On the genericity of Whitehead minimality

Frédérique Bassino, Cyril Nicaud, Pascal Weil

On the genericity of Whitehead minimality*

Frédérique Bassino
Université Paris 13, Sorbonne Paris Cité, LIPN, CNRS, (UMR 7030)
F–93430, Villetaneuse, France. bassino@lipn.univ-paris13.fr

Cyril Nicaud
Université Paris-Est, LIGM, CNRS UMR 8049
F-77454 Marne-la-Vallée, France. nicaud@univ-mlv.fr

Pascal Weil
CNRS, LaBRI, UMR 5800, F-33400 Talence, France. pascal.weil@labri.fr
Univ. Bordeaux, LaBRI, UMR 5800, F-33400 Talence, France

March 6, 2014

Abstract

We show that a finitely generated subgroup of a free group, chosen uniformly at random, is strictly Whitehead minimal with overwhelming probability. Whitehead minimality is one of the key elements of the solution of the orbit problem in free groups. The proofs strongly rely on combinatorial tools, notably those of analytic combinatorics. The result we prove actually depends implicitly on the choice of a distribution on finitely generated subgroups, and we establish it for the two distributions which appear in the literature on random subgroups.

1 Introduction

The problem we consider in this paper is the generic complexity of the *Whitehead minimization problem* for finitely generated subgroups of a free group $F(A)$. Every such subgroup H is a regular subset of $F(A)$ and can be represented uniquely by a finite, edge-labeled graph $\Gamma(H)$ subject to particular constraints, called the Stallings graph of the subgroup; this discrete structure constitutes a natural tool to compute with subgroups, and it also provides a notion of size for H: we denote by $|H|$ the number of vertices of $\Gamma(H)$.

A natural equivalence relation on subgroups is provided by the action of the automorphism group of $F(A)$: the subgroups H and K are in the same orbit if $K = \varphi(H)$ for some automorphism φ of $F(A)$ — that is, H and K are "the same" up to a change of basis in the ambient group. The Whitehead minimization problem consists in finding a minimum size element in the orbit of a given finitely generated subgroup H. This problem is decidable in polynomial time (Roig, Ventura and Weil [16], following an early result of Gersten [7]). We refer the readers to [13] for the usage of this problem in solving the more general orbit membership problem.

*This work was partially supported by the ANR through ANR-2010-BLAN-0204, through ANR-10-LABX-58 and through ANR-JCJC-12-JS02-012-01

Here we are rather interested in the notion of generic complexity, that is, the complexity of the problem when restricted to a generic set of instances (a set of instances such that an instance of size n sits in it with probability tending to 1 when n tends to infinity; precise definitions are given below). Our main result states that the generic complexity of the Whitehead minimization problem is constant, and more precisely, that the set of Whitehead minimal subgroups is generic (see [14] for an early discussion of the generic complexity of this problem, especially in the case of cyclic subgroups).

An implicit element of the discussion of complexity is the notion of size of inputs. In the case of finitely generated subgroups of a free group, we can use either a k-tuple (k fixed) of words which are generators of the subgroup H (and the size of the input is the sum of the lengths of these words), or the Stallings graph of H (and the size is $|H|$). These two ways of specifying the subgroup H give closely related worst-case complexities (because of linear inequalities between the two notions of size), but they can give very different generic complexities: it was shown in [2] that malnormality (an important property of subgroups) is generic if subgroups are specified by a tuple of generators, whereas non-malnormality is generic if subgroups are specified by their Stallings graph. Our results show that Whitehead minimality is generic in both set-ups.

A key ingredient of our proofs is a purely combinatorial characterization of Whitehead minimality in terms of the properties of the graph $\Gamma(H)$ (Proposition 2.2 below), proved in [16], which involves counting the edges labeled by certain subsets of the alphabet in and out of each vertex. This is what allows us to turn the algebraic problem into a combinatorial one, which can be tackled with the methods of combinatorics and theoretical computer science.

Interestingly, the reasons why Whitehead minimality is generic when subgroups are specified by their Stallings graph, and why it is generic when subgroups are specified by a k-tuple of words, are directly opposite. The Stallings graph of the subgroup generated by a k-tuple of words of length at most n generically consists of a small central tree and long loops connecting leaves of the tree, so much of the geometry of the graph is along these long loops, where each vertex is adjacent to only two edges. In contrast, an n-vertex Stallings graph generically has many transitions and each vertex is adjacent to a near-full set of edges.

The origins of this work go back to discussions with Armando Martino and Enric Ventura in 2009.

2 Preliminaries

Let $r > 1$, let A be a finite r-element set and let $F(A)$ be the *free group on* A. We can think of $F(A)$ as the set of *reduced* words on the symmetrized alphabet $\tilde{A} = A \cup \bar{A}$, where $\bar{A} = \{\bar{a} \mid a \in A\}$. Recall that a word is reduced if it does not contain occurrences of the words of the form $a\bar{a}$ or $\bar{a}a$ ($a \in A$). The operation $x \mapsto \bar{x}$ is extended to $\tilde{A}^*$ by letting $\bar{\bar{a}} = a$ and $\overline{ub} = \bar{b}\bar{u}$ for $a \in A$, $b \in \tilde{A}$ and $u \in \tilde{A}^*$.

We denote by $[n]$ the set of positive integers less than or equal to n, and by $\mathcal{R}_n$ (resp. $\mathcal{R}_{\leq n}$) the set of reduced words of length exactly (resp. at most) n. A reduced word u is called *cyclically reduced* if u^2 is reduced, and we let $\mathcal{C}_n$ (resp. $\mathcal{C}_{\leq n}$) be the set of cyclically reduced words of length exactly (resp. at most) n.

2.1 Stallings graph of a subgroup

It is now classical to represent the finitely generated subgroups of a free group by finite rooted edge-labeled graphs, subject to certain combinatorial constraints. An *A-graph* is a finite graph Γ whose edges are labeled by elements of A. It can be seen also as a transition system on alphabet $\tilde{A}$, with the convention that every a-edge from p to q represents an a-transition from p to q

and an $\bar{a}$-transition from q to p. Say that Γ is *reduced* if it is connected and if no two edges with the same label start (resp. end) at the same vertex: this is equivalent to stating that the corresponding transition system is deterministic and co-deterministic. If 1 is a vertex of Γ, we say that $(\Gamma, 1)$ is *rooted* if every vertex, except possibly 1, has valency at least 2.

If H is a finitely generated subgroup of $F(A)$, there exists a unique reduced rooted graph $(\Gamma(H), 1)$, called the *Stallings graph* of H, such that H is exactly the set of reduced words accepted by $(\Gamma(H), 1)$: a reduced word is accepted when it labels a loop starting and ending at 1. Moreover, this graph can be effectively computed given a tuple of reduced words generating H, in time $\mathcal{O}(n \log^* n)$ [19, 20]. We denote by $|H|$ the number of vertices of $\Gamma(H)$, which we interpret as a notion of *size* of H. Observe that if H is the cyclic subgroup generated by a cyclically reduced word w, then $|H|$ is the length of w. This algorithmic construction and the idea of systematically using these graphs to compute with finitely generated subgroups of free groups, go back to Serre's and Stallings' seminal papers ([18] and [19] respectively).

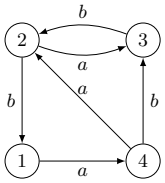

Figure 1: The Stallings graph of $H = \langle aab, ab\bar{a}b, abbb \rangle$. The reduced word $u = aab\bar{b}\bar{a}b$ is in H as it is accepted by $\Gamma(H)$: it labels a path starting from 1 and ending at 1, with edges being used backward when reading a negative letter. Since every vertex has valency at least 2, this graph is cyclically reduced.

We record the following fact, which will be useful in the sequel. Say that an A-graph Γ is *cyclically reduced* if it is reduced and every vertex has valency at least 2. The A-graph in Fig. 1 is cyclically reduced. If H is a finitely generated subgroup of $F(A)$ and $\Gamma(H)$ is not cyclically reduced, then the distinguished vertex 1 has valency 1. Let Γ' be the graph obtained from $\Gamma(H)$ by repeatedly erasing every vertex of valency 1 (and the edges adjacent to them): then Γ' is cyclically reduced and if v is a vertex of Γ', then (Γ', v) is the Stallings graph of some conjugate $H^g = g^{-1}Hg$ of H.

2.2 Whitehead minimality

Say that a subgroup H is *Whitehead minimal* if it has minimum size in its automorphic orbit, that is if $|H| \leq |\varphi(H)|$ for every automorphism φ of $F(A)$. It is *strictly Whitehead minimal* if $|H| < |\varphi(H)|$ for every automorphism φ that is not length preserving (i.e., that is not induced by a permutation of $\tilde{A}$). Strict Whitehead minimality means that H is the only minimum size representative of its orbit, up to a permutation of the letters (that is, up to a relabeling of the edges of its Stallings graph).

Observe, following the discussion at the end of Section 2.1, that if $\Gamma(H)$ is not cyclically reduced, then H is not Whitehead minimal.

A crucial characterization of (strict) Whitehead minimality can be expressed in terms of the so-called *Whitehead automorphisms*. More precisely Whitehead exhibited a finite family $\mathsf{Wh}(A)$ of automorphisms of $F(A)$, with the remarkable property that a subgroup is Whitehead minimal if and only if $|H| \leq |\varphi(H)|$ for every $\varphi \in \mathsf{Wh}(A)$ (this is a result of Whitehead himself for cyclic subgroups, see [13], and of Gersten in the general case [7]).

In this paper we will use a combinatorial formulation of this characterization of Whitehead minimality, which was proved in [16], and which we now explain. We distinguish three kinds of Whitehead automorphisms. Firstly, the length-preserving automorphisms of $F(A)$, which permute the letters of $\tilde{A}$ and for which we always have $|\varphi(H)| = |H|$: they can be disregarded when assessing whether a subgroup is Whitehead minimal. Secondly the inner automorphisms

of the form $g \mapsto g^v = v^{-1}gv$ for some letter $v \in \tilde{A}$. As discussed above, $\Gamma(H)$ is not cyclically reduced if and only if one of these automorphisms satisfies $|\varphi(H)| < |H|$.

The third and last kind of Whitehead automorphisms is in bijection with the set of pairs (Y, v) where Y is a subset of $\tilde{A}$ and v is a letter in $\tilde{A}$ such that $v \in Y$, $\bar{v} \notin Y$ and $2 \leq |Y| \leq 2|A| - 2$. Such a pair (Y, v) is called a *Whitehead descriptor*. The corresponding Whitehead automorphism fixes the letters v and $\bar{v}$ and maps each letter $a \in \tilde{A} \setminus \{v, \bar{v}\}$ to

$$\varphi(a) = v^\lambda a v^\rho \text{ where } \lambda = \begin{cases} -1 & \text{if } \bar{a} \in Y, \\ 0 & \text{otherwise;} \end{cases} \qquad \rho = \begin{cases} 1 & \text{if } a \in Y, \\ 0 & \text{otherwise.} \end{cases}$$

Let Γ be a reduced graph, and let (Y, v) be a Whitehead descriptor. Then we let $\texttt{positive}(\Gamma, Y, v)$ be the set of vertices of Γ with at least one incoming edge labeled by a letter in Y, at least one incoming edge labeled by a letter not in Y, and no incoming edge labeled v. Let also $\texttt{negative}(\Gamma, Y, v)$ be the set of vertices with an incoming edge labeled v, and all other incoming edges labeled by letters in Y.

Example 2.1 Consider the Whitehead descriptor (Y, v) with $v = \overline{a}$ and $Y = \{\overline{a}, b\}$. For the graph Γ depicted on Fig. 1, vertex 1 is in $\texttt{negative}(\Gamma, Y, v)$ since its incoming edges are labeled by b and $\overline{a}$ (obtained by flipping the edge $1 \xrightarrow{a} 4$). Vertex 3 is in $\texttt{positive}(\Gamma, Y, v)$ since its incoming edges are labeled by a, b and $\overline{b}$, one not in Y, one in Y and all different from v. One can also verify that vertices 2 and 4 are neither in $\texttt{positive}(\Gamma, Y, v)$ nor in $\texttt{negative}(\Gamma, Y, v)$. □

The following statement is a reformulation of the Whitehead-Gersten characterization of Whitehead minimality mentioned above in terms of these parameters; it is a consequence of [16, Proposition 2.4].

Proposition 2.2 *A finitely generated subgroup H of $F(A)$ is Whitehead minimal (resp. strictly Whitehead minimal) if and only if it is cyclically reduced and, for every Whitehead descriptor (Y, v), we have $|\,\texttt{positive}(\Gamma(H), Y, v)| \geq |\,\texttt{negative}(\Gamma(H), Y, v)|$ (resp. $|\,\texttt{positive}(\Gamma(H), Y, v)| > |\,\texttt{negative}(\Gamma(H), Y, v)|$).*

Proof. Proposition 2.4 in [16] actually states that, if (Y, v) is a Whitehead descriptor and φ is the corresponding Whitehead automorphism, then $|\varphi(H)| - |H| = |C(H)| - |D(H)|$, where $C(H)$ is the set of vertices of $\Gamma(H)$ with incoming Y-labeled and Y^c-labeled edges, and $D(H)$ is the set of vertices with an incoming v-labeled edge. The intersection $B(H) = C(H) \cap D(H)$ is the set of vertices with an incoming v-labeled edge and some incoming Y^c-labeled edge. Moreover, $\texttt{positive}(\Gamma(H), Y, v)$ is the complement of $B(H)$ in $C(H)$ and $\texttt{negative}(\Gamma(H), Y, v)$ is the complement of $B(H)$ in $D(H)$. The proposition follows immediately. □

2.3 Distributions over finitely generated subgroups

Let S be a countable set, the disjoint union of finite sets S_n ($n \geq 0$), and let $B_n = \bigcup_{i \leq n} S_i$. Typically in this paper, S will be the set of Stallings graphs, of partial injections, of reduced words or of k-tuples of reduced words, and S_n will be the set of elements of S of size n.

A subset X of S is *negligible* if the probability for an element of B_n to be in X, tends to 0 when n tends to infinity; that is, if $\lim_n \frac{|X \cap B_n|}{|B_n|} = 0$.

The notion is refined as follows: we say that X is *exponentially* (resp. *super-polynomially*, *polynomially*) *negligible* if $\frac{|X \cap B_n|}{|B_n|}$ is $\mathcal{O}(e^{-cn})$ for some $c > 0$ (resp. $\mathcal{O}(n^{-k})$ for every positive integer k, $\mathcal{O}(n^{-k})$ for some positive integer k). The set X is exponentially (resp. super-polynomially,

polynomially, simply) *generic* if its complement is exponentially (resp. super-polynomially, polynomially, simply) negligible. We note the following elementary lemma.

Lemma 2.3 *With the above notation, if $C \subseteq S$ satisfies $\liminf_n \frac{|C\cap B_n|}{|B_n|} = p > 0$ and X is exponentially (resp. super-polynomially, polynomially, simply) negligible in S, then so is $X \cap C$ in C.*

Proof. The verification is immediate if we observe that, for n large enough,

$$\frac{|X \cap C \cap B_n|}{|C \cap B_n|} \leq \frac{|X \cap B_n|}{|C \cap B_n|} = \frac{|X \cap B_n|}{|B_n|} \frac{|B_n|}{|C \cap B_n|} \leq \frac{2}{p} \frac{|X \cap B_n|}{|B_n|}.$$

□

Genericity and negligibility can also be defined using the radius n spheres S_n instead of the balls B_n. The same properties are generic or negligible, exponentially, super-polynomially, polynomially or simply, provided $|B_n|$ grows fast enough, see for instance [2, Sec. 2.2.2].

The graph-based distribution. The uniform distribution on the set of size n Stallings graphs was analyzed by Bassino, Nicaud and Weil [3]. Here we summarize the principles of this distribution and the features which will be used in this paper.

In a Stallings graph, each letter labels a partial injection on the vertex set: in fact, such a graph can be viewed as an A-tuple $\vec{f} = (f_a)_{a\in A}$ of partial injections on an n-element set, with a distinguished vertex, and such that the resulting graph (with an a-labeled edge from i to j if and only if $j = f_a(i)$) is connected and has no vertex of valency 1, except perhaps the distinguished vertex. We may even assume that the n-element set in question is $[n]$, with 1 as the distinguished vertex, see [3, Section 1.2] for a precise justification.

Let $\mathcal{I}_n$ denote the set of partial injections on $[n]$ and let $\mathcal{B}_n$ be the set of r-tuples in $\mathcal{I}_n^r$ which define a Stallings graph (recall that $|A| = r$). Let also $\mathcal{D}_n$ be the subset of $\mathcal{B}_n$, of those r-tuples which define a cyclically reduced Stallings graph. Then $\mathcal{D}_n$ (and hence $\mathcal{B}_n$) is generic in $\mathcal{I}_n^r$ [3, Corollary 2.7]

The fundamental observation, used in [3] to achieve this result, is the following: the functional graph of a partial injection $f \in \mathcal{I}_n$ (that is: the pair $([n], E)$ where $i \to j \in E$ whenever $j = f(i)$), is made of cycles and sequences.This allows the use of the analytic combinatorics calculus on exponential generating series (EGS) [6, Sec. II.2]. Recall that, if I_n is the number of partial injections on $[n]$, the corresponding EGS is $I(z) = \sum_{n\geq 0} \frac{1}{n!} I_n z^n$. From [3, Sec. 2.1 and Proposition 2.10], we get

$$I(z) = \frac{1}{1-z} \exp\left(\frac{z}{1-z}\right) \quad \text{and} \quad \frac{I_n}{n!} = \frac{e^{-\frac{1}{2}}}{2\sqrt{\pi}} e^{2\sqrt{n}} n^{-\frac{1}{4}} (1 + o(1)). \tag{1}$$

The formula for $I(z)$ is based on the fact that a partial injection is a set of sequences (whose EGS is $\frac{z}{1-z}$) and of cycles (whose EGS is $\log\left(\frac{1}{1-z}\right)$). We refer the readers to [6, Sec. II.2] and [3] for further details. We use again this calculus in Section 3.1.

The word-based distribution. The distribution more commonly found in the literature (e.g. [11, 9, 10]), which we term *word-based*, originated in the work of Arzhantseva and Ol'shanskiĭ [1]. It is in fact a distribution on the k-tuples $\vec{h} = (h_1, \ldots, h_k)$ of reduced words of length at most n, where k is fixed and n is allowed to grow to infinity; one then considers the subgroup H generated by $\vec{h}$.

This is a reasonable way of defining a distribution on finitely generated subgroups of $F(A)$, and even on rank k subgroups, in spite of the fact that different tuples may generate the same subgroup (see for instance [2, Sec. 3.1]).

The literature also considers Gromov's so-called density model, which uses much larger random tuples (of positive density within $\mathcal{C}_n$). This model is usually considered to study the asymptotic properties of finite group presentations rather than subgroups of $F(A)$ and we will not discuss it here (see for instance [15]).

We will use the following statistics on the number of reduced and cyclically reduced words, which can be easily verified:

$$|\mathcal{R}_m| = 2r(2r-1)^{m-1} \quad \text{and} \quad 2r(2r-1)^{m-2}(2r-2) \leq |\mathcal{C}_m| \leq |\mathcal{R}_m|.$$

Summing over all $m \leq n$, we find that

$$|\mathcal{R}_{\leq n}| = \frac{r}{r-1}\big((2r-1)^n - 1\big) \quad \text{and} \quad 2r\big((2r-1)^{n-1} - 1\big) \leq |\mathcal{C}_{\leq n}| \leq |\mathcal{R}_{\leq n}|.$$

In particular, both $|\mathcal{R}_{\leq n}|$ and $|\mathcal{C}_{\leq n}|$ are $\Theta\big((2r-1)^n\big)$ and $\liminf_n \frac{|\mathcal{C}_{\leq n}|}{|\mathcal{R}_{\leq n}|} > 0$ (see Lemma 2.3).

3 The graph-based distribution

We now study the genericity of strict Whitehead minimality for the graph-based distribution. The proof of Theorem 3.1 below is given in Sections 3.1 and 3.2.

Theorem 3.1 *Strict Whitehead minimality is super-polynomially generic for the uniform distribution over the set of cyclically reduced Stallings graphs.*

3.1 Statistical properties of size n partial injections

If f is a partial injection on $[n]$, we let

- $\texttt{sequence}(f)$ be the number of sequences in the functional graph of f; a sequence has at least one vertex;
- $\texttt{extr}(f) = \{i \in [n] \mid f(i) \text{ is undefined or } i \text{ has no preimage by } f\}$; it is the set of extremities of sequences in the functional graph of f.

We note that, for every $f \in \mathcal{I}_n$, because of length 1 sequences,

$$\texttt{sequence}(f) \leq |\,\texttt{extr}(f)| \leq 2\,\texttt{sequence}(f). \tag{2}$$

Proposition 3.2 *For the uniform distribution, the probability that the number of sequences of a size n partial injection is not in $(\frac{1}{2}\sqrt{n}, 2\sqrt{n})$ is super-polynomially small (of the form $\mathcal{O}(e^{-c\sqrt{n}})$ for some $c > 0$).*

Proof. If $T(z)$ is a formal power series, we denote by $[z^n]T(z)$ the coefficient of z^n in the series. For any $k \geq 0$, let $S^k(z)$, $S^{\leq k}(z)$ and $S^{\geq k}(z)$ be the EGSs of the partial injections having respectively exactly k, at most k and at least k sequences. Observe that an injection with k sequences is a set of k sequences together with a set of cycles; the symbolic method [6, Sec. II.2] therefore yields:

$$S^k(z) = \frac{1}{k!}\left(\frac{z}{1-z}\right)^k \frac{1}{1-z}.$$

The radius of convergence of this series is 1, and Cauchy's estimate for the coefficient of a power series [17, Theorem 10.26] states that for any positive real $\zeta < 1$, we have

$$[z^n]S^k(z) \leq \frac{S^k(\zeta)}{\zeta^n}.$$

Taking $\zeta = 1 - \frac{1}{\sqrt{n}}$ approximatively minimizes the right hand quantity, and after basic computations we obtain that for n large enough,

$$[z^n]S^k(z) \leq \sqrt{n}\, e^{2+\sqrt{n}}\, \frac{n^{\frac{k+1}{2}}}{k!}.$$

Since $S^{\leq \frac{1}{2}\sqrt{n}}(z) = \sum_{k=0}^{\frac{1}{2}\sqrt{n}} S^k(z)$ and $S^{\geq 2\sqrt{n}}(z) = \sum_{k=2\sqrt{n}}^{n} S^k(z)$ we get upper bounds for coefficients of both series by bounding $\sum_{k=0}^{\frac{1}{2}\sqrt{n}} \frac{1}{k!} n^{\frac{k}{2}}$ and $\sum_{k=2\sqrt{n}}^{n} \frac{1}{k!} n^{\frac{k}{2}}$ from above. The term $\frac{1}{k!} n^{\frac{k}{2}}$ is increasing in the first sum and decreasing in the second one, so we can bound each term of each series by its maximum value. This yields the following inequalities:

$$\sum_{k=0}^{\frac{1}{2}\sqrt{n}} \frac{n^{\frac{k}{2}}}{k!} \leq \sum_{k=0}^{\frac{1}{2}\sqrt{n}} \frac{n^{\frac{1}{4}\sqrt{n}}}{(\frac{1}{2}\sqrt{n})!}, \quad [z^n]S^{\leq \frac{1}{2}\sqrt{n}}(z) \leq \frac{n^{\frac{3}{2}+\frac{1}{4}\sqrt{n}}}{(\frac{1}{2}\sqrt{n})!}\, e^{2+\sqrt{n}} \quad \text{and}$$
$$\sum_{k=2\sqrt{n}}^{n} \frac{n^{\frac{k}{2}}}{k!} \leq \sum_{k=2\sqrt{n}}^{n} \frac{n^{\sqrt{n}}}{(2\sqrt{n})!}, \quad [z^n]S^{\geq 2\sqrt{n}}(z) \leq \frac{n^{2+\sqrt{n}}}{(2\sqrt{n})!}\, e^{2+\sqrt{n}}.$$

Using the Stirling bounds [5, Eq. (9.15), p. 54] $n! \geq n^n e^{-n}$ and the asymptotics of I_n in Eq. (1), we obtain upper bounds of the announced form for

$$\frac{[z^n]S^{\leq \frac{1}{2}\sqrt{n}}(z)}{[z^n]I(z)} \quad \text{and} \quad \frac{[z^n]S^{\geq 2\sqrt{n}}(z)}{[z^n]I(z)},$$

respectively the probabilities for a partial injection on $[n]$ to have at most $\frac{1}{2}\sqrt{n}$ and at least $2\sqrt{n}$ sequences. □

We use Proposition 3.2 to bound the number of vertices that are simultaneously extremities for two partial injections.

Proposition 3.3 *For the uniform distribution over size n pairs of partial injections, the probability*

$$\mathbb{P}\left(|\,\mathtt{extr}(f) \cap \mathtt{extr}(f')| \geq \frac{\sqrt{n}}{4(r-1)}\right)$$

is super-polynomially small (of the form $\mathcal{O}(e^{-c\sqrt{n}})$ for some $c > 0$).

Proof. Let f and f' be partial injection on $[n]$. By Proposition 3.2 and Eq. (2), the probability that one of them has more than $4\sqrt{n}$ extremities is super-polynomially small — so we can restrict the analysis to the cases where both f and f' have at most $4\sqrt{n}$ extremities, up to a super-polynomially small error term.

Let $m = \lfloor 4\sqrt{n} \rfloor$. Let E_f and $E_{f'}$ be two sets obtained by adding uniformly at random elements of $[n]$ to $\mathtt{extr}(f)$ and $\mathtt{extr}(f')$ respectively, until $|E_f| = |E_{f'}| = m$. Note that by

symmetry, and since f and f' are chosen independently, both E_f and $E_{f'}$ are uniform and independent size m subsets of $[n]$. Moreover, since $\texttt{extr}(f) \subseteq E_f$ and $\texttt{extr}(f') \subseteq E_{f'}$, we have

$$\mathbb{P}\left(|\,\texttt{extr}(f) \cap \texttt{extr}(f')| \geq \frac{\sqrt{n}}{4(r-1)}\right) \leq \mathbb{P}\left(|E_f \cap E_{f'}| \geq \frac{\sqrt{n}}{4(r-1)}\right).$$

It suffices therefore to show that, super-polynomially generically, the intersection of two m-element subsets of $[n]$ has less than $\frac{\sqrt{n}}{4(r-1)}$ elements. Let $X(n,m,k)$ be the number of pairs of m-subsets whose intersection has size k. Then

$$X(n,m,k) = \binom{n}{k}\binom{n-k}{m-k}\binom{n-m}{m-k}.$$

Therefore the probability that the intersection has size k is

$$\mathbb{P}(|E_f \cap E_{f'}| = k) = \frac{X(n,m,k)}{\binom{n}{m}^2} = k!\binom{m}{k}^2 \frac{(n-m)!^2}{n!(n-2m+k)!}.$$

Note that $\frac{(n-m)!^2}{n!(n-2m+k)!} < (n-m)^{-k}$, that $\binom{m}{k} < 2^m$. Let $\alpha = \frac{1}{4(r-1)}$. Then

$$\mathbb{P}(|E_f \cap E_{f'}| \geq \alpha\sqrt{n}) = \sum_{k=\alpha\sqrt{n}}^{m} \mathbb{P}(|E_f \cap E_{f'}| = k) < 2^{2m} \sum_{k=\alpha\sqrt{n}}^{m} \frac{k!}{(n-m)^k}.$$

Moreover $k \mapsto \frac{k!}{(n-m)^k}$ is decreasing for $k \leq m$ (for n large enough), so we have

$$\mathbb{P}(|E_f \cap E_{f'}| \geq \alpha\sqrt{n}) < 2^{2m} m \frac{(\alpha\sqrt{n})!}{(n-m)^{\alpha\sqrt{n}}} < 2^{8\sqrt{n}} 4\sqrt{n} \left(\frac{\alpha\sqrt{n}}{n - 4\sqrt{n}}\right)^{\alpha\sqrt{n}}.$$

This concludes the proof since the dominant term is of the form $n^{-\frac{\alpha}{2}\sqrt{n}}$. □

3.2 From partial injections to Stallings graph

Notice that if (Y, v) is a Whitehead descriptor, the definitions of the functions $\texttt{negative}(-, Y, v)$ and $\texttt{positive}(-, Y, v)$ make sense for all r-tuple of size n partial injections, even if they do not form a (cyclically reduced) Stallings graph. We will use the following combinatorial bounds to establish Theorem 3.1.

Lemma 3.4 *Let (Y, v) be a Whitehead descriptor and let $\vec{f} = (f_a)_{a\in A} \in \mathcal{I}_n^r$. If $v \in \bar{A}$, we let $f_v = f_{\bar{v}}^{-1}$. Then we have*

$$|\,\texttt{negative}(\vec{f}, Y, v)| \leq \sum_{a \neq v} |\,\texttt{extr}(f_v) \cap \texttt{extr}(f_a)|,$$

$$|\,\texttt{positive}(\vec{f}, Y, v)| \geq \texttt{sequence}(f_v) - \sum_{a \neq v} |\,\texttt{extr}(f_v) \cap \texttt{extr}(f_a)|.$$

Proof. Recall that a vertex p in $\texttt{negative}(\vec{f}, Y, v)$ has an incoming v-edge and all its incoming edges have labels in Y. Since $\bar{v} \notin Y$, it follows that $p \in \texttt{extr}(f_v)$. Moreover, if $a \notin Y$ and $a \neq \bar{v}$ (there exists such an a since $|Y| \leq 2r-2$), p has no incoming a-edge, so $p \in \texttt{extr}(f_a)$. This establishes the first inequality.

Similarly, if $v \in A$ and p is the initial vertex of a sequence of f_v (and hence a v-extremity), and if in addition p is not an a-extremity for any $a \neq v, \bar{v}$, then $p \in \mathtt{positive}(\vec{f}, Y, v)$. Therefore, if $\mathtt{begin}(f_v)$ denotes the set of initial vertices of sequences of f_v, we have

$$\mathtt{begin}(f_v) \setminus \bigcup_{a \neq v, \bar{v}} \mathtt{extr}(f_v) \cap \mathtt{extr}(f_a) \subseteq \mathtt{positive}(\vec{f}, Y, v),$$

and the announced inequality follows since $|\,\mathtt{begin}(f_v)| = \mathtt{sequence}(f_v)$.

If $\bar{v} \in A$ we consider instead the set of final vertices of sequences in $f_{\bar{v}}$. □

Proof of Theorem 3.1. Let $\mathcal{D}_n$ be the set of r-tuples of size n partial injections which define a cyclically reduced Stallings graph, and let $\mathcal{E}_n$ be the set of r-tuples $\vec{f}$ of size n partial injections which fail to satisfy $|\,\mathtt{positive}(\vec{f}, Y, v)| > |\,\mathtt{negative}(\vec{f}, Y, v)|$ for some Whitehead descriptor (Y, v). By Proposition 2.2, we want to show that $\mathcal{E}_n \cap \mathcal{D}_n$ is super-polynomially negligible within $\mathcal{D}_n$.

Since $\mathcal{D}_n$ is generic in the full set of r-tuples of partial injections, namely $\mathcal{I}_n^r$ (see Section 2.3), Lemma 2.3 shows that we only need to show that $\mathcal{E}_n$ is super-polynomially negligible in $\mathcal{I}_n^r$.

For each Whitehead descriptor (Y, v), let $\mathcal{E}_n(Y, v)$ denote the set of r-tuples $\vec{f} \in \mathcal{I}_n^r$ such that $|\,\mathtt{positive}(\vec{f}, Y, v)| \leq |\,\mathtt{negative}(\vec{f}, Y, v)|$. Then $\mathcal{E}_n$ is the (finite) union of the $\mathcal{E}_n(Y, v)$ and it suffices to prove that each $\mathcal{E}_n(Y, v)$ is super-polynomially negligible in $\mathcal{I}_n^r$.

For a fixed Whitehead descriptor (Y, v), Lemma 3.4 shows that

$$\mathbb{P}\Big(\mathcal{E}_n(Y, v)\Big) \leq \mathbb{P}\Big(\,\mathtt{sequence}(f_v) \leq 2 \sum_{a \neq v} |\,\mathtt{extr}(f_v) \cap \mathtt{extr}(f_a)|\Big).$$

We observe that if $|\,\mathtt{extr}(f_v) \cap \mathtt{extr}(f_a)| < \frac{1}{4(r-1)}\sqrt{n}$ for each $a \in A$, $a \neq v, \bar{v}$ and $\mathtt{sequence}(f_v) > \frac{1}{2}\sqrt{n}$, then $2\sum_{a \neq v} |\,\mathtt{extr}(f_v) \cap \mathtt{extr}(f_a)| < \frac{1}{2}\sqrt{n} < \mathtt{sequence}(f_v)$, so that $\vec{f} \notin \mathcal{E}_n(Y, v)$. Therefore, by considering the complements of these properties, we see that $\mathbb{P}(\mathcal{E}_n(Y, v))$ is at most equal to

$$\mathbb{P}\Big(\,\mathtt{sequence}(f_v) \leq \frac{1}{2}\sqrt{n}\Big) + \sum_{a \neq v} \mathbb{P}\Big(|\,\mathtt{extr}(f_v) \cap \mathtt{extr}(f_a)| \geq \frac{1}{4(r-1)}\sqrt{n}\Big).$$

This concludes the proof since each of the summands is super-polynomially small by Propositions 3.2 and 3.3. □

Theorem 3.1 is stated for the uniform distribution on *cyclically reduced* Stallings graphs. One may wonder if a similar result holds for the uniform distribution on Stallings graph. We show the following.

Corollary 3.5 *Strict Whitehead minimality is polynomially, but not super-polynomially, generic for the uniform distribution over Stallings graphs.*

Proof. As per the proof of Theorem 3.1, an r-tuple $\vec{f} \in \mathcal{I}_n^r$ satisfies super-polynomially generically the constraint that $|\,\mathtt{positive}(\vec{f}, Y, v)| > |\,\mathtt{negative}(\vec{f}, Y, v)|$ for any Whitehead descriptor (Y, v), – and hence a Stallings graph $(\Gamma(H), 1)$ super-polynomially generically satisfies the constraint $|\,\mathtt{positive}(\Gamma(H), Y, v)| > |\,\mathtt{negative}(\Gamma(H), Y, v)|$ for any (Y, v).

For H to be strictly Whitehead minimal, $\Gamma(H)$ must also be cyclically reduced. Equivalently, vertex 1 must be of valency at least 2, that is, it must not be an extremity for one letter and isolated (i.e., the extremity of a length 1 sequence) for all other letters.

The probability that a vertex p is an extremity for the partial injection f is $\frac{1}{n}|\,\mathtt{extr}(f)|$, which is $\Theta(\frac{1}{\sqrt{n}})$ by Proposition 3.2. The probability that p is isolated is $\frac{I_{n-1}}{I_n}$, which is $\Theta(\frac{1}{n})$ by Eq. (1). Therefore, vertex 1 is of valency less than 2 with probability $\Theta(n^{-(r-1)-\frac{1}{2}})$, which concludes the proof. □

In other words, the uniform distribution on Stallings graphs exhibits the same behavior as that on cyclically reduced graphs with respect to strict Whitehead minimality, but with a weaker error term.

4 The word-based distribution

Let $k \geq 2$ be a fixed integer. We discuss the genericity of strict Whitehead minimality for the subgroups generated by a random k-tuple of cyclically reduced words and we show the following.

Theorem 4.1 *For the uniform distribution over k-tuples of cyclically reduced words of length at most n, strict Whitehead minimality is exponentially generic.*

4.1 Shape of the Stallings graph

The following elementary statement combines results established in [1, 9] and in [2, Sec. 3.1].

Proposition 4.2 *Let $\alpha \in (0,1)$ and $0 < \beta < \frac{1}{2}\alpha$, let $\vec{h} = (h_1, \ldots, h_k)$ be a tuple of elements of $\mathcal{R}_{\leq n}$ and let H be the subgroup generated by $\vec{h}$. Then, exponentially generically,*

- $\min |h_i| > \lceil \alpha n \rceil$ and the prefixes of the h_i and h_i^{-1} of length $\lfloor \beta n \rfloor$ are pairwise distinct

- the Stallings graph $\Gamma(H)$ consists of a central tree *of height $\lfloor \beta n \rfloor$ – whose vertices can be identified with the prefixes and suffixes of length at most $\lfloor \beta n \rfloor$ of the h_i – and of k* outer loops*, one for each h_i, of length $|h_i| - 2\lfloor \beta n \rfloor$, connecting the leaves of the central tree.*

Proposition 4.2 describes the typical shape of a Stallings graph under the word-based distribution: as β can be taken arbitrarily small and α arbitrarily close to 1, an overwhelming proportion of the vertices are in the outer loops, and in particular have valency exactly two.

4.2 Counting the occurrences of short factors

If u is a word over an alphabet B, we denote by $Z_n(u)$ the function that counts the occurrences of u as a factor in a word in B^n.

Lemma 4.3 *Let B be a finite alphabet with $k \geq 2$ letters and let $u \in B^m$. Then the mean value of $Z_n(u)$ is asymptotically equivalent to $\frac{n}{k^m}$. Moreover, for any $\varepsilon > 0$ there exists a constant $c > 0$ such that*

$$\mathbb{P}\left(\left|Z_n(u) - \frac{n}{k^m}\right| \geq \varepsilon n\right) \leq e^{-cn}.$$

Proof. For $i \in [n+1-m]$, the probability $X_n^{(i)}$ that u is a factor at position i in a random word of length n is k^{-m}, with the convention that the first letter is at position 1. For each $\ell \in [m]$, let $Z_n^{(\ell)}(u) = \sum_j X_n^{(mj+\ell)}$, for $0 \leq j \leq \lfloor \frac{n+1-\ell}{m} \rfloor$. Each $Z_n^{(\ell)}(u)$ is the sum of independent random variables since there is no overlap in the portions of the length n word considered. Therefore $Z_n^{(\ell)}(u)$ follows a binomial law of parameters k^{-m} and $\lfloor \frac{n+1-\ell}{m} \rfloor$: by Hoeffding's inequality [8], it is centered around its mean value which is equivalent to $\frac{n}{mk^m}$, and it

satisfies $\mathbb{P}\left(\left|Z_n^{(\ell)}(u) - \frac{n}{mk^m}\right| > \frac{\varepsilon}{m}n\right) \leq e^{-c_\ell n}$ for some $c_\ell > 0$ and for each n large enough. The announced result follows from the fact that $Z_n(u) = Z_n^{(0)}(u) + \ldots + Z_n^{(m-1)}(u)$. □

Now if u is a reduced word over the alphabet $\tilde{A}$, we denote by $\tilde{Z}_n(u)$ the function that counts the occurrences of u as a factor in a reduced word in $\mathcal{R}_n$.

Lemma 4.4 *Let $u = u_1u_2$ be a reduced word of length 2. Then for any $\varepsilon > 0$ there exists a constant $c > 0$ such that, for n large enough,*

$$\mathbb{P}\left(\tilde{Z}_n(u) > \left(\frac{1}{(2r-1)^2} + \varepsilon\right)(n-1) + 1\right) \leq e^{-cn}$$

and

$$\mathbb{P}\left(\tilde{Z}_n(u) < \left(\frac{2r-2}{(2r-1)^3} - 2\varepsilon\right)(n-1)\right) \leq e^{-cn}$$

Proof. We first consider the case where $u_1 \neq u_2$. The idea is to use Lemma 4.3 via an encoding of reduced words. For every $a \in \tilde{A}$, let φ_a be a bijective map from $\tilde{A} \setminus \{\bar{a}\}$ to $[2r-1]$. Let φ be the map from the set of reduced words to $\tilde{A} \times [2r-1]^*$ defined for every reduced word $z = z_1 \cdots z_n$ by

$$\varphi(z) = (z_1, \varphi_{z_1}(z_2)\varphi_{z_2}(z_3) \cdots \varphi_{z_{n-1}}(z_n)).$$

Observe that for every $n > 0$, φ is a bijection from $\mathcal{R}_n$ to $\tilde{A} \times [2r-1]^{n-1}$, which is computed by an automaton with outputs: the states are the elements of $\tilde{A}$ and for every $a \in \tilde{A}$ and $b \neq \bar{a}$, there is a transition from a to b on input b with output $\varphi_a(b)$. Moreover, the uniform distribution on $\mathcal{R}_n$ is obtained by choosing z_1 uniformly in $\tilde{A}$, z' uniformly in $[2r-1]^{n-1}$, and taking $\varphi^{-1}(z_1, z')$.

We now choose particular functions φ_a: for every $a \neq \bar{u}_1$, we choose $\varphi_a(u_1) = 1$. This way every occurrence of u_1 (except possibly for the first letter of z), is encoded by a 1 (note that the 1s provided by $\varphi_{\bar{u}_1}$ do not encode an occurrence of u_1). We also require that $\varphi_{u_1}(u_2) = 2$ and $\varphi_a(\bar{u}_1) = 3$ for every $a \neq u_1$: thus every occurrence of $u = u_1u_2$ in z translates to an occurrence of 12 in $\varphi(z)$, and every occurrence of $\bar{u}_1$ translates to a 3 in $\varphi(z)$. See Figure 2 for an example.

	a	$\overline{a}$	b	$\overline{b}$
φ_a	1	–	3	2
$\varphi_{\overline{a}}$	–	3	1	2
φ_b	1	3	2	–
$\varphi_{\overline{b}}$	1	3	–	2

z	b	a	$\overline{b}$	$\overline{a}$	b	b	b	a	a	$\overline{b}$	a	b	a	$\overline{b}$	a
$\varphi(z)$	b	**1**	**2**	3	**1**	**2**	2	1	**1**	**2**	1	3	**1**	**2**	1

Figure 2: An example of the encoding used in the proof of Lemma 4.4. The word z above is encoded using the construction associated with the pattern $u = a\bar{b}$: a is always encoded by a 1, $\overline{b}$ by a 2 and the inverse of the first letter, $\overline{a}$, by a 3. An occurrence of u always corresponds to an occurrence of **12** in $\varphi(z)$, but the opposite is not true: there are false positives, which are always preceded by a 3. Note also that an occurrence of **312** does not always correspond to a false positive.

Then for any t, we have $\mathbb{P}(\tilde{Z}_n(u) > t+1) \leq \mathbb{P}(Z_{n-1}(12) > t)$ (the value $t+1$ in the left-hand side of the inequality corresponds to the possibility of an occurrence of u in the leftmost position).

For $t = \left(\frac{1}{(2r-1)^2} + \varepsilon\right)(n-1)$, this yields

$$\mathbb{P}\Big(\tilde{Z}_n(u) > \ (\frac{1}{(2r-1)^2} + \varepsilon)(n-1) + 1\Big) \leq \mathbb{P}\Big(Z_{n-1}(12) > (\frac{1}{(2r-1)^2} + \varepsilon)(n-1)\Big)$$
$$\leq \mathbb{P}\Big(|Z_{n-1}(12) - \frac{n-1}{(2r-1)^2}| \geq \varepsilon(n-1)\Big).$$

The first inequality to be proved then follows from Lemma 4.3 since the pattern 12 is taken in $[2r-1]^{n-1}$ equipped with the uniform distribution.

Observe that counting occurrences of 12 overestimates the number of occurrences of u. More specifically, if a false positive occurs, then the said occurrence of 12 is preceded by a 3 in $\varphi(z)$. Hence, the number of false positives is bounded above by the number of occurrences of 312 in $\varphi(z)$. Therefore $\mathbb{P}(\tilde{Z}_n(u) < t) \leq \mathbb{P}(Z_{n-1}(12) - Z_{n-1}(312) < t)$. Let then $t = \left(\frac{2r-2}{(2r-1)^3} - 2\varepsilon\right)(n-1) = \left(\frac{n-1}{(2r-1)^2} - \varepsilon(n-1)\right) - \left(\frac{n-1}{(2r-1)^3} + \varepsilon(n-1)\right)$. Then

$$\mathbb{P}\left(\tilde{Z}_n(u) < \ \left(\frac{2r-2}{(2r-1)^3} - 2\varepsilon\right)(n-1)\right)$$
$$\leq \mathbb{P}\left(Z_{n-1}(12) - Z_{n-1}(312) < \left(\frac{2r-2}{(2r-1)^3} - 2\varepsilon\right)(n-1)\right)$$
$$\leq \mathbb{P}\left(|Z_{n-1}(12) - \frac{n-1}{(2r-1)^2}| > \varepsilon(n-1)\right)$$
$$+ \mathbb{P}\left(|Z_{n-1}(312) - \frac{n-1}{(2r-1)^3}| > \varepsilon(n-1)\right).$$

The second inequality to be proved again follows from Lemma 4.3.

The case $u = u_1u_1$ is handled in the same fashion, except that we have to set $\varphi_{u_1}(u_1) = 2$ instead of 1. □

Remark 4.5 The statement of Lemma 4.4, and even a slighty stronger statement, can also be obtained using the theory of Markov chains: a reduced word can be seen as a path in a specific Markov chain – where the set of states is $\tilde{A}$, and there is a transition from a to b with probability $\frac{1}{2r-1}$ whenever $a \neq \bar{b}$. The result in Lemma 4.4 then follows from [12, Thm 1.1]. We chose instead to give the elementary and self-contained presentation above. □

4.3 Proof of Theorem 4.1

Let $\alpha \in (0,1)$, $\beta \in (0, \frac{\alpha}{2})$ and $\varepsilon > 0$ be real numbers, to be chosen later. Let $W_{n,\alpha,\beta}$ be the set of k-tuples $\vec{h} = (h_1, \ldots, h_k)$ of reduced words of length at most n, such that $\min |h_i| > \lceil \alpha n \rceil$ and the prefixes of the h_i and h_i^{-1} of length $\lfloor \beta n \rfloor$ are pairwise distinct.

For each word h of length greater than $2\lfloor \beta n \rfloor$, let $\texttt{mid}(h)$ be the factor of h obtained by deleting the length $\lfloor \beta n \rfloor$ prefix and suffix.

Now let (Y, v) be a Whitehead descriptor and let H be the subgroup generated by $\vec{h} \in W_{n,\alpha,\beta}$. We denote by Y^c the complement of Y. The central tree of $\Gamma(H)$ has at most $2k\beta n$ vertices, and the outer loops of $\Gamma(H)$ are labeled by the $\texttt{mid}(h_i)$. All the vertices in these loops have valency 2. Any one of these vertices is in $\texttt{negative}(\Gamma(H), Y, v)$ if and only if it has an incoming v-edge and an outgoing y-edge for some $y \in Y^c \setminus \{v\}$. Let $N = (Y\bar{v} \cup v\bar{Y}) \setminus \{v\bar{v}\}$. Then the number

of negative vertices in the outer loops is equal to the number of occurrences of elements of N as factors in the $\mathtt{mid}(h_i)$. That is:

$$\mathtt{negative}(\Gamma(H), Y, v) \leq \sum_{i=1}^{k} \sum_{xy \in N} \tilde{Z}_{|\,\mathtt{mid}(h_i)|}(xy) + 2k\beta n.$$

By Proposition 4.2, $W_{n,\alpha,\beta}$ is exponentially generic. Moreover, the map $h \mapsto \mathtt{mid}(h)$ turns the uniform distribution on words in $\mathcal{R}_\ell$ ($\ell > \alpha n$) into the uniform distribution on $\mathcal{R}_{\ell-2\lfloor\beta n\rfloor}$: indeed, if $u \in \mathcal{R}_{\ell-2\lfloor\beta n\rfloor}$, then $\mathbb{P}(\mathtt{mid}(h) = u) = (2r-1)^{-2\lfloor\beta n\rfloor}$, which does not depend on u. It follows that the same map also turns the uniform distribution on the set of reduced words of length greater than αn and less than or equal to n, into the uniform distribution on its image. Therefore, exponentially generically, we have

$$\begin{aligned}\mathtt{negative}(\Gamma(H), Y, v) &\leq 2k\beta n + k|N|\Big(\big(\frac{1}{(2r-1)^2} + \varepsilon\big)(1-2\beta)n + 1\Big) \\ &\leq 2k\beta n + 2k(|Y|-1)\Big(\big(1-2\beta\big)\big(\frac{1}{(2r-1)^2} + \varepsilon\big)n + 1\Big).\end{aligned}$$

Similarly, a loop vertex is in $\mathtt{positive}(\Gamma(H), Y, v)$ if it has an incoming x-edge with $x \in Y \setminus \{v\}$ and an outgoing y-edge with $\bar{y} \in Y^c$: if $P = (Y \setminus \{v\})\overline{Y^c} \cup Y^c(\bar{Y} \setminus \{\bar{v}\})$, then the number of positive vertices in the outer loops is equal to the number of occurrences of elements of P as factors in the $\mathtt{mid}(h_i)$. That is, exponentially generically,

$$\begin{aligned}\mathtt{positive}(\Gamma(H), Y, v) &\geq \sum_{i=1}^{k} \sum_{xy \in P} Z_{|\,\mathtt{mid}(h_i)|}(xy) \\ &\geq k|P| \left(\frac{2r-2}{(2r-1)^3} - 2\varepsilon\right)((\alpha - 2\beta)n - 1) \\ &\geq 2k(|Y|-1)(2r-|Y|) \left(\frac{2r-2}{(2r-1)^3} - 2\varepsilon\right)((\alpha - 2\beta)n - 1)\,.\end{aligned}$$

In order to conclude, we only need to show that we can choose α, β and ε such that

$$\begin{aligned}&(2r-|Y|)\Big(\frac{2r-2}{(2r-1)^3} - 2\varepsilon\Big)((\alpha - 2\beta)n - 1) \\ &\qquad > (1-2\beta)\Big(\frac{1}{(2r-1)^2} + \varepsilon\Big)n + 1 + \frac{\beta n}{|Y|-1}.\end{aligned}$$

for all n large enough. The first term is $\Theta(\gamma n)$ with $\gamma = (2r-|Y|)(\frac{2r-2}{(2r-1)^3} - 2\varepsilon)(\alpha - 2\beta)$ and the second term is $\Theta(\delta n)$ with $\delta = (1-2\beta)(\frac{1}{(2r-1)^2} + \varepsilon) + \frac{\beta}{|Y|-1}$, so we need to select α, β and ε such that $\gamma > \delta$. This is possible by continuity, since the limits of these two quantities when $(\alpha, \beta, \varepsilon)$ tends to $(1, 0, 0)$ are respectively $(2r-|Y|)\frac{2r-2}{(2r-1)^3}$ and $\frac{1}{(2r-1)^2}$, and we have $2r - |Y| \geq 2$ and $\frac{2r-2}{2r-1} \geq \frac{2}{3}$, so that $(2r-|Y|)\frac{2r-2}{(2r-1)^3} \geq \frac{4}{3}\,\frac{1}{(2r-1)^2}$.

This establishes that if H is generated by a k-tuple of reduced words, then exponentially generically $\mathtt{positive}(\Gamma(H), Y, v) > \mathtt{negative}(\Gamma(H), Y, v)$ for each Whitehead descriptor. The same exponential genericity holds for k-tuples of cyclically reduced words in view of Lemma 2.3 and the discussion at the end of Section 2.3. Together with Proposition 2.2, this concludes the proof since a subgroup generated by a tuple of cyclically reduced words has a cyclically reduced Stallings graph. □

To complete the picture, we observe that given a random k-tuple of reduced words, instead of cyclically reduced words, there is a non-negligible probability that the graph is not cyclically reduced.

Proposition 4.6 *For the uniform distribution over k-tuples of reduced words of length at most n the Stallings graph is not generically cyclically reduced.*

Proof. Let $\vec{h} = (h_1, \dots, h_k)$ be a random k-tuple of reduced words of length at most n and let $\Gamma(H)$ be the Stalling graph of the subgroup H generated by $\vec{h}$.

We show that with probability tending to $(\frac{1}{2r})^{2k-1}$, $\Gamma(H)$ is not cyclically reduced and, more precisely, there exists a letter $a \in \tilde{A}$ such that every h_i starts with a and ends with $\overline{a}$.

For every pair of letters a and b in $\tilde{A}$, let $\mathcal{R}_{a,b}$ be the set of reduced words that start with a and end by b. Let $R_{a,b}(z)$ be the (ordinary) generating series associated with $\mathcal{R}_{a,b}$ defined by

$$R_{a,b}(z) = \sum_{u \in \mathcal{R}_{a,b}} z^{|u|}.$$

Assume that $b \notin \{a, \bar{a}\}$. Since a word of $\mathcal{R}_{a,b}$ is either ab or a word in some $\mathcal{R}_{a,c}$ ($c \neq \bar{b}$) followed by b, we have

$$R_{a,b}(z) = z^2 + \sum_{c \neq \bar{b}} R_{a,c}(z) z,$$

and similarly

$$R_{a,a}(z) = z^2 + \sum_{c \neq \bar{a}} R_{a,c}(z) z \qquad \text{and} \qquad R_{a,\bar{a}}(z) = \sum_{c \neq a} R_{a,c}(z) z.$$

Now observe that if $b, c \in \tilde{A} \setminus \{a, \bar{a}\}$, then $R_{a,b}(z) = R_{a,c}(z)$ by symmetry. Hence, fixing a letter $b \in \tilde{A} \setminus \{a, \bar{a}\}$, the equations above rewrite as

$$\begin{cases} R_{a,b}(z) &= z^2 + (2r-3) R_{a,b}(z) z + R_{a,a}(z) z + R_{a,\bar{a}}(z) z \\ R_{a,a}(z) &= z^2 + (2r-2) R_{a,b}(z) z + R_{a,a}(z) z \\ R_{a,\bar{a}}(z) &= (2r-2) R_{a,b}(z) z + R_{a,\bar{a}}(z) z. \end{cases}$$

Solving this system yields (thank you `maple`!)

$$\begin{aligned} R_{a,\bar{a}}(z) &= \frac{2z^3(r-1)}{(1-z^2)(1-(2r-1)z)} \\ &= \frac{2r-2}{2r-1} - \frac{1}{2(1-z)} - \frac{r-1}{2r(1+z)} + \frac{1}{2r(2r-1)(1-(2r-1)z)}. \end{aligned}$$

It follows that the number of words of length n in $\mathcal{R}_{a,\bar{a}}$ is asymptotically equivalent to $\frac{1}{2r}(2r-1)^{n-1}$, and the probability that a reduced word of length n begins with a and ends with $\bar{a}$ is asymptotically equivalent to $\frac{1}{(2r)^2}$. This result also holds for words of length at most n, as they are generically of length greater than $\frac{1}{2}n$.

Thus the probability that the k-words of $\vec{h}$ all begin with the same letter a and end with $\bar{a}$ is asymptotivally equivalent to $\frac{1}{(2r)^{2k}}$, and the probability that they all begin with the same letter and end with its opposite is equivalent to $\frac{1}{(2r)^{2k-1}}$, which concludes the proof. □

5 Application to random generation

Proposition 2.2 and the fact that there are finitely many Whitehead descriptors immediately yield algorithms `MinimalityTest` (resp. `StrictMinimalityTest`) to test whether H is (strictly) Whitehead minimal: it suffices to verify whether $\Gamma(H)$ is cyclically reduced (in time at most linear) and to compute, for each Whitehead descriptor (Y, v), $|\,\texttt{positive}(\Gamma(H), Y, v)|$ and $|\,\texttt{negative}(\Gamma(H), Y, v)|$. The time required is linear in $|H|$ for each (Y, v), but the number of Whitehead descriptors is exponential in A: the resulting algorithm is linear in $|H|$ but not in $|A|$.

In this section, our purpose is different: we want to design efficient random generators – in the graph-based or the word-based distribution – for the Stallings graphs of subgroups that are (strictly) Whitehead minimal.

Our algorithms will be rejection algorithms. In general, suppose that S is a countable set, S is the disjoint union of the S_n, and $C \subseteq S$ is such that $\liminf_n \frac{|C \cap B_n|}{|B_n|} = p > 0$ (see Section 2.3 and Lemma 2.3). If `RandomS` is a random generator for elements of S and `TestC` is an algorithm to test whether an element of S is in C, then the algorithm in Figure 3 is a random generator for elements of C.

```
RandomC(n)
1 keep ← False
2 repeat
3 |   x = RandomS(n)
4 |   keep ← TestC(x)
5 until keep == True
6 return x
```

Figure 3: An algorithm to randomly generate an element of C of size n

In such an algorithm, the loop (lines 3–4) is performed in average $\frac{1}{p}$ times. in particular, if both `RandomS` and `TestC` take linear time in average, then so does `RandomC`.

A random generator `RandomStallingsGraph` working in linear average time, is available for the graph-based and the word-based distributions.

- For the graph-based distribution, such an algorithm is given in [3].
- For the word-based distribution, one first generates a k-tuple of reduced words (in linear time); next one applies Touikan's algorithm [20] to compute the associated Stallings graph; it was noted in [4, Theorem 4.1] that the average time complexity of this algorithm is linear.

Following the model of the algorithm in Figure 3, a rejection algorithm to randomly generate Whitehead minimal subgroups is shown in Figure 4.

Similarly, an algorithm `RandomStrictlyWhiteheadMinimalGraph` to randomly generate strictly Whitehead minimal subgroups, is obtained by replacing the call to `MinimalityTest` by a call to `StrictMinimalityTest`. In view of the discussion at the beginning of this section, this yields the following statement.

Proposition 5.1 *For the graph-based and the word-based distributions, the average time complexity of the algorithms* `RandomWhiteheadMinimalGraph` *and* `RandomStrictlyWhiteheadMinimalGraph` *is linear.*

```
RandomWhiteheadMinimalGraph(n,A)
keep ← False
repeat
    Γ = RandomStallingsGraph(n,A)
    keep ← MinimalityTest(Γ)
until keep == True
return Γ
```

Figure 4: An algorithm to randomly generate Whitehead minimal subgroups

References

[1] G. N. Arzhantseva, A. Yu. Ol'shanskiĭ. Generality of the class of groups in which subgroups with a lesser number of generators are free. *Mat. Zametki*, 59:489-496, 638, 1996.

[2] F. Bassino, A. Martino, C. Nicaud, E. Ventura, P. Weil. Statistical properties of subgroups of free groups. *Random Struct. Algorithms*, 42:349-373, 2013.

[3] F. Bassino, C. Nicaud, P. Weil. Random generation of finitely generated subgroups of a free group. *Internat. J. Algebra Comput.*, 18:375-405, 2008.

[4] F. Bassino, C. Nicaud, P. Weil. Generic properties of random subgroups of a free group for general distributions. In *23rd Intern. Meeting on the Analysis of Algorithms*, Discrete Math. Theor. Comput. Sci. Proc., AQ, pp. 155-166, 2012.

[5] W. Feller, *An introduction to probability theory and its applications*, 3rd edition, vol. 1, Wiley, 1968.

[6] Ph. Flajolet, R. Sedgewick. *Analytic combinatorics*. Cambridge Univ. Press, 2009.

[7] S. M. Gersten. On Whitehead's algorithm. *Bull. Amer. Math. Soc.*, 10:281-284, 1984.

[8] W. Hoeffding. Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.*, 58:13-30, 1963.

[9] T. Jitsukawa. Malnormal subgroups of free groups. In *Computational and statistical group theory*, Contemp. Math., vol. 298, pp. 83-95. Amer. Math. Soc., 2002.

[10] I. Kapovich, A. Miasnikov, P. Schupp, V. Shpilrain. Generic-case complexity, decision problems in group theory, and random walks. *J. Algebra*, 264:665-694, 2003.

[11] I. Kapovich, P. Schupp, V. Shpilrain. Generic properties of Whitehead's algorithm and isomorphism rigidity of random one-relator groups. *Pacific J. Math.*, 223:113-140, 2006.

[12] P. Lezaud. Chernoff-type bound for finite Markov chains. *Annals of Applied Probability*, 8:849-867, 1998.

[13] R. C. Lyndon, Paul E. Schupp. *Combinatorial group theory*. Springer, 1977. Ergebnisse der Mathematik und ihrer Grenzgebiete, vol. 89.

[14] A. D. Miasnikov, A. G. Myasnikov. Whitehead method and genetic algorithms. In *Computational and experimental group theory*, Contemp. Math., vol. 349, pp. 89-114. Amer. Math. Soc., 2004.

[15] Y. Ollivier. *A January 2005 invitation to random groups*, Ensaios Matemáticos, vol. 10. Soc. Bras. de Matemática, 2005.

[16] A. Roig, E. Ventura, P. Weil. On the complexity of the Whitehead minimization problem. *Internat. J. Algebra Comput.*, 17:1611-1634, 2007.

[17] W. Rudin. *Real and complex analysis*, 3rd edition, McGraw-Hill 1987.

[18] J.-P. Serre. *Arbres, Amalgames, SL2*, Astérisque, vol. 46. Soc. Math. France, 1977. English translation: *Trees*, Springer Monographs in Mathematics, Springer, 2003.

[19] J. R. Stallings. Topology of finite graphs. *Invent. Math.*, 71:551–565, 1983.

[20] N. W. M. Touikan. A fast algorithm for Stallings' folding process. *Internat. J. Algebra Comput.*, 16:1031–1045, 2006.