(H)$, this is true because when $\phi(H) \geq rk(H)$ it is easily verified that the value of $a_{rk(H)}(H)$ equals the number of quotient in ${}_X(H)$ of rank $rk(H)$ minus 1 (this minus 1 comes from the term $\left(-\frac{1}{n^{rk(H)}}\right)$). We think of this term as offsetting the contribution of $\Gamma_X(H)$ to $a_{rk(H)}(H)$, so $a_{rk(H)}(H)$ equals the number of other quotients in ${}_X(H)$ of rank $rk(H)$.

The second part of the lemma is true because all H-critical subgroups are covered by H (Lemma 1.4.3). □

1.6.1 The Case $i = rk(H) + 1$

The most interesting (and the hardest) case of Theorem 1.1.5 is when $rk(H) = k - 1$. In the previous analysis this corresponds to $i = rk(H) + 1$.

Lemma 1.6.3. *Let $H \leq_{fg} \mathbf{F}_k$. Then,*

(1) $\pi(H) = rk(H) + 1 \Leftrightarrow \phi(H) = rk(H) + 1$

(2) If $\pi(H) = \phi(H) = rk(H) + 1$ then $a_{rk(H)+1}(H)$ equals the number of H-critical subgroups of $\mathbf{F}_k$.

Denote by $\hat{\Gamma} = \Gamma_X(H)$ the associated core graph. By Lemmas 1.6.1 and 1.6.2, we can assume that $\pi(H), \phi(H) \geq rk(H) + 1$. In particular, we can thus assume that H is not contained in any subgroup of rank smaller than $rk(H) + 1$ other than H itself.

The coefficient $a_{rk(H)+1}(H)$ in the expression of $\Phi_H(n)$ is the sum of two expressions:

- The contribution of $\hat{\Gamma}$ which equals $-\left(\binom{v_{\hat{\Gamma}}}{2} - \sum_{j=1}^{k}\binom{e^j_{\hat{\Gamma}}}{2}\right)$
- A contribution of 1 from each core graph of rank $rk(H) + 1$ in ${}_X(H)$

Thus, our goal is to show that the contribution of $\hat{\Gamma}$ is exactly offset by the contribution of the core graphs of rank $rk(H) + 1$ in ${}_X(H)$ in which H is a free factor. This would then yield immediately both parts of Lemma 1.6.3. But the number of subgroups of rank $rk(H) + 1$ (in ${}_X(H)$) in which H is a free factor equals exactly the number of immediate quotients of $\hat{\Gamma}$: Theorem 1.1.1 shows that only immediate quotients of $\hat{\Gamma}$ are subgroups of rank $rk(H) + 1$ in which H is a free factor. On the other hand, (1.3.1) and the assumption that H in not contained in any other subgroup of equal or smaller rank yield that every immediate quotient of $\hat{\Gamma}$ is of rank $rk(H) + 1$ (and H is a free factor in it).

Thus, Lemma 1.6.3 follows from the following lemma.

Lemma 1.6.4. *Assume $\pi(H), \phi(H) > rk(H)$. Then $\hat{\Gamma} = \Gamma_X(H)$ has exactly*

$$\binom{v_{\hat{\Gamma}}}{2} - \sum_{j=1}^{k} \binom{e^j_{\hat{\Gamma}}}{2}$$

immediate quotients.

The intuition behind Lemma 1.6.4 is this: Every immediate quotient of $\hat{\Gamma}$ is generated by identifying some pair of vertices of $\hat{\Gamma}$, and there are exactly $\binom{v_{\hat{\Gamma}}}{2}$ such pairs. But for every pair of equally-labeled edges of $\hat{\Gamma}$, the pair of origins generates the same immediate quotient as the pair of termini. This intuition needs, however, some justification that we now provide.

To this end we use the graph Υ, a concept introduced in [LP10]†. This graph represents the pairs of vertices of $\hat{\Gamma}$ and the equivalence relations between them induced by equally-labeled edges. There are $\binom{v_{\hat{\Gamma}}}{2}$ vertices in Υ, one for each unordered pair of vertices of $\hat{\Gamma}$. The number of directed edges in Υ is $\sum_{j=1}^{k} \binom{e^j_{\hat{\Gamma}}}{2}$, one for each pair of equally-labeled edges in $\hat{\Gamma}$. The edge corresponding to the pair $\{\epsilon_1, \epsilon_2\}$ of j-edges is a j-edge connecting the vertex $\{origin(\epsilon_1), origin(\epsilon_2)\}$ to $\{terminus(\epsilon_1), terminus(\epsilon_2)\}$. For example, when S consists of the commutator word, Υ has $\binom{4}{2} = 6$ vertices and $\binom{2}{2} + \binom{2}{2} = 2$ edges. We illustrate a slightly more interesting case in Figure 1.6.1.

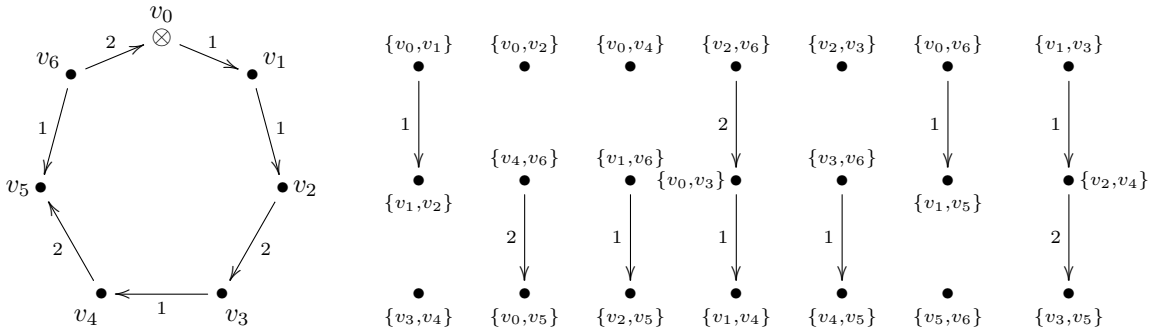


Figure 1.6.1: The graph Υ (on the right) corresponding to $\hat{\Gamma} = \Gamma_X(H)$ (on the left) for $H = \langle x_1^2 x_2 x_1 x_2 {x_1}^{-1} x_2 \rangle$. (The vertices of $\hat{\Gamma}$ are denoted here by $v_0, \ldots, v_6$.)

We denote the set of connected components of Υ by $Comp(\Upsilon)$. The proof of Lemma 1.6.4 will follow from two facts that we show next. Namely, Υ has exactly $\binom{v_{\hat{\Gamma}}}{2} - \sum_{j=1}^{k} \binom{e^j_{\hat{\Gamma}}}{2}$ connected components. Also, there is a one-to-one correspondence between $Comp(\Upsilon)$ and the set of immediate quotients of $\hat{\Gamma}$.

Claim 1.6.5. *If $\pi(H), \phi(H) > rk(H)$, then*

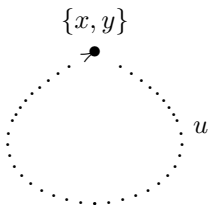
$$\left|Comp(\Upsilon)\right| = \binom{v_{\hat{\Gamma}}}{2} - \sum_{j=1}^{k} \binom{e^j_{\hat{\Gamma}}}{2}$$

Proof. Since Υ has $\binom{v_{\hat{\Gamma}}}{2}$ vertices and $\sum_{j=1}^{k} \binom{e^j_{\hat{\Gamma}}}{2}$ edges, it is enough to show that it is a forest, i.e., it contains no cycles.

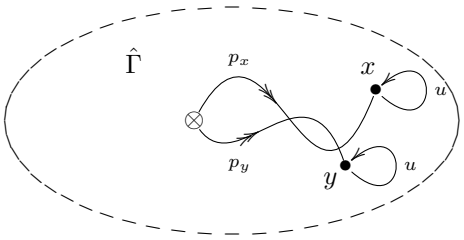
Let $C \in Comp(\Upsilon)$ be some component of Υ. Clearly, every vertex in C (which corresponds to a pair of vertices in $\hat{\Gamma}$) generates the same immediate quotient. Denote this quotient by $\Delta(C)$,

†This is a variation of the classical construction of *pull-back* of graphs (in this case the pull-back of the graph $\hat{\Gamma}$ with itself).

and the corresponding subgroup by J. In particular, $rk(J) = rk(H) + 1$ (recall that under the claim's assumptions, H is not contained in any other subgroup of smaller or equal rank).Assume to the contrary that C contains a cycle. Edges in Υ are directed and labeled, so there is an element $u \in \mathbf{F}_k$ which corresponds to this cycle started, say, at the vertex $\{x, y\}$.



Where do we get as we walk in the core graph $\hat{\Gamma}$ starting at x (resp. y) and following the path corresponding to u? One possibility is that the walk from x returns back to x and likewise for y. Alternatively this u-walk can take us from x to y and from y to x. We consider only the former possibility. The latter case would be handled by considering the walk corresponding to u^2. Let $p_x, p_y \in \mathbf{F}_k$ be words corresponding to some paths from the basepoint of $\hat{\Gamma}$ to x, y respectively. In particular, $p_x u p_x^{-1}, p_y u p_y^{-1} \in H$.



Merging x and y is equivalent to adding the generator $p_x p_y^{-1}$ to H, so that $J = \langle H, p_x p_y^{-1} \rangle$. Since $rk(J) = rk(H) + 1$, we have that $J = H * \langle p_x p_y^{-1} \rangle$. Consider the elements $h_1 = p_x u p_x^{-1} \in H$ and $h_2 = p_y u p_y^{-1} \in H$. The following equality holds:

$$h_1 = p_x u p_x^{-1} = (p_x p_y^{-1}) p_y u p_y^{-1} (p_x p_y^{-1})^{-1} = (p_x p_y^{-1}) h_2 (p_x p_y^{-1})^{-1}$$

This is a contradiction, since we obtained two different expressions for h_1 in the free product $J = H * \langle p_x p_y^{-1} \rangle$. □

Remark 1.6.6. Let $H \leq \mathbf{F}_k$ and $x, y \in V(\Gamma_X(H))$. In the proof of the last claim it was shown that if there is some $1 \neq u \in \mathbf{F}_k$ which is readable as a closed path at both x and y, then the subgroup we obtain by merging them is *not* a free extension of H. We stress that the converse is not true. For example, consider $H = \langle a, bab \rangle \leq \mathbf{F}(\{a, b\})$. Then $\Gamma_{\{a,b\}}(H)$ has three vertices, no pair of which share a common closed path (in other words, the corresponding graph Υ has no cycles). However, $\Gamma_{\{a,b\}}(H)$ has exactly two immediate quotients, none of which is a free extension.

Next we exhibit a one-to-one correspondence between $Comp(\Upsilon)$ and the immediate quotients of $\hat{\Gamma}$. It is very suggestive to try and restore C from $\Delta(C)$ by simply signaling out the pairs of vertices that are identified in $\Delta(C)$. But this is too naive. There may be pairs of vertices not in C that are identified in $\Delta(C)$. For instance, consider C, the rightmost component of Υ in Figure 1.6.1. In $\Delta(C)$ we merge v_1 and v_3 but also v_3 and v_5. Thus v_1 and v_5 are merged and likewise all pairs in the component of $\{v_1, v_5\}$.

However, simple group-theoretic arguments do yield this sought-after result:

Claim 1.6.7. *If $\pi(H), \phi(H) > rk(H)$, then there is a one-to-one correspondence between $Comp(\Upsilon)$ and the set of immediate quotients of $\hat{\Gamma} = \Gamma_X(H)$.*

Proof. Maintaining the above notation, we need to show that the mapping from $C \in Comp(\Upsilon)$ to $\Delta(C)$, the immediate quotient generated by any of the pairs in C, is one to one.

Denote by J the subgroup corresponding to the immediate quotient $\Delta(C)$. Let $\{x, y\}$ be some vertex in C, and $p_x, p_y \in \mathbf{F}_k$ words corresponding to some paths from the basepoint of $\hat{\Gamma}$ to x, y, respectively. Let also $q = p_x p_y^{-1} \in \mathbf{F}_k$. As we saw above,

$$J = \langle H, q \rangle$$

and clearly $q \notin H$.

We claim that any other complementary generator of J over H is in same (H, H)-double-coset of q or of q^{-1} in J. Namely, if $J = \langle H, q' \rangle$ then $q' \in HqH \cup Hq^{-1}H$. To see this, let Y be some basis of H and think of J as the free group over the basis $Y \cup \{q\}$ (this is true because $rk(J) = rk(H) + 1$). Now think of q' as a word in the elements of this basis. Multiplying from the right or left by elements of Y does not affect the (H, H)-double-coset, so assume w.l.o.g. that q' begins and ends with either q or q^{-1}. But then the set $Y \cup \{q'\}$ is Nielsen-reduced with respect to the basis $Y \cup \{q\}$ (see, for instance, the definition in Chapter 1 of [LS70]). As consequence, $Y \cup \{q'\}$ equals $Y \cup \{q\}$ up to taking inverses (Proposition 2.8 therein). Thus $q' = q$ or $q' = q^{-1}$.

So let $\{a, b\}$ be another pair of vertices generating $\Delta(C)$. We show that it belongs to C. Let p_a, p_b be words in $\mathbf{F}_k$ corresponding to paths in $\hat{\Gamma}$ from the basepoint to a, b respectively. We have $\langle H, p_a p_b^{-1} \rangle = J$, so $p_a p_b^{-1} \in HqH \cup Hq^{-1}H$. W.l.o.g. it is in HqH (otherwise switch a and b). So assume $p_a p_b^{-1} = h_1 q h_2$ with $h_1, h_2 \in H$. But $h_1^{-1} p_a$ is also a path from the basepoint of $\hat{\Gamma}$ to a, and likewise $h_2 p_b$ a path to b. Choosing if needed these paths instead of p_a, p_b we can assume that

$$p_a p_b^{-1} = q = p_x p_y^{-1}.$$

Thus,

$$p_a^{-1} p_x = p_b^{-1} p_y.$$

This shows that there is a path in $\hat{\Gamma}$ from a to x corresponding to a path from b to y. This shows precisely that the pair $\{a, b\}$ is in the same component of Υ as $\{x, y\}$, namely, in C. □

This completes the proof of Lemma 1.6.3. This Lemma, together with Lemmas 1.6.2 and 1.6.1, yields Proposition 1.1.9 and thus Theorem 1.1.5.

1.6.2 Further Relations between $\pi(\cdot)$ and $\phi(\cdot)$

Let us take another look now at Conjecture 1.1.10. It posits that the results described in Proposition 1.1.9 hold for all values of $\pi(\cdot)$ and $\phi(\cdot)$. To understand what this means, suppose that H is a free factor in all the quotients in $_X(H)$ of ranks up to $i-1$. What can be said about rank-i quotients in which H is a free factor? The conjecture states that their number exactly offsets the sum of two terms: The contribution to $a_i(H)$ of the quotients of smaller rank and of the term $\frac{-1}{n^{rk(H)}}$ when $i = rk(H)$. For instance, $\pi(H) = 3$ for $H = \langle x_1^2 x_2^2 x_3^2 \rangle$. In particular, H is a free factor of all quotients in $_X(H)$ of rank ≤ 2. There is a single H-critical subgroup ($\mathbf{F}_3$ itself), and additional 13 quotients of rank 3 in which H is a free factor. The contribution of quotients of rank ≤ 2 to $a_3(H)$ is indeed exactly (-13).

Interestingly enough, this is indeed the case for every free factor $H \overset{*}{\leq} \mathbf{F}_k$. In this case, since free factors are measure preserving, we get that $\phi(H) = \infty$, so $a_i(H) = 0$ for every i, and the statement of the previous paragraph holds. For the general case the conjecture states that as long as we consider low-rank quotients and "imprimitivity has not been revealed yet", the situation does not differ from what is seen in the primitive case.

We finish this section by stating another result connecting $\pi(\cdot)$ and $\phi(\cdot)$. It shows an elegant property of both of them that lends further support to our belief in Conjecture 1.1.10.

Two words $w_1, w_2 \in \mathbf{F}_k$ are called *disjoint* (with respect to a given basis) if they share no common letters.

Lemma 1.6.8. *Let $w_1, w_2 \in \mathbf{F}_k$ be disjoint. Then*

$$\pi(w_1 w_2) = \pi(w_1) + \pi(w_2)$$
$$\phi(w_1 w_2) = \phi(w_1) + \phi(w_2)$$

Moreover, $a_{\phi(w_1 w_2)}(w_1 w_2) = a_{\phi(w_1)}(w_1) \cdot a_{\phi(w_2)}(w_2)$, and if part 2 of Conjecture 1.1.10 holds for $H = \langle w_1 \rangle$ and for $H = \langle w_2 \rangle$, then it also holds for $H = \langle w_1 w_2 \rangle$.

This lemma is essentially outside the scope of the present paper, so we only sketch its proof. Let $\alpha_n \in Hom(\mathbf{F}_k, S_n)$ be a random homomorphism chosen with uniform distribution. As w_1 and w_2 are disjoint, the random permutations $\alpha_n(w_1)$ and $\alpha_n(w_2)$ are independent. The claims about the additivity of $\phi(\cdot)$ and the multiplicativity of $a_{\phi(\cdot)}(\cdot)$ are easy to derive by calculating the probability that 1 is a fixed point of $w_1 w_2$. The key fact in this calculation is the aforementioned independence of $\alpha_n(w_1)$ and $\alpha_n(w_2)$.

The other claims in the lemma follow from an analysis of H-critical subgroups. By considering properties of the associated core graphs it is not hard to show that $J \leq \mathbf{F}_k$ is $\langle w_1 w_2 \rangle$-critical iff it is the free product of a $\langle w_1 \rangle$-critical subgroup and a $\langle w_2 \rangle$-critical subgroup.

1.7 Primitive Words and the Profinite Completion

Most of the standard facts below about profinite groups and particularly free profinite groups can be found with proofs in [Wil98] (in particular Section 5.1).

A profinite group is a topological group G with any of the following equivalent properties:

- G is the inverse limit of an inverse system of finite groups.
- G is compact, Hausdorff and totally disconnected.
- G is isomorphic (as a topological group) to a closed subgroup of a Cartesian product of finite groups.
- G is compact and $\bigcap(N | N \lhd_O G) = 1$

The *free profinite group* on a finite set X is a profinite group F together with a map $j : X \to F$ with the following universal property: whenever $\xi : X \to G$ is a map to a profinite group G, there is a unique (continuous) homomorphism $\bar{\xi} : F \to G$ such that $\xi = \bar{\xi} j$. Such F exists for every X and is unique up to a (continuous) isomorphism. We call $j(X)$ a basis of F. It turns out that every two bases of F have the same size which is called the *rank* of F. The free profinite group of rank k is denoted by $\hat{\mathbf{F}}_k$. An element $w \in \widehat{\mathbf{F}_{kk}}$ is *primitive* if it belongs to some basis.

It is a standard fact that $\widehat{\mathbf{F}_{kk}}$ is the profinite completion of $\mathbf{F}_k$ and $\mathbf{F}_k$ is naturally embedded in $\widehat{\mathbf{F}_{kk}}$. Moreover, every basis of $\mathbf{F}_k$ is then also a basis for $\widehat{\mathbf{F}_{kk}}$, so a primitive word $w \in \mathbf{F}_k$ is also primitive as an element of $\widehat{\mathbf{F}_{kk}}$. It is conjectured that the converse also holds:

Conjecture 1.7.1. *A word $w \in \mathbf{F}_k$ is primitive in $\widehat{\mathbf{F}_{kk}}$ iff it is primitive in $\mathbf{F}_k$.*

This conjecture, if true, immediately implies the following one:

Conjecture 1.7.2. *The set of primitive elements in $\mathbf{F}_k$ form a closed set in the profinite topology.*

Conjecture 1.1.4 implies these last two conjectures (it is in fact equivalent to Conjecture 1.7.1, see below): we define measure preserving elements in $\widehat{\mathbf{F}_{kk}}$ as before. Namely, an element $w \in \widehat{\mathbf{F}_{kk}}$ is measure preserving if for any finite group G and a uniformly distributed random (continuous) homomorphism $\hat{\alpha}_G \in Hom(\widehat{\mathbf{F}_{kk}}, G)$, the image $\hat{\alpha}_G(w)$ is uniformly distributed in G. Clearly, an element of $\mathbf{F}_k$ is measure preserving w.r.t $\mathbf{F}_k$ iff this holds w.r.t. $\widehat{\mathbf{F}_{kk}}$.

As in the abstract case, a primitive element of $\widehat{\mathbf{F}_{kk}}$ is measure preserving. Conjecture 1.1.4 would therefore imply that if $w \in \mathbf{F}_k$ is primitive in $\widehat{\mathbf{F}_{kk}}$, then w is also primitive w.r.t. $\mathbf{F}_k$. In particular, Theorem 1.1.5 yields:

Corollary 1.7.3. *Let $S \subset \mathbf{F}_k$ be a finite subset of cardinality $|S| \geq k-1$. Then,*

$$S \text{ can be extended to a basis in } \widehat{\mathbf{F}_{kk}} \iff S \text{ can be extended to a basis in } \mathbf{F}_k$$

In particular, for every $w \in \mathbf{F}_2$:

$$w \text{ is primitive in } \widehat{\mathbf{F}_{k2}} \iff w \text{ is primitive in } \mathbf{F}_2$$

This corollary yields, in turn, Corollary 1.1.6, which states the special case of Conjecture 1.7.2 for $\mathbf{F}_2$.

As shown by Chen Meiri (unpublished), Conjectures 1.7.1 and 1.1.4 are equivalent. With his kind permission we explain this result in this section. Meiri showed that in $\widehat{\mathbf{F}_{kk}}$ primitivity and measure preservation are equivalent (Proposition 1.7.4 below). Thus, $w \in \mathbf{F}_k$ is primitive as an element of $\widehat{\mathbf{F}_{kk}}$ iff it is measure preserving.

Proposition 1.7.4. *[C. Meiri, unpublished] Let w belong to $\widehat{\mathbf{F}_{kk}}$. Then*

$$w \text{ is primitive} \iff w \text{ is measure preserving}$$

Proof. The ($\Rightarrow$) implication is trivial as in the abstract case: for every finite group G and every basis $x_1, \ldots, x_k$ of $\widehat{\mathbf{F}_{kk}}$ there is a bijection

$$\begin{aligned} Hom(\widehat{\mathbf{F}_{kk}}, G) &\xrightarrow{\cong} G^k \\ \alpha_G &\mapsto (\alpha_G(x_1), \ldots, \alpha_G(x_k)) \end{aligned}$$

For the other direction, for every $w \in \widehat{\mathbf{F}_{kk}}$, finite group G and $g \in G$ define

$$\begin{aligned} H_w(G,g) &= \left\{ \alpha_G \in Hom(\widehat{\mathbf{F}_{kk}}, G) \mid \alpha_G(w) = g \right\} \\ E_w(G,g) &= \left\{ \alpha_G \in Epi(\widehat{\mathbf{F}_{kk}}, G) \mid \alpha_G(w) = g \right\} \end{aligned}$$

Now assume $w \in \widehat{\mathbf{F}_{kk}}$ is measure preserving, and let $x \in \widehat{\mathbf{F}_{kk}}$ be any primitive element. For every finite group G we have $|H_w(G,g)| = |G|^{k-1} = |H_x(G,g)|$. The same equality holds for the set of epimorphisms, namely $|E_w(G,g)| = |E_x(G,g)|$. We will show this by induction on $|G|$.

If $|G| = 1$ the claim is trivial. The inductive step goes as follows: if $g \in G$, then

$$\begin{aligned} |E_w(G,g)| &= |H_w(G,g)| - \sum_{g \in H \lneq G} |E_w(H,g)| = \\ &= |H_x(G,g)| - \sum_{g \in H \lneq G} |E_x(H,g)| = |E_x(G,g)| \end{aligned}$$

Now choose a basis $x_1, \ldots, x_k$ of $\widehat{\mathbf{F}_{kk}}$. For every $N \lhd_O \widehat{\mathbf{F}_{kk}}$, $|E_{x_1}(\widehat{\mathbf{F}_{kk}}/N, wN)| = |E_w(\widehat{\mathbf{F}_{kk}}/N, wN)| \geq 1$. If $\alpha \in E_{x_1}(\widehat{\mathbf{F}_{kk}}/N, wN)$ then $wN = \alpha(x_1), \alpha(x_2), \ldots, \alpha(x_k)$ generate $\widehat{\mathbf{F}_{kk}}/N$. A standard compactness argument shows that there are elements $w_2, \ldots, w_k \in \widehat{\mathbf{F}_{kk}}$ such that $\{wN, w_2N, \ldots, w_kN\}$ generate $\widehat{\mathbf{F}_{kk}}/N$ for every $N \lhd_O \widehat{\mathbf{F}_{kk}}$. But then $\{w, w_2, \ldots, w_k\}$ generate $\widehat{\mathbf{F}_{kk}}$ as well. Whenever k elements generate $\widehat{\mathbf{F}_{kk}}$, they generate it freely. Thus $\{w, w_2, \ldots, w_k\}$ is a basis and w is primitive. $\square$

1.8 The Average Number of Fixed Points in $\alpha_n(w)$

As before, let $\alpha_n \in Hom(\mathbf{F}_k, S_n)$ be a uniformly distributed random homomorphism. In (1.1.1) we defined the function $\Phi_{\langle w \rangle}(n) = \Phi_w(n)$ for every $w \in \mathbf{F}_k$. It considers the probability that $\alpha_n(w)$ fixes the element 1 and quantifies its deviation from $\frac{1}{n}$. The choice of the element 1 is arbitrary, of course, and we get the same probability for every element in $1, \ldots, n$. Thus $n\Phi_w(n) + 1$ is the average number of fixed points of the random permutation $\alpha_n(w)$.

Corollary 1.4.2 states that in $\mathbf{F}_2$ there are exactly four possible primitivity ranks of words. This translates through Proposition 1.1.9 to four possibilities for the average number of fixed points in the permutation $\alpha_n(w)$, as summarized by Table 1.1:

$\pi(w)/\phi(w)$	Description	$Prob[\alpha_n(w)(1) = 1]$	Avg # of f.p. of $\alpha_n(w)$
0	$w = 1$	1	n
1	w is a power	$\frac{1}{n} + \frac{a_1(w)}{n} + \sum_{i=2}^{\infty} \frac{a_i(w)}{n^i}$	$1 + a_1(w) + O\left(\frac{1}{n}\right)$
2		$\frac{1}{n} + \frac{a_2(w)}{n^2} + \sum_{i=3}^{\infty} \frac{a_i(w)}{n^i}$	$1 + \frac{a_2(w)}{n} + O\left(\frac{1}{n^2}\right)$
∞	w is primitive	$\frac{1}{n}$	1

Table 1.1: The possibilities for the average number of fixed points of the permutation $\alpha_n(w)$ for some $w \in \mathbf{F}_2$.

Recall that all coefficients $a_i(w)$ are integers (Claim 1.5.1). Moreover, in these cases $a_{\phi(w)}(w)$ counts the $\langle w \rangle$-critical subgroups of $\mathbf{F}_2$, so in particular $a_{\phi(w)}(w) > 0$. We thus obtain

Corollary 1.8.1. *For every word $w \in \mathbf{F}_2$ and every large enough n, the average number of fixed points of $\alpha_n(w)$ is at least 1.*

This leads to the following conjecture, which is a consequence of Conjecture 1.1.10:

Conjecture 1.8.2. *For every word $w \in \mathbf{F}_k$ and every large enough n, the average number of fixed points of $\alpha_n(w)$ is at least 1.*

Proposition 1.1.9 says something about free words in general. If $\phi(w) \leq 2$ for some $w \in \mathbf{F}_k$, then the first non-vanishing coefficient $a_{\phi(w)}(w)$ is positive. Thus,

Corollary 1.8.3. *For every word $w \in \mathbf{F}_k$ the average number of fixed points in $\alpha_n(w)$ is at least $1 - O\left(\frac{1}{n^2}\right)$.*

It is suggestive to ask whether Conjecture 1.8.2 holds for *all* n. Namely, is it true that for every $w \in \mathbf{F}_k$ and every n, the average number of fixed points in $\alpha_n(w)$ is at least 1? By results of Abért ([Abe06]), this statement turns out to be incorrect.

A Note Added in Proof

Remark 1.8.4. After this paper was completed, we learned about the algorithm of Silva and Weil to detect free-factor subgroups in the free group [SW08]. In essence, their algorithm relies on the same phenomenon that we independently noticed here. However, our reasoning is very different, and offers several substantial advantages over the presentation in [SW08]. A more elaborate discussion of the differences between the two approaches appears in Appendix 1.A.

Remark 1.8.5. In subsequent joint work with O. Parzanchevski [PP15], we manage to prove Conjecture 1.1.4 in full. That proof relies on Theorem 1.1.1 and follows the general strategy laid out in the current paper. In particular, we establish Conjectures 1.1.10, 1.7.1, 1.7.2 and 1.8.2.

Acknowledgements

It is a pleasure to thank Nati Linial for his support, encouragement and useful comments. We are also grateful to Aner Shalev for supporting this research and for his valuable suggestions. We would also like to thank Tsachik Gelander, Michael Larsen, Alex Lubotzky, Chen Meiri, Ori Parzanchevski, Iddo Samet and Enric Ventura for their beneficial comments. We would also like to express our gratefulness to the anonymous referee for his many valuable comments.

Appendices

1.A An Algorithm to Detect Free Factors

One of the interesting usages of Theorem 1.1.1 is an algorithm to detect free factor subgroups and consequently, also to detect primitive words in $\mathbf{F}_k$. The algorithm receives as input H and J, two finitely generated subgroups of $\mathbf{F}_k$, and determines whether $H \overset{*}{\leq} J$. The subgroups H and J are given to us by specifying a generating set, where members of the generating sets are words in the elements of the fixed basis X. (Note that the algorithm in particular decides as well whether $H \leq J$, but this is neither hard nor new).

We should mention that ours is not the first algorithm, nor the first graph-theoretic one, for this problem (see Chapter I.2 in [LS70]). We already mentioned (Remark 1.8.4) [SW08], who noticed the basic phenomenon underlying our algorithm, albeit in a very different language. See Remark 1.A.2 below for an explanation of the differences. A well-known algorithm due to Whitehead solves a much more general problem. Namely, for given $2r$ words $w_1, \ldots, w_r, u_1, \ldots, u_r \in \mathbf{F}_k$, it determines whether there is an automorphism $\alpha \in Aut(\mathbf{F}_k)$ such that $\alpha(w_i) = u_i$ for each i ([Whi36a],[Whi36b]. For a good survey see Chapter I.4 at [LS70]. A nice presentation of the restriction of Whitehead's algorithm to our problem appears in [Sta99]). Quite recently, Roig, Ventura and Weil introduced a more clever version of the Whitehead algorithm for the case of detecting primitive words and free factor subgroups [RVW07]. Their version of the algorithm has polynomial time in both the length of the given word w (or the total length of generators of a given subgroup H) and in k, the rank of the ambient group $\mathbf{F}_k$. To the best of our knowledge, their algorithm is currently the best one for this problem, complexity-wise. The algorithm we present is, at least naively, exponential, as we show below (Remark 1.A.1).

So assume we are given two subgroups of finite rank of $\mathbf{F}_k$, H and J, by means of finite generating sets S_H, S_J. Each element of S_H, S_J is assumed to be a word in the letters $X \cup X^{-1}$ (recall that $X = \{x_1, \ldots, x_k\}$ is the given basis of $\mathbf{F}_k$). To find out whether $H \overset{*}{\leq} J$, follow the following steps.

Step 1: Construct Core Graphs and Morphism
First, construct the core graphs $\Gamma = \Gamma_X(H)$ and $\Delta = \Gamma_X(J)$ by the process described in Appendix 1.C. Then, seek a morphism $\eta : \Gamma \to \Delta$. This is a simple process that can be done inductively as follows: η must map the basepoint of Γ to the basepoint of Δ. Now, as long as η is not fully defined, there is some j-edge $e = (u, v)$ in $E(\Gamma)$ for which the image is not known yet, but the image of one of the end points, say $\eta(u)$, is known (recall that Γ is connected). There is at most one possible value that $\eta(e)$ can take, since the star of $\eta(u)$ contains at most one outgoing j-edge. If there is no such edge, we get stuck. Likewise, $\eta(v)$ must equal the terminus of $\eta(e)$, and if $\eta(v)$ was already determined in an inconsistent way, we get a contradiction. If in this process we never get stuck and never reach a contradiction, then η is defined. If this process cannot be carried out, then there is no morphism from Γ to Δ, and hence H is not a subgroup of J (see Claim 1.2.2).

Step 2: Reduce to Two Groups with $H \leq_{\vec{x}} J'$

After constructing the morphism $\eta : \Gamma \to \Delta$, we obtain a new graph from Δ by omitting all edges and all vertices not in the image of η. Namely,

$$\Delta' := \eta(\Gamma)$$

It is easy to see that Δ' is a core-graph, and we denote by J' the subgroup corresponding to Δ'. Obviously, Δ' is a quotient of Γ, so $H \leq_{\vec{x}} J'$. Moreover, it follows from Claim 1.2.5 that

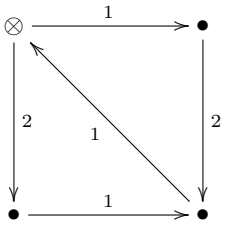
$$H \overset{*}{\leq} J \iff H \overset{*}{\leq} J'.$$

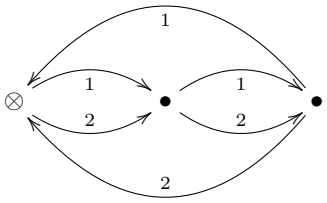
Step 3: Use $\rho_X(H, J')$ to determine whether $H \overset{*}{\leq} J'$
Now calculate $\rho_X(H, J')$ (this is clearly doable because the subgraph of $\mathcal{D}_k$ consisting of quotients of Γ is finite). Thanks to Theorem 1.1.1, $\rho_X(H, J')$ determines whether or not $H \overset{*}{\leq} J'$, and consequently, whether or not $H \overset{*}{\leq} J$.

Remark 1.A.1. The complexity of this algorithm is roughly $O(v^{2t})$, where v is the number of vertices in $\Gamma_X(H)$ and t is the difference in ranks: $t = rk(J) - rk(H)$. Naively, we need to go over roughly all possible sets of t pairs of vertices of $\Gamma_X(H)$ and try to merge them (see Remark 1.3.7). The number of possibilities is at most $\binom{\binom{v}{2}}{t}$, which shows the claimed bound. (In fact, we can restrict to pairs where both vertices are in the same fiber of the morphism $\eta : \Gamma_X(H) \to \Gamma_X(J)$.)

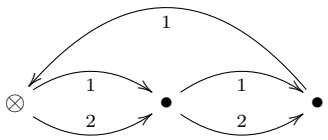
1.A.1 Examples

We illustrate the different phases of the algorithm by two concrete examples. Consider first the groups $H = \langle x_1 x_2 {x_1}^{-1} {x_2}^{-1}, x_2 {x_1}^2 \rangle$ and $J = \langle {x_1}^3, {x_2}^3, x_1 x_2^{-1}, x_1 x_2 x_1 \rangle$, both in $\mathbf{F}_2$. The core graphs of these groups are:





In this case, a morphism η from $\Gamma = \Gamma_X(H)$ to $\Delta = \Gamma_X(J)$ can be constructed. All the vertices of Γ are in the image of η, and only one edge, the long 2-edge at the bottom, is not in $\eta(E(\Gamma))$. Thus Δ' is:

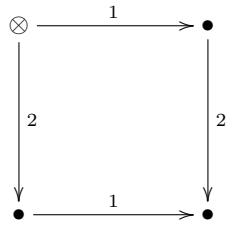


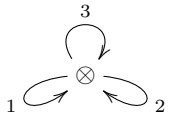
and J' is the corresponding subgroup $J' = \langle {x_1}^3, x_1 x_2^{-1}, x_1 x_2 x_1 \rangle$.

Finally, $rk(H) = 1 - \chi(\Gamma) = 2$ and $rk(J') = 1 - \chi(\Delta') = 3$, and so the difference is $rk(J') - rk(H) = 1$. It can be easily verified that Δ' is indeed an immediate quotient of Γ: simply merge the upper-right vertex of Γ with the bottom-left one to obtain Δ'. Thus

$\rho_X(H, J') = 1 = rk(J') - rk(H)$, and so $H \overset{*}{\leq} J'$ hence $H \overset{*}{\leq} J$.

As a second example, consider the commutator word $w = x_1 x_2 x_1^{-1} x_2^{-1}$. We want to determine whether it is primitive in $\mathbf{F}_3$. We take $H = \langle w \rangle$ and the core graphs are then





Once again, a morphism η from $\Gamma = \Gamma_X(H)$ to $\Delta = \Gamma_X(\mathbf{F}_3)$ can be constructed, and there is a single edge in Δ, the 3-edge, outside the image of η. Thus Δ' is the quotient of Γ which is the bottom graph in Figure 1.3.1, and J' is simply $\mathbf{F}_2$.

Finally, $rk(H) = 1 - \chi(\Gamma) = 1$ and $rk(\mathbf{F}_2) = 2$, and so the difference is $rk(F_2) - rk(H) = 1$. But as we infer from Figure 1.3.1, $\rho_X(H, \mathbf{F}_2) = 2$. Thus $\rho_X(H, \mathbf{F}_2) > rk(\mathbf{F}_2) - rk(H)$ and H is *not* a free factor of $\mathbf{F}_2$. As consequence, w is *not* primitive in $\mathbf{F}_3$. (This example generalizes as follows: if w is a free word containing exactly l different letters, then w is primitive iff we can obtain a wedge-of-loops graph from $\Gamma_X(\langle w \rangle)$ by merging $l - 1$ pairs of vertices.)

Remark 1.A.2. At this point we would like to elaborate on the differences between the algorithm presented here and the one introduced in [SW08]. Silva and Weil's presentation considers automata and their languages. We consider the X-fringe ${}_X(H)$ and introduce the DAG $\mathcal{D}_k$ and the distance function from Definition 1.3.2. Steps 1 and 2 of our algorithm, which reduce the problem in its very beginning to the case where $H \leq_{\vec{x}} J$, have no parallel in [SW08]. Rather, they show that if $H \overset{*}{\leq} J$, then by some sequence of "i-steps" (their parallel of our immediate quotients) on H, of length at most $rk(J) - rk(H)$, one can obtain a core graph which is embedded in $\Gamma_X(J)$ (we make the observation that this embedded core graph can be computed in advance). Besides shedding more light on this underlying phenomenon, our more graph-theoretic approach has another substantial advantage: by considering $\mathcal{D}_k$, turning the fringe ${}_X(H)$ into a directed graph and stating the algorithm in the language of Theorem 1.1.1, we obtain a straight-forward algorithm to identify H-critical subgroups and to compute $\pi(H)$. Moreover, we obtain a straight-forward algorithm to identify all "algebraic extensions" of H (Corollary 1.4.4). In particular, our algorithm to identify algebraic extensions substantially improves the one suggested in [KM02], Theorem 11.3 (and see also remark 11.4 about its efficiency).

1.B The Proof of Lemma 1.3.3

To complete the picture, we prove the upper bound for $\rho_X(H, J)$ stated in Lemma 1.3.3. We need to show that if $H, J \leq_{fg} \mathbf{F}_k$ such that $H \leq_{\vec{x}} J$, then

$$\rho_X(H, J) \quad \leq \quad \mathrm{rk}(J)$$

Proof. We show that $\Delta = \Gamma_X(J)$ can be obtained from $\Gamma = \Gamma_X(H)$ by merging at most $\mathrm{rk}(J)$ pairs of vertices. To see this, denote by m the number of edges in Γ, and choose some order on these edges, $e_1, \ldots, e_m$ so that for every i, there is a path from the basepoint of Γ to e_i traversing only edges among $e_1, \ldots, e_{i-1}$. (So e_1 must be incident with the basepoint, e_2 must be incident either with the basepoint or with the other end of e_1, etc.)

We now expose Δ step by step, each time adding the images of the next edge of Γ and of its end points. Formally, denote by η the (surjective) morphism from Γ to Δ, let Γ_i be the subgraph

of Γ that is the union of the basepoint of Γ together with $e_1, \ldots, e_i$ and their endpoints, and let $\Delta_i = \eta(\Gamma_i)$. We thus have two series of subgraphs

$$\Gamma_0 \subseteq \Gamma_1 \subseteq \ldots \subseteq \Gamma_m = \Gamma$$

and

$$\Delta_0 \subseteq \Delta_1 \subseteq \ldots \subseteq \Delta_m = \Delta$$

with $\Delta_0 = \Gamma_0$ being graphs with a single vertex and no edges.

Assume that $e_i = (u, v)$, and w.l.o.g. that $u \in V(\Gamma_{i-1})$. We then distinguish between three options. A **forced** step is when $\eta(e_i)$ already belongs to Δ_{i-1} and then $\Delta_i = \Delta_{i-1}$. A **free** step takes place when neither $\eta(e_i)$ nor $\eta(v)$ belong to Δ_{i-1}, in which case $\pi_1(\Delta_i) = \pi_1(\Delta_{i-1})$. The third option is that of a **coincidence**. This means that $\eta(e_i)$ does not belong to Δ_{i-1} but $\eta(v)$ does. In this case, Δ_i is obtained from Δ_{i-1} by connecting two vertices by a new edge, and $\pi_1(\Delta_i)$ has rank larger by 1 from the rank of $\pi_1(\Delta_{i-1})$. Since the fundamental group of Δ_0 has rank 0, this shows there are exactly $rk(J)$ coincidences along this process.

Assume the coincidences occurred in steps $j_1, \ldots, j_{\mathrm{rk}(J)}$. If $e_{j_i} = (u, v)$, we let $\tilde{v} \in \eta^{-1}(\eta(v)) \cap V(\Gamma_{i-1})$, and take $\{v, \tilde{v}\}$ to be a pair of vertices of Γ that we merge. (It is possible that $v = w$.) Let $w_i \in \mathbf{F}_k$ be be a word corresponding to this merge in Γ. It is easy to see by induction that Δ_{j_i} corresponds to the subgroup $\langle H, w_1, \ldots, w_i \rangle$. In particular, Δ corresponds to $\left\langle H, w_1, \ldots, w_{\mathrm{rk(J)}} \right\rangle$. We are done because all these words correspond to pairs of vertices in Γ (and see Remark 1.3.7). $\square$

1.C The Folding Algorithm to Construct Core Graphs

Finally, we present a well known algorithm to construct the core graph of a given subgroup $H \leq_{fg} \mathbf{F}_k$. The input to this process is any finite set of words $\{h_1, \ldots, h_r\}$ in the letters $\{x_1, \ldots, x_k\}$ that generate H.

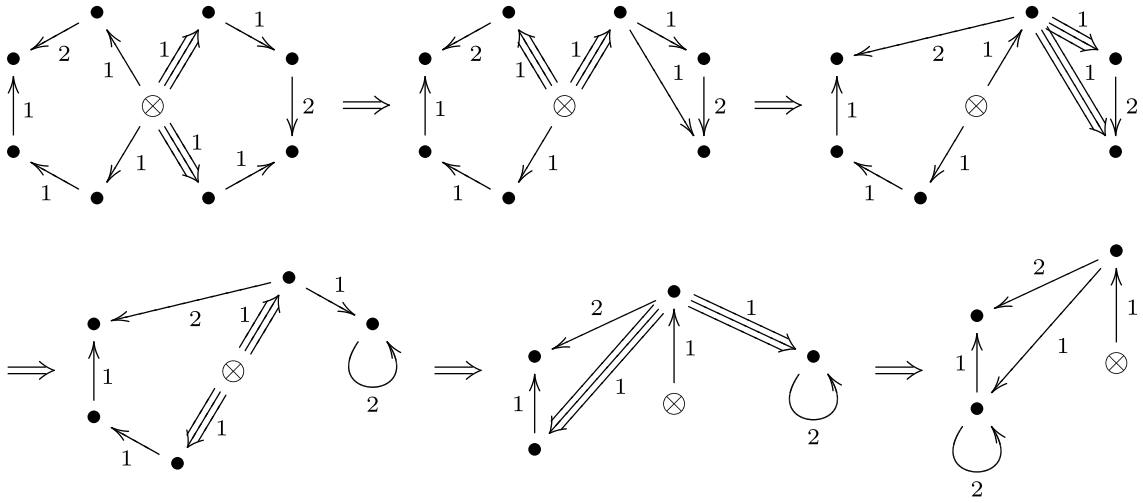


Figure 1.C.1: Generating the core graph $\Gamma_X(H)$ of $H = \langle x_1 x_2 x_1^{-3}, x_1^2 x_2 x_1^{-2} \rangle \leq \mathbf{F}_2$ from the given generating set. We start with the upper left graph which contains a distinct loop at the basepoint for each (reduced) element of the generating set. Then, gradually and at arbitrary order, we merge pairs of equally-labeled edges which share the same origin or the same terminus. One of the possible orders of merging pairs is shown in this figure, and at each phase we mark by triple arrows the pair of edges being merged. The graph in the bottom right is $\Gamma_X(H)$, as it has no equally-labeled edges sharing the same origin or the same terminus.

Every element h_i of the generating set corresponds to some path with directed edges labeled by the x_i's (we assume the element is given in reduced form). Merge these r paths to a single graph by identifying all their $2r$ end-points to a single vertex which is denoted as basepoint. Then, as long as there are two j-labeled edges with the same terminus (resp. origin) for some j, merge the two edges and their origins (resp. termini). Such a step is often referred to as a *Stallings folding*. It is a fairly easy observation that the resulting graph is indeed $\Gamma_X(H)$ and that the order of folding has no significance. To illustrate, we draw in Figure 1.C.1 the folding process by which we obtain the core graph $\Gamma_X(H)$ of $H = \langle x_1 x_2 x_1^{-3}, x_1^2 x_2 x_1^{-2} \rangle \leq \mathbf{F}_2$ from the given generating set.

References

[Abe06] M. Abert. On the probability of satisfying a word in a group. *Journal of Group Theory*, 9:685–694, 2006.

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

[Ger84] S.M. Gersten. On whitehead's algorithm. *Bull. Amer. Math. Soc., New Ser*, 10(2):281–284, 1984.

[GS09] S. Garion and A. Shalev. Commutator maps, measure preservation, and T-systems. *Trans. Amer. Math. Soc.*, 361(9):4631–4651, 2009.

[KM02] I. Kapovich and A. Myasnikov. Stallings foldings and subgroups of free groups. *Journal of Algebra*, 248(2):608–668, 2002.

[LP10] N. Linial and D. Puder. Words maps and spectra of random graph lifts. *Random Structures and Algorithms*, 37(1):100–135, 2010.

[LS70] R.C. Lyndon and P.E. Schupp. *Combinatorial Group Theory*. Springer-Verlag, Berlin; New York, 1970.

[LS08] M. Larsen and A. Shalev. Characters of symmetric groups: sharp bounds and applications. *Inventiones mathematicae*, 174(3):645–687, 2008.

[LS09] M. Larsen and A. Shalev. Words maps and Waring type problems. *J. Amer. Math. Soc.*, 22(2):437–466, 2009.

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

[Nic94] A. Nica. On the number of cycles of given length of a free word in several random permutations. *Random Structures and Algorithms*, 5(5):703–730, 1994.

[PP15] D. Puder and O. Parzanchevski. Measure preserving words are primitive. *Journal of the American Mathematical Society*, 28(1):63–97, 2015.

[RVW07] A. Roig, E. Ventura, and P. Weil. On the complexity of the Whitehead minimization problem. *International journal of Algebra and Computation*, 17(8):1611–1634, 2007.

[Seg09] D. Segal. *Words: notes on verbal width in Groups*. London Mathematical Society, Lecture note Series 361, Cambridge University Press, Cambridge, 2009.

[Sha09] A. Shalev. Words maps, conjugacy classes, and a non-commutative Waring-type theorem. *Annals of Math.*, 170:1383–1416, 2009.

[Sta83] J.R. Stallings. Topology of finite graphs. *Inventiones mathematicae*, 71(3):551–565, 1983.

[Sta99] J.R. Stallings. Whitehead graphs on handlebodies. In J. Cossey, C. F. Miller, W.D. Neumann, and M. Shapiro, editors, *Geometric group theory down under*, pages 317–330. de Gruyter, Berlin, 1999.

[SW08] P. Silva and P. Weil. On an algorithm to decide whether a free group is a free factor of another. *RAIRO - Theoretical Informatics and Applications*, 42(2):395–414, 2008.

[Whi36a] J.H.C. Whitehead. On certain sets of elements in a free group. *Proc. London Math. Soc.*, 41:48–56, 1936.

[Whi36b] J.H.C. Whitehead. On equivalent sets of elements in a free group. *Ann. of Math.*, 37:768–800, 1936.

[Wil98] J.S. Wilson. *Profinite Groups*. Clarendon Press, Oxford, 1998.

Chapter 2

Measure Preserving Words are Primitive

Doron Puder[†] Ori Parzanchevski[‡]
Einstein Institute of Mathematics
Hebrew University, Jerusalem
doronpuder@gmail.com parzan@math.huji.ac.il

Published: Journal of the American Mathematical Society, 28 (1), 2015, pp 63-97.
DOI: 10.1090/S0894-0347-2014-00796-7

Abstract

We establish new characterizations of primitive elements and free factors in free groups, which are based on the distributions they induce on finite groups. For every finite group G, a word w in the free group on k generators induces a *word map* from G^k to G. We say that w is measure preserving with respect to G if given uniform distribution on G^k, the image of this word map distributes uniformly on G. It is easy to see that primitive words (words which belong to some basis of the free group) are measure preserving w.r.t. all finite groups, and several authors have conjectured that the two properties are, in fact, equivalent. Here we prove this conjecture. The main ingredients of the proof include random coverings of Stallings graphs, algebraic extensions of free groups, and Möbius inversions. Our methods yield the stronger result that a subgroup of $\mathbf{F}_k$ is measure preserving if and only if it is a free factor.

As an interesting corollary of this result we resolve a question on the profinite topology of free groups and show that the primitive elements of $\mathbf{F}_k$ form a closed set in this topology.

Contents

[†]Supported by an Advanced ERC Grant and by Adams Fellowship Program of the Israel Academy of Sciences and Humanities.
[‡]Supported by an Advanced ERC Grant.

2.1 Introduction

This paper establishes a new characterization of primitive elements in free groups, which is based on the distributions they induce on finite groups. Let $\mathbf{F}_k$ be the free group on k generators $X = \{x_1, \ldots, x_k\}$, and let $w = \prod_{j=1}^{r} x_{i_j}^{\varepsilon_j}$ ($\varepsilon_j = \pm 1$) be a word in $\mathbf{F}_k$. For every group G, w induces a *word map* from the Cartesian product G^k to G, by substitutions:

$$w : (g_1, \ldots, g_k) \mapsto \prod_{j=1}^{r} g_{i_j}^{\varepsilon j}.$$

The word w is called *measure preserving* with respect to a finite group G if all the fibers of this map are of equal size. Namely, every element in G is obtained by substitutions in w the same number of times. We say that w is *measure preserving* if it is measure preserving w.r.t. every finite group. The last years have seen a great interest in word maps in groups, and the distributions they induce. We refer the reader, for instance, to [Sha09, LS09, AV11, PS13], and to the recent book [Seg09] and survey [Sha13]. Several authors have also studied words which are asymptotically measure preserving on finite simple groups, see e.g. [LS08, GS09, BK13].

The word w is called *primitive* if it belongs to some basis (free generating set) of $\mathbf{F}_k$. It is a simple observation (see 2.1.2 below) that primitive words are measure preserving, and several authors have conjectured that the converse is also true. Namely, that measure preservation implies primitivity[†]. From private conversations we know that this has occurred to the following mathematicians and discussed among themselves: N. Avni, T. Gelander, M. Larsen, A. Lubotzky and A. Shalev. The question was independently raised in [LP10] and also in [AV11], alongside a generalization of it (see Section 2.8).

In [Pud14] the first author proved the conjecture for $\mathbf{F}_2$. Here we prove it in full:

Theorem 2.1.1. *A measure preserving word is primitive.*

A key ingredient of the proof is the extension of the problem from single words to (finitely generated) subgroups of $\mathbf{F}_k$. The concept of primitive words extends naturally to the notion of free factors: Let H be a subgroup of the free group J (in particular, H is free as well). We say that H is a *free factor* of J, and denote this by $H \overset{*}{\leq} J$, if there is a subgroup $H' \leq J$ such that $H * H' = J$. Equivalently, $H \overset{*}{\leq} J$ if and only if some basis of H can be extended to a basis of J. (This in turn is easily seen to be equivalent to the condition that *every* basis of H extends to a basis of J.)

In order to generalize the notion of measure preservation to subgroups, we need to change a little our perspective of word maps. One can think of the word map w as the evaluation map from $\mathrm{Hom}(\mathbf{F}_k, G)$ to G, i.e., $w(\alpha) = \alpha(w)$ for $\alpha \in \mathrm{Hom}(\mathbf{F}_k, G)$. The identification of $\mathrm{Hom}(\mathbf{F}_k, G)$ with G^k depends on the chosen basis, and is due to the fact that a homomorphism from a free group is uniquely determined by choosing the images of the elements of a basis, and these images can be chosen arbitrarily.

In this perspective, w is measure preserving w.r.t. G if the element $\alpha_G(w)$ is uniformly distributed over G, where $\alpha_G \in \mathrm{Hom}(\mathbf{F}_k, G)$ is a homomorphism chosen uniformly at random. If w is primitive then it belongs to some basis, and identifying $\mathrm{Hom}(\mathbf{F}_k, G)$ and G^k according to this basis gives

[†]It is interesting to note that there is an easy abelian parallel to this conjecture. A word $w \in \mathbf{F}_k$ belongs to a basis of $\mathbb{Z}^k \cong \mathbf{F}_k/\mathbf{F}_k'$ if and only if for any group G the associated word map is surjective. See [Seg09], Lemma 3.1.1.

Observation 2.1.2. *A primitive word is measure preserving.*

We can now extend the notion of measure preservation from words to finitely generated subgroups (we write $H \leq_{fg} \mathbf{F}_k$ when H is a finitely generated subgroups of $\mathbf{F}_k$):

Definition 2.1.3. Let $H \leq_{fg} \mathbf{F}_k$. We say that H is *measure preserving* if for every finite group G and $\alpha_G \in \mathrm{Hom}(\mathbf{F}_k, G)$ a random homomorphism chosen with uniform distribution, $\alpha_G\big|_H$ is uniformly distributed in $\mathrm{Hom}(H, G)$.

This can be reformulated in terms of distributions of subgroups: Observe the distribution of the random subgroup $\alpha_G(H) \leq G$, where $\alpha_G \in \mathrm{Hom}(\mathbf{F}_k, G)$ distributes uniformly. Then H is measure preserving if the distribution of $\alpha_G(H)$ is the same as that of the image of a uniformly chosen homomorphism from $\mathbf{F}_{\mathrm{rk}(H)}$ to G (where $\mathrm{rk}(H)$ denotes the rank of H).

As for single words, it is immediate that a free factor is measure preserving, and again it is natural to conjecture that the converse also holds. Since $1 \neq w \in \mathbf{F}_k$ is measure preserving iff $\langle w \rangle$ is measure preserving, this is an extension of the conjecture regarding words. In [Pud14] the first author proved the extended conjecture for subgroups of $\mathbf{F}_k$ of rank $\geq k-1$ (thus proving the conjecture for $\mathbf{F}_2$), but the techniques used in that paper are specialized for the proven cases. In this paper we introduce completely new techniques, which yield the extended conjecture in full:

Theorem 2.1.4. *A measure preserving subgroup is a free factor.*

In Section 2.8 we explain how this circle of ideas is related to the study of profinite groups and decidability questions. In fact, part of the original motivation for this study comes from this relation. In particular we have the following corollary (see also Corollary 2.8.1):

Corollary 2.1.5. *The set P of primitive elements in $\mathbf{F}_k$ is closed in the profinite topology.*

In plain terms, this amounts to the assertion that every non-primitive word in $\mathbf{F}_k$ is contained in a primitive-free coset of a finite index subgroup.

In order to prove Theorem 2.1.4, one needs to exhibit, for each non-primitive word $w \in \mathbf{F}_k$, some "witness" finite group with respect to which w is not measure preserving. Our witnesses are always the symmetric groups S_n. In fact, it is enough to restrict one's attention to the average number of fixed points in the random permutation $\alpha_{S_n}(w)$ (which we also denote by $\alpha_n(w)$). We summarize this in the following stronger version of Theorems 2.1.1 and 2.1.4:

Theorem (2.1.4')**.** *Let $w \in \mathbf{F}_k$, and for every finite group G, let $\alpha_G \in \mathrm{Hom}(\mathbf{F}_k, G)$ denote a random homomorphism chosen with uniform distribution. Then the following are equivalent:*

(1) w is primitive.

(2) w is measure preserving: for every finite group G the random element $\alpha_G(w)$ has uniform distribution.

(3) For every $n \in \mathbb{N}$ the random permutation $\alpha_n(w) = \alpha_{S_n}(w)$ has uniform distribution.

(4) For every $n \in \mathbb{N}$, the expected number of fixed points in the random permutation $\alpha_n(w) = \alpha_{S_n}(w)$ is 1:

$$\mathbb{E}\left[\#\mathrm{fix}\left(\alpha_n(w)\right)\right] = 1$$

(5) For infinitely many $n \in \mathbb{N}$,

$$\mathbb{E}\left[\#\mathrm{fix}\left(\alpha_n(w)\right)\right] \leq 1$$

The analogue properties for f.g. subgroups are equivalent as well. For example, the parallel of property (4) for $H \leq_{fg} \mathbf{F}_k$ is that for every n, the image $\alpha_n(H) \subseteq S_n$ stabilizes on average exactly $n^{1-\mathrm{rk}(H)}$ elements of $\{1, \ldots, n\}$.

We already explained above the implication (1) ⇒ (2), and (2) ⇒ (3) ⇒ (4) ⇒ (5) is evident (recall that a uniformly distributed random permutation has exactly one fixed point on average). The only nontrivial, somewhat surprising part, is the implication (5) ⇒ (1) which is proven in this paper. It turns out that an effective bound can also be obtained:

Proposition 2.1.6. *A word w of length $\ell > 0$ is primitive iff $\mathbb{E}\left[\#\mathrm{fix}\left(\alpha_n(w)\right)\right] = 1$ for $n \leq \ell$.*

An analogue result holds for subgroups (see Corollary 2.6.6).

A key role in our proof is played by the notion of *primitivity rank*, an invariant classifying words and f.g. subgroups of $\mathbf{F}_k$, which was first introduced in [Pud14]: A primitive word $w \in \mathbf{F}_k$ is also primitive in every subgroup containing it (Claim 2.3.9(3)). However, if w is not primitive in $\mathbf{F}_k$, it may be either primitive or non-primitive in subgroups of $\mathbf{F}_k$ containing it. But what is the smallest rank of a subgroup giving evidence to the imprimitivity of w? Informally, how far does one have to search in order to establish that w is not primitive? Concretely:

Definition 2.1.7. The *primitivity rank* of $w \in \mathbf{F}_k$, denoted $\pi(w)$, is

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

If no such J exists, i.e. if w is primitive, then $\pi(w) = \infty$.

More generally, for $H \leq_{fg} \mathbf{F}_k$, the *primitivity rank* of H is

$$\pi(H) = \min\left\{ \mathrm{rk}(J) \;\middle|\; \begin{array}{c} H \leq J \leq \mathbf{F}_k \text{ s.t.} \\ H \text{ is } \mathbf{not} \text{ a free factor of } J \end{array} \right\}.$$

Again, if no such J exists, then $\pi(H) = \infty$. We call a subgroup J for which the minimum is obtained **H-critical**, and denote the set of H-critical subgroups by $\mathrm{Crit}(H)$. The set of w-critical subgroups of a word w is defined analogously.

Note that for $w \neq 1$, $\pi(w) = \pi(\langle w \rangle)$. Let us give a few examples: $\pi(w) = 0$ iff $w = 1$; $\pi(w) = \infty$ iff w is primitive, and $\pi(H) = \infty$ iff H is a free factor; $\pi(w) = 1$ if and only if w is a proper power, namely $w = v^d$ for some $v \in \mathbf{F}_k$ and $d \geq 2$, and then $\mathrm{Crit}(w) = \{\langle v^m \rangle : m \mid d, 1 \leq m < d\}$ (assuming that v itself is not a power). By [Pud14, Lemma 6.8], $\pi\left(x_1^2 \ldots x_r^2\right) = r$ for every $1 \leq r \leq k$. We thus have that π takes all values in $\{0, 1, 2, \ldots, k\} \cup \{\infty\}$, and Claim 2.3.9(3) shows that these are all the values it obtains. The primitivity rank of a word or a subgroup is computable - this is shown in Section 2.4. The distribution of the primitivity rank is discussed in [Pud15a].

In this paper we sometimes find it more convenient to deal with *reduced ranks* of subgroups: $\widetilde{\mathrm{rk}}(H) \stackrel{def}{=} \mathrm{rk}(H) - 1$. We therefore define analogously the *reduced primitivity rank*, $\widetilde{\pi}(\cdot) \stackrel{def}{=} \pi(\cdot) - 1$.

As mentioned above, our main result follows from an analysis of the average number of common fixed points of $\alpha_n(H)$ (where α_n denotes a uniformly distributed random homomorphism in $\mathrm{Hom}(\mathbf{F}_k, S_n)$). In other words, we count the number of elements in $\{1, \ldots, n\}$ stabilized by the images under α_n of *all* elements of H. Theorem 2.1.4' follows from the main result of this analysis:

Theorem 2.1.8. *The average number of common fixed points of $\alpha_n(H)$ is*

$$\frac{1}{n^{\widetilde{\mathrm{rk}}(H)}} + \frac{|\mathrm{Crit}(H)|}{n^{\widetilde{\pi}(H)}} + O\left(\frac{1}{n^{\widetilde{\pi}(H)+1}}\right).$$

In particular, for a word w

$$\mathbb{E}\left[\#\mathrm{fix}\left(\alpha_n(w)\right)\right] = 1 + \frac{|\mathrm{Crit}(w)|}{n^{\widetilde{\pi}(w)}} + O\left(\frac{1}{n^{\widetilde{\pi}(w)+1}}\right).$$

$\pi(w)$	Description of w	$\mathbb{E}[\#\mathrm{fix}(\alpha_n(w))]$
0	$w=1$	n
1	w is a power	$1+\lvert\mathrm{Crit}(w)\rvert+O\left(\frac{1}{n}\right)$
2	E.g. $[x_1,x_2],x_1^2x_2^2$	$1+\frac{\lvert\mathrm{Crit}(w)\rvert}{n}+O\left(\frac{1}{n^2}\right)$
3		$1+\frac{\lvert\mathrm{Crit}(w)\rvert}{n^2}+O\left(\frac{1}{n^3}\right)$
$\vdots$		$\vdots$
k	E.g. $x_1^2\dots x_k^2$	$1+\frac{\lvert\mathrm{Crit}(w)\rvert}{n^{k-1}}+O\left(\frac{1}{n^k}\right)$
∞	w is primitive	1

Table 2.1: Primitivity Rank and Average Number of Fixed Points.

We remark that $\mathrm{Crit}(H)$ is always finite (see Section 2.4). Table 2.1 summarizes the connection implied by Theorem 2.1.8 between the primitivity rank of w and the average number of fixed points in the random permutation $\alpha_n(w)$.

Theorem 2.1.8 implies the following general corollary regarding the family of distributions of S_n induced by word maps:

Corollary 2.1.9. *For a non-primitive $w\in\mathbf{F}_k$ the average number of fixed points in $\alpha_n(w)$ is strictly greater than 1, for large enough n.*

Corollary 2.1.9 is in fact the missing piece (5) $\Rightarrow$ (1) in Theorem 2.1.4'. In addition, it follows from this corollary that for every $w\in\mathbf{F}_k$ and large enough n, the average number of fixed points in $\alpha_n(w)$ is at least one[†]. In other words, primitive words generically induce a distribution of S_n with the fewest fixed points on average.

The results stated above validate completely the conjectural picture described in [Pud14]. Theorem 2.1.8 and its consequences, Corollaries 2.8.1, 2.1.5 and 2.1.9, are stated there as conjectures (Conjectures 1.10, 7.1, 7.2 and 8.2).

The analysis of the average number of fixed points in $\alpha_n(w)$ has its roots in [Nic94]. Nica notices that by studying the various quotients of a labeled cycle-graph (corresponding to w), one can compute a rational expression which gives this average for every large enough n. When $w=v^d$ with d maximal (so v is not a power), he shows that the limit distribution of the number of fixed points in $\alpha_n(w)$ (as $n\to\infty$) is $\delta(d)+O\left(\frac{1}{n}\right)$, where $\delta(d)$ is the number of divisors of d ([Nic94], Corollary 1.3)[‡]. Nica's result follows from Theorem 2.1.8: if $w\neq 1$ is a proper power and $w=v^d$ with $d\geq 2$ maximal, then $\lvert\mathrm{Crit}(w)\rvert=\delta(d)-1$, and if it is not a power then $\widetilde{\pi}(w)\geq 1$.

The results of this paper have interesting implications in the study of expansion in random graphs: In [Pud15a], the first author presents a new approach to showing that random graphs are nearly optimal expanders. A crucial ingredient in the proof is Theorem 2.1.8. More particularly, it was conjectured by Alon [Alo86] that the spectral gap of a random d-regular graph is a.a.s. arbitrarily close to $d-2\sqrt{d-1}$, and this conjecture was generalized by Friedman [Fri03] to non-regular graphs. In [Fri08], Alon's conjecture is proved by highly sophisticated arguments, which are not applicable for the generalized conjecture (as far as is known). The results in [Pud15a] give a simple proof which nearly recovers Friedman's results regarding Alon's conjecture, and can be applied also for the generalized conjecture, giving the best results as of now regarding non-regular graphs.

[†]It is suggestive to ask whether this holds for *all* n. Namely, is it true that for every $w\in\mathbf{F}_k$ and every n, the average number of fixed points in $\alpha_n(w)$ is at least 1? By results of Abért ([Abe06]), this statement turns out to be false.

[‡]Nica's result is in fact more general: the same statement holds not only for fixed points but for cycles of length L for every fixed L.

2.2 Overview of the proof

The proof of our main theorem involves several structures of posets (partially ordered sets) on $\mathfrak{sub}_{fg}(\mathbf{F}_k)$, the set of finitely generated subgroups of $\mathbf{F}_k$. This set has, of course, a natural structure of a poset given by the relation of inclusion. However, there are other interesting partial orders defined on it: the relation of *algebraic extensions*, and the family of relations defined by *covers*. We introduce some notation: If $\preceq$ is some partial order on $\mathfrak{sub}_{fg}(\mathbf{F}_k)$, and $H, J \leq_{fg} \mathbf{F}_k$, we define the *closed interval*

$$[H, J]_{\preceq} = \{L \in \mathfrak{sub}_{fg}(\mathbf{F}_k) \mid H \preceq L \preceq J\}$$

and similarly the open interval $(H, J)_{\preceq} = \{L \mid H \precneqq L \precneqq J\}$, the half-bounded interval $[H, \infty)_{\preceq} = \{L \mid H \preceq L\}$, and so on (see also the glossary).

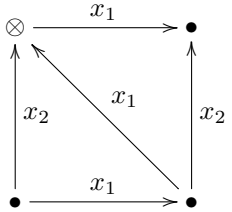
Algebraic Extensions This notion goes back to [Tak51], and was further studied in [KM02, MVW07].

Definition 2.2.1. We say that J is an *algebraic extension* of H, denoted $H \leq_{alg} J$, if $H \leq J$ and H is not contained in any proper free factor of J.

The terminology comes from similarities (that go only to some extent) between this notion and that of algebraic extensions of fields (in this line of thought, J is a *transcendental extension* of H when $H \overset{*}{\leq} J$). We devote Section 2.4 to study this relation. It is clearly reflexive and antisymmetric, but it is also transitive (Claim 2.4.1). In addition, it is very sparse: it turns out that $[H, \infty)_{alg}$, the set of algebraic extensions of H, is finite for every $H \leq_{fg} \mathbf{F}_k$, so in particular $(\mathfrak{sub}_{fg}(\mathbf{F}_k), \leq_{alg})$ is locally finite[†]. It is a simple observation that H-critical subgroups are in particular algebraic extensions of H, i.e. $\mathrm{Crit}(H) \subseteq [H, \infty)_{alg}$. In fact, they are the proper algebraic extensions of minimal rank.

X-cover For every basis $X = \{x_1, \ldots, x_k\}$ of $\mathbf{F}_k$ there is a partial order denoted $\leq_{\vec{x}}$, which is based on the notion of quotients, or surjective morphisms, of *core graphs*. Introduced in [Sta83], core graphs provide a geometric approach to the study of free groups (for an extensive survey see [KM02], and also [MVW07] and the references therein). Given the basis X, Stallings associates with every $H \leq \mathbf{F}_k$ a directed and pointed graph denoted $\Gamma_X(H)$, whose edges are labeled by the elements of X. A full definition appears in Section 2.3, but we illustrate the concept in Figure 2.2.1. It shows the core graph of the subgroup of $\mathbf{F}_2$ generated by $x_1 x_2^{-1} x_1$ and $x_1^{-2} x_2$, with $X = \{x_1, x_2\}$.

Figure 2.2.1: The core graph $\Gamma_X(H)$ where $X = \{x_1, x_2\}$ and $H = \langle x_1 x_2^{-1} x_1, x_1^{-2} x_2 \rangle \leq \mathbf{F}_2$.



The order $\leq_{\vec{x}}$ is defined as follows: for $H, J \leq \mathbf{F}_k$ one has $H \leq_{\vec{x}} J$ iff the associated core graph $\Gamma_X(J)$ is a quotient (as a pointed labeled graph) of the core graph $\Gamma_X(H)$ (see Definition 2.3.3). When $H \leq_{fg} \mathbf{F}_k$, $\Gamma_X(H)$ is finite (Claim 2.3.1(1)), and thus has only finitely many quotients. As it turns out that different groups correspond to different core graphs, this implies that $(\mathfrak{sub}_{fg}(\mathbf{F}_k), \leq_{\vec{x}})$ is locally finite too. We stress that we have here an infinite family of partial orders, one for every choice of basis for $\mathbf{F}_k$. Although the dependency on the basis makes these orders somewhat less universal, they turn out to be the most useful for our purposes.

[†]A *locally finite* poset is one in which every closed interval $[a, b] = \{x : a \leq x \leq b\}$ is finite.

The various relations between subgroups of $\mathbf{F}_k$ are the following:

$$J \in \mathrm{Crit}(H) \;\Rightarrow\; H \leq_{alg} J \;\Rightarrow\; H \leq_{\vec{x}} J \;\Rightarrow\; H \leq J$$

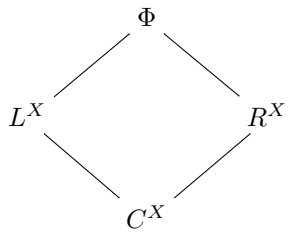
for any $H, J \leq \mathbf{F}_k$ and any basis X (see Sections 2.3 and 2.4).

Recall that the main theorems of this paper follow from Theorem 2.1.8, which estimates the expected number of common fixed points of $\alpha_n(H)$, where $H \leq_{fg} \mathbf{F}_k$ and α_n is a random homomorphism in $\mathrm{Hom}(\mathbf{F}_k, S_n)$. This result is achieved by studying a broader question: For every pair of $H, J \leq_{fg} \mathbf{F}_k$ such that $H \leq J$, we define for $n \in \mathbb{N}$

$$\Phi_{H,J}(n) = \text{The expected number of common fixed points of } \alpha_{J,n}(H), \tag{2.2.1}$$

where $\alpha_{J,n} \in \mathrm{Hom}(J, S_n)$ is a random homomorphism (chosen with uniform distribution). In this perspective, Nica finds $\lim_{n\to\infty} \Phi_{\langle w\rangle,\mathbf{F}_k}(n)$, and shows that it separates powers and non-powers. Theorem 2.1.8 shows that the first two terms in the expansion of $\Phi_{\langle w\rangle,\mathbf{F}_k}(n)$ yield w's primitivity rank, which in particular distinguishes powers ($\pi(w) = 1$) and primitives ($\pi(w) = \infty$). Furthermore, the same holds for subgroups using $\Phi_{H,\mathbf{F}_k}(n)$.

As remarked, in order to understand $\Phi_{H,\mathbf{F}_k}$ we turn to analyze the totality of functions $\Phi_{H,J}$, for various $H \leq J \leq \mathbf{F}_k$. We apply the machinery of Möbius inversions to the incidence algebra arising from the locally finite poset $(\mathfrak{sub}_{fg}(\mathbf{F}_k), \leq_{\vec{x}})$. The local finiteness of the order $\leq_{\vec{x}}$ allows us to "derive" the function Φ and obtain its "right derivation" R^X, its "left derivation" L^X, and its "two sided derivation" C^X (see Section 2.5). For instance, $\Phi_{H,J}$ can be presented as finite sums of R^X:



$$\Phi_{H,J} = \sum_{M \in [H,J]_{\vec{X}}} R^X_{H,M}$$

(here $[H,J]_{\vec{X}}$ is an abbreviation for $[H,J]_{\leq_{\vec{X}}}$, i.e. $[H,J]_{\vec{X}} = \{M \mid H \leq_{\vec{x}} M \leq_{\vec{x}} J\}$).

The proof of Theorem 2.1.8 is then based on a series of lemmas and propositions characterizing Φ and its three derivations:

- (Proposition 2.5.1) The right derivation R^X is supported on algebraic extensions, i.e. if $H \leq_{\vec{x}} M$ but M is not an algebraic extension of H then $R^X_{H,M} \equiv 0$.
- (The discussion in Section 2.6) The random homomorphism $\alpha_{J,n} \in \mathrm{Hom}(J, S_n)$ can be encoded as a random covering-space $\widehat{\Gamma}$ of the core graph $\Gamma_X(J)$, and $\Phi_{H,J}(n)$ can then be interpreted as the expected number of lifts of $\Gamma_X(H)$ into $\widehat{\Gamma}$.
- (Lemmas 2.6.3 and 2.6.4) The left derivation L^X is the expected number of *injective* lifts of the core graph $\Gamma_X(H)$ into the random covering $\widehat{\Gamma}$ of the core graph $\Gamma_X(J)$, and a rational expression can be computed for $L^X_{H,J}$.
- (Proposition 2.7.1 and Section 2.7.1) An analysis involving Stirling numbers of the rational expressions for L^X yields a combinatorial meaning for the two-sided derivation C^X. Using the classification of primitivity rank we then obtain a first-order estimate for the size of $C^X_{H,J}$.
- (Proposition 2.7.2) From C^X we return to R^X (by "left-integration"), obtaining that whenever $H \leq_{alg} M$ we have

  $$R^X_{H,M} = \frac{1}{n^{\widetilde{\mathrm{rk}}(M)}} + O\left(\frac{1}{n^{\widetilde{\mathrm{rk}}(M)+1}}\right)$$

  and by right integration of R^X, we obtain the order of magnitude of Φ, which was our goal.

The paper is arranged as follows: in Section 2.3 the notion of core graphs is explained in details, as well as the partial order $\leq_{\vec{x}}$ and some of the results from [Pud14] which are used here. In Section 2.4 we survey the main properties of algebraic extensions of free groups. Section 2.5 is devoted to recalling Möbius derivations on locally-finite posets and introducing the different derivations of Φ. In Section 2.6 we discuss the connection of the problem to random coverings of graphs and analyze the left derivation L^X. The proof of Theorem 2.1.8 is completed in Section 2.7 via the analysis of the two-sided derivation C^X and the consequence of the latter on the right derivation R^X. Finally, corollaries of our results to the field of profinite groups, and to decidability questions in group theory, are discussed in Section 2.8. We finish with a list of open problems naturally arising from this paper. For the reader's convenience, there is also a glossary of notions and notations at the end of this manuscript.

2.3 Core graphs and the partial order of covers

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 (Stallings) core-graph associated with* H and is denoted by $\Gamma_X(H)$. We recall the notion of the Schreier (right) coset graph of H with respect to the basis X, denoted by $\overline{\Gamma}_X(H)$. This is a directed, pointed and edge-labeled graph. Its vertex set is the set of all right cosets of H in $\mathbf{F}_k$, where the basepoint corresponds to the trivial coset H. For every coset Hw and every basis-element x_j there is a directed j-edge (short for x_j-edge) going from the vertex Hw to the vertex Hwx_j.[†]

The core graph $\Gamma_X(H)$ is obtained from $\overline{\Gamma}_X(H)$ by omitting all the vertices and edges of $\overline{\Gamma}_X(H)$ which are not traced by any reduced (i.e., non-backtracking) path that starts and ends at the basepoint. Stated informally, we trim all "hanging trees" from $\overline{\Gamma}_X(H)$. Formally, $\Gamma_X(H)$ is the induced subgraph of $\overline{\Gamma}_X(H)$ whose vertices are all cosets Hw (with w reduced), such that for some word w' the concatenation ww' is reduced, and $w \cdot w' \in H$. To illustrate, Figure 2.3.1 shows the graphs $\overline{\Gamma}_X(H)$ and $\Gamma_X(H)$ for $H = \langle x_1 x_2 x_1^{-3}, x_1^2 x_2 x_1^{-2} \rangle \leq \mathbf{F}_2$. Note that the graph $\overline{\Gamma}_X(H)$ is $2k$-regular: every vertex has exactly one outgoing j-edge and one incoming j-edge, for every $1 \leq j \leq k$. Every vertex of $\Gamma_X(H)$ has *at most* one outgoing j-edge and *at most* one incoming j-edge, for every $1 \leq j \leq k$.

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

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

However, it is not hard to show that when Γ is a core graph, then $\pi_1^X(\Gamma)$ *is* isomorphic to $\pi_1(\Gamma)$ (e.g. [MVW07]). In this case the labeling gives a canonical identification of $\pi_1(\Gamma)$ as a subgroup of $\mathbf{F}(X)$. It is an easy observation that

$$\pi_1^X\left(\overline{\Gamma}_X(H)\right) = \pi_1^X\left(\Gamma_X(H)\right) = H \tag{2.3.1}$$

This gives a one-to-one correspondence between subgroups of $\mathbf{F}(X) = \mathbf{F}_k$ and core graphs labeled

[†]Alternatively, $\overline{\Gamma}_X(H)$ is the quotient $H\backslash T$, where T is the Cayley graph of $\mathbf{F}_k$ with respect to the basis X, and F_k (and thus also H) acts on this graph from the left. Moreover, this is the covering-space of $\overline{\Gamma}_X(F_k) = \Gamma_X(F_k)$, the bouquet of k loops, corresponding to H, via the correspondence between pointed covering spaces of a space Y and subgroups of its fundamental group $\pi_1(Y)$.

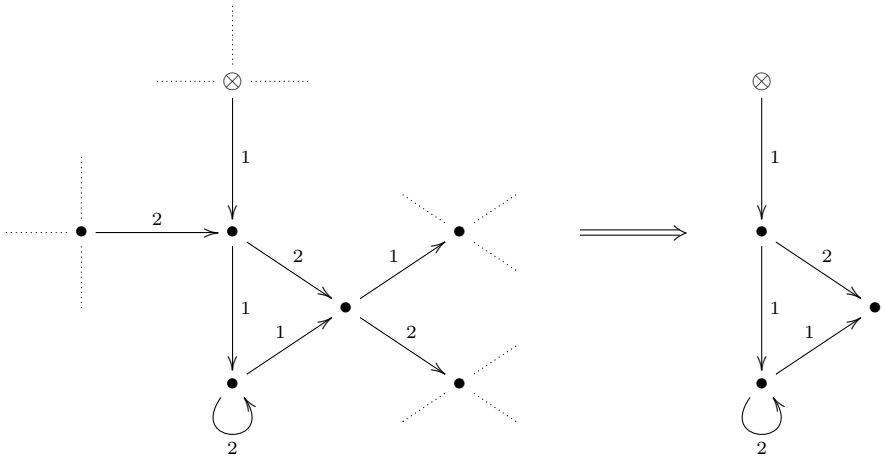


Figure 2.3.1: $\overline{\Gamma}_X(H)$ and $\Gamma_X(H)$ for $H = \langle x_1 x_2 x_1^{-3}, x_1^2 x_2 x_1^{-2} \rangle \leq \mathbf{F}_2$. The Schreier coset graph $\overline{\Gamma}_X(H)$ is the infinite graph on the left (the dotted lines represent infinite 4-regular trees). The basepoint "$\otimes$" corresponds to the trivial coset H, the vertex below it corresponds to the coset Hx_1, the one further down corresponds to $Hx_1^2 = Hx_1x_2x_1^{-1}$, etc. The core graph $\Gamma_X(H)$ is the finite graph on the right, which is obtained from $\overline{\Gamma}_X(H)$ by omitting all vertices and edges that are not traced by reduced closed paths around the basepoint.

by X. Namely, π_1^X and Γ_X are the inverses of each other in a bijection (Galois correspondence)

$$\left\{ \begin{array}{c} \text{Subgroups} \\ \text{of } \mathbf{F}(X) \end{array} \right\} \underset{\pi_1^X}{\overset{\Gamma_X}{\rightleftarrows}} \left\{ \begin{array}{c} \text{Core graphs} \\ \text{labeled by } X \end{array} \right\} \tag{2.3.2}$$

Core graphs were introduced by Stallings [Sta83]. Our definition is slightly different, and closer to the one in [KM02, MVW07] in that we allow the basepoint to be of degree one, and in that our graphs are directed and edge-labeled. We remark that it is possible to study core graphs from a purely combinatorial point of view, as labeled pointed connected graphs satisfying

(1) No two equally labeled edges originate or terminate at the same vertex.

(2) Every vertex and edge are traced by some non-backtracking closed path around the basepoint.

Starting with this definition, every choice of an ordered basis for $\mathbf{F}_k$ then gives a correspondence between these graphs and subgroups of $\mathbf{F}_k$.

In this paper we are mainly interested in finite core graphs, and we now list some basic properties of these (proofs can be found in [Sta83, KM02, MVW07]).

Claim 2.3.1. *Let H be a subgroup of $\mathbf{F}_k$ with an associated core graph $\Gamma = \Gamma_X(H)$. The Euler Characteristic of a graph, denoted $\chi(\cdot)$, is the number of vertices minus the number of edges.*

(1) $\mathrm{rk}(H) < \infty \iff \Gamma$ *is finite.*

(2) $\widetilde{\mathrm{rk}}(H) = -\chi(\Gamma)$.

(3) The correspondence (2.3.2) restricts to a correspondence between $\mathfrak{sub}_{fg}(\mathbf{F}_k)$ and finite core graphs.

Given a finite set of words $\{h_1, \ldots, h_m\} \subseteq \mathbf{F}(X)$ that generate a subgroup H, the core graph $\Gamma_X(H)$ can be algorithmically constructed as follows. Every h_i corresponds to some path with directed edges labeled by the x_j's (we assume the elements are given in reduced forms, otherwise we might need to prune leaves at the end of the algorithm). Merge these m paths to a single graph (bouquet) by identifying all their $2m$ end-points to a single vertex, which is marked as the basepoint. The labeled fundamental group of this graph is clearly H. Then, as long as there are two j-labeled edges with the same terminus (resp. origin) for some j, merge the two edges and their origins (resp. termini). Such a step is often referred to as *Stallings folding*. It is fairly easy to see that each folding step does not change the labeled fundamental group of the graph, that the resulting graph is indeed $\Gamma_X(H)$, and that the order of folding has no significance. To illustrate, we draw in Figure 2.3.2 a folding process by which we obtain the core graph $\Gamma_X(H)$ of $H = \langle x_1 x_2 x_1^{-3}, x_1^2 x_2 x_1^{-2} \rangle \leq \mathbf{F}_2$ from the given generating set.

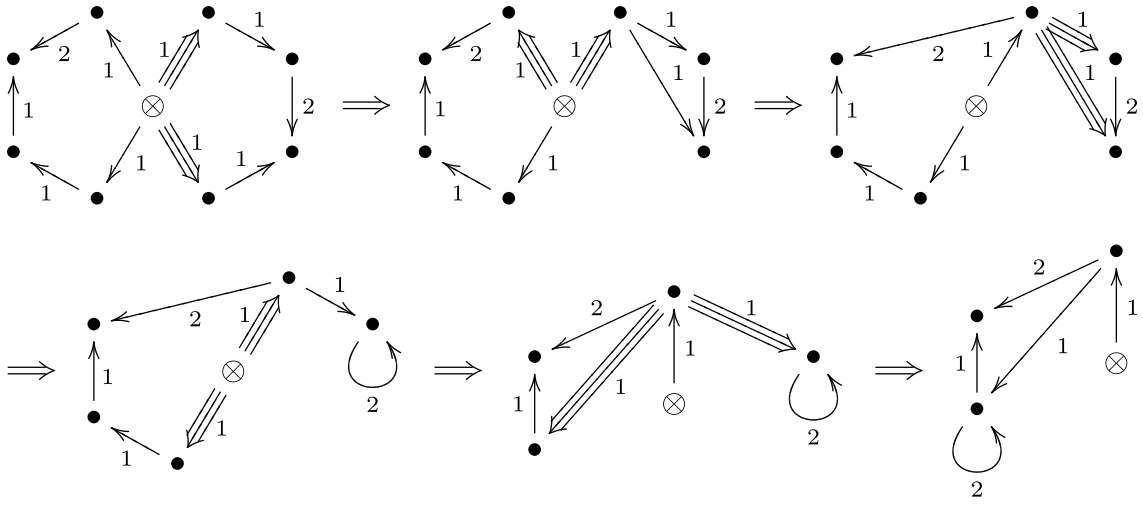


Figure 2.3.2: Constructing the core graph $\Gamma_X(H)$ of $H = \langle x_1 x_2 x_1^{-3}, x_1^2 x_2 x_1^{-2} \rangle \leq \mathbf{F}_2$ from the given generating set. We start with the upper left graph which contains a distinct loop at the basepoint for each (reduced) element of the generating set. Then, at an arbitrary order, we merge pairs of equally-labeled edges which share the same origin or the same terminus (here we mark by triple arrows the pair of edges being merged next). The graph at the bottom right is $\Gamma_X(H)$, as it has no equally-labeled edges sharing the same origin or terminus.

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 2.3.1, each of the following properties is either proven in (some of) [Sta83, KM02, MVW07] or an easy observation:

Claim 2.3.2. *Let $H, J, L \leq \mathbf{F}_k$ be subgroups. Then*

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

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

(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 $\eta^X_{H \to J}$ is injective, then $H \overset{*}{\leq} J$.[‡]

[†]Points (1)-(3) can be formulated by saying that (2.3.2) is in fact an isomorphism of categories, given by the functors π_1^X and Γ_X.

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

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

A special role is played by *surjective* morphisms of core graphs:

Definition 2.3.3. Let $H \leq J \leq \mathbf{F}_k$. Whenever $\eta^X_{H\to J}$ is surjective, we say that $\Gamma_X(H)$ *covers* $\Gamma_X(J)$ or that $\Gamma_X(J)$ *is a quotient of* $\Gamma_X(H)$. We indicate this by $\Gamma_X(H) \twoheadrightarrow \Gamma_X(J)$. As for the groups, we say that H *X-covers* J and denote this by $H \leq_{\vec{x}} J$.

By "surjective" we mean surjective on both vertices and edges. Note that we use the term "covers" even though in general this is *not* a topological covering map (a morphism between core graphs is always locally injective at the vertices, but it need not be locally bijective). In Section 2.6 we do study topological covering maps, and we reserve the term "coverings" for these.

For instance, $H = \langle x_1x_2x_1^{-3}, x_1^2x_2x_1^{-2}\rangle \leq \mathbf{F}_k$ 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 2.3.3. As another example, a core graph Γ X-covers $\Gamma_X(\mathbf{F}_k)$ (which is merely a wedge of k loops) if and only if it contains edges of all k labels.

As implied by the notation, the relation $H \leq_{\vec{x}} J$ indeed depends on the given basis X of $\mathbf{F}_k$. For example, if $H = \langle x_1x_2\rangle$ then $H \leq_{\vec{x}} \mathbf{F}_2$. However, for $Y = \{x_1x_2, x_2\}$, H does not Y-cover $\mathbf{F}_2$, as $\Gamma_Y(H)$ consists of a single vertex and a single loop and has no quotients apart from itself.

It is easy to see that the relation "$\leq_{\vec{x}}$" indeed constitutes a partial ordering of the set of subgroups of $\mathbf{F}_k$. We make a few other useful observations:

Claim 2.3.4. *Let $H, J, L \leq \mathbf{F}_k$ be subgroups. Then*

(1) Whenever $H \leq J$ there exists an intermediate subgroup M such that $H \leq_{\vec{x}} M \overset{*}{\leq} J$.

(2) If one adds the condition that $\Gamma_X(M)$ embeds in $\Gamma_X(J)$, then this M is unique.

(3) If $H \leq_{\vec{x}} J$ and $H \leq_{\vec{x}} L \leq J$, then $L \leq_{\vec{x}} J$.

(4) If H is finitely generated then it X-covers only a finite number of groups. In particular, the poset $(\mathfrak{sub}_{fg}(\mathbf{F}_k), \leq_{\vec{x}})$ is locally finite.

Proof. Point (1) follows from the factorization of the morphism $\eta^X_{H\to J}$ to a surjection followed by an embedding. Indeed, it is easy to see that the image of $\eta^X_{H\to J}$ is a sub-graph of $\Gamma_X(J)$ which is in itself a core graph. Namely, it contains no "hanging trees" (edges and vertices not traced by reduced paths around the basepoint). Let $M = \pi_1^X\left(\operatorname{im}\eta^X_{H\to J}\right)$ be the subgroup corresponding to this sub-core-graph. (1) now follows from points (1) and (4) in Claim 2.3.2. Point (2) follows from the uniqueness of such factorization of a morphism. Point (3) follows from the fact that if $\eta^X_{H\to J} = \eta^X_{L\to J}\circ\eta^X_{H\to L}$ is surjective then so is $\eta^X_{L\to J}$. Point (4) follows from the fact that $\Gamma_X(H)$ is finite (Claim 2.3.1(1)) and thus has only finitely many quotients, and each quotient correspond to a single group (by (2.3.2)). □

In [MVW07], the set of X-quotients of H

$$[H,\infty)_{\vec{X}} = \{J \mid H \leq_{\vec{x}} J\} \tag{2.3.3}$$

is called the *X-fringe* of H. Claim 2.3.4(4) states in this terminology that for every $H \leq_{fg} \mathbf{F}_k$ (and every basis X), $\left|[H,\infty)_{\vec{X}}\right| < \infty$. Note that $[H,\infty)_{\vec{X}}$ always contains the supremum of its elements, namely the group generated by the elements of X which label edges in $\Gamma_X(H)$ (which is $\pi_1^X\left(\operatorname{im}\eta^X_{H\to\mathbf{F}_k}\right)$). (We remark that in the special case of $H = \langle w\rangle$ for some $w \in \mathbf{F}_k$, the set $[\langle w\rangle,\infty)_{\vec{X}}$ appears also in [Tur96] and, in a very different language, in the aforementioned [Nic94].)

It is easy to see that quotients of $\Gamma_X(H)$ are determined by the partition they induce of the vertex set $V(\Gamma_X(H))$. However, not every partition P of $V(\Gamma_X(H))$ corresponds to a quotient core-graph: in the resulting graph, which we denote by ${}^{\Gamma_X(H)}/_P$, two distinct j-edges may have the

same origin or the same terminus. 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 (as demonstrated in Figure 2.3.2) until we obtain a core graph. Since Stallings foldings do not affect π_1^X, the core graph we obtain in this manner is $\Gamma_X(J)$, where $J = \pi_1^X\left(\Gamma_X(H)/P\right)$. The resulting partition $\bar{P}$ of $V(\Gamma_X(H))$ (as the fibers of $\eta^X_{H\to J}$) 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 2.3.3.

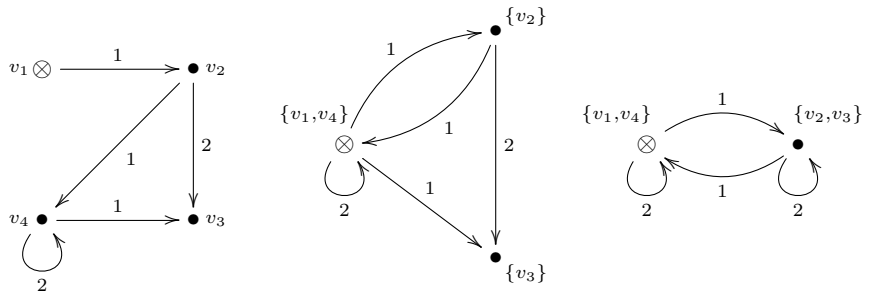


Figure 2.3.3: The left graph is the core graph $\Gamma_X(H)$ of $H = \left\langle x_1x_2x_1^{-3}, x_1^2x_2x_1^{-2}\right\rangle \leq \mathbf{F}_2$. Its vertices are denoted by $v_1,\ldots,v_4$. The graph in the middle is the quotient $\Gamma_X(H)/P$ corresponding to the partition $P = \{\{v_1,v_4\},\{v_2\},\{v_3\}\}$. 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 (illustrated in Figure 2.3.2). The resulting core graph, $\Gamma_X\left(\pi_1^X\left(\Gamma_X(H)/P\right)\right)$, is shown on the right and corresponds to the partition $\bar{P} = \{\{v_1,v_4\},\{v_2,v_3\}\}$.

Thus, there is sense in examining the quotient of a core graph Γ "generated" by some partition P of its vertex set, namely, $\Gamma_X\left(\pi_1^X\left(\Gamma/P\right)\right)$. The most interesting case is that of the "simplest" partitions: those which identify only a single pair of vertices. Before looking at these, we introduce a measure for the complexity of partitions: if $P \subseteq 2^{\mathcal{X}}$ is a partition of some set $\mathcal{X}$, let

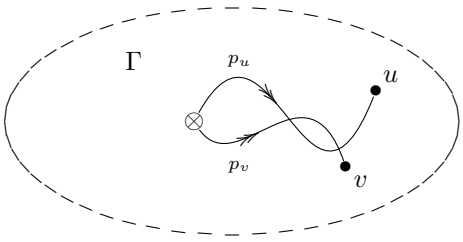
$$\|P\| \stackrel{def}{=} |\mathcal{X}| - |P| = \sum_{B\in P}(|B|-1). \tag{2.3.4}$$

Namely, $\|P\|$ is the number of elements in the set minus the number of blocks in the partition. For example, $\|P\| = 1$ iff P identifies only a single pair of elements. It is not hard to see that $\|P\|$ is also the minimal number of identifications one needs to make in $\mathcal{X}$ in order to obtain the equivalence relation P.

Definition 2.3.5. Let Γ be a core graph and let P be a partition of $V(\Gamma)$ with $\|P\| = 1$, i.e. having a single non-trivial block, of size two. Let Δ be the core graph generated from Γ by P. We then say that Δ is an *immediate quotient* of Γ.

Alternatively, we say that Δ is *generated by identifying a single pair* of vertices of Γ. For instance, the rightmost core graph in Figure 2.3.3 is an immediate quotient of the leftmost one.

The main reason that immediate quotients are interesting is their algebraic significance. Let $H, J \leq \mathbf{F}_k$ with $\Gamma = \Gamma_X(H), \Delta = \Gamma_X(J)$ their core graphs, and assume that Δ is an immediate quotient of Γ obtained by identifying the vertices $u, v \in V(\Gamma)$. Now let $w_u, w_v \in \mathbf{F}_k$ be the words corresponding to some paths p_u, p_v in Γ from the basepoint to u and v respectively (note that these paths are not unique). It is not hard to see that identifying u and v has the same effect as adding the word $w = w_u w_v^{-1}$ to H and considering the generated group. Namely, that $J = \langle H, w\rangle$.



The relation of immediate quotients gives the set of finite core graphs (with edges labeled by $1,\ldots,k$) the structure of a directed acyclic graph $(DAG)^{\dagger}$. This DAG was first introduced in [Pud14], and is denoted by $\mathcal{D}_k$. The set of vertices of $\mathcal{D}_k$ consists of the aforementioned core graphs, and its directed edges connect every core graph to its immediate quotients. Every ordered basis $X=\{x_1,\ldots,x_k\}$ of $\mathbf{F}_k$ determines a one-to-one correspondence between the vertices of this graph and $\mathfrak{sub}_{fg}(\mathbf{F}_k)$.

In the case of finite core graphs, Δ is a quotient of Γ if and only if Δ is reachable from Γ in $\mathcal{D}_k$ (that is, there is a directed path from Γ to Δ). In other words, if $H\leq_{fg}\mathbf{F}_k$ then $H\leq_{\vec{x}}J$ iff $\Gamma_X(J)$ can be obtained from $\Gamma_X(H)$ by a finite sequence of immediate quotients. Thus, for any $H\leq_{fg}\mathbf{F}_k$, the subgraph of $\mathcal{D}_k$ induced by the descendants of $\Gamma_X(H)$ consists of all quotients of $\Gamma_X(H)$, i.e. of all (core graphs corresponding to) elements of $[H,\infty)_{\vec{x}}$. By Claim 2.3.4(4), this subgraph is finite. In Figure 2.3.4 we draw the subgraph of $\mathcal{D}_k$ consisting of all quotients of $\Gamma_X(H)$ when $H=\langle x_1x_2x_1^{-1}x_2^{-1}\rangle$. The edges of this subgraph (i.e. immediate quotients) are denoted by the dashed arrows in the figure.

It is now natural to define a distance function between a finite core graph and each of its quotients:

Definition 2.3.6. Let $H,J\leq_{fg}\mathbf{F}_k$ be subgroups such that $H\leq_{\vec{x}}J$, and let $\Gamma=\Gamma_X(H)$, $\Delta=\Gamma_X(J)$ be the corresponding core graphs. We define the *X-distance* between H and J, denoted $\rho_X(H,J)$ or $\rho(\Gamma,\Delta)$, to be the shortest length of a directed path from Γ to Δ in $\mathcal{D}_k$.

In other words, $\rho_X(H,J)$ is the length of the shortest series of immediate quotients that yields Δ from Γ. There is another useful equivalent definition for the X-distance. To see this, assume that Γ' is generated from Γ by the partition P of $V(\Gamma)$ and let $\eta:\Gamma\twoheadrightarrow\Gamma'$ be the morphism. For every $x,y\in V(\Gamma')$, let $x'\in\eta^{-1}(x),y'\in\eta^{-1}(y)$ be arbitrary vertices in the fibers, and let P' be the partition of $V(\Gamma)$ obtained from P by identifying x' and y'. It is easy to see that the core graph generated from Γ' by identifying x and y is the same as the one generated by P' from Γ. From these considerations we obtain that

$$\rho_X(H,J)=\min\left\{\|P\|\;\middle|\;\begin{array}{c}P\text{ is a partition of }V(\Gamma_X(H))\\ \text{such that }\pi_1^X\left({}^{\Gamma_X(H)}/_P\right)=J\end{array}\right\}. \tag{2.3.5}$$

For example, if Δ is an immediate quotient of Γ then $\rho_X(H,J)=\rho(\Gamma,\Delta)=1$. For $H=\langle x_1x_2x_1^{-1}x_2^{-1}\rangle$, $\Gamma_X(H)$ has four quotients at distance 1 and two at distance 2 (see Figure 2.3.4).

As mentioned earlier, by merging a single pair of vertices of $\Gamma_X(H)$ (and then folding) we obtain the core graph of a subgroup J obtained from H by adding some single generator (thought not every element of $\mathbf{F}_k$ can be added in this manner). Thus, by taking an immediate quotient, the rank of the associated subgroup increases at most by 1 (in fact, it may also stay unchanged or even decrease). This implies that whenever $H\leq_{\vec{x}}J$, one has

$$rk(J)-rk(H)\quad\leq\quad\rho_X(H,J) \tag{2.3.6}$$

In [Pud14] (Lemma 3.3), the distance is bounded from above as well:

†that is, a directed graph with no directed cycles.

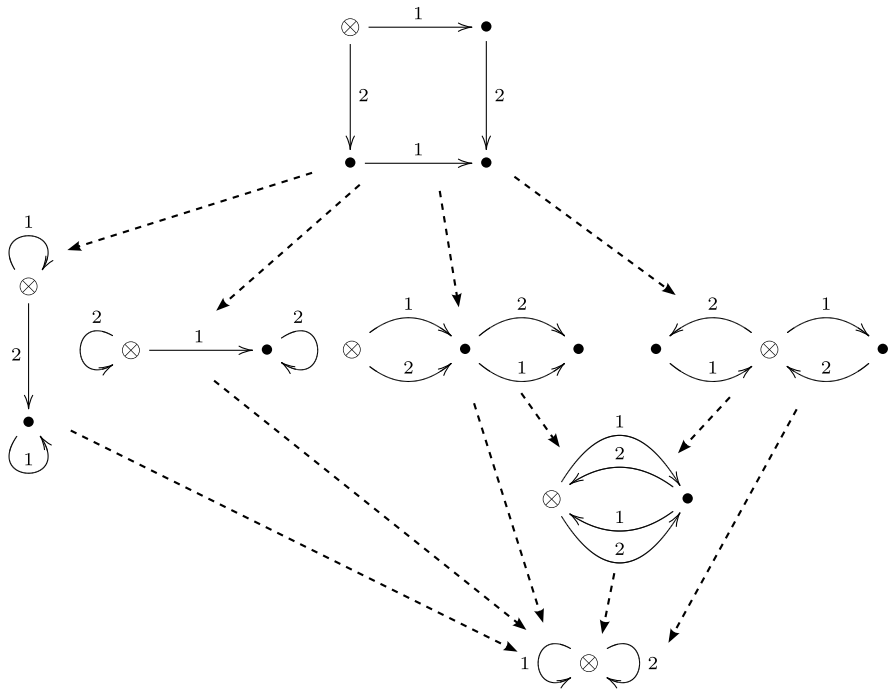


Figure 2.3.4: The subgraph of $\mathcal{D}_k$ induced by $[H,\infty)_{\vec{x}}$, that is, all quotients of the core graph $\Gamma = \Gamma_X(H)$, for $H = \langle x_1 x_2 x_1^{-1} x_2^{-1} \rangle$. The dashed arrows denote immediate quotients, i.e. quotients generated by merging a single pair of vertices. Γ has exactly seven quotients: itself, four immediate quotients, and two quotients at distance 2.

Claim 2.3.7. *Let $H, J \leq_{fg} \mathbf{F}_k$ such that $H \leq_{\vec{x}} J$. Then*

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

We shall make use of the following theorem, which asserts that the lower bound is attained if and only if H is a free factor of J:

Theorem 2.3.8 ([Pud14, Theorem 1.1])**.** *Let $H, J \leq_{fg} \mathbf{F}_k$ and assume further that $H \leq_{\vec{x}} J$. Then $H \overset{*}{\leq} J$ if and only if*

$$\rho_X(H,J) = \mathrm{rk}(J) - \mathrm{rk}(H)$$

In fact, the implication which is needed for our proof is trivial: As mentioned above, merging two vertices in $\Gamma_X(H)$ translates to adding some generator to H. If it is possible to obtain $\Gamma_X(J)$ from $\Gamma_X(H)$ by $rk(J) - rk(H)$ merging steps, this means we can obtain J from H by adding $rk(J) - rk(H)$ complementary generators to H, hence $H \overset{*}{\leq} J$.[†] The other implication is not trivial and constitutes the essence of the proof of Theorem 1.1 in [Pud14]. The difficulty is that when $H \overset{*}{\leq}_{\vec{x}} J$, it is not apriori obvious why it is possible to find $rk(J) - rk(H)$ complementing generators of J from H, so that each of them can be realized by merging a pair of vertices in $\Gamma_X(H)$.

We finish this section with a classical fact about free factors that will be useful in the next section.

Claim 2.3.9. *Let H, J and K be subgroups of $\mathbf{F}_k$.*

(1) If $H \overset{*}{\leq} J$ and $K \leq J$, then $H \cap K \overset{*}{\leq} K$.

[†]This relies on the well known fact that a set of size k which generates $\mathbf{F}_k$ is a basis.

(2) If $H, K \overset{*}{\leq} J$ then $H \cap K \overset{*}{\leq} J$.

(3) If $H \overset{*}{\leq} J$ then H is a free factor of any intermediate group $H \leq M \leq J$.

Proof. Let Y be a basis of J extending a basis Y_0 of H. Then $\Gamma_Y(J)$ and $\Gamma_Y(H)$ are bouquets of $|Y|, |Y_0|$ loops, respectively. It is easy to check that $\Gamma_Y(H \cap K)$ is obtained from $\Gamma_Y(K)$ as follows: first, delete the edges labeled by $Y \setminus Y_0$; then, keep only the connected component of the basepoint; finally, trim all "hanging trees" (see the proof of Claim 2.3.4). Consequently, $\Gamma_Y(H \cap K)$ is embedded in $\Gamma_Y(K)$. Claim 2.3.2(4) then gives (1), and (2) and (3) follow immediately. □

In particular, the last claim shows that if $H \overset{*}{\leq} \mathbf{F}_k$ then $\pi(H) = \infty$ (see Definition 2.1.7). On the other hand, if H is not a free factor of $\mathbf{F}_k$, then obviously $\pi(H) \leq \mathrm{rk}(\mathbf{F}_k) = k$. Thus $\pi(H) \in \{0, 1, 2, \ldots, k\} \cup \{\infty\}$.

2.4 Algebraic extensions and critical subgroups

We now return to the sparsest partial order we consider in this paper, that of algebraic extensions. All claims in this section appear in [KM02, MVW07], except for Lemma 2.4.4. We shall occasionally sketch some proofs in order to allow the reader to obtain better intuition and in order to exemplify the strength of core graphs.

Recall (Definition 2.2.1) that J is an algebraic extension of H, denoted $H \leq_{alg} J$, if $H \leq J$ and H is not contained in any proper free factor of J. For example, consider $H = \langle x_1 x_2 {x_1}^{-1} {x_2}^{-1} \rangle \leq \mathbf{F}_2$. A proper free factor of $\mathbf{F}_2$ has rank at most 1, and H is not contained in any subgroup of rank 1 other than itself (as $x_1 x_2 x_1^{-1} x_2^{-1}$ is not a proper power). Finally, H itself is not a free factor of $\mathbf{F}_2$ (as can be inferred from Theorem 2.3.8 and Figure 2.3.4). Thus, $H \leq_{alg} \mathbf{F}_2$. In fact, we shall see that in this case $[H, \infty)_{alg} = \{H, \mathbf{F}_2\}$.

We first show that "$\leq_{alg}$" is a partial order:

Claim 2.4.1. *The relation "$\leq_{alg}$" is transitive.*

Proof. Assume that $H \leq_{alg} M \leq_{alg} J$. Let $H \leq L \overset{*}{\leq} J$. By Claim 2.3.9(1), $L \cap M \overset{*}{\leq} M$. But $H \leq L \cap M$ and $H \leq_{alg} M$, so $L \cap M = M$, and thus $M \leq L$. So now $M \leq L \overset{*}{\leq} J$, and from $M \leq_{alg} J$ we obtain that $L = J$. □

Next, we show that "$\leq_{alg}$" is dominated by "$\leq_{\vec{x}}$" for every basis X of $\mathbf{F}_k$. Namely, if $H \leq_{alg} J$ then $H \leq_{\vec{x}} J$. This shows, in particular, that the poset $(\mathfrak{sub}_{fg}(\mathbf{F}_k), \leq_{alg})$ is locally-finite.

Claim 2.4.2. *If $H \leq_{alg} J$ then $H \leq_{\vec{x}} J$ for every basis X of $\mathbf{F}_k$.*

Proof. By Claim 2.3.4, there is an intermediate subgroup M such that $H \leq_{\vec{x}} M \overset{*}{\leq} J$, and from $H \leq_{alg} J$ it follows that $M = J$. □

Remark 2.4.3. It is natural to conjecture that the converse also holds, namely that if $H \leq_{\vec{x}} J$ for every basis X of $\mathbf{F}_k$ then $H \leq_{alg} J$. (In fact, this conjecture appears in [MVW07], Section 3.) This is, however, false: it turns out that for $H = \langle x_1^2 x_2^2 \rangle$ and $J = \langle x_1^2 x_2^2, x_1 x_2 \rangle$, $H \leq_{\vec{x}} J$ for every basis X of $\mathbf{F}_2$, but J is *not* an algebraic extension of H [PP14]. However, there are bases of $\mathbf{F}_3$ with respect to which H does not cover J. Hence, it is still plausible that some weaker version of the conjecture holds, e.g. that $H \leq_{alg} J$ if and only if for every embedding of J in a free group F, and for every basis X of F, $H \leq_{\vec{x}} J$. It is also plausible that the original conjecture from [MVW07] holds for $\mathbf{F}_k$ with $k \geq 3$.
In a similar fashion, one can ask whether $H \leq J$ if and only if for some basis X of $\mathbf{F}_k$, $H \leq_{\vec{x}} J$.

Claim 2.4.2 completes the proof of the relations, mentioned in Section 2.2, between the different partial orders we consider in this paper: inclusion, the family $\leq_{\vec{x}}$, and algebraic extensions. Recall that H-critical subgroups are a special kind of algebraic extensions. Thus:

$$\mathrm{Crit}(H) \subseteq [H,\infty)_{alg} \subseteq [H,\infty)_{\vec{X}} \subseteq [H,\infty)_{\leq}.$$

Theorem 2.3.8 and Claim 2.4.2 give the following criterion for algebraic extensions:

Lemma 2.4.4. *Let $H \leq_{fg} \mathbf{F}_k$. The algebraic extensions of H are the elements of $[H,\infty)_{\vec{X}}$ which are* not *immediate quotients of any subgroup in $[H,\infty)_{\vec{X}}$ of smaller rank.*

Proof. Let $J \in [H,\infty)_{\vec{X}}$. If J is an immediate X-quotient of $L \in [H,\infty)_{\vec{X}}$ with $\mathrm{rk}(L) < \mathrm{rk}(J)$, then by Theorem 2.3.8 $H \leq L \overset{*}{\lneq} J$, hence J is not an algebraic extension of H. On the other hand, assume there exists some L such that $H \leq L \overset{*}{\lneq} J$. By Claim 2.3.4(1), there exists M such that $H \leq_{\vec{x}} M \overset{*}{\leq} L \overset{*}{\lneq} J$. By Claim 2.3.4(3), $M \overset{*}{\lneq}_{\vec{X}} J$. From Theorem 2.3.8 it follows that there is a chain of immediate quotients $M = M_0 \leq M_1 \leq \ldots \leq M_r = J$ inside $[H,\infty)_{\vec{X}}$ with $\mathrm{rk}(M_{i+1}) = \mathrm{rk}(M_i) + 1$, and M_{r-1} is the group we have looked for. □

Since the subgraph of $\mathcal{D}_k$ induced by the vertices corresponding to $[H,\infty)_{\vec{X}}$, namely $\Gamma_X(H)$ and its descendants, is finite and can be effectively computed, Lemma 2.4.4 yields a straight-forward algorithm to find all algebraic extensions of a given $H \leq_{fg} \mathbf{F}_k$ (this algorithm was first introduced in [Pud14]). This, in particular, allows one to find all H-critical subgroups, and thus to compute the primitivity rank $\pi(H)$: the subgroups constituting $\mathrm{Crit}(H)$ are those in $(H,\infty)_{alg}$ of minimal rank, which is $\pi(H)$. For instance, Figure 2.3.4 shows that for $H = \langle x_1 x_2 x_1^{-1} x_2^{-1} \rangle$ we have $H = \{H, \mathbf{F}_2\}$. Thus, $\mathrm{Crit}(H) = \{\mathbf{F}_2\}$ and $\pi(H) = 2$ (so $\widetilde{\pi}(H) = 1$).

We conclude this section with yet another elegant result from [KM02, MVW07] that will be used in the proof of Theorem 2.1.8. In the spirit of field extensions, it says that every extension of subgroups of $\mathbf{F}_k$ has a unique factorization to an algebraic extension followed by a free extension (compare this with Claim 2.3.4(1,2)):

Claim 2.4.5. *Let $H \leq J$ be free groups. Then there is a unique subgroup L of J such that $H \leq_{alg} L \overset{*}{\leq} J$. Moreover, L is the intersection of all intermediate free factors of J and the union of all intermediate algebraic extensions of H:*

$$L = \bigcap_{M:\, H \leq M \overset{*}{\leq} J} M = \bigcup_{M:\, H \leq_{alg} M \leq J} M \tag{2.4.1}$$

In particular, the intersection of all free factors is a free factor, and the union of all algebraic extensions is an algebraic extension. Claim 2.4.5 is true in general, but we describe the proof only of the slightly simpler case of finitely generated subgroups. We need only this case in this paper.

Proof. By Claim 2.3.9 and rank considerations, the intersection in the middle of (2.4.1) is by itself a free factor of J. Denote it by L, so we have $H \leq L \overset{*}{\leq} J$. Clearly, L is an algebraic extension of H (otherwise it would contain a proper free factor). But we claim that L contains every other intermediate algebraic extension of H. Indeed, let $H \leq_{alg} M \leq J$. By Claim 2.3.9(1), $H \leq M \cap L \overset{*}{\leq} M$, so $M \cap L = M$, that is $M \leq L$. □

2.5 Möbius inversions

Let $(P,\leq)$ be a locally-finite poset and let A be a commutative ring with unity. Then there exists an *incidence algebra*[†] of all functions from pairs $\{(x,y) \in P \times P \mid x \leq y\}$ to A. In addition to point-wise addition and scalar multiplication, it has an associative multiplication defined by convolution:

[†]The theory of incidence algebras of posets can be found in [Sta97].

$$(f * g)(x,y) = \sum_{z \in [x,y]} f(x,z) g(z,y)$$

(where $x \leq y$ and $[x,y] = \{z \mid x \leq z \leq y\}$). The unit element is the diagonal

$$\delta(x,y) = \begin{cases} 1 & x = y \\ 0 & x \lneq y \end{cases}.$$

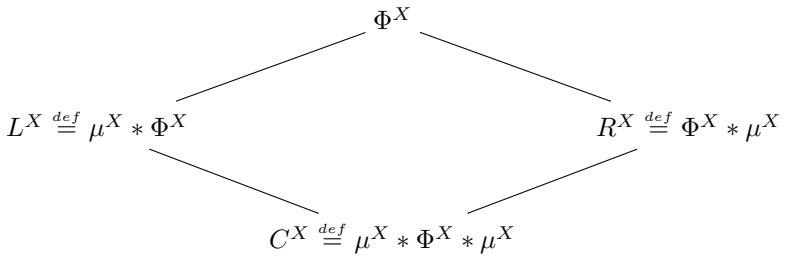
Functions with invertible diagonal entries (i.e. $f(x,x) \in A^\times$ for all $x \in P$) are invertible w.r.t. this multiplication. Most famously, the constant ζ function, which is defined by $\zeta(x,y) = 1$ for all $x \leq y$, is invertible, and its inverse, μ, is called the *Möbius function* of P. This means that $\zeta * \mu = \mu * \zeta = \delta$, i.e., for every pair $x \leq y$

$$\sum_{z \in [x,y]} \mu(z,y) = (\zeta * \mu)(x,y) = \delta(x,y) = (\mu * \zeta)(x,y) = \sum_{z \in [x,y]} \mu(x,z).$$

Let f be some function in the incidence algebra. The function $f * \zeta$, which satisfies $(f * \zeta)(y) = \sum_{z \in [x,y]} f(z)$, is analogous to the right-accumulating function in calculus (for $g : \mathbb{R} \to \mathbb{R}$ this is the function $G(y) = \int_{z \in [x,y]} g(z)\, dz$). Thus, multiplying a function on the right by μ can be thought of as "right derivation". Similarly, one thinks of multiplying from the left by ζ and μ as left integration and left derivation, respectively.

Recall the function Φ (2.2.1), defined for every pair of free subgroups $H, J \leq_{fg} \mathbf{F}_k$ such that $H \leq J$: $\Phi_{H,J}(n)$ is the expected number of common fixed points of $\alpha_{J,n}(H)$, where $\alpha_{J,n} \in \mathrm{Hom}(J, S_n)$ is a random homomorphism chosen with uniform distribution. We think of Φ as a function from the set of such pairs (H,J) into the ring of functions $\mathbb{N} \to \mathbb{Q}$.

Let X be a basis of $\mathbf{F}_k$. We write Φ^X for the restriction of Φ to pairs (H,J) such that $H \leq_{\vec{x}} J$. As "$\leq_{\vec{x}}$" defines a locally finite partial ordering of $\mathfrak{sub}_{fg}(\mathbf{F}_k)$, there exists a matching Möbius function, $\mu^X = \left(\zeta^X\right)^{-1}$ (where $\zeta^X_{H,J} = 1$ for all $H \leq_{\vec{x}} J$). Our proof of Theorem 2.1.8 consists of a detailed analysis of the left, right, and two-sided derivations of Φ^X:



By definition, we have for every f.g. $H \leq_{\vec{x}} J$:

$$\Phi_{H,J} = \sum_{M \in [H,J]^{\rightarrow}_{X}} L^X_{M,J} = \sum_{M,N:\, H \leq_{\vec{x}} M \leq_{\vec{x}} N \leq_{\vec{x}} J} C^X_{M,N} = \sum_{N \in [H,J]^{\rightarrow}_{X}} R^X_{H,N} \tag{2.5.1}$$

Note that (2.5.1) can serve as definitions for the three functions L^X, C^X, R^X: for instance, $L^X = \mu^X * \Phi^X$ is equivalent to $\zeta^X * L^X = \Phi^X$, which is the leftmost equality above.

We begin the analysis of these functions by the following striking observation regarding R^X. Recall (Claim 2.4.2) that if $H \leq_{alg} J$ then $H \leq_{\vec{x}} J$ for every basis X. It turns out that the function R^X is supported on algebraic extensions alone, and moreover, is independent of the basis X.

Proposition 2.5.1. *Let $H, J \leq_{fg} \mathbf{F}_k$.*

(1) If $H \leq_{\vec{x}} J$ but J is not *an algebraic extension of H, then $R^X_{H,J} = 0$.*

(2) $R^X_{H,J} = R^Y_{H,J}$ for every basis Y of $\mathbf{F}_k$, whenever both are defined.

Remark 2.5.2. The only property of Φ we use is that $\Phi_{H,L} = \Phi_{H,J}$ whenever $H \leq L \overset{*}{\leq} J$, which is easy to see from the definition of Φ. Therefore, the proposition holds for the right derivation of every function with this property. In particular, the proposition holds for every "statistical" function, in which the value of (H, J) depends solely on the image of H via a uniformly distributed random homomorphism from J to some group G.

Proof. We show both claims at once by induction on $\left|[H,J]_{\vec{X}}\right|$, the size of the closed interval between H and J. The induction basis is $H = J$. That $H \leq_{alg} H$ is immediate. By (2.5.1), $R^X_{H,H} = \Phi_{H,H}$ and so $R^X_{H,H}$ is indeed independent of the basis X.

Assume now that $\left|[H,J]_{\vec{X}}\right| = r$ and that both claims are proven for every pair bounding an interval of size $< r$. By (2.5.1) and the first claim of the induction hypothesis,

$$R^X_{H,J} = \Phi_{H,J} - \sum_{N \in [H,J)_{\vec{X}}} R^X_{H,N} = \Phi_{H,J} - \sum_{N:\, H \leq_{alg} N \lneqq_{\vec{X}} J} R^X_{H,N} \tag{2.5.2}$$

By Claim 2.3.4(3), $\{N \mid H \leq_{alg} N \lneqq_{\vec{x}} J\} = \{N \mid H \leq_{alg} N \lneqq J\}$, and the latter is independent of the basis X. Furthermore, by the induction hypothesis regarding the second claim, so are the terms $R^X_{H,N}$ in this summation. This settles the second point.

Finally, if J is *not* an algebraic extension of H then let L be some intermediate free factor of J, $H \leq L \overset{*}{\lneqq} J$. As mentioned above, this yields that $\Phi_{H,J} = \Phi_{H,L}$. Therefore,

$$R^X_{H,J} = \Phi_{H,J} - \sum_{N \in [H,J)_{\vec{X}}} R^X_{H,N} = \underbrace{\Phi_{H,L} - \sum_{N \in [H,L]_{\vec{X}}} R^X_{H,N}}_{0 \text{ by definition}} - \sum_{N \in [H,J)_{\vec{X}} \setminus [H,L]_{\vec{X}}} R^X_{H,N}$$

By Claim 2.4.5, all algebraic extensions of H inside the interval $[H,J]_{\vec{X}}$ are contained in L. Hence, every subgroup $N \in [H,J)_{\vec{X}} \setminus [H,L]_{\vec{X}}$ is not an algebraic extension of H, and by the induction hypothesis $R^X_{H,N}$ vanishes. The desired result follows. □

In view of Proposition 2.5.1 we can omit the superscript and write from now on $R_{H,J}$ instead of $R^X_{H,J}$. Moreover, we can write the following "basis independent" equation for every pair of f.g. subgroups $H \leq J$:

$$\Phi_{H,J} = \sum_{N:\, H \leq_{alg} N \leq J} R_{H,N}. \tag{2.5.3}$$

When $H \leq_{\vec{x}} J$ this follows from the proof above. For general $H \leq J$, there is some subgroup L such that $H \leq_{\vec{x}} L \overset{*}{\leq} J$ and every intermediate algebraic extension $H \leq_{alg} N \leq J$ is contained in L (see Claims 2.3.4 and 2.4.5). Therefore,

$$\Phi_{H,J} = \Phi_{H,L} = \sum_{N:\, H \leq_{alg} N \leq L} R_{H,N} = \sum_{N:\, H \leq_{alg} N \leq J} R_{H,N}.$$

It turns out that unlike the function R, the other two derivations of Φ, namely L^X and C^X, do depend on the basis X. However, the latter two functions have combinatorial interpretations. In the next section we show that $\Phi_{H,J}$ and $L^X_{H,J}$ can be described in terms of random coverings of the core graph $\Gamma_X(J)$, and that explicit rational expressions in n can be computed to express these two functions for given H, J (Lemmas 2.6.2 and 2.6.3 below). This, in turn, allows us to analyze the combinatorial meaning and order of magnitude of $C^X_{M,N}$ (Proposition 2.7.1).

Finally, using the fact that R is the "left integral" of C^X, that is $R = \zeta^X * C^X$, we finish the circle around the diagram of Φ's derivations, and use this analysis of Φ, L^X and C^X to prove that for every pair $H \leq_{alg} J$, $R_{H,J}$ does not vanish and is, in fact, positive for large enough n. This alone gives Theorem 2.1.4. The more informative 2.1.8 follows from an analysis of the order of magnitude of $R_{H,J}$ in this case (Proposition 2.7.2).

2.6 Random coverings of core graphs

This section studies the graphs which cover a given core-graph in the topological sense, i.e. $\widehat{\Gamma} \overset{p}{\twoheadrightarrow} \Gamma$ with p locally bijective. We call these graphs (together with their projection maps) *coverings* of Γ. The reader should not confuse this with our notion "covers" from Definition 2.3.3.

We focus on directed and edge-labeled coverings. This means we only consider $\widehat{\Gamma} \overset{p}{\twoheadrightarrow} \Gamma$ such that $\widehat{\Gamma}$ is directed and edge-labeled, and the projection p preserves orientations and labels. When Γ is a core-graph we do *not* assume that $\widehat{\Gamma}$ is a core-graph as well. It may be disconnected, and it need not be pointed. Nevertheless, it is not hard to see that when Γ and $\widehat{\Gamma}$ are finite, for every vertex v in $p^{-1}(\otimes)$, the fiber over Γ's basepoint, we do have a valid core-graph, which we denote by $\widehat{\Gamma}_v$: this is the connected component of v in $\widehat{\Gamma}$, with v serving as basepoint. Moreover, the restriction of the projection map p to $\widehat{\Gamma}_v$ is a core-graph morphism.

The theory of core-graph coverings shares many similarities with the theory of topological covering spaces. The following claim lists some standard properties of covering spaces, formulated for core-graphs.

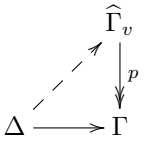
Claim 2.6.1. *Let Γ be a core-graph, $\widehat{\Gamma} \overset{p}{\twoheadrightarrow} \Gamma$ a covering and v a vertex in the fiber $p^{-1}(\otimes)$.*

(1) The group $\pi_1^X(\Gamma)$ acts on the fiber $p^{-1}(\otimes)$, and these actions give a correspondence between coverings of Γ and $\pi_1^X(\Gamma)$-sets.

(2) In this correspondence, coverings of Γ with fiber $\{1,\ldots,n\}$ correspond to actions of $\pi_1^X(\Gamma)$ on $\{1,\ldots,n\}$, i.e., to group homomorphisms $\pi_1^X(\Gamma) \to S_n$.

(3) The group $\pi_1^X\left(\widehat{\Gamma}_v\right)$ is the stabilizer of v in the action of $\pi_1^X(\Gamma)$ on $p^{-1}(\otimes)$ (note that $\pi_1^X\left(\widehat{\Gamma}_v\right)$ and $\pi_1^X(\Gamma)$ are both subgroups of $\mathbf{F}(X)$).

(4) A core-graph morphism $\Delta \to \Gamma$ can be lifted to a core-graph morphism $\Delta \to \widehat{\Gamma}_v$ (i.e., the diagram



can be completed) if and only if $\pi_1^X(\Delta) \subseteq \pi_1^X\left(\widehat{\Gamma}_v\right)$. By the previous point, this is equivalent to saying that all elements of $\pi_1^X(\Delta)$ fix v.

We now turn our attention to random coverings. The vertex set of an n-sheeted covering of a graph $\Gamma = (V, E)$ can be assumed to be $V \times \{1,\ldots,n\}$, so that the fiber above $v \in V$ is $\{v\} \times \{1,\ldots,n\}$. For every edge $e = (u, v) \in E$, the fiber over e then constitutes a perfect matching between $\{v\} \times \{1,\ldots,n\}$ and $\{u\} \times \{1,\ldots,n\}$. This suggests a natural model for random n-coverings of the graph Γ. Namely, for every $e \in E$ choose uniformly a random perfect matching (which is just a permutation in S_n). This model was introduced in [AL02], and is a generalization of a well-known model for random regular graphs (see e.g. [BS87]).[†] Note that the model works equally well for graphs with loops and with multiple edges.

[†]Occasionally these random coverings are referred to as random *lifts* of graphs. We shall reserve this term for its usual meaning.

In fact, there is some redundancy in this model, if we are interested only in isomorphism classes of coverings (two coverings are isomorphic if there is an isomorphism between them that commutes with the projection maps). It is possible to obtain the same distribution on (isomorphism classes of) n-coverings of Γ with fewer random permutations: one may choose some spanning tree T of Γ, associate the identity permutation with every edge in T, and pick random permutations only for edges outside T.

We now fix some $J \leq_{fg} \mathbf{F}_k$, and consider random coverings of its core-graph, $\Gamma_X(J)$. We denote by $\widehat{\Gamma}_X(J)$ a random n-covering of $\Gamma_X(J)$, according to one of the models described above. If $p : \widehat{\Gamma}_X(J) \to \Gamma_X(J)$ is the covering map, then $\widehat{\Gamma}_X(J)$ inherits the edge orientation and labeling from $\Gamma_X(J)$ via p^{-1}. For every i ($1 \leq i \leq n$), we write $\widehat{\Gamma}_X(J)_i$ for the core-graph $\widehat{\Gamma}_X(J)_{(\otimes,i)}$ (the component of $(\otimes, i)$ in $\widehat{\Gamma}_X(J)$ with basepoint $(\otimes, i)$).

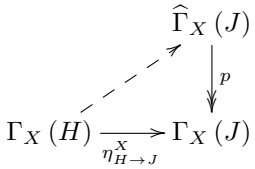
By Claim 2.6.1(2), each random n-covering of $\Gamma_X(J)$ encodes a homomorphism $\alpha_{J,n} \in \mathrm{Hom}(J, S_n)$, via the action of $J = \pi_1^X(\Gamma_X(J))$ on the basepoint fiber. Explicitly, an element $w \in J$ is mapped to a permutation $\alpha_{J,n}(w) \in S_n$ as follows: w corresponds to a closed path p_w around the basepoint of $\Gamma_X(J)$. For every $1 \leq i \leq n$, the lift of p_w that starts at $(\otimes, i)$ ends at $(\otimes, j)$ for some j, and $\alpha_{J,n}(w)(i) = j$.

By the correspondence of actions of J on $\{1, \ldots, n\}$ and n-coverings of $\Gamma_X(J)$, $\alpha_{J,n}$ is a uniform random homomorphism in $\mathrm{Hom}(J, S_n)$. This can also be verified using the "economical" model, as follows: choose some basis $Y = \{y_1, \ldots, y_{\mathrm{rk}(J)}\}$ for J via a choice of a spanning tree T of $\Gamma_X(J)$ and of orientation of the remaining edges, and choose uniformly at random some $\sigma_r \in S_n$ for every basis element y_r. Clearly, $\alpha_{J,n}(y_r) = \sigma_r$.

We can now use the coverings of $\Gamma_X(J)$ to obtain a geometric interpretation of $\Phi_{H,J}$, as follows: let $H \leq J \leq_{fg} \mathbf{F}_k$ and $1 \leq i \leq n$. By 2.6.1(4), the morphism $\eta^X_{H \to J} : \Gamma_X(H) \to \Gamma_X(J)$ lifts to a core-graph morphism $\Gamma_X(H) \to \widehat{\Gamma}_X(J)_i$ iff $H = \pi_1^X(\Gamma_X(H))$ fixes $(\otimes, i)$ via the action of J on the fiber $\otimes \times \{1, \ldots, n\}$. Since this action is given by $\alpha_{J,n}$, this means that $\eta^X_{H \to J}$ lifts to $\widehat{\Gamma}_X(J)_i$ exactly when $\alpha_{J,n}(H)$ fixes i. Recalling that $\Phi_{H,J}(n)$ is the expected number of elements in $\{1, \ldots, n\}$ fixed by $\alpha_{J,n}(H)$, we obtain an alternative definition for it:

Lemma 2.6.2. *Let $\widehat{\Gamma}_X(J)$ be a random n-covering space of $\Gamma_X(J)$ in the aforementioned model from [AL02]. Then,*

$$\Phi_{H,J}(n) = \textit{The expected number of lifts of } \eta^X_{H \to J} \textit{ to } \widehat{\Gamma}_X(J).$$



Note that this characterization of $\Phi_{H,J}$ involves the basis X, although the original definition (2.2.1) does not. One of the corollaries of this lemma is therefore that the average number of lifts does *not* depend on the basis X.

Recall (Section 2.5) the definition of the function L^X, which satisfies $\Phi_{H,J} = \sum_{M \in [H,J]^{\rightarrow}_X} L^X_{M,J}$ for every $H \leq^{\rightarrow}_X J$. It turns out that this derivation of Φ also has a geometrical interpretation. Assume that $\eta^X_{H \to J}$ does lift to $\widehat{\eta}_i : \Gamma_X(H) \to \widehat{\Gamma}_X(J)_i$. By Claim 2.3.4, $\widehat{\eta}_i$ decomposes as a quotient onto $\Gamma_X(M)$, where $M = \pi_1^X(\mathrm{im}\,\widehat{\eta}_i)$, followed by an embedding. Moreover, M lies in $[H, J]^{\rightarrow}_X$. On the other hand, if there is some $M \in [H, J]^{\rightarrow}_X$ such that $\Gamma_X(M)$ is embedded in $\widehat{\Gamma}_X(J)_i$ then such

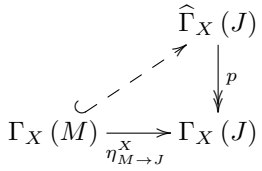
M is unique and $\widehat{\eta}_i$ lifts to the composition of $\eta^X_{H\to M}$ with this embedding. Consequently,

$$\begin{aligned}\Phi_{H,J}(n) &= \text{Expected number of lifts of } \eta^X_{H\to J} \text{ to } \widehat{\Gamma}_X(J)\\ &= \sum_{M\in[H,J]_{\vec{X}}} \text{Expected number of } \textit{injective} \text{ lifts of } \eta^X_{M\to J} \text{ to } \widehat{\Gamma}_X(J).\end{aligned}$$

Taking the left derivations, we obtain:

Lemma 2.6.3. *Let $M \leq_{\vec{x}} J$, and let $\widehat{\Gamma}_X(J)$ be a random n-covering space of $\Gamma_X(J)$ in the aforementioned model from [AL02]. Then,*

$$L^X_{M,J}(n) = \textit{The expected number of } \text{injective } \textit{lifts of } \eta^X_{M\to J} \textit{ to } \widehat{\Gamma}_X(J).$$



Unlike the number of lifts in general, the number of injective lifts does depend on the basis X. For instance, consider $M = \langle x_1x_2\rangle$ and $J = \langle x_1, x_2\rangle = \mathbf{F}_2$. With the basis $X = \{x_1, x_2\}$, the probability that $\eta^X_{M\to J}$ lifts injectively to $\widehat{\Gamma}_X(J)_i$ equals $\frac{n-1}{n^2}$ (Lemma 2.6.4 shows how to compute this). However, with the basis $Y = \{x_1x_2, x_2\}$, the corresponding probability is $\frac{1}{n}$. We also remark that Lemma 2.6.3 allows a natural extension of L^X to pairs M, J such that M does not X-cover J.

Lemma 2.6.3 allows us to generalize the method used in [Nic94, LP10, Pud14] to compute the expected number of fixed points in $\alpha_n(w)$ (see the notations before Theorem 2.1.4'). We claim that for n large enough, $L^X_{M,J}(n)$ is a simple rational expression in n.

Lemma 2.6.4. *Let $M, J \leq_{fg} \mathbf{F}_k$ such that $M \leq_{\vec{x}} J$, and let $\eta = \eta^X_{M\to J}$ be the core-graph morphism. For large enough n,*

$$L^X_{M,J}(n) = \frac{\prod_{v\in V(\Gamma_X(J))} (n)_{|\eta^{-1}(v)|}}{\prod_{e\in E(\Gamma_X(J))} (n)_{|\eta^{-1}(e)|}}, \tag{2.6.1}$$

where $(n)_r$ is the falling factorial $n(n-1)\dots(n-r+1)$, and "large enough n" is $n \geq \max_{e\in E(\Gamma_X(J))} \left|\eta^{-1}(e)\right|$ (so that the denominator does not vanish).

Proof. Let v be a vertex in $\Gamma_X(J)$ and consider the fiber $\eta^{-1}(v)$ in $\Gamma_X(M)$. For every injective lift $\widehat{\eta} : \Gamma_X(M) \hookrightarrow \widehat{\Gamma}_X(J)$, the fiber $\eta^{-1}(v)$ is mapped injectively into the fiber $p^{-1}(v)$. The number of such injections is

$$(n)_{|\eta^{-1}(v)|} = n(n-1)\dots(n-|\eta^{-1}(v)|+1),$$

and therefore the number of injective lifts of $\eta\big|_{V(\Gamma_X(M))}$ into $V\left(\widehat{\Gamma}_X(J)\right)$ is the numerator of (2.6.1).

We claim that any such injective lift has a positive probability of extending to a full lift of η: all one needs is that the fiber above every edge of $\Gamma_X(J)$ satisfy some constraints. To get the exact probability, we return to the more "wasteful" version of the model for a random n-covering of $\Gamma_X(J)$, the model in which we choose a random permutation for every edge of the base graph. Let $\widehat{\eta} : V(\Gamma_X(M)) \hookrightarrow V\left(\widehat{\Gamma}_X(J)\right)$ be an injective lift of the vertices of $\Gamma_X(M)$ as above, and let e be some edge of $\Gamma_X(J)$. If $\widehat{\eta}$ is to be extended to $\eta^{-1}(e)$, the fiber above e in $\widehat{\Gamma}_X(J)$ must contain, for every $(u, v) \in \eta^{-1}(e)$, the edge $(\widehat{\eta}(u), \widehat{\eta}(v))$.

Thus, the random permutation $\sigma \in S_n$ which determines the perfect matching above e in $\widehat{\Gamma}_X(J)$, must satisfy $\left|\eta^{-1}(e)\right|$ non-colliding constraints of the form $\sigma(i) = j$. Whenever $n \geq \left|\eta^{-1}(e)\right|$ (which we assume), a uniformly random permutation in S_n satisfies such constraints with probability

$$\frac{1}{(n)_{|\eta^{-1}(e)|}}.$$

This shows the validity of (2.6.1). □

This immediately gives a formula for $\Phi_{H,J}$ as a rational function:

Corollary 2.6.5. *Let $H, J \leq_{fg} \mathbf{F}_k$ such that $H \leq^{\rightarrow}_{X} J$. Then, for large enough n,*

$$\Phi_{H,J}(n) = \sum_{M \in [H,J]^{\rightarrow}_{X}} L^X_{M,J}(n) = \sum_{M \in [H,J]^{\rightarrow}_{X}} \frac{\prod_{v \in V(\Gamma_X(J))} (n)_{\left|\left(\eta^X_{M \to J}\right)^{-1}(v)\right|}}{\prod_{e \in E(\Gamma_X(J))} (n)_{\left|\left(\eta^X_{M \to J}\right)^{-1}(e)\right|}}.$$

Since H X-covers every intermediate $M \in [H,J]^{\rightarrow}_{X}$, the largest fiber above every edge of $\Gamma_X(J)$ is obtained in $\Gamma_X(H)$ itself. Thus, "large enough n" in this Corollary can be replaced by $n \geq \max_{e \in E(\Gamma_X(J))} \left|\left(\eta^X_{H \to J}\right)^{-1}(e)\right|$.

In fact, Corollary 2.6.5 applies, with slight modifications, to every pair of f.g. subgroups $H \leq J$: Lemma 2.6.2 holds in this more general case, that is $\Phi_{H,J}$ is equal to the expected number of lifts of $\Gamma_X(H)$ to the random n-covering $\widehat{\Gamma}_X(J)$. The image of each lift (with the image of $\otimes$ as basepoint) is a core graph which is a quotient of $\Gamma_X(H)$, and so corresponds to a subgroup M such that $H \leq^{\rightarrow}_{X} M \leq J$. In explaining the rational expression in Lemma 2.6.4 we did not need M to cover J. Thus, for every $H \leq J$, both finitely generated,

$$\Phi_{H,J}(n) = \sum_{M\,:\,H \leq^{\rightarrow}_{X} M \leq J} \frac{\prod_{v \in V(\Gamma_X(J))} (n)_{\left|\left(\eta^X_{M \to J}\right)^{-1}(v)\right|}}{\prod_{e \in E(\Gamma_X(J))} (n)_{\left|\left(\eta^X_{M \to J}\right)^{-1}(e)\right|}}. \tag{2.6.2}$$

Corollary 2.6.5 yields in particular a straight-forward algorithm to obtain a rational expression in n for $\Phi_{H,J}(n)$ (valid for large enough n). For example, consider $H = \left\langle x_1 x_2 x_1^{-1} x_2^{-1} \right\rangle$ and $\mathbf{F}_2 = \langle x_1, x_2 \rangle$. The interval $[H, \mathbf{F}_2]^{\rightarrow}_{X}$ consists of seven subgroups, as depicted in Figure 2.3.4. Following the computation in Corollary 2.6.5, we get that for $n \geq 2$ (we scan the quotients in Figure 2.3.4 top-to-bottom and in each row left-to-right):

$$\begin{aligned}\Phi_{H,\mathbf{F}_2}(n) &= \frac{(n)_4}{(n)_2 (n)_2} + \frac{(n)_2}{(n)_2 (n)_1} + \frac{(n)_2}{(n)_1 (n)_2} \\ &\quad + \frac{(n)_3}{(n)_2 (n)_2} + \frac{(n)_3}{(n)_2 (n)_2} + \frac{(n)_2}{(n)_2 (n)_2} + \frac{(n)_1}{(n)_1 (n)_1} \\ &= \frac{n}{n-1} = 1 + \frac{1}{n} + O\left(\frac{1}{n^2}\right)\end{aligned}$$

This demonstrates Theorem 2.1.8 and Table 2.1 for $H = \left\langle x_1 x_2 x_1^{-1} x_2^{-1} \right\rangle$ (recall the discussion following Lemma 2.4.4, where it is shown that $\pi(H) = 2$ and that $\mathrm{Crit}(H) = \{\mathbf{F}_2\}$).

The explicit computation of Φ yields an effective version of Theorem 2.1.4':

Corollary 2.6.6. *Let $H \leq_{fg} \mathbf{F}_k$, and let ℓ denote the number of edges in $\Gamma_X(H)$. Then $H \overset{*}{\leq} \mathbf{F}_k$ iff $\Phi_{H,\mathbf{F}_k}(n) = n^{-\widetilde{\mathrm{rk}}\,H}$ for $n \leq \ell + \widetilde{\mathrm{rk}}\,H$. In particular, Proposition 2.1.6 follows.*

Proof. Assume that $\Phi_{H,\mathbf{F}_k}(n) = n^{-\widetilde{\mathrm{rk}}\,H}$ holds for $n \le \ell + \widetilde{\mathrm{rk}}\,H$, and denote

$$\Phi'(n) = \sum_{M \in [H,\infty)_{\vec{X}}} \frac{(n)_{|V(\Gamma_X(M))|}}{\prod_{j=1}^{k} (n)_{|E_j(\Gamma_X(M))|}}, \tag{2.6.3}$$

where $E_j(\Gamma)$ are the j-edges in Γ. By Corollary 2.6.5, $\Phi'(n) = \Phi_{H,\mathbf{F}_k}(n)$ for $n \ge n_0 = \max_{j=1..k} |E_j(\Gamma_X(H))|$, and in particular $\Phi'(n) = n^{-\widetilde{\mathrm{rk}}\,H}$ for $n_0 \le n \le \ell + \widetilde{\mathrm{rk}}\,H$. We proceed to show that $\Phi'(n) \equiv n^{-\widetilde{\mathrm{rk}}\,H}$, which implies $\Phi_{H,\mathbf{F}_k}(n) = n^{-\widetilde{\mathrm{rk}}\,H}$ for $n \ge n_0$. The conclusion then follows by Theorem 2.1.4' (which is proved in the next section).

The number of j-edges in every quotient of $\Gamma_X(H)$ is at most $E_j(\Gamma_X(H))$, so that $\Phi'(n)\, g(n)$ is a polynomial for $g(n) = \prod_{j=1}^{k} (n)_{|E_j(\Gamma_X(H))|}$. We would like to establish

$$\Phi'(n)\, g(n)\, n^{\widetilde{\mathrm{rk}}(H)} \equiv g(n), \tag{2.6.4}$$

and we note that $\deg g = \ell$, and $\deg \Phi' \le \max_{M \in [H,\infty)_{\vec{X}}} -\widetilde{\mathrm{rk}}\,(M) \le 0$ follows from Claim 2.3.1(2) (assuming $H \neq id$). Therefore, the degrees of both sides of (2.6.4) are at most $\ell + \widetilde{\mathrm{rk}}\,H$, and it suffices to show they agree at $\ell + \widetilde{\mathrm{rk}}\,H + 1 = \ell + \mathrm{rk}\,H$ points. We already know that they agree for $n_0 \le n \le \ell + \widetilde{\mathrm{rk}}\,H$. For $0 \le n < n_0$ it is clear that $g(n) = 0$. It turns out that the l.h.s. vanishes as well for these values of n. Expanding the l.h.s. gives

$$n^{\widetilde{\mathrm{rk}}\,H} \cdot \sum_{M \in [H,\infty)_{\vec{X}}} (n)_{|V(\Gamma_X(M))|} \prod_{j=1}^{k} \left(n - |E_j(\Gamma_X(M))|\right)_{|E_j(\Gamma_X(H))| - |E_j(\Gamma_X(M))|}, \tag{2.6.5}$$

and each term in the sum vanishes for $0 \le n < n_0$: Choose $1 \le j \le k$ for which $|E_j(\Gamma_X(H))| = n_0$. For each $M \in [H,\infty)_{\vec{X}}$ either $|E_j(\Gamma_X(M))| \le n$, in which case $(n - |E_j(\Gamma_X(M))|)_{n_0 - |E_j(\Gamma_X(M))|} = 0$, or $|E_j(\Gamma_X(M))| > n$; as different j-edges must have different origins, the latter implies that $|V(\Gamma_X(M))| > n$, hence $(n)_{|V(\Gamma_X(M))|}$ vanishes. □

Remark 2.6.7. The discussion in this section suggests a generalization of our analysis to finite groups G other than S_n. For any (finite) faithful G-set S, one can consider a random $|S|$-covering of $\Gamma_X(J)$. The fiber above every edge is chosen according to the action on S of a (uniformly distributed) random element of G. In this more general setting we also get a one-to-one correspondence between $\mathrm{Hom}(\mathbf{F}_k, G)$ and $|S|$-coverings. Although the computation of L^X and of Φ might be more involved, this suggests a way of analyzing words which are measure preserving w.r.t. G.

2.7 The proof of Theorem 2.1.8

The last major ingredient of the proof of our main result, Theorem 2.1.8, is an analysis of C^X, the double-sided derivation of Φ. Recall Definition 2.3.6 where the X-distance $\rho_X(H,J)$ was defined for every $H, J \le_{fg} \mathbf{F}_k$ with $H \le_{\vec{x}} J$.

Proposition 2.7.1. *Let $M, N \le_{fg} \mathbf{F}_k$ satisfy $M \le_{\vec{x}} N$. Then*

$$C^X_{M,N}(n) = O\left(\frac{1}{n^{\widetilde{\mathrm{rk}}(M) + \rho_X(M,N)}}\right)$$

Section 2.7.1 is dedicated to the proof of this proposition. Before getting there, we show how it practically finishes the proof of our main result. We do this with the following final step:

Proposition 2.7.2. *Let $H, N \leq_{fg} \mathbf{F}_k$ satisfy $H \leq_{alg} N$. Then*

$$R_{H,N}(n) = \frac{1}{n^{\widetilde{\mathrm{rk}}(N)}} + O\left(\frac{1}{n^{\widetilde{\mathrm{rk}}(N)+1}}\right)$$

Proof. Let X be some basis of $\mathbf{F}_k$. Recall that $R = \zeta^X * C^X$, i.e.

$$R_{H,N}(n) = \sum_{M \in [H,N]_{\vec{X}}} C^X_{M,N}(n).$$

For $M = N$ we have $C^X_{N,N}(n) = R_{N,N}(n) = \Phi_{N,N}(n) = n^{-\widetilde{\mathrm{rk}}(N)}$ (the last equality follows from the fact that m independent uniform permutations fix a point with probability n^{-m}). For any other M, i.e. $M \in [H,N)_{\vec{X}}$, the fact that N is an algebraic extension of H means that M is *not* a free factor of N and therefore, by Theorem 2.3.8 (and (2.3.6)), $\rho_X(M,N) \geq \widetilde{\mathrm{rk}}(N) - \widetilde{\mathrm{rk}}(M) + 1$. Proposition 2.7.1 then shows that

$$C^X_{M,N}(n) \in O\left(\frac{1}{n^{\widetilde{\mathrm{rk}}(M)+\rho_X(M,N)}}\right) \subseteq O\left(\frac{1}{n^{\widetilde{\mathrm{rk}}(N)+1}}\right).$$

Hence,

$$R_{H,N}(n) = C^X_{N,N}(n) + \sum_{M \in [H,N)_{\vec{X}}} C^X_{M,N}(n) = \frac{1}{n^{\widetilde{\mathrm{rk}}(N)}} + O\left(\frac{1}{n^{\widetilde{\mathrm{rk}}(N)+1}}\right).$$

□

The proof of Theorem 2.1.8 is now at hand. For every $H, J \leq_{fg} \mathbf{F}_k$ with $H \leq J$, by (2.5.3) and Proposition 2.7.2,

$$\begin{aligned}
\Phi_{H,J}(n) &= \sum_{N\,:\,H \leq_{alg} N \leq J} R_{H,N}(n) \\
&= R_{H,H}(n) + \sum_{N\,:\,H \lneq_{alg} N \leq J} R_{H,N}(n) \\
&= \frac{1}{n^{\widetilde{\mathrm{rk}}(H)}} + \sum_{N\,:\,H \lneq_{alg} N \leq J} \frac{1}{n^{\widetilde{\mathrm{rk}}(N)}} + O\left(\frac{1}{n^{\widetilde{\mathrm{rk}}(N)+1}}\right).
\end{aligned}$$

For $J = \mathbf{F}_k$ we can be more concrete. Recall that the H-critical groups, $\mathrm{Crit}(H)$, are the algebraic extensions of H of minimal rank (other than H itself), and this minimal rank is $\pi(H)$. Therefore,

$$\begin{aligned}
\Phi_{H,\mathbf{F}_k}(n) &= \frac{1}{n^{\widetilde{\mathrm{rk}}(H)}} + \sum_{N \in (H,\infty)_{alg}} \frac{1}{n^{\widetilde{\mathrm{rk}}(N)}} + O\left(\frac{1}{n^{\widetilde{\mathrm{rk}}(N)+1}}\right) \\
&= \frac{1}{n^{\widetilde{\mathrm{rk}}(H)}} + \frac{|\mathrm{Crit}(H)|}{n^{\widetilde{\pi}(H)}} + O\left(\frac{1}{n^{\widetilde{\pi}(H)+1}}\right).
\end{aligned}$$

This establishes our main results: Theorem 2.1.8, Theorem 2.1.4 and all their corollaries.

2.7.1 The analysis of $C^X_{M,N}$

In this subsection we look into C^X, the double-sided derivation of Φ, and establish Proposition 2.7.1, which bounds the order of magnitude of $C^X_{M,N}$. Recall that by definition $C^X = L^X * \mu^X$, which is equivalent to

$$L^X_{M,J} = \sum_{N \in [M,J]_{\vec{X}}} C^X_{M,N} \qquad (\forall M \leq_{\vec{x}} J) \tag{2.7.1}$$

We derive a combinatorial meaning of $C^X_{M,N}$ from this relation. To obtain this, we further analyze the rational expression (2.6.1) for $L^X_{M,J}$ and write it as a formal power series. Then, using a combinatorial interpretation of the terms in this series, we attribute each term to some $N \in [M,J]^{\rightarrow}_{X}$, and show that for every $N \in [M,J]^{\rightarrow}_{X}$, the sum of terms attributed to N is nothing but $C^X_{M,N}$. Finally, we use this combinatorial interpretation of $C^X_{M,N}$ to estimate its order of magnitude.

Rewriting $L^X_{M,J}$ as a power series in n^{-1}

Consider the numerator and denominator of (2.6.1): these are products of expressions of the type $(n)_r$. It is a classical fact that

$$(n)_r = \sum_{j=1}^{r} (-1)^{r-j} \begin{bmatrix} r \\ j \end{bmatrix} n^j$$

where $\begin{bmatrix} r \\ j \end{bmatrix}$ is the *unsigned Stirling number of the first kind.* That is, $\begin{bmatrix} r \\ j \end{bmatrix}$ is the number of permutations in S_r with exactly j cycles (see, for instance, [vLW01], Chapter 13).

We introduce the notation $[r]_j \stackrel{def}{=} \begin{bmatrix} r \\ r-j \end{bmatrix}$, which is better suited for our purposes. The cycles of a permutation $\sigma \in S_r$ constitute a partition P_σ of $\{1,\ldots,r\}$. We define $\|\sigma\| = \|P_\sigma\|$ (recall (2.3.4)), and it is immediate that $[r]_j$ is the number of permutations $\sigma \in S_r$ with $\|\sigma\| = j$. It is also easy to see that $\|\sigma\|$ is the minimal number of transpositions needed to be multiplied in order to obtain σ. Therefore, $[r]_j$ is the number of permutations in S_r which can be expressed as a product of j transpositions, but no less. In terms of this notation, we obtain

$$(n)_r = n^r \sum_{j=0}^{r-1} (-1)^j [r]_j \, n^{-j}.$$

The product of several expressions of this form, namely $(n)_{r_1} (n)_{r_2} \ldots (n)_{r_\ell}$, can be written as a polynomial in n whose coefficients have a similar combinatorial meaning, as follows. Let X be a set, and $\varphi : X \to \{1,\ldots,\ell\}$ some function with fibers of sizes $\left|\varphi^{-1}(i)\right| = r_i$ $(1 \le i \le \ell)$. We denote by

$$\mathrm{Sym}_\varphi(X) = \{\sigma \in \mathrm{Sym}(X) \mid \varphi \circ \sigma = \varphi\}$$

the set of permutations $\sigma \in \mathrm{Sym}(X)$ subordinate to the partition of X induced by the fibers of φ, i.e., such that $\varphi(\sigma(x)) = \varphi(x)$ for all $x \in X$. We define

$$[X]^\varphi_j = \left|\{\sigma \in \mathrm{Sym}_\varphi(X) \,:\, \|\sigma\| = j\}\right|,$$

the number of φ-subordinate permutations with $\|\sigma\| = j$. Put differently, $[X]^\varphi_j$ counts the permutations counted in $\left[\,|X|\,\right]_j$ which satisfy, in addition, that every cycle consists of a subset of some fiber of φ. With this new notation, one can write:

$$(n)_{r_1} (n)_{r_2} \ldots (n)_{r_\ell} = \prod_{i=1}^{l} \left(n^{r_i} \sum_{m=0}^{r_i-1} (-1)^m [r_i]_m \, n^{-m} \right) = n^{|X|} \sum_{j=0}^{|X|} (-1)^j [X]^\varphi_j \, n^{-j}$$

Turning back to (2.6.1), we let V_M and E_M denote the sets of vertices and edges, respectively, of $\Gamma_X(M)$. We denote by η the morphism $\eta^X_{M\to J}$, and use it implicitly also for its restrictions to V_M and E_M, which should cause no confusion. We obtain

$$L^X_{M,J}(n) = \frac{n^{|V_M|} \sum_{j=0}^{|V_M|} (-1)^j [V_M]^\eta_j \, n^{-j}}{n^{|E_M|} \sum_{j=0}^{|E_M|} (-1)^j [E_M]^\eta_j \, n^{-j}},$$

which by Claim 2.3.1(2) equals

$$L^X_{M,J}(n) = n^{-\widetilde{\mathrm{rk}}(M)} \frac{\sum_{j=0}^{|V_M|} (-1)^j [V_M]^\eta_j n^{-j}}{\sum_{j=0}^{|E_M|} (-1)^j [E_M]^\eta_j n^{-j}}. \tag{2.7.2}$$

Consider the denominator of (2.7.2) as a power series $Q\left(n^{-1}\right)$. Its free coefficient is $[E_M]^\eta_0 = 1$. This makes it relatively easy to get a formula for its inverse $1/Q\left(n^{-1}\right)$ as a power series. In general, if $Q(x) = 1 + \sum_{i=1}^\infty a_i x^i$, then

$$\frac{1}{Q(x)} = \frac{1}{1 - \sum_{i=1}^\infty (-a_i) x^i} = \sum_{t=0}^\infty \left(\sum_{i=1}^\infty (-a_i) x^i \right)^t = \\ = \sum_{t=0}^\infty \sum_{j_1, j_2, \ldots, j_t \geq 1} (-1)^t a_{j_1} \cdot \ldots \cdot a_{j_t} x^{\sum_{i=1}^t j_i}.$$

In the denominator of (2.7.2) we have $a_i = (-1)^i [E_M]^\eta_i$, and the resulting expression needs to be multiplied with the numerator $\sum_{j=0}^{|V_M|} (-1)^j [V_M]^\eta_j n^{-j}$. In total, we obtain

$$L^X_{M,J}(n) = \\ \sum_{t=0}^\infty \sum_{\substack{j_0 \geq 0 \\ j_1, \ldots, j_t \geq 1}} (-1)^{t + \sum_{i=0}^t j_i} [V_M]^\eta_{j_0} \cdot [E_M]^\eta_{j_1} \cdot \ldots \cdot [E_M]^\eta_{j_t} n^{-\widetilde{\mathrm{rk}}(M) - \sum_{i=0}^t j_i}. \tag{2.7.3}$$

The combinatorial meaning and order of magnitude of $C^X_{M,N}$

The expression (2.7.3) is a bit complicated, but it presents $L^X_{M,J}(n)$ as a sum (with coefficients $\pm n^{-s}$) of terms with a combinatorial interpretation: the term $[V_M]^\eta_{j_0} \cdot [E_M]^\eta_{j_1} \cdot \ldots \cdot [E_M]^\eta_{j_t}$ counts $(t+1)$-tuples of η-subordinate permutations. The crux of the matter is that this interpretation allows us to attribute each tuple to a specific subgroup $N \in [M, J]^{\vec{}}_X$. This is done as follows.

Let $(\sigma_0, \sigma_1, \ldots, \sigma_t)$ be a $(t+1)$-tuple of permutations such that $\sigma_0 \in \mathrm{Sym}_\eta(V_M)$ and $\sigma_1, \ldots, \sigma_t \in \mathrm{Sym}_\eta(E_M) \setminus \{\mathrm{id}\}$ (we exclude $\mathrm{id} \in \mathrm{Sym}(E_M)$, which is the only permutation counted in $[E_M]^\eta_0$). Consider the graph $\Gamma = {}^{\Gamma_X(M)}\!/_{\langle \sigma_0, \ldots, \sigma_t \rangle}$, which is the quotient of $\Gamma_X(M)$ by all identifications of pairs of the form $v, \sigma_0(v)$ $(v \in V_M)$ and $e, \sigma_i(e)$ $(e \in E_M,\ 1 \leq i \leq t)$[†]. Since Γ is obtained from $\Gamma_X(M)$ by identification of elements with the same η-image, η induces a well defined morphism $\Gamma \to \Gamma_X(J)$. Thus, every closed path in Γ projects to a path in $\Gamma_X(J)$, giving $\pi_1^X(\Gamma) \leq \pi_1^X(\Gamma_X(J)) = J$. We denote $N = N_{\sigma_0, \sigma_1, \ldots, \sigma_t} = \pi_1^X(\Gamma)$. As usual (see Figures 2.3.2, 2.3.3), we can perform Stallings foldings on Γ until we obtain the core graph corresponding to N, $\Gamma_X(N)$. Obviously we have $M \leq^{\vec{}}_X N$, and by Claim 2.3.4(3) also $N \leq^{\vec{}}_X J$. Thus, we always have $N = N_{\sigma_0, \sigma_1, \ldots, \sigma_t} \in [M, J]^{\vec{}}_X$. To summarize the situation:

$$\Gamma_X(M) \twoheadrightarrow \Gamma = {}^{\Gamma_X(M)}\!/_{\langle \sigma_0, \ldots, \sigma_1 \rangle} \xrightarrow{folding} \Gamma_X(N) \xrightarrow{\eta^X_{N \to J}} \Gamma_X(J) \\ \eta^X_{M \to N} \tag{2.7.4}$$

[†]For the definition of the quotient of a graph by identifications of vertices see the discussion preceding Figure 2.3.3. Although we did not deal with merging of edges before, this is very similar to merging vertices. Identifying a pair of edges means identifying the pair of origins, the pair of termini and the pair of edges. In terms of the generated core graph (see Section 2.3), identifying a pair of edges is equivalent to identifying the pair of origins and\or the pair of termini.

Our next move is to rearrange (2.7.3) according to the intermediate subgroups $N \in [M,J]_{\vec{X}}$ which correspond to the tuples counted in it. For any $N \in [M,J]_{\vec{X}}$ we denote by $\mathcal{T}^X_{M,N,J}$ the set of tuples $(\sigma_0,\sigma_1,\ldots,\sigma_t)$ such that $N_{\sigma_0,\sigma_1,\ldots,\sigma_t} = N$, i.e.

$$\mathcal{T}^X_{M,N,J} = \left\{ (\sigma_0,\sigma_1,\ldots,\sigma_t) \left| \begin{array}{c} t\in\mathbb{N},\ \sigma_0 \in \mathrm{Sym}_\eta(V_M) \\ \sigma_1,\ldots,\sigma_t \in \mathrm{Sym}_\eta(E_M)\setminus\{\mathrm{id}\} \\ \pi_1^X\left(\Gamma_X(M)/\langle\sigma_0,\sigma_1,\ldots,\sigma_t\rangle\right) = N \end{array} \right. \right\}.$$

The terms in (2.7.3) which correspond to a fixed $N \in [M,J]_{\vec{X}}$ thus sum to

$$\widetilde{C}^X_{M,J}(N) = \sum_{(\sigma_0,\sigma_1,\ldots,\sigma_t)\in\mathcal{T}^X_{M,N,J}} \frac{(-1)^{t+\sum_{i=0}^t \|\sigma_i\|}}{n^{\widetilde{\mathrm{rk}}(M)+\sum_{i=0}^t\|\sigma_i\|}}, \tag{2.7.5}$$

and (2.7.3) becomes

$$L^X_{M,J} = \sum_{N\in[M,J]_{\vec{X}}} \widetilde{C}^X_{M,J}(N) \tag{2.7.6}$$

The equation (2.7.6) looks much like (2.7.1), with $\widetilde{C}^X_{M,J}(N)$ playing the role of $C^X_{M,N}$. In order to establish equality between the latter two, we must show that $\widetilde{C}^X_{M,J}(N)$ does not depend on J. Fortunately, this is not hard: it turns out that

$$\widetilde{C}^X_{M,J}(N) = \widetilde{C}^X_{M,N}(N) \qquad \left(\forall N\in[M,J]_{\vec{X}}\right), \tag{2.7.7}$$

and the r.h.s. is, of course, independent of J. This equality follows from $\mathcal{T}^X_{M,N,J} = \mathcal{T}^X_{M,N,N}$, which we now justify. The only appearance J makes in the definition of $\mathcal{T}^X_{M,N,J}$ is inside $\eta = \eta^X_{M\to J}$, which is to be σ_i-invariant (for $0\le i\le n$), i.e., σ_i must satisfy $\eta^X_{M\to J}\circ\sigma_i = \eta^X_{M\to J}$. If $(\sigma_0,\ldots,\sigma_t)\in\mathcal{T}^X_{M,N,J}$ then $\eta^X_{M\to N}\circ\sigma_i = \eta^X_{M\to N}$ follows from the fact that $\Gamma_X(N)$ is a quotient of $\Gamma_X(M)/\langle\sigma_i\rangle$. On the other hand, if $(\sigma_0,\ldots,\sigma_t)\in\mathcal{T}^X_{M,N,N}$ then we have $\eta^X_{M\to N}\circ\sigma_i = \eta^X_{M\to N}$, hence also (see (2.7.4))

$$\eta^X_{M\to J}\circ\sigma_i = \eta^X_{N\to J}\circ\eta^X_{M\to N}\circ\sigma_i = \eta^X_{N\to J}\circ\eta^X_{M\to N} = \eta^X_{M\to J}.$$

Writing $\widetilde{C}^X_{M,N} \overset{def}{=} \widetilde{C}^X_{M,N}(N)$, we have by (2.7.1), (2.7.6), and (2.7.7)

$$C^X * \zeta^X = L^X = \widetilde{C}^X * \zeta^X$$

which shows that $C^X = \widetilde{C}^X$, as desired.

We approach the endgame. Let $(\sigma_0,\sigma_1,\ldots,\sigma_t)\in\mathcal{T}^X_{M,N,J} = \mathcal{T}^X_{M,N,N}$, and consider the partition P of $V(\Gamma_X(H))$, obtained by identifying v and v' whenever $\sigma_0(v) = v'$, or $\sigma_i(e) = e'$ for some $1\le i\le t$ and edges e,e' whose origins are v and v', respectively. Since P can clearly be obtained by $\sum_{i=0}^t\|\sigma_i\|$ identifications, we have $\|P\|\le\sum_{i=0}^t\|\sigma_i\|$ (a strong inequality can take place - for example, one can have $\sigma_1=\sigma_2$). Since $(\sigma_0,\sigma_1,\ldots,\sigma_t)\in\mathcal{T}^X_{M,N,J}$ we have $\pi_1^X\left(\Gamma_X(H)/P\right) = N$, and thus by (2.3.5) we obtain

$$\rho_X(H,J) \le \|P\| \le \sum_{i=0}^t \|\sigma_i\|.$$

From (2.7.5) (recall that $\widetilde{C}^X_{M,J}(N) = \widetilde{C}^X_{M,N} = C^X_{M,N}$) we now have

$$C^X_{M,N}(n) = O\left(\frac{1}{n^{\widetilde{\mathrm{rk}}(M)+\rho_X(M,N)}}\right),$$

and Proposition 2.7.1 is proven.

2.8 Primitive words in the profinite topology

Theorem 2.1.4 has some interesting implications to the study of profinite groups. In fact, some of the original interest in the conjecture that is proven in this paper stems from these implications.

Let $\widehat{\mathbf{F}_k}$ denote the profinite completion of the free group $\mathbf{F}_k$. A basis of $\widehat{\mathbf{F}_k}$ is a set $S \subset \widehat{\mathbf{F}_k}$ such that every map from S to a profinite group G admits a unique extension to a continuous homomorphism $\widehat{\mathbf{F}_k} \to G$. It is a standard fact that $\mathbf{F}_k$ is embedded in $\widehat{\mathbf{F}_k}$, and that every basis of $\mathbf{F}_k$ is also a basis of $\widehat{\mathbf{F}_k}$ (see for example [Wil98]). An element of $\widehat{\mathbf{F}_k}$ is called *primitive* if it belongs to a basis of $\widehat{\mathbf{F}_k}$.

It is natural to ask whether an element of $\mathbf{F}_k$, which is primitive in $\widehat{\mathbf{F}_k}$, is already primitive in $\mathbf{F}_k$. In fact, this was conjectured by Gelander and by Lubotzky, independently. Theorem 2.1.4 yields a positive answer, as follows. An element $w \in \widehat{\mathbf{F}_k}$ is said to be *measure preserving* if for any finite group G, and a uniformly distributed random (continuous) homomorphism $\hat{\alpha}_G \in \mathrm{Hom}_{cont}\left(\widehat{\mathbf{F}_k}, G\right)$, the image $\hat{\alpha}_G(w)$ is uniformly distributed in G. By the natural correspondence $\mathrm{Hom}_{cont}\left(\widehat{\mathbf{F}_k}, G\right) \cong \mathrm{Hom}\left(\mathbf{F}_k, G\right)$, an element of $\mathbf{F}_k$ is measure preserving w.r.t. $\mathbf{F}_k$ iff it is so w.r.t. $\widehat{\mathbf{F}_k}$. As in $\mathbf{F}_k$, a primitive element of $\widehat{\mathbf{F}_k}$ is easily seen to be measure preserving. Theorem 2.1.4 therefore implies that if $w \in \mathbf{F}_k$ is primitive in $\widehat{\mathbf{F}_k}$, then it is also primitive in $\mathbf{F}_k$. In other words:

Corollary 2.8.1. *Let P denote the set of primitive elements of $\mathbf{F}_k$, and let $\widehat{P}$ denote the set of primitive elements of $\widehat{\mathbf{F}_k}$. Then*

$$P = \widehat{P} \cap \mathbf{F}_k.$$

As $\widehat{P}$ is a closed set in $\widehat{\mathbf{F}_k}$, this immediately implies Corollary 2.1.5, which states that P is closed in the profinite topology. In fact, there is also a direct proof to Corollary 2.1.5 from Theorem 2.1.8: one has to find, for every non-primitive word $w \in \mathbf{F}_k$, some $H \leq_{\text{f.i.}} \mathbf{F}_k$ such that the coset wH contains no primitives. By Theorem 2.1.8 there exists n so that w does not induce uniform distribution on S_n. For this n, let

$$H = \bigcap_{\alpha: \mathbf{F}_k \to S_n} \ker \alpha$$

and then wH is a primitive-free coset (as all words in the same coset of H induce the exact same measure on S_n).

This circle of ideas has a natural generalization. Observe the following five equivalence relations on the elements of $\mathbf{F}_k$:

- $w_1 \overset{A}{\sim} w_2$ if w_1 and w_2 belong to the same $\mathrm{Aut}\,\mathbf{F}_k$-orbit.
- $w_1 \overset{B}{\sim} w_2$ if w_1 and w_2 belong to the same $\overline{\mathrm{Aut}\,\mathbf{F}_k}$-orbit (where $\overline{\mathrm{Aut}\,\mathbf{F}_k}$ is the closure of $\mathrm{Aut}\,\mathbf{F}_k$ in $\mathrm{Aut}\,\widehat{\mathbf{F}_k}$).
- $w_1 \overset{C}{\sim} w_2$ if w_1 and w_2 belong to the same $\mathrm{Aut}\,\widehat{\mathbf{F}_k}$-orbit.
- $w_1 \overset{C'}{\sim} w_2$ if w_1 and w_2 have the same "statistical" properties, namely if they induce the same distribution on any finite group.
- $w_1 \overset{C''}{\sim} w_2$ if the evaluation maps $ev_{w_1}, ev_{w_2} : \mathrm{Epi}\left(F_k, G\right) \to G$ have the same images for every finite group G.

It is not hard to see that $(A) \Rightarrow (B) \Rightarrow (C) \Rightarrow (C') \Rightarrow (C'')$ (namely, that if $w_1 \overset{A}{\sim} w_2$ then $w_1 \overset{B}{\sim} w_2$, and so on). The only nontrivial implication is $(C') \Rightarrow (C'')$, which can be shown by induction on the size of G. In an unpublished manuscript, C. Meiri gave a one-page proof that (C), (C') and (C'') in fact coincide (in fact, these three coincide for all elements of $\widehat{\mathbf{F}_k}$).

From this perspective, our main result shows that in the case that w_1 is primitive, all five relations coincide, and it is natural to conjecture that they in fact coincide for all elements in $\mathbf{F}_k$[†]. Showing that $(A) \Leftarrow (B)$ would imply that $\mathrm{Aut}\,\mathbf{F}_k$-orbits in $\mathbf{F}_k$ are closed in the profinite topology, and the stronger statement $(A) \Leftarrow (C)$ would imply that words which lie in different $\mathrm{Aut}\,\mathbf{F}_k$-orbits can be told apart using statistical methods.

The analysis which is carried out in this paper does not suffice for the general case. For example, consider the words $w_1 = x_1x_2x_1x_2^{-1}$ and $w_2 = x_1x_2x_1^{-1}x_2^{-1}$. They belong to different $\mathrm{Aut}\,\mathbf{F}_2$-orbits, as $w_2 \in \mathbf{F}_2'$ but $w_1 \notin \mathbf{F}_2'$, but induce the same distribution on S_n for every n: their images under a random homomorphism are a product of a random permutation (σ) and a random element in its conjugacy class ($\tau\sigma\tau^{-1}$ for w_1, and $\tau\sigma^{-1}\tau^{-1}$ for w_2). However, while S_n do not distinguish between these two words, other groups do (in fact these words induce the same distribution on G precisely when every element in G is conjugate to its inverse, see [PS13] for a discussion of this).

These questions also play a role in the theory of decidability in infinite groups. A natural extension of the word-problem and the conjugacy-problem, is the following *automorphism-problem*: given a group G generated by S, and two words $w_1, w_2 \in F(S)$, can it be decided whether w_1 and w_2 belong to the same $\mathrm{Aut}\,G$-orbit in G? Whitehead's algorithm [Whi36a, Whi36b] gives a concrete solution when $G = \mathbf{F}_k$. Showing that $(A) \Leftarrow (B)$ would provide an alternative decision procedure for $\mathbf{F}_k$.

More generally, and in a similar fashion to the conjugacy problem, it can be shown that if

(1) G is finitely presented

(2) $\mathrm{Aut}\,G$ is finitely generated

(3) $\mathrm{Aut}\,G$-orbits are closed in the profinite topology

then the automorphism-problem in G is decidable. For the free group (1) and (2) are known, and (3) is exactly the conjectured coincidence $(A) \Leftrightarrow (B)$.

2.9 Open questions

We mention some open problems that naturally arise from the discussion in this paper.

- Section 2.8 shows how the questions about primitive elements can be extended to all $\mathrm{Aut}\,\mathbf{F}_k$-orbits in $\mathbf{F}_k$ (is it true that $(A) \Leftrightarrow (B)$, and even the stronger equivalence $(A) \Leftrightarrow (C)$?). More generally, can statistical properties tell apart two subgroups $H_1, H_2 \leq_{fg} \mathbf{F}_k$ which belong to distinct $\mathrm{Aut}\,\mathbf{F}_k$-orbits? This would be a further generalization of Theorem 2.1.4.

- It is also interesting to consider words which are measure preserving w.r.t. other types of groups. For instance, does Theorem 2.1.4 still hold if we replace "finite groups" by "compact Lie groups", and study Haar-measure preserving words? Is there a single compact Lie group which suffices? Within finite groups, we showed that measure preservation w.r.t. S_n implies primitivity. Is it still true if we replace S_n by some other infinite family of finite groups (e.g. $\mathrm{PSL}_n(q)$, or solvable groups)?

- Is it true that
$$[H,\infty)_{\leq} = \bigcup_{\substack{\boldsymbol{X} \text{ is a} \\ \text{basis of } \mathbf{F}_k}} [H,\infty)_{\vec{X}}$$
and under which assumptions does the following hold
$$[H,\infty)_{alg} = \bigcap_{\substack{\boldsymbol{X} \text{ is a} \\ \text{basis of } \mathbf{F}_k}} [H,\infty)_{\vec{X}}$$

[†]In [AV11], for example, the authors indeed ask whether $(C') \Rightarrow (A)$.

(see Remark 2.4.3)?

- The distribution induced by w on a finite group G is a class function, and so is a linear combination of the characters of G (for more on this point of view e.g. [AV11, PS13]). In particular, $\Phi_{\langle w\rangle,\mathbf{F}_k}(n)-1$ is the coefficient of the *standard* character of S_n. The first nonzero term of $\Phi_{\langle w\rangle,\mathbf{F}_k}-1$ encodes the primitivity rank and number of critical subgroups of w. Can the next terms be given an algebraic interpretation, and can they be estimated? (Such an estimation may contribute further to the study of expansion in graphs, which started in [Pud15a].) What about the coefficients of other characters of S_n or of any other (family of) groups?

Acknowledgments

It is a pleasure to thank our advisors Nati Linial and Alex Lubotzky for their support, encouragement and useful comments. We are also grateful to Aner Shalev for supporting this research and for his valuable suggestions. We would also like to thank Uri Bader, Tsachik Gelander, Chen Meiri, Paul Nelson and Iddo Samet for their beneficial comments. We have benefited much from the mathematical open source community, and in particular from GAP [GAP13], and its free group algorithms package [Sie12].

Glossary

		Reference	Remarks
$H\leq_{fg}\mathbf{F}_k$	finitely generated		
$H\overset{*}{\leq}J$	free factor		
$H\leq_{alg}J$	algebraic extension	Definition 2.2.1	
$H\leq_{\vec{x}}J$	H X-covers J	Definition 2.3.3	$H\overset{X}{\twoheadrightarrow}J$ in [Pud14]
$\mathfrak{sub}_{fg}(\mathbf{F}_k)$	the set of finitely generated subgroups of $\mathbf{F}_k$		
$[H,J]_{\preceq}$	$\{L\mid H\preceq L\preceq J\}$		$\preceq$ is either one of $\leq,\overset{*}{\leq},\leq_{alg}$ or $\underset{\vec{x}}{}$ (standing for $\leq_{\vec{x}}$)
$[H,J)_{\preceq}$	$\{L\mid H\preceq L\precneqq J\}$		
$[H,\infty)_{\preceq}$	$\{L\mid H\preceq L\}$		
$[H,\infty)_{\vec{x}}$	the X-quotients of H		$\mathcal{O}_X(H)$, or X-frigne in [MVW07]
$[H,\infty)_{alg}$	algebraic extensions of H		$AE(H)$ in [MVW07]
$\pi(H)$	primitivity rank of H	Definition 2.1.7	$\widetilde{\pi}(H)=\pi(H)-1$
$\mathrm{Crit}(H)$	H-critical groups		
$\Gamma_X(H)$	X-labeled core graph of H		
$\rho_X(H,J)$	X-distance	Definition 2.3.6	$H\leq_{\vec{x}}J$
$\eta^X_{H\to J}$	the morphism $\Gamma_X(H)\to\Gamma_X(J)$	Claim 2.3.2	$H\leq J$
$\alpha_{J,n}$	a uniformly chosen random homomorphism in $\mathrm{Hom}(J,S_n)$		$J\leq_{fg}\mathbf{F}_k$
$\Phi_{H,J}(n)$	the expected number of common fixed points of $\alpha_{J,n}(H)$	(2.2.1)	$H\leq J$

References

[Abe06] M. Abert. On the probability of satisfying a word in a group. *Journal of Group Theory*, 9:685–694, 2006.

[AL02] A. Amit and N. Linial. Random graph coverings I: General theory and graph connectivity. *Combinatorica*, 22(1):1–18, 2002.

[Alo86] N. Alon. Eigenvalues and expanders. *Combinatorica*, 6(2):83–96, 1986.

[AV11] A. Amit and U. Vishne. Characters and solutions to equations in finite groups. *Journal of Algebra and Its Applications*, 10(4):675–686, 2011.

[BK13] T. Bandman and B. Kunyavskii. Criteria for equidistribution of solutions of word equations on $SL(2)$. *Journal of Algebra*, 382:282–302, 2013.

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

[Fri03] J. Friedman. Relative expanders or weakly relatively ramanujan graphs. *Duke Mathematical Journal*, 118(1):19–35, 2003.

[Fri08] J. Friedman. *A proof of Alon's second eigenvalue conjecture and related problems*, volume 195 of *Memoirs of the AMS*. AMS, september 2008.

[GAP13] The GAP Group. *GAP – Groups, Algorithms, and Programming, Version 4.6.5*, 2013.

[GS09] S. Garion and A. Shalev. Commutator maps, measure preservation, and T-systems. *Trans. Amer. Math. Soc.*, 361(9):4631–4651, 2009.

[KM02] I. Kapovich and A. Myasnikov. Stallings foldings and subgroups of free groups. *Journal of Algebra*, 248(2):608–668, 2002.

[LP10] N. Linial and D. Puder. Words maps and spectra of random graph lifts. *Random Structures and Algorithms*, 37(1):100–135, 2010.

[LS08] M. Larsen and A. Shalev. Characters of symmetric groups: sharp bounds and applications. *Inventiones mathematicae*, 174(3):645–687, 2008.

[LS09] M. Larsen and A. Shalev. Words maps and Waring type problems. *J. Amer. Math. Soc.*, 22(2):437–466, 2009.

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

[Nic94] A. Nica. On the number of cycles of given length of a free word in several random permutations. *Random Structures and Algorithms*, 5(5):703–730, 1994.

[PP14] O. Parzanchevski and D. Puder. Stallings graphs, algebraic extensions and primitive elements in F_2. *Mathematical Proceedings of the Cambridge Philosophical Society*, 157(1):1–11, 2014.

[PS13] O. Parzanchevski and G. Schul. On the Fourier expansion of word maps. *Bull. London Math. Soc.*, 2013. doi:10.1112/blms/bdt068.

[Pud14] D. Puder. Primitive words, free factors and measure preservation. *Israel Journal of Mathematics*, 201(1):25–73, 2014.

[Pud15a] D. Puder. Expansion of random graphs: New proofs, new results. *Inventiones Mathematicae*, 2015. to appear. arXiv:1212.5216.

[Seg09] D. Segal. *Words: notes on verbal width in Groups*. London Mathematical Society, Lecture note Series 361, Cambridge University Press, Cambridge, 2009.

[Sha09] A. Shalev. Words maps, conjugacy classes, and a non-commutative Waring-type theorem. *Annals of Math.*, 170:1383–1416, 2009.

[Sha13] A. Shalev. Some results and problems in the theory of word maps. In L. Lovász, I. Ruzsa, V.T. Sós, and D. Palvolgyi, editors, *Erdős Centennial (Bolyai Society Mathematical Studies)*. Springer, 2013.

[Sie12] C. Sievers. *Free Group Algorithms – a GAP package, Version 1.2.0*, 2012.

[Sta83] J.R. Stallings. Topology of finite graphs. *Inventiones mathematicae*, 71(3):551–565, 1983.

[Sta97] R.P. Stanley. *Enumerative Combinatorics*, volume 1 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 1997.

[Tak51] M. Takahasi. Note on chain conditions in free groups. *Osaka Math. J*, 3(2):221–225, 1951.

[Tur96] E.C. Turner. Test words for automorphisms of free groups. *Bulletin of the London Mathematical Society*, 28(3):255–263, 1996.

[vLW01] J.H. van Lint and R.M. Wilson. *A course in combinatorics*. Cambridge Univ Pr, 2001.

[Whi36a] J.H.C. Whitehead. On certain sets of elements in a free group. *Proc. London Math. Soc.*, 41:48–56, 1936.

[Whi36b] J.H.C. Whitehead. On equivalent sets of elements in a free group. *Ann. of Math.*, 37:768–800, 1936.

[Wil98] J.S. Wilson. *Profinite Groups*. Clarendon Press, Oxford, 1998.

Chapter 3

Expansion of Random Graphs: New Proofs, New Results

Doron Puder[†]

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

Published in an electronic format: Inventiones Mathematicae, posted on 2014, DOI: 10.1007/s00222-014-0560-x

Abstract

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

Let Γ be a random d-regular graph on n vertices, and let λ be the largest absolute value of a non-trivial eigenvalue of its adjacency matrix. It was conjectured by Alon [Alo86] that a random d-regular graph is "almost Ramanujan", in the following sense: for every $\varepsilon > 0$, 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 Ω, we study the spectrum of Γ, a random n-covering of Ω. Let λ be the largest absolute value of a non-trivial eigenvalue of Γ. Extending Alon's conjecture to this more general model, Friedman [Fri03] conjectured that for every $\varepsilon > 0$, a.a.s. $\lambda < \rho + \varepsilon$, where ρ is the spectral radius of the universal cover of Ω. When Ω is regular we get the same bound as before: $\rho + 1$, and for an arbitrary Ω, we prove a nearly optimal upper bound of $\sqrt{3}\rho$. This is a substantial improvement upon all known results (by Friedman, Linial-Puder, Lubetzky-Sudakov-Vu

[†]Supported by Adams Fellowship Program of the Israel Academy of Sciences and Humanities, and an Advanced ERC Grant.

and Addario-Berry-Griffiths).

Contents

3.1 Introduction

Random d-regular graphs

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

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

The eigenvalue $\lambda_1 = d$ corresponds to constant functions and is considered the *trivial* eigenvalue of Γ. Let $\lambda(\Gamma)$ be the largest absolute value of a non-trivial eigenvalue of Γ, i.e. $\lambda(\Gamma) = \max\{\lambda_2, -\lambda_n\}$. $\lambda(\Gamma)$ This value measures the spectral expansion of the graph: the smaller $\lambda(\Gamma)$ is, the better expander Γ is (see Appendix 3.B for details).

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

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

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

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

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

Theorem 3.1.1. *Fix $d \geq 3$ and let Γ be a random d-regular simple graph on n vertices chosen at uniform distribution. Then*

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

asymptotically almost surely[†].

For d even, or d odd large enough, we obtain a better bound of $2\sqrt{d-1}+0.84$. The same result, for d even, holds also for random d-regular graphs in the *permutation model* (see below). In fact, we first prove the result stated in Theorem 3.1.1 for random graphs in this model (with d even). The derivation of Theorem 3.1.1 for the uniform model and d even is then immediate by results of Wormald [Wor99] and Greenhill et al. [GJKW02] showing the *contiguity*[‡] of different models of random regular graphs (see Appendix 3.A). Finally, we derive the case of odd d relying on the even case and a contiguity argument in which we lose some in the constant and get 1 instead of 0.84 (Section 3.6.2).

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

We stress that even after Alon's conjecture is established, many open questions remain concerning $\lambda(\Gamma)$. In fact, very little is known about the distribution of $\lambda(\Gamma)$. A major open question is the following: what is the probability that a random d-regular graph is *Ramanujan*, i.e. that $\lambda(\Gamma) \leq 2\sqrt{d-1}$? There are contradicting experimental pieces of evidence (in [MNS08] it is conjectured that this probability tends to 27% as n grows; simulations in [HLW06, Section 7] suggest it may be larger than 50%) . However, even the following, much weaker conjecture 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 $\left[-2\sqrt{d-1}, 2\sqrt{d-1}\right]$). Still, the original problem remains open. We hope our new approach may eventually contribute to answering these open questions.

Random coverings of a fixed base graph

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

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

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

It is therefore natural to measure the spectrum of any graph Γ against the spectral radius of its covering tree. Several authors call graphs whose non-trivial spectrum is bounded by this

[†]For small d's a better bounds are attainable - see the table in Section 3.6.2.

[‡]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 $\sup\{|\lambda| \mid \lambda \in \operatorname{Spec} M\}$.

value *Ramanujan*, generalizing the regular case. Many of the results and questions regarding the spectrum of d-regular graphs extend to this general case. For example, an analogue of Alon-Boppana's Theorem is given in Proposition 3.1.2.

Ideally, one would like to extend Alon's conjecture on nearly-Ramanujan graphs to every infinite tree T with finite quotients, and show that most of its quotients are nearly Ramanujan. However, as shown in [LN98], there exist trees T with some minimal finite quotient Ω which is not Ramanujan. All other finite quotients of T are then coverings of Ω, and inherit the 'bad' eigenvalues of this quotient (we elaborate a bit more in Appendix 3.A) . Such examples invalidate the obvious analogue of Alon's conjecture.

But what if we ignore this few, fixed, 'bad' eigenvalues originated in the minimal quotient Ω and focus only on the remaining, 'new' eigenvalues of each larger quotient? In this sense, a generalized version of Alon's conjecture is indeed plausible. Instead of studying the spectrum of a random finite quotient of T, one may consider the spectrum of a random finite covering of a fixed finite graph. This is the content of the generalized Conjecture of Friedman appearing here as Conjecture 3.1.3.

In order to describe this conjecture precisely, let us first describe the random model we consider. This is a generalization of the permutation model for random regular graphs, which generates families of graphs with a common universal covering tree. A random graph Γ in the permutation model $\mathcal{P}_{n,d}$ can be equivalently thought of as a random n-sheeted covering space of the bouquet with $\frac{d}{2}$ loops. In a similar fashion, fix a finite, connected base graph Ω, and let Γ be a random n-covering space of Ω. More specifically, Γ is sampled as follows: its set of vertices is $V(\Omega) \times [n]$. A permutation $\sigma_e \in S_n$ is then chosen uniformly and independently at random for every edge $e = (u, v)$ of Ω, and for every $i \in [n]$ the edge $((u, i), (v, \sigma_e(i)))$ is introduced in Γ[†]. We denote this model by $\mathcal{C}_{n,\Omega}$ (so that $\mathcal{C}_{n,B_{\frac{d}{2}}} = \mathcal{P}_{n,d}$, where $B_{\frac{d}{2}}$ is the bouquet with $\frac{d}{2}$ loops). For example, all bipartite d-regular graphs on $2n$ vertices cover the graph •≡• with two vertices and d edges connecting them. Various properties of random graphs in the $\mathcal{C}_{n,\Omega}$ model were thoroughly examined over the last decade (e.g. [AL02, ALM02, Fri03, LR05, AL06, BL06, LP10]). From now on, by a "random n-covering of Ω" we shall mean a random graph in the model $\mathcal{C}_{n,\Omega}$.

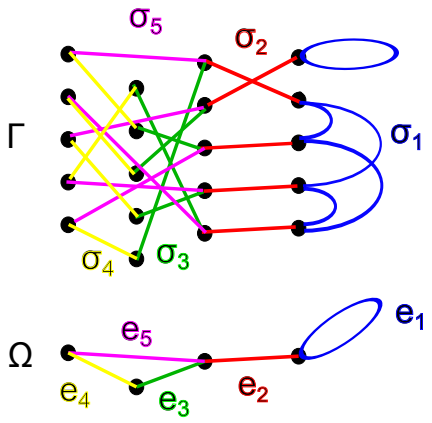


Figure 3.1.1: A 5-covering of a base graph using permutations.

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

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

A_Γ
M_Γ

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

(A third possible operator is the *Laplacian* - see Appendix 3.B.) With a suitable inner product, each of these operators is self-adjoint and therefore admits a real spectrum (see Appendix 3.B for the relations of these spectra to expansion properties of Γ).

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

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

[†]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.

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

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

the eigenvalue 1 corresponding to the constant function. Every finite covering Γ of Ω shares the same Perron-Frobenius eigenvalue, and moreover, inherits the entire spectrum of Ω (with multiplicity): Let $\pi : \Gamma \to \Omega$ be the covering map, sending the vertex (v,i) to v and the edge $((u,i),(v,j))$ to (u,v). Every eigenfunction $f : V(\Omega) \to \mathbb{C}$ of any operator on $l^2(V(\Omega))$ as above, can be pulled back to an eigenfunction of Γ, $f \circ \pi$, with the same eigenvalue. Thus, every eigenvalue of Ω (with multiplicity) is trivially an eigenvalue of Γ as well. We denote by $\lambda_A(\Gamma)$ the largest absolute value of a *new* eigenvalue of A_Γ, namely the largest one not inherited from Ω. Equivalently, this is the largest absolute eigenvalue of an eigenfunction of Γ which sums to zero on every fiber of π. In a similar fashion we define $\lambda_M(\Gamma)$, the largest absolute value of a new eigenvalue of M_Γ. Note that in the regular case (i.e. when Ω is d-regular), $A_\Gamma = d \cdot M_\Gamma$, and so $\lambda_A(\Gamma) = d \cdot \lambda_M(\Gamma)$. Moreover, when $\Omega = B_{\frac{d}{2}}$ is the bouquet, $\lambda_A(\Gamma) = \lambda(\Gamma)$. $\lambda_A(\Gamma)$ $\lambda_M(\Gamma)$

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

Proposition 3.1.2. *Let Γ be an n-covering of Ω. Then*

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

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

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

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

Conjecture 3.1.3 (Friedman, [Fri03])**.** *Let Ω be a finite connected graph. If Γ is a random n-covering of Ω, then for every $\varepsilon > 0$,*

$$\lambda_A(\Gamma) < \rho_A(\Omega) + \varepsilon$$

asymptotically almost surely, and likewise

$$\lambda_M(\Gamma) < \rho_M(\Omega) + \varepsilon$$

asymptotically almost surely.

Since $\lambda_A(\Gamma)$ and $\lambda_M(\Gamma)$ provide an indication for the quality of expansion of Γ (see Appendix 3.B), Conjecture 3.1.3 asserts that if the base graph Ω is a good (nearly optimal) expander then with high probability so is its random covering Γ.

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

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

The following theorems differ from Conjecture 3.1.3 only by a small additive or multiplicative factor, and are nearly optimal by Proposition 3.1.2. They pose a substantial improvement upon all former results, both in the special case of a d-regular base-graph Ω and, to a larger extent, in the general case of any finite base-graph.

Theorem 3.1.4. *Let Ω be an arbitrary finite connected graph, and let Γ be a random n-covering of Ω. Then for every $\varepsilon > 0$,*

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

asymptotically almost surely, and similarly

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

asymptotically almost surely.

For the special case where Ω is regular, we obtain the same bound as in the case of the bouquet (Theorem 3.1.1 for d even):

Theorem 3.1.5. *Let Ω be a finite connected d-regular graph ($d \geq 3$) and let Γ be a random n-covering of Ω. Then 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 Ω consisting of two vertices and d edges connecting them.

Corollary 3.1.6. *Let Γ 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 Γ are a.a.s. within $\left[-2\sqrt{d-1} - 0.84, 2\sqrt{d-1} + 0.84\right]$. The result applies also to random simple bipartite regular graphs: see appendix 3.A.

To put Theorems 3.1.1, 3.1.4 and 3.1.5 in context, Table 3.1 summarizes the results mentioned above for the different cases in question, with respect to the adjacency operator A_Γ.

Finally, let us stress that alongside the different models for random d-regular graphs, random coverings of a fixed, good expander, are probably the most natural other source for random, good expanders ("good" expanders are *sparse* graphs with high quality of expansion). Other known models for random graphs do not necessarily have this property. For example, the Erdös-Rényi model $G(n,p)$, 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.

[†]Again, for small values of d a better bound is reachable - see Sections 3.6.2 and 3.6.3.

The base-graph Ω	Any graph	d-regular	$B_{\frac{d}{2}}$ = a bouquet of $\frac{d}{2}$ loops
		$\rho = 2\sqrt{d-1}$	$\rho = 2\sqrt{d-1}$
Deterministic lower bound for $\lambda_A(\Gamma)$	$\rho - o_n(1)$ [Gre95]	$\rho - o_n(1)$ [Ser90]	$\rho - o_n(1)$ (Alon-Boppana) [Nil91]
Conjectured probabilistic upper bound	$\rho + \varepsilon$ [Fri03]		$\rho + \varepsilon$ [Alo86]
Probabilistic upper bounds, ordered by asymptotic strength for growing ρ	$\sqrt{\mathfrak{pf}(\Omega)\,\rho} + \varepsilon$ [Fri03] $\Longrightarrow$	$\sqrt{d\rho} + \varepsilon$	$\sqrt{d\rho} + \varepsilon$ [BS87]
	$3 \cdot \mathfrak{pf}(\Omega)^{1/3} \rho^{2/3} + \varepsilon$ [LP10] $\Longrightarrow$	$3 \cdot d^{1/3} \rho^{2/3} + \varepsilon$	
		$C \cdot \max(\lambda(\Omega), \rho) \log \rho$ [LSV11]	
		$265{,}000 \cdot \rho$ [ABG10]	$6{,}200 \cdot \rho$ [FKS89, DJPP11]
	$\sqrt{3} \cdot \boldsymbol{\rho + \varepsilon}$ **(Thm 3.1.4)**		
			$\rho + 2\log d + c$ [Fri91]
		$\boldsymbol{\rho + 0.84}$ **(Thm 3.1.5)**	$\boldsymbol{\rho + 0.84}$ **(Thm 3.1.1)**
			$\rho + \varepsilon$ [Fri08]

Table 3.1: Our results compared with former ones. As above, Ω is the connected base-graph and $\rho = \rho_A(\Omega)$ is the spectral radius of its universal covering tree. The results are ordered by their asymptotic strength.

3.2 Overview of the Proof

In this section we present the outline of the proof of Theorems 3.1.1, 3.1.4 and 3.1.5 (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].

Step I: The trace method

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

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

$\mathrm{Spec}(\mathrm{A}_\Gamma)$

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

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

Let $w = x_{j_1}^{\varepsilon_1} \dots x_{j_t}^{\varepsilon_t} \in \mathcal{CP}_t(\Omega) \subseteq \left(X \cup X^{-1}\right)^t$ be a closed path in the base graph Ω, beginning (and terminating) at some vertex $v \in V(\Omega)$. (Here $\varepsilon_i = \pm 1$ and x_j^{-1} means the 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 Γ, 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 $w(\sigma_1, \dots, \sigma_k)$ denote the permutation obtained by composing $\sigma_1, \dots, \sigma_k$ according to w, namely, $w(\sigma_1, \dots, \sigma_k) = \sigma_{j_1}^{\varepsilon_1} \dots \sigma_{j_t}^{\varepsilon_t} \in S_n$. Then j is the image of i under this permutation: $j = w(\sigma_1, \dots, \sigma_k)(i) = \sigma_{j_1}^{\varepsilon_1} \dots \sigma_{j_t}^{\varepsilon_t}(i)$[†]. Thus, the i-th lift of w is a closed path if and only if i is a fixed point of the permutation $w(\sigma_1, \dots, \sigma_k)$, and the number of closed paths in Γ projecting onto w is equal to the number of fixed points of $w(\sigma_1, \dots, \sigma_k)$. We denote this number by $\mathcal{F}_{w,n} = \mathcal{F}_{w,n}(\sigma_1, \dots, \sigma_k)$. $\mathcal{F}_{w,n}$

Claim 3.2.1. *For every even $t \in \mathbb{N}$,*

$$\mathbb{E}\left[\lambda_A(\Gamma)^t\right] \leq \sum_{w \in \mathcal{CP}_t(\Omega)} \left[\mathbb{E}\left[\mathcal{F}_{w,n}\right] - 1\right]. \tag{3.2.1}$$

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

Proof. Since t is even,

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

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

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

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

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

A word $w \in \mathbf{F}_k$ is *primitive* if it belongs to a *basis*[§] of $\mathbf{F}_k$. For a given w, one can also ask *primitive, basis*

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

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

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

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

Definition 3.2.2. The *primitivity rank* of $w \in \mathbf{F}_k$, denoted $\pi(w)$, is $\pi(w)$

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

If no such J exists, i.e. if w is primitive in $\mathbf{F}_k$, then $\pi(w) = \infty$.

A subgroup J for which the minimum is obtained is called *w-critical*, and the set of w-critical subgroups is denoted $\mathrm{Crit}(w)$. $\mathrm{Crit}(w)$

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

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

Theorem 3.2.3. *[PP15, Thm 1.7] For every $w \in \mathbf{F}_k$, the expected number of fixed points in $w(\sigma_1, \ldots, \sigma_k)$ is*

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

In particular, it is also shown that $\mathrm{Crit}(w)$ is always finite. The three leftmost columns in Table 3.2 summarize the connection implied by Theorem 3.2.3 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 3.2.3 at hand, we can use the primitivity rank to split the summation in (3.2.1). We shall use the notation $\mathcal{CP}_t^m(\Omega) = \{w \in \mathcal{CP}_t(\Omega) \mid \pi(w) = m\}$ for the subsets we obtain by splitting $\mathcal{CP}_t(\Omega)$ according to primitivity ranks: $\mathcal{CP}_t^m(\Omega)$

$$\begin{aligned} \mathbb{E}\left[\lambda_A(\Gamma)^t\right] &\leq \sum_{w \in \mathcal{CP}_t(\Omega)} (\mathbb{E}[\mathcal{F}_w] - 1) = \\ &= \sum_{m=0}^{k} \sum_{w \in \mathcal{CP}_t^m(\Omega)} \left(\frac{|\mathrm{Crit}(w)|}{n^{m-1}} + O\left(\frac{1}{n^m}\right) \right) \end{aligned} \tag{3.2.2}$$

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

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

The error term $O\left(\frac{1}{n^m}\right)$ in (3.2.2) 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 (3.2.2), one needs a uniform bound for all closed paths of length t with a given primitivity rank in Ω. Namely, for every m one needs to control the $O(\cdot)$ term for all $w \in \mathcal{CP}_t^m(\Omega)$ simultaneously. The third stage is therefore the following proposition:

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

Proposition (Follows from Prop. 3.5.1 and Claim 3.5.2)**.** *Let $t = t(n)$ and $w \in \left(X \cup X^{-1}\right)^t$. If $t^{2k+2} = o(n)$ then*

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

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

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

$$\mathbb{E}\left[\lambda_A\left(\Gamma\right)^t\right] \leq \left(1 + o_n\left(1\right)\right)\sum_{m=0}^{k}\frac{1}{n^{m-1}}\sum_{w\in\mathcal{CP}_t^m(\Omega)}\left|\mathrm{Crit}\left(w\right)\right|. \tag{3.2.3}$$

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 $\sum_{w\in\mathcal{CP}_t^m(\Omega)}\left|\mathrm{Crit}\left(w\right)\right|$ 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 $\sum_{w\in\mathcal{CP}_t^0(\Omega)}\left|\mathrm{Crit}\left(w\right)\right| = \left|\mathcal{CP}_t^0\left(\Omega\right)\right|$. 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 $\left|\mathcal{CP}_t^0\left(\Omega\right)\right|$ is exactly $\rho = \rho_A\left(\Omega\right)$, the spectral radius of the covering tree (see Claim 3.4.12). For larger m we obtain the following upper bound:

Theorem (Theorem 3.4.11)**.** *Let Ω be a finite, connected graph with $k \geq 2$ edges, and let $m \in \{1, \ldots, k\}$. Then*

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

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

Theorem (Follows from Corollaries 3.4.5 and 3.4.10 and Theorem 3.8.5)**.** *Let Ω be a finite, connected d-regular graph ($d \geq 3$) with k edges, and let $m \in \{0, 1, \ldots, k\}$. Then*

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

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

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

Remark 3.2.4. In fact, in the case of the bouquet, the growth rates in the statement of the last theorem remain the same if we assume every word has only a single critical subgroup. That is, the r.h.s. gives also the growth rate of the *number* of words in $\left(X \cup X^{-1}\right)^t$ with primitivity rank m - see Theorem 3.8.5.

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

Table 3.2: Primitivity rank, the average number of fixed points, the exponential growth rate of $\sum_{w\in\mathcal{CP}_t^m\left(B_{\frac{d}{2}}\right)}|\operatorname{Crit}(w)|$, and bounds on the exponential growth rate of $\sum_{w\in\mathcal{CP}_t^m(\Omega)}|\operatorname{Crit}(w)|$.

Table 3.2 summarizes the content of Theorems 3.2.3, 3.4.11 and 3.8.5†.

Whereas in the special case of the bouquet we count words in $\mathbf{F}_k$ of a given length and a given primitivity rank, the case of a general graph concerns the equivalent question for words which in addition belong to some fixed subgroups of $\mathbf{F}_k$. (There is one such subgroup for each vertex v of Ω: the one consisting of the words which correspond to closed paths at v.) The fact that the bounds in Corollary 3.4.5 are better than those in Theorem 3.4.11 explains the gap between Theorems 3.1.1 and 3.1.5 which are tight up to a small *additive* constant, and Theorem 3.1.4 which is tight up to a small *multiplicative* factor (assuming, of course, that Conjecture 3.1.3 is true).

Step V: Some analysis

The final step is fairly simple and technical: it consists of analyzing the upper bounds we obtain from (3.2.3) together with Theorem 3.4.11 and Corollary 3.4.5. We seek the value of t (as a function of n) which yields the best bounds.

The paper is arranged as follows. Section 3.3 provides some basic facts about the concepts of core graphs and algebraic extensions which are used throughout this paper. In Section 3.4 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 3.4.1, then for the intermediate case of an arbitrary regular base graph in Section 3.4.2, and finally for the most general case in Section 3.4.3). The third step of the proof, where the error term from Theorem 3.2.3 is dealt with, is carried out in

†The number $2k-2+\frac{2}{2k-3}$ in the last row of the table is the exponential growth rate of the set of primitives in $\mathbf{F}_k$, namely of $\left|\mathcal{CP}_t^\infty\left(B_{\frac{d}{2}}\right)\right|$. (Primitive words have no critical subgroups.) This result is not necessary for the current work, and is established in a separate paper [PW14], using completely different techniques. We use it here only to show that our bounds for $\sum_{w\in\mathcal{CP}_t^m\left(B_{\frac{d}{2}}\right)}|\operatorname{Crit}(w)|$ are tight - see Section 3.8.

Section 3.5, where we have to recall some more details from [PP15]. Section 3.6 completes the proof of Theorems 3.1.1 and 3.1.5 and addresses the source of the gap between Theorem 3.1.1 and Friedman's result. In Section 3.7 we complete the proof of Theorem 3.1.4. We end with results on the accurate exponential growth rate of words with a given primitivity rank in $\mathbf{F}_k$ (Section 3.8), 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 3.A), and on the theory of spectral expansion of non-regular graphs (Appendix 3.B).

3.3 Preliminaries: Core Graphs and Algebraic Extensions

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

3.3.1 Algebraic extensions

Let $H \leq J$ be subgroups of $\mathbf{F}_k$. We say that J is an *algebraic extension* of H and denote $H \leq_{\text{alg}} J$, if there is *no* intermediate subgroup $H \leq L \lneq J$ 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 $H \leq_{alg} M \overset{*}{\leq} J$ (where $\overset{*}{\leq}$ denotes a free factor). Moreover, if $H \leq \mathbf{F}$ is a finitely generated subgroup, it has only finitely many algebraic extensions in $\mathbf{F}$. Thus, every group containing H is a free extension of one of finitely many extensions of H, which is a well known theorem of Takahasi [Tak51]. For more information we refer the interested reader to [KM02, PP15] and especially to [MVW07].

algebraic extension $H \leq_{\text{alg}} J$

$\overset{*}{\leq}$

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

Claim 3.3.1. *[Pud14, Cor. 4.4] Every w-critical subgroup is an algebraic extension of $\langle w \rangle$ (the subgroup generated by w).*

More precisely, $\text{Crit}(w)$ *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, $\langle w \rangle \lneq H$. If H is not an algebraic extension of $\langle w \rangle$, then there is a proper intermediate free factor $\langle w \rangle \leq L \lneq_{ff} H$. Since w is not primitive in H, it is also not primitive in L (e.g. [Pud14, 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 $\langle w \rangle$.

3.3.2 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 3.3.1.

$\Gamma_X(H)$

To understand how $\Gamma_X(H)$ is constructed, recall first the notion of the Schreier (right) coset graph of H with respect to the basis X, denoted by $\overline{\Gamma}_X(H)$. This is a directed, pointed and edge-labeled graph. Its vertex set is the set of all right cosets of H in $\mathbf{F}_k$, where the basepoint corresponds

$\overline{\Gamma}_X(H)$

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

to the trivial coset H. For every coset Hw and every basis-element x_j there is a directed j-edge (short for x_j-edge) going from the vertex Hw to the vertex Hwx_j.[†]

The core graph $\Gamma_X(H)$ is obtained from $\overline{\Gamma}_X(H)$ by omitting all the vertices and edges of $\overline{\Gamma}_X(H)$ which are not traced by any reduced (i.e., non-backtracking) path that starts and ends at the basepoint. Stated informally, we trim all "hanging trees" from $\overline{\Gamma}_X(H)$. To illustrate, Figure 3.3.1 shows the graphs $\overline{\Gamma}_X(H)$ and $\Gamma_X(H)$ for $H = \langle x_1x_2x_1^{-3}, x_1^2x_2x_1^{-2}\rangle \leq \mathbf{F}_2$.

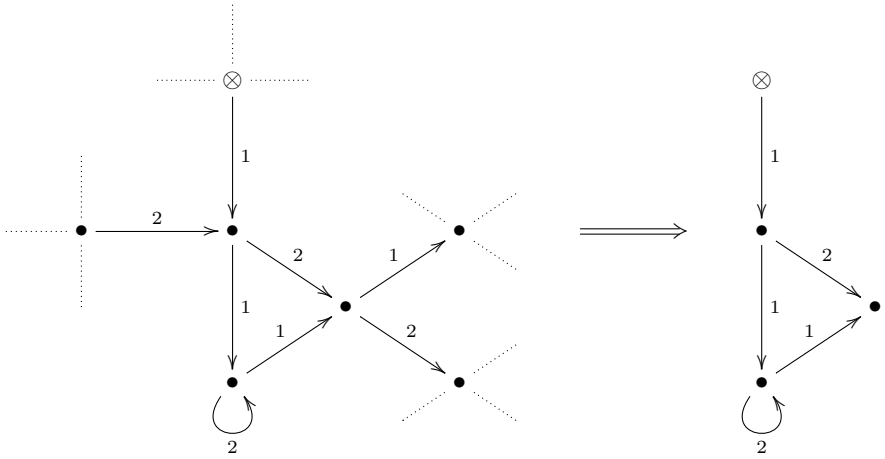
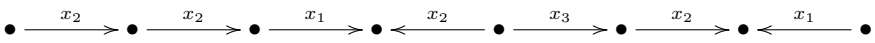
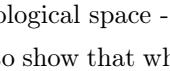


Figure 3.3.1: $\overline{\Gamma}_X(H)$ and $\Gamma_X(H)$ for $H = \langle x_1x_2x_1^{-3}, x_1^2x_2x_1^{-2}\rangle \leq \mathbf{F}_2$. The Schreier coset graph $\overline{\Gamma}_X(H)$ is the infinite graph on the left (the dotted lines represent infinite 4-regular trees). The basepoint "$\otimes$" corresponds to the trivial coset H, the vertex below it corresponds to the coset Hx_1, the one further down corresponds to $Hx_1^2 = Hx_1x_2x_1^{-1}$, etc. The core graph $\Gamma_X(H)$ is the finite graph on the right, which is obtained from $\overline{\Gamma}_X(H)$ by omitting all vertices and edges that are not traced by reduced closed paths around the basepoint.

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



corresponds to the word $x_2^2x_1x_2^{-1}x_3x_2x_1^{-1}$. The set of all words obtained from closed paths around the basepoint in Γ is a subgroup of $\mathbf{F}(X)$ which we call the *labeled fundamental group* of Γ, and denote by $\pi_1^X(\Gamma)$. Note that $\pi_1^X(\Gamma)$ need not be isomorphic to $\pi_1(\Gamma)$, the standard fundamental group of Γ viewed as a topological space - for example, take $\Gamma =$  . $\pi_1^X(\Gamma)$

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

$$\pi_1^X\left(\overline{\Gamma}_X(H)\right) = \pi_1^X\left(\Gamma_X(H)\right) = H \tag{3.3.1}$$

This gives a one-to-one correspondence between subgroups of $\mathbf{F}(X) = \mathbf{F}_k$ and core graphs labeled

[†]Alternatively, $\overline{\Gamma}_X(H)$ is the quotient $H\backslash T$, where T is the Cayley graph of $\mathbf{F}_k$ with respect to the basis X, and F_k (and thus also H) acts on this graph from the left. Moreover, this is the covering-space of $\overline{\Gamma}_X(F_k) = \Gamma_X(F_k)$, the bouquet of k loops, corresponding to H, via the correspondence between pointed covering spaces of a space Y and subgroups of its fundamental group $\pi_1(Y)$.

by X. Namely, π_1^X and Γ_X are the inverses of each other in a bijection (Galois correspondence)

$$\left\{ \begin{array}{c} \text{Subgroups} \\ \text{of } \mathbf{F}(X) \end{array} \right\} \underset{\pi_1^X}{\overset{\Gamma_X}{\rightleftarrows}} \left\{ \begin{array}{c} \text{Core graphs} \\ \text{labeled by } X \end{array} \right\}. \tag{3.3.2}$$

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

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

Claim 3.3.2. *Let H be a subgroup of $\mathbf{F}_k$ with an associated core graph $\Gamma = \Gamma_X(H)$.*

(1) $\mathrm{rk}(H) < \infty \iff \Gamma$ *is finite.*

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

(3) The correspondence (3.3.2) restricts to a correspondence between finitely generated subgroups of $\mathbf{F}_k$ 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 3.3.2, each of the following properties is either proven in (some of) [Sta83, KM02, MVW07, Pud14] or an easy observation:

Claim 3.3.3. *Let $H, J, L \leq \mathbf{F}_k$ be subgroups. Then*

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

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

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

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

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

3.4 Counting Words and Critical Subgroups

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

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

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

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

[†]Points (1)-(3) can be formulated by saying that (3.3.2) is in fact an isomorphism of categories, given by the functors π_1^X and Γ_X.

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

Lemma 3.4.1. *Let $w \in \mathbf{F}_k$ and let $N \leq_{f.g.} \mathbf{F}_k$ be a proper algebraic extension of $\langle w \rangle$.* Then the closed 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 3.5.3 and e.g. [PP15, Claim 4.2]. We repeat the simple argument here.) Otherwise, let J be the subgroup of N corresponding to the subgraph Δ traced by w (so $\Delta = \operatorname{im} \eta^X_{\langle w \rangle \to N}$), and $J = \pi_1^X(\Delta)$, see Section 3.3.2 and in particular Claim 3.3.3). Then $w \in J \lneq_{ff} N$ (Claim 3.3.3), contradicting the fact that N is an algebraic extension of $\langle w \rangle$.

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

Finally, assume that e is not separating, and w traverses it exactly once, so that the path corresponding to w in $\Gamma_X(N)$ is $w_1 e w_2$ (with w_1, w_2 avoiding e; we think of e as oriented according to the direction of w). Choose a spanning tree T of $\Gamma_X(N)$ which avoids e to obtain a basis for N as follows. There are $r = \operatorname{rk}(N)$ 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 $\langle w \rangle \overset{*}{\leq} N$, a contradiction.

It is enough to show that $u_1 \in \langle w, u_2, \ldots, u_r \rangle$ (see footnote on Page 85). 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

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

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

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

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

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

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

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

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

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

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

so $e \leq 3m - 1$. □

3.4.1 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

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

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

Proposition 3.4.3. *Let $k \geq 2$ and $m \in \{1, 2, \ldots, k\}$. Then*

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

Put differently, the lim sup is bounded by $\max\left\{\sqrt{2k-1}, 2m-1\right\}$ (we present it in a lengthier way to stress the threshold phenomenon). In fact, this is not only an upper bound but the actual exponential growth rate - see Theorem 3.8.2.

Proof. Note that

$$\begin{aligned} \sum_{\substack{w \in \mathbf{F}_k: \\ |w|=t \,\&\, \pi(w)=m}} |\mathrm{Crit}\,(w)| &= \sum_{J \leq \mathbf{F}_k:\, \mathrm{rk}(J)=m} |\{w \in \mathbf{F}_k \,|\, |w| = t,\, J \in \mathrm{Crit}\,(w)\}| \\ &\leq \sum_{J \leq \mathbf{F}_k:\, \mathrm{rk}(J)=m} |\{w \in \mathbf{F}_k \,|\, |w| = t,\, \langle w \rangle \lesssim_{alg} J\}| \qquad (3.4.1) \\ &\leq \sum_{J \leq \mathbf{F}_k:\, \mathrm{rk}(J)=m} \left| \left\{ w \in J \;\middle|\; \begin{array}{c} |w| = t,\, w \,\text{traces each edge} \\ \text{of } \Gamma_X\,(J) \text{ at least twice} \end{array} \right\} \right|, \end{aligned}$$

where the first inequality stems from Claim 3.3.1 and the second from Lemma 3.4.1. We continue to bound the latter sum. For each $J \leq \mathbf{F}_k$ let $\nu_t\,(J)$ denote the corresponding summand:

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

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

- First, let us bound the number of unlabeled and unoriented connected pointed graphs with δt edges and rank m (here the rank of a connected graph is $e - v + 1$). As in the proof of Lemma 3.4.1, 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 $< \delta t$. Now, $p_{2,1}$ leaves $p_{1,1} \cup p_{1,2}$ at one of $\leq \delta t$ possible vertices and goes on for $< \delta t$ new edges. This goes on $2m$ times in total

(afterward, the ends of $p_{i,1}$ and $p_{i,2}$ are connected by an edge). In total, there are at most $\left((\delta t)^2\right)^{2m} = (\delta t)^{4m}$ possible unlabeled pointed graphs of rank m with δt edges[†].

- Next, we bound the number of labelings of each such graph Γ (here, the labeling includes also the orientation of each edge). Label some edge (there are $2k$ options) and then gradually label edges adjacent to at least one edge which is already labeled (at most $2k-1$ possible labels for each edge). Over all the number of possible labelings of Γ is $\leq 2k \cdot (2k-1)^{\delta t - 1}$.
- For a given labeled core-graph Γ, let $J = \pi_1^X(\Gamma)$ be the corresponding subgroup. We claim that $\nu_t(J) \leq \left(4t^2\right)^{3m-1} \cdot (2m-1)^{(1-2\delta)t}$. Indeed, note first that if the basepoint $\otimes$ is a leaf, then every reduced w must first follow the string from $\otimes$ to the first "topological" vertex (vertex of degree ≥ 3), and then return to the string only in its final steps back to $\otimes$. So we can assume w traces a leaf-free graph of rank m and at most δt edges. A reduced word $w \in J$ which traces every edge at least twice, also traverses any topological edge at least twice, each time in one shot (without backtracking). Each time w traces some topological edge $\widetilde{e}$ in Γ, it begins in one of $\leq t$ possible positions (in w), and from ≤ 2 possible directions of $\widetilde{e}$. So there $\leq 4t^2$ possible ways in which w traces $\widetilde{e}$ for the first two times. By Claim 3.4.2(2) there are at most $3m-1$ topological edges, and so at most $\left(4t^2\right)^{3m-1}$ possibilities for how w traces each topological edge of Γ for the first two times. The rest of w is of length (at most) $(1-2\delta)t$, and in every step there are at most $2m-1$ ways to proceed, by Claim 3.4.2(1).

Hence,

$$\begin{aligned}
\sum_{\substack{J \leq \mathbf{F}_k : rk(J)=m \\ |\Gamma_X(J)|=\delta t}} \nu_t(J) &\leq (\delta t)^{4m} \cdot 2k\,(2k-1)^{\delta t-1} \cdot \left(4t^2\right)^{3m-1} (2m-1)^{(1-2\delta)t} \\
&\leq c \cdot t^{10m-2} \cdot \left[(2k-1)^{\delta} (2m-1)^{1-2\delta}\right]^t \\
&= c \cdot t^{10m-2} \cdot \left[\left(\frac{2k-1}{(2m-1)^2}\right)^{\delta} (2m-1)\right]^t . \qquad (3.4.2)
\end{aligned}$$

Recall that $\delta \in \left[0, \frac{1}{2}\right]$ and $\delta t \in \mathbb{N}$. We bound $\sum_{J \leq \mathbf{F}_k : rk(J)=m} \nu_t(J)$ by $\frac{t}{2}$ times the maximal possible value of the r.h.s. of (3.4.2) (when going over all possible values of δ). When $2m-1 \leq \sqrt{2k-1}$, the r.h.s. of (3.4.2) is largest when $\delta = \frac{1}{2}$, so we get overall

$$\sum_{J \leq \mathbf{F}_k : rk(J)=m} \nu_t(J) \leq c \cdot t^{10m-1} \cdot \left[\sqrt{2k-1}\right]^t . \qquad (3.4.3)$$

For $2m-1 \geq \sqrt{2k-1}$, the r.h.s. of (3.4.2) is largest when $\delta = 0$, so we get overall

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

The proposition follows. □

The next step is to deduce an analogue result for non-reduced words. To this goal, we use an extended version of the well known *cogrowth formula* due to Grigorchuk [Gri77] and Northshield [Nor92]. Let Γ be a connected d-regular graph. Let $b_{\Gamma,v}(t)$ denote the number of *cycles* of length t at some vertex v in Γ, and let $n_{\Gamma,v}(t)$ denote the size of the smaller set of *non-backtracking cycles* of length t at v. The spectral radius of A_Γ, denoted $\mathrm{rad}(\Gamma)$[‡], is equal to $\limsup_{t\to\infty} b_{\Gamma,v}(t)^{1/t}$

(in particular, this limit does not depend on v). The *cogrowth* of Γ is defined as $\operatorname{cogr}(\Gamma) = \limsup_{t\to\infty} n_{\Gamma,v}(t)^{1/t}$, and is also independent of v. $\operatorname{cogr}(\cdot)$

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

$$g(\alpha) = \begin{cases} 2\sqrt{d-1} & \alpha \le \sqrt{d-1} \\ \frac{d-1}{\alpha} + \alpha & \alpha \ge \sqrt{d-1} \end{cases}. \tag{3.4.4}$$

Another way to view the parameters $\operatorname{rad}(\Gamma)$ and $\operatorname{cogr}(\Gamma)$ 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[†] of T_d belonging to S.

For our needs we introduce (in a separate paper - [Pud15b])[‡] an extended formula applying to other types of subsets S of $V(T_d)$, which do not necessarily correspond to a fiber of a covering map of a graph. Even more generally, we extend the formula to a class of functions on $V(T_d)$ (this extends the previous case if S is identified with its characteristic function $\mathbb{1}_S$):

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

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

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 3.4.4. *[Extended Cogrwoth Formula [Pud15b]] Let $d \ge 3$, $f : V(T_d) \to \mathbb{R}$, $\beta_f(t)$ and $\nu_f(t)$ as above. If $\nu_f(t) \le \overline{c} \cdot \alpha^t$ then*

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

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 3.4.5. *For every $k \ge 2$ and $m \in \{1, \dots, k\}$,*

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

Proof. Consider the the Cayley graph of $\mathbf{F}_k$ which is a $2k$-regular tree. Every vertex corresponds to a word in $\mathbf{F}_k$, and we let $f_m(w) = \mathbb{1}_{\pi(w)=m} |\operatorname{Crit}(w)|$. The corollary then follows by applying Theorem 3.4.4 on f_m, using Proposition 3.4.3. □

In Section 3.8 it is shown (Theorem 3.8.5) that the bound in Corollary 3.4.5 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 [Pud15b].

[†]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.

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

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

[‡]The results in [Pud15b] include a new proof of the original cogrowth formula.

Remark 3.4.6. Interestingly, the threshold of $\sqrt{2k-1}$ shows up twice, apparently independently, both in Proposition 3.4.3 and in the (extended) cogrowth formula.

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

Claim 3.4.7.

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

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

$$\operatorname{rad}(\Gamma)=\|A_\Gamma\|=\limsup_{t\to\infty}\sqrt[t]{c_\Gamma(t,u,v)} \tag{3.4.5}$$

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

$$c_\Gamma(t,u,v)=\left\langle A_\Gamma^t\delta_u,\delta_v\right\rangle_1\le\left\|A_\Gamma^t\delta_u\right\|\cdot\|\delta_v\|\le\|A_\Gamma\|^t\cdot\|\delta_u\|\cdot\|\delta_v\|=\operatorname{rad}(\Gamma)^t \tag{3.4.6}$$

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

The words of length t reducing to 1 are exactly the closed paths of length t at the basepoint of the $2k$-regular tree T_{2k}. So the number we seek is
$\limsup_{t\to\infty}\sqrt[t]{c_{T_{2k}}(t,v,v)}$, which therefore equals $\operatorname{rad}(T_{2k})=2\sqrt{2k-1}$. □

3.4.2 An arbitrary regular base-graph Ω

We proceed with the observation that when Ω is d-regular (but not necessarily the bouquet), the bounds from Corollary 3.4.5 generally apply. We begin with a few claims that will be useful also in the next subsection dealing with irregular base graphs.

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

Lemma 3.4.8. *Let Ω be a finite, connected graph. Then $\pi(w)\in\{0,1,\ldots,\operatorname{rk}(\Omega),\infty\}$ for every $w\in\mathcal{CP}_t(\Omega)$.*

Proof. Recall from Section 3.2 that we denote $k=|E(\Omega)|$ and orient each of the k edges arbitrarily and label them by $x_1,\ldots,x_k$. With the orientation and labeling of its edges, Ω becomes a non-pointed X-labeled graph, where $X=\{x_1,\ldots,x_k\}$. (This is not a core-graph, for it has no basepoint and may have leaves.) So every path in Ω of length t can be regarded as an element of $\left(X\cup X^{-1}\right)^t$ and (after reduction) of $\mathbf{F}_k=\mathbf{F}(X)$. If a word $w\in\mathcal{CP}_t(\Omega)$ begins (and ends) at $v\in V(\Omega)$, then $w\in J_v$, where $J_v=\pi_1^X(\Omega_v)$ is the subgroup of $\mathbf{F}_k$ corresponding to the X-labeled graph Ω pointed at v. The rank of J_v is independent of v and equals $\operatorname{rk}(\Omega)$. It is easy to see that $J_v\overset{*}{\le}\mathbf{F}_k$ (recall that '$\overset{*}{\le}$' 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 3.4.1. This basis can then be extended to a basis of $\mathbf{F}_k$ by the x_i's associated with the edges inside the spanning tree. So if w is primitive in J_v, is it also primitive in $\mathbf{F}_k$ and $\pi(w)=\infty$. Otherwise, $\pi(w)\le\operatorname{rk}(J_v)=\operatorname{rk}(\Omega)$. □ J_v,Ω_v

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

Claim 3.4.9. *In $w\in\mathcal{CP}_t(\Omega)$ is a cycle around the vertex v and $\langle w\rangle\lneq_{alg}N$, then $N\le J_v$.*

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

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

Corollary 3.4.10. *Let Ω be a finite, connected d-regular graph, and let $m \in \{0, 1, \ldots, \mathrm{rk}(\Omega)\}$. Then*

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

Proof. First, for words with $\pi(w) = 0$, that is, words reducing to 1, their number is $|V(\Omega)|$ times the number of cycles of length t at a fixed vertex in the d-regular tree. Thus, as in the proof of Claim 3.4.7,

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

For $m \geq 1$, since the extended cogrowth formula (Theorem 3.4.4) applies here too, it is enough to prove that for *reduced* words we have:

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

From Claim 3.4.9 we deduce that every critical subgroup is necessarily a subgroup of $J_v = \pi_1^X(\Omega_v)$ for some vertex $v \in V(\Omega)$. As in the proof of Proposition 3.4.3, we denote

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

for every $J \leq \mathbf{F}_k$, and as in (3.4.1), we obtain the bound:

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

We carry the same counting argument as in the proof of Proposition 3.4.3:

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

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

We conclude as in the proof of Proposition 3.4.3.

□

3.4.3 An arbitrary base-graph Ω

We now return to the most general case of an arbitrary connected base graph Ω. Theorem 3.4.11 below is needed for proving the bound on the new spectrum of the adjacency operator on Γ, the random covering of Ω in the $\mathcal{C}_{n,\Omega}$ model (the first part of Theorem 3.1.4). The small variation needed for the second part of this theorem, dealing with the Markov operator, is discussed in Section 3.7.1.

Recall that T denotes the universal covering of Ω (and of Γ), and $\rho = \rho_A(\Omega)$ denotes the spectral radius of its adjacency operator. Recall also that we denote $k = |E(\Omega)|$ and orient each of the k edges arbitrarily and label them by $x_1,\ldots,x_k$. With the orientation and labeling of its edges, Ω becomes a non-pointed X-labeled graph, where $X = \{x_1,\ldots,x_k\}$. Every path in Ω of length t can be regarded as an element of $\left(X \cup X^{-1}\right)^t$ and (after reduction) of $\mathbf{F}_k = \mathbf{F}(X)$. We also denoted $\mathrm{rk}(\Omega) = |E(\Omega)| - |V(\Omega)| + 1$ and showed that $\pi(w) \in \{0,1,\ldots,\mathrm{rk}(\Omega),\infty\}$ for every $w \in \mathcal{CP}_t(\Omega)$ (Lemma 3.4.8). The main theorem of this subsection is the following: T

Theorem 3.4.11. *Let Ω be a finite, connected graph, and let $m \in \{1,\ldots,\mathrm{rk}(\Omega)\}$. Then*

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

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

Claim 3.4.12.

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

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

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

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

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

To show there is actual equality, repeat the argument from Claim 3.4.7. □

We return to the proof of Theorem 3.4.11. By Claim 3.3.1,

$$\begin{aligned}\sum_{w\in\mathcal{CP}_t^m(\Omega)} |\mathrm{Crit}(w)| &= \sum_{\substack{N\leq\mathbf{F}_k:\\ \mathrm{rk}(N)=m}} |\{w\in\mathcal{CP}_t(\Omega) \mid N\in\mathrm{Crit}(w)\}| \\ &\leq \sum_{\substack{N\leq\mathbf{F}_k:\\ \mathrm{rk}(N)=m}} |\{w\in\mathcal{CP}_t(\Omega) \mid \langle w\rangle \lesssim_{alg} N\}| \qquad (3.4.7)\end{aligned}$$

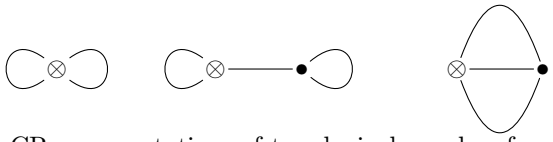


Figure 3.4.1: The three CR representatives of topological graphs of rank 2: Figure-Eight, Barbell and Theta.

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

$$\beta_t(N) = \left|\{w \in \mathcal{CP}_t(\Omega) \mid \langle w \rangle \lesssim_{alg} N\}\right|.$$

Note that while a non-reduced element $w \in \mathcal{CP}_t(\Omega)$ 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 $\overline{\Gamma}_X(N)$.

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

Claim 3.4.13. *If $N \leq \mathbf{F}_k$ is CR then*

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

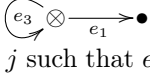
Proof. The Schreier graphs of N and of any conjugate of it differ only by the basepoint. If N' is some conjugate of N and $w' \in \mathcal{CP}_t(\Omega)$ satisfies $\langle w' \rangle \lesssim_{alg} N'$, then the path corresponding to w' in the Schreier graph $\overline{\Gamma}_X(N')$ must visit all vertices and edges of the core of $\overline{\Gamma_X(N')}$, and in particular the basepoint of $\overline{\Gamma}_X(N)$ (by Lemma 3.4.1). So there is some cyclic rotation w of w' satisfying $\langle w \rangle \lesssim_{alg} N$ (clearly, w also belongs to $\mathcal{CP}_t(\Omega)$). On the other hand, each such w has at most t possible cyclic rotations, each of which corresponds to one w' and one N'. □

Next, we classify the subgroups $N \leq \mathbf{F}_k$ according to their "topological" core graph Λ. As implied in the short discussion preceding Claim 3.4.2, 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 3.4.13 allows us to restrict to one CR representative from each conjugacy class of subgroups in $\mathbf{F}_k$, we also restrict attention to one CR representative Λ of each "conjugacy class" of topological core graphs. Ignoring the basepoints, any Λ' in the "conjugacy class" of Λ retracts to this representative. For example, we need exactly three such representatives in rank 2, as shown in Figure 3.4.1.

The following proposition is the key step in the proof of Theorem 3.4.11.

Proposition 3.4.14. *Let Λ be a pointed finite connected graph without vertices of degree 1 or 2 except for possibly the basepoint, and let δ denote its maximal degree. Then the sum of $\beta_t(N)$ over all subgroup $N \leq \mathbf{F}_k$ whose core graph is topologically Λ is at most*

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

Proof. Denote $r = |E(\Lambda)|$. Order and orient the edges of Λ $\{e_1, e_2, \ldots, e_r\}$ so that e_1 emanates from $\otimes$, and for every $i \geq 2$, e_i emanates either from $\otimes$ or from a vertex which is the beginning or endpoint of one of $e_1, \ldots, e_{i-1}$. 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:  , where $v_0 = v_3$ are $\otimes$ and $v_1 = v_2$ are $\bullet$. Also, denote by $\text{beg}(i)$ the smallest index j such that e_i begins at

v_j, so e_i is a directed edge from $v_{\mathrm{beg}(i)}$ to v_i and $\mathrm{beg}\,(\mathrm{i}) < i$. In our example, $\mathrm{beg}\,(1) = \mathrm{beg}\,(3) = 0$ and $\mathrm{beg}\,(2) = 1$.

Note that each N corresponding to Λ is determined by the paths (words in $\mathbf{F}_k$) associated with $e_1, \ldots, e_r$. From Claim 3.4.9 it follows one can restrict to subgroups N which are subgroups of J_v for some $v \in V(\Omega)$. So fix some $v_0 \in V(\Omega)$ and also some $\widehat{v}_0 \in V(T)$ which projects to v_0. We claim that every subgroup $N \leq J_{v_0}$ corresponding to Λ is completely determined by a set of vertices $\widehat{v}_1, \ldots, \widehat{v}_r$ in T: the topological edge in $\Gamma_X(N)$ associated with e_i corresponds to the path in T from $\widehat{v}_{\mathrm{beg(i)}}$ to $\widehat{v}_i$. (There are some constraints on the choices of the $\widehat{v}_i$'s. For example, if $v_i = v_j$ then $\widehat{v}_i$ and $\widehat{v}_j$ must belong to the same fiber of the projection map $p : T \to \Omega$. However, as we only bound from above, we ignore these constraints.) So instead of summing over all possible N's, we go through all possible choices of vertices $\widehat{v}_1, \ldots, \widehat{v}_r$ in T.

The counting argument that follows resembles the one in Proposition 3.4.3. Fix a particular $N \leq J_{v_0}$ corresponding to Λ and let $\widehat{v}_1, \ldots, \widehat{v}_r$ be the corresponding vertices in T. By Lemma 3.4.1, if $w \in \left(X \cup X^{-1}\right)^t$ satisfies $\langle w \rangle \lesssim_{alg} N$, then its reduced form traverses every topological edge of $\Gamma_X(N)$ at least twice. For each i, assume that w first traverses the topological edge associated with e_i starting at position $\tau_{i,1}$ (the position is in w, namely $0 \leq \tau_{i,1} \leq t-1$), and in $\ell_{i,1}$ steps, and then from position $\tau_{i,2}$ in $\ell_{i,2}$ steps (recall that w is not reduced so $\ell_{i,2}$ may be different from $\ell_{i,1}$). The directions of these traverses are $\varepsilon_{i,1}, \varepsilon_{i,2} \in \{\pm 1\}$. In total, there are less than t^{2r} options for the $\tau_{i,j}$'s, less than t^{2r} options for the $\ell_{i,j}$'s and less than 2^{2r} options for the $\varepsilon_{i,j}$'s: a total of less than $\left(4t^4\right)^r$ options. There are $t - \ell_{1,1} - \ell_{1,2} - \ldots - \ell_{r,1} - \ell_{r,2}$ remaining steps, and these are divided to at most $4r$ segments (we can always assume one of the $\tau_{i,1}$'s equals 0). Denote the lengths of these segments by $q_1, \ldots, q_{4r}$ (some may be 0). The i'th segment reduces to some path in $\Gamma_X(N)$, with at most $(\delta-1)^{q_i}$ possibilities (recall that δ marks the maximal degree of a vertex in Λ). Overall, there are at most $(\delta-1)^{q_1+\ldots+q_{4r}} \leq (\delta-1)^t$ options to choose the reduced paths traced by these $4r$ segments in w. Given such a reduced path for the i'th segment, let $\widehat{x}_i, \widehat{y}_i \in V(T)$ be suitable vertices in the tree such that the reduced path lifts to the unique reduced path from $\widehat{x}_i$ to $\widehat{y}_i$.

Now, we sum over all subgroups N corresponding to Λ and all words $w \in \mathcal{CP}_t(\Omega)$ with $\langle w \rangle \lesssim_{alg} N$. By adding a factor of $|V(\Omega)| \left(4t^4\right)^r \cdot (\delta-1)^t$ we assume we already know v_0 and $\widehat{v}_0$, the $\tau_{i,j}$'s, $\ell_{i,j}$'s, $\varepsilon_{i,j}$'s, the q_i's and the reduced $4r$ paths. Moreover, conditioning on knowing $\widehat{v}_1, \ldots, \widehat{v}_r$, we also know the $\widehat{x}_i$'s and the $\widehat{y}_i$'s. Recall that $c_\Gamma(t,u,v)$ denotes the number of paths of length t in a graph Γ from the vertex u to the vertex v, and that by (3.4.6), $c_T(t,u,v) \leq \rho^t$ for every $u, v \in V(T)$. For each $i = 1, \ldots, r$ and $j = 1, 2$, there are $c_T\left(\ell_{i,j}, \widehat{v}_{\mathrm{beg}(i)}, \widehat{v}_i\right)$ possible subwords corresponding to the j'th traverse of e_i (even if $\varepsilon_{i,j} = -1$, because $c_T\left(\ell_{i,j}, \widehat{v}_{\mathrm{beg}(i)}, \widehat{v}_i\right) = c_T\left(\ell_{i,j}, \widehat{v}_i, \widehat{v}_{\mathrm{beg}(i)}\right)$). Similarly, there are at most $c_T\left(q_i, \widehat{x}_i, \widehat{y}_i\right)$ subwords corresponding to the the i'th intermediate segment. Thus, if $\alpha = |V(\Omega)| \cdot \left(4t^4\right)^r \cdot (\delta-1)^t$ then

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

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

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

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

$$\begin{aligned}\sum_{\widehat{v}_r \in V(T)} f(r) &= \sum_{\widehat{v}_r \in V(T)} c_T\left(\ell_{r,1}, \widehat{v}_{\mathrm{beg}(r)}, \widehat{v}_r\right) c_T\left(\ell_{r,2}, \widehat{v}_{\mathrm{beg}(r)}, \widehat{v}_r\right) \\ &= \sum_{\widehat{v}_r \in V(T)} c_T\left(\ell_{r,1}, \widehat{v}_{\mathrm{beg}(r)}, \widehat{v}_r\right) c_T\left(\ell_{r,2}, \widehat{v}_r, \widehat{v}_{\mathrm{beg}(r)}\right) \\ &\overset{(*)}{=} c_T\left(\ell_{r,1}+\ell_{r,2}, \widehat{v}_{\mathrm{beg}(r)}, \widehat{v}_{\mathrm{beg}(r)}\right) \leq \rho^{\ell_{r,1}+\ell_{r,2}}.\end{aligned}$$

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

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

□

Finally, we are in position to establish the upper bounds stated in Theorem 3.4.11. Fix $m \in \{1, 2, \ldots, \mathrm{rk}(\Omega)\}$. Then by (3.4.7) and Claim 3.4.13,

$$\begin{aligned}\sum_{w \in \mathcal{CP}_t^m(\Omega)} |\mathrm{Crit}(w)| &\leq \sum_{\substack{N \leq \mathbf{F}_k: \\ \mathrm{rk}(N)=m}} \beta_t(N) \\ &\leq \sum_{\substack{[N] \in \mathrm{ConjCls}(\mathbf{F}_k, m) \\ N \text{ is CR}}} t\beta_t(N) \qquad (3.4.8)\end{aligned}$$

where the final summation is over all conjugacy classes of subgroups of rank m in $\mathbf{F}_k$, and for each class N is a CR representative. Moreover, we choose these representatives N so that if $[N_1]$ and $[N_2]$ correspond the same non-pointed topological graph, the representatives N_1 and N_2 correspond to the same *pointed* topological graph Λ.

Finally, split the summation of the CR representatives N by their topological graph Λ. By Claim 3.4.2, each such Λ has maximal degree at most $2m$ and at most $3m-1$ edges, so by Proposition 3.4.14, the N's corresponding to each Λ contribute to the summation in (3.4.8) at most

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

This finishes the proof of Theorem 3.4.11 as there is a finite number of topological graphs Λ of rank m. □

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

In this section we establish the third step of the proofs of Theorems 3.1.1, 3.1.4, and 3.1.5, as introduced in the overview of the proof (Section 3.2). Recall that according to Theorem 3.2.3, for every $w \in \mathbf{F}_k$ the following holds:

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

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

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

Achieving such a bound requires more elaborated details from the proof of Theorem 3.2.3, which appears in [PP15]. We therefore begin with recalling relevant concepts and results from [PP15]. We then present the proof of Proposition 3.5.1 in Section 3.5.5.

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

Claim 3.5.2. *Let $w \in \left(X \cup X^{-1}\right)^t$ satisfy $\pi(w) = 0$ (so w reduces to 1). If $n > t^2$ then*

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

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

3.5.1 The partial order "covers"

In Section 3.3.2 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 3.3.3). A special role is played by *surjective* morphisms of core graphs:

Definition 3.5.3. Let $H \leq J \leq \mathbf{F}_k$. Whenever the morphism $\eta^X_{H \to J} : \Gamma_X(H) \to \Gamma_X(J)$ is surjective, we say that $\Gamma_X(H)$ *covers* $\Gamma_X(J)$ or that $\Gamma_X(J)$ *is a quotient of* $\Gamma_X(H)$. As for the groups, we say that H *X-covers* J and denote this by $H \leq_{\vec{x}} J$. $H \leq_{\vec{x}} J$

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

For instance, $H = \langle x_1 x_2 x_1^{-3}, x_1^2 x_2 x_1^{-2} \rangle \leq \mathbf{F}_k$ X-covers the group $J = \langle x_2, x_1^2, x_1 x_2 x_1 \rangle$, the corresponding core graphs of which are the leftmost and rightmost graphs in Figure 3.5.1. As another example, a core graph Γ X-covers $\Gamma_X(\mathbf{F}_k)$ (which is merely a wedge of k loops) if and only if it contains edges of all k labels.

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

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

Claim 3.5.4. *If $H \leq \mathbf{F}_k$ is a f.g. subgroup then it X-covers only a finite number of groups. In particular, the partial order "$\leq_{\vec{x}}$" restricted to f.g. subgroups of $\mathbf{F}_k$ is locally finite.*

Proof. The claim follows from the fact that $\Gamma_X(H)$ is finite (Claim 3.3.2(1)) and thus has only finitely many quotients. Each quotient corresponds to a single group, by (3.3.2). □

3.5.2 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 $\eta^X_{H\to J}$). However, not every partition P of $V(\Gamma_X(H))$ corresponds to a quotient core-graph. Indeed, Δ, the graph we obtain after merging the vertices grouped together in P, might not be a core-graph: two distinct j-edges may have the same origin or the same terminus. (For a combinatorial description of core-graphs see e.g. [Pud14, Claim 2.1].) Then again, when a partition P of $V(\Gamma_X(H))$ yields a quotient which is not a core-graph, we can perform Stallings foldings[†] until we obtain a core graph. We denote the resulting core-graph by[‡] $\Gamma_X(H)/P$. Since Stallings foldings do not affect π^X_1, this core graph $\Gamma_X(H)/P$ is $\Gamma_X(J)$, where $J=\pi^X_1(\Delta)$. The resulting partition $\bar{P}$ of $V(\Gamma_X(H))$ (the blocks of which are the fibers of $\eta^X_{H\to J}$) 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 3.5.1. $\Gamma_X(H)/P$

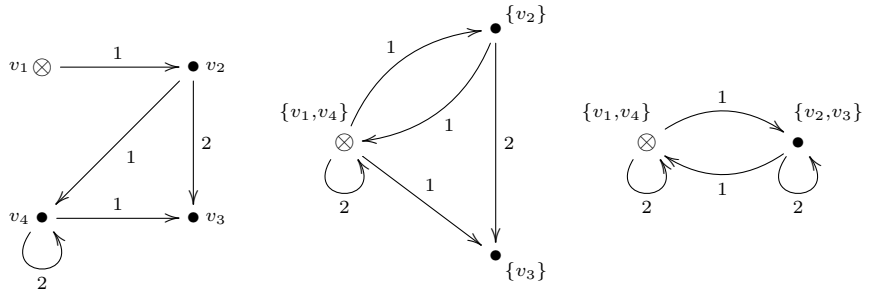


Figure 3.5.1: The left graph is the core graph $\Gamma_X(H)$ of $H=\langle x_1x_2x_1^{-3}, x_1^2x_2x_1^{-2}\rangle \le \mathbf{F}_2$. Its vertices are denoted by $v_1,\ldots,v_4$. The graph in the middle is the quotient corresponding to the partition $P=\{\{v_1,v_4\},\{v_2\},\{v_3\}\}$. This is not a core graph as there are two 1-edges originating at $\{v_1,v_4\}$. In order to obtain a core quotient-graph, we use the Stallings folding process and identify these two 1-edges and their termini. The resulting core graph, $\Gamma_X(H)/P$, is shown on the right and corresponds to the partition $\bar{P}=\{\{v_1,v_4\},\{v_2,v_3\}\}$.

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

$$\|P\| \stackrel{def}{=} |\mathcal{X}| - |P| = \sum_{B\in P}(|B|-1). \tag{3.5.1}$$

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

Definition 3.5.5. Let $H,J\le_{fg}\mathbf{F}_k$ be subgroups such that $H\le_{\vec{x}} J$, and let $\Gamma=\Gamma_X(H)$, $\Delta=\Gamma_X(J)$ be the corresponding core graphs. We define the *X-distance* between H and J, denoted $\rho_X(H,J)$ or $\rho(\Gamma,\Delta)$ as $\rho_X(H,J)$

$$\rho_X(H,J)=\min\left\{\|P\| \;\middle|\; \begin{array}{c} P \text{ is a partition of } V(\Gamma_X(H)) \\ \text{s.t. } \Gamma_X(H)/P=\Gamma_X(J)\end{array}\right\}. \tag{3.5.2}$$

[†]A folding means merging two equally-labeled edges with the same origin or with the same terminus. See also Figure 3.5.1. For a fuller description of Stallings foldings we refer the reader to [Pud14, PP15].

[‡]In [PP15], the notation $\Gamma_X(H)/P$ was used to denote something a bit different (the unfolded graph Δ).

For example, the rightmost core graph in Figure 3.5.1 is a quotient of the leftmost one, and the distance between them is 1. For a more geometric description of this distance function, as well as more details and further examples, we refer the readers to [Pud14, PP15].

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

Theorem 3.5.6. *[[Pud14],Theorem 1.1 and Lemma 3.3] Let $H, J \leq_{fg} \mathbf{F}_k$ such that $H \leq_{\vec{x}} J$. Then*

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

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

This theorem is used, in particular, in the proof in [PP15] of Theorem 3.2.3.

So far the partitions considered here were partitions of the vertex set

$V(\Gamma_X(H))$. However, it is also possible to identify (merge) different *edges* in $\Gamma_X(H)$, as long as they share the same label, and then, as before, perform the folding process to obtain a valid core graph. Moreover, it is possible to consider several partitions $P_1, \ldots, P_r$, each one *either* of the vertices *or* of the edges of $\Gamma_X(H)$, identify vertices and edges according to these partitions and then fold. We denote the resulting core graph by ${}^{\Gamma_X(H)}\!/_{\langle P_1,\ldots,P_r\rangle}$. It is easy to see that one 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: ${}^{\Gamma_X(H)}\!/_{\langle P_1,\ldots,P_r\rangle}$

$$\rho_X(H,J) = \min\left\{ \|P_1\| + \ldots + \|P_r\| \;\middle|\; \begin{array}{c} P_i: \text{ a partition of } V(\Gamma_X(H)) \text{ or of } E(\Gamma_X(H)) \\ \text{s.t. } {}^{\Gamma_X(H)}\!/_{\langle P_1,\ldots,P_r\rangle} = \Gamma_X(J) \end{array} \right\}. \quad (3.5.3)$$

3.5.3 From random elements of S_n to random subgroups

Recall that Theorem 3.2.3 estimates $\mathbb{E}[\mathcal{F}_{w,n}]$, the expected number of fixed points of $w(\sigma_1,\ldots,\sigma_k)$, where $\sigma_1,\ldots,\sigma_k \in S_n$ are chosen independently at random in uniform distribution. The first step in its proof consists of a *generalization of the problem to subgroups:*

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

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

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

$$\mathbb{E}[\mathcal{F}_{w,n}] = \Phi_{\langle w\rangle,\mathbf{F}_k}(n). \quad (3.5.4)$$

3.5.4 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 $[x,y]_{\preceq} \overset{def}{=} \{z \mid x \preceq z \preceq y\}$ is finite). Here we skip the general definition and define these inversions directly in the special case of interest (for a more general point of view see [PP15]).

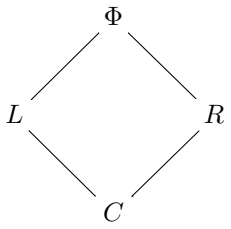
In our case, the poset in consideration is $\mathfrak{sub}_{fg}(\mathbf{F}_k) = \{H \leq \mathbf{F}_k \mid H \text{ is f.g.}\}$, and the partial order is $\leq_{\vec{x}}$, which is indeed locally-finite (Claim 3.5.4). We define three derivations of the function Φ defined in Section 3.5.3: the left one (L), the right one (R) and the two-sided one (C). These are usually formally defined by convolution of Φ with the Möbius function of $\mathfrak{sub}_{fg}(\mathbf{F}_k)_{\leq_{\vec{x}}}$ (see [PP15])

but here we define them in an equivalent simpler way: these are the functions satisfying, for every $H \leq_{\vec{x}} J$,

$$\Phi_{H,J}(n) = \sum_{M \in [H,J]_{\vec{x}}} L_{M,J}(n) = \sum_{M,N:\, H \leq_{\vec{x}} M \leq_{\vec{x}} N \leq_{\vec{x}} J} C_{M,N}(n) = \sum_{N \in [H,J]_{\vec{x}}} R_{H,N}(n). \tag{3.5.5}$$

Note that the summations in (3.5.5) are well defined because the order is locally finite. To see that (3.5.5) can indeed serve as the definition for the three new functions, use induction on $|[H,J]|$: for example, for any $H \leq_{\vec{x}} J$, $L_{H,J}(n) = \Phi_{H,J}(n) - \sum_{M \in [H,J)_{\vec{x}}} L_{M,J}(n)$ and all pairs (M,J) on the r.h.s. satisfy $|[M,J]| < |[H,J]|$.

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



Proposition 3.5.7 ([PP15], Proposition 5.1)**.** *The function R is supported on algebraic extensions.*

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

$$\Phi_{H,J}(n) = \sum_{N:\, H \leq_{alg} N \leq J} R_{H,N}(n). \tag{3.5.6}$$

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

Claim 3.5.8. *Let $H, J \leq \mathbf{F}_k$ be f.g. subgroups such that $H \leq_{\vec{x}} J$. 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.*

After some involved combinatorial arguments, one obtains from this the following expression for $C_{M,N}(n)$: Denote by $\mathrm{Sym}(S)$ the set of permutations of a given set S. Every permutation $\sigma \in \mathrm{Sym}(S)$ defines, in particular, a partition on S whose blocks are the cycles of σ. By abuse of notation we denote by σ both the permutation and the corresponding partition. For instance, one can consider its "norm" $\|\sigma\|$ (see (3.5.1); this is also the minimal length of a product of transpositions that gives the permutation σ). We also use V_M and E_M as short for $V(\Gamma_X(M))$ and $E(\Gamma_X(M))$, respectively. V_M, E_M

Proposition 3.5.9 ([PP15], Section 7.1)**.** *Let $M, N \leq \mathbf{F}_k$ be f.g. subgroups with $M \leq_{\vec{x}} N$. Consider the set*

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

Then

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

The derivation of the main result of [PP15] (Theorem 3.2.3) from Theorem 3.5.6 and Propositions 3.5.7 and 3.5.9 is short: see the beginning of Section 7 in [PP15].

3.5.5 Proving the uniform bound for the error term

We now have all the tools required for proving Proposition 3.5.1. Namely, we now prove that every $1 \neq w \in \mathbf{F}_k$ of length t and every $n > t^2$,

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

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

Proof. [of Proposition 3.5.1] Recall (Section 3.5.3) that $\mathbb{E}\left[\mathcal{F}_{w,n}\right] = \Phi_{\langle w\rangle,\mathbf{F}_k}\left(n\right)$ and this quantity is given by some rational expression in n (for large enough n, say $n \geq |w|$, see Claim 3.5.8). This expression can be expressed as a Taylor series in $\frac{1}{n}$, so write

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

where $a_s\left(w\right) \in \mathbb{R}$ (in fact these are integers: see [Pud14, Claim 5.1] and also the sequel of the current proof). By Theorem 3.2.3, $a_0 = 1$, $a_1 = a_2 = \ldots = a_{\pi(w)-2} = 0$ and $\alpha_{\pi(w)-1} = |\mathrm{Crit}\left(w\right)|$ (unless $\pi\left(w\right) = 1$ in which case $a_0 = 1 + |\mathrm{Crit}\left(w\right)|$). So our goal here is to bound the remaining coefficients $a_s\left(w\right)$ for $s \geq \pi\left(w\right)$.

The discussion in Section 3.5.4 yields the following equalities:

$$\begin{aligned}\mathbb{E}\left[\mathcal{F}_{w,n}\right] &= \Phi_{\langle w\rangle,\mathbf{F}_k}\left(n\right) = \sum_{N\colon \langle w\rangle \leq_{alg} N \leq \mathbf{F}_k} R_{\langle w\rangle,N}\left(n\right) = \\ &= \sum_{M,N\colon \langle w\rangle \leq_{\vec{X}} M \leq_{\vec{X}} N} C_{M,N}\left(n\right) = \sum_{M\colon \langle w\rangle \leq_{\vec{X}} M} \sum_{N\colon M \leq_{\vec{X}} N} C_{M,N}\left(n\right)\end{aligned}$$

From Proposition 3.5.9 we obtain that for a fixed M,

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

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

$$\mathcal{P}_{M,q} = \left\{ \left(\sigma_0,\ldots,\sigma_r\right) \middle| \begin{array}{c} r \in \mathbb{N},\ \sigma_0 \in \mathrm{Sym}\left(V_M\right) \\ \sigma_1,\ldots,\sigma_r \in \mathrm{Sym}\left(E_M\right)\setminus\{\mathrm{id}\} \\ \|\sigma_0\| + \ldots + \|\sigma_r\| = q \end{array} \right\}, \tag{3.5.7}$$

so that

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

Hence,

$$a_s\left(w\right) = \sum_{i=1}^{s+1} \sum_{\substack{M\colon \langle w\rangle \leq_{\vec{X}} M \\ \mathrm{rk}(M)=i}} \left(-1\right)^{s-(i-1)} \sum_{(\sigma_0,\ldots,\sigma_r)\in\mathcal{P}_{M,s-(i-1)}} \left(-1\right)^r. \tag{3.5.8}$$

In what follows we ignore the alternating signs of the summands in (3.5.8) and bound $|a_s\left(w\right)|$ by

$$|a_s\left(w\right)| \leq \sum_{i=1}^{s+1} \sum_{\substack{M\colon \langle w\rangle \leq_{\vec{X}} M \\ \mathrm{rk}(M)=i}} \left|\mathcal{P}_{M,s-(i-1)}\right|. \tag{3.5.9}$$

Claim: For every $M \leq_{fg} \mathbf{F}_k$ with $\langle w \rangle \leq_{\vec{x}} M$, we have $|\mathcal{P}_{M,q}| \leq t^{2q}$.

Proof of Claim: Fix M and denote $b_q = |\mathcal{P}_{M,q}|$. Clearly, $b_0 = 1$, and we proceed by induction on q. Let $q \geq 1$. We split the set $\mathcal{P}_{M,q}$ by the value of σ_r. For $r = 0$ there are at most

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

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

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

We proceed with the proof of the proposition. For a given $w \in \left(X \cup X^{-1}\right)^t$ there are at most $\binom{|V_{\langle w \rangle}|}{2}^\beta \leq \binom{t}{2}^\beta$ partitions of norm β of $V_{\langle w \rangle}$, and so at most $\binom{t}{2}^\beta$ subgroups M of rank β with $\langle w \rangle \leq_{\vec{x}} M$ (see Theorem 3.5.6). Hence from (3.5.9) we obtain,

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

Finally,

$$\begin{aligned} \left|\mathbb{E}\left[\mathcal{F}_{w,n}\right] - 1 - \frac{|\mathrm{Crit}\,(w)|}{n^{\pi(w)-1}}\right| &= \left|\sum_{s=\pi(w)}^{\infty} \frac{a_s(w)}{n^s}\right| \leq \sum_{s=\pi(w)}^{\infty} \frac{|a_s(w)|}{n^s} \\ &\leq \sum_{s=\pi(w)}^{\infty} \frac{t^{2s+2}}{n^s} = t^2 \cdot \left(\frac{t^2}{n}\right)^{\pi(w)} \cdot \frac{n}{n-t^2}. \end{aligned}$$

This finishes the proof. $\square$

3.6 Completing the Proof for Regular Graphs

In this section we complete the proofs of Theorems 3.1.1 and 3.1.5. In addition, we explain (in Section 3.6.4) the source of the gap between these results on the one hand and Friedman's result and Conjecture 3.1.3 on the other.

3.6.1 Proof of Theorem 3.1.1 for d even

We begin with the case of even d in Theorem 3.1.1. We show that a random d-regular graph Γ on n vertices in the permutation model (a random n-covering of the bouquet with $\frac{d}{2}$ loops) satisfies a.a.s. $\lambda(\Gamma) < 2\sqrt{d-1} + 0.84$, where $\lambda(\Gamma)$ is the largest non-trivial eigenvalue of A_Γ. As explained in more details in Appendix 3.A, this yields the same result for a uniformly random d-regular simple graph.

So let $d = 2k$ and $n, t = t(n)$ be such that $n > t^2$ and t is even. The base graph Ω is the bouquet with k loops, so $\mathcal{CP}_t(\Omega) = \left(X \cup X^{-1}\right)^t$. By (3.2.2), Proposition 3.5.1 and Claim 3.5.2,

$$\begin{aligned}
\mathbb{E}\left[\lambda(\Gamma)^t\right] &\leq \sum_{w \in (X \cup X^{-1})^t} \left(\mathbb{E}\left[\mathcal{F}_w\right] - 1\right) = \\
&= \sum_{m=0}^{k} \sum_{\substack{w \in \left(X \cup X^{-1}\right)^t: \\ \pi(w) = m}} \left(\frac{|\mathrm{Crit}(w)|}{n^{m-1}} + O\left(\frac{1}{n^m}\right)\right) \\
&\leq \sum_{m=0}^{k} \frac{1}{n^{m-1}} \sum_{\substack{w \in \left(X \cup X^{-1}\right)^t: \\ \pi(w) = m}} \left(|\mathrm{Crit}(w)| + \frac{t^{2+2m}}{n - t^2}\right) \\
&\leq \left(1 + \frac{t^{2+2k}}{n - t^2}\right) \sum_{m=0}^{k} \frac{1}{n^{m-1}} \sum_{\substack{w \in \left(X \cup X^{-1}\right)^t: \\ \pi(w) = m}} |\mathrm{Crit}(w)|
\end{aligned}$$

Let $\varepsilon > 0$. For $m \in \{0, 1, \ldots, k\}$, Corollary 3.4.5 (for $m \geq 1$) and Claim 3.4.7 (for $m = 0$) yield that for large enough t,

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

where $g(\cdot)$ is defined as in (3.4.4):

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

Thus

$$\begin{aligned}
\mathbb{E}\left[\lambda(\Gamma)^t\right] &\leq \left(1 + \frac{t^{2+2k}}{n - t^2}\right) \sum_{m=0}^{k} \frac{\left[g(2m-1) + \varepsilon\right]^t}{n^{m-1}} \\
&\leq \left(1 + \frac{t^{2+2k}}{n - t^2}\right) \cdot (k+1) \cdot \\
&\quad \cdot \left[\max \left\{ \begin{array}{c} n^{1/t}\left[g(-1) + \varepsilon\right], g(1) + \varepsilon, \frac{g(3)+\varepsilon}{n^{1/t}} \cdots \\ \ldots, \frac{g(2k-3)+\varepsilon}{\left(n^{1/t}\right)^{k-2}}, \frac{2k+\varepsilon}{\left(n^{1/t}\right)^{k-1}} \end{array} \right\}\right]^t
\end{aligned} \tag{3.6.1}$$

Recall that Γ is a random graph on n vertices. In order to obtain the best bound, t needs to be chosen to minimize the maximal summand in the r.h.s. of (3.6.1). 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 $n^{1/t} \approx c$, then all $k+1$ elements in the set in the r.h.s. of (3.6.1) are strictly less than $2\sqrt{d-1} + 0.835$ (for small enough ε). Thus, for large enough t, $\mathbb{E}\left[\lambda(\Gamma)^t\right] \leq \left[2\sqrt{d-1} + 0.835\right]^t$. A standard application of Markov's inequality then shows that $\mathrm{Prob}\left[\lambda(\Gamma) < 2\sqrt{d-1} + 0.84\right] \underset{n \to \infty}{\to} 1$.

Indeed, for $d \geq 26$, one can set $n^{1/t} = e^{\frac{2}{5\sqrt{d-1}}}$. Simple analysis shows that for $d \geq 26$, $e^{\frac{2}{5\sqrt{d-1}}} < 1 + \frac{5}{12\sqrt{d-1}}$, so the element corresponding to $m = 0$ is at most $2\sqrt{d-1} \cdot e^{\frac{2}{5\sqrt{d-1}}} <$

$2\sqrt{d-1}\left(1+\frac{5}{12\sqrt{d-1}}\right)=2\sqrt{d-1}+\frac{5}{6}<2\sqrt{d-1}+0.835$. This first element is clearly larger than all other elements corresponding to m such that $2m-1\leq\sqrt{d-1}$. Among all other values of m, the maximal element is obtained when $2m-1\approx 4.55\sqrt{d-1}$, but its value is bounded from above by $1.94\sqrt{d-1}+0.4$ (again, by simple analysis). For all remaining $d's$ $(4,6,\ldots,24)$, it can be checked case by case that choosing $n^{1/t}$ so that $n^{1/t}\cdot 2\sqrt{d-1}=2\sqrt{d-1}+0.8$ works (and see the table in Section 3.6.2). □

3.6.2 From even d to odd d

In this subsection we derive the statement of Theorem 3.1.1 for d odd from the now established statement for d even. We showed that for d even we have a.a.s. $\lambda(\Gamma)<2\sqrt{d-1}+0.84$. The idea is that every upper bound applying to some value of d also applies to $d-1$.

As explained in Appendix 3.A, by contiguity results from [GJKW02], it is enough to show the $2\sqrt{d-1}+1$ upper bound for random graphs Γ in a random model denoted $\mathcal{G}^*_{n,d}$ (the result for random simple graphs than follows immediately).

Claim 3.6.1. *Let $d\geq 3$ be odd. Assume that a random $(d+1)$-regular graph Γ in the permutation model satisfies a.a.s. $\lambda(\Gamma)<C$. Then a random d-regular graph Γ in $\mathcal{G}^*_{n,d}$ also satisfies a.a.s. $\lambda(\Gamma)<C$.*

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

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

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

where the summation is over all edges e in the random perfect matching m, and e^+ and e^- mark the two endpoints of e. Let R denote the random summation $2\sum_{e\in m}f(e^+)f(e^-)$. 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,

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

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

Similarly,

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

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

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

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

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

For smaller values of odd d we use tighter results for $d+1$. For example, we seek the smallest constant c for which a bound of $2\sqrt{4-1} + c$ can be obtained for 4-regular graphs in our methods. In order to minimize $\max\left\{n^{1/t} \cdot 2\sqrt{d-1}, 2\sqrt{d-1}, \frac{4}{n^{1/t}}\right\}$ (see (3.6.1)), we choose $n^{1/t} = \sqrt{\frac{4}{2\sqrt{d-1}}}$ to get an upper bound of 3.723 (compared with $2\sqrt{d-1} = 3.464$, so here $c \approx 0.259$). For $d = 3$ this bound is useless (it is larger than the trivial bound of 3).

The following table summarizes the bounds we obtain for $d \leq 20$ in the scenario of Theorem 1. This can be carried on to establish Theorem 3.1.1 for $d \leq 40$.

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

Remark 3.6.2. Of course, the method presented here to derive the statement of Theorem 3.1.1 for odd d's from the statement for even d's works only because of the small additive constant we have in the result. To obtain a tight result (Friedman's Theorem) in our approach, we will need another method to work with odd d's.

One plausible direction is as follows. We may construct a random d-regular graph with d odd using $k = \frac{d-1}{2}$ random permutations plus one random perfect matching. If we label the edges corresponding to the perfect matching by b, and orient the edges corresponding to the permutations and label them by $a_1, \ldots, a_k$, the graphs become Schreier graphs of subgroups of $\mathbf{F}_k * \mathbb{Z}/2\mathbb{Z} = \left\langle a_1, \ldots, a_k, b \,\middle|\, b^2 = 1 \right\rangle$. It is conceivable that the machinery we developed for the free group (and especially, Theorem 3.2.3) can be also developed for this kind of free products.

3.6.3 Proof of Theorem 3.1.5

The only change upon the previous case (Theorem 3.1.1 with d even) is that the summation in (3.6.1) over the primitivity rank m does not stop at $k = \frac{d}{2}$ but continues until $\mathrm{rk}(\Omega) = |V(\Omega)|\left(\frac{d}{2} - 1\right) + 1$. However, when $m > k$, it follows from Corollary 3.4.10 that the corresponding term inside the max operator is $\frac{d}{\left(n^{1/t}\right)^{m-1}}$ which is strictly less than $\frac{d}{\left(n^{1/t}\right)^{d/2-1}}$ (for every choice of t and n), but this latter term is already there in (3.6.1). Thus, the maximal term is remained unchanged, and we obtain the same bound overall as in the even case of Theorem 3.1.1, 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 $\sqrt{3 \cdot 2\sqrt{d-1}} \approx 2.913$.

3.6.4 The source of the gap

It could be desirable to use the approach presented in this paper and replace the constant 1 in Theorem 3.1.1 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 3.2) is in bounding $\left\{\mathbb{E}\left[\lambda\left(\Gamma\right)^t\right]\right\}^{1/t}$ by $\left\{\mathbb{E}\left[\sum_{\mu\in\mathrm{Spec}(A_\Gamma)\setminus\{d\}}\mu^t\right]\right\}^{1/t}$. As long as $t=\theta\left(\log n\right)$, our loss here is bounded, and if $t\gg\log n$ we lose nothing. This is because

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

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 Γ satisfy $|\mu|\geq\rho-\delta$ (here $\rho=2\sqrt{d-1}$). If $t\in o\left(\log n\right)$ then $n^{1/t}$ tends to infinity, and thus

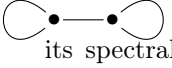
$$\left\{\sum_{\mu\in\mathrm{Spec}(A_\Gamma)\setminus\{d\}}\mu^t\right\}^{1/t}>\left\{\varepsilon n\left(\rho-\delta\right)^t\right\}^{1/t}\underset{n\to\infty}{\to}\infty.$$

Our proof proceeds by bounding this t-th moment of the non-trivial spectrum. Let us stress that as long as $t=t\left(n\right)$ is small enough in terms of n so that the error term in Proposition 3.5.1 is negligible ($t=o\left(n^{1/(2+2k)}\right)$ suffices), the upper bound our technique yields for $\mathbb{E}\left[\sum_{\mu\in\mathrm{Spec}(A_\Gamma)\setminus\{d\}}\mu^t\right]$ is tight. In particular, for large enough d, and $t\approx c\log n$ with a suitable constant $c=c\left(d\right)$,

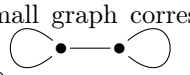
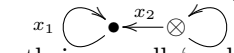
$$\left\{\mathbb{E}\left[\sum_{\mu\in\mathrm{Spec}(A_\Gamma)\setminus\{d\}}\mu^t\right]\right\}^{1/t}\approx2\sqrt{d-1}+0.84.$$

To see why, note that all relevant steps of the proof yield equalities or tight bounds: the second step, which relies on Theorem 3.2.3, has only equalities so it is surely tight. In the third step, we prove that the error term is $o_n\left(1\right)$ 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 $\sum_{w\in(X\cup X^{-1})^t:\,\pi(w)=m}\left|\mathrm{Crit}\left(w\right)\right|$ is, in fact, the correct value (see Theorem 3.8.5). 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.

What is, then, the source of this gap? It seems, therefore, that the reason the bound we get for $\lambda\left(\Gamma\right)$ is not tight lies in rare events that enlarge $\mathbb{E}\left[\lambda\left(\Gamma\right)^t\right]$ substantially. For example, in the permutation model every vertex of Γ is isolated with probability $\frac{1}{n^k}$, so overall there are on average $\frac{1}{n^{k-1}}$ isolated vertices. Each such vertex is responsible to an additional eigenvalue d, alongside the trivial one. These rare events alone contribute $\frac{1}{n^{k-1}}\cdot d^t$ to $\mathbb{E}\left[\lambda\left(\Gamma\right)^t\right]$. For example, for $d=4$ ($k=2$) and $n^{1/t}\approx1.075$ as in the table in Section 3.6.2, isolated vertices contribute about $\left[\frac{4}{(1.075)}\right]^t\approx3.721^t$ to $\mathbb{E}\left[\lambda\left(\Gamma\right)^t\right]$, which is roughly the bound we obtain in this case.

There are other, slightly more complicated, rare events that contribute much to $\mathbb{E}\left[\lambda\left(\Gamma\right)^t\right]$. Consider, for instance, the event that when $d=4$ the random graph Γ contains the subgraph  . If this subgraph is completed to a 4-regular graph by attaching a tree to each vertex, its spectral radius becomes 3.5. Since this resulting graph topologically covers (the connected component of the subgraph in) Γ, we get a non-trivial eigenvalue which is at least 3.5 (but normally

very close to 3.5). On average, there are $\frac{2}{n}$ such subgraphs in Γ, so they contribute about $\frac{2}{n} \cdot 3.5^t$ to $\mathbb{E}\left[\lambda(\Gamma)^t\right]$. When $n^{1/t}$ is small enough, this is strictly larger than $\left[2\sqrt{d-1}\right]^t \approx 3.464^t$.

Each such small graph corresponds to a few particular subgroups of $\mathbf{F}_k$. For example, the subgraph  corresponds to one of four subgroups, one of which is 

.One therefore needs to realize which are all these 'bad' subgroups, show their overall 'probability' is small (the average number of appearances of $H \leq \mathbf{F}_k$ in Γ is exactly $L_{H,\mathbf{F}_k}$), and somehow omit their contribution to $\mathbb{E}\left[\lambda(\Gamma)^t\right]$. This would be relatively easy if our analysis of $\mathbb{E}[\mathcal{F}_w]$ were based on $\mathbb{E}[\mathcal{F}_w] = \sum_{M \in [\langle w \rangle, \mathbf{F}_k]_{\overrightarrow{X}}} L_{M,\mathbf{F}_k}$, but it is based, instead, on $\mathbb{E}[\mathcal{F}_w] = \sum_{N \in [\langle w \rangle, \mathbf{F}_k]_{\overrightarrow{X}}} R_{\langle w \rangle, N}$ (see Section 3.5). 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 3.6.3. These 'bad', rare events are parallel to the notion of *tangles* in [Fri08].

3.7 Completing the Proof for Arbitrary Graphs

The completion of the proof of Theorem 3.1.4 is presented in this Section. We begin with the proof of the first statement of the theorem which concerns the spectrum of the adjacency operator of Γ, the random n-covering of the fixed base graph Ω. The variations needed in order to establish the statement about the Markov operator are described in Section 3.7.1.

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

$$\lambda_A(\Gamma) < \sqrt{3} \cdot \rho + \varepsilon. \tag{3.7.1}$$

As in the proof of Theorem 3.1.1 (the beginning of Section 3.6), let $n, t = t(n)$ be so that $n > t^2$ and t is even. Using (3.2.2), Proposition 3.5.1, Claim 3.5.2 and Lemma 3.4.8, one obtains

$$\begin{aligned}
\mathbb{E}\left[\lambda_A(\Gamma)^t\right] &\leq \sum_{w \in \mathcal{CP}_t(\Omega)} (\mathbb{E}[\mathcal{F}_w] - 1) = \\
&\leq \left(1 + \frac{t^{2+2\,\mathrm{rk}(\Omega)}}{n - t^2}\right) \sum_{m=0}^{\mathrm{rk}(\Omega)} \frac{1}{n^{m-1}} \sum_{w \in \mathcal{CP}_t^m(\Omega)} |\mathrm{Crit}(w)|
\end{aligned}$$

Let $\varepsilon > 0$. From Theorem 3.4.11 and Lemma 3.4.12 it follows now that for t even and large enough,

$$\begin{aligned}
\mathbb{E}\left[\lambda_A(\Gamma)^t\right] &\leq \left(1 + \frac{t^{2+2\,\mathrm{rk}(\Omega)}}{n - t^2}\right) \left[n \cdot [\rho + \varepsilon]^t + \sum_{m=1}^{\mathrm{rk}(\Omega)} \frac{[(2m-1) \cdot \rho + \varepsilon]^t}{n^{m-1}}\right] \cdot \\
&\leq \left(1 + \frac{t^{2+2\,\mathrm{rk}(\Omega)}}{n - t^2}\right) (1 + \mathrm{rk}(\Omega)) \cdot \\
&\quad \cdot \left[\max\left\{n^{1/t}[\rho + \varepsilon], \rho + \varepsilon, \frac{3\rho + \varepsilon}{n^{1/t}}, \frac{5\rho + \varepsilon}{\left(n^{1/t}\right)^2}, \ldots, \frac{(2\,\mathrm{rk}(\Omega) - 1)\,\rho + \varepsilon}{\left(n^{1/t}\right)^{\mathrm{rk}(\Omega)-1}}\right\}\right]^t
\end{aligned} \tag{3.7.2}$$

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 $\left(1 + \frac{t^{2+2\,\mathrm{rk}(\Omega)}}{n-t^2}\right)^{1/t} \underset{n \to \infty}{\to} 1$, and the maximal value inside the set in (3.7.2) is then $\sqrt{3}(\rho + \epsilon)$. Again, a standard application of Markov inequality finishes the proof. □

3.7.1 The spectrum of the Markov operator

After establishing the first statement of Theorem 3.1.4, 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 Γ. The goal is to show that for every $\varepsilon > 0$

$$\lambda_M(\Gamma) < \sqrt{3} \cdot \rho_M(\Omega) + \varepsilon \tag{3.7.3}$$

asymptotically almost surely.

As we note in Appendix 3.B, the Markov operator is given by $B_\Gamma D_\Gamma^{-1}$, where B_Γ is the adjacency matrix and D_Γ the diagonal matrix with the degrees of vertices in the diagonal. This is conjugate to and thus share the same spectrum with $Q_\Gamma = D_\Gamma^{-1/2} B_\Gamma D_\Gamma^{-1/2}$, but the latter has the advantage of being symmetric, so we work with it.

The (u,v) entry of Q_Γ equals $\frac{1}{\sqrt{\deg(u)\deg(v)}}$ times the number of edges between u and v. For every path w in Γ 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 $f(w) = \frac{1}{\sqrt{\deg v_0}\cdot \deg v_1 \cdot \ldots \cdot \deg v_{t-1} \cdot \sqrt{\deg v_t}}$. It is easy to see that $[Q_\Gamma^t]_{u,v}$ equals the sum of $f(w)$ over all paths w of length t from u to v, and thus

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

Moreover, note that when a path from the covering Γ projects to the base graph Ω, its weight does not change. Using this fact, we can imitate step I from Section 3.2 to obtain, for t even,

$$\begin{aligned} \lambda_M(\Gamma)^t &\leq \sum_{\mu \in \mathrm{Spec}(M_\Gamma)} \mu^t - \sum_{\mu \in \mathrm{Spec}(M_\Omega)} \mu^t = \sum_{w \in \mathcal{CP}_t(\Gamma)} f(w) - \sum_{w \in \mathcal{CP}_t(\Omega)} f(w) = \\ &= \sum_{w \in \mathcal{CP}_t(\Omega)} f(w) \left[\mathcal{F}_{w,n}(\sigma_1, \ldots, \sigma_k) - 1\right]. \end{aligned}$$

The second and third steps remain the same, obtaining

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

The next modification needs take place in the fourth step, where instead of bounding $\sum_{w \in \mathcal{CP}_t(\Omega):\, \pi(w) = m} |\mathrm{Crit}(w)|$, one needs to bound $\sum_{w \in \mathcal{CP}_t(\Omega):\, \pi(w) = m} f(w) |\mathrm{Crit}(w)|$. But the exact same proofs work if we merely replace $\rho_A(\Omega)$ with $\rho_M(\Omega)$. Theorem 3.4.11 becomes

$$\limsup_{t \to \infty} \left[\sum_{w \in \mathcal{CP}_t^m(\Omega)} f(w) \left|\mathrm{Crit}(w)\right|\right]^{1/t} \leq (2m-1) \cdot \rho_M(\Omega). \tag{3.7.4}$$

and likewise, Lemma 3.4.12 becomes

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

Similarly, the definition of $\beta_t(N)$ (preceding Claim 3.4.13) should be modified to

$$\beta_t(N) = \sum_{w \in \mathcal{CP}_t(\Omega):\, \langle w \rangle \lneq_{alg} N} f(w)$$

and in the proof of Claim 3.4.13 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 3.4.14 we sometimes replace a path with its inverse and use the symmetry of the operator. This is the reason for working with Q_Γ rather than with M_Γ. 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 (3.7.4) are exactly those in Theorem 3.4.11 only with $\rho_M(\Omega)$ instead of $\rho_A(\Omega)$, the final step of the proof (which appears in Section 3.7) also remains unchanged.

3.8 The Distribution of Primitivity Ranks

In this subsection we show that the upper bounds from Proposition 3.4.3 and Corollary 3.4.5 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 3.1.1, the fourth step of the proof, where words and critical subgroups are counted, yields a tight bound. Thus, the origin of the gap between our result and Friedman's lies elsewhere (see Section 3.6.4).

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

Theorem 3.8.1. *[PW14] For every $k \geq 3$, let $p_k(t)$ denote the number of primitive words of length t in $\mathbf{F}_k$. Then,*

$$\lim_{t\to\infty} \sqrt[t]{p_k(t)} = 2k-3.$$

For $\mathbf{F}_2$ it is known that this exponential growth rate equals $\sqrt{3}$ ([Riv04]). These results show that the portion of primitive words among all words of length t decays exponentially fast[†]. They are used in the following theorem, which states that the upper bounds from Proposition 3.4.3 are accurate.

Theorem 3.8.2. *Let $k \geq 2$ and $m \in \{1, 2, \ldots, k\}$. Let* $c_{k,m}(t)$

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

Then,

$$\limsup_{t\to\infty} c_{k,m}(t)^{1/t} = \begin{cases} \sqrt{2k-1} & 2m-1 \leq \sqrt{2k-1} \\ 2m-1 & 2m-1 \geq \sqrt{2k-1} \end{cases}. \tag{3.8.1}$$

In fact, as the proof shows, for $m \geq 2$, we can replace the lim sup 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 3.8.3. *A generic word in $\mathbf{F}_k$ has primitivity rank k.*

Proof. [of Theorem 3.8.2] The r.h.s. of (3.8.1) is an upper bound for the lim sup by Proposition 3.4.3. It remains to show that for every $m \in \{1, \ldots, k\}$, there is some subset of words with primitivity rank m and growth rate $\max\left(\sqrt{2k-1}, 2m-1\right)$.

Consider first the case $2m-1 > \sqrt{2k-1}$. Take any subset of the generators $S \subseteq X$ of size m and consider the subgroup $H = \mathbf{F}(S)$. Its core graph is a bouquet of m loops. The number of words of length t in H is $2m \cdot (2m-1)^{t-1}$. By Theorem 3.8.1, 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 then $(2m-1)$ (by Proposition 3.4.3).

[†]That primitive words in $\mathbf{F}_k$ are negligible in this sense follows from the earlier results [BV02], [BMS02b, Thm 10.4] and [Shp05], where the exponential growth rate from Theorem 3.8.1 is shown to be $\leq 2k-2-o_k(1)$.

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 $2m-1 \le \sqrt{2k-1}$. Consider subgroups of the form $H = \langle x_1, \dots, x_{m-1}, u\rangle$ where u is a cyclically reduced word of length $\sim \frac{t}{2}$ such that its first and last letters are *not* one of $\{x_1^{\pm 1}, \dots, x_{m-1}^{\pm 1}\}$. Then, $\Gamma_X(H)$ has the form of a bouquet of $m-1$ small loops of size 1 and one large loop of size $\sim \frac{t}{2}$. Now consider the word $w = w(u) = x_1^2 x_2^2 \dots x_{m-1}{}^2 u^2$. Obviously, the growth rate of the number of possible u's (as a function of t) is $\sqrt{2k-1}$, hence also the growth rate of the number of different w's. It can be shown that w is *not* primitive in H, using the primitivity criterion from Theorem 3.5.6 ([Pud14, Thm 1.1]). (In fact, it follows from [PP15, 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, $\pi(w) \le m$. In general, the primitivity rank might be strictly smaller. For example, for $m = 3$ and $u = x_3 x_1^2 x_2^2 x_3$, we have $\pi(w) = 2$ because w is not a proper power yet is not primitive in $\langle x_3, x_1^2 x_2^2\rangle$. However, we claim that for a generic u, the primitivity rank of w is exactly m.

Indeed, if this is not the case, then there is some $\tilde{m} < m$ such that the growth rate of words w as above with $\pi(w) = \tilde{m}$ is $\sqrt{2k-1}$. By the proof of Proposition 3.4.3 and especially (3.4.2) , it follows that most of these words ($w = w(u)$ with $\pi(w) = \tilde{m}$) have an algebraic extension N of rank $\tilde{m}$ such that the number of edges in $\Gamma_X(N)$ is close to $\frac{t}{2}$. (By (3.4.2), the total number of words of length t with an algebraic extension N of rank $\tilde{m}$ and δt edges in $\Gamma_X(N)$, for some $\delta < \frac{1}{2}$, grows strictly slower than $\sqrt{2k-1}$.) So almost all these words $w = w(u)$ trace twice every edge of some $\Gamma_X(N)$ of rank $\tilde{m}$ with roughly $\frac{t}{2}$ edges. In particular, each such $w = w(u)$ traces twice some topological edge in $\Gamma_X(N)$ of length at least $\frac{1}{2(3m-1)}t$. This implies that there is some linear-size two overlapping subwords of u or of u^{-1}. But for a generic u, the longest subword appearing twice in u or in u^{-1} has length of order $\log t$.

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

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

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

Remark 3.8.4. It follows from the proofs of Proposition 3.4.3 and Theorem (3.8.2) that while for $2m-1 > \sqrt{2k-1}$ the main source for words with $\pi(w) = m$ is in subgroups with core graphs of minimal size (and their conjugates), the main source for $2m-1 < \sqrt{2k-1}$ is in subgroups with core graphs of maximal size, namely of size roughly $\frac{t}{2}$.

Recall that in the proof of Theorem 3.1.1 we used bounds on the number of *not-necessarily-reduced* words (and their critical subgroups). Here, too, the bounds from Corollary 3.4.5 are accurate for every value of m:

Theorem 3.8.5. *Let $k \ge 2$ and $m \in \{0, 1, 2, \dots, k, \infty\}$. Let* $b_{k,m}(t)$

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

Then for $m = 0$ we have

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

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

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

Finally, for $m = \infty$ we have

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

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

Proof. For $m = 0$ this is (the proof of) Claim 3.4.7 (evidently, there are no odd-length words reducing to 1). For $1 \le m$ with $2m - 1 \le \sqrt{2k-1}$ the same proof (as in Claim 3.4.7) can be followed as long as we present at least one even-length and one odd-length words with primitivity rank m. And indeed, as mentioned above (and see [Pud14, Lemma 6.8]), $\pi\left(x_1^2 x_2^2 \dots x_m^2\right) = \pi\left(x_1^3 x_2^2 \dots x_m^2\right) = m$. If $2m - 1 > \sqrt{2k-1}$, the statement follows from the statements on reduced words (Theorems 3.8.2 and 3.8.1) and an application of the extended cogrowth formula [Pud15b] (here a bit more elaborated results from [Pud15b], not mentioned in Theorem 3.4.4, are required). □

The result of the last theorem are summarized in Table 3.2.

3.9 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 3.2.3? This would probably require not ignoring the alternating signs in (3.5.8). As explained in Section 3.6.4, this may be the seed to closing the gap in the result of Theorem 3.1.1.
- Is it possible to generalize the techniques in this paper (and even more so the ones from [PP15]) to odd values of d? (See Remark 3.6.2).
- 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 3.4.11 and 3.1.4? 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

3.A Contiguity and Related Models of Random Graphs

Random d-regular graphs

In this paper, the statement of Theorem 3.1.1 is first proved for the permutation model of random d-regular graphs with d even. We then derive Theorem 3.1.1, 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 80) of different models of random regular graphs.

In particular, they describe the following model: consider dn labeled points, with d points in each of n buckets, and take a random perfect matching of the points. Letting the buckets be vertices and each pair represent an edge, one obtains a random d-regular graph. This model is denoted $\mathcal{G}^*_{n,d}$. It is shown [GJKW02, Theorem 1.3] that $\mathcal{G}^*_{n,d}$ is contiguous to the permutation model $\mathcal{P}_{n,d}$ (for d even). If Γ is a random d-regular graph in $\mathcal{G}^*_{n,d}$, the event that Γ is a simple graph (with no loops nor multiple edges) has positive probability, bounded away from 0. Moreover, within this event, simple graphs are distributed uniformly[†]. Thus, for even values of d, Theorem 3.1.1 follows from the corresponding result for the permutation model. The derivation of the odd case also uses contiguity results, as explained in Section 3.6.2.

Random d-regular bipartite graphs

As an immediate corollary from Theorem 3.1.5 we deduced that a random d-regular bipartite graph is "nearly Ramanujan" in the sense that besides its two trivial eigenvalues $\pm d$, all other eigenvalues are at most $2\sqrt{d-1}+0.84$ in absolute value a.a.s. (Corollary 3.1.6). Our proof works in the model $C_{n,\Omega}$ (here Ω is the graph with 2 vertices and d parallel edges connecting them). However, by the results of [Ben74], the probability that our graph has no multiple edges is bounded away from zero (asymptotically it is $e^{-\binom{d}{2}}$). Thus, our result applies also to the model of d random *disjoint* perfect matchings between two sets of n vertices. This model, in turn, is contiguous to the uniform model of bipartite (vertex-labeled) d-regular simple graphs (for $d \geq 3$: see [MRRW97, Section 4][‡]), so our result applies in the latter model as well.

Random coverings of a fixed graph

In Theorem 3.1.4 we consider random n-coverings of a fixed graph Ω in the model $\mathcal{C}_{n,\Omega}$, where a uniform random permutation is generated for every edge of Ω. An equivalent model is attained if we cover some spanning tree of Ω by n disjoint copies and then choose a random permutation for every edge outside the tree (that is, the same automorphism-types of non-labeled graphs are obtained with the same distribution). In fact, picking a basepoint $\otimes \in V(\Omega)$, there is yet another description for this model: The classification of n-sheeted coverings of Ω by the action of $\pi_1(\Omega, \otimes)$ on the fiber $\{\otimes\} \times [n]$ above $\otimes$ shows that $\mathcal{C}_{n,\Omega}$ is equivalent to choosing uniformly at random an action of the free group $\pi_1(\Omega, \otimes)$ on $\{\otimes\} \times [n]$.

A different but related model uses the classification of *connected*, pointed coverings of $(\Omega, \otimes)$ by the corresponding subgroups of $\pi_1(\Omega, \otimes)$. A random n-covering is thus generated by choosing a

[†]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 3.1.1 applies both to the uniform model of labeled graph and to the uniform model of unlabeled graphs.

[‡]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.

random subgroup of index n. However, it seems that this model is contiguous to $\mathcal{C}_{n,\Omega}$ if $\mathrm{rk}(\Omega) \geq 2$ (note that the random covering Γ in $\mathcal{C}_{n,\Omega}$ is a.a.s. connected provided that $\mathrm{rk}(\Omega) \geq 2$). Indeed, the only difference is that in the new model, the probability of every connected graph Γ from $\mathcal{C}_{n,\Omega}$ is proportional to $\frac{1}{|\mathrm{Aut}(\Gamma)|}$. When $\mathrm{rk}(\Omega) \geq 2$, it seems that a.a.s. $|\mathrm{Aut}(\Gamma)| = 1$, which would show that our result applies to this model as well.

Finally, there is another natural model that comes to mind: given a periodic infinite tree, namely a tree that covers some finite graph, one can consider a random (simple) graph Γ with n vertices covered by this tree (with uniform distribution among all such graphs with n vertices, for suitable n's only). One can then analyze $\lambda(\Gamma)$, the largest absolute value of an eigenvalue besides[†] $\mathfrak{pf}(\Gamma)$. (This generalizes the uniform model on d-regular graphs.) Occasionally, all the quotients of some given periodic tree T cover the same finite "minimal" graph Ω. Interestingly, Lubotzky and Nagnibeda [LN98] showed that there exist such T's with a minimal quotient Ω which is *not* Ramanujan (in the sense that $\lambda(\Omega)$ is strictly larger than $\rho(T)$, the spectral radius of T). Since all the quotients of T inherit the eigenvalues of Ω, their $\lambda(\cdot)$ is also bounded away from $\rho(T)$ (from above). Hence, the corresponding version of Conjecture 3.1.3 is false in this general setting.

3.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 3.1.4 shows that if the base graph Ω is a good (nearly optimal) expander, then a.a.s. so are its random coverings. We would like to thank Ori Parzanchevski for his valuable assistance in writing this appendix.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

The spectral gap is then $1-\mu(\Gamma)$ here $\mu(\Gamma)=\max\{\mu_2,-\mu_m\}$. Up to a possible affine transformation, the spectrum of M_Γ 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* Δ_Γ^+ on $\left(\ell^2(V(\Gamma)),1\right)$:

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

The Laplacian equals $D_\Gamma-A_\Gamma$, where D_Γ 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 Γ is finite. In the finite case, the spectrum is

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

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

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

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

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

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

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

In the d-regular case, this amounts to the usual mixing lemma: $\left|E(S,T)-d\frac{|S|\cdot|T|}{m}\right|\leq\lambda(\Gamma)\sqrt{|S|\cdot|T|}$. If one takes $T=V\setminus S$, one can attain a bound on the Cheeger constant of Γ (see (3.B.1)).

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

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

Moreover, consider the *conductance* of Γ

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

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

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

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

$$h(\Gamma) = \min_{\substack{\emptyset \neq S \subseteq V \\ |S| \leq \frac{|V|}{2}}} \frac{|E(S, V \setminus S)|}{|S|}. \tag{3.B.1}$$

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

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

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

References

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

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

[AL02] A. Amit and N. Linial. Random graph coverings I: General theory and graph connectivity. *Combinatorica*, 22(1):1–18, 2002.

[AL06] A. Amit and N. Linial. Random lifts of graphs: edge expansion. *Combinatorics, Probability and Computing*, 15(03):317–332, 2006.

[ALM02] A. Amit, N. Linial, and J. Matoušek. Random lifts of graphs: independence and chromatic number. *Random Structures & Algorithms*, 20(1):1–22, 2002.

[Alo86] N. Alon. Eigenvalues and expanders. *Combinatorica*, 6(2):83–96, 1986.

[AM85] N. Alon and V.D. Milman. $\lambda 1$, isoperimetric inequalities for graphs, and superconcentrators. *Journal of Combinatorial Theory, Series B*, 38(1):73–88, 1985.

[Bar99] L. Bartholdi. Counting paths in graphs. *Enseign. Math., II. Sér.*, 45(1-2):83–131, 1999.

[Ben74] E. A. Bender. The asymptotic number of non-negative integer matrices with given row and column sums. *Discrete Mathematics*, 10(2):217–223, 1974.

[BL06] Y. Bilu and N. Linial. Lifts, discrepancy and nearly optimal spectral gap. *Combinatorica*, 26(5):495–519, 2006.

[BM86] B. Bollobás and B.D. McKay. The number of matchings in random regular graphs and bipartite graphs. *Journal of Combinatorial Theory, Series B*, 41(1):80–91, 1986.

[BMS02b] A.V. Borovik, A.G. Myasnikov, and V. Shpilrain. Measuring sets in infinite groups. In *Computational and Statistical Group Theory*, Contemp. Math., pages 21–42, Las Vegas, NV/Hoboken, NJ, 2002. American Mathematical Society.

[Bog08] O. 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. In *Foundations of Computer Science, 1987., 28th Annual Symposium on*, pages 286–294. IEEE, 1987.

[Bur87] M. Burger. Cheng's inequality for graphs. *preprint*, 1987.

[BV02] J. Burillo and E. Ventura. Counting primitive elements in free groups. *Geometriae Dedicata*, 93(1):143–162, 2002.

[Chu97] F.R.K. Chung. *Spectral graph theory*. Number 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. In *Proceedings of the twenty-first annual ACM symposium on Theory of computing*, pages 587–598. ACM, 1989.

[Fri91] J. Friedman. On the second eigenvalue and random walks in randomd-regular graphs. *Combinatorica*, 11(4):331–362, 1991.

[Fri03] J. Friedman. Relative expanders or weakly relatively ramanujan graphs. *Duke Mathematical Journal*, 118(1):19–35, 2003.

[Fri08] J. Friedman. *A proof of Alon's second eigenvalue conjecture and related problems*, volume 195 of *Memoirs of the AMS*. AMS, september 2008.

[GJKW02] C. Greenhill, S. Janson, J.H. Kim, and N.C. Wormald. Permutation pseudographs and contiguity. *Combinatorics, Probability and Computing*, 11(03):273–298, 2002.

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

[Gri77] R.I. Grigorchuk. Symmetric random walks on discrete groups. *Uspekhi Matematicheskikh Nauk*, 32(6):217–218, 1977.

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

[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*, 39:188–204, 1999.

[HLW06] S. Hoory, N. Linial, and A. Wigderson. Expander graphs and their applications. *Bulletin of the American Mathematical Society*, 43(4):439–562, 2006.

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

[KM02] I. Kapovich and A. Myasnikov. Stallings foldings and subgroups of free groups. *Journal of Algebra*, 248(2):608–668, 2002.

[Lei82] F.T. Leighton. Finite common coverings of graphs. *Journal of Combinatorial Theory, Series B*, 33(3):231–238, 1982.

[LN98] A. Lubotzky and T. Nagnibeda. Not every uniform tree covers ramanujan graphs. *Journal of Combinatorial Theory, Series B*, 74(2):202–212, 1998.

[LP10] N. Linial and D. Puder. Words maps and spectra of random graph lifts. *Random Structures and Algorithms*, 37(1):100–135, 2010.

[LPS88] A. Lubotzky, R. Phillips, and P. Sarnak. Ramanujan graphs. *Combinatorica*, 8(3):261–277, 1988.

[LR05] N. Linial and E. Rozenman. Random lifts of graphs: perfect matchings. *Combinatorica*, 25(4):407–424, 2005.

[LSV11] E. Lubetzky, B. Sudakov, and V. Vu. Spectra of lifted ramanujan graphs. *Advances in Mathematics*, 227(4):1612–1645, 2011.

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

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

[Mar88] G.A Margulis. Explicit group-theoretic constructions of combinatorial schemes and their applications in the construction of expanders and concentrators. *Problemy Peredachi Informatsii*, 24(1):51–60, 1988.

[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(2):231–244, 2008.

[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(1):44–62, 1994.

[MRRW97] M.S.O. Molloy, H. Robalewska, R.W. Robinson, and N.C. Wormald. 1-factorizations of random regular graphs. *Random Structures and Algorithms*, 10(3):305–321, 1997.

[MSS13] A. Marcus, D.A. Spielman, and N. 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. In G.N. Arzhantseva, L. Bartholdi, J. Burillo, and E. Ventura, editors, *Geometric group theory*, pages 225–253. Trends Math., Birkhauser, 2007.

[Nil91] A. Nilli. On the second eigenvalue of a graph. *Discrete Mathematics*, 91(2):207–210, 1991.

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

[Nor04] S. Northshield. Cogrowth of arbitrary graphs. In Kaimanovich V., editor, *Random walks and geometry*, pages 501–513. de Gruyter, 2004.

[PP15] D. Puder and O. Parzanchevski. Measure preserving words are primitive. *Journal of the American Mathematical Society*, 28(1):63–97, 2015.

[PR12] O. Parzanchevski and R. Rosenthal. Simplicial complexes: spectrum, homology and random walks. *arXiv preprint arXiv:1211.6775*, 2012.

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

[Pud14] D. Puder. Primitive words, free factors and measure preservation. *Israel Journal of Mathematics*, 201(1):25–73, 2014.

[Pud15b] D. Puder. Notes on the cogrowth formula: the regular, biregular and irregular cases. preprint, 2015+.

[PW14] D. Puder and C. Wu. Growth of the primitives elements in free groups. *Journal of London Mathematical Society*, 90(1):89–104, 2014.

[Riv04] I. Rivin. A remark on 'counting primitive elements in free groups' (by j. burillo and e. ventura). *Geometriae Dedicata*, 107(1):99–100, 2004.

[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*, 372:91–98, 2005.

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

[Sta83] J.R. Stallings. Topology of finite graphs. *Inventiones mathematicae*, 71(3):551–565, 1983.

[Tak51] M. Takahasi. Note on chain conditions in free groups. *Osaka Math. J*, 3(2):221–225, 1951.

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

Chapter 4

Growth of Primitive Elements in Free Groups

Doron Puder[†]
Einstein Institute of Mathematics
Hebrew University, Jerusalem
doronpuder@gmail.com

Conan Wu
Department of Mathematics,
Princeton University
shuyunwu@princeton.edu

Published: Journal of the London Mathematical Society, 90 (1), 2014, pp 89-104.
DOI: 10.1112/jlms/jdu009

Abstract

In the free group F_k, an element is said to be primitive if it belongs to a free generating set. In this paper, we describe what a generic primitive element looks like. We prove that up to conjugation, a random primitive word of length N contains one of the letters exactly once asymptotically almost surely (as $N \to \infty$).

This also solves a question from the list 'Open problems in combinatorial group theory' [Baumslag-Myasnikov-Shpilrain 02']. Let $p_{k,N}$ be the number of primitive words of length N in F_k. We show that for $k \geq 3$, the exponential growth rate of $p_{k,N}$ is $2k-3$. Our proof also works for giving the exact growth rate of the larger class of elements belonging to a proper free factor.

2010 Mathematics Subject Classification: 20E05 (Primary) 05A16 (Secondary)

Contents

[†]Supported by the Adams Fellowship Program of the Israel Academy of Sciences and Humanities, and an Advanced ERC Grant.

4.1 Introduction

Let F_k be the free group on k generators $X = \{x_1, \ldots, x_k\}$ ($k \geq 2$). Elements in F_k are represented by reduced words in the alphabet $X^{\pm 1} = \left\{x_1^{\pm 1} x_2^{\pm 1} \cdots, x_k^{\pm 1}\right\}$. A word $w \in F_k$ is called *primitive* if it belongs to some free generating set. We let $P_{k,N}$ denote the set of primitive elements of word length N in F_k. It is known (see, for example, [BMS02b]) that as $N \to \infty$ the set of primitive words is exponentially small in F_k. Namely, the exponential growth rate

$P_{k,N}$

$$\rho_k \stackrel{\text{def}}{=} \limsup_{N\to\infty} \sqrt[N]{|P_{k,N}|}$$

ρ_k

is strictly smaller than that of the whole free group F_k, which is $2k-1$. As observed in [Riv04], $\rho_2 = \sqrt{3}$, which gives the only case where the growth rate is known. For $k \geq 3$, various upper bounds on ρ_k have been established [BV02, BMS02b, Shp05]. The best upper bound to date is due to Shpilrain [Shp05] who showed $\rho_k \leq \lambda_k$, where λ_k is the greatest real root of $\lambda\left(\lambda^2 - 1\right)(\lambda - (2k-2)) + 1$. Here $\lambda_k < 2k-2$ for each k, but λ_k approaches $2k-2$ in the limit. A simple lower bound of $\rho_k \geq 2k-3$ stems from the fact that every word of the form $x_1 w(x_2, x_3, \cdots, x_k)$, where w is a word of length $N-1$ in $\left\{x_2^{\pm 1}, \ldots, x_k^{\pm 1}\right\}$, forms a free generating set together with $\{x_2, x_3, \cdots, x_k\}$, hence is primitive.

The exact value of ρ_k is the content of one of the open questions, attributed to M. Wicks, in [BMS02a, Problem F17] (see also the active website [BMS, Problem F19]). Here we answer the question and show the following tight result:

Theorem 4.1.1. *For all $k \geq 3$,*

$$\rho_k = \lim_{N\to\infty} \sqrt[N]{|P_{k,N}|} = 2k-3.$$

Moreover, there are positive constants c_k and C_k such that

$$c_k \cdot N \cdot (2k-3)^N \leq |P_{k,N}| \leq C_k \cdot N \cdot (2k-3)^N .$$

Remark 4.1.2. The second statement of Theorem 4.1.1 can be sharpened to $|P_{k,N}| = (1 + o_N(1)) \cdot \widehat{C_k} \cdot N \cdot (2k-3)^N$ for a specific constant $\widehat{C_k}$ which can be computed. This can be inferred from Theorem 4.1.3 and the analysis in Proposition 4.3.1 below.

The above theorem follows from an analysis of conjugacy classes of primitives in free groups. A word $w = a_1 a_2 \cdots a_N$ is called *cyclically reduced* if $a_1 \neq a_N^{-1}$. Such words, up to a cyclic permutation of letters, uniquely represent conjugacy classes in F_k. Hence for $w \in F_k$ we call the conjugacy class $[w]$ *the cyclic word* associated with w. Let the *cyclic length* of w, denoted by $|w|_c$, be the length of the cyclically reduced representatives of $[w]$.

$[w]$

$|w|_c$

There is a stark difference between the behavior of $P_{2,N}$ and that of $P_{k,N}$ when $k \geq 3$: whereas in F_2 'most' long primitives are conjugates of short ones, it turns out that for higher rank free groups the generic primitive word is nearly cyclically-reduced. In particular, the growth of the set of primitive elements is the same as that of primitive conjugacy classes (cyclic words) with respect to cyclic length. (This is the content of Proposition 4.3.1 below.)

Consider the set

$C_{k,N}$

$$C_{k,N} = \{[w] \,|\, w \in F_k \text{ is primitive and } |w|_c = N\}.$$

We compare the size of $C_{k,N}$ with its subset of cyclic-words in which some letter $x \in X$ appears exactly once (either as itself or its inverse), namely the set

$L_{k,N}$

$$L_{k,N} = \{[w] \,|\, \text{some } x \in X \text{ appears in } w \text{ exactly once}\} \subseteq C_{k,N}.$$

The size of $L_{k,N}$ can be easily approximated as[†] $|L_{k,N}| \approx 2k\,(2k-2)\,(2k-3)^{N-2}$. So that

$$\limsup_{N\to\infty} \sqrt[N]{|L_{k,N}|} = 2k-3.$$

Theorem 4.1.3. *For $k \geq 2$*

$$\limsup_{N\to\infty} \sqrt[N]{|C_{k,N}|} = 2k-3.$$

For $k \geq 3$,

$$\limsup_{N\to\infty} \sqrt[N]{|C_{k,N} \setminus L_{k,N}|} < 2k-3.$$

Moreover,

$$|C_{k,N}| = (1+o_N(1)) \cdot \frac{2k\,(2k-2)}{(2k-3)^2}\,(2k-3)^N.$$

The second statement of theorem 4.1.3 means that except for an exponentially small set, all primitive cyclic-words contain one of the letters exactly once. When $k \geq 3$ the first and last statements are an immediate consequence of the second one and the approximated size of $L_{k,N}$ as given above.

Note that the first statement of theorem 4.1.3 is also valid for $k = 2$: the exponential growth rate of conjugacy classes of primitives in F_2 is 1. This special case was already shown in [MS03, Prop 1.4]: it turns out the size of $C_{2,N}$ is exactly $4\varphi(n)$, where $\varphi(\cdot)$ is the Euler function. Whereas $\rho_2 = \sqrt{3}$ is strictly larger than 1, for all $k \geq 3$ the growth of primitive cyclic-words is the same as the growth of primitive words.

A natural question along the same vein would be to estimate the growth of the larger set $S_{k,N}$ consisting of words in F_k which are contained in a proper free factor (clearly, $P_{k,N} \subseteq S_{k,N}$). Our proof of Theorem 4.1.1 also applies to this question and yields that $S_{k,N}$ has the same exponential growth rate as $P_{k,N}$: $S_{k,N}$

Corollary 4.1.4. *For $k \geq 3$ we have*

$$\lim_{N\to\infty} \sqrt[N]{|S_{k,N}|} = \lim_{N\to\infty} \sqrt[N]{|S_{k,N} \setminus P_{k,N}|} = 2k-3.$$

We show that $\lim_{N\to\infty} \sqrt[N]{|S_{k,N}|} \leq 2k-3$ in Section 4.3.4. This requires only a small variation on the proof of Theorem 4.1.1. The lower bound is, again, easier, and follows immediately from the fact that primitives are exponentially negligible in F_k (this fact follows from Theorem 4.1.1 but also, as mentioned above, from previous results concerning the growth of primitives). Indeed, this fact shows that most words in any size $k-1$ subset of the letters are non-primitive. We conclude that the number of non-primitive words in $S_{k,N}$ grows at least as fast as $(2k-3)^N$. Thus $S_{k,N}$ is indeed larger than $P_{k,N}$ in a non-negligible manner, namely,

$$\lim_{N\to\infty} \sqrt[N]{|S_{k,N} \setminus P_{k,N}|} \geq 2k-3 = \lim_{N\to\infty} \sqrt[N]{|P_{k,N}|}.$$

Remark 4.1.5. For a different proof showing that $\lim_{N\to\infty} \sqrt[N]{|S_{k,N} \setminus P_{k,N}|} = 2k-3$, see [Pud15a, Thm 8.2] (in the terminology therein, every word in $S_{k,N} \setminus P_{k,N}$ has primitivity rank $\leq k-1$). In the techniques of that paper (especially [Pud15a, Prop. 4.3]), it can be shown that a generic word in $S_{k,N} \setminus P_{k,N}$ is, up to conjugation, a word in some $(k-1)$-subset of the letters of X.

[†]This expression is very close to the truth, except that we double count words in which two or more letters appear exactly once. The exact cardinality of $L_{k,N}$ can be obtained by an application of the inclusion-exclusion formula. Note that the share of doubly-counted words is exponentially negligible in $L_{k,N}$: it is of exponential order $(2k-5)^N$.

Our proofs rely on a thorough analysis of the Whitehead algorithm to detect primitive elements. To a lesser extent, we also use a characterization of primitive elements based on the distribution they induce on finite groups. In Section 4.2 we give some background on Whitehead algorithm and describe the graphs used in it, called Whitehead graphs. We then divide the set of primitives into finitely many classes according to certain properties of their Whitehead graphs. Most of these classes turn out to be of negligible size, but we postpone the somewhat technical proof of this fact to Section 4.4. In Section 4.3 we give some background on the aforementioned "statistical" characterization of primitives, estimate the size of the remaining classes of primitives and complete the proofs of Theorems 4.1.1, 4.1.3 and of Corollary 4.1.4. We end with some open questions in Section 4.5.

Acknowledgements

We would like to thank Tsachik Gelander for bringing our attention to the question. We thank Warren Dicks, Ilya Kapovich, Nati Linial, Shahar Mozes and Alexey Talambutsa for beneficial comments. We also thank the anonymous referee for his valuable comments. The second author would like to thank the Hebrew University for providing hospitality and stimulating mathematical environment during which part of this work was conducted.

4.2 Whitehead Graphs

In [Whi36a], Whitehead introduced the first algorithm to detect primitive words in $\mathbf{F}_k$ (and more generally subsets of bases of $\mathbf{F}_k$). (Subsequently, in [Whi36b], he solved a more general question: Given two words $w_1, w_2 \in F_k$, when does there exist an automorphism $\phi \in \mathrm{Aut}(F_k)$ mapping w_1 to w_2? Note that w_1 is primitive if and only if there is an automorphism mapping it to a single-letter word.) Along the years it has become the most standard way of detecting primitive elements. Stallings generalized the algorithm in order to detect words belonging to free factors of $\mathbf{F}_k$ [Sta99]. For other algorithms to detect primitives see, e.g., [LS70, Chapter I.2] or [Pud14].

The algorithm is based on the following construction: Let M_k be a 3-manifold which is the connected sum of k copies of $\mathbb{S}^1 \times \mathbb{S}^2$. Clearly, we have $\pi_1(M_k) = F_k$. Fix a set of k disjoint 2-spheres $S_1, S_2, \cdots, S_k$, one corresponding to each summand, so that $\widehat{M}_k = M_k \setminus \bigcup_{i=1}^k S_i$ is simply connected with $2k$ boundary components $S_1^+, S_1^-, S_2^+, \cdots, S_k^-$. The manifold M_k may be visualized as the double of a handlebody H_k with $\{S_i\}$ being the double of a cut system of H_k (a cut system is a set of disjoint discs that cuts the handlebody into a ball). For every $w \in F_k = \pi_1(M_k)$, the cyclic word $[w]$ can be realized as a simple curve in M_k. Conversely, given any oriented curve in M_k one can write down a cyclic word in F_k by reading off the sequence of spheres the curve intersects, with signs. Hence we get a bijective correspondence between cyclic words $[w]$ and homotopy classes of oriented simple curves in M_k.

Given any proper non-empty subset $\mathcal{U} \subset \left\{S_1^+, S_1^-, \cdots, S_k^-\right\}$, there is an embedded 2-sphere $S_{\mathcal{U}}$ in $\widehat{M}_k$ separating the boundary components in $\mathcal{U}$ from those not in $\mathcal{U}$. For every $v \in X^{\pm 1}$ denote by S_v the corresponding boundary component of $\widehat{M}$ (so $S_{x_i} = S_i^+$ and $S_{x_i^{-1}} = S_i^-$). If there exists some $v = x_j^{\varepsilon} \in X^{\pm 1}$ such that $S_v \notin \mathcal{U}$ and $S_{v^{-1}} \in \mathcal{U}$ then $S_{\mathcal{U}}$ is an essential non-separating sphere[†] in $\widehat{M}_k \cup S_j$. The **Whitehead automorphism** $\varphi_{(\mathcal{U},v)}$ of F_k is then defined by replacing the sphere S_j by $S_{\mathcal{U}}$ and writing each cyclic word as the intersection pattern of the corresponding curve with the new set of spheres. In the example illustrated above, $S_3^+ \notin \mathcal{U}$ and $S_3^- \in \mathcal{U}$ hence we may replace S_3 with $S_{\mathcal{U}}$. Writing down $\varphi_{(\mathcal{U},v)}$ formally one gets:

- $\varphi_{(\mathcal{U},v)}(v) = v$; $\varphi_{(\mathcal{U},v)}(v^{-1}) = v^{-1}$;

[†]Namely, a non-contractible embedding of a sphere which does not separate $\widehat{M}_k \cup S_j$ into two connected components.

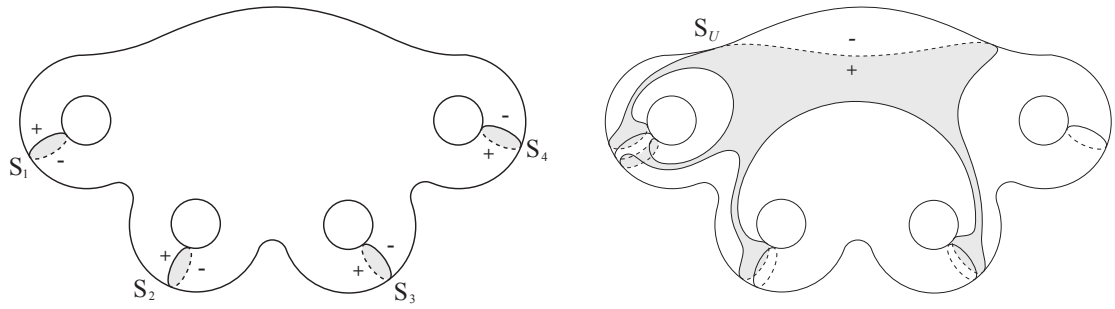
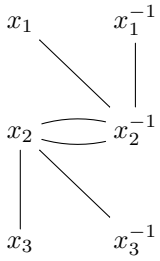


Figure 4.2.1: Spheres S_i in H_k and $S_{\mathcal{U}}$ with $\mathcal{U} = \left\{S_1^+, S_1^-, S_2^+, S_3^-\right\}$

- for $u \neq v$,
  - $\varphi_{(\mathcal{U},v)}(u) = u$ if $S_u, S_{u^{-1}} \notin \mathcal{U}$;
  - $\varphi_{(\mathcal{U},v)}(u) = vuv^{-1}$ if $S_u, S_{u^{-1}} \in \mathcal{U}$;
  - $\varphi_{(\mathcal{U},v)}(u) = vu$ and $\varphi_{(\mathcal{U},v)}(u^{-1}) = uv^{-1}$ if $S_u \in \mathcal{U}$, $S_{u^{-1}} \notin \mathcal{U}$;

By forgetting the order in which the spheres are intersected and looking only at the arcs connecting boundary components in $\widehat{M}_k$ one gets a finite graph with $2k$ vertices labeled $X^{\pm 1} = \left\{x_1^{\pm 1}, \cdots, x_k^{\pm 1}\right\}$. This is called the **Whitehead graph** of the cyclic word $[w]$, denoted by $\Gamma(w)$. For example, $\Gamma(w)$ for $w = x_1 x_2^2 x_3^{-1} x_2^{-2} \in F_3$ is:

$\Gamma(w)$



Going from manifolds to graphs, to every $(\mathcal{U}, v)$ defined as above corresponds a partition of the vertices by $Z = \{x_u \,|\, S_u \in \mathcal{U}\}$ and $Y = X^{\pm 1} \backslash (Z \cup \{v\})$. Denote $\phi_{Y,Z,v} = \varphi_{(\mathcal{U},v)}$, and notice that $X^{\pm 1} = Y \amalg Z \amalg \{v\}$. The following theorem, part of the foundation for Whitehead's algorithm, plays a central role in our argument:

$\phi_{Y,Z,v}$

Theorem 4.2.1. *[Sta99, Thm 2.4] If w is contained in a proper free factor of F_k, then $\Gamma(w)$ has a cut vertex.*

Namely, there exists a vertex v such that $\Gamma(w) \backslash \{v\}$ is disconnected. This includes the case where $\Gamma(w)$ is itself disconnected. Note that, in particular, all primitive elements are contained in a rank one free factor, hence have Whitehead graphs with cut vertices.

Note that the cyclic length $|w|_c$ of w is the number of edges in the Whitehead graph corresponding to a cyclically reduced representative. A natural candidate for a length reducing Whitehead automorphism is therefore to replace the sphere corresponding to the cut vertex v by one that separates a connected component of $\Gamma(w) \backslash \{v\}$. Indeed, we have the following:

Proposition 4.2.2. *[Sta99, Prop 2.3] Let v be a cut-vertex of $\Gamma(w)$, and let Y and Z be a non-trivial partition of the remaining vertices so that there are no edges between Y and Z, and $v^{-1} \in Z$. Then*

$$|\phi_{Y,Z,v}(w)|_c = |w|_c - E(Y, v).$$

Here $E(Y,v)$ is the number of edges connecting v to Y. For instance, for $w = x_1x_2^2x_3^{-1}x_2^{-2}$ as above, there are two possible cut-vertices: x_2 and x_2^{-1}. If one chooses $v = x_2$ and $Z = \{x_1, x_1^{-1}, x_2^{-1}\}$, then $[\phi_{Y,Z,v}(w)] = [x_2x_1x_2x_3^{-1}x_2^{-2}] = [x_1x_2x_3^{-1}x_2^{-1}]$ has cyclic length 4.

Moreover, it is easy to see that if w is contained in a proper free factor then it is almost always possible to find a triplet (Y,Z,v) as in Proposition 4.2.2 with $E(Y,v) > 0$: the only exceptions are $|w|_c \leq 1$ or when w is a word in a proper subset of the letters, say $x_1, \dots, x_j$ ($j < k$), and it does not belong to a proper free factor in $F(x_1, \dots, x_j)$. This is the crux of the Whitehead algorithm to detect primitives: since the second case cannot occur for primitive elements with $|w|_c > 1$, if w is primitive one can always apply a sequence of Whitehead automorphisms according to cut vertices in the Whitehead graph, until it becomes a (conjugate of a) single-letter word.

Our proof of Theorem 4.1.1 (and of Corollary 4.1.4) relies on a rigorous analysis of the possible triplets (Y,Z,v). We say that a triplet (Y,Z,v) is *valid* for the cyclic word $[w]$ if it satisfies the statement in Proposition 4.2.2 (namely, if v is a cut-vertex of $\Gamma(w)$, Y and Z are a non-trivial partition of the remaining vertices with $E(Y,Z) = 0$, and $v^{-1} \in Z$). Let $A_{Y,Z,v}$ denote the set of all cyclic words having (Y,Z,v) as a valid triplet; namely $A_{Y,Z,v}$

$$A_{Y,Z,v} = \{[w] \mid (Y,Z,v) \text{ is a valid triplet for } w\}$$

and $A^N_{Y,Z,v}$

$$A^N_{Y,Z,v} = \{[w] \in A_{Y,Z,v} \,|\, |w|_c = N\}.$$

By Theorem 4.2.1,

$$C_{k,N} \subseteq \bigcup_{(Y,Z,v)} A^N_{Y,Z,v}$$

taking the union over all possible triplets partitioning $X^{\pm 1}$ (with $Y, Z \neq \emptyset$ and $v^{-1} \in Z$).

We proceed by bounding the growth of primitives in $A_{Y,Z,v}$ for each of the finitely many triplets (Y,Z,v). Intuitively, the cut vertex and partition will restrict the number of possible ways to connect vertices, hence result in a smaller growth rate. In the extreme case, if both sets Y and Z contain roughly half of the elements of $X^{\pm 1}$, namely $|Y| \approx k$ and $|Z| \approx k$; then from any vertex in Y one can connect only to another vertex in $Y \cup \{v\}$, resulting in $\sim k$ choices. If we ignore the possibility of going through v, then the possible number of such cyclic words would be roughly only k^N, which amounts to an exponential growth rate of k (note that this applies to the whole set $A_{Y,Z,v}$ and not only to the primitives in it). Hence we should expect $A_{Y,Z,v}$ to grow faster for triplets (Y,Z,v) where one of Y, Z is almost all of $X^{\pm 1}$.

Indeed it turns out that $A_{Y,Z,v}$ is negligible unless one of Y, Z is very small:

Proposition 4.2.3. *Every triplet (Y,Z,v) satisfies*

$$\limsup_{N\to\infty} \sqrt[N]{\left|A^N_{Y,Z,v}\right|} < 2k-3,$$

unless $min(|Y|,|Z|) = 1$ or $Y = \{x, x^{-1}\}$ for some letter x.

The proof of this proposition involves some careful analysis in various cases, and we postpone it to Section 4.4. In Section 4.3 we assume this proposition, give the precise growth rates for the remaining essential partitions and obtain our theorems.

4.3 Proof of Theorems

In this section we complete the proofs of Theorems 4.1.1 and 4.1.3 and of Corollary 4.1.4.

4.3.1 Primitives and cyclic primitives

Here, we present the observation that, unlike in F_2 , for $k \geq 3$ it suffices to count conjugacy classes containing primitive words.

Proposition 4.3.1. *For $k \geq 3$, if $|C_{k,N}| \leq C \cdot (2k-3)^N$ for some $C > 0$ as follows from Theorem 4.1.3, then*

$$|P_{k,N}| \leq D \cdot N \cdot (2k-3)^N$$

for some $D > 0$.

Proof. Each $w \in P_{k,N}$ is of the form

$$uw'u^{-1}$$

where $w' \in F_k$ is cyclically reduced and primitive. Let ℓ be the word length of u, so that $0 \leq \ell \leq \frac{N-1}{2}$ and $|w|_c = |w'|_c = N - 2\ell$.

Since w' is primitive, in particular, it is not a proper power, hence each of its cyclic shifts is different. Namely, the cyclic word $[w']$ is represented by exactly $N - 2\ell$ distinct cyclically reduced words. On the other hand, u can be any word of length ℓ as long as the first letter of u^{-1} and the last letter of u do not cancel out their adjacent letters in w'. There are $(2k-1)^{\ell-1}(2k-2)$ such words. Therefore,

$$\begin{aligned}
|P_{k,N}| &= \sum_{\ell=0}^{\lfloor \frac{N-1}{2} \rfloor} (N-2\ell) \, |C_{k,N-2\ell}| \, (2k-2) \, (2k-1)^{\ell-1} \\
&\leq N \cdot \sum_{\ell=0}^{\lfloor \frac{N-1}{2} \rfloor} |C_{k,N-2\ell}| \, (2k-1)^{\ell} \\
&\leq N \cdot \sum_{\ell=0}^{\lfloor \frac{N-1}{2} \rfloor} C \cdot (2k-3)^{N-2\ell} \, (2k-1)^{\ell} \\
&= N \cdot C \cdot (2k-3)^N \sum_{\ell=0}^{\lfloor \frac{N-1}{2} \rfloor} \left(\frac{2k-1}{(2k-3)^2} \right)^{\ell}.
\end{aligned}$$

For $k \geq 3$, $\left(\frac{2k-1}{(2k-3)^2} \right) \leq \frac{5}{9} < 1$. Bounding the geometric series, we deduce that $|P_{k,N}| \leq D \cdot N \cdot (2k-3)^N$. □

The proposition shows that Theorem 4.1.1 follows from Theorem 4.1.3: For the lower bound in Theorem 4.1.1 recall that the number of cyclically reduced primitive words of length N with one of the letters appearing exactly once is

$$N \cdot |L_{k,N}| = (1 - o_N(1)) \cdot N \cdot 2k \cdot (2k-2) \cdot (2k-3)^{N-2}.$$

To complete the proofs of Theorems 4.1.1 and 4.1.3, it remains to bound from above the growth of cyclic primitives. Before starting the proof we present in Section 4.3.2 a couple of useful facts which will be used in the sequel.

4.3.2 Ingredients for bounding cyclic primitives

First, we give some background on a line of thought regarding primitive words which is different from Whitehead's and leads to a measure-theoretic characterization of primitives. Let $w = x_{i_1}^{\varepsilon_1} x_{i_2}^{\varepsilon_2} \cdots x_{i_N}^{\varepsilon_N}$ be a word in F_k. For every group G, w induces a *word map* from the Cartesian product G^k to G, by substitutions:

$$w : (g_1, \ldots, g_k) \mapsto g_{i_1}^{\varepsilon_1} g_{i_2}^{\varepsilon_2} \cdots g_{i_N}^{\varepsilon_N}.$$

When G is finite (compact) and G^k is given the uniform (Haar, resp.) measure, the push forward by w of this measure results in a new measure on G, which we denote by G_w. It is an easy observation that if w_1 and w_2 are in the same $\mathrm{Aut}\, F_k$-orbit of F_k, then they induce the same measure on every finite or compact group, namely $G_{w_1} = G_{w_2}$ (see [PP15, Observation 1.2]). In particular, if w is primitive, then $G_w = G_{x_1}$ which is clearly the uniform (Haar) measure on G. G_w

It is natural to ask whether the converse also holds. Namely, if $G_{w_1} = G_{w_2}$ for every finite (compact) group, does it imply that w_1 and w_2 are in the same $\mathrm{Aut}\, F_k$-orbit? This conjecture is still wide open. However, the special case concerning primitives was settled in [PP15]. It is shown there that if G_w is uniform for every finite group G, then w is primitive. In the heart of the argument in [PP15] lies a result about the distributions induced by words on the symmetric groups S_n. We re-formulate it as follows:

Theorem 4.3.2. *[PP15, Thm 1.7] Let $w \in F_k$. For every $n \in \mathbb{N}$ let $\sigma_{w,n}$ be a random permutation in S_n distributed according to $(S_n)_w$. Then w is non-primitive if and only if there exists some n_0 such that for all $n > n_0$ we have*

$$\mathbb{E}(|\mathrm{Fix}\,(\sigma_{w,n})|) > 1,$$

where Fix *(σ) denotes the set of fixed points of σ.*

Note that for w primitive, $\sigma_{w,n}$ is a uniformly distributed random permutation in S_n, hence the expected number of fixed points is exactly 1. From this theorem we derive the following fact which will be useful in the argument. We say that w_1 and w_2 are *letter disjoint* words if their reduced forms use disjoint subsets of the alphabet X.

Proposition 4.3.3. *Let $w_1, w_2 \in F_k$ be letter disjoint words. If the concatenation $w_1 w_2$ is primitive, then at least one of w_1 or w_2 is primitive.*

Proof. Since the push-forward measure by w is a class function, the probability $Pr(\sigma_{w,n}(i) = i)$ is independent of i (here $i \in \{1, \ldots, n\}$), and likewise, $Pr(\sigma_{w,n}(i) = j)$ is independent of i and j as long as $i \neq j$. Thus, $\mathbb{E}(|\mathrm{Fix}(\sigma_{w,n})|) > 1$ if and only if $Pr(\sigma_{w,n}(1) = 1) > \frac{1}{n}$. Let $p(w,n) = Pr(\sigma_{w,n}(1) = 1)$.

Since w_1 and w_2 are letter disjoint, they induce independent push-forward measures on S_n. If both words are non-primitive then for large enough n, both $p(w_1, n) > \frac{1}{n}$ and $p(w_2, n) > \frac{1}{n}$, which implies

$$\begin{aligned} p\,(w_1 w_2, n) &= Pr(\sigma_{w_1 w_2, n}(1) = 1) \\ &= \sum_{j=1}^{n} Pr(\sigma_{w_1,n}(1) = j) \cdot Pr(\sigma_{w_2,n}(j) = 1) \\ &= p(w_1,n)p(w_2,n) + (n-1)\,\frac{1-p(w_1,n)}{n-1} \cdot \frac{1-p(w_2,n)}{n-1} \\ &= \frac{1}{n} + \frac{n}{n-1} \cdot \left(p(w_1,n) - \frac{1}{n}\right) \cdot \left(p(w_2,n) - \frac{1}{n}\right) > \frac{1}{n}. \end{aligned}$$

This contradicts the assumption that $w_1 w_2$ is primitive. □

Recall that by Proposition 4.2.3 primitives from $A_{Y,Z,v}$ for most triplets (Y, Z, v) are negligible. We make some simple observations about the remaining three types of triplets:

- If $|Y| = 1$, say $Y = \{a\}$, and $w \in A_{Y,Z,v}$ then each appearance of a is followed by v^{-1} and each appearance of a^{-1} is preceded by v. It is not hard to see that the growth rate here is at least $\sqrt{(2k-3)^2 + 1} > 2k - 3$. Indeed, consider the $(2k-2)\,(2k-3)$ ordered reduced pairs of letters not containing $a^{\pm 1}$ and, in addition, the pair av^{-1}. Each one of these pairs

can be followed by one of $(2k-3)(2k-3)+1$ of these pairs, which shows the lower bound. Since every possible pair of letters is followed by one of less than $(2k-1)^2$ possible pairs, the growth rate is strictly less than $2k-1$. In fact, the exact growth rate is the largest (real) root of $\lambda^5-(2k-3)\lambda^4-3\lambda^3+(2k-3)\lambda^2+3\lambda+(2k-3)$, which tends to $2k-3$ as $k\to\infty$.

- If $Y=\left\{a,a^{-1}\right\}$ then every instance of $a^{\pm 1}$ in a word from $A_{Y,Z,v}$ is in the form $\ldots va^m v^{-1}\ldots$ for some $0\neq m\in\mathbb{Z}$. The exponential growth rate here is the largest (real) root of $\lambda^4-(2k-2)\lambda^3+(2k-4)\lambda^2+(2k-2)\lambda-6k+11$, which again approaches $2k-3$ from above as $k\to\infty$. Again, looking at pairs of letters one can easily infer the growth rate is strictly less than $2k-1$.

- Finally, if $|Z|=1$, namely $Z=\left\{v^{-1}\right\}$, then v^{-1} is followed only by v^{-1}, and v is preceded only by v. So $A_{Y,Z,v}$ consists of all cyclic words not containing $v^{\pm 1}$ (together with $\{v^m \mid m\in\mathbb{Z}\}$). Hence, the growth rate is exactly $(2k-3)$.

In particular, this analysis gives rise to the following naive bound:

Corollary 4.3.4. *Let $k\geq 2$. Every triplet (Y,Z,v) satisfies*

$$\limsup_{N\to\infty}\sqrt[N]{\left|A^N_{Y,Z,v}\right|}<2k-1.$$

4.3.3 Proof of Theorem 4.1.3

The moral of the following proof will be that even when the set $A_{Y,Z,v}$ has exponential growth rate larger than $2k-3$, every cyclic *primitive* $[w]\in A_{Y,Z,v}$ can be shortened 'fast enough' by the corresponding Whitehead automorphism $\phi_{Y,Z,v}$. Recall that the second statement of Theorem 4.1.3 is that for $k\geq 3$,

$$\limsup_{N\to\infty}\sqrt[N]{|C_{k,N}\setminus L_{k,N}|}<2k-3. \tag{4.3.1}$$

As mentioned in Section 4.1, when $k\geq 3$ the other two statements of the theorem follow from (4.3.1), and for $k=2$, the relevant statement ($\limsup_{N\to\infty}\sqrt[N]{|C_{2,N}|}=1$) is already known [MS03, Prop. 1.4]. Therefore, we shall prove (4.3.1) by induction on k, assuming only that $k\geq 3$ and that for $k-1$ we have

$$\limsup_{N\to\infty}\sqrt[N]{|C_{k-1,N}|}=2k-5.$$

Assume then that $k\geq 3$, and let $M_{k,N}\subseteq C_{k,N}\setminus L_{k,N}$ be the set of cyclic primitive words such that either $M_{k,N}$

- $[w]\in A^N_{Y,Z,v}$ with $Y=\{x\}$ and x appearing at least 4 times in $[w]$, or
- $[w]\in A^N_{Y,Z,v}$ with $Y=\left\{x,x^{-1}\right\}$ and $[w]$ containing at least 2 instances of vx^mv^{-1} (for any $0\neq m\in\mathbb{Z}$).

Let $M^c_{k,N}$ denote the complement of $M_{k,N}$ inside $C_{k,N}\setminus L_{k,N}$. We proceed by showing that the exponential growth rates of $|M_{k,N}|$ and $|M^c_{k,N}|$ are both strictly less than $2k-3$. $M^c_{k,N}$

Lemma 4.3.5.

$$\limsup_{N\to\infty}\sqrt[N]{\left|M^c_{k,N}\right|}<2k-3.$$

Proof. In light of Proposition 4.2.3 we only need to consider cyclic primitive words $[w] \in M^c_{k,N}$ with a Whitehead partition (Y, Z, v) where $\min\{|Y|, |Z|\} = 1$ or $Y = \{x, x^{-1}\}$.

If $|Z| = 1$ then by definition $Z = \{v^{-1}\}$, and every cyclic word in $A_{Y,Z,v}$ is either a power of v or a word in the alphabet $X^{\pm 1} \setminus \{v^{\pm 1}\}$. The set $\{v^n\}$ is clearly negligible. In the latter case, $[w]$ is primitive in the letters $X \setminus \{v^{\pm 1}\}$. It follows from the induction hypothesis that the exponential growth rate of this set of cyclic primitives is $2k - 5 < 2k - 3$.

Assume next that $Y = \{x, x^{-1}\}$ and $[w]$ has exactly one instance of $vx^m v^{-1}$ for some $|m| \geq 2$ (the case $|m| = 1$ is impossible as $M^c_{k,N} \cap L_{k,N} = \emptyset$). Pick a representative of $[w]$ of the form $[x^m w']$, where w' has length $< N$ and does not contain the letter x. By Proposition 4.3.3, w' is primitive in F_{k-1} hence this set has exponential growth rate $2k - 5$.

The remaining case is $Y = \{x\}$ and $x^{\pm 1}$ appearing exactly twice or thrice in $[w]$ (again, if $x^{\pm 1}$ appears only once, then $[w]$ belongs to $L_{k,N}$). Consider first the case where $x^{\pm 1}$ appears exactly **twice**:
We can write $[w]$ in the form

$$\left[(u_1 x u_2)^{\pm 1} w_1 (u_1 x u_2)^{\pm 1} w_2\right],$$

where u_1, u_2 are maximal sequence of letters preceding and following both instances of $x^{\pm 1}$, respectively.

Let ℓ_i be the length of u_i. Up to a factor of N^3, which is negligible in terms of exponential growth rates, we know ℓ_1, ℓ_2, $|w_1|$ and $|w_2|$. There are about $(2k-3)^{\ell_1+\ell_2}$ options for the values of u_1 and u_2. The automorphism $\psi \in \mathrm{Aut}(\mathbf{F}_k)$ which maps $x \mapsto u_1^{-1} x u_2^{-1}$ and leaves unchanged the remaining letters, maps $[w]$ to the primitive cyclic word

$$[w'] = \left[x^{\pm 1} w_1 x^{\pm 1} w_2\right].$$

Let $N' = N - 2(\ell_1 + \ell_2)$ be the length of w'. We claim that the number of possible w' is bounded above by $C \cdot (2k - 3 - \varepsilon)^{N'}$ for some $C, \varepsilon > 0$, ε small. This will suffice as the number of possible $[w]$ is then bounded by some polynomial in N times

$$(2k-3-\varepsilon)^{N'} \cdot (2k-3)^{\ell_1+\ell_2} = (2k-3-\varepsilon)^{N-2\ell_1-2\ell_2} \cdot \sqrt{2k-3}^{2\ell_1+2\ell_2} \leq (2k-3-\varepsilon)^N .$$

Firstly, if one of w_1 or w_2 is trivial, then as in the preceding case, the number of options for $[w']$ is at most some constant times $(2k-5)^{N'}$. So, assume $w_1, w_2 \neq 1$. The word $[w']$ belongs to some $A^{N'}_{Y',Z',v'}$. By the maximality of u_1 and u_2, each of the vertices x and x^{-1} in the Whitehead graph $\Gamma(w')$ has at least two neighbors in $X^{\pm 1} \setminus \{x, x^{-1}\}$. Hence, $Y' \neq \{x\}, \{x^{-1}\}, \{x, x^{-1}\}$. Also, it is not possible that $v' = x^{\pm 1}$ and $Z' = \left\{(v')^{-1}\right\}$, since cyclic words corresponding to this triplet are either words in $X \setminus \{v'\}$ or powers of v'. Hence, the triplet (Y', Z', v') induces some non-trivial partition of $(X \setminus \{x\})^{\pm 1}$ (both Y' and Z' intersect $(X \setminus \{x\})^{\pm 1}$). Hence, w_1 and w_2 are words (albeit not cyclic) corresponding to some non-trivial triplet partitioning $(X \setminus \{x\})^{\pm 1}$. But by Corollary 4.3.4, the exponential growth of such subsets is strictly less than $2(k-1) - 1 = 2k - 3$.

Finally, consider the case where $Y = \{x\}$ and $x^{\pm 1}$ appears exactly **three times** in $[w]$:
The proof that this subset grows slower than $(2k-3)^N$ is very similar to the previous case. This time, each such word is of the form

$$\left[(u_1 x u_2)^{\pm 1} w_1 (u_1 x u_2)^{\pm 1} w_2 (u_1 x u_2)^{\pm 1} w_3\right]$$

with u_1, u_2 maximal. It can be shortened via an automorphism to

$$[w'] = \left[x^{\pm 1} w_1 x^{\pm 1} w_2 x^{\pm 1} w_3\right]$$

of length N'. Again, up to a polynomial factor of N^4 we know $\ell_1, \ell_2, |w_1|, |w_2|$ and $|w_3|$, and we claim that the number of options for $[w']$ is bounded by some constant times $(2k-3-\varepsilon)^{N'}$ with

$\varepsilon > 0$. Hence the total number of options for $[w]$ is bounded by some polynomial in N times

$$(2k-3-\varepsilon)^{N-3\ell_1-3\ell_2} \cdot (2k-3)^{\ell_1+\ell_2} \leq (2k-3-\varepsilon)^N$$

for ε small enough. Indeed, if two of w_1, w_2 and w_3 are trivial, we are again in the same situation as in the case $Y = \{x, x^{-1}\}$. Otherwise, the exact same argument as before shows that $[w'] \in A^{N'}_{Y',Z',v'}$ for some triplet (Y', Z', v') partitioning $(X \setminus \{x\})^{\pm 1}$ non-trivially.

This covered all the cases of $M^c_{k,N}$ and hence the lemma is established. □

Now we move on to the remaining set $M_{k,N}$. The idea is to shorten such words by applying appropriate Whitehead automorphisms until the result falls outside of $M_{k,*}$ (i.e. outside of $M_{k,n}$ for all n). To achieve this goal we first consider cyclic words $[w] \in M_{k,N}$ such that the corresponding automorphism $\phi_{Y,Z,v}$ maps them into $L_{k,*}$: $[\phi_{Y,Z,v}(w)] \in L_{k,*}$. Denote this subset by $\widetilde{L}_{k,N} \subseteq M_{k,N}$. $\widetilde{L}_{k,N}$ We claim that:

Lemma 4.3.6.

$$\limsup_{N\to\infty} \sqrt[N]{\left|\widetilde{L}_{k,N}\right|} < 2k-3.$$

Proof. Consider first the words $[w] \in \widetilde{L}_{k,N}$ with triplet (Y, Z, v) where $Y = \{x\}$. The effect of $\phi = \phi_{Y,Z,v}$ on $[w]$ is precisely that: each instance of the form $\dots x v^{-1} \dots$ becomes simply $\dots x \dots$, and each instance of $\dots v x^{-1} \dots$ turns into $\dots x^{-1} \dots$. In particular, the only letter whose number of appearances in $[w]$ is changed by ϕ is v. By definition, $[w] \in \widetilde{L}_{k,N} \subseteq M_{k,N} \subseteq C_{k,N} \setminus L_{k,N}$, so the letter which $[\phi(w)]$ contains exactly once is necessarily $v^{\pm 1}$. Hence, aside for one, all occurrences of $v^{\pm 1}$ in $[w]$ are as part of either xv^{-1} or vx^{-1}. We deduce that $[w]$ is of the form

$$[vw']$$

with w' being a word of length $N-1$ in $2(k-2)$ building blocks of length 1: $\left(X \setminus \left\{x^{\pm 1}, v^{\pm 1}\right\}\right)^{\pm 1}$ and 2 building blocks of length 2: $\left(xv^{-1}\right)^{\pm 1}$. (In other words, w' is any word in F_{k-1} but where one of the letters is of length 2). This kind of words clearly has exponential growth rate strictly less than $2k-3$. (To be precise, the growth rate is the larger root of $\lambda^2 - (2k-4)\lambda - 2$.)

The complement of this latter subset inside $\widetilde{L}_{k,N}$ consists of primitive cyclic words belonging to $A^N_{Y,Z,v}$ with $Y = \{x, x^{-1}\}$. This time, ϕ turns each instance of $\dots v x^m v^{-1} \dots$ into $\dots x^m \dots$. The same arguments as before show that if $[w]$ is such a word, then

$$[w] = [vw']$$

where w' is composed of building blocks from $\left(X \setminus \left\{x^{\pm 1}, v^{\pm 1}\right\}\right)^{\pm 1}$ together with $\left\{vx^m v^{-1} \,\middle|\, 0 \neq m \in \mathbb{Z}\right\}$. Every letter in w' is followed by one of at most $(2k-4)$ possible letters, showing this type of words also has exponential growth rate $< 2k-3$, and thus completing the proof. □

Lemma 4.3.7.

$$\limsup_{N\to\infty} \sqrt[N]{|M_{k,N}|} < 2k-3.$$

Proof. Every $[w] \in M_{k,N}$ is equipped with some triplet (Y, Z, v) from the definition of $M_{k,N}$ (so $Y = \{x\}$ or $Y = \{x, x^{-1}\}$ for some $x \in X^{\pm 1}$). First, we observe that the corresponding Whitehead automorphism $\phi_{Y,Z,v}$ shortens $[w]$ by at least 4. There are in total $2k(2k-2)$ triplets with $Y = \{x\}$ and $k(2k-2)$ triplets with $Y = \{x, x^{-1}\}$. Let $\mathcal{W}$ denote the set of these $2k(2k-2) + k(2k-2) = 6k(k-1)$ possible Whitehead automorphisms.

In other words, for every $[w] \in M_{k,N}$ there exists $\phi_1 \in \mathcal{W}$ such that $|\phi_1(w)|_c = N' \leq N-4$. If $[\phi_1(w)] \in M_{k,N'}$ we apply the corresponding automorphism $\phi_2 \in \mathcal{W}$ and obtain a cyclic word of

length $\leq N-8$. Since for all n we have $C_{k,n} = L_{k,n} \cup M_{k,n} \cup M_{k,n}^c$, one can continue this process until the resulting cyclic word is either in $M_{k,n}^c$ or in $\widetilde{L}_{k,n}$ for some $2 \leq n \leq N$ (note that each cyclic word in $M_{k,*}$ is of length ≥ 8). Let $[\widehat{w}] \in M_{k,n}^c \cup \widetilde{L}_{k,n}$ be the cyclic-word we obtain this way, i.e.

$$[w] \stackrel{\phi_1}{\to} [w_1] \stackrel{\phi_2}{\to} [w_2] \to \ldots \stackrel{\phi_r}{\to} [\widehat{w}]\,.$$

Each element in $M_{k,N}$ is uniquely determined by r, $n = |\widehat{w}|_c$, $[\widehat{w}] \in M_{k,n}^c \cup \widetilde{L}_{k,n}$ and $(\phi_1, \phi_2, \cdots, \phi_r) \in \mathcal{W}^r$, where $r \leq R = \left\lfloor \frac{N-n}{4} \right\rfloor$. For each n, the number of possible tuples $(\phi_1, \phi_2, \cdots, \phi_r)$ is

$$\sum_{i=0}^{R} |\mathcal{W}|^i = \sum_{i=0}^{R} [6k\,(k-1)]^i \leq \frac{[6k\,(k-1)]^{R+1} - 1}{6k\,(k-1) - 1} \leq 2\,[6k\,(k-1)]^R\,.$$

By Lemmas 4.3.5 and 4.3.6, the possible number of $[\widehat{w}]$ of length n is bounded from above by $C \cdot (2k-3-\varepsilon)^n$ for some $C, \varepsilon > 0$. Summing over all possible values of n we obtain:

$$\begin{aligned} |M_{k,N}| &\leq 2C \cdot \sum_{n=2}^{N} (2k-3-\varepsilon)^n \cdot [6k\,(k-1)]^R \\ &\leq 2C \cdot \sum_{n=2}^{N} (2k-3-\varepsilon)^n \cdot (\sqrt[4]{6k(k-1)})^{N-n}. \end{aligned}$$

For $k \geq 3$ we can pick ε small enough so that

$$\sqrt[4]{6k\,(k-1)} < 2k-3-\varepsilon,$$

hence

$$|M_{k,N}| < 2C \cdot N \cdot (2k-3-\varepsilon)^N\,.$$

□

This completes the proof of (4.3.1), hence also of Theorem 4.1.3. Theorem 4.1.1 now follows by Proposition 4.3.1.

4.3.4 The growth of non-primitives belonging to free factors

Finally, let us say some words about the variation of the proof required for Corollary 4.1.4. Recall that $S_{k,N}$ denotes the set of words of length N in $\mathbf{F}_k$ belonging to a proper free factor. We ought to show that $|S_{k,N}|$ grows exponentially with base $(2k-3)$. We already mentioned on Page 126 why $(2k-3)$ is a lower bound. To show it is also an upper bound, we repeat similar arguments as above[†]:

Firstly, the same argument as in Section 4.3.1 shows the exponential growth of $S_{k,N}$ is the same as the exponential growth of $\overline{S_{k,N}}$, the set of cyclic-words of length N belonging to a proper free factor. By Theorem 4.2.1, each $[w] \in \overline{S_{k,N}}$ belongs to some $A_{Y,Z,v}^N$. For most triplets, Proposition 4.2.3 shows they grow slower than $(2k-3)^N$. When $Z = \left\{v^{-1}\right\}$, $A_{Y,Z,v}^N \subseteq \overline{S_{k,N}}$ and grows exponentially with base $(2k-3)$.

$\overline{S_{k,N}}$

Consider next words in $A_{Y,Z,v}^N$ such that either

- $Y = \left\{x, x^{-1}\right\}$ and there is exactly one instance of vx^mv^{-1}, or

[†]In fact, the proof of Corollary 4.1.4 alone could be shorter than the proof of Theorem 4.1.3. The same shorter proof would show that $\limsup_{N\to\infty} \sqrt[N]{C_{k,N}} = 2k-3$. In other words, much of the complexity of the analysis in Sections 4.3.2 and 4.3.3 is required only for showing the stronger result that the growth rate of $\left|C_{k,N} \setminus L_{k,N}\right|$ is strictly smaller than $2k-3$.

- $Y = \{x\}$ and there are up to three instances of $x^{\pm 1}$.

It is evident that this set of words has exponential growth rate $(2k-3)$.

The remaining words from $\overline{S_{k,N}}$, which we denote by $Q_{k,N}$, can be described in a similar fashion to the cyclic words from $M_{k,N}$. In a similar argument as in Lemma 4.3.7, we can shorten each word from $Q_{k,N}$ by the corresponding Whitehead automorphisms until we get a word outside $Q_{k,N}$. Since we have already seen that $\left|\overline{S_{k,N}} \setminus Q_{k,N}\right|$ has exponential growth rate $(2k-3)$, we can complete the proof of Corollary 4.1.4 in the same manner we proved Lemma 4.3.7. $\square$

4.4 Most Triplets are Negligible

The last section is dedicated to proving Proposition 4.2.3, stating that for most triplets, the set $A_{Y,Z,v}$ has exponential growth rate strictly smaller than $(2k-3)$. This is done by way of considering different cases according to the cardinalities $|Y|$ and $|Z|$, and treating each case separately. To simplify the notation we denote $|Y|$ and $|Z|$ by y and z, respectively. Note that $y+z = 2k-1$. The assumptions of Proposition 4.2.3 are that $y, z \geq 2$ and $Y \neq \{x, x^{-1}\}$. The main technique is to rely on the following intuitive lemma. We call a set of words in $\mathbf{F}_k$ *Markovian* if it is closed under taking prefixes and if to every $x \in X^{\pm 1}$ corresponds a fixed subset $\Sigma_x \subseteq X^{\pm 1}$ of letters which can follow x. Namely, if $w \in A$ is of length N and terminates with x, one can extend it to a word in A of length $(N+1)$ by appending one of the letters from Σ_x. Obviously, the sets $A_{Y,Z,v}$ are all Markovian (to be precise, the set of all cyclically reduced representatives of the cyclic words in some $A_{Y,Z,v}$ is Markovian).

Lemma 4.4.1. *Let A be a* Markovian *set of words in $\mathbf{F}_k$ and let $\alpha > 1$. Assume that for each letter $x \in X^{\pm 1}$ there is some $1 \leq r = r(x) \in \mathbb{N}$ such that x is followed by one of less than α^r possible r-tuples of letters. Then the exponential growth of A is less than α.*

Proof. For every $x \in X^{\pm 1}$ and $1 \leq i \leq r(x)$ let $T_{x,i}$ denote the number of possible i-tuples which can follow x in words from A. (In particular, $T_{x,r(x)} < \alpha^{r(x)}$.) For $w_1, w_2 \in A$ we say that w_2 is an *i-extension* of w_1 if w_1 is a prefix of w_2 and $|w_2| - |w_1| = i$.

Let A_N be the set of words of length N in A. Define a subset $B \subseteq A$ by the following recursive rules: $A_1 \subseteq B$ and if $w \in B$ terminates with x, then

- all i-extensions of w for $1 \leq i \leq r(x)-1$ do *not* belong to B (there are $T_{x,1}+T_{x,2}+\ldots+T_{x,r(x)-1}$ such words), and
- all the $T_{x,r(x)}$ words which are $r(x)$-extensions of w belong to B.

Define $f : A \to \mathbb{R}$ as follows: for every $w \in B$ terminating with the letter x,

- set $\mathrm{f}(w) = 1$, and
- for every $1 \leq i \leq r(x) - 1$ and every i-extension u of w, set

$$f(u) = \frac{\left(T_{x,r(x)}\right)^{i/r(x)}}{T_{x,i}} < \frac{\alpha^i}{T_{x,i}}.$$

Now, set $g(N) = \sum_{w \in A_N} f(w)$. For $w \in B$ terminating with x and $1 \leq i \leq r(x)$, the sum of f over all i-extensions of w is less than α times the sum over all $(i-1)$-extensions of w. We obtain that $g(N+1) < \alpha \cdot g(N)$, so the exponential growth rate of g is $< \alpha$. We end the proof by claiming that $c < \frac{g(N)}{|A_N|} < C$ for some positive constants c, C. Indeed, one can set

$$c = \min_{\substack{x \in X^{\pm 1} \\ 1 \leq i \leq r(x)}} \frac{\left(T_{x,r(x)}\right)^{i/r(x)}}{T_{x,i}}, \quad C = \max_{\substack{x \in X^{\pm 1} \\ 1 \leq i \leq r(x)}} \frac{\left(T_{x,r(x)}\right)^{i/r(x)}}{T_{x,i}}.$$

$\square$

We now return to the proof of Proposition 4.2.3. We shall use Lemma 4.4.1 for the sets $A_{Y,Z,v}$ with $\alpha = 2k-3$.

Case 1: $|Y|, |Z| \geq 4$: Let $[w] \in A_{Y,Z,v}$ and let $x \in X^{\pm 1}$ appear in w. If $x \in Y$ then either the inverse of the following letter is v or it belongs to $Y \setminus \{x\}$, so there are y options, and $y = 2k-1-z \leq 2k-5$. The same argument applies for $x \in Z$. Finally, the only letter that cannot follow v is v^{-1}, so v is followed by:

- one of y letters from Y, which are followed in turn by one of y letters, or
- one of $z-1$ letters from $Z \setminus \left\{v^{-1}\right\}$, which are followed in turn by one of z letters, or
- v, which is followed by one of $2k-1$ letters.

Overall, there are $y^2 + z(z-1) + (2k-1)$ possibilities for the two letters following v. It is easy to see that under the assumptions in the current case, this expression is largest when $y = 2k-5$ and $z = 4$. But even in this case,

$$y^2 + z(z-1) + (2k-1) = (2k-5)^2 + 12 + (2k-1) < (2k-3)^2,$$

(note that if $y, z \geq 4$ then $k \geq 5$). This completes the proof.

Case 2: $|Y| = 3$: Assume first that $k \geq 5$, so $2k-3 \geq 7$. As in the previous proof, let $x \in X^{\pm 1}$ be a letter in a word from $A_{Y,Z,v}$. If $x \in Y$, the following letter is one of $y = 3$ possibilities. If $x \in Z$ there are at most $z = 2k-4$ possibilities. Finally, if $x = v$, then v is followed by

- one of 3 letters from Y, which are followed in turn by one of 3 letters, or
- one of $2k-5$ letters from $Z \setminus \left\{v^{-1}\right\}$, which are followed in turn by one of $2k-4$ letters, or
- v, which is followed by one of $2k-1$ letter.

Overall, there are $3^2 + (2k-5)(2k-4) + (2k-1)$ possibilities for the two letters following v. This is strictly less than $(2k-3)^2$ for $k \geq 5$.

Suppose next that $k = 4$, so now $2k-3 = 5$. Any $x \in X^{\pm 1} \setminus \{v\}$ is followed by at most 4 possible letters. As for v itself, we need to distinguish between two cases: either Y does not contain a letter and its inverse, in which case w.l.o.g. $Y = \{a, b, c\}$ and $Z = \left\{a^{-1}, b^{-1}, c^{-1}, v^{-1}\right\}$; or w.l.o.g. $Y = \left\{a, a^{-1}, b\right\}$ and $Z = \left\{b^{-1}, v^{-1}, c^{\pm 1}\right\}$. In the first case an easy computation[†] shows that v is followed by one of 115 possible triplets of letters, and we are done as $115 < 5^3$. In the second case, v is followed by one of 617 possible quadruplets, and $617 < 5^4$.

Finally, if $y = 3$ and $k = 3$ (k cannot be smaller than 3 if $y = 3$), then $2k-3 = 3$. The partition is, up to name changes, $Y = \left\{a, a^{-1}, b\right\}$ and $Z = \left\{b^{-1}, v^{-1}\right\}$. An easy computation shows that any letter $x \neq v$ is followed by at most 8 possible pairs of letters, and v is followed by at most 17,883 possible 9-tuples (and $17{,}833 < 3^9 = 19{,}683$).[‡]

Case 3: $|Z| = 3$ Assume first that $k \geq 4$, hence $2k-3 \geq 5$. If $x \in Y$, it is followed by the inverse of one of $2k-5$ letters from $Y \setminus \{x\}$ or by v^{-1}, a total of $2k-4$ possibilities. A letter from Z is followed by one of $3 < 2k-3$ letters (v^{-1} and two letters whose inverse belongs to Z).

To analyze the number of possibilities after v, we distinguish between two cases:

(1) Assume that Z contains a letter and its inverse, i.e. $Z = \left\{v^{-1}, a, a^{-1}\right\}$. In this case,

- Every $x \in Y$, is followed either by one of $2k-5$ letters from $Y \setminus \left\{x^{-1}\right\}$ which are then followed by one of $(2k-4)$ letters, or by v^{-1} which is followed by one of 3 possible letters: a total of $(2k-5)(2k-4)+3 = 4k^2-18k+23$ possible pairs.

[†]Computation of this kind can be easily carried out in some Excel-type spreadsheet program.

[‡]Alternatively, one can show that the exponential growth rate here equals the largest real root of $\lambda^5 - 3\lambda^4 + \lambda^3 - \lambda^2 - \lambda + 7$, which is about 2.68.

- The letter a is followed either by one of two letters from Z (a or v^{-1}), which are in turn followed by one of 3 possible letters, or by v which is followed by one of $2k-1$ letters: a total of $2 \cdot 3 + (2k-1) = 2k+5$ possible pairs. The same computation holds for a^{-1}.
- Every pair of letters following v starts either with some $x \in Z$ ($2 \cdot 3$ options), by v ($2k-1$ options) or by some $x \in Y$ ($(2k-4)^2$ options): a total of $6 + 2k - 1 + (2k-4)^2 = 4k^2 - 14k + 21$ possible pairs.

We can now count the number of possible triplets of letters following v:

- $2(2k+5)$ triplets begin with a or a^{-1}.
- $(2k-4)\left(4k^2-18k+23\right)$ triplets begin with some $x \in Y$.
- $4k^2 - 14k + 21$ triplets begin with v.

The total number of possible triplets following v is, therefore, $8k^3 - 48k^2 + 108k - 61$ which is strictly less than $(2k-3)^3$ when $k \geq 4$.

(2) The other possibility is that Z does not contain a letter and its inverse, hence $Z = \left\{v^{-1}, a, b\right\}$. A similar computation shows that v is followed in this case by one of $8k^3 - 56k^2 + 160k - 145$ which is again strictly less than $(2k-3)^3$ when $k \geq 4$.

Finally, if $k = 3$ then $Z = \left\{v^{-1}, a, b\right\}$ and $Y = \left\{a^{-1}, b^{-1}\right\}$ (for otherwise $Y = \left\{x, x^{-1}\right\}$). Another technical computation shows that a^{-1} and b^{-1} are followed by at most 2 possible letters, v^{-1} by at most 7 possible pairs of letters, a and b by at most $237 < 3^5$ 5-tuples, and v by at most $41{,}372{,}449 < 3^{16}$ 16-tuples of letters[†].

Case 4: $|Y| = 2$ Since $Y \neq \left\{x, x^{-1}\right\}$, assume w.l.o.g. that $Y = \{a, b\}$. The possibility $k = 3$ was already dealt with in the previous case, so assume $k \geq 4$. Similar calculations to those above show that:

- a and b are followed by one of $2 < (2k-3)$ letters.
- a^{-1} and b^{-1} are followed by one of $\left(4k^2-14k+16\right) < (2k-3)^2$ pairs of letters.
- Any letter x such that $x, x^{-1} \in Z$ is followed by one of $\left(4k^2-16k+21\right) < (2k-3)^2$ pairs of letters.
- v^{-1} is followed by one of $\left(4k^2-16k+19\right) < (2k-3)^2$ pairs of letters.
- v is followed by one of $\left(8k^3-44k^2+106k-91\right)$ triplets of letters. This is strictly less than $(2k-3)^3$ for $k \geq 5$. For $k = 4$, a concrete computation shows v is followed by one of $613 < 5^4$ possible 4-tuples of letters.

Case 5: $|Z| = 2$ Assume w.l.o.g. that $Z = \left\{a, v^{-1}\right\}$. The case $k = 3$ was handled in the case $|Y| = 3$, so assume $k \geq 4$. Again, the following formulas can be easily computed:

- a and v^{-1} are each followed by one of $2 < (2k-3)$ letters.
- a^{-1} is followed by one of $\left(4k^2-14k+14\right) < (2k-3)^2$ pairs of letters.
- Any letter x such that $x, x^{-1} \in Y$ is followed by one of $\left(4k^2-16k+19\right) < (2k-3)^2$ pairs of letters.

[†]Alternatively, the exponential growth rate here equals the largest real root of $\lambda^4 - 2\lambda^3 - 4\lambda^2 + 2\lambda + 7$, which is about 2.85.

- v is followed by one of $\left(8k^3 - 40k^2 + 80k - 51\right)$ triplets of letters. This is strictly less than $(2k-3)^3$ for $k \geq 6$. For $k = 5$, a concrete computation shows v is followed by one of $1{,}951 < 7^4$ possible 4-tuples of letters, and for $k = 4$, v is followed by one of $557 < 5^4$ possible 4-tuples of letters.

This finishes the proof of Proposition 4.2.3. □

4.5 Open Questions

Finally, we mention the following closely related questions which are still open:

Question 4.5.1. *What can be said about the growth of* $\mathrm{Aut}(F_k)$ *with respect to standard generating sets such as*
i) Nielsen moves?
ii) Whitehead automorphisms?

Question 4.5.2. *What is the smallest possible exponential growth rate of* $P_{k,N}$ *(or* $C_{k,N}$*) with respect to arbitrary finite generating sets of* F_k *(not necessarily bases)?*

Question 4.5.3. *What is the growth of* $C_{k,N} \setminus L_{k,N}$*? What does a generic primitive cyclic element containing every letter at least twice (or not at all) look like? Is it, up to permuting the letters, of the form* $x_1 w(x_1 x_2, x_3, x_4, \ldots, x_k)$*? (In particular, the latter set of words shows that the growth of* $C_{k,N} \setminus L_{k,N}$ *is at least* λ_k*, the largest root of* $\lambda^3 - (2k-5)\lambda^2 - \lambda - (2k-3)$*, which satisfies* $\lambda_k \searrow 2k-5$ *as* $k \to \infty$*.)*

Question 4.5.4. *What is the growth of other* $\mathrm{Aut}(\mathbf{F}_k)$*-orbits in* $\mathbf{F}_k$*? Which orbits, other than that of the primitives, have the largest growth?*
We conjecture the following is true: For $w \in \mathbf{F}_k$*, let* $\mu(w)$ *denote the minimal (positive) number of instances of some letter* $x \in X$ *in any element of the* $\mathrm{Aut}(\mathbf{F}_k)$*-orbit of* w *(this number does not depend on* x*). Then the growth of the set of cyclic words in the orbit of* w *is* $\sqrt[\mu(w)]{2k-3}$*. If true, this shows that unless* w *is primitive, most words in its orbit are conjugates of small words, so that the growth of the orbit is always* $\sqrt{2k-1}$*.*

References

[BMS] G. Baumslag, A. Myasnikov, and V. Shpilrain. Open problems in combinatorial and geometric group theory. *http://www.grouptheory.info/*.

[BMS02a] G. Baumslag, A. Myasnikov, and V. Shpilrain. Open problems in combinatorial group theory. *Contemporary Mathematics*, 296:1–38, 2002.

[BMS02b] A.V. Borovik, A.G. Myasnikov, and V. Shpilrain. Measuring sets in infinite groups. In *Computational and Statistical Group Theory*, Contemp. Math., pages 21–42, Las Vegas, NV/Hoboken, NJ, 2002. American Mathematical Society.

[BV02] J. Burillo and E. Ventura. Counting primitive elements in free groups. *Geometriae Dedicata*, 93(1):143–162, 2002.

[LS70] R.C. Lyndon and P.E. Schupp. *Combinatorial Group Theory*. Springer-Verlag, Berlin; New York, 1970.

[MS03] A. Myasnikov and V. Shpilrain. Automorphic orbits in free groups. *Journal of Algebra*, 269(1):18–27, 2003.

[PP15] D. Puder and O. Parzanchevski. Measure preserving words are primitive. *Journal of the American Mathematical Society*, 28(1):63–97, 2015.

[Pud14] D. Puder. Primitive words, free factors and measure preservation. *Israel Journal of Mathematics*, 201(1):25–73, 2014.

[Pud15a] D. Puder. Expansion of random graphs: New proofs, new results. *Inventiones Mathematicae*, 2015. to appear. arXiv:1212.5216.

[Riv04] I. Rivin. A remark on 'counting primitive elements in free groups' (by j. burillo and e. ventura). *Geometriae Dedicata*, 107(1):99–100, 2004.

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

[Sta99] J.R. Stallings. Whitehead graphs on handlebodies. In J. Cossey, C. F. Miller, W.D. Neumann, and M. Shapiro, editors, *Geometric group theory down under*, pages 317–330. de Gruyter, Berlin, 1999.

[Whi36a] J.H.C. Whitehead. On certain sets of elements in a free group. *Proc. London Math. Soc.*, 41:48–56, 1936.

[Whi36b] J.H.C. Whitehead. On equivalent sets of elements in a free group. *Ann. of Math.*, 37:768–800, 1936.

Chapter 5

Stallings Graphs, Algebraic Extensions and Primitive Elements in $\mathbf{F}_2$

Ori Parzanchevski[†] Doron Puder[‡]
Einstein Institute of Mathematics
Hebrew University, Jerusalem
parzan@math.huji.ac.il doronpuder@gmail.com

Published: Mathematical Proceedings of the Cambridge Philosophical Society, 157 (1), 2014, pp 1-11. DOI: 10.1017/S0305004114000097

Abstract

We study the free group of rank two from the point of view of Stallings core graphs. The first half of the paper examines primitive elements in this group, giving new and self-contained proofs for various known results about them. In particular, this includes the classification of bases of this group. The second half of the paper is devoted to constructing a counterexample to a conjecture by Miasnikov, Ventura and Weil, which seeks to characterize algebraic extensions in free groups in terms of Stallings graphs.

Contents

[†]Supported by an Advanced ERC Grant, and the ISF.
[‡]Supported by the Adams Fellowship Program of the Israel Academy of Sciences and Humanities, and an Advanced ERC Grant.

5.1 Introduction

Let $\mathbf{F}$ be a finitely generated free group. A subgroup J of $\mathbf{F}$ is said to be an *algebraic extension* of another subgroup H, if $H \leq J$ and there does not exist an intermediate subgroup $H \leq M \lneq J$ such that M is a proper free factor of J. We denote this by $H \leq_{alg} J$. This notion, which was formulated independently by several authors (and already appears in [Tak51]), is central to 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 the algebraic extensions of H, which is a well known theorem of Takahasi [Tak51]. For proofs of the mentioned facts, as well as a general survey of algebraic extensions, we refer the reader to [MVW07].

$H \leq_{alg} J$

Given a basis X of $\mathbf{F}$ and $H \leq \mathbf{F}$, we denote by $\Gamma_X(H)$ the *Stallings core graph* of H with respect to X. This is a pointed, directed, X-labeled graph, such that the words formed by closed paths around the basepoint are precisely the elements of H, and which is minimal with respect to this property. One way to construct this graph is by taking the Schreier right coset graph of H in $\mathbf{F}$ w.r.t. X and then deleting all "hanging trees", i.e., all edges which are not traced by some non-backtracking loop around the basepoint. Figure 5.1.1 demonstrates the core graph of $H = \langle ab^{-1}a, a^{-2}b \rangle$ for $X = \{a, b\}$ and $\mathbf{F} = \mathbf{F}(X)$. We refer to [Sta83, KM02, MVW07, Pud14] for further background on Stallings graphs.

$\Gamma_X(H)$

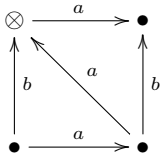


Figure 5.1.1: The core graph $\Gamma_X(H)$ where $X = \{a, b\}$ and $H = \langle ab^{-1}a, a^{-2}b \rangle \leq \mathbf{F}(X)$.

Given the basis X, and two subgroups $H, J \leq \mathbf{F}$, there is a graph morphism (which preserves the basepoint, directions and labeling) from $\Gamma_X(H)$ to $\Gamma_X(J)$ if and only if $H \leq J$. Such a morphism is unique, when it exists. Given $H, J \leq \mathbf{F}$, we say that H *X-covers* J if $H \leq J$ and the morphism from $\Gamma_X(H)$ to $\Gamma_X(J)$ is onto. We denote this by $H \leq_{\vec{x}} J$. (In [MVW07] this is indicated by saying that J is a "X-principal overgroup" of H, and by the notation $J \in \mathcal{O}_X(H)$.)

X-covers $\leq_{\vec{x}}$

It is not hard to see (e.g. [MVW07, prop. 3.7], or [PP15, claim 4.2]) that if $H \leq_{alg} J$, then $H \leq_{\vec{x}} J$ for every basis X of $\mathbf{F}$. The following conjecture, raised in [MVW07], asks whether the converse also holds.

Conjecture ([MVW07, §5(1)]). *If $H \leq J \leq \mathbf{F}$ and $H \leq_{\vec{x}} J$ for every basis X of $\mathbf{F}$ then J is an algebraic extension of H.*

The main result of this paper is a counterexample to this conjecture:

Proposition (Prop. 5.4.1). *Let $\mathbf{F}_2 = \mathbf{F}(a, b)$ be the free group on two generators, $H = \langle a^2b^2 \rangle$, and $J = \langle a^2b^2, ab \rangle$. Then $H \leq_{\vec{x}} J$ for every basis X of $\mathbf{F}_2$, but J is not an algebraic extension of H.*

The relation "$H \leq_{\vec{x}} J$" is basis-dependent, while the relation "$H \leq_{\vec{x}} J$ for every basis X" is intrinsic, as is "$H \leq_{alg} J$". Proposition 5.4.1 means that the latter two relations are different, and this raises the intriguing question of understanding the algebraic significance of "covering with respect to all bases".

The proof of Proposition 5.4.1 follows from a thorough analysis of Stallings graphs, using classical results (e.g. [Nie17, Coh72, CMZ81, OZ81]) on primitive elements and bases of $\mathbf{F}_2$. It turns out that these results can also be proven by appealing solely to Stallings graphs, and we use the opportunity to provide self-contained proofs for them in Section 5.3. Section 5.2 recalls some basic facts about

Stallings graphs and foldings, and presents two auxiliary lemmas which will be used later on. Finally, the proof of the counterexample (Proposition 5.4.1) is given in Section 5.4, and some concluding remarks in Section 5.5.

5.2 Stallings Graphs

We assume that the reader is familiar with the theory of Stallings foldings, but recall the basic facts. If Γ is a pointed, directed, X-labeled graph, we denote by $\pi_1^X(\Gamma)$ the subgroup of $\mathbf{F} = \mathbf{F}(X)$ consisting of the words which appear as closed loops around the basepoint of Γ. The operators π_1^X and Γ_X constitute a bijection between subgroups of $\mathbf{F}(X)$ and X-labeled core graphs, which matches f.g. subgroups to finite graphs. $\pi_1^X(\Gamma)$

If Γ is a finite (pointed, directed) X-labeled graph, and $\pi_1^X(\Gamma) = H$, then $\Gamma_X(H)$ is obtained from Γ by repeatedly performing one of the following operations, in any order, until neither of them is possible:

(1) *Folding* - merging two edges with the same label, and the same origin or terminus (and thus merging also the other ends).

(2) *Trimming* - deleting a leaf which is not the basepoint, and the edge which leads to it.

The following lemma shows that under certain conditions only foldings are necessary in this process:

Lemma 5.2.1. *Let Γ be a finite, pointed, directed, X-labeled graph such that at every vertex, except possibly the basepoint, there are at least two types of edges (the type of an edge consists of its label and direction). Then the core graph $\Gamma_X(H)$ of $H = \pi_1^X(\Gamma)$ is obtained from Γ by foldings alone (i.e. without trimming).*

Proof. Evidently, Γ cannot have leaves, except for possibly the basepoint. Folding steps do not decrease the number of types of edges at a vertex, so that the property in the statement still holds after every folding step, and no new leaves are created throughout the process. □

This simple lemma will prove out to be extremely useful. It already plays a role in Lemma 5.2.3, which characterizes X-covering in simple extensions.

Definition 5.2.2. Let Γ be a pointed and directed X-labeled graph and let $w \in \mathbf{F}$. We say that *w appears in Γ* if there exist paths p_1, p_2 in Γ such that p_1 starts at the basepoint, p_2 terminates at the basepoint, and $w = p_1 p_2$ (i.e. $p_1 p_2$ is the presentation of w as a reduced word in $X \cup X^{-1}$). *appears in*

For example, for H in Figure 5.1.1, a^3 and $a^2 b a^{-1}$ appear in $\Gamma_X(H)$, but $a^2 b^2$ does not. Notice that if w appears in Γ, s.t. $\pi_1^X(\Gamma) = H$, and Γ satisfies the conditions of Lemma 5.2.1, then w appears in $\Gamma_X(H)$ as well. This will play a significant part in Section 5.4.

Lemma 5.2.3. *Let $H \leq \mathbf{F}$, $w \in \mathbf{F}$ and $J = \langle H, w \rangle$. Then $H \leq_{\vec{x}} J$ if and only if w appears in $\Gamma_X(H)$.*

Proof. Assume first that w appears in $\Gamma_X(H)$, and let p_1, p_2 be as in Definition 5.2.2. Denote by Γ the graph obtained from $\Gamma_X(H)$ by identification of p_1's endpoint and p_2's start-point. We have $\pi_1^X(\Gamma) = J$, and the (pointed, directed, labeled) map from $\Gamma_X(H)$ to Γ is onto. Since Γ satisfies the conditions of Lemma 5.2.1, $\Gamma_X(J)$ is obtained from it by foldings alone. We have now that $\Gamma_X(H)$ maps onto Γ, which maps onto $\Gamma_X(J)$, and by transitivity it follows that $H \leq_{\vec{x}} J$.

Assume now that w does not appear in $\Gamma_X(H)$. Let p_1 be the maximal path beginning at the basepoint of $\Gamma_X(H)$ which is a prefix of w, and denote by v_1 its endpoint. Let p_2 be the maximal path ending at the basepoint of $\Gamma_X(H)$ which is a suffix of w, and v_2 its beginning. If $w = p_1 w' p_2$, take $\Gamma = \Gamma_X(H) \bigcup p_{w'}$ where $p_{w'}$ is a path labeled by w', whose beginning is attached to v_1, and whose endpoint to v_2 (see Figure 5.2.1). Now $\pi_1^X(\Gamma) = J$ and Γ has no foldable edges nor leaves, i.e. $\Gamma = \Gamma_X(J)$. Thus $\Gamma_X(H)$ is a subgraph of $\Gamma_X(J)$, and in particular does not map onto it. (In fact, since the map from $\Gamma_X(H)$ to $\Gamma_X(J)$ is injective, H is a free factor of J.) □

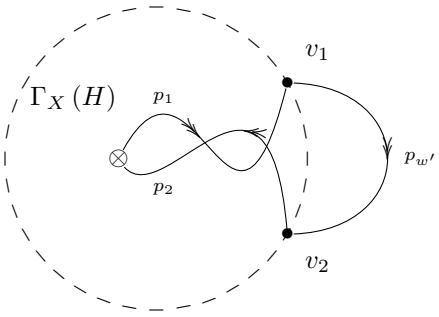


Figure 5.2.1: $\Gamma = \Gamma_X(H) \bigcup p_{w'}$

Remark 5.2.4. With some further work, the basic idea of Lemma 5.2.3 can lead to an algorithm to detect primitive words and free factors in $\mathbf{F}$. See [Pud14, Thm 1.1].

5.3 Primitives in $\mathbf{F}_2$

In this section we give new proofs for the classical theorems on primitive words and bases of $\mathbf{F}_2$ [Nie17, Coh72, CMZ81, OZ81]. Throughout the section X denotes the basis $\{a, b\}$ of $\mathbf{F}_2 = \mathbf{F}(a, b)$.

We start with the following lemma, which reduces the classification of bases of $\mathbf{F}_2$ to that of cyclically reduced (henceforth: CR) bases. *CR*

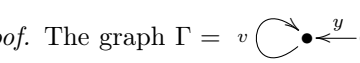
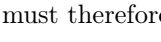
Lemma 5.3.1. *Let $Y = \{\overline{u}, \overline{v}\}$ be any basis of $\mathbf{F}_2$.*

(1) Write $\overline{u} = xux^{-1}$ † and $\overline{v} = yvy^{-1}$ with u, v CR. Then either x is a prefix of y or y is a prefix of x.

(2) Assume that x is a prefix of y, and write $\overline{u} = xux^{-1}$ and $\overline{v} = xwvw^{-1}x^{-1}$. Then w is a prefix of some power of u or of u^{-1} (which implies that $w^{-1}uw$ is a cyclic rotation of u).

(3) The basis $(xw)^{-1} Y xw = \left\{w^{-1}uw, v\right\}$ is CR.

Therefore, any basis of $\mathbf{F}_2$ is of the form $\left\{xux^{-1}, xwvw^{-1}x^{-1}\right\}$ where w is a prefix of some power of $u^{\pm 1}$, $\left\{w^{-1}uw, v\right\}$ is a CR basis, and x is any word s.t. xux^{-1} and $xwvw^{-1}x^{-1}$ are reduced.

Proof. The graph $\Gamma =$  satisfies $\pi_1^X(\Gamma) = \langle \overline{u}, \overline{v} \rangle = \mathbf{F}_2$. It also satisfies the conditions of Lemma 5.2.1, and must therefore fold into $\Gamma_X(\mathbf{F}_2) =$  . The only vertex at which folding may occur is the basepoint $\otimes$, and after the first identification of edges the only possible folding place is at the identified termini. Continuing in this manner shows that for Γ to fold into $\Gamma_X(\mathbf{F}_2)$, the shorter word among x, y must be completely merged with a prefix of the longer one, giving (1). By the same arguments, the graph $\Gamma' =$  must fold into $\Gamma_X(\mathbf{F}_2)$, and for this to happen w must wind itself completely around u, or around u^{-1} (i.e. w must be a prefix of some power of u or of u^{-1}). It follows that $w^{-1}uw$ is a cyclic rotation of u, and $\left\{w^{-1}uw, v\right\}$ is a CR basis. □

Moving on to CR bases, we have the following:

Proposition 5.3.2. *Let $\{u, v\}$ be a CR basis of $\mathbf{F}_2$, such that $|u| + |v| \geq 3$ ‡, and $|u| \leq |v|$. Then either u is a prefix or a suffix of v, or u^{-1} is.*

†By "write" we mean that xux^{-1} is a reduced expression of $\overline{u}$ - no cancellation is needed. This convention will repeat throughout the paper, and we will not mention it again.

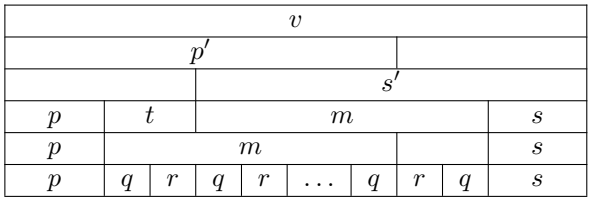
‡Here $|w|$ is the length of w as a reduced word in $X \cup X^{-1}$.

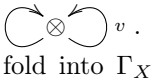
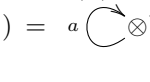
Proof. Since $|v| \geq 2$ and v is not a proper power, v contains both a and b (possibly with negative exponents), and thus $\langle v \rangle \leq_{\vec{x}} \mathbf{F}_2$. Since $\mathbf{F}_2 = \langle u, v \rangle$, Lemma 5.2.3 implies that u appears in $\Gamma_X(\langle v \rangle)$, which is just a cycle labeled by v as a word in $X \cup X^{-1}$.

Let p' be the maximal prefix of u which is a path emanating from the basepoint of $\Gamma_X(\langle v \rangle)$. Since $|u| \leq |v|$, this means that p' is a prefix of v or of v^{-1}. By inverting v if necessary we assume that p' is a prefix of v. Let s' be the maximal suffix of u which is a path ending at the basepoint of $\Gamma_X(\langle v \rangle)$. Since u is CR, s' cannot be a suffix of v^{-1}, and must be a suffix of v.

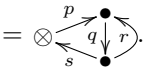
Let m be the middle part of u where p' and s' overlap (it may be empty: $|m| = |p'| + |s'| - |u| \geq 0$). Write $p' = pm$, $s' = ms$, which means that $u = pms$ (see Figure 5.3.1). Thus, if p is empty then $u = pms = s'$ is a suffix of v, and if s is empty then u is a prefix of v. We proceed to show that they cannot be both nonempty.

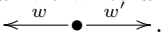
Figure 5.3.1: Illustration of the decomposition of v.



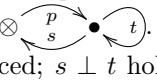
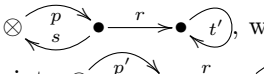
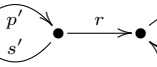
Let $\Gamma =$  . Since $\pi_1^X(\Gamma) = \langle u, v \rangle = \mathbf{F}_2$ and Γ satisfies the conditions of Lemma 5.2.1, it must fold into $\Gamma_X(\mathbf{F}_2) =$  , and we will show that this cannot happen if $p, s \neq 1$.

Assume therefore that $p, s \neq 1$. Since $|v| \geq |u|$ we can write $v = ptms$, and $t \neq 1$ since otherwise $u = v$ (this shows, in particular, that $|v| > |u|$). Since pm is a prefix of v, m is a prefix of tm, which means that m is a prefix of some (positive) power of t (see again Figure 5.3.1). We consider two cases:

Case (i): m is not a power of t. In this case $t = qr$ with $q, r \neq 1$, and $m = (qr)^n q$ with $n \geq 0$ (see Figure 5.3.1; $n = 0$ corresponds to the possibility that p' and s' do not overlap in v). Since $u = p(qr)^n qs$ and $v = p(qr)^{n+1} qs$, Γ folds into $\Gamma' =$  .

We aim to show that no folding can occur in Γ', but let us first introduce the following notations: for $w \in \mathbf{F}(X)$, we denote by w_1 the first letter of w as a reduced word in $X \cup X^{-1}$, and for two words w, w' we write $w \perp w'$ to indicate that $w_1 \neq (w')_1$. Namely, $w \perp w'$ implies that no folding occurs in  . ⊥

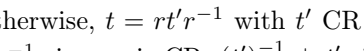
Returning to Γ', we have $p \perp s^{-1}$ since u (or equivalently v) is CR, so no folding occurs at $\otimes$. Since $v = p(qr)^{n+1} qs$ is reduced, $p^{-1}, r^{-1} \perp q$ and $q^{-1} \perp r, s$. We also have $r \perp s$, for otherwise $p's_1 = p(qr)^n qr_1$ would be a common prefix of $u = p's$ and $v = p(qr)^{n+1} qs$, contradicting the maximality of p'. Finally, $r^{-1} \perp p^{-1}$ follows in the same way from the maximality of s', and we conclude that Γ' cannot be folded any further, i.e. $\Gamma' = \Gamma_X(\langle u, v \rangle)$, which contradicts $\Gamma_X(\langle u, v \rangle) = \Gamma_X(\mathbf{F}_2)$.

Case (ii): m equals a power of t, $m = t^n$ $(n \geq 0)$, so that $u = pt^n s$ and $v = pt^{n+1} s$. This time Γ folds into $\Gamma' =$  . We have $p \perp s^{-1}$ as before; $p^{-1} \perp t$ and $t^{-1} \perp s$ follows from $v = pt^{n+1} s$ being reduced; $s \perp t$ holds, since otherwise $pt^n s_1 = pt^n t_1$ would be a common prefix of u and v, contradicting the maximality of $p' = pt^n$; likewise, $p^{-1} \perp t^{-1}$ by the maximality of s'. Now, if $n > 0$ then $t \perp t^{-1}$ since $v = pt^{n+1} s$ is reduced, and if $n = 0$ then $p^{-1} \perp s$ since $u = ps$ is reduced. In either case, Γ' cannot fold into $\Gamma_X(\mathbf{F}_2)$: For $n = 0$, assuming that Γ' folds at all, $\Gamma_X(\langle u, v \rangle) =$  , where $t = rt'r^{-1}$ with t' CR. Thus, $\Gamma_X(\langle u, v \rangle) \neq \Gamma_X(\mathbf{F}_2)$. For $n > 0$, Γ' folds into  where r is the common suffix of p and s^{-1}, so that $p = p'r$, $s = s'r$. If $p', s' \neq 1$ we are done, and $p' = s' = 1$ is impossible since $p \perp s^{-1}$. If $p' = 1 \neq s'$

then the graph folds into  where s'' is CR and $s' = ys''y^{-1}$, and likewise for $p' \neq 1 = s'$. □

Definition. A word $w \in \mathbf{F}(a,b)$ is *monotone* if for every letter (a or b) all the exponents of this letter in w have the same sign. *monotone*

Proposition 5.3.3. *A CR primitive word in $\mathbf{F}_2$ is monotone.*

Proof. Let u be a CR primitive. By Lemma 5.3.1, possibly applying some cyclic rotation to u, we can complete it to a CR basis $\{u, v\}$. We show that both u and v are monotone, by induction on $|u| + |v|$. The base case $|u| + |v| = 2$ is trivial. Assume that $|u| \leq |v|$. Using Proposition 5.3.2, and perhaps replacing u, v, or both of them by their inverses (which does not affect monotonicity), we can write $v = ut$. Now $\{u, t\}$ is a basis with $|u| + |t| < |u| + |v|$, and we claim that t is CR as well. Otherwise, $t = rt'r^{-1}$ with t' CR and $r \neq 1$, and we have $\Gamma_X(\langle u, t\rangle) =$  , as $u \perp u^{-1}$ since u is CR, $(t')^{-1} \perp t'$ since t' is, and all other relevant pairs since $v = urt'r^{-1}$ is. This, of course, contradicts $\langle u, t\rangle = \mathbf{F}_2$.

Therefore, by the induction hypothesis u and t are monotone. Assume first that $|u| \leq |t|$. Since $v = ut$ is CR, u^{-1} cannot be neither a prefix nor a suffix of t. Thus, by Proposition 5.3.2 u must be a prefix or a suffix of t, and in either case v is monotone. The same argument applies to the case $|u| > |t|$. □

We stress the following observation made in the proof:

Corollary 5.3.4. *Let $\{u, v\}$ be a CR basis of $\mathbf{F}_2$ with u a prefix of v, and write $v = ut$. Then $\{u, t\}$ is again a CR basis.*

This leads to a constructive description of all CR bases of $\mathbf{F}_2$:

Proposition 5.3.5. *Any CR basis of $\mathbf{F}_2$ is obtained as follows: given a pair of positive co-prime integers (p, q), there is a unique sequence of pairs*

$$(p,q) = (p_0,q_0),(p_1,q_1),\ldots,(p_\ell,q_\ell) = (1,1) \tag{5.3.1}$$

which is the result of applying the Euclidean g.c.d. algorithm (i.e. if $p_i < q_i$ then $p_{i+1} = p_i$ and $q_{i+1} = q_i - p_i$, and vice-versa). Let $X_\ell = \{u_\ell, v_\ell\}$ be one of the four bases $\left\{a^{\pm 1}, b^{\pm 1}\right\}$, and define $X_i = \{u_i, v_i\}$ iteratively for $i = \ell - 1 \ldots 0$ by

$$(u_i, v_i) = \begin{cases} (u_{i+1} \quad , \ v_{i+1}u_{i+1}) & p_i < q_i \\ (u_{i+1}v_{i+1} \ , \ v_{i+1} \quad) & q_i < p_i. \end{cases} \tag{5.3.2}$$

Finally, take X_0, conjugate its elements by any common prefix or suffix (thus cyclically rotating both of them), and possibly replace one of them by its inverse.

Proof. This construction certainly gives a CR basis of $\mathbf{F}_2$ (CR follows from monotonicity), and it remains to show that every CR basis is thus obtained. This is done by reversing the process, as follows. Let $\{x_0, y_0\}$ be a CR basis. Discarding the trivial bases $\left\{a^{\pm 1}, b^{\pm 1}\right\}$, we can assume (by inversion if necessary) that one of x_0 or y_0 is a prefix/suffix of the other (by Proposition 5.3.2).

Lemma. *x_0, y_0 can be rotated by a a common prefix/suffix, so that a sequence of CR bases $\{x_1, y_1\}, \ldots, \{x_\ell, y_\ell\}$ with the following properties is obtained:*

(1) For every $0 \leq i \leq \ell - 1$, the shorter of x_i, y_i is a suffix of the longer one.

(2) Each basis is obtained from the previous one by

$$(x_{i+1}, y_{i+1}) = \begin{cases} (x_i, t) & |x_i| < |y_i| \text{ and } y_i = tx_i \\ (t, y_i) & |y_i| < |x_i| \text{ and } x_i = ty_i. \end{cases} \tag{5.3.3}$$

(3) $\{x_\ell, y_\ell\}$ *is one of the four bases* $\left\{a^{\pm 1}, b^{\pm 1}\right\}$.

Proof of the Lemma. If we do not perform the rotation of x_0, y_0 by a common prefix/suffix, the same holds, but with the exception that at each stage the shorter among x_i, y_i may be a prefix of the longer one, and not a suffix: this follows from by Proposition 5.3.2, and Corollary 5.3.4, which ensures that all of these bases are CR. Assume that the process first fails at step m, i.e. the shorter among x_m, y_m is not a suffix of the longer one, and assume for simplicity that x_m is shorter, so that $y_m = x_m t$.

Since x_0, y_0 are products of positive powers of x_m and $y_m = x_m t$, it follows that x_m is a common prefix of both of them. Therefore, $\{x'_0, y'_0\} = \left\{x_m^{-1} x_0 x_m, x_m^{-1} y_0 x_m\right\}$ is a cyclic rotation of $\{x_0, y_0\}$. Let $\{x'_i, y'_i\}_{i=0\ldots m}$ be the bases obtained by (5.3.3) from $\{x'_0, y'_0\}$. Since the expression of x'_i, y'_i as words in x'_m, y'_m is the same as that of x_i, y_i as words in x_m, y_m, we still have that at every step until m the shorter of x'_i, y'_i is a suffix of the longer one. In fact, $x'_i = x_m^{-1} x_i x_m$ and likewise for y'_i, for all $i \leq m$. Now, assertion *(1)* holds for step m as well, as $x'_m = x_m$ and $y'_m = x_m^{-1} (x_m t) x_m = t x_m = t x'_m$.

We continue in this manner: at the next step at which assertion *(1)* fails we replace $\{x'_0, y'_0\}$ by $\{x''_0, y''_0\}$ which resolves that step, and note that x''_0, y''_0 is still a cyclic rotation of the original x_0, y_0 by a common prefix/suffix, and that no new failures of *(1)* were introduced by this change for the previous steps. Repeating this for every failure of *(1)* guarantees that it hold throughout the process. □

We continue the proof of the proposition, assuming that x_0, y_0 were inverted and rotated according to the Lemma, and $\{x_1, y_1\}, \ldots, \{x_\ell, y_\ell\}$ are the bases obtained by (5.3.3). The sequence of integer pairs $(|x_0|, |y_0|), \ldots, (|x_\ell|, |y_\ell|) = (1, 1)$ is then the sequence obtained by the Euclidean algorithm for $(|x_0|, |y_0|)$ (as in (5.3.1)), and in particular this shows that $|x_0|, |y_0|$ are co-prime. Thus, if one takes $(p, q) = (|x_0|, |y_0|)$ and $(u_\ell, v_\ell) = (x_\ell, y_\ell)$, and follows (5.3.2) as explained in the statement of the proposition, the process in (5.3.3) is reversed, and one obtains $(u_0, v_0) = (x_0, y_0)$. □

Corollary 5.3.6. *For a CR basis* $\{u, v\}$*, regard* u *and* v *as cyclic words, and assume (by inverting* u *if necessary) that one of them is a subword of the other. Then one of* a, b *always appears (in both* u *and* v*) with exponent* ε *for some fixed* $\varepsilon \in \{1, -1\}$*, and the other letter always appears with exponent* m *or* $m + 1$ *for some* $m \in \mathbb{Z}$.

Proof. Let $X_i = \{u_i, v_i\}$ $(0 \leq i \leq \ell)$ be the bases constructed in Proposition 5.3.5 to give $X_0 = \{u, v\}$. Assume, for simplicity, that $X_\ell = \{u_\ell, v_\ell\} = \{a, b\}$, and that in the first step the first option in (5.3.2) holds, so that $X_{\ell-1} = \{a, ba\}$. Let r be the number of times the first option in (5.3.2) holds before it fails, i.e. $X_{\ell-2} = \left\{a, ba^2\right\}, \ldots, X_{\ell-r} = \{a, ba^r\}$ (possibly $r = 1$). If $r = \ell$ then the statement holds. Otherwise, $X_{\ell-r-1} = \{aba^r, ba^r\}$, and now every cyclic word which is a product of the elements of $X_{\ell-r-1}$ (with positive exponents only) clearly satisfies the statement of the corollary with $\varepsilon = 1$ and $m = r$. Since u and v are such words, we are done. □

5.4 The counterexample

Let $X = \{a, b\}$ and $\mathbf{F}_2 = \mathbf{F}(X)$. In this section we prove the following:

Proposition 5.4.1. *Let* $H = \left\langle a^2 b^2 \right\rangle$*, and* $J = \left\langle a^2 b^2, ab \right\rangle$*. Then* $H \leq^{\rightarrow}_{Y} J$ *for every basis* Y *of* $\mathbf{F}_2$*, but* J *is not an algebraic extension of* H.

Proof. First, as H is a free factor of J (since $J = H * \langle ab \rangle$), it is clear that J is not an algebraic extension of H, and it is left to show that H covers J with respect to every basis $Y = \{u, v\}$. For any automorphism φ of $\mathbf{F}_2$, $H \leq^{\rightarrow}_{X} J$ iff $\varphi(H)$ covers $\varphi(J)$ w.r.t. the basis $\varphi(X) = \{\varphi(a), \varphi(b)\}$. As $\varphi(X)$ achieves all bases of $\mathbf{F}_2$, what we seek to show is equivalent to the assertion that $\left\langle u^2 v^2 \right\rangle \leq^{\rightarrow}_{X} \left\langle u^2 v^2, uv \right\rangle$ for every basis $\{u, v\}$.

By Lemma 5.2.3, showing that $\langle u^2v^2\rangle$ X-covers $\langle u^2v^2, uv\rangle$ is equivalent to verifying that uv appears in $\Gamma_X\left(\langle u^2v^2\rangle\right)$. For the case where u and v are CR this is shown in Lemma 5.4.3, and the case where only one of them is CR is handled in Lemma 5.4.4. For the general case, let $Y = \{\overline{u}, \overline{v}\}$ be the base at hand, and write $\overline{u} = xux^{-1}$ and $\overline{v} = yvy^{-1}$ with u, v CR. By Lemma 5.3.1, x is a prefix of y, or vice-versa. Thus, if w is the shorter among x, y then $w^{-1}Yw$ is a basis with one CR element, which was already handled. Inferring from this the result for the original Y is done in Lemma 5.4.2. For this we need an additional technical assumption on $\Gamma_X\left(\langle w^{-1}\overline{u}^2\overline{v}^2w\rangle\right)$, which is seen to hold in Lemmas 5.4.3 and 5.4.4. □

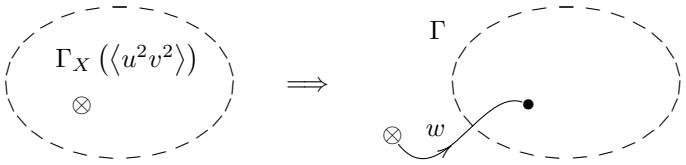
Lemma 5.4.2. *Let $\{\overline{u}, \overline{v}\}$ be a basis of $\mathbf{F}_2$ such that $\overline{u}$ and $\overline{v}$ share a common prefix w and a common suffix w^{-1}, and write $\overline{u} = wuw^{-1}$ and $\overline{v} = wvw^{-1}$. If*

(1) uv appears in $\Gamma_X\left(\langle u^2v^2\rangle\right)$, and

(2) either u_1 or $\left(v^{-1}\right)_1$ emanates from the basepoint of $\Gamma_X\left(\langle u^2v^2\rangle\right)$,

then $\overline{uv}$ appears in $\Gamma_X\left(\langle \overline{u}^2\overline{v}^2\rangle\right)$.

Proof. Observe the graph Γ, which is obtained by attaching a path labeled by w to the basepoint of $\Gamma_X\left(\langle u^2v^2\rangle\right)$ and moving the basepoint to the origin of the w-path:



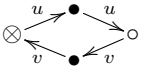
The graph Γ folds into $\Gamma_X\left(\pi_1^X(\Gamma)\right) = \Gamma_X\left(\langle \overline{u}^2\overline{v}^2\rangle\right)$, since if satisfies the conditions of Lemma 5.2.1: the only vertex that needs checking is the gluing place, and there the conditions hold by assumption *(2)* and the fact that $w^{-1} \perp u, v^{-1}$ (as $\overline{u}, \overline{v}$ are reduced). Finally, since uv appears in $\Gamma_X\left(\langle u^2v^2\rangle\right)$, $\overline{uv} = wuvw^{-1}$ appears in Γ, and thus also in its folding $\Gamma_X\left(\langle \overline{u}^2\overline{v}^2\rangle\right)$. □

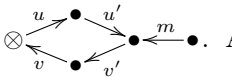
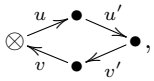
Lemma 5.4.3. *If $\{u, v\}$ is a CR basis of $\mathbf{F}_2$ then*

(1) uv appears in $\Gamma_X\left(\langle u^2v^2\rangle\right)$, and

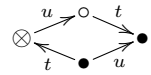
(2) u_1 or $\left(v^{-1}\right)_1$ emanates from the basepoint of $\Gamma_X\left(\langle u^2v^2\rangle\right)$.

Proof. If $|u| + |v| = 2$ then the claims hold, so assume that $|u| + |v| \geq 3$. Since $\Gamma_X\left(\langle u^2v^2\rangle\right) = \Gamma_X\left(\langle v^{-2}u^{-2}\rangle\right)$ and $\Gamma_X\left(\langle u^2v^2, uv\rangle\right) = \Gamma_X\left(\langle v^{-2}u^{-2}, v^{-1}u^{-1}\rangle\right)$, one can look at $\left\{v^{-1}, u^{-1}\right\}$ instead of $\{u, v\}$ (this also does not affect assertion *(2)*), and thus it is enough to handle the cases where $|u| < |v|$.

Observe the graph $\Gamma = \otimes$  , which obviously satisfies $\pi_1^X(\Gamma) = \langle u^2v^2\rangle$. At the black vertices there can be no folding, as $u^{-1} \perp u$ follows from u being CR, and likewise for v. In what follows we will continue to mark by $\bullet$ vertices at which we already know that no folding can occur, and by $\circ$ vertices at which we do not know this.

Assume first that u^{-1} is not a prefix of v, and m is the maximal common prefix of u^{-1} and v. Writing $u = u'm^{-1}$ and $v = mv'$, Γ folds into $\otimes$  . After trimming one obtains $\Gamma' = \otimes$  , which satisfies the conditions of Lemma 5.2.1: $u^{-1} \perp u'$ since u is CR and u' is a prefix of u, and likewise for v and v'^{-1}; $u'^{-1} \perp v'$ by the maximality of m. Therefore, Γ' folds

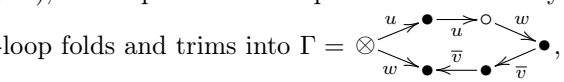
into $\Gamma_X\left(\left\langle u^2v^2\right\rangle\right)$. Since uv appears in Γ', and both u_1 and $\left(v^{-1}\right)_1$ leave its basepoint, the same holds after any foldings, and in particular in $\Gamma_X\left(\left\langle u^2v^2\right\rangle\right)$.

Assume now that u^{-1} is a prefix of v and write $v = u^{-1}t$. Now Γ folds and trims into  , as $u \perp t$ and $t^{-1} \perp u^{-1}$ follow from $v = u^{-1}t$ being CR. Let m be the maximal common prefix of u^{-1} and tu^{-1}, and write $u = \overline{u}m^{-1}$, $tu^{-1} = mq$ (note that $\left|tu^{-1}\right| > \left|u^{-1}\right| \geq |m|$). The last graph then folds and trims into $\Gamma' =$  (with $\overline{u}$ possibly empty, in which case $\Gamma' =$ ). This graph satisfies the conditions of Lemma 5.2.1: $q^{-1} \perp t$ since $tu^{-1} = mq$ is CR, being a cyclic rotation of v, and if $\overline{u}$ is not empty then $\overline{u}^{-1} \perp q$ by maximality of m. Since $uv = t$ appears in Γ' and $\left(t^{-1}\right)_1 = \left(v^{-1}\right)_1$ leaves its basepoint, the same holds for $\Gamma_X\left(\left\langle u^2v^2\right\rangle\right)$, which is obtained from it by foldings. □

Lemma 5.4.4. *If $\{u, v\}$ is a basis of $\mathbf{F}_2$ with u or v CR then*

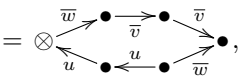
(1) uv appears in $\Gamma_X\left(\left\langle u^2v^2\right\rangle\right)$, and

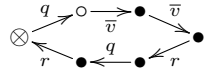
(2) u_1 or $\left(v^{-1}\right)_1$ emanates from the basepoint of $\Gamma_X\left(\left\langle u^2v^2\right\rangle\right)$.

Proof. If both u and v are CR then we are done by Lemma 5.4.3. Again, by replacing u and v with v^{-1} and u^{-1} respectively we can assume that u is CR and v is not (here u is not necessarily shorter). Writing $v = w\overline{v}w^{-1}$ with $\overline{v}$ CR (and $w \neq 1$), w is a prefix of some power of u or u^{-1} by Lemma 5.3.1. The graph formed by a single u^2v^2-loop folds and trims into $\Gamma =$  , where at all the black vertices there can be no folding since u and $\overline{v}$ are CR and v is reduced.

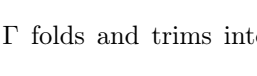
If w is a prefix of a (positive) power of u then at $\circ$ there is no folding as well since u is CR, so that $\Gamma_X\left(\left\langle u^2v^2\right\rangle\right)$ is obtained from Γ by foldings. Therefore to establish that uv appears in $\Gamma_X\left(\left\langle u^2v^2\right\rangle\right)$ it is enough to show that it appears in Γ. This is not obvious in first sight, but it is true: w is a prefix of a positive power of u, so that uw is a prefix of u^2w, hence $uv = uw\overline{v}w^{-1}$ indeed appears in Γ. Finally, both u_1 and $\left(v^{-1}\right)_1 = w_1$ leave the basepoint of Γ and thus also of $\Gamma_X\left(\left\langle u^2v^2\right\rangle\right)$.

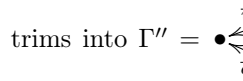
We assume now that w is a prefix of some power of u^{-1}, and observe six cases.

Case (i): $2|u| \leq |w|$, so that $w = u^{-2}\overline{w}$ with $\overline{w}$ possibly empty. In this case Γ folds and trims into $\Gamma' =$  , which satisfies Lemma 5.2.1 (even if $\overline{w} = 1$). Now $uv = u^{-1}\overline{wvw}^{-1}u^2$ appears in Γ': $\overline{vw}^{-1}u^2$ is a suffix of Γ' oriented clockwise, and $u^{-1}\overline{w}$ is a prefix of Γ' oriented counterclockwise since $\overline{w}$ is a prefix of $u^{-1}\overline{w}$. In addition, $\left(v^{-1}\right)_1 = w_1 = \left(u^{-1}\right)_1$ leaves $\otimes$.

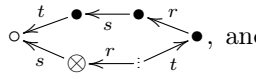
Case (ii): $|u| < |w| < 2|u|$, so that we can write $u = qr$ and $w = r^{-1}q^{-1}r^{-1}$ with $q, r \neq 1$. Now Γ folds and trims into $\Gamma' =$  , and no folding can occur at the black vertices. If at $\circ$ there is no folding as well, then $uv = r^{-1}\overline{v}rqr$ appears in Γ' and $\left(v^{-1}\right)_1 = w_1 = \left(r^{-1}\right)_1$ leaves $\otimes$, yielding the same for $\Gamma_X\left(\left\langle u^2v^2\right\rangle\right)$.
Assume now that there is folding at $\circ$, so that q^{-1} and $\overline{v}^2$ have a common prefix. If this prefix is shorter than $|\overline{v}|$ than the $\overline{v}rqr$ part in Γ' survives in $\Gamma_X\left(\left\langle u^2v^2\right\rangle\right)$, and thus $uv = r^{-1}\overline{v}rqr$ still appears in it. Otherwise, $\overline{v}$ is a prefix of q^{-1}, so that $r^{-1}\overline{v}$ is a prefix of Γ' oriented CCW, and rqr is a suffix of Γ' oriented CW, so that uv appears already in the lower half of Γ'. Finally, this half survives in $\Gamma_X\left(\left\langle u^2v^2\right\rangle\right)$ since q^{-1} cannot overlap with r, since $u = qr$ is monotone by Proposition 5.3.3. In both cases $\left(v^{-1}\right)_1 = \left(r^{-1}\right)_1$ still leaves $\otimes$.

Case (iii): $w = u^{-1}$. The reasoning here is as in the previous case with $r = 1$.

In cases ***(iv) − (vi)*** w is a proper prefix of u^{-1}, and we write $u = qr$ and $w = r^{-1}$ (with $r, q \neq 1$). Here Γ folds and trims into $\Gamma' =$  , with folding possible only at $\circ$, and the folding cannot reach past $\vdots$ due to the monotonicity of $u = qr$. This already shows that $\left(v^{-1}\right)_1 = \left(r^{-1}\right)_1$ must leave the basepoint of the final Stallings graph $\Gamma_X\left(\left\langle u^2v^2\right\rangle\right)$. It is left to show that the folding and trimming at $\circ$ does not prevent $uv = q\overline{v}r$ from appearing in $\Gamma_X\left(\left\langle u^2v^2\right\rangle\right)$. If the lower half of Γ' survives the folding and trimming then uv certainly appears in it. We assume therefore that there is folding at $\circ$, and that it encompasses either all of $\overline{v}$ to the right of $\circ$ or all of rq to the left of it (i.e. it reaches the lower half of Γ').

Case (iv): $|\overline{v}| \leq |q|$. By our assumption, $\overline{v}$ is a prefix of q^{-1}, so $q = \overline{q}\overline{v}^{-1}$ ($\overline{q}$ may be empty). Γ' then folds and trims into $\Gamma'' =$  . Now $uv = q\overline{v}r = \overline{q}r$ and since the folding from $\circ$ downward must stop at $\vdots$ (or earlier), the $\bullet \xleftarrow{\overline{q}} \otimes \xleftarrow{r} \vdots$ part survives in $\Gamma_X\left(\left\langle u^2v^2\right\rangle\right)$ (since $|\overline{v}| < \left|\overline{q}^{-1}r^{-1}\overline{v}\right|$) and we are done.

Case (v): $|q| < |\overline{v}| \leq |rq|$. Now we can assume that $\overline{v}$ is a prefix of $q^{-1}r^{-1}$, so that $r = st$ and $\overline{v} = q^{-1}t^{-1}$ (possibly with $s = 1$). In this case $uv = q\overline{v}r = t^{-1}r$ already appears in the $\otimes \xleftarrow{r} \vdots$ part of Γ', which always survives due to monotonicity.

Case (vi): $|rq| < |\overline{v}|$. Now we can assume that $q^{-1}r^{-1}$ is a prefix of $\overline{v}$. Therefore, $uv = q\overline{v}r$ is a suffix of $\overline{v}r$, and thus appears in the $\otimes \xleftarrow{r} \vdots \xleftarrow{\overline{v}} \bullet$ part in Γ'. If $\overline{v}$ is not a prefix of $q^{-1}r^{-1}q^{-1}$ then the folding from $\circ$ downward stops before reaching this part, and we are done. We thus add the assumption that $\overline{v}$ is a prefix of $q^{-1}r^{-1}q^{-1}$. Since $|rq| < |\overline{v}|$ we can write $q = st$ so that $\overline{v} = q^{-1}r^{-1}t^{-1} = t^{-1}s^{-1}r^{-1}t^{-1}$. Now Γ' folds and trims into $\Gamma'' =$  , and $uv = r^{-1}t^{-1}r$ appears in $\otimes \xleftarrow{r} \vdots \xrightarrow{t} \bullet$, which survives any further folding since $|s| < \left|t^{-1}s^{-1}r^{-1}\right|$. □

5.5 Epilogue

While the original conjecture [MVW07, §5(1)] fails, it is plausible that some modification of it holds. One possible option is the following:

Conjecture 5.5.1. *Let $H \leq J$ be subgroups of the free group $\mathbf{F}$. Then $H \leq_{alg} J$ iff $H \leq_{\vec{x}} J$ for every free extension $\mathbf{F}'$ of $\mathbf{F}$, and every basis X of $\mathbf{F}'$.*

Since the relation $H \leq_{alg} J$ does not depend on the ambient group, one direction holds as before. But in contrast with the original conjecture, the example in Section 5.4 is no longer a counterexample: let $\mathbf{F} = \mathbf{F}(a, b)$, $H = \left\langle a^2b^2\right\rangle$ and $J = \left\langle ab, a^2b^2\right\rangle$. For $\mathbf{F}' = \mathbf{F}(a, b, c)$ and $X = \left\{a, cb^{-1}, cbc^{-1}\right\}$, H does not X-cover J: denote $x = a$, $y = cb^{-1}$ and $z = cbc^{-1}$. Then written in this basis, $H = \left\langle x^2y^{-1}z^2y\right\rangle$ and $J = \left\langle x^2y^{-1}z^2y, xy^{-1}zy\right\rangle$. By Lemma 5.2.3, $H \leq_{\vec{x}} J$ iff $xy^{-1}zy$ appears in $\Gamma_{\{x,y,z\}}\left(\left\langle x^2y^{-1}z^2y\right\rangle\right)$, which is not the case.

Another plausible option is that the original conjecture from [MVW07, §5(1)] holds for free groups of rank three or more, as it is clear that the counterexample exploits many idiosyncrasies of $\mathbf{F}_2$. If this is true, then Conjecture 5.5.1 follows as well.

References

[CMZ81] M. Cohen, W. Metzler, and A. Zimmermann. What does a basis of $F(a, b)$ look like? *Mathematische Annalen*, 257(4):435–445, 1981.

[Coh72] H. Cohn. Markoff forms and primitive words. *Mathematische Annalen*, 196(1):8–22, 1972.

[KM02] I. Kapovich and A. Myasnikov. Stallings foldings and subgroups of free groups. *Journal of Algebra*, 248(2):608–668, 2002.

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

[Nie17] J. Nielsen. Die isomorphismen der allgemeinen, unendlichen gruppe mit zwei erzeugenden. *Mathematische Annalen*, 78(1):385–397, 1917.

[OZ81] R. P. Osborne and H. Zieschang. Primitives in the free group on two generators. *Inventiones Mathematicae*, 63(1):17–24, 1981.

[PP15] D. Puder and O. Parzanchevski. Measure preserving words are primitive. *Journal of the American Mathematical Society*, 28(1):63–97, 2015.

[Pud14] D. Puder. Primitive words, free factors and measure preservation. *Israel Journal of Mathematics*, 201(1):25–73, 2014.

[Sta83] J.R. Stallings. Topology of finite graphs. *Inventiones mathematicae*, 71(3):551–565, 1983.

[Tak51] M. Takahasi. Note on chain conditions in free groups. *Osaka Math. J*, 3(2):221–225, 1951.

Epilogue

The results elaborated in the five manuscripts included in this thesis constitute important progress in their lines of research. In most chapters I also presented some intriguing open questions sparked by the different results. In this short epilogue, let me mention two questions which I consider to be the most interesting ones.

- **Separating $\mathrm{Aut}(\mathbf{F}_k)$-orbits:** A very challenging open question is the following. The (now resolved) question about primitive words and measure preservation is actually a special case of a more general problem. The primitives constitute a single $\mathrm{Aut}(\mathbf{F}_k)$-orbit in $\mathbf{F}_k$. In the same manner that they induce the same measure on every finite (or compact) group, it is an easy observation that any two words belonging to the same $\mathrm{Aut}(\mathbf{F}_k)$-orbit induce the same measure on every finite (compact) group. But does the converse hold? Namely, if w_1 and w_2 belong to different orbits, is there necessarily some (finite? compact?) group on which they induce different measures? Not only is this question very natural, it also has some interesting implications. For example, a positive answer would provide us with an algorithm to solve the automorphism decidability problem for free groups. Namely, to determine whether two given words belong to the same $\mathrm{Aut}(F_k)$-orbit. A famous algorithm of Whitehead [Whi36b] solves this decidability problem, but much is left to be desired in terms of complexity and more generally in terms of our understanding of $\mathrm{Aut}(\mathbf{F}_k)$-orbits in $\mathbf{F}_k$. (See [PP15, Sec. 8], or Section 2.8 here, for details).

- **Distribution of second eigenvalue:** I hope that the methods I developed in [Pud15a] (Chapter 3) can be improved to produce a new, simplified proof of Alon's conjecture (Friedman's theorem) and to yield a complete proof of the generalized Alon-Friedman conjecture regarding random graph coverings. However, as aforementioned, even after Alon's conjecture is established, many open questions remain concerning λ, the second (absolute value of an) eigenvalue of a random d-regular graph on n vertices. In fact, very little is known about the distribution of λ. A major open question is the following: what is the probability that a random d-regular graph is Ramanujan, i.e. that $\lambda \leq 2\sqrt{d-1}$? Numeric simulations were conducted by several researchers (see, for instance, [MNS08] and [HLW06]) but different pieces of evidence suggest contradicting answers. We hope our new approach may eventually contribute to answering these open questions. In fact, even the following, much weaker question is not known: are there infinitely many Ramanujan d-regular graphs for every $d \geq 3$? The 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 $d \geq 3$. Still, the original problem remains open.

References

[HLW06] S. Hoory, N. Linial, and A. Wigderson. Expander graphs and their applications. *Bulletin of the American Mathematical Society*, 43(4):439–562, 2006.

[LPS88] A. Lubotzky, R. Phillips, and P. Sarnak. Ramanujan graphs. *Combinatorica*, 8(3):261–277, 1988.

[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(1):51–60, 1988.

[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(2):231–244, 2008.

[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(1):44–62, 1994.

[MSS13] A. Marcus, D.A. Spielman, and N. Srivastava. Interlacing families i: Bipartite ramanujan graphs of all degrees. *arXiv preprint arXiv:1304.4132*, 2013.

[PP15] D. Puder and O. Parzanchevski. Measure preserving words are primitive. *Journal of the American Mathematical Society*, 28(1):63–97, 2015.

[Pud15a] D. Puder. Expansion of random graphs: New proofs, new results. *Inventiones Mathematicae*, 2015. to appear. arXiv:1212.5216.

[Whi36b] J.H.C. Whitehead. On equivalent sets of elements in a free group. *Ann. of Math.*, 37:768–800, 1936.