The primitivity index function for a free group, and untangling closed curves on hyperbolic surfaces. *With an appendix by Khalid Bou-Rabee.*

By NEHA GUPTA

Department of Mathematics, FAS, Harvard University, One Oxford Street, Cambridge, MA 02138
e-mail neha@math.harvard.edu

and ILYA KAPOVICH

Department of Mathematics, University of Illinois at Urbana-Champaign,
1409 West Green Street, Urbana, IL 61801
email kapovich@math.uiuc.edu

(*Received*)

Abstract

Motivated by the results of Scott and Patel about "untangling" closed geodesics in finite covers of hyperbolic surfaces, we introduce and study primitivity, simplicity and non-filling index functions for finitely generated free groups. We obtain lower bounds for these functions and relate these free group results back to the setting of hyperbolic surfaces. An appendix by Khalid Bou-Rabee connects the primitivity index function $f_{prim}(n, F_N)$ to the residual finiteness growth function for F_N.

The second author was partially supported by the NSF grants DMS-1405146 and DMS-1710868. Both authors acknowledge support from U.S. National Science Foundation grants DMS 1107452, 1107263, 1107367 "GEAR Network".

1. *Introduction*

Let Σ be a compact connected surface with a hyperbolic metric ρ and with (possibly empty) geodesic boundary. In [52, 53] Scott proved that $\pi_1(\Sigma)$ is *subgroup separable* or *LERF*, meaning that for every finitely generated subgroup $K \leq \pi_1(\Sigma)$ and every $g \in \pi_1(\Sigma)$ such that $g \notin K$ there exists a subgroup $H \leq \pi_1(\Sigma)$ of finite index in $\pi_1(\Sigma)$ such that $K \leq H$ but $g \notin H$. (Scott's result dealt with the case of a closed surface S since in the case $\partial S \neq \emptyset$, the group $\pi_1(S)$ is free and hence known to be subgroup separable by a much older result of Hall [29]). In the same work [52] Scott showed that if γ is a closed geodesic on (Σ, ρ) then there exists a finite cover $\widehat{\Sigma} \to \Sigma$ such that γ lifts to a simple closed geodesic in $\widehat{\Sigma}$, where $\widehat{\Sigma}$ is given the hyperbolic structure obtained by the pull-back of ρ. As customary in the context of hyperbolic surfaces, the term "closed geodesic" here assumes that the curve in question is not a proper power in the fundamental group

2010 *Mathematics Classification.* Primary 20F65, Secondary 37D, 53C, 57M.

of the surface. Recently Patel [44] obtained quantitative versions of Scott's subgroup separability result and of his result about lifting a closed geodesic to a simple one in a finite cover. Thus she proved that for every Σ as above there exists a hyperbolic metric ρ_0 on Σ such that every closed geodesic of length L on (Σ, ρ_0) lifts to a simple closed geodesic in some finite cover of Σ of degree $\leq 16.2L$. Since the length functions on $\pi_1(\Sigma)$ coming from any two hyperbolic structures on Σ are bi-Lipschitz equivalent, it follows that for any hyperbolic structure ρ on Σ there is some constant $c > 0$ such that every closed geodesic of length L on (Σ, ρ) lifts to a simple closed geodesic in some finite cover of Σ of degree $\leq cL$. Motivated by these results, if ρ is a hyperbolic structure on Σ, for every closed geodesic γ on (Σ, ρ) we denote by $\deg_{\Sigma,\rho}(\gamma)$ the smallest degree of a finite cover of Σ such that γ lifts to a simple closed geodesic in that cover. For $L \geq sys(\rho)$ (where $sys(\rho)$ is the shortest length of a closed geodesic on (Σ, ρ)) put $f_{\Sigma,\rho}(L)$ to be the maximum of $\deg_{\Sigma,\rho}(\gamma)$ taken over all closed geodesics γ on (Σ, ρ) of length $\leq L$. Patel's result mentioned above implies that for every hyperbolic structure ρ on Σ there is $c > 0$ such that $f_{\Sigma,\rho}(L) \leq cL$ for all $L \geq sys(\rho)$.

A simple closed geodesic on a hyperbolic surface is a particular example of a non-filling curve. Thus for a hyperbolic surface (Σ, ρ) as above and for a closed geodesic γ on Σ we can also define $\deg^{fill}_{\Sigma,\rho}(\gamma)$ to be the smallest degree of a finite cover of Σ such that γ lifts to a non-filling closed geodesic in that cover. Then put $f^{fill}_{\Sigma,\rho}(L)$ to be the maximum of $\deg^{fill}_{\Sigma,\rho}(\gamma)$ taken over all closed geodesics γ on (Σ, ρ) of length $\leq L$. Thus, in view of Patel's result, we have $f^{fill}_{\Sigma,\rho}(L) \leq f_{\Sigma,\rho}(L) \leq cL$ for all $L \geq sys(\rho)$. However, up to now, nothing has been known about lower bounds for $f_{\Sigma,\rho}(L)$ or $f^{fill}_{\Sigma,\rho}(L)$. (Note that the first place where the question about quantitative properties of $f_{\Sigma,\rho}(L)$ was raised, although somewhat indirectly, appears to have been the paper of Rivin [48]).

In general, obtaining lower bounds for quantitative results related to residual finiteness is quite difficult, and is usually harder than obtaining upper bounds. Recently there has been a significant amount of research regarding quantitative aspects of residual finiteness; see, for example [8, 5, 11, 12, 13, 14, 10, 17, 27, 37, 38, 44, 48, 19, 9, 16, 18]. We will discuss some of these results in more detail below.

Let $N \geq 2$ be an integer and let F_N be the free group of rank N. If A is a free basis of F_N, for an element $g \in F_N$ we denote by $|g|_A$ the freely reduced length of g over A and we denote by $||g||_A$ the cyclically reduced length of g over A. A classic result of Marshall Hall [29], mentioned above, (see also [34] for a modern proof using Stallings subgroup graphs) proves that finitely generated free groups are subgroup separable. More precisely, Hall proved that if $K \leq F_N$ is a finitely generated subgroup and $g \in F_N - K$ then there exists a subgroup $H \leq F_N$ of finite index such that $g \notin H$, $K \leq H$, and, moreover, K is a free factor of H. It is not hard to adapt the proof of this result to show that for every $g \in F_N, g \neq 1$ there exists a subgroup $H \leq F_N$ of finite index such that $g \in H$ and that g is a *primitive* element of H, that is, that g belongs to some free basis of H. In fact, a simple argument using Stallings subgroup graphs (see Proposition 3·5 below) shows that if A is a free basis of F_N and w is a nontrivial cyclically reduced word in $F(A)$ of length n then there exists a subgroup $H \leq F_N$ with $[F_N : H] = n$ such that $w \in H$ is a primitive element of H. For a nontrivial element $g \in F_N$ we define the *primitivity index* $d_{prim}(g) = d_{prim}(g; F_N)$ as the minimum of $[F_N : H]$ where H varies over all subgroups of finite index in F_N containing g as a primitive element. Given a free basis A of F_N, for $n \geq 1$ we then define $f_{prim}(n) = f_{prim}(n; F_N)$ as the maximum of $d_{prim}(g)$ where g

varies over all nontrivial freely reduced words of length $\leq n$ in $F_N = F(A)$ which are not proper powers in F_N. It is not hard to see that $f_{prim}(n)$ does not depend on the choice of a free basis A of F_N; we call $f_{prim}(n)$ the *primitivity index function* for F_N. Thus $f_{prim}(n)$ is the smallest monotone non-decreasing function such that for every nontrivial root-free $g \in F_N$ we have $d_{prim}(g) \leq f_{prim}(|g|_A)$.

A nontrivial element $g \in F_N$ is called *simple* in F_N if g belongs to some proper free factor of F_N. A nontrivial element $g \in F_N$ is called *filling* in F_N if g does not belong to a vertex group of a nontrivial splitting of F_N over the trivial or maximal infinite cyclic subgroup. See Section 2·3 for more precise definitions and a discussion of these notions. Note that for $1 \neq g \in F_N$, if g is primitive then g is simple, and if g is simple then g is non-filling. For a nontrivial element $g \in F_N$ let $d_{simp}(g) = d_{simp}(g; F_N)$ be the smallest index $[F_N : H]$ where H varies over all subgroups of finite index in F_N such that $g \in H$ and that g is simple in H. Finally, let $d_{fill}(g) = d_{fill}(g; F_N)$ be the smallest index $[F_N : H]$ where H varies over all subgroups of finite index in F_N such that $g \in H$ and that g is non-filling in H. Thus by definition, we have $d_{fill}(g) \leq d_{simp}(g) \leq d_{prim}(g)$. For $n \geq 1$ we then define the *simplicity index function* $f_{simp}(n) = f_{simp}(n; F_N)$ as the maximum of $d_{simp}(g)$ where g varies over all nontrivial freely reduced words of length $\leq n$ in $F_N = F(A)$ that are not proper powers in F_N. Also, for $n \geq 1$ we then define the *non-filling index function* $f_{fill}(n) = f_{fill}(n; F_N)$ as the maximum of $d_{fill}(g)$ where g varies over all nontrivial freely reduced words of length $\leq n$ in $F_N = F(A)$ that are not proper powers in F_N.

In view of Proposition 3·5 mentioned above, for every nontrivial $g \in F_N$ we have $d_{simp}(g) \leq d_{prim}(g) \leq ||g||_A \leq |g|_A$, and hence $f_{fill}(n) \leq f_{simp}(n) \leq f_{prim}(n) \leq n$ (see Lemma 3·6 for details).

In general, we are interested in the following types of questions:

- Understand the actual asymptotics of the "worst-case" index functions $f_{fill}(n)$, $f_{simp}(n)$, $f_{prim}(n)$ for free groups and of their geometric counterparts $f_{\Sigma,\rho}(L)$ or $f^{fill}_{\Sigma,\rho}(L)$.
- Find specific sequences of elements in free groups or curves on surfaces realizing this "worst-case" behavior or at least exhibiting reasonably fast growth of the corresponding index and degree functions.
- Understand the asymptotics of the indexes $d_{prim}(g_n), d_{simp}(g_n), d_{fill}(g_n)$ and of $\deg_{\Sigma,\rho}(\gamma_n)$, $\deg^{fill}_{\Sigma,\rho}(\gamma_n)$ for various "natural" sequences of group elements $g_n \in F_N$ or closed geodesics γ_n on (Σ, ρ).
- Understand the relationship between the index functions for free groups and the degree functions for surfaces, and relate both to other functions measuring quantitative aspects of residual properties of free and surface groups.

Our first main result provides a lower bound for $f_{fill}(n; F_N)$; see Theorem 6·2 below:

THEOREM 1·1. *Let $N \geq 2$ and let $F_N = F(A)$ where $A = a_1, \ldots, a_N$. Then there exists a constant $c > 0$ and an integer $M \geq 1$ such that for all $n \geq M$ we have*

$$f_{prim}(n) \geq f_{simp}(n) \geq f_{fill}(n) \geq c\frac{\log n}{\log\log n}.$$

For a finitely generated group G equipped with a finite generating set A, the *residual finiteness growth function* $\mathrm{RF}_G(n)$ is defined as the smallest number d such that for every nontrivial element $g \in G$ of word-length $\leq n$ with respect to A there exists a subgroup of index at most d in G that does not contain g.

In an appendix to this paper, for a free group F_N with a free basis A, Khalid Bou-Rabee relates $f_{prim}(n, F_N)$ to the residual finiteness growth function $\mathrm{RF}_{F_N}(n)$. Namely, he shows in Theorem A1 below that for $n \geq 1$ one has $f_{prim}(4n+4, F_N) \geq \mathrm{RF}_{F_N}(n)$. Using a recent result of Kozma and Thom [38] about lower bounds for $\mathrm{RF}_{F_N}(n)$, Bou-Rabee then shows in Corollary A2 below that for all sufficiently large n one has

$$f_{prim}(4n+4) \geq \exp\left(\left(\frac{\log(n)}{C\log\log(n)}\right)^{1/4}\right).$$

Note that this lower bound behaves almost like $n^{1/4}$. Moreover, if we assume Babai's Conjecture on the diameter of Cayley graphs of permutation groups, then for all sufficiently large n we have an almost linear lower bound:

$$f_{prim}(4n+4) \geq n^{\frac{1}{C\log\log(n)}}.$$

Bou-Rabee's homological trick used in Theorem A1 does not work for the index functions $f_{simp}(n)$ and $f_{fill}(n)$. Thus for these functions the lower bound given by Theorem 1·1 remains the best known bound.

We also obtain a bound from below on $d_{simp}(w_n)$ and $d_{fill}(w_n)$ where w_n is a "random" freely reduced word in $F(A)$ of length $n >> 1$.

THEOREM 1·2. *Let $N \geq 2$ and let $F_N = F(A)$ where $A = \{a_1, \dots, a_N\}$.*

Then there exist constants $c(N) > 0$, $D_1(N) > 1$, $1 > D_2(N) > 0$ such that for $n \geq 1$ and for a freely reduced word $w_n \in F(A)$ of length n chosen uniformly at random from the sphere $S(n)$ of radius n in $F(A)$ we have

$$1 - P_{\mu_n}\left(d_{simp}(w_n) \geq c\log^{1/3} n\right) = O\left((D_1)^{-n^{D_2}}\right)$$

and

$$1 - P_{\mu_n}\left(d_{fill}(w_n) \geq c\log^{1/5} n\right) = O\left((D_1)^{-n^{D_2}}\right)$$

so that

$$\lim_{n\to\infty} P_{\mu_n}\left(d_{simp}(w_n) \geq c\log^{1/3} n\right) = 1$$

and

$$\lim_{n\to\infty} P_{\mu_n}\left(d_{fill}(w_n) \geq c\log^{1/5} n\right) = 1$$

Here μ_n is the uniform probability distribution on the n-sphere $S(n) \subseteq F_N = F(A)$. See Convention 7·1 for our use of the big-O notation.

It remains an interesting question to understand the actual behavior of $d_{simp}(w_n)$ and $d_{fill}(w_n)$ on "random" elements $w_n \in F_N$ and, in particular, to see if $d_{simp}(w_n)$ and $d_{fill}(w_n)$ admit sublinear upper bounds.

Finally, in Section 9 we relate the above results for free groups to the original motivating questions about the degree functions for hyperbolic surfaces. Thus, using Theorem 1·2, we obtain (see Theorem 9·6 below):

COROLLARY 1·3. *Let (Σ, ρ) be a compact connected hyperbolic surface with $b \geq 1$ geodesic boundary components. Then there exists $C' > 0$ such that for all sufficiently large L we have*

$$f_{\Sigma,\rho}(L) \geq f^{fill}_{\Sigma,\rho}(L) \geq C'\frac{\log L}{\log\log L}.$$

Similarly, using Theorem 1·1, we obtain (see Theorem 9·3 below):

COROLLARY 1·4. *Let Σ be a compact connected surface with a hyperbolic structure ρ and with (possibly empty) geodesic boundary. Let $\Sigma_1 \subseteq \Sigma$ be a compact connected subsurface with ≥ 3 boundary components, each of which is a geodesic in (Σ, ρ). Let $x \in \Sigma_1$ and let A be a free basis of $\pi_1(\Sigma_1, x)$.*

Let $w_n \in F(A) = \pi_1(\Sigma_1, x)$ be a freely reduced word of length n over $A^{\pm 1}$ generated by a simple non-backtracking random walk on $F(A) = \pi_1(\Sigma_1, x)$. Let γ_n be the closed geodesic on (Σ, ρ) in the free homotopy class of w_n.

Then there exist constants $c > 0, K' \geq 1$ such that

$$\lim_{n\to\infty} Pr(\deg_{\Sigma,\rho}(\gamma_n) \geq c \log^{1/3} n) = 1$$

and such that with probability tending to 1 as $n \to \infty$ we have that $w_n \in \pi_1(\Sigma, x)$ is not a proper power and that $n/K' \leq \ell_\rho(\gamma_n) \leq K'n$.

In the original November 2014 version of this paper we used Corollary 1·4 to obtain, for all sufficiently large L, a lower bound

$$f_{\Sigma,\rho}(L) \geq c \log^{1/3} L,$$

where (Σ, ρ) is a closed hyperbolic surface. At the time this was the only known lower bound for $f_{\Sigma,\rho}(L)$. Motivated by our work, Jonah Gaster [**26**] subsequently obtained a linear lower bound $f_{\Sigma,\rho}(L) \geq cL$ and exhibited a specific sequence of curves γ_n in Σ, living in a pair-of-pants subsurface of Σ, realizing this lower bound. Since these curves are already non-filling in Σ and have $\deg^{fill}_{\Sigma,\rho}(\gamma_n) = 1$, Gaster's proof does not provide any lower bounds for $f^{fill}_{\Sigma,\rho}(L)$. Thus for the moment the lower bound for $f^{fill}_{\Sigma,\rho}(L)$ given by Corollary 1·3 remains the best bound known. In Section 9 we also relate our results to the versions of $f_{\Sigma,\rho}(L)$ and $f^{fill}_{\Sigma,\rho}(L)$ that do not involve a hyperbolic metric and use the geometric intersection number $i([\gamma], [\gamma])$ instead of the hyperbolic length of γ in their definitions.

Also, in Section 4 we prove algorithmic computability of the indexes $d_{prim}(g, F_N)$ $d_{simp}(g, F_N)$, $d_{fill}(g, F_N)$ and of the corresponding index functions $f_{prim}(n), f_{simp}(n), f_{fill}(n)$; see Theorem 4·14 and Theorem 4·18 below.

A recent paper of Puder [**45**] (see also [**46**, **47**] for related work) introduces the notion of a *primitivity rank* $\pi(g)$ for an element $g \in F_N$. Namely, $\pi(g)$ is defined as the smallest rank of a subgroup $H \leq F_N$ such that $g \in H$ but g is not primitive in H. Puder proves in [**45**, Corollary 4.2] that for an element $g \in F_N$ one has either $\pi(g) = \infty$ or $0 \leq \pi(g) \leq N$, and that every integer between 0 and N does occur as a value of $\pi(g)$ for some g. He also defines and studies the primitivity rank $\pi(H)$ for a finitely generated subgroup $H \leq F_N$, where $\pi(H)$ is defined as the minimum rank of J such that $H \leq J \leq F_N$ and that H is not a proper free factor of J. These notions are related to and in some sense dual to our definitions of $d_{prim}(g)$ and $d_{simp}(g)$, but the precise connection of our results with Puder's work remains to be understood. Malestein and Putman [**41**] obtained a number of lower bound results (in terms of k) for the minimal self-intersection number of nontrivial elements of the k-term of the lower central series and the derived series of a surface group. It would be interesting to see if their methods can be used to obtain lower bounds for the function $f_{\Sigma,\rho}$. It would also be interesting to investigate if looking inside

the lower central series and the derived series of F_N may produce new lower bounds for $f_{prim}(n)$ and $f_{simp}(n)$.

We are grateful to Yuliy Baryshnikov for providing us with a proof of Lemma 5·1. We then used the idea of the proof of Lemma 5·1 to obtain Proposition 5·7, which plays a crucial role in the proof of our main results. We are also grateful to Igor Rivin for suggesting to try to apply our free group results to untangling closed curves on hyperbolic surfaces, and to Priyam Patel for suggesting to apply our results to the degree functions based in the self-intersection number rather than the length of a curve. We thank Kasra Rafi for the suggestion to consider $\deg^{fill}_{\Sigma,\rho}$ and $f^{fill}_{\Sigma,\rho}$. We thank Nathan Dunfield and Chris Leininger for many useful conversations. We are grateful to Andreas Thom, Gady Kozma, Doron Puder and Khalid Bou-Rabee for helpful feedback. We are particularly grateful to the referee of the original version of this paper for pointing out that our methods implied a much better lower bound for $f_{prim}(n)$ and $f_{simp}(n)$ than the one we originally had in mind. We are also grateful to the referee of the current version for numerous detailed helpful suggestions.

2. *Preliminaries*

2·1. *Graphs and Edge Paths*

The exposition in this subsection follows that of [**36**].

DEFINITION 2·1. *A graph is a 1-dimensional cell-complex. The 0-cells of* Γ *are called vertices and we denote the set of vertices of* Γ *by* $V\Gamma$. *The open 1-cells of* Γ *are called topological edges of* Γ *and the set of topological edges are denoted by* $E_{top}\Gamma$.

Every topological edge of Γ is homeomorphic to the open interval $(0,1)$ and thus, when viewed as a 1-manifold, admits two possible orientations. An *oriented edge* of Γ is a topological edge with a choice of orientation on it. We denote by $E\Gamma$ the set of all oriented edges of Γ. If $e \in E\Gamma$ is an oriented edge, we denote by $\bar{e}$ the same underlying edge with the opposite orientation. Note that for every $e \in E\Gamma$ we have $\bar{e} \neq e$ and $\bar{\bar{e}} = e$; thus $^{-}\colon E\Gamma \to E\Gamma$ is an involution with no fixed points.

Since Γ is a cell-complex, every oriented edge $e \in E\Gamma$ comes equipped with the orientation-preserving attaching map $j_e : [0,1] \to \Gamma$ such that j_e maps $(0,1)$ homeomorphically to e and such that $j_e(0)$, $j_e(1) \in V\Gamma$ (though not necessarily distinct). For $e \in E\Gamma$ we call $j_e(0)$ the *initial vertex* of e, denoted $o(e)$, and we call $j_e(1)$ the *terminal vertex* of e, denoted $t(e)$. Thus, by definition, $o(\bar{e}) = t(e)$ and $t(\bar{e}) = o(e)$.

For any vertex $x \in V\Gamma$, the *degree of* x in Γ denoted by $deg(x)$ is the cardinality of the set $\{e \in E\Gamma | o(e) = x\}$.

An *orientation* on a graph Γ is a partition $E\Gamma = E_+\Gamma \sqcup E_-\Gamma$ such that for an edge $e \in E\Gamma$ we have $e \in E_+\Gamma$ if and only if $\bar{e} \in E_-\Gamma$. If Γ is a graph with an orientation, and $\Delta \subseteq \Gamma$ is a subgraph, then Δ inherits an *induced orientation* from Γ by setting $E_+\Delta := E_+\Gamma \cap E\Delta$ and $E_-\Delta := E_-\Gamma \cap E\Delta$. Whenever we are dealing with a graph, equipped with an orientation, and a subgraph of that graph, we will always assume that the subgraph is given the induced orientation.

An *edge-path* p in Γ is a sequence of edges $e_1, e_2, \ldots, e_k$ with $e_i \in E\Gamma$ for all i and $o(e_j) = t(e_{j-1})$ for all $2 \leq j \leq k$. The length $|p|$, of the path p is the number of edges in p, that is $|p| = k$. We put $o(p) = o(e_1)$, and $t(p) = t(e_k)$. We define $p^{-1} := e_k, e_{k-1}, \ldots, e_1$.

A path p in a graph Γ is *reduced* if it does not contain any sub-paths of the form e, e^{-1} where $e \in E\Gamma$ is an edge.

Note that if Γ is a graph and $x \in V\Gamma$ is a vertex, there is a canonical identification of $\pi_1(\Gamma, x)$ with the set of reduced edge-paths from x to x in Γ. We will use this identification throughout the paper.

DEFINITION 2·2. *For two graphs Γ_1 and Γ_2, a morphism or a graph-map $f : \Gamma_1 \to \Gamma_2$ is a continuous map f such that $f(V\Gamma_1) \subseteq V\Gamma_2$ and such that the restriction of f to any topological edge $e \in \Gamma_1$ is a homeomorphism between e and some topological edge e' of Γ_2. Thus a morphism $f : \Gamma_1 \to \Gamma_2$ naturally defines functions $f : E\Gamma_1 \to E\Gamma_2$ and $f : V\Gamma_1 \to V\Gamma_2$ such that for any $e \in E\Gamma_1$ we have $f(\bar{e}) = \overline{f(e)} \in E\Gamma_2$, $o(f(e)) = f(o(e))$ and $t(f(e)) = f(t(e))$.*

DEFINITION 2·3. *Let Γ be a graph and $x \in V\Gamma$. Then the core of Γ at x is defined as:*

$$Core(\Gamma, x) = \cup\{p \,|\, \text{where } p \text{ is a reduced path in } \Gamma \text{ from } x \text{ to } x\}.$$

Note that $Core(\Gamma, x)$ is a connected subgraph of Γ containing x. If $Core(\Gamma, x) = \Gamma$ we say that Γ is a core graph with respect to x. *The graph $Core(\Gamma, x)$ has no degree 1 vertices except possibly x itself.*

We say that a graph Γ is a core graph *if Γ is connected and for every vertex $x \in V\Gamma$ we have $Core(\Gamma, x) = \Gamma$.*

If a graph T is a tree then for vertices $v, v' \in VT$ we denote by $[v, v']_T$ the unique reduced edge-path from v to v' in T.

PROPOSITION-DEFINITION 2·4. *Let Γ be a connected graph, and $x \in V\Gamma$. Choose a maximal subtree $T \subseteq \Gamma$, and an orientation $E\Gamma = E_+\Gamma \sqcup E_-\Gamma$. For $e \in E\Gamma$ define $[x, o(e)]_T$ to be the unique reduced path in T from x to $o(e)$, and let $s_e := [x, o(e)]_T \, e \, [t(e), x]_T$. Let $S_T := \{s_e \,|\, e \in E_+\Gamma - E_+T\}$. Then $\pi_1(\Gamma, x)$ is free and S_T is a free basis of $\pi_1(\Gamma, x)$.*

We call S_T the free basis of $\pi_1(\Gamma, x)$ dual to T.

We need to explicitly say how to rewrite elements of $\pi_1(\Gamma, x)$ in terms of the basis S_T, both as freely reduced words and cyclically reduced words.

PROPOSITION 2·5. *Let Γ be a connected graph, let $x \in V\Gamma$ and let $T \subseteq \Gamma$ be a maximal subtree. Suppose $E_+\Gamma - E_+T = \{e_1, \ldots, e_m\}$ where $e_i \neq e_j$ for $i \neq j$, so that $S_T = \{s_{e_i} | 1 \leq i \leq m\}$. Then:*

(i) Rewriting γ as a freely reduced word in S_T: Delete from γ all edges of T and replace each $e_i^{\pm 1}$ by $s_{e_i}^{\pm 1}$. The result is a freely reduced word over S_T representing $\gamma \in \pi_1(\Gamma, x)$.

(ii) Rewriting γ as a cyclically reduced word in S_T: First cyclically reduce the edge-path γ by removing the maximal initial and terminal segments of γ that cancel in the concatenation $\gamma\gamma$. The result is a subpath γ_1 of γ such that γ_1 is a closed cyclically reduced path (though γ_1 maybe based at a vertex different from x). Now apply the previous procedure to γ_1: delete all edges of T and replace each $e_i^{\pm 1}$ by $s_{e_i}^{\pm 1}$. The result is the cyclically reduced form of $\gamma \in \pi_1(\Gamma, x)$ over S_T.

2·2. *Graphs and subgroups*

In a seminal paper from 1983 Stallings [**56**] used labeled graphs to study subgroups of free groups. We give a brief exposition of the relevant definitions and results below and refer the reader to[**34**] for details.

Recall that we fix for the free group $F_N = F(A) = F(a_1, \dots, a_N)$ (where $N \geq 2$), a distinguished free basis $A = \{a_1, \dots, a_N\}$. If w is a word in $\Upsilon = A \sqcup A^{-1}$, we will denote by $\underline{w}$ the freely reduced word in Υ obtained from w by performing all possible (if any) free reductions.

DEFINITION 2·6. *An A-graph Γ consists of an underlying oriented graph where every edge $e \in E\Gamma$ is labeled by a letter $\mu(e) \in A \sqcup A^{-1}$ in such a way that $\mu(\bar{e}) = (\mu(e))^{-1}$. Multiple edges between vertices and loops at a vertex are allowed. An A-graph Γ is said to be folded if there does not exist a vertex x and two distinct edges e_1, e_2 with $o(e_1) = o(e_2) = x$ such that $\mu(e_1) = \mu(e_2)$. Otherwise Γ is said to be non-folded.*

An A-graph Γ is said to be *A-regular* if for every vertex $x \in V\Gamma$ and for every a_i, there is precisely one outgoing edge at x labeled by a_i and precisely one incoming edge at x labeled by a_i (thus, in particular, an A-regular graph is folded).

If Γ is an A-graph and $p = e_1, \dots, e_k$ is an edge-path in Γ, then p has a label which is a word in $A \sqcup A^{-1}$ and we denote this label by $\mu(p) = \mu(e_1)\mu(e_2)\dots\mu(e_k)$. The definitions immediately imply:

LEMMA 2·7. *An A-graph Γ is folded if and only if the label of every reduced path in Γ is a freely reduced word.*

DEFINITION 2·8. *For any two A-graphs Γ_1 and Γ_2, a map $f : \Gamma_1 \to \Gamma_2$ is an A-morphism if f is a graph-map such that $\mu(e) = \mu(f(e))$.*

For $F_N = F(a_1, \dots, a_N)$ we define the *standard N-rose* R_N to be the wedge of N loop-edges each labeled by $a_1, \dots, a_N$ respectively, at a vertex x_0. Then $F(A) = \pi_1(R_N, x_0)$.

For Γ an A-graph, $x \in V\Gamma$ and μ as before, we can define a map $\mu_\# : \pi_1(\Gamma, x) \to F(A)$ as $p \mapsto \underline{\mu(p)}$. This map is a group homomorphism.

NOTATION 2·9. *For Γ an A-graph, $x \in V\Gamma$ we say that (Γ, x) represents the subgroup $H := \mu_\#(\pi_1(\Gamma, x)) \leq F(A)$.*

PROPOSITION-DEFINITION 2·10. [56, 34] *Let $H \leq F(A)$. Then there exists a connected, folded A-graph Γ with $x_0 \in V\Gamma$ such that $\Gamma = Core(\Gamma, x_0)$ and (Γ, x_0) represents*

$$H = \{\mu(p) \,|\, p \text{ is a reduced path in } \Gamma \text{ from } x_0 \text{ to } x_0\} \leq F(A)$$

Moreover, such a (Γ, x_0) is unique. This graph (Γ, x_0) is called the Stallings subgroup graph of H *with respect to A.*

If (Γ, x_0) is the Stallings subgroup graph for H, then the labeling map $\mu : \pi_1(\Gamma, x_0) \to H$ is a group isomorphism. If $T \subseteq \Gamma$ is a maximal tree and $S_T = \{s_e | e \in E_+(\Gamma - T)\}$ is the dual free basis of $\pi_1(\Gamma, x_0)$, then $\mu(S_T) = \{\mu(s_e) | e \in E_+(\Gamma - T)\}$ is a free basis of H.

2·3. *Primitive, simple and non-filling elements*

DEFINITION 2·11 (Primitive and simple elements). *In the free group F_N, a non-trivial element $g \in F_N$ is called primitive in F_N if g belongs to some free basis of F_N.*

In the free group F_N, a non-trivial element $g \in F_N$ is called simple in F_N if g belongs to a proper free factor of F_N.

DEFINITION 2·12 (Non-filling elements). *An element $g \in F_N$ is said to be* non-filling *in F_N if there exists a splitting of F_N as an amalgamated free product $F_N = K *_C L$ or as an HNN-extension $F_N = \langle K, t | t^{-1}Ct = C' \rangle$, such that $C \leq F_N$ is either trivial or a*

maximal cyclic subgroup, such that in the $F_N = K *_C L$ *case* $C \neq K, C \neq L$, *and such that* $g \in K$.

An element $g \in F_N$ *is said to be* filling *in* F_N *if* g *is not non-filling.*

REMARK 2·13. *Note that if* $g \in F_N$ *is primitive, then it is also simple. Similarly, if* $g \in F_N$ *is simple, then* g *is non-filling.*

Also, for elements of F_N *the properties of being primitive, being simple and being non-filling are preserved under applying arbitrary automorphisms of* F_N.

The following known key fact relates the property of being filling in F_N to the compactification $\overline{\mathrm{CV}}_N$ of the projectivized Culler-Vogtmann Outer space CV_N. This compactification consists of the projective classes of all minimal "very small" isometric actions of F_N on $\mathbb{R}$-trees. An isometric action of a group G on an $\mathbb{R}$-tree T is called *very small* if for every nondegenerate segment of T the setwise stabilizer of that segment in G is either trivial or maximal infinite cyclic in G, and if the setwise stabilizer of every tripod in T is trivial. For example, the Bass-Serre tree corresponding to a splitting of F_N as an amalgamated product or an HNN-extension over a maximal infinite cyclic subgroup is a very small F_N-tree. See [33, 4] for a more detailed explanation of the relevant terminology. Also, if T is an $\mathbb{R}$-tree equipped with an isometric action of a group G, for $g \in G$ we denote $||g||_T := \inf_{x \in T} d(x, gx)$; the quantity $||g||_T$ is called the *translation length* of g in T.

PROPOSITION 2·14. [33, 55] *Let* $1 \neq g \in F_N$. *Then the following conditions are equivalent:*

(i) The element g *is filling in* F_N.

(ii) For every minimal very small isometric action of F_N *on a nontrivial simplicial tree* T *we have* $||g||_T > 0$.

(iii) For every minimal very small isometric action of F_N *on a nontrivial* $\mathbb{R}$*-tree* T *we have* $||g||_T > 0$.

Proof. The proof of this statement is implicit in [33, 55] but we sketch the argument for completeness.

Part (3) directly implies part (2). Since the simplicial splittings that appear in Definition 2·12 are very small, part (2) also directly implies part (1).

To see that part (1) implies part (3), suppose that $1 \neq g \in F_N$ is filling but that there exists a minimal very small isometric action of F_N on a nontrivial $\mathbb{R}$-tree T we have $||g||_T = 0$. Then a result of Bestvina and Feighn [4] (see also a paper of Guirardel [28]) implies that there exists a very small minimal simplicial F_N-tree T' with $||g||_{T'} = 0$. Taking the quotient graph of groups T'/F_N and collapsing all edges except one in this graph gives us a splitting of F_N as in Definition 2·12 such that g is conjugate to a vertex group element for that splitting. This contradicts the assumption that g is filling in F_N. Thus (1) implies (3), as required.

□

2·4. *Whitehead Graphs*

We now describe the relationship between simple elements, primitive elements, and Whitehead graphs.

DEFINITION 2·15. *[Whitehead graph] Let* $F_N = F(A)$ *be as before and let* $w \in F_N$ *be*

a nontrivial cyclically reduced word. Let c be the first letter of w. The word wc is then freely reduced.

The Whitehead graph of w *with respect to A, denoted by $Wh_A(w)$, is an undirected graph whose set of vertices $V(Wh_A(w)) = \Upsilon$. Edges are added as follows: For $a, b \in V(Wh_A(w))$, there is an undirected edge joining a^{-1} and b if ab or $b^{-1}a^{-1}$ occurs as a subword of wc.*

Note that if $\tilde{w}$ is a cyclic permutation of w or of w^{-1} then $Wh_A(w) = Wh_A(\tilde{w})$.

For an arbitrary $1 \neq g \in F_N$, we put $Wh_A(g) := Wh_A(w)$, where w is the cyclically reduced form of g in $F(A)$.

Recall that a *cut vertex* in a graph Δ is a vertex x such that $\Delta - \{x\}$ is disconnected. Note that if Δ has at least one edge and is disconnected, then Γ does possess a cut vertex; namely any end-vertex of an edge of Δ is a cut vertex in this case.

Generalizing a result of Whitehead, Stallings established the relationship between simple elements and Whitehead graphs [57]:

PROPOSITION 2·16. [57] *Let $F_N = F(A)$, where $N \geq 2$ and let $g \in F(A)$ be a cyclically reduced word. If g is simple, then the Whitehead graph $Wh_A(g)$ has a cut vertex.*

Notice that Remark 2·13 implies that if $g \in F(A)$ is primitive, then $Wh_A(g)$ has a cut vertex.

REMARK 2·17. *Stallings' definition of Whitehead graphs differs slightly from our definition. Assume the same setting as in Definition 2·15. Stallings adds an edge from a^{-1} to b for each occurrence of a subword ab in wc. Let us call the Whitehead graph of a cyclically reduced word w under Stallings' definition Γ, and the corresponding graph under our definition Γ_1. It is clear that $V(\Gamma) = V(\Gamma_1)$. Further it is easily checked that $x \in V(\Gamma)$ is a cut-vertex in Γ if and only if $x \in V(\Gamma_1)$ is a cut-vertex in Γ_1. Thus Proposition 2·16 holds for our definition of Whitehead graphs just as well.*

Finally, note that if a graph has a reduced circuit that contains all the vertices, then the graph can not have a cut vertex. This observation applies, for instance, to the Whitead graph of an element $g \in F_N$ when the string $a_N^2 a_1^2 a_2^2 \dots a_N^2$ occurs as a subword of a cyclically reduced form of g. In this case g is not simple (and hence not primitive) as its Whitehead graph does not have a cut vertex. We state this fact explicitly as a corollary of Proposition 2·16:

COROLLARY 2·18. *Let $F_N = F(A)$, where $N \geq 2$ and $A = \{a_1, \dots, a_N\}$. If a cyclically reduced word $w \in F(A)$ contains the subword $a_N^2 a_1^2 a_2^2 \dots a_N^2$ then w is not simple (and hence not primitive) in $F(A)$.*

The Whitehead graph, as defined above, records the information about two-letter subwords in the cyclically reduced form w of a nontrivial element $g \in F_N = F(A)$. There are also generalizations of the Whitehead graph recording the information about k-letter subwords of w, where $k \geq 2$ is a fixed integer. These generalizations are commonly known as "Rauzy graphs" or "initial graphs" and naturally occur in the study of geodesic currents on free groups [30, 31, 32].

We do not formally define these "level k" versions of the Whitehead graph in this paper, because we only need the following specific statement related to the $k = 3$ case:

PROPOSITION 2·19. [21] *Let $F_N = F(A)$, where $N \geq 2$ and $A = \{a_1, \ldots, a_N\}$. Let w be a nontrivial cyclically reduced word in $F(A)$ such that for every freely reduced word $v \in F(A)$ with $|v| = 3$ the word v occurs as a subword of a cyclic permutation of w or of w^{-1}.*

Then w is filling in F_N (and, in particular, w is non-simple and non-primitive in F_N).

3. *Primitivity, Simplicity, and Non-Filling Index Functions*

In 1949 Marshall Hall Jr. proved in [29] that any finitely generated subgroup of a free group F_N is a free factor of a finite index subgroup of F_N. We state the result in a more precise form, as stated in [56]:

PROPOSITION 3·1. [56] *Let $\alpha_1, \ldots, \alpha_k, \beta_1, \ldots, \beta_l$ be elements of a free group F_N. Let S be the subgroup of F_N generated by $\{\alpha_1, \ldots, \alpha_k\}$. Suppose $\beta_i \notin S$ for $i = 1, \ldots, l$. Then there exists a subgroup S' of finite index in F_N, such that $S \subset S'$, $\beta_i \notin S'$ for $i = 1, \ldots, l$, and there exists a free basis of S' having a subset that is a free basis of S.*

If we pick $g \neq 1 \in F_N$ and apply the above result to the infinite cyclic subgroup $S = \langle g \rangle$, we get that there must exist a finite index subgroup S' of F_N such that g is a primitive element in S' (and hence $g \in S'$ is non-simple and non-filling in S').

This fact motivates the following definition:

DEFINITION 3·2. *[Primitivity, simplicity and non-filling indexes]*

Let $N \geq 2$ be an integer and let F_N be a free group of rank N. Let $1 \neq g \in F_N$.

Define the primitivity index $d_{prim}(g) = d_{prim}(g, F_N)$ *of g in F_N to be the smallest possible index for a subgroup $L \leq F_N$ containing g as a primitive element.*

Define the simplicity index $d_{simp}(g) = d_{simp}(g, F_N)$ *to be the smallest possible index for a subgroup $L \leq F_N$ containing g as a simple element.*

Finally, define the non-filling index $d_{fill}(g) = d_{fill}(g, F_N)$ *to be the smallest possible index for a subgroup $L \leq F_N$ containing g as a non-filling element.*

As noted above, Proposition 3·1 implies that for every nontrivial $g \in F_N$ we have $d_{fill}(g) \leq d_{simp}(g) \leq d_{prim}(g) < \infty$.

DEFINITION 3·3 (Primitivity, simplicity and non-filling index functions). *Let F_N be a free group of rank $N \geq 2$ and let A be a free basis of F_N. For any $n \geq 1$ define the* primitivity index function *for F_N as*

$$f_{prim}(n) = f_{prim}(n; F_N) := \max_{\substack{1 \leq |g|_A \leq n,\, g \neq 1 \\ g \text{ not a proper power in } F_N}} d_{prim}(g)$$

Similarly, for $n \geq 1$ define the simplicity index function *for F_N as*

$$f_{simp}(n) = f_{simp}(n; F_N) := \max_{\substack{1 \leq |g|_A \leq n,\, g \neq 1 \\ g \text{ not a proper power in } F_N}} d_{simp}(g)$$

Finally, for for $n \geq 1$ define the non-filling index function *for F_N as*

$$f_{fill}(n) = f_{fill}(n; F_N) := \max_{\substack{1 \leq |g|_A \leq n,\, g \neq 1 \\ g \text{ not a proper power in } F_N}} d_{fill}(g)$$

It is easy to see that the definitions of $f_{prim}(n; F_N)$, $f_{simp}(n; F_N)$ and $f_{fill}(n; F_N)$ do not depend on the choice of a free basis A of F_N. Note that $f_{prim}(n)$ is the smallest

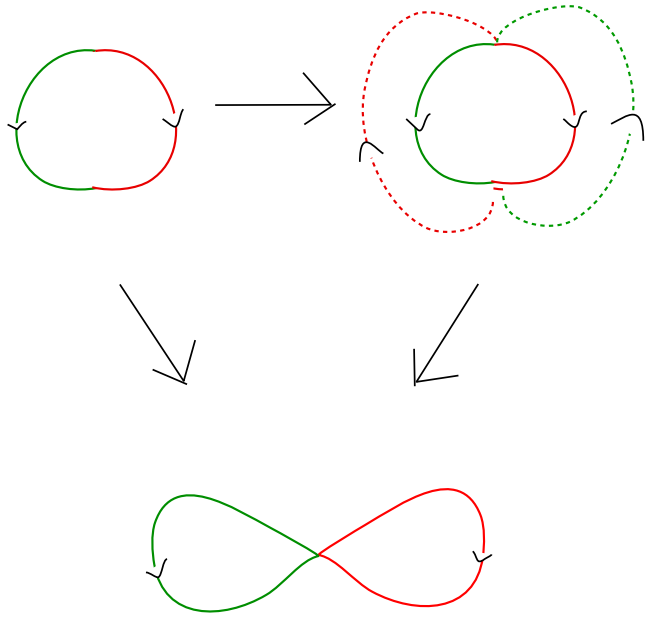

Fig. 1. Proof by Picture for Proposition 3·5

monotone non-decreasing function such that for every non-trivial root-free $g \in F_N$ we have $d_{prim}(g) \leq f_{prim}(|g|_A)$; similar reformulations hold for $f_{simp}(n)$ and $f_{fill}(n)$. We recall the following well-known fact, which is Lemma 8.10 in [34]:

LEMMA 3·4. *Let Γ be a finite folded A-graph. Then there exists a finite folded A-regular graph Γ' such that Γ is a subgraph of Γ' and such that $V\Gamma = V\Gamma'$.*

PROPOSITION 3·5. *For every non-trivial cyclically reduced word $w \in F(A)$ of length n, there exists a finite index subgroup $H \leq F(A)$ of index n such that $w \in H$ is primitive in H.*

Proof. Take the word w of length n and write it on a circle of simplicial length n. Pick a vertex x as the base vertex. Call this graph (Γ_w, x). By Lemma 3·4 we can complete this graph to a finite cover (Γ'_w, x) of the N-rose without adding any extra vertices. Thus (Γ'_w, x) has n vertices and represents a subgroup H of F_N of index precisely n. The fact that w is realized as the label of a simple closed curve in (Γ'_w, x) implies that w is a primitive element in H. It is clear that $w \in H$ by definition of H. Note that since (Γ', x) has no extra vertices, a maximal tree T of (Γ, x) consists of all but one edge of the simple closed curve representing w. Let $e \in E_+\Gamma' - T$. Then $\mu(s_e) = w$ and hence w is primitive. See Figure 1 for a pictorial proof.

□

Proposition 3·5, together with the definitions, directly implies:

LEMMA 3·6. *Let $N \geq 2$ and let F_N be free of rank N. Then the following hold:*
(i)If $1 \neq g \in F_N = F(A)$ then

$$d_{fill}(g) \leq d_{simp}(g) \leq d_{prim}(g) \leq ||g||_A \leq |g|_A = n.$$

(ii)For every $n \geq 1$ we have

$$f_{fill}(n) \leq f_{simp}(n) \leq f_{prim}(n) \leq n.$$

(iii)Let $1 \neq g \in F_N$ and let $\alpha \in Aut(F_N)$. Then $d_{prim}(g) = d_{prim}(\alpha(g))$, $d_{simp}(g) = d_{simp}(\alpha(g))$ and $d_{fill}(g) = d_{fill}(\alpha(g))$.

(iv)If $1 \neq g \in F_N$ and $k \geq 1$ is an integer, then $d_{simp}(g^k) \leq d_{simp}(g)$ and $d_{fill}(g^k) \leq d_{fill}(g)$.

In particular, part (3) of the above lemma shows that for g_1, g_2 conjugate non-trivial elements of F_N, we have $d_{prim}(g_1) = d_{prim}(g_2)$, $d_{simp}(g_1) = d_{simp}(g_2)$ and $d_{fill}(g_1) = d_{fill}(g_2)$.

As noted above, if $1 \neq g \in F_N$ and $k \geq 1$ is an integer, then $d_{simp}(g^k) \leq d_{simp}(g)$ and $d_{fill}(g^k) \leq d_{fill}(g)$. However, the function $d_{prim}(g)$ does not behave well under taking powers, as demonstrated by the following lemma:

LEMMA 3·7. *For any $a_i \in \{a_1, \dots, a_N\}$, and any positive integer n, $d_{prim}(a_i^n) = n$.*

Proof. As noted above, for every nontrivial $g \in F_N$ we have $d_{simp}(g) \leq d_{prim}(g) \leq ||g||_A$. Thus $d_{prim}(a_i^n) \leq ||a_i^n||_A = n$. We need to show that $d_{prim}(a_i^n) \geq n$.

Let $d = d_{prim}(a_i^n)$ and let $H \leq F_N$ be a subgroup of index d such that $a_i^n \in H$ and that a_i^n is a primitive element of H. Let $(\Gamma, *)$ be the d-fold cover of R_N corresponding to H, so that for the covering map $p : \Gamma \to R_N$ have $\pi_1(\Gamma, *) \cong H$ and $p_\# = \mu : \pi_1(\Gamma, *) \to H \leq F_N = \pi_1(R_N, x_0)$ is an isomorphism.

The fact that $a_i^n \in H$ implies that there exists a reduced closed path γ from $*$ to $*$ in Γ with $\mu(\gamma) = a_i^n$. Since a_i^n is primitive in H, the element γ is primitive in $\pi_1(\Gamma, *)$.

Since a_i^n is cyclically reduced, the closed path γ is also cyclically reduced. We claim that γ is a simple closed path in Γ. Indeed, suppose not. Then $\gamma = \gamma_1^k$ where $k \geq 2$ and where γ_1 is a simple closed path at $*$ in Γ with label $a_i^{n/k}$. Therefore γ is a proper power in $\pi_1(\Gamma, *)$, which contradicts the fact that γ is primitive in $\pi_1(\Gamma, *)$. Thus indeed γ is a simple closed path in Γ with label a_i^n. This means that the full p-preimage of the i-th petal of R_N, labeled a_i, in Γ consists of $\geq n$ distinct topological edges. Therefore the degree d of the cover $p : \Gamma \to R_N$ satisfies $d \geq n$.

Thus $d = d_{prim}(a_i^n) \geq n$. Since we already know that $d_{prim}(a_i^n) \leq n$, it follows that $d_{prim}(a_i^n) = n$, as required. □

Avoiding the bad behavior of $d_{prim}(g)$ under taking powers of g, demonstrated by Lemma 3·7, is the main reason why in Definition 3·2 we take the maximum over all root-free nontrivial elements $g \in F_N$ with $|g|_A \leq n$ rather than over all nontrivial $g \in F_N$ with $|g|_A \leq n$.

4. *Algorithmic computability of $d_{prim}(g)$, $d_{simp}(g)$, and $d_{fill}(g)$*

In this section we will establish algorithmic computability of $d_{prim}(g)$, $d_{simp}(g)$, and $d_{fill}(g)$. Consequently, we will also establish the algorithmic computability of $f_{prim}(n)$, $f_{simp}(n)$, and $f_{fill}(n)$.

We first need to recall some basic definitions and facts related to Whitehead automorphisms and Whitehead's algorithm. We only briefly cover this topic here and refer the reader for further details to [**40**, pp. 30-35] and to [**43**, **35**, **32**, **49**] for some of the recent developments. As before, $F_N = F(A) = F(a_1, \dots, a_N)$ is the free group of rank $N \geq 2$ with a free basis $A = \{a_1, \dots, a_N\}$.

DEFINITION 4·1 (Whitehead automorphisms). *A* Whitehead automorphism *τ of $F_N = F(A)$ with respect to A is an automorphism τ of $F(A)$ of one of the following types:*

(i) There exists a permutation t of $\Upsilon = A \sqcup A^{-1}$ such that $\tau|_\Upsilon = t$. In this case τ is called a relabeling automorphism *or a* Whitehead automorphism of the first kind.

(ii) There exists an element $a \in \Upsilon$ which we call the multiplier *such that for any $x \in \Upsilon$, $\tau(x) \in \{x, xa, a^{-1}x, a^{-1}xa\}$. In this case τ is called a* Whitehead automorphism of the second kind.

Note that since $\tau \in Aut(F(A))$, if τ is a Whitehead automorphism of the second kind with multiplier a, then $\tau(a) = a$. Also for any $a \in \Upsilon$, the inner automorphism corresponding to conjugation by a is a Whitehead automorphism of the second kind.

DEFINITION 4·2 (Automorphically minimal and Whitehead minimal elements). *An element $g \in F(A) = F_N$ is* automorphically minimal *in $F(A)$ with respect to a basis A of F_N if, for every $\varphi \in Aut(F(A))$ we have $||g||_A \leq ||\varphi(g)||_A$.*

An element $g \in F(A)$ is Whitehead minimal *in $F(A)$ with respect to a free basis A if, for every Whitehead automorphism τ of $F(A)$ we have $||g||_A \leq ||\tau(g)||_A$. For an element $g \in F(A)$ we say that $\tilde{g} \in F(A)$ is a* Whitehead minimal form *of g with respect to A if $\tilde{g}$ is Whitehead minimal with respect to A and there exists an automorphism $\varphi \in Aut(F(A))$ such that $\varphi(g) = \tilde{g}$.*

Note that neither Whitehead automorphisms of the first kind nor inner automorphisms change the cyclically reduced length of an element.

The following proposition summarizes the key known facts regarding Whitehead's algorithm (see [**61**] for the original proof by Whitehead and see [**40**, Proposition 4.17] for a modern exposition):

PROPOSITION 4·3 (Whitehead's Theorem). *Let $N \geq 2$ and let $F_N = F(A)$ be free of rank N with a free basis A. Then:*

(i) An element $g \in F(A)$ is automorphically minimal in $F(A)$ with respect to a basis A if and only if g is Whitehead minimal in $F(A)$ with respect to A. (Hence $g \in F(A)$ is not automorphically minimal with respect to A if and only if there exists a Whitehead automorphism τ such that $||\tau(g)||_A < ||g||_A$).

(ii) Whenever $u, v \in F(A)$ are Whitehead minimal with respect to A such that the orbits $Aut(F(A))u = Aut(F(A))v$ (so that, in particular, $||u||_A = ||v||_A$), then there exists a sequence of Whitehead automorphisms $\tau_1, \ldots, \tau_m$ of $F(A)$ with respect to A such that $\tau_m ... \tau_1(u) = v$ and that $||\tau_i ... \tau_1(u)||_A = ||u||_A$ for $i = 1, ..., m$.

Note that part (2) of Proposition 4·3 holds even if u, v are conjugate in $F(A)$ since conjugation by an element of $A^{\pm 1}$ is a Whitehead automorphism.

4·1. *Algorithmic computability of $d_{prim}(g)$ and $d_{simp}(g)$*

The following useful lemma explicitly states the relationship between primitivity, simplicity and Whitehead minimality:

LEMMA 4·4. *Let $1 \neq w \in F(A) = F_N$.*

(i) w primitive in $F(A)$ if and only if every (equivalently, some) Whitehead minimal form $\widetilde{w}$ of w has $||\widetilde{w}||_A = 1$.

(ii) w is simple in $F(A)$ if and only if some Whitehead minimal form $\widetilde{w}$ of w misses an $a_i^{\pm 1}$.

(iii)w is simple in $F(A)$ if and only if every Whitehead minimal cyclically reduced form $\widetilde{w}$ of w misses an $a_i^{\pm 1}$.

Proof.

Part (1) of the lemma is well-known and follows directly from Proposition 4·3.

If some Whitehead minimal form $\widetilde{w}$ of w misses an $a_i^{\pm 1}$, then w is simple in $F(A)$ as $w \in F(B)$ where $B = A - \{a_i\}$ and $F(B)$ is a proper free factor of $F(A)$.

Conversely, suppose that w is simple in $F(A)$. Then there exists an automorphism φ of $F(A)$ such that the cyclically reduced form $\widehat{w}$ of $\varphi(w)$ misses $a_N^{\pm 1}$.

Claim 1. We claim that some Whitehead minimal form of $\widehat{w}$ also misses $a_N^{\pm 1}$.

We prove this claim by induction on $||\widehat{w}||_A$. If $||\widehat{w}||_A = 1$, then the claim clearly holds. Suppose now that $||\widehat{w}||_A = m > 1$ and that the claim has been established for all nontrivial cyclically reduced words in $F(a_1, \ldots, a_{N-1})$ of length $\leq m-1$.

If $\widehat{w}$ is already Whitehead minimal in $F(A)$ then we are done as the claim holds in this case.

If $\widehat{w}$ is not Whitehead minimal in $F(A)$ then there exists a Whitehead automorphism τ of $F(A)$ such that $||\tau(\widehat{w})||_A < ||\widehat{w}||_A$. Note first that since the cyclically reduced length of $\widehat{w}$ changes under τ, we must have that τ is a Whitehead automorphisms of the second kind that is not an inner automorphism.

Let $a \in \Upsilon = A \sqcup A^{-1}$ be the multiplier of τ. If $a = a_N^{\pm 1}$, since $\widehat{w}$ is a cyclically reduced word in $F(A)$ that misses the letter $a_N^{\pm 1}$, the definition of a Whitehead automorphism implies that there can be no cancellation in $\tau(\widehat{w})$ between the letters $\{a_1, \ldots, a_{N-1}\}$ when a cyclically reduced form of $\tau(\widehat{w})$ is computed. Hence $||\tau(\widehat{w})||_A \geq ||\widehat{w}||_A$, contrary to the fact that $||\tau(\widehat{w})||_A < ||\widehat{w}||_A$. Therefore $a \in \{a_1, \ldots, a_{N-1}\}^{\pm 1}$. We then define a Whitehead automorphism τ' of $F(a_1, \ldots, a_{N-1})$ with respect to $\{a_1, \ldots, a_{N-1}\}$ as $\tau' = \tau|_{\{a_1, \ldots, a_{N-1}\}}$. Hence $\tau(\widehat{w}) = \tau'(\widehat{w})$. Thus $\tau(\widehat{w})$ still misses $a_N^{\pm 1}$ and $||\tau(\widehat{w})||_A < ||\widehat{w}||_A = m$. Applying the inductive hypothesis to $\tau(\widehat{w})$, we conclude that some Whitehead minimal form $\widetilde{w}$ of $\tau(\widehat{w})$ in $F(A)$ misses $a_N^{\pm 1}$. Then $\widetilde{w}$ is also a Whitehead minimal form of $\widehat{w}$, and Claim 1 is verified.

Thus we have established part (2) of the lemma.

To see that part (3) holds, note that if every Whitehead minimal cyclically reduced form $\widetilde{w}$ of w misses an $a_i^{\pm 1}$ then w is simple in $F(A)$.

Now suppose w is simple in $F(A)$. From (2) we know that there is a $\widetilde{w}$ Whitehead minimal cyclically reduced form of w that misses $a_N^{\pm 1}$. Let w' be another Whitehead minimal cyclically reduced form of w in $F(A)$. Then $Aut(F(A))w' = Aut(F(A))\widetilde{w}$, and so by part (2) of Proposition 4·3, there exists a sequence of Whitehead automorphisms $\tau_1, \ldots, \tau_m$ of $F(A)$ with respect to A such that $\tau_m ... \tau_1(\widetilde{w}) = w'$ and that $||\tau_i ... \tau_1(\widetilde{w})||_A = ||w'||_A$ for $i = 1, ..., m$.

For $j = 0, 1, \ldots, m$ denote $w_j = \tau_j ... \tau_1(\widetilde{w})$, where $w_0 = \widetilde{w}$.

Claim 2. We claim that for each $j = 0, \ldots, m$ the cyclically reduced form of w_j misses some $a_i^{\pm 1}$.

We will establish Claim 2 by induction on j.

If $j = 0$ then $w_0 = w$ and there is nothing to prove. Suppose now that $j \geq 1$ and that the claim has been verified for w_{j-1}.

Thus the cyclically reduced form of w_{j-1} misses some $a_i^{\pm 1}$. If τ_j is a Whitehead automorphism of the first kind, it is clear that the cyclically reduced form of $\tau_j(w_{j-1}) = w_j$ still misses some $a_k^{\pm 1}$ (this $a_k^{\pm 1}$ is not necessarily $a_i^{\pm 1}$). Suppose now that τ_j is a Whitehead automorphism of the second kind. The restriction that $||\tau_j(w_{j-1})||_A = ||w_{j-1}||$

forces the condition that either $\tau_j(w_{j-1})$ is equal to w_{j-1} after cyclic reduction, or else τ_j is a Whitehead automorphism of the second kind with multiplier $a \in B \sqcup B^{-1}$ where $B = \{x \in A \sqcup A^{-1} | x$ occurs in the cyclically reduced form of $w_{j-1}\}$ (in particular $a \neq a_i^{\pm 1}$). In both cases we see that the cyclically reduced form of w_j still misses $a_i^{\pm 1}$, as required. This completes the inductive step and the proof of Claim 2.

Applying Claim 2 with $j = m$ shows that the cyclically reduced form of $w' = w_m$ misses some $a_i^{\pm 1}$, and part (3) of the lemma is proved. □

PROPOSITION 4·5. *Let $1 \neq g \in H \leq F(A)$, where H is a proper free factor of $F(A)$. Then the following hold:*

(i) The element g is primitive in H if and only if g is primitive in F_N.

(ii) There is an algorithm which decides, given $g \in F(A)$, whether or not $g \in F(A)$ is primitive.

(iii) There is an algorithm which given $g \in F(A)$, whether or not $g \in F(A)$ is simple.

Proof.

We first prove part (1). The "only if" direction is obvious. Thus we assume that $g \in H$ is primitive in F_N.

Let $K \leq F_N$ be such that $F_N = H * K$. Let $\mathcal{B}_H = \{h_1, \dots, h_l\}$ be a free basis for H, and $\mathcal{B}_K = \{k_1, \dots, k_m\}$ be a free basis for K. Then $\mathcal{B}_F = \{h_1, \dots, h_l, k_1, \dots, k_m\}$ is a free basis for F_N (here $l + m = n$).

Since $g \in H$, then g is a freely reduced word over $\mathcal{B}_H$, with cyclically reduced form w. We prove that g is primitive in H by induction on the length m of w.

If w has length 1, then g is primitive in H, as required. If w has length $m > 1$, then the fact that w is primitive in F_N implies that w is not Whitehead minimal in F_N with respect to the free basis $\mathcal{B}_F$ of F_N. Hence there exists a Whitehead automorphism τ of F_N with respect to $\mathcal{B}_F$ such that $||\tau(w)||_{\mathcal{B}_F} < m$. (Note that at this point we do not yet know that $\tau(w) \in H$ since τ is a Whitehead automorphism of F_N, and not of H).

By the same argument as in the proof of Lemma 4·4, we see that there exists a Whitehead automorphism τ' of $H = F(\mathcal{B}_H)$ such that $\tau'(w) = \tau(w)$. Then $\tau(w) = \tau'(w) \in H$ is primitive in F_N with $||\tau(w)||_{\mathcal{B}_F} < m$. Therefore by the inductive hypothesis the element $\tau(w) = \tau'(w)$ is primitive in H. Since $\tau' \in Aut(H)$, it follows that w is also primitive in H, as required. Thus part (1) of the proposition holds.

To prove parts (2) and (3) for $g \in F(A) = F(a_1, \dots, F_N)$, let $\widetilde{g}$ be a Whitehead minimal form of g in $F(A)$ (such $\widetilde{g}$ exists by Proposition 4·3. By part (1) of Lemma 4·4, $||\widetilde{g}||_A = 1$ if and only if g is primitive in $F(A)$. By part (3) of Lemma 4·4, $\widetilde{g}$ misses some $a_i^{\pm 1}$ if and only if w is simple in $F(A)$. □

REMARK 4·6. *The algorithm described in part (2) of Proposition 4·5 is due to Whitehead* [61]. *The first algorithms for deciding whether an element of F_N is simple in F_N were provided by Stallings* [57] *and Stong* [58] *in 1990s. Their algorithms are somewhat different from the algorithm given in part (3) of Proposition 4·5 above, but they are also based on using Whitehead's algorithm.*

DEFINITION 4·7 (Principal quotient). *Following the terminology of* [34], *for a finite connected A-graph Γ_1 and a folded A-graph Γ_2, we say that Γ_2 is a* principal quotient *of Γ_1 if there exists a surjective A-morphism $\Gamma_1 \to \Gamma_2$.*

DEFINITION 4·8. *Let $w \in F_N = F(A)$ be a nontrivial cyclically reduced word. We*

denote by C_w the A-graph which is a simplicial circle subdivided into $n = ||w||_A$ topological edges, such that the label of the closed path of length n corresponding to going around this circle once from some vertex $*$ to $*$ is the word w.

By definition, the graph C_w has a distinguished base-vertex $*$. Thus a principal quotient of C_w also come equipped with a distinguished base-vertex. We say that (Γ, x) is a principal quotient *of C_w if Γ is a finite connected folded A-graph, if $x \in V\Gamma$ and if there exists a surjective A-morphism $f : C_w \to \Gamma$ such that $f(*) = x$.*

Note that if (Γ, x) is a principal quotient of C_w, then there exists a unique path $\gamma_{w,x}$ in Γ starting with x and with label w, and, moreover, this path is closed and passes through every topological edge of Γ.

The following lemma is an immediate corollary of the definitions:

LEMMA 4·9. *The following hold:*

(i) Let Γ_1 be a finite connected A-graph and Γ_2 be a finite folded A-graph. Then Γ_2 is a principal quotient of Γ_1 if and only if Γ_2 can be obtained from Γ_1 by the following procedure: choose some partition $V\Gamma_1 = V_1 \sqcup \cdots \sqcup V_m$ (with all $V_i \neq \emptyset$), then for each $i = 1, \ldots, m$ collapse V_i to a single vertex to get an A-graph Γ_1', and then fold the graph Γ_1' to obtain Γ_2.

(ii) If $w \in F_N = F(A)$ is a nontrivial cyclically reduced word and Γ is a finite connected folded A-graph, then Γ is a principal quotient of C_w if and only if Γ is a core graph and there exists a closed path γ_w in Γ with label w such that γ_w passes through every topological edge of Γ.

A priori it is unclear that the functions $f_{prim}(n)$ and $f_{simp}(n)$ are even computable for a given F_N. We now give an algorithm that calculates $d_{prim}(g)$ and $d_{simp}(g)$ for any non-trivial g. This would then show that the functions $f_{prim}(n)$ and $f_{simp}(n)$ are indeed algorithmically computable.

DEFINITION 4·10. *Let $1 \neq g \in F_N = F(A)$ and let $w \in F(A)$ be the cyclically reduced form of g. We denote by $\mathcal{G}_0(w)$ the set of all finite connected folded basepointed A-graphs (Γ, x) such that there exists a closed path γ from x to x labeled w with the property that γ passes through every topological edge of Γ at least once and such that either the labeling map $\Gamma \to R_N$ is not a covering (that is, there exists a vertex of Γ of degree $< 2N$), or the labeling map $\Gamma \to R_N$ is a covering and the element $\gamma \in \pi_1(\Gamma, x)$ is simple in $\pi_1(\Gamma, x)$.*

We denote by $\mathcal{G}(w)$ the set of all finite connected folded basepointed A-graphs (Γ, x) such that there exists a closed path γ from x to x labeled w with the property that γ passes through every topological edge of Γ at least once and such that the element $\gamma \in \pi_1(\Gamma, x)$ is primitive in $\pi_1(\Gamma, x)$.

Let $(\Gamma, x) \in \mathcal{G}(w)$ or $(\Gamma, x) \in \mathcal{G}_0(w)$. Since w is cyclically reduced and γ passes through every topological edge of Γ at least once, every vertex of Γ has degree ≥ 2, so that Γ is a core graph.

Note further that the condition that γ is simple in $\pi_1(\Gamma, x)$ is equivalent to the condition that w is simple in the subgroup $H \leq F_N$ represented by (Γ, x). This follows from the fact that the labeling map gives an isomorphism $\mu : \pi_1(\Gamma, x) \to H$, with $\mu(\gamma) = w$.

We recall the following basic fact:

LEMMA 4·11 ([34], p.13). *Let Γ be a folded connected A-graph and let Γ' be a con-*

nected subgraph of Γ. Let $*$ be a vertex of Γ'. If $H' \leq F(A)$ is the subgroup represented by $(\Gamma', *)$ and H is the subgroup represented by $(\Gamma, *)$, then H' is a free factor of H.

REMARK 4·12. *In the setting of Lemma 4·11, $\pi_1(\Gamma', *)$ is a free factor of $\pi_1(\Gamma, *)$.*

PROPOSITION 4·13. *Let $1 \neq g \in F_N = F(A)$ and let $w \in F(A)$ be the cyclically reduced form of g. Then the following hold:*

(i) The number $d_{prim}(g)$ equals to the minimum of $\#V\Gamma$, taken over all $(\Gamma, x) \in \mathcal{G}(w)$.

(ii) The number $d_{simp}(g)$ equals to the minimum of $\#V\Gamma$, taken over all $(\Gamma, x) \in \mathcal{G}_0(w)$.

Proof. We give a proof of part (2). The proof of part (1) is very similar in nature. However, it additionally involves using part (1) of Proposition 4·5 to prove one of the inequalities. For $1 \neq g \in F_N = F(A)$ and $w \in F(A)$ the cyclically reduced form of g, let $\overline{d_{simp}}(g) = \min_{(\Gamma,x)\in\mathcal{G}_0(w)} \#V\Gamma$. First suppose that $H \leq F_N$ such that $[F_N : H] = d_{simp}(g) = d_{simp}(w)$, and that $w \in H$ is simple in H. Let (Γ, x) be the graph representing H as in Proposition-Definition 2·10. We have that $\#V\Gamma = d_{simp}(w)$. Since $w \in H$, there exists a path γ from x to x in Γ with label w. Also since $w \in H$ is simple in H, $\gamma \in \pi_1(\Gamma, x)$ is simple in $\pi_1(\Gamma, x)$. Let $\Gamma' \subseteq \Gamma$ be the subgraph spanned by γ. Then γ is a path from x to x in Γ' that passes through every topological edge in Γ' at least once. If $\Gamma' = \Gamma$, then the labeling map $\Gamma' \to R_N$ is a covering. Since γ is simple in $\Gamma = \Gamma'$, we have $(\Gamma', x) \in \mathcal{G}_0(w)$. Since $\#V\Gamma' = \#V\Gamma = d_{simp}(g)$, we have that $\overline{d_{simp}}(g) \leq d_{simp}(g)$. If $\Gamma' \neq \Gamma$, then $\#V\Gamma' \leq \#V\Gamma$ and $\#E\Gamma - \#E\Gamma' \geq 1$. From Remark 4·12, (Γ', x) is a proper free factor of (Γ, x). In this case the labeling map $\Gamma' \to R_N$ is not a covering and $(\Gamma', x) \in \mathcal{G}_0(w)$. Thus $\overline{d_{simp}}(g) \leq d_{simp}(g)$.

Conversely suppose that $(\Gamma, x) \in \mathcal{G}_0(w)$ with $\#V\Gamma = \overline{d_{simp}}(g)$. Let γ be the closed path from x to x labeled by w such that γ passes through every topological edge of Γ at least once. If the labeling map $\Gamma \to R_N$ is a covering then $\gamma \in \pi_1(\Gamma, x)$ is simple in $\pi_1(\Gamma, x)$ by definition of $\mathcal{G}_0(w)$. Let H be the subgroup represented by (Γ, x). H is then a subgroup of F_N of index $\overline{d_{simp}}(g)$ with $w \in H$ and w simple in H. Hence $d_{simp}(g) = d_{simp}(w) \leq \overline{d_{simp}}(g)$. If the labeling map $\Gamma \to R_N$ is not a covering, we use Lemma 3·4 to complete (Γ, x) to a finite cover $(\widehat{\Gamma}, x)$ of R_N without adding any extra vertices and by adding at least one edge. Again from Remark 4·12, (Γ, x) is a proper free factor of $(\widehat{\Gamma}, x)$. Hence $\gamma \in \pi_1(\widehat{\Gamma}, x)$ is simple in $\pi_1(\widehat{\Gamma}, x)$. Let H be the subgroup represented by $(\widehat{\Gamma}, x)$. We have shown that $w \in H$ is simple in H. Since $\#V\widehat{\Gamma} = \#V\Gamma = \overline{d_{simp}}(g)$, we see that $d_{simp}(g) \leq \overline{d_{simp}}(g)$. □

We can now prove:

THEOREM 4·14. *Let $F_N = F(A)$, where $N \geq 2$ and where $A = \{a_1, \ldots, a_N\}$ is a free basis of F_N. Then:*

(i) There exists an algorithm that, given $1 \neq g \in F_N$, computes $d_{prim}(g)$ and $d_{simp}(g)$.

(ii) There exists an algorithm that, for every $n \geq 1$ computes $f_{prim}(n)$ and $f_{simp}(n)$

Proof.

Let $1 \neq g \in F_N$ and let w be the cyclically reduced form of g. Note that a finite connected folded base-pointed A-graph (Γ, x) admits a closed path γ from x to x labeled w and passing through every topological edge of Γ at least once if and only if (Γ, x) is a principal quotient of C_w with x being the image of the base-vertex $*$ of C_w.

Therefore we can algorithmically find all the graphs in $\mathcal{G}_0(w)$ as follows: List all partitions on VC_w. For each partition of VC_w as a disjoint union of nonempty subsets

$V_1, \dots V_m$, collapse V_i to a single vertex for $i = 1, \dots, m$, and fold the resulting graph to obtain a principal quotient (Γ, x) of C_w, with x being the image of the base-vertex $*$ of C_w. Let γ be the path from x to x in Γ labeled w (so that, by construction, γ passes through every topological edge of Γ at least once). Then check whether the labeling map $\Gamma \to R_N$ is a covering, that is, whether it is true that every vertex of Γ has degree $2N$. If $\Gamma \to R_N$ is not a covering, the graph (Γ, x) belongs to $\mathcal{G}_0(w)$. If $\Gamma \to R_N$ is a covering, check, using the algorithm from part (3) of Proposition 4·5, whether or not $\gamma \in \pi_1(\Gamma, x)$ is simple in the finite rank free group $\pi_1(\Gamma, x)$. If $\gamma \in \pi_1(\Gamma, x)$ is simple in $\pi_1(\Gamma, x)$, we conclude that the graph (Γ, x) belongs to $\mathcal{G}_0(w)$, and $\gamma \in \pi_1(\Gamma, x)$ is not simple in $\pi_1(\Gamma, x)$, we conclude that he graph (Γ, x) does not belong to $\mathcal{G}_0(w)$. Performing this procedure for each partition of VC_w as a disjoint union of nonempty subsets produces the finite set $\mathcal{G}_0(w)$. Proposition 4·13 then implies that $d_{simp}(g) = d_{simp}(w) = \min\{\#V\Gamma : (\Gamma, x) \in \mathcal{G}_0(w)\}$.

The algorithm for computing $d_{prim}(g) = d_{prim}(w)$ is similar. We first find all the graphs in $\mathcal{G}(w)$ as follows. Enumerate all partitions of VC_w as a disjoint union of nonempty subsets. For each such partition $V_1, \dots V_m$ collapse each V_i, $i = 1, \dots, m$, to a vertex and then fold the result to get a principal quotient (Γ, x) of C_w. There is a path γ from x to x in Γ labeled w. Then check, using the algorithm from part (2) of Proposition 4·5,, whether or not $\gamma \in \pi_1(\Gamma, x)$ is primitive in the free group $\pi_1(\Gamma, x)$. If yes, we conclude that $(\Gamma, x) \in \mathcal{G}(w)$ and if not, we conclude that $(\Gamma, x) \notin \mathcal{G}(w)$. This procedure algorithmically computes the set $\mathcal{G}(w)$.

Proposition 4·13 then implies that $d_{prim}(g) = d_{prim}(w) = \min\{\#V\Gamma : (\Gamma, x) \in \mathcal{G}(w)\}$. Thus part (1) of the theorem is verified.

Part (2) now follows directly from part (1) using the definitions of $f_{prim}(n)$ and $f_{simp}(n)$.

□

REMARK 4·15. *The complexity of the algorithms for computing $d_{simp}(g)$ and $d_{prim}(g)$ given in part (1) of Theorem 4·14 is super-exponential in $n = ||g||_A$. The reason is that enumerating all principal quotients of the graph C_w requires listing all partitions of the n-element set VC_w. The* Bell number *B_n, which is the number of all partitions of an n-element set, grows roughly as n^n.*

4·2. *Algorithmic computability of $d_{fill}(g)$*

We now want to give an algorithm for computing $d_{fill}(g)$. Computationally this algorithm is not nearly as nice as the algorithms for computing $d_{simp}(g)$ and $d_{prim}(g)$ described above.

We briefly recall here some definitions and notations related to the Outer space. We refer the reader to [**28**, **33**, **59**] for more details. Let $N \geq 2$ be an integer. The *unprojectivized Outer space* cv_N is the set of all of F_N-equivariant isometry classes of $\mathbb{R}$-trees T such that T is equipped with a free discrete minimal isometric action of F_N. The *projectivized Outer space* CV_N consists of the projective classes $[T]$ where $T \in \mathrm{cv}_N$. Here for $T \in \mathrm{cv}_N$ the projective class $[T]$ of T is the set of all $cT \in \mathrm{cv}_N$ where $c \in \mathbb{R}_{\geq 0}$. Here cT is the same set as T, with the same action of F_N, but where the metric on cT is the multiple by c of the metric on T.

The space $\overline{\mathrm{cv}}_N$ is the closure of cv_N in the equivariant Gromov-Hausdorff convergence topology. It is known that $\overline{\mathrm{cv}}_N$ consists precisely of all of F_N-equivariant isometry classes of $\mathbb{R}$-trees T such that T is equipped with a free minimal very small isometric action of F_N. The space $\overline{\mathrm{CV}}_N$ is the set of all projective classes $[T]$ where $T \in \mathrm{cv}_N$ (the projective

class $[T]$ for $T \in \overline{\text{cv}}_N$ is defined similarly as above, as the set of f all $cT \in \overline{\text{cv}}_N$ where $c \in \mathbb{R}_{\geq 0}$).

The following result provides a useful characterization of filling elements:

PROPOSITION 4·16. *Let $1 \neq g \in F_N$. Then $g \in F_N$ is filling if and only if $Stab_{Out(F_N)}([g])$ is finite.*

Proof. Solie [54, Lemma 2.42, Lemma 2.44] proves that if $g \in F_N$ is non-filling then $Stab_{Out(F_N)}([g])$ is infinite. Thus the the "if" direction of Proposition 4·16 holds.

Let us now prove the "only if" direction. Suppose $Stab_{Out(F_N)}([g])$ is infinite. Choose a basepoint $[T_0] \in \text{CV}_N$. Since the action of $Out(F_N)$ on CV_N is properly discontinuous and since $\overline{\text{CV}}_N$ is compact, it follows that there exist an infinite sequence of distinct elements $\varphi_n \in Stab_{Out(F_N)}([g])$ and a point $[T] \in \overline{\text{CV}}_N - \text{CV}_N$ such that $\lim_{n\to\infty}[T_0]\varphi^n = [T]$. Then for some sequence of scalars $c_n \geq 0$ with $c_n \to 0$ as $n \to \infty$ we have $\lim_{n\to\infty} c_n T_0 \varphi_n = T$ in $\overline{\text{cv}}_N$. Since $\varphi_n([g]) = [g]$, it follows that $||g||_T = \lim_{n\to\infty} c_n ||\varphi_n(g)||_{T_0} = 0$. Then by Proposition 2·14 the element g is not filling in F_N, as required. □

PROPOSITION 4·17. *Let $F_N = F(A)$, where $N \geq 2$ and where $A = \{a_1, \ldots, a_N\}$ is a free basis of F_N. Then there exists an algorithm that, given a nontrivial element $g \in F_N$, decides whether or not g is filling in F_N.*

Proof.

Let $g \in F_N = F(A)$ be a nontrivial freely reduced word. By a result of McCool [42] the group $Stab_{Out(F_N)}([g])$ is finitely generated and, moreover, we can algorithmically compute a finite generating set $Y = \{\psi_1, \ldots, \psi_k\}$ of $Stab_{Out(F_N)}([g])$.

In view of Proposition 4·16 we next need to determine if $H := \langle Y \rangle \leq Out(F_N)$ is finite. Wang and Zimmermann [60] prove that for $N > 2$, the maximum order of a finite subgroup of $Out(F_N)$ is $2^N N!$. Also, the word problem for $Out(F_N)$ is solvable (even solvable in polynomial time [50]). Thus we then start building the Cayley graph $Cay(H; Y)$ of H with respect to Y. Using solvability of the word problem in $Out(F_N)$, for any finite k we can algorithmically construct the ball $B(k)$ of radius k cantered at identity in $Cay(H; Y)$. We construct the balls $B(2^N N!)$ and $B(1 + 2^N N!)$. By the result of Wang and Zimmermann mentioned above, the group H is finite if and only if $B(2^N N!) = B(1 + 2^N N!)$.

Thus we can algorithmically decide whether or not $Stab_{Out(F_N)}([g])$ is finite, and hence, by Proposition 4·16, whether or not g is filling in F_N.

□

THEOREM 4·18. *Let $F_N = F(A)$, where $N \geq 2$ and where $A = \{a_1, \ldots, a_N\}$ is a free basis of F_N. Then:*

(i) There exists an algorithm that, given $1 \neq g \in F_N$, computes $d_{fill}(g)$.

(ii) There exists an algorithm that, for every $n \geq 1$ computes $f_{fill}(n)$

Proof. Part (2) follows directly from part (1) and from the definition of $f_{fill}(n)$. Thus we only need to establish part (1).

Given $g \in F_N$, let w be the cyclically reduced form of g. Let C_w and its principle quotients be as in Definitions 4·8, 4·7. Enumerate all principle quotients of C_w as $\{\Gamma_1, \ldots, \Gamma_k\}$. For each Γ_i with $1 \leq i \leq k$, two possibilities arise:

Case (i) (Γ_i *is not a finite cover of* R_N): In this case, we call Γ_i a "success". In this

case we can complete Γ_i to a finite cover Γ_i' of R_N and now $\pi_1(\Gamma_i)$ is a free factor of $\pi_1(\Gamma_i')$. Hence w is simple in in the subgroup represented by Γ_i' i.e. w is not filling in the subgroup represented by Γ_i'.

Case (ii) (Γ_i *is a finite cover of* R_N): In this case there is a closed loop γ_i in Γ_i with label w. We then use the algorithm from Proposition 4·17 to check whether γ_i is filling in $\pi_1(\Gamma_i)$. If γ_i is not filling in $\pi_1(\Gamma_i)$, and we call Γ_i a success.

Finally, observe that $d_{fill}(g) = \min\{V\Gamma_i \,|\, \Gamma_i \text{ is a "success"}\}$ where this equality is established in a manner similar to that in Proposition 4·13. Thus part (1) of the theorem is proved. □

5. *Special words and finite covers*

The main goal of this section is to find a suitable sufficient condition implying that a given freely reduced word is filling in a given finite index subgroup of F_N represented by a finite cover of the rose R_N. Similarly we find a suitable sufficient condition implying that a given freely reduced word is not simple in a subgroup of F_N represented by a given finite cover of the rose R_N.

These goals are accomplished by constructing "simplicity blocking" and "filling forcing" words in F_N of controlled length, provided by Proposition 5·12 and Proposition 5·7 below. Since the proofs of these Propositions are somewhat technical, we first illustrate the idea of their proof by obtaining a related simpler statement, given in Lemma 5·1 below. The proof of Lemma 5·1 is due to Yuliy Baryshnikov. We then adapt the idea of this proof to obtain Proposition 5·7 and Proposition 5·12.

LEMMA 5·1. *Let $N \geq 2$. Then there exists a constant $c_0 = c_0(N) > 0$ with the following property. Let $(\Gamma, *)$ be a connected d-fold cover of the N-rose R_N, where $d \geq 1$. Then there exists a freely reduced word $v = v(\Gamma)$ with $|v| \leq c_0 d^2$ such that for every vertex $x \in V\Gamma$ the path $p(x, v)$ from x labeled by v in Γ passes through every topological edge of Γ at least once.*

Proof. The graph Γ is a connected $2N$-regular graph with d vertices and Nd topological edges. We can view Γ as a directed graph where the directed edges are labeled by elements of A (and without using A^{-1}). Then Γ is a connected directed graph where the in-degree of every vertex is equal to N, which is also equal to the out-degree of every vertex. Hence there exists an Euler circuit in Γ beginning and ending at $*$ consisting of edges labeled by elements of A that transverses each topological edge exactly once. Let v_1 be the label of this Euler circuit. Then v_1 is freely reduced and no a_i^{-1} occurs in v_1 for $i = 1, \ldots, N$. Enumerate the vertices as $V\Gamma = \{x_1, x_2, \ldots, x_d\}$ with $* = x_1$. Starting at the vertex x_2 follow a path p_1 with label v_1. Denote the terminal vertex of p_1 by z_1. Let p_1' be an Euler circuit in Γ starting and ending at z_1 and consisting only of edges labeled by elements of A. Let v_2 be the label of this path p_1'. Note that since we only consider positively labeled edges, the path $p_2 = p_1 p_1'$ is reduced and its label $v_1 v_2$ is a positive (and hence freely reduced) word over A. We now inductively define a positive word v_{i+1} over A given that the positive words $v_1, \ldots, v_i$ where $i \in \{1, \ldots, d-1\}$ have already been defined. Starting at vertex x_{i+1} we follow a path p_i with label $v_1 \ldots v_i$. Denote the terminal vertex of the path p_i by z_i. Let p_i' be an Euler circuit at z_i that transverses every positively labeled edge exactly once. Let v_{i+1} be the label of this path p_i'. We define our word $v := v_1 v_2 \ldots v_d$. Since following a path with label $v_1 \ldots v_i$ at any vertex v_i already

passes through every topological edge of Γ at least once, so does following a path with label v. Since each $|v_i| = Nd$ for $1 = 1, \dots, d$, we have that $|v| = Nd^2$. $\square$

5·1. *Simplicity blocking words and finite covers*

In the above proof the concatenation argument always produces reduced edge-paths because we only deal with edges and paths labeled by positive words over A. By contrast, in proving Proposition 5·7 simple concatenation does not always work as it may result in paths that are not reduced. Also, instead of paths labeled by v passing through every edge of Γ, we need to ensure a more complicated condition which implies that all paths labeled by v in Γ pass through a certain "simplicity-blocking" path $\alpha(\Gamma, T)$, which is defined below.

DEFINITION 5·2. *Let Γ be a finite connected folded A-graph, let $T \subseteq \Gamma$ be a maximal tree in Γ and let S_T be the corresponding basis of $\pi_1(\Gamma, *)$. Let $u = y_1 \dots y_n$ be a nontrivial freely reduced word over $S_T^{\pm 1}$. Thus each y_i corresponds to an edge $e_i \in E(\Gamma - T)$. We define a reduced path $\delta(u)$ in Γ as*

$$\delta(u) := [*, o(e_1)]_T e_1 [t(e_1), o(e_2)]_T e_2 \dots \dots e_n [t(e_n), *]_T.$$

Note that if $d = \#V\Gamma$ then T has $\le d-1$ topological edges and hence $|\delta(u)| \le n + (n+1)(d-1) = nd + d - 1 = d(n+1) - 1$.

DEFINITION 5·3. *Let $(\Gamma, *)$ be a finite folded core graph with a base-vertex $*$. Let $T \subseteq \Gamma$ be a maximal subtree in Γ. Let $S_T = \{b_1, \dots, b_r\}$ be the basis of $\pi_1(\Gamma, *)$ dual to T.*

Define a reduced edge-path $\alpha(\Gamma, T)$ from $*$ to $*$ in Γ as

$$\alpha(\Gamma, T) := \delta(b_r^2 b_1^2 \dots b_r^2).$$

REMARK 5·4. *Note that the path $\alpha(\Gamma, T)$ is reduced and represents the element $b_r^2 b_1^2 \dots b_r^2$ in $\pi_1(\Gamma, *)$. The following proposition demonstrates the "simplicity-blocking" property of $\alpha(\Gamma, T)$. The word $b_r^2 b_1^2 \dots b_r^2$ has length $2r+2$ and hence $|\alpha(\Gamma, T)| \le d(2r+3) - 1$ where $d = \#V\Gamma$. In particular, if Γ is a d-fold cover of the rose R_N, then $r = d(N-1)+1$ and*

$$|\alpha(\Gamma, T)| \le d(2d(N-1)+3) - 1 \le 2d^2(N-1) + 4d.$$

PROPOSITION 5·5. *Let Γ be as in Definition 5·3 with T a maximal tree in Γ. Let S_T and $\alpha(\Gamma, T)$ be as before. Let $\gamma \in \pi_1(\Gamma, *)$ be such that γ is represented by a cyclically reduced circuit in Γ containing $\alpha(\Gamma, T)$ as a subpath. Then γ is not simple in $\pi_1(\Gamma, *)$.*

Proof. We first use Proposition 2·5 to rewrite γ as a cyclically reduced word w in $S_T = \{b_1, \dots, b_r\}$. Then the occurrence of $\alpha(\Gamma, T)$ in γ produces an occurrence of the reduced word $b_r^2 b_1^2 \dots b_r^2$ in w. Hence, by Corollary 2·18, in this case γ is not simple in $F(b_1, \dots, b_r) = \pi_1(\Gamma, *)$.

$\square$

Note that Definition 2·3 of a core graph implies that if Γ is a finite connected core graph, then Γ does not have any degree-1 vertices.

LEMMA 5·6. *Let Γ be a finite connected core graph with d vertices. Suppose that $\pi_1(\Gamma)$ has rank ≥ 2. Then for any any two edges $e_1, e_2 \in E(\Gamma)$, there exists a reduced path $p(e_1, e_2)$ starting at e_1, ending at e_2, and with $|p(e_1, e_2)| \le 3d$.*

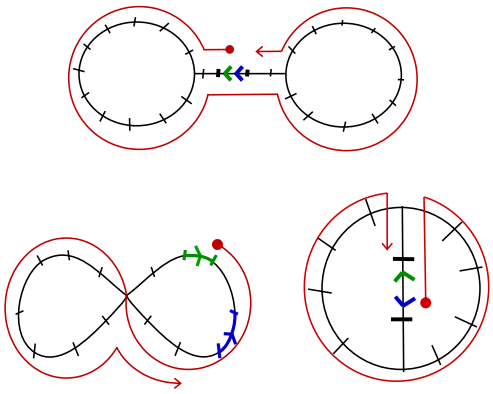

Fig. 2. Proof by picture for Lemma 5·6

Proof. Pick a graph $\Gamma' \subseteq \Gamma$ such that Γ' is a finite, connected, core graph with $\pi_1(\Gamma')$ of rank 2 and $e_1, e_2 \in E\Gamma'$. Then there are precisely three possibilities for Γ'. It can be the wedge of two circles, or a theta-graph (a circle with a line segment joining two points on the circle), or a barbell graph (two circles attached to two ends of a line segment). We will show that the result holds for the graph Γ', and hence holds for our graph Γ. Our proof is essentially going to be a proof by picture for each of these three cases. In Figure 2, green edges (or arrows) indicate e_1 and blue edges (or arrows) indicate e_2. We indicate the path $p(e_1, e_2)$ in red with the $\bullet$ representing the starting point of $p(e_1, e_2)$ and the $\rightarrow$ representing the direction. The path $p(e_1, e_2)$ starts at $o(e_1)$ and ends at $t(v_1)$. We call a "cusp" any vertex of Γ' of degree ≥ 3 in Γ'. The idea behind finding this path $p(e_1, e_2)$ is always to travel along e_1 to the nearest cusp. Then if one is required to go back on the same path one has already been on to get to e_2, one instead travels along a disjoint loop at the cusp. Now one can go back to e_2 and the path $p(e_1, e_2)$ will be reduced. If after traveling from e_1 to the cusp one can get to e_2 without compromising the fact that the path $p(e_1, e_2)$ is reduced, then one simply goes to e_2 and the path $p(e_1, e_2)$ so obtained is reduced. From Figure 2 we see that the result holds.

□

The following fact plays a key role in the proof of Theorem 1·2:

PROPOSITION 5·7. *Let $N \geq 2$. Then there exists a constant $c_0 = c_0(N) > 0$ with the following property. Let $(\Gamma, *)$ be a connected d-fold cover of the N-rose R_N, where $d \geq 1$ and let $T \subseteq \Gamma$ be a maximal subtree of Γ. Then there exists a freely reduced word $v = v(\Gamma, T)$ with $|v| \leq c_0 d^3$ such that for every vertex $x \in V\Gamma$ the path $p(x, v)$ from x labeled by v in Γ contains $\alpha(\Gamma, T)$ as a subpath.*

Proof. Let us begin by enumerating the vertices of $V\Gamma = x_1, x_2, \ldots, x_d$. Let $H \leq F_N$ be the subgroup of index d that is represented by $(\Gamma, *)$. Let T be a maximal tree in $(\Gamma, *)$ and let $S_T = \{b_1, b_2, \ldots, b_r\}$ be the corresponding basis of $\pi_1(\Gamma, *)$. By Remark 5·4, we have $|\alpha(\Gamma, T)| \leq 2d^2(N-1) + 4d$.

Let e be the first edge of the path $\alpha(\Gamma, T)$. Starting at the vertex $x_1 \in V\Gamma$, there exists a unique path $[x_1, *]_T$ of length $\leq d-1$ with terminal edge e_1 (say). Lemma 5·6 then

gives us a reduced path $p(e_1, e) = e_1 p' e$ of length $\leq 3d$. Let the word v_1 be the label of the path $p_1 = [x_1, *]_T p' \alpha(\Gamma, T)$. Note that $|v_1| = |p_1| \leq 2d^2(N-1) + 8d - 3$.

Starting at the vertex x_2 we follow a path p_1' that has label v_1. Let e_2 be the terminal edge of the path p_1'. Then from Lemma 5·6, the path $p(e_2, e) = e_2 p_1'' e$ is reduced with $|p(e_2, e)| \leq 3d$, and hence $|p_1''| \leq 3d - 2$. Let the word v_2 be the label of the path $p_2 = p_1'' \alpha(\Gamma, T)$. Now the path $p_1' p_2 = p_1' p_1'' \alpha(\Gamma, T)$ is reduced. Notice that $|v_2| = |p_2| \leq 2d^2(N-1) + 7d - 1$. We now define inductively a sequence of words and paths as follows: Suppose we have already defined our words $v_1, v_2, \ldots, v_{i-1}$ which are respectively the labels of reduced paths $p_1, \ldots, p_{i-1}$. Starting at vertex x_i we follow the path p_{i-1}' labeled by the word $v_1 v_2 \ldots v_{i-1}$. Let e_i be the terminal edge of the path p_{i-1}'. Then the path $p(e_i, e) = e_i p_{i-1}'' e$ is reduced with $|p_{i-1}''| \leq 3d - 2$. Let the word v_i be the label of the reduced path $p_i = p_{i-1}'' \alpha(\Gamma, T)$. Now the path $p_{i-1}' p_i = p_{i-1}' p_{i-1}'' \alpha(\Gamma, T)$ is reduced. Let the word $v := v_1 v_2 \ldots v_d$. Then notice that at any vertex x_i with $1 \leq i \leq d$, the path $p_{i-1}' p_i$ is a reduced path labeled by $v_1 \ldots v_i$ that already contains the subpath $\alpha(\Gamma, T)$. Thus for $i = 1, \ldots, d$ the path starting at x_i labeled by the word $v_1 \ldots v_d$ also contains the subpath $\alpha(\Gamma, T)$. Since for all $2 \leq i \leq d$, $|v_i| \leq 2d^2(N-1) + 7d - 1$, we have that $|v| \leq 2d^3(N-1) + 7d^2 - 2 \leq (2N+5)d^3$. Thus with $c_0 = 2N + 5$, we are done.

□

If Γ is a finite connected graph and $T \subseteq \Gamma$ is a maximal subtree, then, following the conventions of Bass-Serre theory, we denote

$$\pi_1(\Gamma, T) := \langle E\Gamma | e\bar{e} = 1 \text{ for all } e \in E\Gamma, e = 1 \text{ for all } e \in ET \rangle.$$

If Γ is equipped with an orientation, then $\pi_1(\Gamma, T)$ is canonically isomorphic to the free group $F(E_+\Gamma - E_+T)$. Note also that $\pi_1(\Gamma, T)$ is isomorphic to the fundamental group of the quotient space Γ/T (where T is collapsed to a point).

The freely reduced word $v = v(\Gamma, T)$ in $F(A)$ can be viewed as a "simplicity blocking" word for the elements of the fundamental group of a d-fold cover Γ of R_N.

COROLLARY 5·8. *Let $N \geq 2$ and let $c_0 = c_0(N) > 0$ be the constant provided by Proposition 5·7.*

Let $d \geq 1$, let Γ be a connected d-fold cover of the N-rose R_N and let $T \subseteq \Gamma$ be a maximal tree in Γ. Let $* \in V\Gamma$, let γ be a reduced edge-path from $*$ to $*$ in Γ and let γ' be the cyclically reduced form of the path γ (so that the label of γ' is a cyclically reduced word in $F(A)$). Suppose that the label of γ' contains as a subword the word $v = v(\Gamma, T)$ with $|v| \leq c_0 d^3$ provided by Proposition 5·7.

Then $\gamma \in \pi_1(\Gamma, *)$ does not belong to a proper free factor of $\pi_1(\Gamma, *)$.

Proof. From definitions $\gamma \in \pi_1(\Gamma, *)$. Using the tree T we can obtain a free basis $S_T = \{b_1, \ldots, b_r\}$ of $\pi_1(\Gamma, T)$. Then Proposition 2·5 tells us how to rewrite γ in terms of the basis S_T, both as freely reduced word and as a cyclically reduced word. Let $\alpha(\Gamma, T)$ be as before. Then for the label of γ' to contain the word v, we must have that the cyclically reduced form of γ' in terms of S_T contains $b_r^2 b_1^2 \ldots b_r^2$ as a subword. Now from Corollary 2·18 we know that γ' is not simple in $\pi_1(\Gamma, T)$. Finally from Lemma 3·6 γ is not simple in $\pi_i(\Gamma, T)$, that is, $\gamma \in \pi_1(\Gamma, *)$ does not belong to a proper free factor of $\pi_1(\Gamma, *)$.

□

5·2. *Filling forcing words and finite covers*

To proceed further we will once again adapt the idea of proof of Lemma 5·1 to produce a "filling-forcing" path $\beta(\Gamma, T)$ of controlled length.

CONVENTION 5·9. *If $F(B)$ is a free group with $|B| = r \geq 2$, then the total number of freely reduced words of length 3 in $F(B)$ are $L = 2r(2r-1)^2$. Let $\{u_1, \dots, u_L\}$ be the set of all freely reduced words of length 3 in $B^{\pm 1}$. Define a freely reduced word $u_B := u_1 y_1 u_2 y_2 u_3 y_3 \dots y_{L-1} u_L$ where each y_i is either the empty word, or $y_i \in B^{\pm 1}$. Namely, whenever the concatenation $u_j u_{j+1}$ is reduced to begin with, we define y_j to be the empty word. If this concatenation is not reduced, then we can always choose $y_j \in B^{\pm 1}$ so that $u_j y_j u_{j+1}$ is reduced in $F(B)$. Note that $|u|_B \leq 3L + L - 1 = 4L - 1$.*

We now define the path $\beta(\Gamma, T)$ as follows:

DEFINITION 5·10. *Let $(\Gamma, *)$ be a finite connected folded core graph with a base-vertex $*$. Let $T \subseteq \Gamma$ be a maximal subtree in Γ with $E_+(\Gamma - T) = \{e_1, \dots, e_r\}$, and let $S_T = \{b_1, \dots, b_r\}$ be the basis of $\pi_1(\Gamma, *)$ dual to T. We put*

$$\beta(\Gamma, T) := \delta(u_{S_T}).$$

Thus $\beta(\Gamma, T)$ is a reduced edge-path from $*$ to $*$ in Γ representing the element u_{S_T} in $\pi_1(\Gamma, *)$. Recall that u_{S_T} has length $4L - 1 = 8r(2r-1)^2 - 1$. Therefore

$$|\beta(\Gamma, T)| \leq 8rd(2r-1)^2 - 1$$

where $d = \#V\Gamma$. In particular, if Γ is a d-fold cover of R_N then $r = d(N-1) + 1$ and

$$|\beta(\Gamma, T)| \leq 8d(d(N-1)+1)(2d(N-1)+1)^2 - 1 \leq 500 d^4 N^3.$$

The following proposition demonstrates the a "filling-forcing" property of the path $\beta(\Gamma, T)$

PROPOSITION 5·11. *Let Γ be as in Definition 5·10 with T a maximal tree. Let S_T and $\beta(\Gamma, T)$ be as before. Let $\gamma \in \pi_1(\Gamma, *)$ be such that γ is represented by a cyclically reduced circuit in Γ containing $\beta(\Gamma, T)$ as a subpath. Then γ is filling in $\pi_1(\Gamma, *)$.*

Proof. We first use Proposition 2·5 to rewrite γ as a cyclically reduced word w in $S_T = \{b_1, \dots, b_r\}$. Then the occurrence of $\beta(\Gamma, T)$ in γ produces an occurrence of the reduced word $u_1 y_1 u_2 y_2 u_3 y_3 \dots y_{l-1} u_l$ in w. Since every reduced word of length 3 now occurs in w, by Proposition 2·19 γ is filling in $F(b_1, \dots, b_r) = \pi_1(\Gamma, *)$.

□

We are now in a position to prove a key proposition that is used in the proofs of Theorem 1·1 and Theorem 1·2:

PROPOSITION 5·12. *Let $N \geq 2$. Then there exists a constant $c_1 = c_1(N) > 0$ with the following property. Let $(\Gamma, *)$ be a connected d-fold cover of the N-rose R_N, where $d \geq 1$ and let $T \subseteq \Gamma$ be a maximal subtree of Γ. Then there exists a freely reduced word $w = w(\Gamma, T)$ with $|w| \leq c_1 d^5$ such that for every vertex $x \in V\Gamma$ the path $p(x, w)$ from x labeled by w in Γ contains $\beta(\Gamma, T)$ as a subpath.*

Proof. Let us begin by enumerating the vertices of $V\Gamma = \{x_1, x_2, \dots, x_d\}$. Let $H \leq F_N$ be the subgroup of index d that is represented by $(\Gamma, *)$. We have seen above that $|\beta(\Gamma, T)| \leq 500 d^4 N^3$.

Let e be the first edge of the path $\beta(\Gamma, T)$. Starting at the vertex $x_1 \in V\Gamma$, there exists a unique path $[x_1, *]_T$ of length $\leq d-1$ with terminal edge e_1 (say). Lemma 5·6 then gives us a reduced path $p(e_1, e) = e_1 p' e$ of length $\leq 3d$. Let the word w_1 be the label of the path $p_1 = [x_1, *]_T p' \beta(\Gamma, T)$. Note that $|w_1| \leq 500d^4N^3 + 3d$.

Starting at the vertex x_2 we follow a path p_1' that has label w_1. Let e_2 be the terminal edge of the path p_1'. Then from Lemma 5·6, there is a reduced path $p(e_2, e) = e_2 p_1'' e$ with $|p(e_2, e)| \leq 3d$ and $|p_1''| \leq 3d-2$. Let the word w_2 be the label of the path $p_2 = p_1'' \beta(\Gamma, T)$. Thus $|w_2| = |p_2| \leq 500d^4N^3 + 3d$.

Now the path $p_1' p_2 = p_1' p_1'' \beta(\Gamma, T)$ is reduced, starts at x_2, ends in $\beta(\Gamma, T)$, has label $w_1 w_2$ and has length

$$|p_1' p_2| = |w_1 w_2| \leq 2(500d^4N^3 + 3d).$$

We proceed inductively as follows. For $2 \leq i \leq d$ suppose that we have already constructed freely reduced words $w_1, \ldots, w_{i-1} \in F_N = F(A)$ of length $|w_j| \leq 500d^4N^3 + 3d$ such that the word $w_1 \ldots w_{i-1}$ is freely reduced and such that reading $w_1 \ldots w_{i-1}$ from the vertex x_{i-1} gives a reduced path in Γ ending in $\beta(\Gamma, T)$.

Starting at vertex x_i we follow the path p_{i-1}' labeled by the word $w_1 w_2 \ldots w_{i-1}$. Let e_i be the terminal edge of the path p_{i-1}'. Then the path $p(e_i, e) = e_i p_{i-1}'' e$ is reduced with $|p_{i-1}''| \leq 3d-2$. Let the word w_i be the label of the reduced path $p_i = p_{i-1}'' \beta(\Gamma, T)$. We again have $|w_i| \leq 500d^4N^3 + 3d$. Now the path $p_{i-1}' p_i = p_{i-1}' p_{i-1}'' \beta(\Gamma, T)$ is reduced, starts with x_i and ends in $\beta(\Gamma, T)$, completing the inductive step.

Finally let $w := w_1 w_2 \ldots w_d$. Then w is freely reduced, has $|w| \leq 500d^5N^3 + 3d^2 \leq 1000N^3d^5$. By construction w has the property that for $i = 1, \ldots, d$ reading w from x_i gives a path in Γ containing $\beta(\Gamma, T)$ as a subpath. We put $w(\Gamma, T) := w$ and $c_1 = 1000N^3$. The conclusion of the proposition now holds. □

The freely reduced word $w = w(\Gamma, T)$ in $F(A)$ can be viewed as a "filling forcing" word for the elements of the fundamental group of a d-fold cover Γ of R_N.

6. *A lower bound for the non-filling index function*

REMARK 6·1. *Let $F_N = F(a_1, \ldots, a_N)$ be free of rank $N \geq 2$, as before. It is well-known (see, for example, [39]) that for an integer $d \geq 1$ there are $\leq (d!)^N$ subgroups of index d in F_N. Indeed, every subgroup of index d in F_N can be uniquely represented by a finite connected folded $2N$-regular A-graph on vertices $1, \ldots, d$, where 1 is viewed as a base-vertex. Every such graph Γ is uniquely specified by choosing an ordered N-tuple of permutations in S_d. Indeed, if $\sigma_1, \ldots, \sigma_N \in S_d$, we construct Γ with $V\Gamma = \{1, \ldots, d\}$ by putting an edge from j to $\sigma_i(j)$ labeled by a_i for $1 \leq i \leq N$, and $1 \leq j \leq d$.*

Thus indeed F_N has $\leq (d!)^N$ subgroups of index d and it has $\leq d(d!)^N$ subgroups of index $\leq d$.

THEOREM 6·2. *Let $N \geq 2$ and let $F_N = F(A)$ where $A = a_1, \ldots, a_N$. Then there exists a constant $c > 0$ and an integer $M \geq 1$ such that for all $n \geq M$ we have*

$$f_{prim}(n) \geq f_{simp}(n) \geq f_{fill}(n) \geq c\frac{\log n}{\log\log n}.$$

Proof.

Let $d \geq 1$ be an integer. Denote $m(d) = m := d(d)!^N$. Enumerate all the subgroups of F_N of index $\leq d$ as $H_1, \ldots, H_m$ (we do allow repetitions in this list since the actual

number of such distinct subgroups is $< m(d)$. Let $\Gamma_1, \ldots, \Gamma_m$ be the base-pointed finite covers of the rose R_N representing the subgroups $H_1, \ldots, H_m$.

For $i = 1, \ldots, m$ let $w_i \in F(A)$ be the freely reduced "filling forcing" word with $|w_i| \le c_1 d^5$ corresponding to Γ_i as provided by Proposition 5·12. We can now construct a freely reduced and cyclically reduced word

$$z_d := w_1 u_1 w_2 u_2 \ldots u_{m-1} w_m u_m$$

where each u_i is either the empty word or $u_i \in \{a_1, \ldots, a_N\}^{\pm 1}$. Then

$$||z_d|| \le c_1 m d^5 = c_1 d^6 (d!)^N.$$

We claim that $d_{fill}(z_d) > d$. Indeed, suppose not, that is suppose that $d_{fill}(z_d) \le d$. Then there exists $1 \le i \le m$ such that $z_d \in H_i$ and that z_d is a non-filling element of $H_i = \pi_1(\Gamma_i, *)$. Let γ be the path in Γ_i from $*$ to $*$ labeled by z_d. By Proposition 5·12 the fact that z_d is cyclically reduced and contains w_i as subword implies that γ contains the path $\beta(\Gamma_i, T)$ as a subword. Hence, by Proposition 5·11, γ is a filling element in $\pi_1(\Gamma_i, *)$, yielding a contradiction. Thus indeed $d_{fill}(z_d) > d$.

Now for $d \ge 1$ let $n_d := c_1 d^6 (d!)^N$. We also put $n_0 = 1$. Then for every integer $d \ge 0$ we have $f_{fill}(n_d) > d$. By Stirling's formula, there is $C > 0$ such that for all sufficiently large $d \ge 1$ we have

$$d \ge C \frac{\log n_d}{\log \log n_d} \qquad (\dagger)$$

Similarly, using a standard calculus argument we see that for all sufficiently large d we have

$$\frac{\log(n_{d-1})}{\log \log(n_{d-1})} \ge \frac{1}{2} \frac{\log(n_d)}{\log \log(n_d)}. \qquad (\ddagger)$$

Let $d_0 \ge 2$ be such that for all $d \ge d_0$ the inequalities (†) and (‡) hold and that the function the function $\dfrac{\log x}{\log \log x}$ is monotone increasing on the interval $[n_{d_0 - 1}, \infty)$.

Now let $n \ge n_{d_0+1}$ be an arbitrary integer. There exists a unique $d \ge 0$ such that $n_{d-1} < n \le n_d$. Since $f_{fill}(n)$ is a non-decreasing function, we get that $f_{fill}(n) \ge f_{fill}(n_{d-1}) > d - 1$ and $d - 1 \ge d_0$.

Then

$$f_{fill}(n) \ge f_{fill}(n_{d-1}) > d - 1 \ge C \frac{\log(n_{d-1})}{\log \log(n_{d-1})} \ge \frac{C}{2} \frac{\log(n_d)}{\log \log(n_d)} \ge \frac{C}{2} \frac{\log n}{\log \log n},$$

and the conclusion of the theorem follows. □

7. *Non-backtracking simple random walk on F_N*

Convention 7·1. *In this paper we use the standard big-O and big-Θ conventions. For functions $f, g : \mathbb{N} \to \mathbb{R}$ we write $f = O(g)$ (or sometimes $f(n) = O(g(n))$) if there exist an integer $n_0 \ge 1$ and a constant $C > 0$ such that for all integers $n \ge n_0$ we have $|f(n)| \le C|g(n)|$. For such f, g we write $f = \Theta(g)$ if $f = O(g)$ and $g = O(f)$. In particular, if $f(n) = O(g(n))$ and $\lim_{n\to\infty} g(n) = 0$ then $\lim_{n\to\infty} f(n) = 0$.*

Recall that we set for the free group $F_N = F(A) = F(a_1, \ldots, a_N)$ (where $N \ge 2$) a distinguished free basis $A = \{a_1, \ldots, a_N\}$. Put $\Upsilon = A \cup A^{-1}$.

DEFINITION 7·2. *We consider the following finite-state Markov chain $\mathcal{X}$. The set of states for $\mathcal{X}$ is Υ. For $x, y \in \Upsilon$, the transition probability $P_{x,y}$ from x to y is defined as:*

$$P_{x,y} := \begin{cases} \frac{1}{2N-1}, & \text{if } y \neq x^{-1} \\ 0, & \text{if } y = x^{-1} \end{cases}.$$

Let M be the transition matrix of $\mathcal{X}$. That is, M is a $2N \times 2N$ matrix with columns and rows indexed by Υ where for $x, y \in \Upsilon$ the entry $m_{x,y}$ in M is equal to 1 if $y \neq x^{-1}$ and is equal to 0 if $y = x^{-1}$.

We summarize the following elementary properties of $\mathcal{X}$, which easily follow from the definitions:

LEMMA 7·3. *Let $N \geq 2$ and $\mathcal{X}$ be as in Definition 7·2. Then:*

(i) $\mathcal{X}$ is an irreducible aperiodic finite-state Markov chain.

(ii) The uniform probability distribution μ_1 on Υ is stationary for $\mathcal{X}$.

(iii) The matrix M is an irreducible aperiodic nonnegative matrix with the Perron-Frobenius eigenvalue $\lambda = 2N - 1$.

Proof. For any $x, y \in \Upsilon$ there exists $z \in \Upsilon$ such that xzy is a freely reduced word. Hence $P_{x,z}P_{z,y} > 0$, which means that $\mathcal{X}$ is an irreducible Markov chain. The fact that for every $x \in \Upsilon$, we have $P_{x,x} > 0$ implies that $\mathcal{X}$ is aperiodic. Thus (1) is verified.

Part (2) easily follows from the definition of $\mathcal{X}$ by direct verification.

Part (1) implies that M is an irreducible aperiodic nonnegative matrix. Therefore, by the basic Perron-Frobenius theory, the spectral radius $\lambda := \max\{|\lambda_*| : \lambda_* \in \mathbb{C} \text{ is an eigenvalue of } M\}$ is a positive real number which is itself an eigenvalue of M called the Perron-Frobenius eigenvalue of M. It is also known that λ admits an eigenvector with strictly positive coordinates, and that any other eigenvalue of M admitting such an eigenvector is equal to λ. It is easy to see from the definition of M that for the vector v with all entries equal to 1 we have $Mv = (2N-1)v$. Therefore $\lambda = 2N - 1$, as claimed. □

Let $\Omega = \Upsilon^{\mathbb{N}} = \{\omega = x_1, x_2, \ldots | x_i \in \Upsilon\}$. We put the discrete topology on Υ and the product topology on Ω so that Ω becomes a compact Hausdorff space. For every finite word $\sigma \in \Upsilon^*$ the *cylinder* $Cyl(\sigma) \subseteq \Omega$ consists of all sequences $\omega \in \Omega$ with σ as the initial segment. For each $\sigma \in \Upsilon^*$ the set $Cyl(\sigma)$ is compact and open in Ω and the sets $\{Cyl(\sigma) | \sigma \in \Upsilon^*\}$ provide a basis for the product topology on Ω.

By using the uniform distribution μ_1 on Υ as the initial distribution for $\mathcal{X}$, the Markov chain $\mathcal{X}$ defines a Borel probability measure μ on Ω via the standard convolution formula:

For $\sigma = x_1 \ldots x_n \in \Upsilon^*$,

$$\mu(Cyl(\sigma)) = \mu_1(x_1)P_{x_1,x_2} \ldots P_{x_{n-1},x_n}.$$

Note that the support of μ is exactly ∂F_N, that is, the set of all semi-infinite freely reduced words $\omega = x_1, x_2, \ldots$ over Υ.

CONVENTION 7·4. *For $\sigma \in \Upsilon^*$ (where Υ^* is the set of all words over the alphabet Υ) we denote $\mu(\sigma) := \mu(Cyl(\sigma))$. Also, for the remainder of this section we denote $\lambda := 2N - 1$.*

The following is a direct corollary of the definitions:

LEMMA 7·5. *Let $\sigma = x_1 \dots x_n \in \Upsilon^*$, where $n \geq 1$. Then*

$$\mu(\sigma) = \begin{cases} \frac{1}{2N(2N-1)^{n-1}}, & \text{if } \sigma \text{ is freely reduced,} \\ 0, & \text{if } \sigma \text{ is not freely reduced.} \end{cases}$$

NOTATION 7·6. *Let $v, w \in \Upsilon^*$. We denote by $\langle v, w \rangle$ the number of times the word v occurs as a subword of w.*

For $n \geq 1$ let $S(n)$ be the set of all freely reduced words of length n in Υ^* (so that $\#(S(n)) = 2N(2N-1)^{n-1} = \frac{2N}{2N-1}\lambda^n$), and let μ_n be the uniform probability distribution on $S(n)$.

The following statement is a special case, when applied to $\mathcal{X}$, of Proposition 3.13 in [20] (which in turn is based on the proof of the main result of Dinwoodie [23]).

PROPOSITION 7·7. *Let $\varepsilon > 0$ and $0 < \ell < 1$. Then there exist constants $C_1 > 1$ and $C_2 > 0$, depending on ε and ℓ, with the following property. Let $n \geq 1$ and $\sigma \in \Upsilon^*$ be a freely reduced word be such that $|\sigma| = \ell \log_\lambda n = \ell \log n / \log \lambda$. Then for $w_n \in S(n)$ we have*

$$1 - P_{\mu_n}(|\langle \sigma, w_n \rangle - n\mu(\sigma)| < n^{\varepsilon + (1-\ell)/2}) = O(C_1^{-n^{C_2}}),$$

and therefore, since $\lambda = 2N - 1$ and $\mu(\sigma) = \frac{2N-1}{2N}\lambda^{-|\sigma|} = \frac{2N-1}{2N} n^{-\ell}$,

$$1 - P_{\mu_n}(\left|\langle \sigma, w_n \rangle - \frac{2N-1}{2N} n^{1-\ell}\right| < n^{\varepsilon + (1-\ell)/2}) = O(C_1^{-n^{C_2}}),$$

COROLLARY 7·8. *Let $\varepsilon > 0$ and $0 < \ell < 1$. Let constants $C_1 = C_1(\varepsilon, \ell) > 1$ and $C_2 = C_2(\varepsilon, \ell) > 0$ be the constants provided by Proposition 7·7.*

(i)Let $n \geq 1$ and let $E_n \subseteq S(n)$ consist of those $w_n \in S(n)$ such that for every freely reduced $\sigma \in \Upsilon^*$ with $|\sigma| = \ell \log_\lambda n = \ell \log n / \log \lambda$ we have

$$\left|\langle \sigma, w_n \rangle - \frac{2N-1}{2N} n^{1-\ell}\right| < n^{\varepsilon + (1-\ell)/2},$$

Then

$$1 - P_{\mu_n}(w_n \in E_n) = O\left(n^\ell C_1^{-n^{C_2}}\right).$$

(ii)Suppose that $\varepsilon > 0, 0 < \ell < 1$ are chosen so that $\ell < 1 - 2\varepsilon$, and thus $1 - \ell > \varepsilon + (1-\ell)/2$. Let $H_n \subseteq S(n)$ consist of all $w_n \in S(n)$ such that for every freely reduced σ with $|\sigma| = \ell \log_\lambda n$ we have

$$\langle \sigma, w_n \rangle \geq \frac{2N-1}{4N} n^{1-\ell}.$$

Then for $n \geq n_0$ we have

$$1 - P_{\mu_n}(w_n \in H_n) = O(n^\ell C_1^{-n^{C_2}}).$$

Proof. For every freely reduced σ with $|\sigma| = \ell \log_\lambda n$ let $E'_{n,\sigma}$ consist of all $w_n \in S(n)$ such that $|\langle \sigma, w_n \rangle - n\mu(\sigma)| \geq n^{\varepsilon + (1-\ell)/2}$. Thus, by Proposition 7·7, for every such σ we have $P_{\mu_n}(E'_{n,\sigma}) = O(C_1^{-n^{C_2}})$.

Suppose $w_n \notin E_n$. Then there exists freely reduced $\sigma \in \Upsilon^*$ with $|\sigma| = \ell \log_\lambda n$ such that $w_n \in E'_{n,\sigma}$. Since there are $O(n^\ell)$ freely reduced words σ with $|\sigma| = \ell \log_\lambda n$, it follows that $P_{\mu_n}(S(n) \setminus E_n) = O\left(n^\ell C_1^{-n^{C_2}}\right)$. Hence $1 - P_{\mu_n}(E_n) = O\left(n^\ell C_1^{-n^{C_2}}\right)$, as required, and part (1) of Corollary 7·8 is verified.

Part (2) now directly follows from part (1). □

NOTATION 7·9. *For a freely reduced word $w \in \Upsilon^*$ let $\iota(w)$ be the maximal initial segment of w such that $(\iota(w))^{-1}$ is a terminal segment of w. Let $\tilde{w}$ be the word obtained by removing the initial and terminal segments of w of length $|\iota(w)|$. Thus $\tilde{w}$ is the cyclically reduced form of w.*

The following facts are well-known and easy to check by a direct counting argument; see [2] for details:

LEMMA 7·10. *The following hold:*

(i)For every $0 < \varepsilon_0 < 1$ there exists $C_0 > 1$ such that for $w_n \in S(n)$

$$1 - P_{\mu_n}(|\iota(w_n)| \leq \varepsilon_0 n) = O(C_0^{-n}).$$

(ii)There is $C > 1$ such that for $w_n \in S(n)$

$$1 - P_{\mu_n}(w_n \text{ is not a proper power in } F_N) = O(C^{-n}).$$

8. *Bounding below the simplicity and the non-filling index for random elements*

Recall that for a nontrivial element $g \in F_N$ we denote by $d_{simp}(g)$ the smallest $d \geq 1$ such that there exists a subgroup $H \leq F_N$ with $[F_N : H] \leq d$ such that $g \in H$ and, moreover, that g belongs to a proper free factor of H. Similarly, for $g \in F_N - \{1\}$ we denote by $d_{prim}(g)$ the smallest $d \geq 1$ such that there exists a subgroup $H \leq F_N$ with $[F_N : H] \leq d$ such that $g \in H$ and, moreover, that g is primitive in H. As we have seen, for every $g \in F_N - \{1\}$ we have $d_{simp}(g) \leq d_{prim}(g) \leq ||g||_A$, where $A = \{a_1, \dots, a_n\}$ is a free basis of F_N. Recall that for $n \geq 1$ we denote by μ_n the uniform probability distribution on the sphere $S(n) \subseteq F(A) = F_N$.

For the remainder of the paper we adopt the convention that whenever we mention a word of length $t \geq 0$ where t is not necessarily an integer, we actually mean a word of length $\lfloor t \rfloor$.

We can now prove Theorem 1·2 from the Introduction:

THEOREM 1·2. *Let $N \geq 2$ and let $F_N = F(A)$ where $A = \{a_1, \dots, a_N\}$.*

Then there exist constants $c(N) > 0$, $D_1(N) > 1$, $1 > D_2(N) > 0$, such that for $n \geq 1$ and for a freely reduced word $w_n \in F(A)$ of length n chosen uniformly at random from the sphere $S(n)$ of radius n in $F(A)$ we have

$$1 - P_{\mu_n}\left(d_{simp}(w_n) \geq c \log^{1/3} n\right) = O\left((D_1)^{-n^{D_2}}\right)$$

and

$$1 - P_{\mu_n}\left(d_{fill}(w_n) \geq c \log^{1/5} n\right) = O\left((D_1)^{-n^{D_2}}\right)$$

so that

$$\lim_{n\to\infty} P_{\mu_n}\left(d_{simp}(w_n) \geq c \log^{1/3} n\right) = 1$$

and

$$\lim_{n\to\infty} P_{\mu_n}\left(d_{fill}(w_n) \geq c \log^{1/5} n\right) = 1$$

Proof.

Choose $\varepsilon > 0$ and $0 < \ell < 1$ such that $\ell < 1 - 2\varepsilon$ (for concreteness we can take $\ell = 1/2$

and $\varepsilon = 1/5$). Thus $1 - \ell > \varepsilon + (1-\ell)/2 > 0$. Let $n_0 \geq 1$ be such that for all $n \geq n_0$ we have

$$\frac{2N-1}{4N}(0.99n)^{1-\ell} \geq (0.99n)^{\varepsilon+(1-\ell)/2} \geq 1.$$

(The choice of the number 0.99 here is essentially arbitrary, and the argument would also work if 0.99 is replaced by any other number sufficiently close to 1.) Let $C_1 > 1$ and $C_2 > 0$ be the constants provided by Corollary 7·8. Note that we can assume that $0 < C_2 < 1$ since decreasing C_2 preserves the validity of the conclusion of Corollary 7·8.

For $w_n \in S(n)$ denote by w'_n the subword of w_n obtained by removing the initial and terminal segments of length $0.005n$ from w_n. Then $|w'_n| = 0.99n$ so that $w'_n \in S(0.99n)$. Since the uniform distribution on $A^{\pm 1}$ is stationary for the Markov chain $\mathcal{X}$, it follows that under the map $S(n) \to S(0.99n)$, $w_n \mapsto w'_n$ the uniform distribution μ_n on $S(n)$ projects to the uniform distribution $\mu_{0.99n}$ on $S(0.99n)$.

Let H'_n be the event that for $w_n \in S(n)$ the word w'_n satisfies the property that for every freely reduced word $\sigma \in F(A)$ with $|\sigma| = \ell \log_\lambda(0.99n)$ we have

$$\langle \sigma, w'_n \rangle \geq 1.$$

Since for $n \geq n_0$ we have $\frac{2N-1}{4N}(0.99n)^{1-\ell} \geq (0.99n)^{\varepsilon+(1-\ell)/2} \geq 1$, Corollary 7·8 implies that

$$1 - P_{\mu_n}(H'_n) = O((0.99n)^\ell C_1^{-(0.99n)^{C_2}}) = O\left(n^\ell (C_1)^{-0.99^{C_2} n^{C_2}}\right) = O\left((C'_1)^{-n^{C'_2}}\right),$$

where $C'_1 = (C_1+1)/2$ and $C'_2 = C_2/2$ (for the last inequality we use the fact that $0 < C_2 < 1$). Note that $C'_1 > 1$ and $1 > C'_2 > 0$.

Let $Q_n \subseteq S(n)$ be the event that for $w_n \in S(n)$ we have $\iota(w_n) \leq 0.001n$. Lemma 7·10 implies that $P_{\mu_n}(Q_n) \geq 1 - O(C_0^{-n})$ for some constant $C_0 > 1$. Now let H''_n be the set of all $w_n \in H'_n$ such that $\iota(w_n) \leq 0.001n$, that is, $H''_n = H'_n \cap Q_n$.

Then

$$1 - P_{\mu_n}(H''_n) = O\left((C'_1)^{-n^{C'_2}}\right) - O(C_0^{-n}) \geq_{n\to\infty} = O\left((D_1)^{-n^{D_2}}\right),$$

where $D_1 = \min\{C_0, C'_1\}$ and $D_2 = \min\{C'_2, 1\} = C'_2$, so that $D_1 > 1$ and $1 > D_2 > 0$.

We choose $c > 0$ such that $c_0 c^3 \leq \frac{\ell}{2\log(2N-1)}$, where $c_0 > 0$ is the constant provided by Proposition 5·7.

Let $n \geq n_0$ and let $w_n \in S(n)$ be such that $w_n \in H''_n$.

Since $\iota(w_n) \leq 0.001n$ and since w'_n is the subword of w_n obtained by removing the initial and terminal segments of length $0.005n$ from w_n, it follows that w'_n is a subword of the cyclically reduced form $\tilde{w}_n$ of w_n.

Let $d = d_{simp}(w_n) = d_{simp}(\tilde{w}_n)$. We claim that $d \geq c \log^{1/3} n$.

Indeed, suppose not, that is, suppose that $d < c \log^{1/3} n$. Let (Γ, x_0) be a d-fold cover of the N-rose $(R_N, *)$ such that $\tilde{w}_n$ lifts to a loop γ_n from x_0 to x_0 in Γ such that γ_n belongs to a proper free factor of $\pi_1(\Gamma, x_0)$. Note that since $\tilde{w}_n$ is cyclically reduced, the closed path γ_n is also cyclically reduced.

Let T be a maximal subtree of Γ and let $v = v(\Gamma, T)$ be the freely reduced word in $F(A)$ with $|v| \leq c_0 d^3$ provided by Proposition 5·7. Thus $|v| \leq c_0 d^3 \leq c_0 c^3 \log n$.

By definition of H''_n, the fact that $w_n \in H''_n$ implies that the word w'_n contains as subwords all freely reduced words in $F(A)$ of length

$$\ell \log_\lambda(0.99n) = \frac{\ell}{\log(2N-1)}(\log n - |\log 0.99|)$$

There is $n_1 \geq n_0$ such that for all $n \geq n_1$ we have

$$\frac{\ell}{\log(2N-1)}(\log n - |\log 0.99|) \geq \frac{\ell}{2\log(2N-1)} \log n.$$

Hence for $n \geq n_1$ the word w_n' contains as subwords all freely reduced words of length $\frac{\ell}{2\log(2N-1)} \log n$. Since $|v| \leq c_0 c^3 \log n \leq \frac{\ell}{2\log(2N-1)} \log n$, it follows that w_n' contains v as a subword.

Recall that w_n' is a subword of the cyclically reduced form $\tilde{w}_n$ of w_n. Therefore, by Proposition 5·7, the path γ_n in Γ, labeled by $\tilde{w}_n$, contains $\alpha(\Gamma, T)$ as a subpath. Hence, by Corollary 5·8, γ_n does not belong to a proper free factor of $\pi_1(\Gamma, x_0)$, yielding a contradiction. Thus $d = d_{simp}(w_n) \geq c \log^{1/3} n$, as claimed.

We have verified that for every $w_n \in H_n''$, where $n \geq n_1$, we have $d_{simp}(w_n) \geq c \log^{1/3} n$, and we also know that

$$1 - P_{\mu_n}(H_n'') = O\left((D_1)^{-n^{D_2}}\right).$$

The conclusion of Theorem 1·2 regarding $d_{simp}(w_n)$ is established.

The proof of the conclusion of Theorem 1·2 regarding $d_{fill}(w_n)$ is identical, with Proposition 5·11 and Proposition 5·12 used instead of Proposition 5·7 and Corollary 5·8. We leave the details to the reader.

□

9. *Untangling closed geodesics on hyperbolic surfaces*

9·1. *Lower bounds for* $\deg_{\Sigma,\rho}$ *and* $f_{\Sigma,\rho}$ *for hyperbolic surfaces.*

We need the following well-known fact:

LEMMA 9·1. *Let S be a compact connected surface with $b \geq 2$ boundary components such that $\pi_1(S)$ is free of rank ≥ 2. Let γ be an essential simple closed curve (possible peripheral) on S and let $x \in S$ be a base-point for S. Then the element of $\pi_1(S, x)$ given by any loop at x corresponding to γ belongs to a proper free factor of $\pi_1(S, x)$.*

Proof.

Without loss of generality we may assume that $x \in \gamma$.

By assumption, we have $\pi_1(S, x) = F_m$ with $m \geq 2$. Since S has $b \geq 2$ boundary components, it follows that every boundary component (when realized as a loop at x) represents a primitive element of F_m.

Let γ be an essential simple closed curve on S. If γ is peripheral, then γ is a primitive element of F_m and thus belongs to a proper free factor of F_m.

Suppose now that γ is non-peripheral. Then cutting S along γ yields a nontrivial splitting of $F_m = \pi_1(S)$ as an amalgamated product (if γ is separating) or as an HNN-extension (if γ is non-separating) over $\langle\gamma\rangle = \mathbb{Z}$. Suppose that γ is separating, and it cuts S into two compact surfaces S_1 and S_2 with $S_1 \cap S_2 = \gamma$ and $S_1 \cup S_2 = S$, each of $\pi_1(S_1), \pi_1(S_2)$ is free of rank ≥ 2. Thus $F_m = \pi_1(S, x) = \pi_1(S_1, x) *_\gamma \pi_1(S_2, x)$. The fact that $b \geq 2$ means that at least one of S_1, S_2 has ≥ 2 boundary components. Assume for concreteness that S_1 has ≥ 2 boundary components. Then γ is primitive in $\pi_1(S_1, x)$. Thus we can find a free basis $a_1, \ldots, a_m$ of $\pi_1(S_1, x)$ such that $m \geq 2$ and $\gamma = a_m$. Also choose a free basis $b_1, \ldots, b_k$ of $\pi_1(S_2, x)$, where $k \geq 2$. Let $v \in F(b_1, \ldots, b_k) = \pi_1(S_2, x)$ be the freely reduced word equal to γ in $\pi_1(S_2, x)$. Then the above splitting of $\pi_1(S, x)$ can be written as $\pi_1(S, x) = F(a_1, \ldots, a_m) *_{a_m = v} F(b_1, \ldots, b_k)$. By eliminating the generator

a_m from this presentation, we see that $\pi_1(S,x) = F(a_1,\dots,a_{m-1},b_1,\dots,b_k)$. Thus $\gamma = v(b_1,\dots,b_k)$ belongs to a proper free factor $F(b_1,\dots,b_k)$ of $\pi_1(S,x)$, as required. The case where γ is non-separating is similar, and we leave the details to the reader.

Note that there is a general result (see, for example, [4, Lemma 4.1] and [55, Proposition 5.1]) which says that whenever the free group F_N (with $N \geq 2$) splits nontrivially as an amalgamated free product or an HNN-extension over a maximal infinite cyclic subgroup $\langle g \rangle$, then g belongs to a proper free factor of F_N. □

The following proposition relates the degree function $\deg_{\Sigma,\rho}(\gamma)$ for curves in hyperbolic surfaces discussed in the Introduction, with the simplicity index d_{simp} in free groups for curves contained in suitable subsurfaces:

PROPOSITION 9·2. *Let (Σ,ρ) be a compact connected hyperbolic surface with (possibly empty) geodesic boundary. Let $\Sigma_1 \subseteq \Sigma$ be a compact connected subsurface with ≥ 3 boundary components, each of which is a geodesic in (Σ,ρ). Let $x \in \Sigma_1$ be a base-point. Then for every nontrivial element $g \in \pi_1(\Sigma_1,x)$ represented by a closed geodesic γ_g on Σ we have*

$$\deg_{\Sigma,\rho}(\gamma_g) \geq d_{simp}(g;\pi_1(\Sigma_1,x)).$$

Proof. By assumption $\pi_1(\Sigma_1,x) \cong F_m$ is free of rank $m \geq 2$. The fact that Σ_1 is a subsurface of Σ with geodesic boundary implies that if $g \in \pi_1(\Sigma_1,x)$ is a nontrivial element, then the shortest geodesic in Σ in the free homotopy class of g is contained in Σ_1. Indeed, the universal cover $X := \widetilde{(\Sigma_1,x)}$ is a convex $\pi_1(\Sigma_1,*)$-invariant subset of $\widetilde{(\Sigma,x)} = \mathbb{H}^2$. Therefore for every nontrivial element $g \in \pi_1(\Sigma_1,x)$ the axis $Axis(g)$ of g in $\mathbb{H}^2$ is contained in X. The image of $Axis(g)$ in Σ is the unique closed geodesic in the free homotopy class of g; the fact that $Axis(g) \subseteq X$ implies that this closed geodesic is contained in Σ_1, as claimed.

Now let $1 \neq g \in \pi_1(\Sigma_1,x)$ and γ_g be as in the assumptions of the proposition. Thus γ_g is contained in Σ_1.

Let $d = \deg_{\Sigma,\rho}(\gamma_g)$. Let $p : \widehat{\Sigma} \to \Sigma$ be a d-fold cover of Σ such that γ_g lifts to a simple closed geodesic $\widehat{\gamma}_g$ in $\widehat{\Sigma}$. Let $\widehat{\Sigma}_1 \subseteq \widehat{\Sigma}$ be the connected component of the full preimage $p^{-1}(\Sigma_1)$ of Σ_1 containing $\widehat{\gamma}_g$. Then $p : \widehat{\Sigma}_1 \to \Sigma_1$ is a d'-fold cover of Σ_1 with $d' \leq d$. Pick a base-point $x' \in \widehat{\Sigma}_1$ such that $p(x') = x$.

The cover $p : (\widehat{\Sigma}_1,x') \to (\Sigma_1,x)$ corresponds to a subgroup $H \leq \pi_1(\Sigma_1,x)$ of index d', such that $p_\#(\pi_1(\widehat{\Sigma}_1,x)) = H$, and that $p_\#$ maps $\pi_1(\widehat{\Sigma}_1,x')$ isomorphically to H.

Since $\widehat{\Sigma}_1$ is a cover of Σ_1, the surface $\widehat{\Sigma}_1$ has ≥ 2 boundary components and $\pi_1(\widehat{\Sigma}_1)$ is free of rank ≥ 2. By Lemma 9·1, the fact that $\widehat{\gamma}_g$ is an essential simple closed curve on $\widehat{\Sigma}_1$ implies that $\widehat{\gamma}_g$ corresponds an element $w \in \pi_1(\widehat{\Sigma}_1,x')$ which belongs to a proper free factor of $\pi_1(\widehat{\Sigma}_1,x')$. Since $p(\widehat{\gamma}_g) = \gamma_g$, we have $p_\#(w) = g \in H$. Since $p_\#$ maps $\pi_1(\widehat{\Sigma}_1,x')$ isomorphically to H, we conclude that g belongs to a proper free factor of H. Thus $H \leq \pi_1(\Sigma_1,x)$, $[\pi_1(\Sigma_1,x) : H] = d'$ and g belongs to a proper free factor of H. Therefore $d' \geq d_{simp}(g;\pi_1(\Sigma_1,x))$. Therefore

$$\deg_{\Sigma,\rho}(\gamma_g) = d \geq d' \geq d_{simp}(g;\pi_1(\Sigma_1,x)),$$

as required.

□

THEOREM 9·3. *Let Σ be a compact connected surface with a hyperbolic structure ρ and*

with (possibly empty) geodesic boundary. Let $\Sigma_1 \subseteq \Sigma$ be a compact connected subsurface with ≥ 3 boundary components, each of which is a geodesic in (Σ, ρ). Let $x \in \Sigma_1$ and let A be a free basis of $\pi_1(\Sigma_1, x)$.

Let $w_n \in F(A) = \pi_1(\Sigma_1, x)$ be a freely reduced word of length n over $A^{\pm 1}$ generated by a simple non-backtracking random walk on $F(A) = \pi_1(\Sigma_1, x)$. Let γ_n be the closed geodesic on (Σ, ρ) in the free homotopy class of w_n.

Then there exist constants $c > 0, K' \geq 1$ such that

$$\lim_{n\to\infty} Pr(\deg_{\Sigma,\rho}(\gamma_n) \geq c \log^{1/3} n) = 1$$

and such that with probability tending to 1 as $n \to \infty$ we have that $w_n \in \pi_1(\Sigma, x)$ is not a proper power and that $n/K' \leq \ell_\rho(\gamma_n) \leq K'n$.

Proof. As we have seen in the proof of Proposition 9·2, the fact that Σ_1 is a subsurface of Σ with geodesic boundary implies that if $g \in \pi_1(\Sigma_1, *)$ is a nontrivial element, then the shortest geodesic in Σ in the free homotopy class of g is contained in Σ_1.

By Theorem 1·2 and Lemma 7·10, there exist an integer $n_0 \geq 1$ such that for $n \geq n_0$, with probability tending to 1 as $n \to \infty$ we have that w_n is not a proper power in $F(A)$, that $0.99n \leq ||w_n||_A \leq n = |w_n|_A$ and $d_{simp}(w_n; F(A)) \geq c \log^{1/3} n$, where $c = c(A) > 0$ is the constant provided by Theorem 1·2 for the free group $F_m = F(A)$.

Proposition 9·2 now implies that with probability tending to 1 as $n \to \infty$ we have

$$\deg_{\Sigma,\rho}(\gamma_n) \geq d_{simp}(w_n; F(A)) \geq c \log^{1/3} n.$$

Finally, the fact that Σ_1 has geodesic boundary in (Σ, ρ) also implies that there exists a constant $K \geq 1$ such that for every nontrivial element $g \in \pi_1(\Sigma_1, x)$ represented by a closed geodesic γ on (Σ, ρ) we have $||g||_A/K \leq \ell_\rho(\gamma) \leq K||g||_A$. Since with probability tending to 1 as $n \to \infty$ we have $0.99n \leq ||w_n||_A \leq n = |w_n|_A$, it follows that for all sufficiently large n with with probability tending to 1 as $n \to \infty$ we have $0.99n/K \leq \ell_\rho(\gamma_n) \leq Kn$, as required.

□

REMARK 9·4. *Theorem 9·3 directly implies (e.g. by taking Σ_1 to be a suitable pair-of-pants subsurface) that if (Σ, ρ) is a compact connected hyperbolic surface of genus ≥ 2 with (possibly empty) geodesic boundary, then there exists $c' = c'(\Sigma) > 0$ such that for every $L \geq sys(\rho)$ we have $f_\rho(L) \geq c'(\log L)^{1/3}$.*

9·2. *Lower bounds for $\deg^{fill}_{\Sigma,\rho}$ and $f^{fill}_{\Sigma,\rho}$ for hyperbolic surfaces.*

Our results about the behavior of d_{fill} in free groups can also be used to obtain information about $\deg^{fill}_{\Sigma,\rho}$ for compact hyperbolic surfaces.

LEMMA 9·5. *Let (Σ, ρ) be a compact connected hyperbolic surface with $b \geq 1$ geodesic boundary components. Then the following hold:*

(i) If γ is a non-filling closed geodesic on (Σ, ρ) , then γ represents a non-filling element of the free group $\pi_1(\Sigma)$.

(ii) For any closed geodesic γ on (Σ, ρ) we have $\deg^{fill}_{\Sigma,\rho}(\gamma) \geq d_{fill}(\gamma, \pi_1(\Sigma))$.

Proof. To see that (1) holds, let γ be a non-filling closed geodesic on (Σ, ρ). Then either γ is contained in a proper compact connected subsurface Σ_1 of (Σ, ρ) with geodesic boundary or $\Sigma - \gamma$ is a union of disks, peripheral annuli and non-peripheral annuli

$A_1, \dots, A_k$ (where $k \geq 1$). In the latter case the simple closed geodesics $\alpha_1, \dots, \alpha_k$ homotopic to the core curves of $A_1, \dots, A_k$ are disjoint from γ, and we put Σ_1 to be the surface obtained by cutting Σ open along the curves $\alpha_1, \dots, \alpha_k$.

In either case, cutting Σ open along the boundary of Σ_1 provides a nontrivial graph-of-groups decomposition of $\pi_1(\Sigma)$ with maximal cyclic edge groups and such that γ belongs to a vertex group of this decomposition. Hence γ is non-filling in $\pi_1(\Sigma)$. Thus (1) holds.

For (2), let γ be a closed geodesic on (Σ, ρ). Let $d = \deg^{fill}_{\Sigma,\rho}(\gamma)$ and let $\widehat{\Sigma} \to \Sigma$ be a degree-d cover such that γ lifts to a closed non-filling geodesic $\widehat{\gamma}$ on $\widehat{\Sigma}$. This cover corresponds to a subgroup $H = \pi_1(\Sigma_1) \leq \pi_1(\Sigma)$ of index d containing the element γ. The fact that $\widehat{\gamma}$ is a non-filling curve in Σ_1 implies, by part (1) of this lemma, that γ is a non-filling element of $H = \pi_1(\Sigma_1)$. Therefore, by definition, $d_{fill}(\gamma, \pi_1(\Sigma)) \leq d = \deg^{fill}_{\Sigma,\rho}(\gamma)$, as required. □

THEOREM 9·6. *Let (Σ, ρ) be a compact connected hyperbolic surface with $b \geq 1$ geodesic boundary components. Then there exists $C' > 0$ such that for all sufficiently large L we have*

$$f^{fill}_{\Sigma,\rho}(L) \geq C' \frac{\log L}{\log \log L}.$$

Proof. Let $\pi_1(\Sigma) = F_N = F(A)$ where $A = \{a_1, \dots, a_N\}$ with $N \geq 2$. The universal cover $X = (\tilde{\Sigma}, \tilde{\rho})$ is a convex $\pi_1(\Sigma)$-invariant subset of $\mathbb{H}^2$. Therefore the orbit map $F(A) \to \mathbb{H}^2$, $w \mapsto w*$ (where $* \in \mathbb{H}^2$ is some basepoint) is a $\pi_1(\Sigma)$-equivariant quasi-isometry. Hence there exists $K \geq 1$ such that for every closed geodesic γ on (Σ, ρ) representing an element $w \in \pi_1(\Sigma)$ we have $||w||_A / K \leq \ell_\rho(\gamma) \leq K||w||_A$.

By Theorem 6·2 there exists a sequence of nontrivial cyclically reduced elements $w_n \in F(A)$ such that $||w_n||_A = n$ and that for all sufficiently large n we have

$$d_{fill}(w_n, F(A)) \geq C \frac{\log n}{\log \log n},$$

where $C > 0$ is the constant provided by Theorem 6·2. By Lemma 9·5, it follows that for all sufficiently large n we have

$$\deg^{fill}_{\Sigma,\rho}(\gamma) \geq d_{fill}(w_n, F(A)) \geq C \frac{\log n}{\log \log n}.$$

Since $||w||_A / K \leq \ell_\rho(\gamma) \leq K||w||_A$, the statement of the theorem now follows. □

THEOREM 9·7. *Let (Σ, ρ) be a compact connected hyperbolic surface with $b \geq 1$ geodesic boundary components. Let $A = \{a_1, \dots, a_N\}$ be a free basis of $\pi_1(\Sigma, x)$, so that $\pi_1(\Sigma) = F(A)$. Let $w_n \in F(A) = \pi_1(\Sigma, x)$ be a freely reduced word of length n over $A^{\pm 1}$ generated by a simple non-backtracking random walk on $F(A)$. Let γ_n be the closed geodesic on (Σ, ρ) in the free homotopy class of w_n.*

Then there exist constants $c_1 > 0, K_1 \geq 1$ such that

$$\lim_{n \to \infty} Pr(\deg^{fill}_{\Sigma,\rho}(\gamma_n) \geq c_1 \log^{1/5} n) = 1$$

and such that with probability tending to 1 as $n \to \infty$ we have that $w_n \in \pi_1(\Sigma, x)$ is not a proper power and that $n/K_1 \leq \ell_\rho(\gamma_n) \leq K_1 n$.

Proof. The proof is essentially identical to the proof of Theorem 9·3, and we leave the details to the reader. □

9·3. *Degree and index functions based on the geometric intersection number*

Let Σ be a compact connected surface admitting some hyperbolic structure (so that $\pi_1(\Sigma)$ is free of rank ≥ 2). If ρ is a hyperbolic metric on Σ and γ is a closed geodesic with respect to ρ on Σ, we denote by $d_\rho(\gamma)$ the smallest degree of a finite cover of Σ such that γ lifts to a simple closed geodesic in that cover.

We adopt the following conventions regarding the geometric intersection number for curves on surfaces. Let S is a compact surface and $\alpha, \beta : \mathbb{S}^1 \to S$ be homotopically nontrivial closed curves on S. Then the geometric intersection number $i([\alpha], [\beta])$ is defined as the minimum cardinality $|(\alpha_1 \times \beta_1)^{-1}(\Delta)|$ where $\Delta \subseteq S \times S$ is the diagonal and where α_1, β_1 vary over all closed curves in the free homotopy classes $[\alpha], [\beta]$ respectively. It is well-know that if ρ is a hyperbolic structure on S (where we always assume that the boundary curves of S, if any, are geodesic with respect to ρ) and if α, β are distinct closed primitive (i.e. not proper powers) geodesics on S with respect to ρ then $i([\alpha], [\beta]) = |(\alpha \times \beta)^{-1}(\Delta)|$. See [22] for a proof in the case of simple closed geodesics, and see p. 143 in [7] and p. 99 in [6] for the general case.

Denote by $\mathcal{C}_\Sigma$ the set of free homotopy classes of essential closed curves on Σ that are not proper powers in $\pi_1(\Sigma)$. For $[\gamma] \in \mathcal{C}_\Sigma$ denote by $d_\Sigma([\gamma])$ the smallest degree of a finite cover of Σ such that a representative of $[\gamma]$ lifts to a simple closed curve in that cover. Note that if ρ is a hyperbolic metric on Σ, then for every $[\gamma] \in \mathcal{C}_\Sigma$ there exists a unique closed ρ-geodesic $\gamma \in [\gamma]$ and $d_\rho(\gamma) = d_\Sigma([\gamma])$. Moreover, as noted above, in this case the geometric intersection number $i([\gamma], [\gamma])$ is realized by γ.

For an integer $m \geq 1$ we define $f_\Sigma(m)$ as the maximum of $d_\Sigma([\gamma])$ where $[\gamma]$ varies over all elements of $\mathcal{C}_\Sigma$ with $i([\gamma], [\gamma]) \leq m$. Similarly, for $[\gamma] \in \mathcal{C}_\Sigma$ denote by $d_\Sigma^{fill}([\gamma])$ the smallest degree of a finite cover of Σ such that a representative of $[\gamma]$ lifts to a non-filling closed curve in that cover. Then define $f_\Sigma^{fill}(m)$ as the maximum of $d_\Sigma^{fill}([\gamma])$ where $[\gamma]$ varies over all elements of $\mathcal{C}_\Sigma$ with $i([\gamma], [\gamma]) \leq m$. Since simple curves are non-filling, we always have $d_\Sigma([\gamma]) \geq d_\Sigma^{fill}([\gamma])$ and hence $f_\Sigma(m) \geq f_\Sigma^{fill}(m)$.

A result of Basmajian [3, Theorem 1.1] (which also can be derived from the results of Bonahon [7]) states:

PROPOSITION 9·8. *Let (Σ, ρ) be a connected compact hyperbolic surface with a (possibly empty) geodesic boundary. Then there exists a constant $K = K(\Sigma, \rho) \geq 1$ such that for every closed geodesic γ on (Σ, ρ) we have*

$$i([\gamma], [\gamma]) \leq K \ell_\rho(\gamma)^2.$$

Theorem 9·6 can be used to derive a lower bound for f_Σ:

THEOREM 9·9. *Let Σ be a compact connected surface admitting some hyperbolic structure. Then there exist a constant $c = c(\Sigma) > 0$ and an integer $m_0 \geq 1$ such that for all $m \geq m_0$ we have*

$$f_\Sigma(m) \geq f_\Sigma^{fill}(m) \geq c \frac{\log m}{\log \log m}.$$

Proof. Fix a hyperbolic metric ρ on Σ. By Proposition 9·8, there exists a constant $K = K(\rho) > 0$ such that for every $[\gamma] \in \mathcal{C}_\Sigma$ we have $i([\gamma], [\gamma]) \leq K \ell_\rho([\gamma])^2$. Let $C' = C'(\Sigma, \rho) > 0$ be the constant provided by Theorem 9·6. Then Theorem 9·6 implies that there exist a sequence of closed geodesics γ_n on (Σ, ρ) and an integer $n_0 \geq 1$ such that for every $n \geq n_0$ we have $\ell_\rho(\gamma_n) \leq n$ and $d_\Sigma^{fill}([\gamma_n]) \geq C' \frac{\log n}{\log \log n}$. Therefore $i(\gamma_n, \gamma_n) \leq K \ell_\rho(\gamma_n)^2 \leq K n^2$ for all $n \geq n_0$.

Fix an integer $n_1 \geq n_0$ such that for all integers $n \geq n_1$ we have $(n+1)^2 \leq 2n^2$.

Let $m \geq Kn_1^2$ be an integer. Choose an integer $n \geq n_1$ such that $Kn^2 \leq m \leq K(n+1)^2$. Then

$$i([\gamma_n],[\gamma_n]) \leq Kn^2 \leq m \leq K(n+1)^2 \leq 2Kn^2$$

and $n \geq \frac{\sqrt{m}}{\sqrt{2K}}$.

Therefore $i([\gamma_n],[\gamma_n]) \leq m$ and

$$d_\Sigma^{fill}([\gamma_n]) \geq C' \frac{\log n}{\log\log n} \geq C' \frac{\log \frac{\sqrt{m}}{\sqrt{2K}}}{\log\log \frac{\sqrt{m}}{\sqrt{2K}}} = C' \frac{\frac{1}{2}\log m - \log\sqrt{2K}}{\log\left(\frac{1}{2}\log m - \log\sqrt{2K}\right)},$$

and the statement of Theorem 9·9 follows. □

REMARK 9·10.

Note that the linear upper bound for $f_{\Sigma,\rho}(m)$, obtained by Patel [44] does not directly imply any upper bound for $f_\Sigma(m)$. The reason is that on a fixed hyperbolic surface there are arbitrarily long simple closed geodesics (which thus have self-intersection number 0). The lower bound for f_Σ given by Theorem 9·9 was the first bound (upper or lower) known for f_Σ. Subsequent to our paper and in part motivated by it, Aougab, Gaster, Patel and Sapir [1] proved that $f_\Sigma(m) = \Theta(m)$, that is $f_\Sigma(m)$ has precisely linear growth in m.

Appendix A. Estimating the primitivity index function from below by the residual finiteness growth function function

by Khalid Bou-Rabee
City College of the City University of New York

In this appendix we relate the primitivity index function $f_{prim}(n;F_N)$ to the residual finiteness growth function introduced in [8]. Applying deep results of Gady Kozma and Andreas Thom [38] then improves the lower bounds for the primitivity index function to almost linear.

We first recall the residual finiteness growth function. Let G be a finitely generated, residually finite group. The *divisibility function* $\mathrm{D}(g) = \mathrm{D}(g;G)$ is the minimum $[G:H]$ where H varies over all subgroups of finite index in G with $g \notin H$. For a fixed finite generating set $A \subset G$ the *residual finiteness growth function* is $\mathrm{RF}_{G,A}(n) := \max\{\mathrm{D}(g;G) : g \in G, |g|_A \leq n, g \neq 1\}$. Here $|g|_A$ is the word-length of g with respect to the word metric on G corresponding to A. In the case where G is a nonabelian free group F_N with word-length $|\cdot|_A$ given by a free basis A, we simply use this basis and denote the function by $\mathrm{RF}_G(n)$.

Next, we recall the primitivity index function introduced by Gupta and Kapovich above. Fix a free group F_N of finite rank $N \geq 2$ with a free basis $A = \{a_1,\ldots,a_N\}$. The *primitivity index* $d_{prim}(g) = d_{prim}(g;F_N)$ of an element $g \in F_N \setminus \{1\}$ is the minimum $[F_N : H]$ where H varies over all subgroups of finite index in F_N containing g as a primitive element. Recall that the *primitivity index function* is

$$f_{prim}(n;F_N) = f_{prim}(n) := \max\{d_{prim}(g) : g \in G, |g|_A \leq n, g \neq 1, g \text{ is not a proper power }\}.$$

THEOREM A1. *Let $G = F_N$ be a free group of finite rank $N \geq 2$. Then $\mathrm{RF}_G(n) \leq f_{prim}(4n+4)$ for all $n \geq 1$.*

Proof. For each $n \geq 1$ let w_n be an element in F_N with $|w_n|_A \leq n$ such that $\mathrm{D}_G(w_n) = \mathrm{RF}_G(n)$. In the free group F_N commutativity is a transitive relation on the set of all nontrivial elements, and therefore there exists $a \in A$ such that $[w_n, a] \neq 1$. Also, in a free group any two non-commuting elements freely generate a free subgroup of rank two. Thus w_n, a freely generate a free subgroup of rank 2 in F_N, and hence $\gamma_n := [w_n, w_n^a] \neq 1$. (In [13, 15] the property, that for every nontrivial $w \in F_N$ there exists $a \in A$ such that $[w, w^a] \neq 1$, is referred to as F_N being *1-malabelian*). Note that $|[w_n, w_n^a]|_A \leq 4n + 4$. Since γ_n is a nontrivial commutator in F_N, a result of Schützenberger [51] then implies that γ_n is not a proper power in F_N.

Let H be a finite-index subgroup of G with γ_n primitive in H. If $w_n \in H$ and $w_n^a \in H$, then $[w_n, w_n^a] \in [H, H]$, and thus $[w_n, w_n^a]$ cannot be primitive in H. Hence, w_n or w_n^a is not in H. In either case, it follows that $[G : H] \geq D_G(w_n) = \mathrm{RF}_G(n)$. Since H was an arbitrary finite-index subgroup for which $[w_n, w_n^a]$ is primitive, it follows that $\mathrm{RF}_G(n) \leq f_{prim}(4n + 4)$, as desired. □

A result of Kozma and Thom [38] about lower bounds for $\mathrm{RF}_{F_N}(n)$ now directly implies:

COROLLARY A2. *Let $G = F_N$ be free of finite rank $N \geq 2$. There exists a constant $C > 0$ such that for all sufficiently large n we have*

$$f_{prim}(4n + 4) \geq \exp\left(\left(\frac{\log(n)}{C \log\log(n)}\right)^{1/4}\right).$$

If we assume Babai's Conjecture on the diameter of Cayley graphs of permutation groups, then for all sufficiently large n we have $f_{prim}(4n + 4) \geq n^{\frac{1}{C \log\log(n)}}$.

At the time of this writing, for a nonabelian free group G, the best upper and lower bounds for $f_{prim}(n)$ and $\mathrm{RF}_G(n)$ have the same asymptotic behavior. Is it true that $f_{prim}(n)$ and $\mathrm{RF}_G(n)$ have the same asymptotic behavior?

REFERENCES

[1] T. Aougab, J. Gaster, P. Patel, J. Sapir, *Building hyperbolic metrics suited to closed curves and applications to lifting simply*, to appear in Math. Research Letters; arXiv:1603.06303v1, March 2016

[2] G. Arzhantseva and A. Ol'shanskii, *Generality of the class of groups in which subgroups with a lesser number of generators are free*, (Russian) Mat. Zametki **59** (1996), no. 4, 489–496; translation in: Math. Notes **59** (1996), no. 3-4, 350–355

[3] A. Basmajian, *Universal length bounds for non-simple closed geodesics on hyperbolic surfaces.* J. Topol. **6** (2013), no. 2, 513–524

[4] M. Bestvina and M. Feighn, *Outer limits*, preprint, 1994; http://andromeda.rutgers.edu/~feighn/papers/outer.pdf

[5] I. Biringer, K. Bou-Rabee, M. Kassabov, F. Matucci. *Intersection growth in groups*, Transact. Amer. Math. Soc., to appear; arXiv:1309.7993;

[6] F. Bonahon, *Bouts des variétés hyperboliques de dimension 3*, Ann. of Math. (2) **124** (1986), no. 1, 71–158

[7] F. Bonahon, *The geometry of Teichmüller space via geodesic currents*, Invent. Math. **92** (1988), no. 1, 139–162

[8] K. Bou-Rabee, *Quantifying residual finiteness*, J. Algebra **323** (2010), no. 3, 729–737

[9] K. Bou-Rabee and Y. Cornulier, *Systolic growth of linear groups*, Proc. Amer. Math. Soc. **144** (2016), no. 2, 529–533

[10] K. Bou-Rabee, M. Hagen and P. Patel, *Residual finiteness growths of virtually special groups,* Math. Z. **279** (2015), no. 1-2, 297–310

[11] K. Bou-Rabee, and T. Kaletha, *Quantifying residual finiteness of arithmetic groups.* Compos. Math. **148** (2012), no. 3, 907–920

[**12**] K. Bou-Rabee, and D. B. McReynolds *Bertrand's postulate and subgroup growth.* J. Algebra **324** (2010), no. 4, 793–819

[**13**] K. Bou-Rabee, and D. B. McReynolds, *Asymptotic growth and least common multiples in groups.* Bull. Lond. Math. Soc. **43** (2011), no. 6, 1059–1068

[**14**] K. Bou-Rabee, and D. B. McReynolds, *Extremal behavior of divisibility functions*, Geom. Dedicata **175** (2015), 407–415

[**15**] K. Bou-Rabee and D. B. McReynolds, *Characterizing linear groups in terms of growth properties*, Michigan Math. J. **65** (2016), no. 3, 599–611

[**16**] K. Bou-Rabee and A. Myropolska, *Groups with near exponential residual finiteness growth*, Israel J. Math., to appear; arXiv:1509.01372

[**17**] K. Bou-Rabee and B. Seward, *Arbitrarily large residual finiteness growth*, J. Reine Angew. Math. **710** (2016), 199–204

[**18**] K. Bou-Rabee and D. Studenmund, *Full residual finiteness growths of nilpotent groups*, Israel J. Math. **214** (2016), no. 1, 209–233

[**19**] N. V. Buskin, *Efficient separability in free groups.* (Russian) Sibirsk. Mat. Zh. **50** (2009), no. 4, 765–771; translation in Sib. Math. J. **50** (2009), no. 4, 603–608

[**20**] D. Calegari and J. Maher, *Statistics and compression of scl*, Ergodic Theory Dynam. Systems **35** (2015), no. 1, 64–110

[**21**] C. Cashen and J. Manning, *Virtual geometricity is rare*, LMS J. Comput. Math. **18** (2015), no. 1, 444–455

[**22**] A. Casson and S. Bleiler, *Automorphisms of surfaces after Nielsen and Thurston.* London Mathematical Society Student Texts, 9; Cambridge University Press, Cambridge, 1988

[**23**] I. Dinwoodie. *Expectations for nonreversible Markov chains.* J. Math. Anal. Appl. **220** (1998), 585–596

[**24**] S. Dowdall and S. Taylor, *Hyperbolic extensions of free groups*, Geom. Topol., to appear; arXiv:1406.2567

[**25**] B. Farb and D. Margalit, *A primer on mapping class groups.* Princeton Mathematical Series, 49. Princeton University Press, Princeton, NJ, 2012

[**26**] J. Gaster, *Lifting curves simply*, Internat. Math. Res. Notices IMRN **2016** (2016), no. 18, 5559–5568

[**27**] R. Gimadeev and M. Vyalyi, *Identical relations in symmetric groups and separating words with reversible automata,* Computer science theory and applications, Lecture Notes in Comput. Sci., vol. 6072, Springer, Berlin, 2010, pp. 144–155

[**28**] V. Guirardel, *Approximations of stable actions on R-trees,* Comment. Math. Helv., 73(1):89-121, 1998.

[**29**] M. Hall, *Coset representations in free groups*, Trans. Amer. Math. Soc. **67** (1949), 421–432

[**30**] I. Kapovich, *The frequency space of a free group,* Internat. J. Alg. Comput. **15** (2005), no. 5-6, 939–969

[**31**] I. Kapovich, *Currents on free groups,* Topological and Asymptotic Aspects of Group Theory (R. Grigorchuk, M. Mihalik, M. Sapir and Z. Sunik, Editors), AMS Contemporary Mathematics Series, vol. 394, 2006, pp. 149-176

[**32**] I. Kapovich, *Clusters, currents and Whitehead's algorithm*, Experimental Mathematics **16** (2007), no. 1, pp. 67-76

[**33**] I. Kapovich and M. Lustig, *Intersection form, laminations and currents on free groups*, Geom. Funct. Anal. (GAFA) **19** (2010), no. 5, pp. 1426-1467

[**34**] I. Kapovich and A. Myasnikov, *Stallings foldings and the subgroup structure of free groups* , J. Algebra **248** (2002), no 2, pp. 608–668

[**35**] I. Kapovich, P. Schupp, and V. Shpilrain, *Generic properties of Whitehead's algorithm and isomorphism rigidity of random one-relator groups.* Pacific J. Math. **223** (2006), no. 1, 113–140

[**36**] I. Kapovich and C. Pfaff, *A Train track Directed Random Walk on* $Out(F_r)$, Internat. J. Algebra Comput. **25** (2015), no. 5, 745–798.

[**37**] M. Kassabov and F. Matucci, *Bounding the residual finiteness of free groups*, Proc. Amer. Math. Soc. **139** (2011), no. 7, 2281–2286

[**38**] G. Kozma and A. Thom, *Divisibility and Laws in Finite Simple Groups*, to appear in Math. Ann. ; arXiv:1403.2324

[**39**] A. Lubotzky, and D. Segal, *Subgroup growth.* Progress in Mathematics, 212. Birkhäuser Verlag, Basel, 2003

[**40**] R. Lyndon and P. Schupp, *Combinatorial group theory*, Reprint of the 1977 edition. Classics in Mathematics. Springer-Verlag, Berlin, 2001

[**41**] J. Malestein, and A. Putman, *On the self-intersections of curves deep in the lower central series of a surface group.* Geom. Dedicata **149** (2010), 73–84

[**42**] J. McCool, *Some finitely presented subgroups of the automorphism group of a free group,* J. Algebra **35** (1975), 205213

[**43**] A. Myasnikov, and V. Shpilrain, *Automorphic orbits in free groups.* J. Algebra **269** (2003), no. 1, 18–27

[**44**] P. Patel, *On a theorem of Peter Scott*, Proc. Amer. Math. Soc. **142** (2014), no. 8, 2891–2906

[**45**] D. Puder, *Primitive words, free factors and measure preservation.* Israel J. Math. **201** (2014), no. 1, 25–73

[**46**] D. Puder, and O. Parzanchevski, *Measure preserving words are primitive.* J. Amer. Math. Soc. **28** (2015), no. 1, 63–97

[**47**] D. Puder, and C. Wu, *Growth of primitive elements in free groups.* J. Lond. Math. Soc. (2) **90** (2014), no. 1, 89–104

[**48**] I. Rivin, *Geodesics with one self-intersection, and other stories*, Adv. Math. **231** (2012), no. 5, 2391–2412

[**49**] A. Roig, E. Ventura, and P. Weil, *On the complexity of the Whitehead minimization problem.* Internat. J. Algebra Comput. **17** (2007), no. 8, 1611–1634

[**50**] S. Schleimer, *Polynomial-time word problems.* Comment. Math. Helv. **83** (2008), no. 4, 741–765

[**51**] M.-P. Schützenberger, *Sur l'quation* $a^{2+n} = b^{2+m}c^{2+p}$ *dans un groupe libre.* C. R. Acad. Sci. Paris **248** (1959), 2435–2436

[**52**] Peter Scott, *Subgroups of surface groups are almost geometric*, J. London Math. Soc. (2), **17** (1978), no. 3, 555–565

[**53**] Peter Scott, *Correction to: "Subgroups of surface groups are almost geometric"* [J. London Math. Soc. (2) 17 (1978), no. 3, 555–565], J. London Math. Soc. (2) **32** (1985), no. 2, 217–220

[**54**] B. Solie, *Algorithmic and statistical properties of filling elements of a free group, and quantitative residual properties of* Γ*-limit groups*, PhD Thesis, (2011)

[**55**] B. Solie, *Genericity of filling elements* Internat. J. Algebra Comput. **22** (2012), no. 2

[**56**] J. R. Stallings, *Topology of finite graphs.* Invent. Math. **71** (1983), 552-565

[**57**] J. R. Stallings, *Whitehead graphs on handlebodies.* Geometric group theory down under (Canberra, 1996), 317–330, de Gruyter, Berlin, 1999

[**58**] R. Stong, *Diskbusting elements of the free group.* Math. Res. Lett. **4** (1997), no. 2-3, 201–21

[**59**] K. Vogtmann, *On the geometry of outer space.* Bull. Amer. Math. Soc. **52** (2015), no. 1, 27–46

[**60**] S. Wang, and B. Zimmermann, *The maximum order of finite groups of outer automorphisms of free groups,* Math. Z. **216**(1994), no. 1, 83–87

[**61**] J. H. C. Whitehead, *On equivalent sets of elements in a free group*, Ann. of Math. (2) **37**(1936), no. 4, 782–800