

Word Maps and Spectra of Random Graph Lifts

Nati Linial,¹ Doron Puder²

¹*School of Computer Science and Engineering, The Hebrew University of Jerusalem, Jerusalem, Israel; e-mail: nati@cs.huji.ac.il*

²*Institute of Mathematics, The Hebrew University of Jerusalem, Jerusalem, Israel; e-mail: doronpuder@gmail.com*

Received 7 June 2008; accepted 3 February 2009

Published online 12 January 2010 in Wiley InterScience (www.interscience.wiley.com).

DOI 10.1002/rsa.20304

ABSTRACT: We study here the spectra of random lifts of graphs. Let G be a finite connected graph, and let the infinite tree T be its universal cover space. If λ_1 and ρ are the spectral radii of G and T respectively, then, as shown by Friedman (Graphs Duke Math J 118 (2003), 19–35), in almost every n -lift H of G , all “new” eigenvalues of H are $\leq O(\lambda_1^{1/2} \rho^{1/2})$. Here we improve this bound to $O(\lambda_1^{1/3} \rho^{2/3})$. It is conjectured in (Friedman, Graphs Duke Math J 118 (2003) 19–35) that the statement holds with the bound $\rho + o(1)$ which, if true, is tight by (Greenberg, PhD thesis, 1995). For G a bouquet with $d/2$ loops, our arguments yield a simple proof that almost every d -regular graph has second eigenvalue $O(d^{2/3})$. For the bouquet, Friedman (2008), has famously proved the (nearly?) optimal bound of $2\sqrt{d-1} + o(1)$.

Central to our work is a new analysis of formal words. Let w be a formal word in letters $g_1^{\pm 1}, \dots, g_k^{\pm 1}$. The word map associated with w maps the permutations $\sigma_1, \dots, \sigma_k \in S_n$ to the permutation obtained by replacing for each i , every occurrence of g_i in w by σ_i . We investigate the random variable $X_w^{(n)}$ that counts the fixed points in this permutation when the σ_i are selected uniformly at random. The analysis of the expectation $\mathbb{E}(X_w^{(n)})$ suggests a categorization of formal words which considerably extends the dichotomy of primitive vs. imprimitive words. A major ingredient of our work is a second categorization of formal words with the same property. We establish some results and make a few conjectures about the relation between the two categorizations. These conjectures suggest a possible approach to (a slightly weaker version of) Friedman’s conjecture.

As an aside, we obtain a new conceptual and relatively simple proof of a theorem of A. Nica (Nica, Random Struct Algorithms 5 (1994), 703–730), which determines, for every fixed w , the limit distribution (as $n \rightarrow \infty$) of $X_w^{(n)}$. A surprising aspect of this theorem is that the answer depends only on the largest integer d so that $w = u^d$ for some word u . © 2010 Wiley Periodicals, Inc. Random Struct. Alg., 37, 100–135, 2010

Keywords: word map; random graphs; graph lifts; Alon’s conjecture; spectrum of graphs

Correspondence to: D. Puder

© 2010 Wiley Periodicals, Inc.

1. INTRODUCTION

Let $G = (V, E)$ be some fixed finite connected graph with $E = \{g_1, \dots, g_k\}$, and let $\lambda_1 \geq \lambda_2 \dots \geq \lambda_{|V|}$ be the eigenvalues of its adjacency matrix. We think of the edges as being oriented, though the results do not depend on the orientation chosen. We recall that $L_n(G)$ denotes the probability space of n -lifts of G (i.e., graphs that have an n -fold cover map onto G). A graph $H \in L_n(G)$, has vertex set $V \times \{1, \dots, n\}$. For every (oriented) edge $g_i = (u, v)$, we choose independently and uniformly a random permutation, $\sigma_i \in S_n$, and introduce an edge between (u, j) to $(v, \sigma_i(j))$ for all j . For background on lifts and random lifts, see [1–4]. In particular [21] shows how to construct regular graph lifts with a nearly optimal spectral gap.

The projection $\pi : H \rightarrow G$ given by $\pi(u, j) = u$ is the cover map associated with $H \in L_n(G)$. Every eigenfunction f of (the adjacency matrix of) G is pulled back by π to an eigenfunction $f \circ \pi$ of H . Therefore, every eigenvalue of G is also an eigenvalue of H . Such an eigenvalue of H is considered “old” while all other ones are “new” (including, possibly, duplicates of the old eigenvalues corresponding to “new” eigenfunctions). Let T be the (infinite) universal cover of G (and H). We consider $l_2(V(T))$, the Hilbert space of square-summable real functions on T 's vertices, i.e., the functions $f : V(T) \rightarrow \mathbb{R}$ with $\sum_{v \in V(T)} f^2(v) < \infty$. If A_T is the (infinite) adjacency matrix of T , we consider the linear operator corresponding to A_T . Namely, $f \rightarrow g$ where $g(v) = \sum_{u \sim v} f(u)$. This is a bounded linear operator on $l_2(V(T))$, and we denote its spectral radius by $\rho = \rho(T) = \rho(G)$. (Recall that, by definition, $\rho = \sup\{|\nu| : \nu \text{ is in the spectrum of } A_T\}$). Our main result is a new bound on H 's new eigenvalues in terms of λ_1 (the largest eigenvalue of G), and ρ .

Let $\mu_{\max} := \max\{|\mu| : \mu \text{ is a new eigenvalue of } H\}$. In [5], Friedman showed that $\mu_{\max} \leq \lambda_1^{1/2} \rho^{1/2} + o_n(1)$ for almost every H . We improve this bound as follows:

Theorem 1. *Almost every random n -lift H of G satisfies:*

$$\mu_{\max} \leq O(\lambda_1^{1/3} \rho^{2/3})$$

More specifically,

$$\mu_{\max} < \max\left(1, 3\left(\frac{\rho}{\lambda_1}\right)^{2/3}\right) \cdot \lambda_1^{1/3} \rho^{2/3} + \epsilon \quad (1)$$

almost surely for every $\epsilon > 0$.

Remark 2. Let Γ be a (not necessarily finite) connected graph, and let v be a vertex in Γ . We denote by $t_s(v)$ the number of closed paths of length s that start and end at v . It is well known that the spectral radius of Γ equals $\limsup_{s \rightarrow \infty} t_s(v)^{1/s}$. In particular, this value is independent of the choice of v . (These facts may be proven by an easy variation on the proof of Proposition 3.1 in [6].) Returning to our notation, observe that a path that starts and ends at a vertex $v \in V(T)$ is projected to a path of the same length that starts and ends at the corresponding vertex of G . Consequently, $\rho \leq \lambda_1$ always holds.

Our proof of Theorem 1 suggests an approach that may lead to an even better (nearly optimal) bound $\mu_{\max} \leq O(\rho)$. This plan depends on an unresolved conjecture that we present shortly. It follows from Lubotzky and Greenberg [7] that this statement cannot hold

with any bound smaller than $\rho - o(1)$. It is shown in [7] that for every infinite tree T and for every $\epsilon > 0$, there exists a constant $c = c(\epsilon, T) > 0$, such that if T is the universal covering space of a finite graph Γ , then at least $c|V(\Gamma)|$ of Γ 's eigenvalues exceed $\rho(T) - \epsilon$. Thus for G fixed, and for $\epsilon > 0$ there exists an n_ϵ such that $\mu_{\max}(H) > \rho(G) - \epsilon$ for every $n \geq n_\epsilon$ and every $H \in L_n(G)$. (Since the infinite d -regular tree T_d has spectral radius $\rho(T_d) = 2\sqrt{d-1}$ ([8]), this extends the Alon-Boppana bound [9] that $\lambda_2 \geq 2\sqrt{d-1} - o_n(1)$ for every n -vertex d -regular graph).

The ‘‘permutation model’’ of random d -regular graphs (for d even) is a special case of random lifts of graphs. In the permutation model, n -vertex d -regular graphs are generated through a random n -lift of a bouquet of $d/2$ loops. Thus, our result, as well as Friedman’s, extend earlier work on random d -regular graphs. Namely, Friedman’s result states that $\lambda(G) \leq \sqrt{2d\sqrt{d-1}} + o(1)$ for almost every d -regular graph, which is a slight improvement of an old result of Broder and Shamir [10]. In this special case, Theorem 1 states that $\lambda(G) = O(d^{2/3})$ holds almost surely, and the tentative proof strategy mentioned above would yield $\lambda(G) = O(d^{1/2})$ almost surely. In particular, we obtain the following corollary (which is, of course, substantially weaker than the one proven in [11]):

Corollary 3. *If G is d -regular and $d \geq 107$, then*

$$\mu_{\max} < \lambda_1^{1/3} \rho^{2/3} + \epsilon = [4d(d-1)]^{1/3} + \epsilon.$$

almost surely for every $\epsilon > 0$.

A major tool in this area is the *Trace Method* which goes back to Wigner [12]. It is based on a natural connection between graph spectra and word-maps. This approach underlies the work of Broder-Shamir [10] and of Friedman [5].

Let w be a (not necessarily reduced) formal word in the letters $g_1^{\pm 1}, \dots, g_k^{\pm 1}$. For every k -tuple $(\sigma_1, \dots, \sigma_k)$ of permutations in S_n , we form the permutation $w(\sigma_1, \dots, \sigma_k) \in S_n$, by replacing g_1, \dots, g_k with $\sigma_1, \dots, \sigma_k$ in the expression of w . For instance, if $w = g_2 g_1^2 g_2^{-1} g_3$, then $w(\sigma_1, \sigma_2, \sigma_3) = \sigma_2 \sigma_1^2 \sigma_2^{-1} \sigma_3$. The correspondence between w and the permutation $w(\sigma_1, \dots, \sigma_k) \in S_n$ is called a word map. Such maps can be evaluated in groups other than S_n as well (we refer to this briefly in Section 2). The study of word maps has a long history in group theory (see [13] and the references therein). Our perspective is mostly combinatorial and probabilistic.

For fixed formal word w we denote by $X_w^{(n)}$ a random variable on S_n^k which is defined by:

$$X_w^{(n)}(\sigma_1, \dots, \sigma_k) = \# \text{ of fixed points of } w(\sigma_1, \dots, \sigma_k). \tag{2}$$

Now let H be an n -lift of G and let A_G, A_H be the adjacency matrices of G, H resp. We denote by μ a running index for the ‘‘new’’ eigenvalues of H . For every $t \geq 1$, the trace of A_H^t equals the number of closed paths of length t in H . This number can also be expressed as $(\sum_\mu \mu^t) + (\sum_{i=1}^{|V(G)|} \lambda_i^t)$. Therefore, for t even we obtain:

$$\mu_{\max}^t \leq \sum_\mu \mu^t = \left(\sum_\mu \mu^t + \sum_{i=1}^{|V(G)|} \lambda_i^t \right) - \sum_{i=1}^{|V(G)|} \lambda_i^t = \text{tr}(A_H^t) - \text{tr}(A_G^t)$$

Every closed path in H is a lift of a closed path in G . Since the edges of G are labeled g_1, \dots, g_k , every (closed) path in G corresponds to some formal word w in $g_1^{\pm 1}, \dots, g_k^{\pm 1}$. The

closed lifts of this path are in 1 : 1 correspondence with the fixed points of $w(\sigma_1, \dots, \sigma_k)$, so that their number is $X_w^{(n)}(\sigma_1, \dots, \sigma_k)$. Let $\mathcal{CP}_t(G)$ denote the set of all closed paths of length t in G (i.e., $|\mathcal{CP}_t(G)| = \text{tr}(A_G^t)$). The above inequality now becomes:

$$\mu_{\max}^t \leq \text{tr}(A_H^t) - \text{tr}(A_G^t) = \sum_{w \in \mathcal{CP}_t(G)} [X_w^{(n)}(\sigma_1, \dots, \sigma_k) - 1]$$

Taking expectations, we obtain:

$$\mathbb{E}(\mu_{\max}^t) \leq \sum_{w \in \mathcal{CP}_t(G)} [\mathbb{E}(X_w^{(n)}) - 1] \quad (3)$$

Equation (3) shows the significance of $\mathbb{E}(X_w^{(n)}) - 1$ in the study of spectra in random lifts. If we let $\Phi_w(n)$ equal $\frac{\mathbb{E}(X_w^{(n)}) - 1}{n}$, it turns out that for every w , Φ_w can be expressed as a power series in $\frac{1}{n}$. Namely,

$$\Phi_w(n) = \frac{\mathbb{E}(X_w^{(n)}) - 1}{n} = \sum_{i=0}^{\infty} a_i(w) \frac{1}{n^i}, \quad (4)$$

where the $a_i(w)$ are integers. (This fact appears in [14], but we present (Lemma 4) a new and simpler proof). This induces a categorization of words in \mathbf{F}_k , the free group with generators g_1, \dots, g_k . Namely, $\phi(w)$ is the smallest index i for which $a_i(w) \neq 0$, or ∞ if $\mathbb{E}(X_w^{(n)}) \equiv 1$.

We consider next (Section 2.2) another categorization of the words in \mathbf{F}_k , which does not depend on a word-map to specific groups such as S_n . To every $w \in \mathbf{F}_k$ we associate $\beta(w)$ which is a non-negative integer or ∞ . The categorizations induced by both $\phi(w)$ and $\beta(w)$ extend the dichotomy between primitive and imprimitive words (Recall that w is called imprimitive if $w = u^d$ for some $u \in \mathbf{F}_k$ and $d \geq 2$). Without going into the (somewhat lengthy) definition, let us say that the main step in both [10] and [5] can be viewed as the observation that for $i = 0, 1$, $\phi(w) = i$ iff $\beta(w) = i$. Our aforementioned conjecture states in this language that $\phi(w) = \beta(w)$ for every word w (Conjecture 15). These relations between ϕ and β allow us to bound the sum in the r.h.s of (3): We can bound the contribution of w to this sum in terms of $\phi(w)$. This is complemented by bounding the number of words $w \in \mathcal{CP}_t(G)$ with a given value of $\beta(w)$ which bound is stated in terms of ρ . Indeed, a key step in the present paper (Lemma 20) can be interpreted as a partial proof of the claim that $\beta(w) = 2$ iff $\phi(w) = 2$.

As an aside to our work we obtain a new conceptual and relatively simple proof of a theorem of A. Nica [14], which determines for every fixed w the limit distribution of $X_w^{(n)}$ as $n \rightarrow \infty$ (see Theorem 25). We carry out a similar analysis for all higher moments of $X_w^{(n)}$, and use the method of moments to derive Nica's result. A surprising aspect of this theorem is that the limit distribution depends *only* on the largest integer d such that $w = u^d$ for some $u \in \mathbf{F}_k$. Nica's full result (which we derive by the same argument) concerns not only fixed points but applies just as well to the number of L -cycles for any fixed $L \geq 1$.

The paper is arranged as follows. We begin (Section 2) with our analysis of word maps and introduce the two new categorizations of formal words. Based on this analysis, we prove Theorem 1 in Section 3. In Section 4 we deal with the distribution of the number of L -cycles in $w(\sigma_1, \dots, \sigma_k)$ and present our new proof for Nica's Theorem. For the reader interested only in this new proof, this section is mostly self-contained with only occasional references to earlier parts of the article. There are numerous open problems and conjectures raised in this article, some of which we collect in Section 5.

2. WORD MAPS AND THE LEVEL OF PRIMITIVITY OF A WORD

We begin with some notation. We denote by Σ_k the set of all finite words in letters $g_1^{\pm 1}, \dots, g_k^{\pm 1}$ (though we occasionally use the letters a, b, c, \dots instead). The quotient of Σ_k modulo reduction of words is \mathbf{F}_k , the set of elements of the free group on k generators. For instance, the set $\mathcal{CP}_t(G)$ introduced before Eq. (3) is a subset of Σ_k , so it may contain different words which are equivalent as members of \mathbf{F}_k .

For every group P and every word $w \in \Sigma_k$, the word map $w : P^k \rightarrow P$ is defined by substitutions and composition. For $p_1, \dots, p_k \in P$, the element $w(p_1, \dots, p_k)$ is obtained by substituting p_i for each occurrence of g_i in w and this for every $1 \leq i \leq k$. Clearly, the word map of w is invariant under reductions, so we can regard w as an element in \mathbf{F}_k .

Most research on word maps concerns the range of certain fixed words w in a group P . More specifically, for P finite, it is of interest to understand the distribution induced on P by the word map $w : P^k \rightarrow P$ and the uniform distribution on P^k . This perspective makes it natural to consider an equivalence relation on words (beyond that of reduction).

To introduce this equivalence relation, we now recall some simple terminology from combinatorial group theory. There are three elementary Nielsen transformations defined on the free group \mathbf{F}_k : (i) Exchanging any two generators g_i and g_j for some $i \neq j$, (ii) Replacing some g_i with g_i^{-1} , (iii) Replacing any g_i by $g_i g_j$, for some $i \neq j$. We recall (e.g. [15], Theorem 3.2) that these transformations generate the automorphism group \mathbf{A}_k of \mathbf{F}_k . We say that two words $w_1, w_2 \in \mathbf{F}_k$ are equivalent, and denote $w_1 \sim w_2$, if they belong to the same orbit of \mathbf{A}_k . Obviously, “ \sim ” is an equivalence relation. It is quite clear that for every finite group P , every two equivalent words $w_1, w_2 \in \mathbf{F}_k$ induce the same distribution on P . We do not know whether the converse is true as well, and we state a specific problem (Conjecture 17) in this vein.

Given a word w and the distribution it induces on a group P , it is of interest to consider how far this distribution is from the uniform distribution. The two gradings of words $\phi(\cdot)$ and $\beta(\cdot)$ can be viewed as our attempts to capture this intuition. Both parameters associate a non-negative integer or ∞ with every $w \in \mathbf{F}_k$, and they tend to grow as the aforementioned distance decreases. Of course, the distribution furthest away from the uniform distribution corresponds to the word $w = 1$. Indeed, $\beta(w) = 0$ iff $\phi(w) = 0$ iff $w = 1$ (Lemma 12). Also, $\beta(w) = 1$ iff $\phi(w) = 1$ iff w is imprimitive (Lemma 13). In this case the range of w contains only powers of certain exponent. (Recall that $w \in \mathbf{F}_k$ is called imprimitive if $w = u^d$ for some word u and $d \geq 2$.) At the other end of the scale, both $\beta(w)$ and $\phi(w)$ equal ∞ for words that are \sim -equivalent to a single-letter word. Clearly, such words always induce the uniform distribution on P . Another important property is that both β and ϕ are invariant under “ \sim ”.

The definition of $\phi(w)$ depends on the word map for the symmetric group S_n (see Section 2.1). The definition of $\beta(w)$ is more involved and is based on a certain analysis of w as a formal word without reference to groups (see Section 2.2). In fact, we have arrived at our definition of $\beta(w)$ through our study of $\phi(w)$. Some proven results and extensive numerical simulations suggest that $\phi(w) = \beta(w)$ for every w (Section 2.3). One advantage of the parameter β over ϕ is that we can bound the number of words with fixed value of β - see Section 3.1.

Since both $\phi(w)$ and $\beta(w)$ offer an extension of the primitive-imprimitive dichotomy for words, we tend to think of them as quantifying “the level of primitivity” of a word (this level is 0 if $w = 1$, it is 1 if w is imprimitive, and ≥ 2 for primitive words - see Section 2.3).

2.1. The Word Map $w : S_n^k \rightarrow S_n$ and $\phi(\cdot)$

In this section we present a method to calculate the expectation of $X_w^{(n)}$, the number of fixed points in $w(\sigma_1, \dots, \sigma_k)$ (defined in (2)). We count the fixed points in $w(\sigma_1, \dots, \sigma_k)$ for all k -tuples $(\sigma_1, \dots, \sigma_k) \in S_n^k$ and divide by $(n!)^k$. This calculation is carried out through a certain categorization of all fixed points. We note that similar considerations appear in [14].

We begin with some technicalities. Let $w = g_{i_1}^{\alpha_1} g_{i_2}^{\alpha_2} \dots g_{i_m}^{\alpha_m} \in \Sigma_k$, where $i_1, \dots, i_m \in \{1, \dots, k\}$ and $\alpha_1, \dots, \alpha_m \in \{-1, 1\}$, and let $\sigma_1, \dots, \sigma_k \in S_n$. Assume that $s_0 \in \{1, \dots, n\}$ is a fixed point of $w(\sigma_1, \dots, \sigma_k)$. Associated with s_0 is the following closed trail:

$$s_0 \xrightarrow{\sigma_{i_1}^{\alpha_1}} s_1 \xrightarrow{\sigma_{i_2}^{\alpha_2}} s_2 \xrightarrow{\sigma_{i_3}^{\alpha_3}} \dots \xrightarrow{\sigma_{i_{m-1}}^{\alpha_{m-1}}} s_{m-1} \xrightarrow{\sigma_{i_m}^{\alpha_m}} s_0$$

with $s_1, \dots, s_{m-1} \in \{1, \dots, n\}$, and $s_b \bmod m = \sigma_{i_b}^{\alpha_b}(s_{b-1})$ ($b = 1, \dots, m$).

Note that for the sake of convenience, we compose permutations from left to right. This is inconsequential for the analysis of the variables $X_w^{(n)}$ since $w(\sigma_1, \dots, \sigma_k)$ with left-to-right composition is the inverse of $w(\sigma_1^{-1}, \dots, \sigma_k^{-1})$ with right-to-left composition, and thus both have the same cycle structure.

We categorize fixed points according to their associated trails. Let $s_0 \rightarrow \dots \rightarrow s_{m-1} \rightarrow s_0$ and $s'_0 \rightarrow \dots \rightarrow s'_{m-1} \rightarrow s'_0$ be the trails of the fixed points s_0 and s'_0 in $w(\sigma_1, \dots, \sigma_k)$ and $w(\sigma'_1, \dots, \sigma'_k)$, respectively. These two trails are placed in the same category, if they have the same coincidence pattern, that is, if for every $i, j \in \{0, \dots, m-1\}$, $s_i = s_j \Leftrightarrow s'_i = s'_j$.

Each closed trail consists of m integers, or points, possibly with repetitions, and each category of trails uniquely corresponds to some partition of these m points. Consequently, there are at most $B(m)$ categories, where the m -th Bell Number, $B(m)$, is the number of partitions of an m element set. This bound is, however, not tight. For instance, if $s_{a+1} = \sigma_j(s_a)$ and $s_{b+1} = \sigma_j^{-1}(s_b)$, then $s_a = s_{b+1} \Leftrightarrow s_{a+1} = s_b$. Therefore, not every partition corresponds to a realizable category of trails.

It is convenient to associate a directed edge-colored graph Γ with each category. Vertices in Γ correspond to blocks in the partition that defines Γ 's category. In other words, Γ has as many vertices as the number of distinct integers among the s_b 's. There is a directed edge labeled j from one vertex (=block) to another, whenever the trails include an arrow labeled σ_j (resp. σ_j^{-1}) from a point in the first (second) block to a point in the second (first) one.

Of special importance is the graph associated with the finest possible partition which we call the universal graph. Two points in the trail are merged in this partition if and only if they are merged in every realizable partition. If w is cyclically reduced (i.e., no two consecutive letters are inverses, nor are the first and last letter), this is the partition where all m points in the trail are distinct. To illustrate, we draw in Fig. 1 the universal graph of three different words.

All other graphs are now easily derived as quotients of the universal graph, or partitions of its vertices. A quotient graph has one vertex per each block in the partition. It has a j -labeled directed edge (j -edge for short) from block v_1 to block v_2 , if the universal graph contains a j -edge from a vertex in v_1 to a vertex in v_2 . A quotient is not realizable if it contains two distinct j -edges with common head and different tails or vice-versa. We denote by \mathcal{Q}_w the set of all realizable quotients.

To illustrate, we draw (Fig. 2) all the realizable quotient graphs of the universal graph of the commutator word (one of the graphs in Fig. 1). Note that a four element set has 15 partitions (the fourth Bell number, $B(4) = 15$), of which only 7 are realizable in this case.

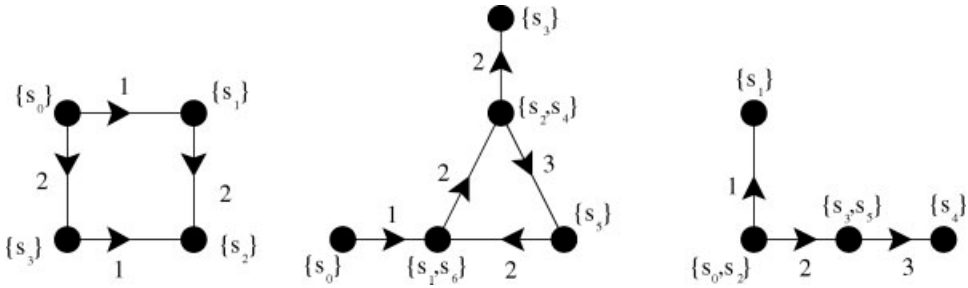


Fig. 1. From left to right: the universal graphs of $w = g_1g_2g_1^{-1}g_2^{-1}$ (the commutator word), of $w = g_1g_2g_2g_2^{-1}g_3g_2g_1^{-1}$, and of $w = g_1g_1^{-1}g_2g_3g_3^{-1}g_2^{-1}$.

These graphs suggest a simple formula for the number of fixed points in each category. Let v_Γ , (e_Γ) be the number of vertices (edges) in the graph Γ , and e_Γ^j be the number of j -edges (and so $e_\Gamma = \sum_{j=1}^k e_\Gamma^j$). To count the number of fixed points in Γ 's category, or the number of realizations of Γ , we first label Γ 's vertices by distinct numbers from $\{1, \dots, n\}$ (i.e., specify the values of s_0, \dots, s_{m-1}). This can be done in $n(n-1) \dots (n-v_\Gamma+1)$ ways. For each $j = 1, \dots, k$ there are $(n-e_\Gamma^j)!$ permutations that are consistent with the e_Γ^j values in the permutation σ_j that are already determined. Thus, the number of realizations of Γ is:

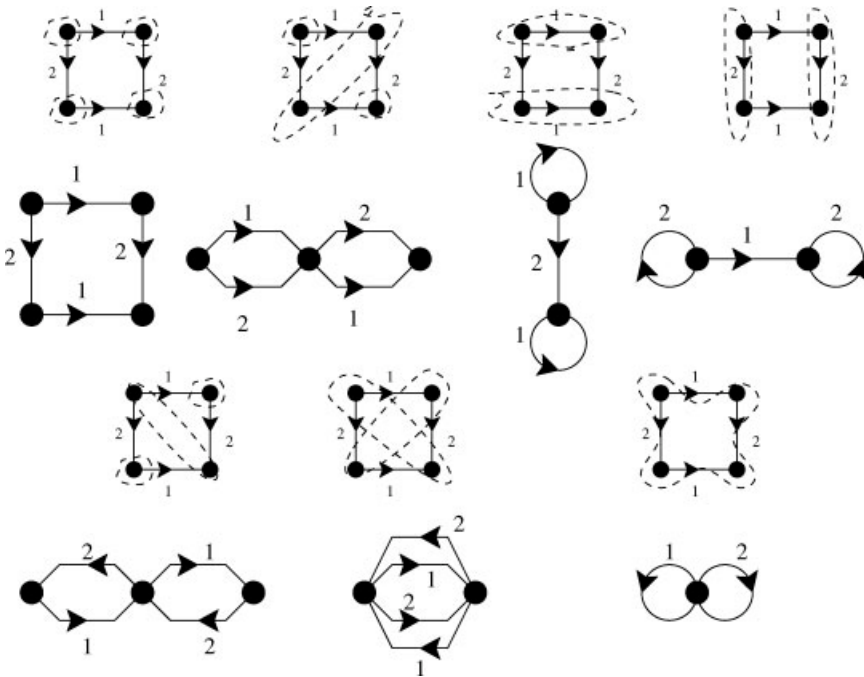


Fig. 2. The 7 different graphs in \mathcal{Q}_w , the set of graphs of categories of fixed points when $w = g_1g_2g_1^{-1}g_2^{-1}$. Each graph is drawn together with the partition of the universal graph which yields it. (We do not specify the block corresponding to each vertex).

$$N_\Gamma(n) = n(n-1) \dots (n - v_\Gamma + 1) \prod_{j=1}^k [n - e_\Gamma^j]! \tag{5}$$

A formula for the expectation of $X_w^{(n)}$ is now at hand:

$$\begin{aligned} \mathbb{E}(X_w^{(n)}) &= \frac{1}{(n!)^k} \sum_{\sigma_1, \dots, \sigma_k \in S_n} X_w^{(n)}(\sigma_1, \dots, \sigma_k) = \frac{1}{(n!)^k} \sum_{\Gamma \in \mathcal{Q}_w} N_\Gamma(n) \\ &= \sum_{\Gamma \in \mathcal{Q}_w} \frac{n(n-1) \dots (n - v_\Gamma + 1)}{\prod_{j=1}^k n(n-1) \dots (n - e_\Gamma^j + 1)} \\ &= \sum_{\Gamma \in \mathcal{Q}_w} \left(\frac{1}{n}\right)^{e_\Gamma - v_\Gamma} \frac{\prod_{l=1}^{v_\Gamma-1} \left(1 - \frac{l}{n}\right)}{\prod_{j=1}^k \prod_{l=1}^{e_\Gamma^j-1} \left(1 - \frac{l}{n}\right)} \end{aligned} \tag{6}$$

(The third equality holds only for n that is $\geq e_\Gamma^j$ for all j and Γ .)

We illustrate these calculations for w the commutator word $g_1 g_2 g_1^{-1} g_2^{-1}$. If we go over the graphs in Fig. 2 in clockwise order starting at the upper-left graph, (6) becomes:

$$\begin{aligned} \mathbb{E}(X_w^{(n)}) &= \frac{n(n-1)(n-2)(n-3)}{[n(n-1)][n(n-1)]} + \frac{n(n-1)(n-2)}{[n(n-1)][n(n-1)]} + \\ &+ \frac{n(n-1)}{[n(n-1)][n]} + \frac{n(n-1)}{[n][n(n-1)]} + \frac{n}{[n][n]} + \\ &+ \frac{n(n-1)}{[n(n-1)][n(n-1)]} + \frac{n(n-1)(n-2)}{[n(n-1)][n(n-1)]} = \frac{n}{n-1} \end{aligned}$$

In (4) we defined $\Phi_w(n)$ to be $\frac{\mathbb{E}(X_w^{(n)})-1}{n}$. The following lemma states an important property of Φ_w :

Lemma 4. *Associated with every $w \in \mathbf{F}_k$ is a power series $\sum_{i=0}^\infty a_i(w)x^i$ that has a positive radius of convergence where the coefficients $a_i(w)$ are integers. In particular, for every w and for every sufficiently large n , there holds $\Phi_w(n) = \sum_{i=0}^\infty a_i(w) \frac{1}{n^i}$.*

Proof. As we saw in (6), for large enough n (e.g., $n \geq |w|$ suffices),

$$\Phi_w(n) = -\frac{1}{n} + \sum_{\Gamma \in \mathcal{Q}_w} \left(\frac{1}{n}\right)^{e_\Gamma - v_\Gamma + 1} \frac{\prod_{l=1}^{v_\Gamma-1} \left(1 - \frac{l}{n}\right)}{\prod_{j=1}^k \prod_{l=1}^{e_\Gamma^j-1} \left(1 - \frac{l}{n}\right)}. \tag{7}$$

Since every Γ is a connected graph, we have $e_\Gamma - v_\Gamma + 1 \geq 0$. The lemma follows when we individually consider the expression corresponding to each Γ . (See the proof of Lemma 19 for a thorough analysis of these expressions). ■

By Lemma 4, the contribution of every $w \in \mathcal{CP}_i(G)$ in (3), is $\frac{a_i(w)+o(1)}{n^{i-1}}$ for some nonnegative integer i . This induces the following useful grading of words:

$$\phi(w) := \begin{cases} \text{the smallest integer } i \text{ with } a_i(w) \neq 0 & \text{if } \mathbb{E}(X_w^{(n)}) \not\equiv 1 \\ \infty & \text{if } \mathbb{E}(X_w^{(n)}) \equiv 1 \end{cases} \tag{8}$$

Recall that although the construction of Φ_w depends on the actual representation of w (a reduction of w usually changes \mathcal{Q}_w), this function captures some features of the distribution of the image of w on S_n . Thus, Φ_w as well as $\phi(w)$, are invariant not only under reduction, but also under “ \sim ”.

With this new terminology we can reinterpret both [10] and [5] as follows: Both papers rely on the fact that $\phi(w) = 0$ iff w reduces to the empty word, and that $\phi(w) = 1$ iff w is imprimitive (see Lemmas 12 and 13). To study the new spectrum of random lifts, one proceeds as follows: The number of words of these two kinds can be bounded in terms of ρ (the spectral radius of the universal cover) alone (and does not depend on λ_1 , the spectral radius of the base graph). Finally, the rest of the words (which are, in fact, the vast majority) contribute to the summation in (3) only $O(\frac{1}{n})$ each.

Here we extend these ideas and seek (with partial success) a similar characterization for all words with $\phi(w) = i$ for fixed i . Our analysis of the new spectrum extends these arguments and refines them. We further split the above-mentioned third set and attain an improved bound on the contributions of these subsets to the sum in Eq. (3).

2.2. More on the Level of Primitivity: $\beta(\cdot)$

We now start our second attempt at capturing the “level of primitivity” of w by means of the parameter $\beta(w)$. We begin with some definitions.

Recall the notion of a trail from Section 2.1. Consider then the following trail through $w = g_{i_1}^{\alpha_1} g_{i_2}^{\alpha_2} \dots g_{i_{|w|}}^{\alpha_{|w|}}$ (this time, the s_i 's should not be thought of as numbers but rather as abstract symbols, and the trail deliberately ends with $s_{|w|}$ and not with s_0):

$$s_0 \xrightarrow{g_{i_1}^{\alpha_1}} s_1 \xrightarrow{g_{i_2}^{\alpha_2}} s_2 \xrightarrow{g_{i_3}^{\alpha_3}} \dots \xrightarrow{g_{i_{|w|-1}}^{\alpha_{|w|-1}}} s_{|w|-1} \xrightarrow{g_{i_{|w|}}^{\alpha_{|w|}}} s_{|w|}$$

As in our former definition of a realizable category of trails, we say that a partition of $\{s_0, \dots, s_{|w|}\}$ is realizable if the following conditions hold: Whenever $i_h = i_l$ and $\alpha_h = \alpha_l$, s_{h-1} is in the same block with s_{l-1} (we denote $s_{h-1} \equiv s_{l-1}$) iff $s_h \equiv s_l$. Likewise, whenever $i_h = i_l$ and $\alpha_h = -\alpha_l$, $s_{h-1} \equiv s_l \Leftrightarrow s_h \equiv s_{l-1}$. (In other words, a partition is realizable whenever it traces a trail of some point, fixed or not, through $w(\sigma_1, \dots, \sigma_k)$ for some $\sigma_1, \dots, \sigma_k \in S_n$ and some n .)

As before, to each realizable partition of $\{s_0, \dots, s_{|w|}\}$ corresponds a directed edge-colored graph Γ , which is a quotient of the graph of the trail. According to our former notation, $\Gamma \in \mathcal{Q}_w$ whenever $s_0 \equiv s_{|w|}$. We now concentrate on the number of pairs of s_i 's that should be merged in order to yield a specific Γ .

Definition 5. Let Γ be the quotient graph corresponding to some realizable partition of $\{s_0, \dots, s_{|w|}\}$. We say that set of pairs $\{\{s_{j_1}, s_{k_1}\}, \dots, \{s_{j_r}, s_{k_r}\}\}$ generates Γ , if Γ corresponds to the finest realizable partition in which $s_{j_i} \equiv s_{k_i} \forall i = 1, \dots, r$.

Example 1. Let us return to the commutator word $w = g_1 g_2 g_1^{-1} g_2^{-1}$. The trail here is

$$s_0 \xrightarrow{g_1} s_1 \xrightarrow{g_2} s_2 \xrightarrow{g_1^{-1}} s_3 \xrightarrow{g_2^{-1}} s_4.$$

In Fig. 3 we revisit the seven quotient graphs from Fig. 2 (the seven graphs in \mathcal{Q}_w), and specify a smallest generating set for each of them.

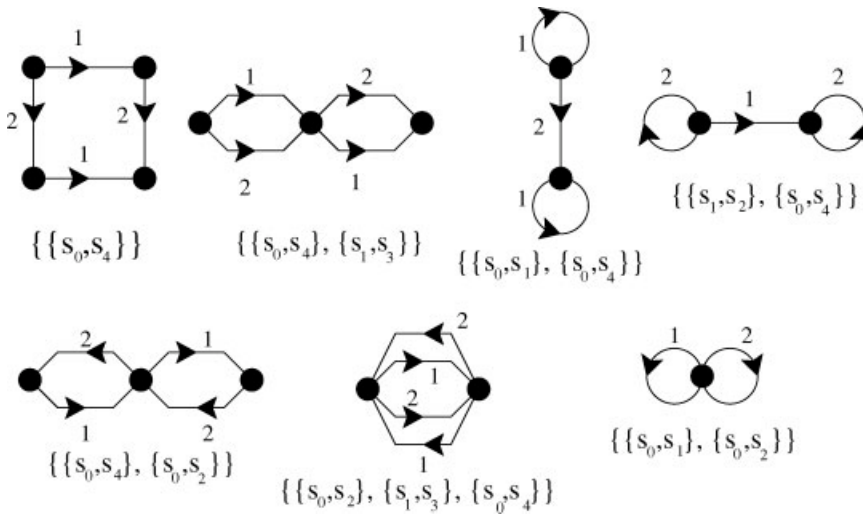


Fig. 3. A smallest generating set of each of the seven graphs in \mathcal{Q}_w , the set of graphs of categories of fixed points when $w = g_1 g_2 g_1^{-1} g_2^{-1}$.

We denote by $\chi(\Gamma) = e_\Gamma - v_\Gamma + 1$ the Euler characteristic of Γ . It turns out that there is a tight connection between $\chi(\Gamma)$ and the smallest size of a generating set of Γ :

Lemma 6. *Let Γ be the quotient graph corresponding to some realizable partition of $\{s_0, \dots, s_{|w|}\}$. The smallest cardinality of a generating set for Γ is $\chi(\Gamma)$.*

Proof. It is quite easy to construct a set \hat{S} of $\chi(\Gamma)$ pairs that generates Γ . To this end, we adopt the original terminology of [10]. As we follow the path of w through Γ , each move has one of three types. In a **free** step we traverse a new edge and reach a new vertex, and so one vertex and one edge are added to the partial graph. In a **coincidence** a new edge leads us to an “old” vertex, so we gain one new edge and no new vertices. In a **forced** step we traverse an old edge (necessarily to an old vertex), so the numbers of vertices and edges remain unchanged. Consequently, $\chi(\Gamma)$ equals the number of coincidences in this walk. We introduce into \hat{S} one pair for each coincidence. If the j -th step is a coincidence in which we reach a vertex in Γ representing the block s_{i_1}, \dots, s_{i_r} ($i_1 \leq \dots \leq i_r < j$), we add $\{s_j, s_{i_1}\}$ to \hat{S} . Clearly, the cardinality of \hat{S} is $\chi(\Gamma)$ and it generates Γ .

To see that Γ has no generating set smaller than \hat{S} , consider \mathcal{S} , the collection of all generating sets of Γ of the smallest possible cardinality. We claim that \hat{S} is the lexicographically first member of \mathcal{S} . Concretely, write each pair under consideration as $\{s_i, s_j\}$ with $i > j$. Now sort the pairs in each $S \in \mathcal{S}$ in increasing lexicographic order and let $T \in \mathcal{S}$ be the lexicographically first member of \mathcal{S} . Our claim is that $T = \hat{S}$. Observe that if $\{s_i, s_j\} \in T$ then there is no index $h < i$ with $h \neq j$ with $\{s_h, s_j\} \in T$. Otherwise we could replace the pair $\{s_i, s_j\}$ with the pair $\{s_h, s_j\}$ and generate the same quotient as does T with a lexicographically smaller set of pairs.

It is helpful to consider for each $1 \leq i \leq |w|$ the graph Γ_i that is the quotient of $s_0 \rightarrow \dots \rightarrow s_i$ generated by the initial segments of T that includes only those pairs in T where both indices are $\leq i$. We claim that Γ_{i-1} is a subgraph of Γ_i for all i . If there is no

TABLE 1. The Number of Graphs in \mathcal{Q}_w in Any Specified Character and Type, when $w = g_1 g_2 g_1^{-1} g_2^{-1}$

χ	0	1	2	3	4
Type A	0	1	4	1	0
Type B	0	0	1	0	0

pair in T where s_i is the larger member, this is clear. If $\{s_i, s_j\}$ is in T then the index j is uniquely defined by the above remark. In this case, we need to show that the identification of s_i with s_j does not entail any additional identification (which would violate the inclusion $\Gamma_{i-1} \subseteq \Gamma_i$). How can such an identification occur? Only if the label of the edge (s_{i-1}, s_i) agrees with that of an edge e incident with the block that contains s_j (and has the correct orientation). Let s_ν be a member of the block at the other end of e . Clearly $\nu < i$. But now again we can generate the quotient generated by T by a lexicographically smaller set, i.e., replace the pair $\{s_i, s_j\}$ by $\{s_{i-1}, s_\nu\}$.

We can again recognize the three types of steps by observing how the graphs Γ_i grow at each step. If Γ_i stays unchanged, this is a forced move. If only an edge is added, this is a coincidence and in a free step one vertex and one edge are added. It is exactly at each coincidence step that vertices at T get merged. But this is precisely what we did in constructing \hat{S} , so that $\hat{S} = T$, as claimed. ■

The following categorization of the quotient graphs in \mathcal{Q}_w turns out to be very useful:

Definition 7. Let w be a word in Σ_k . We say that a quotient graph $\Gamma \in \mathcal{Q}_w$ has type A, if one of the smallest generating sets for Γ contains the pair $\{s_0, s_{|w|}\}$. Otherwise, we say Γ has type B.

Given a word w , we classify the graphs in \mathcal{Q}_w according to their characteristics and type. Note that $\chi(\Gamma) \leq |w|$, since every Γ has at most $|w|$ edges. We illustrate this again with the seven graphs of the commutator word: The figure-eight graph with one vertex and two edges has type B. The other six graphs have type A (their generating sets specified in Fig. 3 purposely include $\{s_0, s_4\}$). Table 1 shows the whole census.

We are now ready for the second definition for a word’s “level of primitivity”. Let w be a word in Σ_k . We define $\beta(w)$ to be the smallest characteristic of a type-B graph in \mathcal{Q}_w . Namely,

$$\beta(w) := \min(\{\chi(\Gamma) : \Gamma \in \mathcal{Q}_w \text{ and } \Gamma \text{ has type B}\} \cup \{\infty\}) \tag{9}$$

Example 2. For the commutator word $\beta(w) = 2$.

In the next few lemmas we establish several properties of $\beta(\cdot)$ which are clearly desirable. Among others, $\beta(\cdot)$ is proved to be invariant under reductions, and hence well defined as a function on \mathbf{F}_k .

Lemma 8. $\beta(\cdot)$ is invariant under cyclic shifts.

Proof. Let $w \in \Sigma_k$, and let w' be some cyclic shift of w . There is an obvious bijection between \mathcal{Q}_w and $\mathcal{Q}_{w'}$, obtained by applying the appropriate cyclic shift on the indices of

the s_i 's in the blocks' names. (E.g, if w' is attained from w by a right cyclic shift of two positions, replace each label s_i in w by s'_j in w' where $j \equiv i + 2 \pmod{|w|}$.) We claim that in addition, each $\Gamma \in \mathcal{Q}_w$ has the same type in \mathcal{Q}_w as its matching quotient Γ' in $\mathcal{Q}_{w'}$. It suffices to show that if Γ has type A in \mathcal{Q}_w , then Γ' have type A in $\mathcal{Q}_{w'}$ (It then follows by symmetry that Γ has type A in \mathcal{Q}_w iff Γ' has type A in $\mathcal{Q}_{w'}$).

Let S be a smallest generating set of Γ with $\{s_0, s_{|w|}\} \in S$. Given S , we can generate Γ gradually, through a series of quotients. To proceed from the quotient Γ_i to the next quotient, we add a pair $\{s_j, s_r\} \in S$. To determine Γ_{i+1} we carry out all necessary identifications and only them. Formally, Γ_{i+1} is the finest realizable quotient of Γ_i in which the pair $\{s_j, s_r\}$ is merged. It is easily verified that the final quotient is Γ regardless of the order at which the pairs in S are introduced. But merging $\{s_0, s_{|w|}\}$ in w , and merging $\{s'_0, s'_{|w|}\}$ in w' , yield equivalent quotient graphs (these are the universal graphs of w and of w' , as defined in Section 2.1 around Fig. 1). We can now proceed by applying the same series of quotients as described above on both universal graphs, to conclude that Γ' is of type A with respect to w' as well. ■

Lemma 9. *Let $w \in \Sigma_k$, and let w' be its reduced form. Then $\beta(w) = \beta(w')$.*

Proof. We need to show that β does not change when a letter and its inverse are inserted consecutively into a word. But Lemma 8 says that β is invariant under cyclic shifts, so it suffices to show that $\beta(w) = \beta(w')$ for $w = g_1^{\alpha_1} g_2^{\alpha_2} \dots g_m^{\alpha_m}$ and $w' = g_1^{\alpha_1} g_2^{\alpha_2} \dots g_m^{\alpha_m} g_j g_j^{-1}$. To see this, we define below for every quotient $\Gamma \in \mathcal{Q}_w$ the subset $\epsilon(\Gamma) \subseteq \mathcal{Q}_{w'}$ of all consistent extensions of Γ . The set $\epsilon(\Gamma)$ contains a certain member $\delta(\Gamma)$ which plays a special role. The relevant properties of ϵ and δ are:

- The union of the images of ϵ is all of $\mathcal{Q}_{w'}$.
- If $\Gamma_1 \neq \Gamma_2 \in \mathcal{Q}_w$, then $\epsilon(\Gamma_1)$ and $\epsilon(\Gamma_2)$ are disjoint.
- If $\Gamma \in \mathcal{Q}_w$ has type A, then all members in $\epsilon(\Gamma)$ have type A.
- For every $\Gamma \in \mathcal{Q}_w$, one graph $\delta(\Gamma) \in \epsilon(\Gamma)$ has the same Euler characteristic and the same type as Γ . All other members in $\epsilon(\Gamma)$ have Euler characteristic $\chi(\Gamma) + 1$.

It should be clear that these properties prove the lemma.

If $v \in V(\Gamma)$ is the vertex corresponding to s_m , then $\epsilon(\Gamma)$ is the set of all extensions of $\Gamma \in \mathcal{Q}_w$ where there is a j -edge (an edge labeled j) starting at v . If Γ already has such an edge, then we can attain a graph $\Gamma' \in \mathcal{Q}_{w'}$ by adding s_{m+2} to the block containing s_m , and adding s_{m+1} to the block at the end of this j -edge. We then define $\epsilon(\Gamma) = \{\delta(\Gamma)\} = \{\Gamma'\} \approx \{\Gamma\}$ (We use “ \approx ” to denote equality as vertex-unlabeled-graphs.)

Otherwise, $\epsilon(\Gamma)$ includes all the $(v_\Gamma - e_\Gamma^j + 1)$ different possible extensions of Γ with such an edge. We only need to specify the other vertex of this new edge, that corresponds to s_{m+1} . In the graph $\delta(\Gamma)$ the vertex s_{m+1} is new and so is the j -edge (s_m, s_{m+1}) . Clearly, $\chi(\delta(\Gamma)) = \chi(\Gamma)$, as claimed. Otherwise this additional edge can go from s_m to any of the $v_\Gamma - e_\Gamma^j$ vertices in Γ which are not tails of a j -edge. Such graphs clearly have characteristics $\chi(\Gamma) + 1$.

We prove the first two properties of ϵ by recovering, for every $\Gamma' \in \mathcal{Q}_{w'}$ the (unique) graph $\Gamma \in \mathcal{Q}_w$ with $\Gamma' \in \epsilon(\Gamma)$. We consider the $(m + 1)$ -st step in the path of w' through Γ' (the step from s_m to s_{m+1}), and use the notations of Lemma 6. If this step is free, then Γ is obtained from Γ' by deleting the vertex corresponding to s_{m+1} and the edge (s_m, s_{m+1}) (as well as, of course, omitting s_{m+2} from its block). If it is a coincidence, then clearly $\Gamma \approx \Gamma' \setminus (s_m, s_{m+1})$. Otherwise, it is forced and so $\Gamma \approx \Gamma'$.

We want to show next that if $\Gamma \in \mathcal{Q}_w$ has type A, then all graphs in $\epsilon(\Gamma)$ have type A as well. By assumption Γ is generated by a set S of cardinality $|S| = \chi(\Gamma)$ and $\{s_0, s_m\} \in S$. Note that $s_m \equiv s_{m+2}$ in every quotient of w' . Therefore, when we consider generating sets for graphs in $\mathcal{Q}_{w'}$, the vertices s_m and s_{m+2} play the exact same role. We therefore define S' to be the set of pairs that is attained by replacing each occurrence of s_m in S with s_{m+2} . Clearly, S' generates the graph $\delta(\Gamma) \in \epsilon(\Gamma)$. For any other graph in $\epsilon(\Gamma)$, we add to S' the pair $\{s_{m+1}, s_i\}$ where s_i ($i \leq m$) corresponds to the vertex which (s_m, s_{m+1}) goes to. In each of these cases we found a smallest generating set S' that includes the pair $\{s_0, s_{m+2}\}$, so all members of $\epsilon(\Gamma)$ have type A.

Finally, we need to show that if $\Gamma \in \mathcal{Q}_w$ has type B, then so does $\delta(\Gamma)$. So suppose $\delta(\Gamma)$ has type A, with a smallest generating set S' that contains the pair $\{s_0, s_{m+2}\}$. Let us construct a set of pairs S by replacing each occurrence of s_{m+2} in S' by s_m . If S' contains some pair $\{s_{m+1}, s_i\}$, then clearly s_{m+1} is not a new vertex in $\delta(\Gamma)$, and we are necessarily in the case where $\Gamma \approx \delta(\Gamma)$ (recall that Γ and $\delta(\Gamma)$ have the same characteristic). In this case the edge (s_m, s_{m+1}) is not new, so it is merged with some (s_{r-1}, s_r) (or (s_{r+1}, s_r)) for some $r < m$. Thus, we can replace each s_{m+1} in S' with s_r . This is a contradiction since S is a smallest generating set of Γ which therefore has type A. ■

Note that from Lemmas 8 and 9 it follows that β is invariant under cyclic reduction as well, or under conjugation. Similar arguments show that it is also invariant under the equivalence relation “ \sim ”, but we do not include the proof. In the following lemma we state an important property of type-B quotient graphs. This property plays a crucial role in the sequel, where we introduce a bound to the number of words with some fixed value of $\beta(\cdot)$.

Lemma 10. *Let $\Gamma \in \mathcal{Q}_w$ have type B. As we trace the path of w through Γ , every edge in Γ is traversed at least twice.*

Proof. We show that if some edge e is traversed only once, then Γ has type A. Lemma 8 allows us to assume that e is the last step in the path of w , i.e., the step from $s_{|w|-1}$ to $s_{|w|}$. In the proof of Lemma 6, we constructed \hat{S} , a generating set of Γ of smallest cardinality, with one pair for each coincidence in the path of w through Γ . Here, the last coincidence corresponds to the pair $\{s_0, s_{|w|}\}$. Therefore Γ has type A. ■

Remark 11. The converse is not true. There are quotients of type A where every edge is traversed more than once. Consider the word $w = ababa$. One of the quotient graphs in \mathcal{Q}_w is a figure-eight with one vertex and two loops. Each edge in this quotient is traversed twice or thrice, but the quotient has type A. It is generated by the two pairs $\{s_0, s_2\}, \{s_0, s_5\}$.

2.3. Some Connections Between $\phi(w)$ and $\beta(w)$

We now turn to examine the relation between ϕ and β . We first observe that $a_i(w)$ (the coefficient of $\frac{1}{n^i}$ in the power series form of $\Phi_w(n)$), is completely determined by quotients in \mathcal{Q}_w with characteristic $\leq i$. This is easily verified by considering the contribution of each $\Gamma \in \mathcal{Q}_w$ in (7). We are now able to use our new perspective and show that (as mentioned earlier) for $i = 0, 1$, $\phi(w) = i \Leftrightarrow \beta(w) = i$.

Lemma 12. *For every $w \in \mathbf{F}_k$, $\phi(w) = 0 \Leftrightarrow \beta(w) = 0 \Leftrightarrow w = 1$.*

Proof. By Lemma 6, the only quotient graph of w with characteristic 0 is the graph Γ generated by the empty set. Now $\phi(w) = 0 \Leftrightarrow a_0(w) > 0 \Leftrightarrow \Gamma \in \mathcal{Q}_w \Leftrightarrow s_0 \equiv s_{|w|}$ in $\Gamma \Leftrightarrow w$ reduces to 1. If $\Gamma \in \mathcal{Q}_w$ then Γ has type B by definition. Thus $\beta(w) = 0 \Leftrightarrow \Gamma \in \mathcal{Q}_w$. ■

Lemma 13. *For every $w \in \mathbf{F}_k$, $\phi(w) = 1 \Leftrightarrow \beta(w) = 1 \Leftrightarrow w$ is imprimitive.*

Proof. Let $w \in \mathbf{F}_k$ be in reduced form and assume $w \neq 1$. By Lemma 12, all $\Gamma \in \mathcal{Q}_w$ have a positive characteristic. The definition of $\Phi_w(n)$ clearly yields that $a_1(w) = |\{\Gamma \in \mathcal{Q}_w : \chi(\Gamma) = 1\}| - 1$. In this case the single pair $\{s_0, s_{|w|}\}$ is a smallest generating set for the universal graph (defined in Section 2.1), so at least one quotient has characteristic 1. Obviously, any other quotient with $\chi = 1$ is not generated by $\{s_0, s_{|w|}\}$, and thus (by Lemma 6) has type B. Thus $\phi(w) = 1 \Leftrightarrow \beta(w) = 1 \Leftrightarrow$ such additional quotients exist. We complete the proof by showing that the latter is true iff w is imprimitive.

Let w' be the cyclic reduction of w . It is easy to verify that $w \sim w'$ whence $\phi(w) = \phi(w')$ and w is primitive iff so is w' . Lemmas 8 and 9 yield that $\beta(w) = \beta(w')$ as well. Thus we can assume, for simplicity, that w is cyclically reduced.

In this case, merging s_0 and $s_{|w|}$ implies no other identifications, and the universal graph is a cycle of length $|w|$. If w is imprimitive, there is some $u \in \mathbf{F}_k$ and $d \geq 2$ such that $w = u^d$. Clearly, u is cyclically reduced as well, and the universal graph of u is a cycle of length $|u|$ which is an additional quotient of characteristic 1 in \mathcal{Q}_w .

On the other hand, since w is cyclically reduced, every vertex in every $\Gamma \in \mathcal{Q}_w$ has degree ≥ 2 . Thus, if $\chi(\Gamma) = 1$, it is necessarily a cycle. As the path of w through Γ is non-backtracking and $w \neq 1$, it consists of tracing this cycle some $d \geq 1$ times. If $d = 1$, Γ is the universal graph. Otherwise, if we let u denote the word corresponding to a single traversal of the cycle, then $w = u^d$. ■

The contents of Lemmas 12 and 13 appear in different language in [10], in [14] and in [5]. But the relation between $\phi(\cdot)$ and $\beta(\cdot)$ goes deeper. For instance, for the single-letter word $w = a$ both $\phi(w) = \beta(w) = \infty$. The reason for $\phi(a) = \infty$ is that the expected number of fixed points in a random permutation equals 1. On the other hand, Lemma 10 implies that $\beta(a) = \infty$. Also, as already mentioned, both ϕ and β are invariant under “ \sim ”, so they are both infinite on the entire equivalence class of a under “ \sim ”. (In particular, every w in which some letter appears exactly once belongs to this class.)

The following lemma expands even further the relation between $\phi(\cdot)$ and $\beta(\cdot)$. The next natural step would be to prove that $\phi(w) = 2 \Leftrightarrow \beta(w) = 2$. This, in other words, says that for primitive words, $\beta(w) \geq 3$ iff $a_2(w) = 0$. This is, at present, still beyond our reach and we content ourselves with a weaker statement.

Lemma 14. *Let $w \in \mathbf{F}_k$ have $\beta(w) \geq 3$. Then $a_2(w) \leq 0$.*

Proof. For simplicity we assume that w is cyclically reduced. (Again, this assumption is possible because both $\beta(w)$ and $\Phi_w(n)$ are invariant under cyclic reductions of w .) When $\beta(w) \geq 3$, there is only one graph $\hat{\Gamma} \in \mathcal{Q}_w$ of characteristic 1 (the universal graph), and all the quotient graphs of characteristic 2 have type A. To find the contribution of $\hat{\Gamma}$ to

$a_2(w)$ expand the expression $\frac{\prod_{l=1}^{v_{\hat{\Gamma}}-1} (1-lx)}{\prod_{j=1}^k \prod_{l=1}^{e_{\hat{\Gamma}}^j-1} (1-lx)}$ to first order. It follows that this contribution is $-\binom{v_{\hat{\Gamma}}}{2} + \sum_{j=1}^k \binom{e_{\hat{\Gamma}}^j}{2}$. We need to show that there are at most $\binom{v_{\hat{\Gamma}}}{2} - \sum_{j=1}^k \binom{e_{\hat{\Gamma}}^j}{2}$ graphs $\Gamma \in \mathcal{Q}_w$ with $\chi(\Gamma) = 2$.

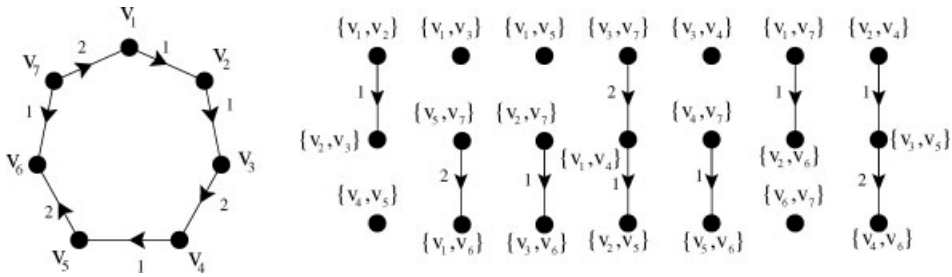


Fig. 4. The graph Υ (on the right) corresponding to the universal graph $\hat{\Gamma}$ (on the left) for $w = g_2^2 g_1 g_2 g_1^{-1} g_2$. ($\hat{\Gamma}$'s vertices are denoted here by v_1, \dots, v_7 while the s_i labels are omitted.)

Since $\hat{\Gamma}$ is generated by $\{s_0, s_{|w|}\}$, and every quotient $\Gamma \in \mathcal{Q}_w$ of characteristic 2 has type A, Γ is generated from $\hat{\Gamma}$ by a single pair of vertices of $\hat{\Gamma}$. The total number of pairs is $\binom{v_{\hat{\Gamma}}}{2}$, but clearly different pairs may generate the same Γ . For instance, for any two j -edges, the pair of heads generates the same quotient as the pair of tails.

To understand the full picture, we introduce a graph Υ , which captures this kind of dependency between pairs. The graph Υ has $\binom{v_{\hat{\Gamma}}}{2}$ vertices labeled by the pairs of vertices of $\hat{\Gamma}$, and has $\sum_{j=1}^k \binom{e_{\hat{\Gamma}}^j}{2}$ edges, one for each pair of same-color edges in $\hat{\Gamma}$. The edge corresponding to the pair $\{\epsilon_1, \epsilon_2\}$ of j -edges, is a j -edge from the vertex $\{\text{head}(\epsilon_1), \text{head}(\epsilon_2)\}$ to $\{\text{tail}(\epsilon_1), \text{tail}(\epsilon_2)\}$. We illustrate this in Fig. 4.

We claim that Υ has no cycles. As w is assumed to be cyclically reduced, $\hat{\Gamma}$ is simply a cycle and the path of w through it is a simple cycle. Now say that $\{x_1, y_1\}, \{x_2, y_2\}, \dots, \{x_r, y_r\}, \{x_1, y_1\}$ is a simple cycle in Υ , whose edges compose some word u . This u corresponds to two nonbacktracking paths in $\hat{\Gamma}$ in one of the two following ways. Either u is a path from x_1 to itself and a path from y_1 to itself (whence it is some cyclic shift of w or of w^{-1}). Or u is a path from x_1 to y_1 as well as a path from y_1 to x_1 . (See Fig. 5.)

In the former case, some cyclic shift of w equals another cyclic shift of w or of w^{-1} . But w is primitive, so it is not invariant under any cyclic shift. To see that w cannot equal

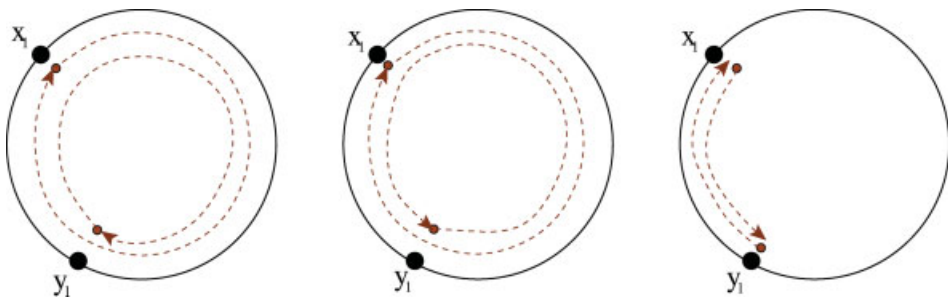


Fig. 5. A cycle in Υ yields one of these three scenarios. The thick line stands for the universal graph $\hat{\Gamma}$. The broken lines (arrows) stand for copies of u , the subword composed by the edges of the cycle in Υ . In the left and middle scenarios, u is a path from x_1 to itself and a path from y_1 to itself (in the left graph the paths are equally oriented, in the middle one they have inconsistent orientation). The right graph describes the case where u is a path from x_1 to y_1 as well as a path from y_1 to x_1 . All three scenarios are impossible.

a cyclic shift of its inverse, we need to show that w^{-1} is not a subword of w^2 . To see this, let $w = g_1 \dots g_m$ and assume to the contrary that $\forall j = 1, \dots, m \ g_j^{-1} = g_{s-j}$ for some $m + 1 \leq s \leq 2m$ (here $g_{i+m} = g_i$ for $i = 1, \dots, m$). If s is even, say $s = 2q$, we conclude (for $j = q$) that $g_q = g_q^{-1}$ which is impossible. If $s = 2q + 1$ we conclude for $j = q$ that $g_q^{-1} = g_{q+1}$, so that the word w is not reduced.

In the latter case, either $u = u^{-1}$, which is impossible by the same argument, or w is a cyclic shift of u^2 , which again contradicts its being primitive. This rules out the possibility of a cycle in Υ .

Clearly, two pairs of vertices in $V(\hat{\Gamma})$ which belong to the same connected component in Υ , generate the same Γ . Thus the number of different Γ 's in \mathcal{Q}_w with characteristic 2 is at most the number of connected components in Υ . But Υ has no cycles, so the number of connected components is exactly $\binom{v_{\hat{\Gamma}}}{2} - \sum_{j=1}^k \binom{e_{\hat{\Gamma}}^j}{2}$. ■

Note that this last proof yields a surjective function from the connected components of Υ to type-A graphs of characteristics 2 in \mathcal{Q}_w . To prove that $\phi(w) = 2 \Leftrightarrow \beta(w) = 2$ it suffices to show that this function is injective (which we believe is true). That would yield that the number of type-A quotients of characteristic 2 exactly balances off the negative contribution of $\hat{\Gamma}$ to $a_2(w)$. Thus, $a_2(w) \neq 0$ (or more precisely $a_2(w) > 0$) iff there is a type-B graph of characteristic 2, i.e., iff $\beta(w) = 2$.

In fact, we believe that something similar happens in general. Namely, for every integer $i \geq 0$, if \mathcal{Q}_w has no type-B graphs of characteristic $< i$, then the contributions from all the type-A graphs of characteristic $\leq i$. (i.e., the graphs generated from the universal graph by fewer than i pairs) balances out. Hence $a_0(w) = a_1(w) = \dots = a_{i-1}(w) = 0$, and $a_i(w)$ equals the number of type-B quotients of characteristic i .

There are three kinds of supporting evidence to this belief. As we saw, it is valid for $i = 0, 1$, and we have a good understanding why it should hold for $i = 2$ as well. Also, we have carried out extensive numerical simulations to test it for $i = 2, 3, 4$ without any failure. Finally, consider a word w such that $w \sim a$. Such a word has only type-A quotients, and we know that $a_i(w) = 0$ for all i . In this case, therefore, the type-A quotients of characteristic $\leq i$ indeed balance the contribution of each other to $a_0(w), \dots, a_i(w)$. (We note that the vanishing of all $a_i(w)$ even in this specialized case is not obvious).

Here, then, is a formal statement of this main conjecture:

Conjecture 15. For every $w \in \mathbf{F}_k$,

$$\phi(w) = \beta(w) \tag{10}$$

Moreover, if $\phi(w) = i$, then

$$a_i(w) = |\{\Gamma \in \mathcal{Q}_w : \chi(\Gamma) = i \text{ and } \Gamma \text{ has type B}\}| \tag{11}$$

Conjecture 15 is the main missing link in our tentative proof to the claim that $\mathbb{E}(\mu_{\max}) < O(\rho)$. We return to this in Section 3. Note that this conjecture implies that the first non-zero coefficient in the power series of $\Phi_w(n)$ is positive. This means that $E(X_w^{(n)}) \geq 1$ for every $w \in \mathbf{F}_k$ and every large enough n .

Remark 16. It might be tempting to suspect something even stronger, namely that for every $w \in \mathbf{F}_k$ and $n \geq 1$, $E(X_w^{(n)}) \geq 1$. However, this stronger assertion is false. (We would

like to thank Miklós Abért for showing us the invalidity of this claim. The main ideas of the proof can be found in [16].

Finally, we present a second conjecture based on other simulations we conducted. These simulations suggest that the only words for which $\phi(w) = \infty$ are those mentioned above:

Conjecture 17. $E(X_w^{(n)}) \equiv 1$ iff w is equivalent (\sim) to a single-letter word.

3. THE LARGEST NEW EIGENVALUE IN A RANDOM LIFT OF A GRAPH

In this section, we apply our findings concerning formal words and word maps on S_n to study the new eigenvalues in random lifts of graphs. Our main result is Theorem 1 which says that $\mu_{\max} \leq O(\lambda_1^{1/3} \rho^{2/3})$ almost surely.

Recall the definition of $\phi(w)$ for $w \in \mathbf{F}_k$ [Eq. (8)]. Namely, $\Phi_w(n) = \frac{a_{\phi(w)}(w)+o(1)}{n^{\phi(w)}}$. We first seek an improved estimate of the $o(1)$ term in the numerator.

Lemma 18. *Let $w \in \Sigma_k$ and let $i \geq 0$ be some integer. Then:*

$$|\{\Gamma \in \mathcal{Q}_w : \chi(\Gamma) = i\}| \leq |w|^{2i}$$

Proof. As we saw in Lemma 6, each $\Gamma \in \mathcal{Q}_w$ with characteristic i is generated by some set of i pairs. There are $\binom{|w|+1}{2} \leq |w|^2$ pairs to choose from, and the claim follows. ■

Lemma 19. *If $\phi(w) = i$ and if $n \geq 3|w|^2$, then*

$$\Phi_w(n) \leq \frac{1}{n^i} \left(a_i(w) + \frac{|w|^{2i+4}}{n} \right). \tag{12}$$

Proof. As mentioned in the beginning of Section 2.3, $a_i(w)$, the coefficient of $\frac{1}{n^i}$ in the power series of $\Phi_w(n)$, is completely determined by quotients in \mathcal{Q}_w of characteristic $\leq i$. We now bound the contribution to $a_i(w)$ of every such quotient.

To this end, we analyze the contribution of Γ to $\Phi_w(n)$, as specified in (7). For the sake of convenience, we let x equal $\frac{1}{n}$, and express this contribution as $x^{e_{\Gamma}-v_{\Gamma}+1} \cdot \frac{\prod_{l=1}^{v_{\Gamma}-1} (1-lx)}{\prod_{j=1}^k \prod_{r=1}^{e_{\Gamma}^j-1} (1-lx)}$.

For small x we can expand the fraction in this expression as a power series $\sum_{r=0}^{\infty} b_r x^r$. Write $\prod_{l=1}^{v_{\Gamma}-1} (1-lx) = 1 + \sum_{r \geq 1} c_r x^r$, and $\prod_{j=1}^k \prod_{l=1}^{e_{\Gamma}^j-1} (1-lx) = 1 + \sum_{r \geq 1} d_r x^r$. Then:

$$1 + \sum_{r=1}^{\infty} c_r x^r = \left[\sum_{r=0}^{\infty} b_r x^r \right] \left[1 + \sum_{r=1}^{\infty} d_r x^r \right]$$

Thus $b_0 = 1$ and for $r \geq 1$, $b_r = c_r - d_r - b_1 d_{r-1} - \dots - b_{r-1} d_1$. We have

$$|c_r| = \sum_{1 \leq y_1 < \dots < y_r \leq v_{\Gamma}-1} y_1 \cdots y_r \leq \left[\sum_{y=1}^{v_{\Gamma}-1} y \right]^r \leq \left(\frac{v_{\Gamma}^2}{2} \right)^r \leq \frac{|w|^{2r}}{2^r}$$

Similarly, $|d_r| \leq \frac{|w|^{2r}}{2^r}$. A simple induction now shows that $|b_r| \leq |w|^{2r}$:

$$\begin{aligned} |b_r| &= |c_r - d_r - b_1 d_{r-1} - \dots - b_{r-1} d_1| \leq \\ &\leq \frac{|w|^{2r}}{2^r} + \frac{|w|^{2r}}{2^r} + \frac{|w|^2 |w|^{2r-2}}{2^{r-1}} + \frac{|w|^4 |w|^{2r-4}}{2^{r-2}} + \dots + \frac{|w|^{2r-2} |w|^2}{2} = |w|^{2r} \end{aligned}$$

Now consider the coefficient $a_i(w)$. There is at most one quotient in \mathcal{Q}_w of characteristic 0 (Lemma 18) which contributes at most $|w|^{2i}$ to $a_i(w)$; There are at most $|w|^2$ quotients of characteristic 1 which contribute at most $|w|^{2i-2}$ each, etc. There are no quotients with characteristic greater than $|w|$, so we have at most contribution of $|w|^{2i}$ of quotients of every characteristic $0 \leq \chi \leq |w|$. Thus $a_i(w) \leq (|w| + 1)|w|^{2i}$.

This yields the following:

$$\begin{aligned} \Phi_w(n) &= \sum_{i=\phi(w)}^{\infty} a_i(w) \frac{1}{n^i} \leq a_{\phi(w)}(w) \frac{1}{n^{\phi(w)}} + \sum_{i=\phi(w)+1}^{\infty} \frac{(|w| + 1)|w|^{2i}}{n^i} = \\ &= \frac{1}{n^{\phi(w)}} \left[a_{\phi(w)}(w) + \frac{(|w| + 1)|w|^{2\phi(w)+2}}{n} \frac{n}{n - |w|^2} \right] \end{aligned}$$

The lemma now follows because $n \geq 3|w|^2$. ■

The set of possible values for $\beta(w)$ is $\{0, 1, \dots, |w|\} \cup \{\infty\}$, and we now split the sum over $w \in \mathcal{CP}_t(G)$ in (3) according to $\beta(w)$. This yields:

$$\begin{aligned} \mathbb{E}(\mu_{\max}^t) &\leq \sum_{w \in \mathcal{CP}_t(G)} [\mathbb{E}(X_w^{(n)}) - 1] = \sum_{w \in \mathcal{CP}_t(G)} n \cdot \Phi_w(n) \\ &= \sum_{i \in \{0, 1, \dots, t\} \cup \{\infty\}} \sum_{\substack{w \in \mathcal{CP}_t(G) \\ \beta(w)=i}} n \cdot \Phi_w(n) \end{aligned}$$

The statement of (the unproved) Conjecture 15 implies that the sum over w with $\beta(w) = \infty$ vanishes, since $\beta(w) = \infty$ yields $\phi(w) = \infty$ and hence $\Phi_w(n) \equiv 0$. We suspect that it should be possible to bound the number of words with $\beta(w) = i$ (Some results along these lines are proved in Section 3.1). This, combined with the statement of Conjecture 15 would have allowed us to bound the contribution of each $0 \leq i \leq t$ to the above sum.

This problem is still open, so instead we split the set $\mathcal{CP}_t(G)$ into four parts:

$$\mathbb{E}(\mu_{\max}^t) \leq \sum_{\substack{w \in \mathcal{CP}_t(G) \\ \beta(w)=0}} n \cdot \Phi_w(n) + \sum_{\substack{w \in \mathcal{CP}_t(G) \\ \beta(w)=1}} n \cdot \Phi_w(n) + \sum_{\substack{w \in \mathcal{CP}_t(G) \\ \beta(w)=2}} n \cdot \Phi_w(n) + \sum_{\substack{w \in \mathcal{CP}_t(G) \\ \beta(w) \geq 3}} n \cdot \Phi_w(n) \tag{13}$$

Using Lemmas 12 and 13 we can bound the value of $\Phi_w(n)$ when $\beta(w) = 0$ or 1. For these values $\beta(w) = \phi(w)$ and Lemma 19 can be applied. If $\beta(w) = 2$, then $\phi(w) \geq 2$, and we can use (12) in its worst case, i.e. when $\phi(w) = 2$. Finally, if $\beta(w) \geq 3$, the following lemma shows that $\Phi_w(n) \leq O(\frac{1}{n^3})$.

Lemma 20. *Let $w \in \mathbf{F}_k$ have $\beta(w) \geq 3$. Then*

$$\Phi_w(n) \leq \frac{1}{n^3} \left(a_3(w) + \frac{|w|^{10}}{n} \right)$$

Proof. The assumption $\beta(w) \geq 3$ yields that $a_0(w) = a_1(w) = 0$ and that $a_2(w) \leq 0$ (Lemma 14). Thus clearly $\Phi_w(n) \leq \sum_{i=3}^{\infty} a_i(w) \frac{1}{n^i}$, and the claim is an immediate consequence of the analysis in Lemma 19. (This is true since the proof of Lemma 19 does not take full advantage of the assumption that $\phi(w) = i$, but rather that $a_j(w) \leq 0$ for $j < i$.) ■

To proceed with our analysis of Eq. (13), we now bound the number of words in $\mathcal{CP}_t(G)$ with $\beta(w) = i$ for $i = 0, 1, 2$.

3.1. The Number of Words $w \in \mathcal{CP}_t(G)$ with Fixed $\beta(w)$

Our proof for this bound extends an idea that originated with [6] and was later developed in [5]. Recall that ρ denotes the spectral radius of T , the universal cover of the base graph G (as well as of any lift of G). Buck found a bound expressed in terms of ρ for the number of words in $\mathcal{CP}_t(G)$ that reduce to 1. Friedman used a similar method to bound the number of imprimitive words in $\mathcal{CP}_t(G)$. We further develop the method in order to bound the number of words in $\mathcal{CP}_t(G)$ with $\beta(w) = 2$.

We present the three cases ($i = 0, 1, 2$) one by one. The case $i = 0$ is indeed the simplest, and things get more complicated as i grows. (However, it does seem that a general bound can be proven for the number of words in $\mathcal{CP}_t(G)$ for any fixed value of $\beta(\cdot)$.) We first note that A_T , the (infinite) adjacency matrix of T , is a self-adjoint bounded operator on the Hilbert space $l_2(V(T))$. Consequently, its operator norm equals its spectral radius, i.e., $\|A_T\| = \rho$. (The same argument shows that $\|A_T^l\| = \rho^l$ for any integer $l > 0$). For every $v_1, v_2 \in V(T)$ the number of paths of length l from v_1 to v_2 is $A_T^l(v_1, v_2)$, which can be bounded by $\|A_T^l\| = \rho^l$.

For every $x \in V(G)$, we arbitrarily choose some vertex $v_1 = v_1(x)$ in the fiber of x in T . For every path γ in G that starts at x , we consider the lift of γ that starts at v_1 . We denote the tail of the lifted path by $v_\gamma = v_\gamma(x)$.

For $i = 0$, recall that $\beta(w) = 0 \Leftrightarrow w$ reduces to 1 (Lemma 12). Thus, every $w \in \mathcal{CP}_t(G)$ with $\beta(w) = 0$ corresponds to some path in G of length t which reduces to 1 (a nullhomotopic path). But these are exactly the paths which lift to closed paths in T as well. Thus:

$$|\{w \in \mathcal{CP}_t(G) : \beta(w) = 0\}| = \sum_{x \in V(G)} A_T^t(v_1, v_1) \leq \sum_{x \in V(G)} \rho^t = |V(G)|\rho^t \quad (14)$$

For $i = 1$ we want to count the number of imprimitive words in $\mathcal{CP}_t(G)$. If w is imprimitive, then $w = u^d$ (equality in \mathbf{F}_k) for some $u \in \mathbf{F}_k$ and $d \geq 2$. Suppose that the path of the cyclically reduced form of u in G starts at $x \in V(G)$. Since the path of w visits x , there is some cyclic shift of w that starts at x . Thus, by adding a factor of $|w| = t$ to our eventual bound, we can assume w begins at x .

We now let γ be the path in G of the cyclically reduced form of u (a loop from x to itself). We divide the path of w through G to three parts:

1. a path homotopic to γ of length l_1
2. a path homotopic to γ of length l_2
3. a path homotopic to γ^{d-2} of length l_3

with $l_1 + l_2 + l_3 = t$.

These three parts lift to the following paths in T :

1. a path from v_1 to v_γ of length l_1
2. a path from v_1 to v_γ of length l_2
3. a path from v_1 to $v_{\gamma^{d-2}}$ of length l_3

Thus, we can bound the total number of imprimitive words in $\mathcal{CP}_t(G)$ as follows:

$$\begin{aligned} & |\{w \in \mathcal{CP}_t(G) : \beta(w) = 1\}| \leq \\ & \leq t \cdot \sum_{d=2}^t \sum_{x \in V(G)} \sum_{\substack{\gamma \text{ is a reduced} \\ \text{loop from } x \text{ to } x}} \sum_{l_1+l_2+l_3=t} A_T^{l_1}(v_1, v_\gamma) A_T^{l_2}(v_1, v_\gamma) A_T^{l_3}(v_1, v_{\gamma^{d-2}}) \end{aligned}$$

(the initial factor of t accounts for the cyclic shift of w).

Using the symmetry of A_T , the inequality $A_T^{l_3}(v_1, v_{\gamma^{d-2}}) \leq \rho^{l_3}$ and changing the order of summations, we obtain:

$$\begin{aligned} & |\{w \in \mathcal{CP}_t(G) : \beta(w) = 1\}| \\ & \leq t \cdot \sum_{d=2}^t \sum_{x \in V(G)} \sum_{l_1+l_2+l_3=t} \rho^{l_3} \sum_{v_\gamma \in \text{Fib}(x) \setminus \{v_1\} \subset V(T)} A_T^{l_1}(v_1, v_\gamma) A_T^{l_2}(v_\gamma, v_1) \\ & \leq t(t-1) \sum_{x \in V(G)} \sum_{l_1+l_2+l_3=t} \rho^{l_3} A_T^{l_1+l_2}(v_1, v_1) \\ & \leq t(t-1) \sum_{x \in V(G)} \sum_{l_1+l_2+l_3=t} \rho^{l_3} \rho^{l_1+l_2} \\ & = t(t-1)|V(G)| \binom{t}{2} \rho^t \leq |V(G)| t^4 \rho^t \end{aligned} \tag{15}$$

(The crux of the matter in this calculation is the second step which eliminates the need to sum over closed paths γ).

Next, we bound the number of words in $\mathcal{CP}_t(G)$ with $\beta(w) = 2$. To this end we introduce a lemma that deals with quotient graphs of smallest characteristic among all type-B quotients. Clearly, it is such quotients that determine $\beta(w)$.

Let G be a graph, $w \in \mathcal{CP}_t(G)$ and $\Gamma \in \mathcal{Q}_w$. Each label s_i that appears in a block that is associated with a vertex in Γ corresponds to a vertex in G . Therefore, a vertex in Γ may correspond to several vertices in G . We next show that for Γ as above each vertex in Γ corresponds to a single vertex from G .

Lemma 21. *Let G be a graph and $w \in \mathcal{CP}_t(G)$. If $\Gamma \in \mathcal{Q}_w$ has type B and $\chi(\Gamma) = \beta(w)$, then all labels that appear in a block from a vertex in Γ correspond to the same vertex in G .*

Proof. The proof proceeds by showing that otherwise there is another type-B quotient in \mathcal{Q}_w with smaller characteristic. Indeed, let $\{s_{i_1}, \dots, s_{i_r}\}$ be a block in Γ where the labels $\{s_{i_1}, \dots, s_{i_k}\}$ ($k < r$) correspond to the vertex v in G , while the labels $\{s_{i_{k+1}}, \dots, s_{i_r}\}$ correspond to other vertices in G . Define the partition $\bar{\Gamma}$ by splitting this block to $\{s_{i_1}, \dots, s_{i_k}\}$

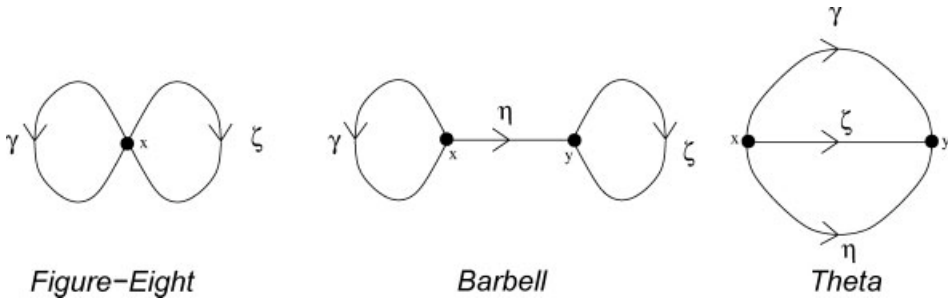


Fig. 6. The three possible shapes of quotient graphs of a cyclically reduced word with characteristic 2. The edges γ , ζ and η denote subwords.

and $\{s_{i_{k+1}}, \dots, s_{i_r}\}$. We claim that this is (i) a realizable quotient in \mathcal{Q}_w (ii) of characteristic $\chi(\Gamma) - 1$ and (iii) of type B as well.

To see (i), recall that every letter in w corresponds to some edge in G . As the two parts of the split block correspond to disjoint sets of vertices in G , there is no j such that both of them are heads (or tails) of a j -edge. Thus, the realizability of Γ yields the realizability of $\bar{\Gamma}$. In deriving $\bar{\Gamma}$ from Γ , we have increased the number of vertices by one with no additional edges, whence (ii) is proved. To show (iii), note that any generating set of $\bar{\Gamma}$ can be extended to a generating set of Γ by adding $\{s_{i_1}, s_{i_{k+1}}\}$, so that if $\bar{\Gamma}$ has type A, so does Γ . ■

As we saw (Lemmas 8 and 9) if w' is the cyclic reduction of w , then $\beta(w') = \beta(w)$. Moreover, every type-B $\Gamma \in \mathcal{Q}_w$ with $\chi(\Gamma) = \beta(w)$ can be generated from some $\Gamma' \in \mathcal{Q}_{w'}$ with $\chi(\Gamma') = \chi(\Gamma)$, through a series of δ -operations (as in Lemma 9). Since w' is cyclically reduced, every vertex in Γ' has degree ≥ 2 . (Indeed, Γ' is obtained from Γ by successive elimination of vertices of degree one in the graph).

Now assume $\beta(w) = \chi(\Gamma) = 2$. There are three possible shapes that Γ' can have: Figure-Eight, Barbell or Theta (see Fig. 6). For a cost of an additional factor of t as above, we may assume the path of w through Γ begins at the vertex x specified in each of the diagrams. (More accurately, it begins at the vertex of Γ corresponding to x through the series of δ -operations.) By Lemma 21, x corresponds to a unique vertex in G which we call x as well (by abuse of notation). This vertex x in G marks the starting point of w . We now analyze each case separately, and using the notations in Fig. 6, we trace the path of w' through Γ' .

Assume first that Γ' has the shape of a Figure-Eight. In this case w' can be expressed using γ, ζ in a reduced expression in which each of them appears at least twice (Lemma 10). For any fixed reduced expression in γ, ζ , we specify certain two appearances of γ and certain two appearances of ζ . The path of w through Γ can be then divided to at most seven parts: four parts for the chosen appearances of γ and ζ , and three parts for sequences of the rest of the expression (we can always choose the first two characters in the expression, but we may be forced to have spaces between the second and the third, between the third and the fourth and after the fourth). We then proceed to a calculation as in (15).

To illustrate, let $w' = \gamma\gamma\zeta\gamma^{-1}\zeta^{-1}\zeta^{-1}\gamma$. We split w' to seven parts as shown in the following bracketing: $w' = (\gamma)(\gamma)(\gamma)(\zeta)(\gamma^{-1})(\zeta^{-1})(\zeta^{-1}\gamma)$. The corresponding lengths are: l_1 steps for the first γ , l_2 for the second, l_3 steps for the next γ (which is considered “a space”), l_4 steps for ζ , l_5 for the space γ^{-1} , l_6 for ζ^{-1} and l_7 for the space $\zeta^{-1}\gamma$. We can now bound the total number of words which reduce cyclically (and with a possible cyclic

shift) to this expression in some γ and ζ . (The first factor of t in the calculation accounts for the initial cyclic shift of w .)

$$\begin{aligned} & \{ \{ w \in \mathcal{CP}_t(G) : w \text{ reduces cyclically to } \gamma\gamma\gamma\zeta\gamma^{-1}\zeta^{-1}\zeta^{-1}\gamma \text{ for some } \gamma, \zeta \} \} \\ & \leq t \cdot \sum_{x \in V(G)} \sum_{\gamma, \zeta} \sum_{l_1 + \dots + l_7 = t} A_T^{l_1}(v_1, v_\gamma) A_T^{l_2}(v_1, v_\gamma) A_T^{l_3}(v_1, v_\gamma) A_T^{l_4}(v_1, v_\zeta) \\ & \qquad \qquad \qquad \cdot A_T^{l_5}(v_\gamma, v_1) A_T^{l_6}(v_\zeta, v_1) A_T^{l_7}(v_1, v_{\zeta^{-1}\gamma}) \end{aligned}$$

(The second sum is over all γ and ζ - two distinct nonempty reduced loops from x to itself in G .) We proceed as before:

$$\begin{aligned} & \leq t \cdot \sum_{x \in V(G)} \sum_{l_1 + \dots + l_7 = t} \rho^{l_3 + l_5 + l_7} \sum_{v_\gamma \in \text{Fib}(x) \setminus \{v_1\}} A_T^{l_1}(v_1, v_\gamma) A_T^{l_2}(v_\gamma, v_1) \cdot \sum_{v_\zeta \in \text{Fib}(x) \setminus \{v_1\}} A_T^{l_4}(v_1, v_\zeta) A_T^{l_6}(v_\zeta, v_1) \\ & \leq t \cdot \sum_{x \in V(G)} \sum_{l_1 + \dots + l_7 = t} \rho^{l_3 + l_5 + l_7} \rho^{l_1 + l_2} \rho^{l_4 + l_6} \leq |V(G)| t^7 \rho^t \end{aligned}$$

This bound was calculated for a specific reduced expression in γ, ζ . The total number of possible expressions is less than 3^t . (w.l.o.g every expression begins with γ , and it contains a total of between four and t components. For r components there are at most 3^{r-1} possible continuations, and $3^3 + 3^4 + \dots + 3^{t-1} < 3^t$). Thus, we can bound the total number of words in $\mathcal{CP}_t(G)$ with $\beta(w) = 2$ and which have a type-B Eight-Figure quotient graph, by $|V(G)| t^7 3^t \rho^t$.

For the Barbell and Theta the analysis is similar, but their contribution is negligible relative to the contribution of the Figure-Eight. This time we construct a reduced expression in three subwords: γ, ζ , and η , but the possible number of expressions is bounded by 2^t (the same argument as above, only this time every subword has only two possible subsequent subwords). We need to specify two occurrences of each of the three letters this time, so we may need to split the path of w' to $6 + 5 = 11$ parts. The bound is therefore $|V(G)| t^{11} 2^t \rho^t$, the asymptotic comparison is clearly $|V(G)| t^{11} 2^t \rho^t \ll |V(G)| t^7 3^t \rho^t$.

To illustrate the calculation, consider the following (reduced) expression for the Barbell: $w' = \gamma\eta\zeta\zeta\zeta\eta^{-1}\gamma^{-1}\eta\zeta\eta^{-1}\gamma$. The number of words in $\mathcal{CP}_t(G)$ which reduce cyclically to such an expression can be bounded by:

$$\begin{aligned} & \leq t \cdot \sum_{x, y, \gamma, \zeta, \eta} \sum_{l_1, \dots, l_8} A_T^{l_1}(v_1(x), v_\gamma(x)) A_T^{l_2}(v_1(x), v_\eta(x)) A_T^{l_3}(v_1(y), v_\zeta(y)) \cdot \\ & \qquad \qquad \qquad \cdot A_T^{l_4}(v_1(y), v_\zeta(y)) A_T^{l_5}(v_1(y), v_\zeta(y)) A_T^{l_6}(v_\eta(x), v_1(x)) \cdot \\ & \qquad \qquad \qquad \cdot A_T^{l_7}(v_\gamma(x), v_1(x)) A_T^{l_8}(v_1(x), v_{\eta\zeta\eta^{-1}\gamma}(x)) \\ & \leq t \cdot \sum_{l_1, \dots, l_8} \rho^{l_5 + l_8} \sum_x \rho^{l_1 + l_7} \sum_y \rho^{l_3 + l_4} \sum_{v_\eta \in \text{Fib}(y)} A_T^{l_2}(v_1(x), v_\eta) A_T^{l_6}(v_\eta, v_1(x)) \\ & \leq t \cdot \sum_{l_1, \dots, l_8} \rho^{t - l_2 - l_6} \sum_x \sum_{v_\eta \in V(T)} A_T^{l_2}(v_1(x), v_\eta) A_T^{l_6}(v_\eta, v_1(x)) \\ & \leq |V(G)| t^8 \rho^t \end{aligned}$$

[Here x and y are vertices of G , γ (resp. ζ) is a reduced loop from x (resp. y) to itself, and η a reduced path from x to y , and $l_1 + \dots + l_8 = t$].

TABLE 2. A Bound for the Size of Each of the Four Subsets of $\mathcal{CP}_t(G)$, and a Bound for the Value of $\Phi_w(n)$ for Each Word in the Subset

The Value of $\beta(w)$	The Size of the Subset \leq	$\Phi_w(n) \leq$
0	$ V(G) \cdot \rho^t$	$1 + \frac{t^4}{n}$
1	$ V(G) t^4 \cdot \rho^t$	$\frac{1}{n}(t^2 + \frac{t^6}{n})$
2	$ V(G) t^7 \cdot 3^t \cdot \rho^t$	$\frac{1}{n^2}(t^4 + \frac{t^8}{n})$
≥ 3	$ V(G) \cdot \lambda_1^t$	$\frac{1}{n^3}(3t^6 + \frac{t^{10}}{n})$

To sum up, we have for large enough t :

$$|\{w \in \mathcal{CP}_t(G) : \beta(w) = 2\}| \leq (1 + o(1))|V(G)|t^7 3^t \rho^t \tag{16}$$

Remark 22. These calculations suggest that for every integer $r \geq 2$, the dominant figure among quotient graphs of characteristic r (after cyclic reduction) is a bouquet with r loops. Thus, for large enough t , the number of words in $\mathcal{CP}_t(G)$ with $\beta(w) = r$ is less than $(1 + o(1))|V(G)|t^{4r-1}(2r - 1)^t \rho^t$.

Remark 23. Note that this counting argument is quite wasteful, and involves a good deal of overcounting. For instance, in the case $i = 1$, we counted each word of the form $w = u^4$ twice: once with the root u and $d = 4$, and once with the root u^2 and $d = 2$. It seems that in fact, we have bounded $\sum_{\substack{w \in \mathcal{CP}_t(G) \\ \beta(w)=i}} a_i(w)$.

3.2. The Proof of Theorem 1

We now have all the necessary ingredients for a proof of Theorem 1. Recall that in (13), we split the set of words $\mathcal{CP}_t(G)$ to four subsets according to the value of $\beta(\cdot)$. In Table 2, we collect the following information for each subset: A bound on the subset's size and a bound on the value of $\Phi_w(n)$ for the words in the subset. This table highlights the significance of the proved and unproved relations between $\phi(\cdot)$ and $\beta(\cdot)$ to the analysis of the sum in (13). The value of $\phi(\cdot)$ yields the bounds in the right column of the table, whereas $\beta(\cdot)$ is used to derive the bounds in the middle one.

The number of words with $\beta(w) = 0$ was bounded in (14), and the bound for $\Phi_w(n)$ is from Lemma 19 (since $a_0(w) = 0$ whenever $\beta(w) = 0$). In (15) we bounded the number of imprimitive words in $\mathcal{CP}_t(G)$, and Lemma 19 yields again a bound for $\Phi_w(n)$ for imprimitive w . (By Lemma 18, $a_1(w) \leq t^2$). The number of words with $\beta(w) = 2$ was bounded in (16) (for t large enough), and this time we bound $\Phi_w(n)$ using the fact that $\phi(w) \geq 2$ for every word in this subset. The size of the fourth set is bounded by the total size of $\mathcal{CP}_t(G)$

$$|\mathcal{CP}_t(G)| = \text{tr}(A_G^t) = \sum_{i=1}^{|V(G)|} \lambda_i^t \leq |V(G)|\lambda_1^t,$$

and the bound on $\Phi_w(n)$ in this case comes from Lemma 20 and the analysis of $a_i(w)$ in the proof of Lemma 19.

We now select t to minimize the resulting upper bound on $\mathbb{E}(\mu_{\max}^t)$. It is not hard to see that the optimal t is $\Theta(\log n)$. Consequently, all terms $\frac{t^4}{n}, \dots, \frac{t^{10}}{n}$ are $o_n(1)$ and can be ignored. The information in Table 2 is now combined with (13) to yield [for $t = \Theta(\log n)$]:

$$\begin{aligned} \mathbb{E}(\mu_{\max}^t) &\leq (1 + o_n(1)) \\ &\quad \cdot \left(n \cdot |V(G)| \cdot \rho^t + |V(G)| \cdot t^6 \cdot \rho^t + |V(G)| \cdot t^{11} \cdot 3^t \cdot \frac{\rho^t}{n} + |V(G)| \cdot 3t^6 \cdot \frac{\lambda_1^t}{n^2} \right) \\ &\leq (1 + o_n(1)) |V(G)| \cdot t^{11} \cdot 4[\max(\rho \cdot n^{1/t}, \rho, 3\rho \cdot n^{-1/t}, \lambda_1 \cdot n^{-2/t})]^t \end{aligned}$$

To balance the first and the last terms, we set $n^{1/t} \approx \rho^{-1/3} \lambda_1^{1/3}$ (recall that t is an even integer, so we cannot guarantee exact equality here). Since $t \rightarrow \infty$ with n , the constant and polynomial factors can be replaced by $(1 + o_n(1))^t$. We obtain

$$\begin{aligned} \mathbb{E}(\mu_{\max}^t) &\leq \left[\lambda_1^{1/3} \rho^{2/3} \cdot (1 + o_n(1)) \cdot \max \left(1, \left(\frac{\rho}{\lambda_1} \right)^{1/3}, 3 \left(\frac{\rho}{\lambda_1} \right)^{2/3}, 1 \right) \right]^t \\ &= \left[\lambda_1^{1/3} \rho^{2/3} \cdot (1 + o_n(1)) \cdot \max \left(1, 3 \left(\frac{\rho}{\lambda_1} \right)^{2/3} \right) \right]^t \end{aligned}$$

(the equality holds because $\rho \leq \lambda_1$ is always true, see Remark 2).

The statement of Theorem 1 now follows from a standard application of Markov’s inequality. Obviously, for every $\epsilon > 0$, $3 \cdot \lambda_1^{1/3} \rho^{2/3} + \epsilon$ may serve as an absolute upper bound.

Remark 24. Here is a sketch of our proposed approach to Friedman’s Conjecture. Say we seek to prove that all new eigenvalues are, almost surely $O(\rho^{1-\epsilon} \lambda_1^\epsilon)$ for every $\epsilon > 0$.

To prove a bound of $O(\rho^{\frac{r}{r+1}} \lambda_1^{\frac{1}{r+1}})$ one would have to show that $\beta(w) = i \Leftrightarrow \phi(w) = i$ for every $i \leq r$. We believe that this can be shown in a way similar to our proof of Theorem 1. In addition, it would be necessary to follow up on Remark 22, and establish a bound on the number of words $w \in \mathcal{CP}_t(G)$ with given $\beta(w)$.

4. THE NUMBER OF L -CYCLES IN $w(s_1, \dots, s_k)$

In this section, we introduce a new conceptual and relatively simple proof of a Theorem of A. Nica [14]. In (2) we defined the random variable $X_w^{(n)}$ which counts the number of fixed points in $w(\sigma_1, \dots, \sigma_k)$ for fixed w . We extend this concept and for every integer $L \geq 1$ denote by $X_{w,L}^{(n)}$ a random variable on S_n^k which is defined by:

$$X_{w,L}^{(n)}(\sigma_1, \dots, \sigma_k) = \# \text{ of cycles of length } L \text{ of } w(\sigma_1, \dots, \sigma_k) \tag{17}$$

($X_{w,1}^{(n)}$ is a new notation for $X_w^{(n)}$).

Nica’s theorem says that the variables $X_{w,L}^{(n)}$ have, for fixed w and L and for $n \rightarrow \infty$, a limit distribution which can be computed explicitly. (Unless otherwise stated, the distribution on S_n^k is always uniform.)

Theorem 25. *Let $1 \neq w \in \mathbf{F}_k$ and suppose that $w = u^d$, with u primitive. Then for every integer $L \geq 1$, the random variable $X_{w,L}^{(n)}$ defined in (17) has, for $n \rightarrow \infty$, a limit distribution, which is given by:*

$$X_{w,L}^{(n)} \xrightarrow{\text{dis}} \sum_{h \in H(d,L)} hZ_{1/Lh} \tag{18}$$

where $H(d, L)$ is the set

$$H(d, L) = \left\{ h > 0 : h \mid d \text{ and } \gcd\left(\frac{d}{h}, L\right) = 1 \right\}, \tag{19}$$

$Z_m \sim \text{Poi}(m)$ (a variable with Poisson distribution with parameter m), and “ $\xrightarrow{\text{dis}}$ ” denotes convergence in distribution.

In particular, this limit distribution depends only on d and L (and not on u).

Note that in the case that w is primitive (i.e. $d = 1$), the limit distribution is simply Poisson with parameter $1/L$.

Our proof of Theorem 25 is based on the method of moments and provides, in particular, explicit expressions for the moments of $X_{w,L}^{(n)}$:

Corollary 26. *For every $1 \neq w \in \mathbf{F}_k$, $L \geq 1$ and $r \geq 1$ there exists a rational function $\Psi_{w,L,r}$ that is defined on a neighborhood of 0 such that $\mathbb{E}([X_{w,L}^{(n)}]^r) = \Psi_{w,L,r}(1/n)$ for sufficiently large n . Consequently, the limit $\lim_{n \rightarrow \infty} \mathbb{E}([X_{w,L}^{(n)}]^r)$ exists and equals $\Psi_{w,L,r}(0)$. In addition, if $w = u^d$, with u primitive, then this limit equals the corresponding moment of the sum in (18). In particular, $\Psi_{w,L,1}(0) = \lim_{n \rightarrow \infty} \mathbb{E}(X_{w,L}^{(n)}) = \frac{|H(d,L)|}{L}$.*

(Note that this is a generalization of the function $\Phi_w(n)$ defined in (4): $n \cdot \Phi_w(n) + 1 = \mathbb{E}(X_{w,1}^{(n)}) = \Psi_{w,1,1}(1/n)$.)

The contents of Nica’s work is Theorem 25 and Corollary 26 for the first moment.

The main idea of the proof was already illustrated in Section 2, where we analyzed the expectation of $X_{w,1}^{(n)}$. Recall Eq. (4) where $\Phi_w(n)$ is expressed as a power series $\sum_{i=0}^{\infty} a_i(w) \frac{1}{n^i}$. We already know (Lemma 12) that when $w \neq 1$, $a_0(w) = 0$, and so $\lim_{n \rightarrow \infty} \mathbb{E}(X_{w,1}^{(n)}) = a_1(w) + 1$. But this is exactly the number of graphs $\Gamma \in \mathcal{Q}_w$ with $\chi(\Gamma) = 1$ (see, for instance, Lemma 13).

If w is cyclically reduced, then the only graphs in \mathcal{Q}_w with $\chi = 1$ are cycles (again, see the proof of Lemma 13). Such a cycle consists of a cyclic concatenation of several copies of u (the primitive word such that $w = u^d$). The number of copies of u in the cycle has to divide d , hence

$$\lim_{n \rightarrow \infty} \mathbb{E}(X_{w,1}^{(n)}) = |H(d, 1)| = \# \text{ of divisors of } d$$

But we can indeed restrict our discussion to cyclically reduced words. The justification for this is the following. Let $1 \neq w \in \mathbf{F}_k$, and let w' be its cyclic reduction. We have already mentioned that $w \sim w'$ and so they induce the same distribution on S_n . In particular, $X_{w,L}^{(n)}$ and $X_{w',L}^{(n)}$ are equally distributed. It is also quite evident that w and w' share an identical exponent of their primitive root (i.e., if $w = xw'x^{-1}$ for some $x \in \mathbf{F}_k$ and $w' = u^d$ with u

primitive, then xux^{-1} is primitive too, and $w = (xux^{-1})^d$. Thus, the validity of Theorem 25 and Corollary 26 for w' , yields their validity for w as well.

Below, we extend this argument to obtain the limit of the expectation of $[X_{w,L}^{(n)}]^r$ for every L and r . We essentially use the same way we counted fixed points in $w(\sigma_1, \dots, \sigma_k)$ for all k -tuples $(\sigma_1, \dots, \sigma_k) \in S_n^k$, to count L -cycles and sequences of L -cycles, which we call lists of cycles. The point is that the total number of lists of length r of L -cycles, divided by $(n!)^k$, equals the expectation of $[X_{w,L}^{(n)}]^r$, the r -th factorial moment of $X_{w,L}^{(n)}$. Once we know how to calculate the limits of the factorial moments, the limits of the regular moments are at easy reach. To finish, we show that these limits equal the corresponding moments in the r.h.s of (18), and use the method of moments to conclude the proof.

4.1. Lists of Trails and Their Categories

We begin by generalizing some of the notions from Section 2.1. Let $1 \neq w \in \mathbf{F}_k$ be cyclically reduced, $n \geq 1$ an integer, $s_0 \in \{1, \dots, n\}$, and $\sigma_1, \dots, \sigma_k \in S_n$. The trail of s_0 through $w(\sigma_1, \dots, \sigma_k)$ is the sequence of images of s_0 under $w(\sigma_1, \dots, \sigma_k)$. Namely, if $w = g_{i_1}^{\alpha_1} g_{i_2}^{\alpha_2} \dots g_{i_m}^{\alpha_m}$ (with $\alpha_i \in \{-1, 1\}$) in reduced form, then associated with s_0 is the following path:

$$s_0 \xrightarrow{\sigma_{i_1}^{\alpha_1}} s_1 \xrightarrow{\sigma_{i_2}^{\alpha_2}} s_2 \xrightarrow{\sigma_{i_3}^{\alpha_3}} \dots \xrightarrow{\sigma_{i_m}^{\alpha_m}} s_m$$

with $s_1, \dots, s_m \in \{1, \dots, n\}$, and $s_b = \sigma_{i_b}^{\alpha_b}(s_{b-1})$ ($b = 1, \dots, m$). (Recall that we compose permutations from left to right.)

Likewise, we can speak of the trail through some power of w . For example, the trail of s_0 through $w^3(\sigma_1, \dots, \sigma_k)$ is

$$s_0 \xrightarrow{\sigma_{i_1}^{\alpha_1}} s_1 \xrightarrow{\sigma_{i_2}^{\alpha_2}} \dots \xrightarrow{\sigma_{i_m}^{\alpha_m}} s_m \xrightarrow{\sigma_{i_1}^{\alpha_1}} s_{m+1} \xrightarrow{\sigma_{i_2}^{\alpha_2}} \dots \xrightarrow{\sigma_{i_m}^{\alpha_m}} s_{2m} \xrightarrow{\sigma_{i_1}^{\alpha_1}} s_{2m+1} \xrightarrow{\sigma_{i_2}^{\alpha_2}} \dots \xrightarrow{\sigma_{i_m}^{\alpha_m}} s_{3m}$$

with s_1, \dots, s_m as before, and $s_{m+1}, \dots, s_{3m} \in \{1, \dots, n\}$ satisfying the obvious constraints.

We recall that two trails were placed in the same category if they have the same coincidence pattern. This notion can be extended to our present, more general context, in an obvious way. Namely, every category is associated with some directed edge-colored graph. Moreover, we can define categories not only of single trails, but also of lists of trails, and again, associate a graph to each category. The nature of this graph is exactly as described in Section 2.1.

To illustrate, let $w = g_1 g_2 g_1^{-1} g_2^{-1}$ be the commutator word, $n \geq 8$ an integer and $\sigma_1, \sigma_2 \in S_n$ such that the following trails are realized by $w(\sigma_1, \sigma_2)$ and $w^2(\sigma_1, \sigma_2)$, respectively:

$$\begin{array}{ccccccc} 1 & \xrightarrow{\sigma_1} & 3 & \xrightarrow{\sigma_2} & 7 & \xrightarrow{\sigma_1^{-1}} & 3 & \xrightarrow{\sigma_2^{-1}} & 8 \\ 4 & \xrightarrow{\sigma_1} & 6 & \xrightarrow{\sigma_2} & 8 & \xrightarrow{\sigma_1^{-1}} & 5 & \xrightarrow{\sigma_2^{-1}} & 5 & \xrightarrow{\sigma_1} & 8 & \xrightarrow{\sigma_2} & 3 & \xrightarrow{\sigma_1^{-1}} & 1 & \xrightarrow{\sigma_2^{-1}} & 2 \end{array}$$

We denote the nodes of the trail through w by s_0^1, \dots, s_4^1 and the nodes through w^2 by s_0^2, \dots, s_8^2 . Then the graph associated with the category of this list of trails through w, w^2 is shown in Fig. 7.

Although the notions here have a wider scope, we limit our discussion to categories of trails which represent cycles in $w(\sigma_1, \dots, \sigma_k)$: an L -cycle is represented by a closed trail

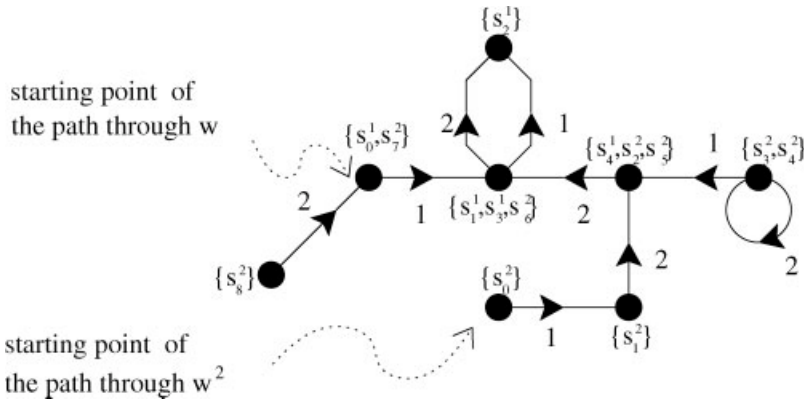


Fig. 7. An example of a graph associated with a certain category of a list of trails through w, w^2 , where $w = g_1 g_2 g_1^{-1} g_2^{-1}$ is the commutator word.

through $w^L(\sigma_1, \dots, \sigma_k)$. In fact, we are interested in counting cycles of a given length, so for our purposes we can confine ourselves to categories of a list of r trails through $w^L(\sigma_1, \dots, \sigma_k)$, for some $L, r \geq 1$.

First, let us analyze the categories that represent a single L -cycle. A closed trail through $w^L(\sigma_1, \dots, \sigma_k)$ represents an L -cycle only if it does not represent any smaller cycle. That is, in the closed trail

$$s_0 \xrightarrow{\sigma_{i_1}^{\alpha_1}} s_1 \xrightarrow{\sigma_{i_2}^{\alpha_2}} \dots \xrightarrow{\sigma_{i_m}^{\alpha_m}} s_m \xrightarrow{\sigma_{i_1}^{\alpha_1}} \dots \xrightarrow{\sigma_{i_m}^{\alpha_m}} s_{2m} \xrightarrow{\sigma_{i_1}^{\alpha_1}} \dots \xrightarrow{\sigma_{i_{m-1}}^{\alpha_{m-1}}} s_{Lm-1} \xrightarrow{\sigma_{i_m}^{\alpha_m}} s_0,$$

the L labels $s_0, s_m, s_{2m}, \dots, s_{(L-1)m}$ must all be distinct.

Once again, the graph of each category is a quotient of the universal graph. In this case, the universal graph is $C_{L \cdot m}$, the cycle of length $L \cdot m$. A partition of its vertices corresponds to some category of an L -cycle if and only if it does not create ‘‘collisions’’ of same-color edges, and keeps apart all L vertices corresponding to $s_0, s_m, s_{2m}, \dots, s_{(L-1)m}$. We demonstrate this in Fig. 8.

The most general case in our discussion comes up when we turn to calculate the r -th moment of $X_{w,L}^{(n)}$. To this end we consider categories of lists of r L -cycles through $w(\sigma_1, \dots, \sigma_k)$. The universal graph $\tilde{\Gamma}_{w,L,r}$ which represents r ordered cycles of length L each, consists of a disjoint union of r copies of $C_{L \cdot m}$. We name the vertices of the first cycle in $\tilde{\Gamma}_{w,L,r}$ by s_0^1, \dots, s_{Lm-1}^1 , the vertices of the second cycle by s_0^2, \dots, s_{Lm-1}^2 and so on until s_0^r, \dots, s_{Lm-1}^r for the r -th cycle.

We are interested in quotients (or partitions of the vertices) of $\tilde{\Gamma}_{w,L,r}$ that represent r distinct L -cycles. This means that the vertices

$$s_0^1, s_m^1, s_{2m}^1, \dots, s_{(L-1)m}^1, s_0^2, s_m^2, \dots, s_{(L-1)m}^2, \dots, s_0^r, s_m^r, \dots, s_{(L-1)m}^r$$

(a total of $L \cdot r$ vertices) should be kept apart in every partition, as illustrated in Fig. 9.

4.2. Formulas for the Moments of $X_{w,L}^{(n)}$

The method we use to calculate the expectation of $[X_{w,L}^{(n)}]^r$ is quite similar to the one we used for $\mathbb{E}(X_{w,1}^{(n)})$. We first calculate the ‘‘factorial moments’’ of $X_{w,L}^{(n)}$ and derive the regular

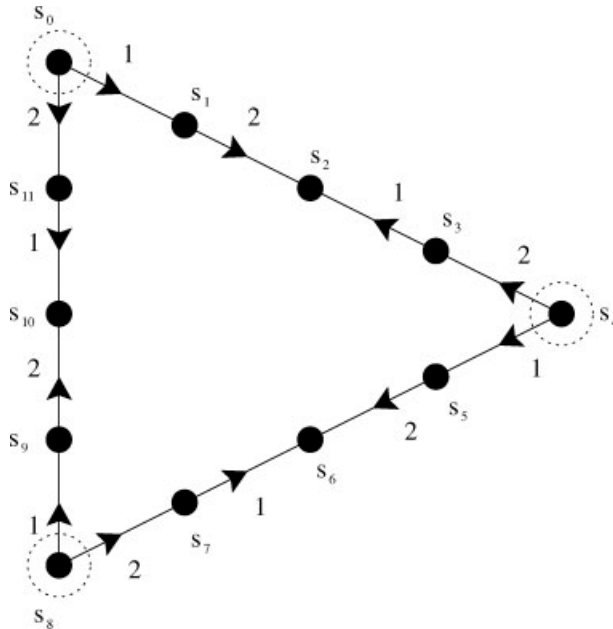


Fig. 8. The universal graph of a 3-cycle when $w = g_1 g_2 g_1^{-1} g_2^{-1}$. In a realizable partition of the vertices that represents a 3-cycle, the 3 circled vertices (s_0, s_4, s_8) must belong to distinct blocks.

moments from them (the r -th factorial moment of a random variable X is defined as $\mathbb{E}([X]_r)$, where $[X]_r$ is the “falling factorial”, namely $[X]_r = X(X - 1) \dots (X - r + 1)$).

$[X_{w,L}^{(n)}]_r$ counts lists of r L -cycles in $w(\sigma_1, \dots, \sigma_k)$. As in the case of the first moment, we calculate its expectation by counting the total number of lists of r L -cycles in $w(\sigma_1, \dots, \sigma_k)$ for all k -tuples $(\sigma_1, \dots, \sigma_k) \in S_n^k$ and dividing by $(n!)^k$.

The counting is carried out by classifying these lists into categories. Each list of r L -cycles with specified starting point for each cycle, belongs to some category. These categories are the quotients of the universal graph $\tilde{\Gamma}_{w,L,r}$, which we denote by $\mathcal{Q}_{w,L,r}$ (e.g., $\mathcal{Q}_{w,1,1}$ is the same set as \mathcal{Q}_w). To recap, the set $\mathcal{Q}_{w,L,r}$ can be generated as follows:

We first draw $\tilde{\Gamma}_{w,L,r}$, the universal graph of r ordered L -cycles of w , which consists of r disjoint cycles each of which has $L \cdot |w|$ vertices. $\mathcal{Q}_{w,L,r}$ consists of quotient graphs that are

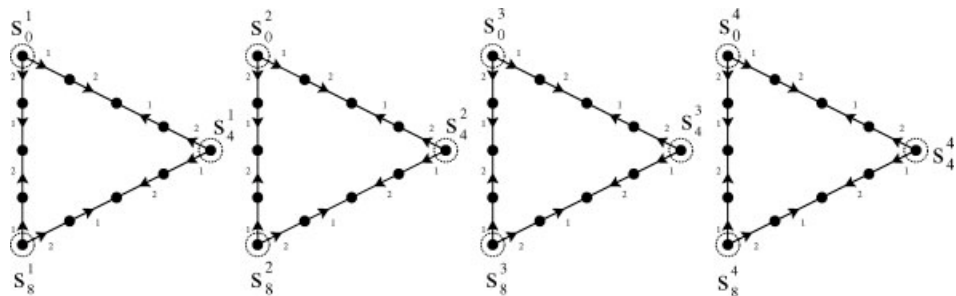


Fig. 9. The universal graph $\tilde{\Gamma}_{w,3,4}$ of four 3-cycles where $w = g_1 g_2 g_1^{-1} g_2^{-1}$. In a realizable partition of the vertices that represents four 3-cycles, the 12 circled vertices must belong to distinct blocks. (We omit the blocks corresponding to the rest of the vertices.)

generated by partitions of the vertices of $\tilde{\Gamma}_{w,L,r}$. A quotient graph is included in $\mathcal{Q}_{w,L,r}$ if it is realizable, and if in the corresponding partition each of the $r \cdot L$ vertices that represent the r L -cycles is in a different block.

Let Γ be some graph in $\mathcal{Q}_{w,L,r}$. A realization of Γ is a k -tuple of permutations $\sigma_1, \dots, \sigma_k \in S_n$, a list of r L -cycles of $w(\sigma_1, \dots, \sigma_k)$ and a specified starting point for each cycle, such that they belong to Γ 's category. The number of realizations of Γ is the same as in (5), namely:

$$N_\Gamma(n) = n(n-1) \dots (n - v_\Gamma + 1) \prod_{j=1}^k (n - e_\Gamma^j)! \tag{20}$$

Since every list of r L -cycles is counted L^r times (there are L^r ways to choose the starting points), we have:

$$\begin{aligned} \mathbb{E}([X_{w,L}^{(n)}]_r) &= \frac{1}{(n!)^k} \sum_{(\sigma_1, \dots, \sigma_k) \in S_n^k} [X_{w,L}^{(n)}(\sigma_1, \dots, \sigma_k)]_r \\ &= \frac{1}{(n!)^k} \cdot \frac{1}{L^r} \sum_{\Gamma \in \mathcal{Q}_{w,L,r}} N_\Gamma(n) \end{aligned} \tag{21}$$

$$= \frac{1}{L^r} \sum_{\Gamma \in \mathcal{Q}_{w,L,r}} \frac{n(n-1) \dots (n - v_\Gamma + 1)}{\prod_{j=1}^k n(n-1) \dots (n - e_\Gamma^j + 1)} \tag{22}$$

$$= \frac{1}{L^r} \sum_{\Gamma \in \mathcal{Q}_{w,L,r}} [n]_{v_\Gamma} \prod_{j=1}^k \frac{1}{[n]_{e_\Gamma^j}}$$

Note that the equality between (21) and (22) holds only for n large enough. Indeed $N_\Gamma(n) = 0$ if $n < e_\Gamma^j$ for some $\Gamma \in \mathcal{Q}_{w,L,r}$ and $j \in \{1, \dots, k\}$. This holds for $n \geq \max_{j=1, \dots, k} (e_j(\tilde{\Gamma}_{w,L,r}))$.

For every $L \geq 1$ and $r \geq 1$, (22) thus yields a rational function in n , which, for sufficiently large n , is the r -th factorial moment of $X_{w,L}^{(n)}$. It is convenient to rewrite (22) as a function of $\frac{1}{n}$:

$$\mathbb{E}([X_{w,L}^{(n)}]_r) = \frac{1}{L^r} \sum_{\Gamma \in \mathcal{Q}_{w,L,r}} \left(\frac{1}{n}\right)^{e_\Gamma - v_\Gamma} \frac{\prod_{t=1}^{v_\Gamma-1} (1 - \frac{t}{n})}{\prod_{j=1}^k \prod_{t=1}^{e_\Gamma^j-1} (1 - \frac{t}{n})} \tag{23}$$

We can now define a rational function $\psi_{w,L,r}$ by:

$$\psi_{w,L,r}(x) = \frac{1}{L^r} \sum_{\Gamma \in \mathcal{Q}_{w,L,r}} x^{\chi(\Gamma)-1} \frac{\prod_{t=1}^{v_\Gamma-1} (1 - tx)}{\prod_{j=1}^k \prod_{t=1}^{e_\Gamma^j-1} (1 - tx)} \tag{24}$$

and so $\mathbb{E}([X_{w,L}^{(n)}]_r) = \psi_{w,L,r}(\frac{1}{n})$ for $n \geq \max_{j=1, \dots, k} (e_j(\tilde{\Gamma}_{w,L,r}))$.

The following lemma shows that $\psi_{w,L,r}$ is well defined on a neighborhood of 0.

Lemma 27. For each $\Gamma \in \mathcal{Q}_{w,L,r}$, $\chi(\Gamma) \geq 1$.

Proof. Note that in $\tilde{\Gamma}_{w,L,r}$ every vertex has degree 2, since w is cyclically reduced. This degree can not decrease in a quotient. Therefore, every vertex in every $\Gamma \in \mathcal{Q}_{w,L,r}$ has degree at least 2, and the lemma follows. ■

It is well-known how to express the regular moments as linear combinations of the factorial moments. Thus, we can derive (an efficiently computable) rational function $\Psi_{w,L,r}$ (which is a linear combination of $\psi_{w,L,1}, \dots, \psi_{w,L,r}$), such that $\mathbb{E}([X_{w,L}^{(n)}]^r) = \Psi_{w,L,r}(1/n)$ for sufficiently large n . This function $\Psi_{w,L,r}$ is obviously defined on a neighborhood of 0, which proves the first part of Corollary 26.

4.3. Proving Theorem 25 With the Method of Moments

The proof of Theorem 25 is based on the method of moments. Under certain mild conditions, a probability distribution is determined by its moments, or as here, a limit distribution is determined by the limits of the moments.

4.3.1. Some Facts From the Method of Moments. A probability measure μ on \mathbb{R} is said to be determined by its moments if it has finite moments $\alpha_r = \int_{-\infty}^{\infty} x^r \mu(dx)$ of all orders, and μ is the only probability measure with these moments. We quote Theorem 30.2 from [17]:

Theorem 28. *Let X and X_n ($n \in \mathbb{N}$) be random variables, and suppose that the distribution of X is determined by its moments, that the X_n have moments of all order, and that $\lim_{n \rightarrow \infty} \mathbb{E}(X_n^r) = \mathbb{E}(X^r)$ for every $r \in \mathbb{N}$. Then*

$$X_n \xrightarrow{\text{dis}} X.$$

Note that if X and the X_n are integer-valued then $X_n \xrightarrow{\text{dis}} X$ is equivalent to $Pr(X_n = k) \rightarrow Pr(X = k)$ for every integer k .

The relation between regular moments and factorial moments implies:

Corollary 29. *The statement of Theorem 28 holds where moments are replaced by factorial moments.*

In this section, we use Corollary 29 to prove Theorem 25. That $X_{w,L}^{(n)}$ has moments of all order is evident. We still need to show that the r.h.s of (18) is determined by its moments, and that the r -th factorial moment of $X_{w,L}^{(n)}$ indeed converges to the r -th factorial moment of this random variable.

Theorem 30.1 from [17] provides a sufficient condition for a probability measure μ to be determined by its moments, namely, that the power series $\sum_r \alpha_r t^r / r!$ where α_r is the r -th moment of μ has a positive radius of convergence. (This series is the moment generating function of μ , when the latter exists.) For μ a Poisson distribution (with any parameter), this power series converges for all real t (e.g., [18], section 4), so μ is determined by its moments. A convolution (a summation) of several Poisson distributions is itself Poisson (whose parameter is the sum of parameters), and thus satisfies the condition as well.

In particular, recall that the r.h.s. of (18) is $\sum_{h \in H(d,L)} h Z_{1/Lh}$, where $Z_m \sim Poi(m)$. If we omit the constant h of every term in this sum, we obtain $\sum_{h \in H(d,L)} Z_{1/Lh}$, whose distribution is simply $Poi(\sum_{h \in H(d,L)} 1/Lh)$, which is determined by its moments. According to the

definition of $H(d, L)$ in (19), each $h \in H(d, L)$ satisfies $1 \leq h \leq d$. Thus, if we denote by α_r the r -th moment of $\sum_{h \in H(d, L)} Z_{1/Lh}$ and by β_r the r -th moment of $\sum_{h \in H(d, L)} hZ_{1/Lh}$, then $\alpha_r \leq \beta_r \leq d^r \alpha_r$. Consequently, the series $\sum_r \beta_r t^r / r!$ has radius of convergence that is $\geq \frac{1}{d}$ that of the series $\sum_r \alpha_r t^r / r!$. But the latter converges for all real t , hence so does $\sum_r \beta_r t^r / r!$, and the distribution of $\sum_{h \in H(d, L)} hZ_{1/Lh}$ is determined by its moments.

To conclude the proof of Theorem 25, we need to show that the (factorial) moments of $X_{w, L}^{(n)}$ indeed converge to the respective moments of the r.h.s. of (18). But what is the limit of $\mathbb{E}([X_{w, L}^{(n)}]_r)$? By Lemma 27, the limit of each term in the r.h.s. of (23) is either 0 (if $\chi(\Gamma) \geq 2$), or 1 (if $\chi(\Gamma) = 1$). Therefore,

$$\lim_{n \rightarrow \infty} \mathbb{E}([X_{w, L}^{(n)}]_r) = \frac{1}{L^r} |\{\Gamma \in \mathcal{Q}_{w, L, r} : \chi(\Gamma) = 1\}| \tag{25}$$

As explained in the proof of Lemma 27, the equality $\chi(\Gamma) = 1$ holds for some $\Gamma \in \mathcal{Q}_{w, L, r}$, iff every vertex in Γ has degree 2, i.e., iff Γ is a disjoint union of cycles.

We denote by $\mathcal{C}_{w, L, r}$ the subset of $\mathcal{Q}_{w, L, r}$ consisting of all graphs which are a disjoint union of cycles. (25) now becomes:

$$\lim_{n \rightarrow \infty} \mathbb{E}([X_{w, L}^{(n)}]_r) = \frac{1}{L^r} |\mathcal{C}_{w, L, r}| \tag{26}$$

4.3.2. The Graphs in $\mathcal{C}_{w, L, r}$. Let $w \in \mathbf{F}_k$ be cyclically reduced and equal u^d with u primitive and $d \geq 1$. A graph $\Gamma \in \mathcal{C}_{w, L, r}$ has a very specific structure: Each cycle c in Γ must be a cyclic concatenation of several copies of u (every cycle in Γ looks like $\tilde{\Gamma}_{u, b, 1}$ for some positive integer b).

To see this, recall that each cycle c in Γ represents a closed trail through w^L (at least one closed trail). Hence there is some vertex x in c and some orientation on c , such that if we leave x in this orientation and go exactly $d \cdot L$ times through u , we get back to x (possibly after tracing c several times). Since u is primitive, it cannot be invariant under cyclic shift of any length $l < |u|$, and the size of c must divide $|u|$. (In fact, the integer $|c|/|u|$ divides $d \cdot L$).

Moreover, all closed trails that are represented in c , go in the same direction (and thus also start in one of the $|c|/|u|$ head vertices of u). We already saw in the proof of Lemma 14 that for u primitive, u^{-1} is not a subword of u^2 . This rules out the possibility of “finding u in the opposite direction.”

This analysis of the structure of the graphs in $\mathcal{C}_{w, L, r}$ yields an important corollary, which ultimately explains why the limit distribution of $X_{w, L}^{(n)}$ depends solely on d and L , and not on u :

Corollary 30. *Let $w_1, w_2 \in \mathbf{F}_k$ be cyclically reduced and equal u_1^d and u_2^d , respectively, with u_1 and u_2 primitive and $d \geq 1$. Then*

$$|\mathcal{C}_{w_1, L, r}| = |\mathcal{C}_{w_2, L, r}|$$

Proof. The analysis above shows that the inner structure of u is completely irrelevant to the graphs in $\mathcal{C}_{w, L, r}$, and there is a natural bijection between $\mathcal{C}_{w_1, L, r}$ and $\mathcal{C}_{w_2, L, r}$: simply replace each copy of u_1 by a copy of u_2 . ■

4.3.3. *The Simple Case of a Primitive Word.* The material in this section is not necessary for the proof and deals with the special case of primitive words. Our hope is that this special case makes it easier to follow the general proof. So let us assume that $u = w$ and $d = 1$. The universal graph $\tilde{\Gamma}_{w,L,r}$ is a disjoint union of r cycles, each of which consists of L copies of u . We have a total of $r \cdot L$ copies of u , and their $r \cdot L$ initial vertices are kept separated in each quotient. It is easily verified that the only quotient of $\tilde{\Gamma}_{w,L,r}$ which is a disjoint union of cycles is $\tilde{\Gamma}_{w,L,r}$ itself, and thus $|\mathcal{C}_{w_1,L,r}| = 1$. By (26) we have $\lim_{n \rightarrow \infty} \mathbb{E}([X_{w,L}^{(n)}]_r) = \frac{1}{L^r}$.

On the other hand, when w is primitive, the r.h.s. of (18) is simply $Z_{1/L}$. Let X be an integer-valued non-negative random variable and let $f_X(t)$ be its generating function

$$f_X(t) = \sum_{j=0}^{\infty} Pr(X = j)t^j.$$

Under certain mild conditions and in particular if $f_X(t)$ is analytic on \mathbb{R} ,

$$\mathbb{E}([X]_r) = f_X^{(r)}(1) \quad \forall r \in \mathbb{N}.$$

The generating function of $Z_{1/L}$ is $f(t) = e^{-\frac{1}{L}} e^{\frac{t}{L}}$. Thus,

$$\mathbb{E}([Z_{1/L}]_r) = f^{(r)}(1) = \frac{1}{L^r}$$

which completes the proof of Theorem 25 for w primitive.

4.3.4. *The General Case.* To complete the proof of Theorem 25, we show that for every $w = u^d \in \mathbf{F}_k$, $L \geq 1$ and $r \geq 1$, the series $\mathbb{E}([X_{w,L}^{(n)}]_r)$ converges, for $n \rightarrow \infty$, to $\mathbb{E}([Y_{d,L}]_r)$, where $Y_{d,L}$ is the r.h.s. of (18), i.e. $Y_{d,L} = \sum_{h \in H(d,L)} hZ_{1/Lh}$.

Let X_1 and X_2 be non-negative integer-valued variables with generating functions $f_{X_1}(t), f_{X_2}(t)$. If $Y = X_1 + X_2$, then clearly Y 's generating function is $f_Y(t) = f_{X_1}(t) \cdot f_{X_2}(t)$. The generating function of hZ_m ($h \in \mathbb{N}, Z_m \sim Poi(m)$) is $f_{hZ_m}(t) = e^{-m} \cdot e^{mt^h}$. Therefore,

$$f_{Y_{d,L}} = \prod_{h \in H(d,L)} e^{-\frac{1}{Lh}} e^{\frac{h}{Lh}}$$

and if $H(d, L) = \{h_1, \dots, h_p\}$, then

$$\begin{aligned} \mathbb{E}([Y_{d,L}]_r) &= f_{Y_{d,L}}^{(r)}(t)|_{t=1} \\ &= \sum_{\substack{r_1 + \dots + r_p = r \\ r_j \geq 0}} \binom{r}{r_1 \dots r_p} \prod_{j=1}^p f_{h_j Z_{1/Lh_j}}^{(r_j)}(t)|_{t=1} \end{aligned} \tag{27}$$

By (26), the series $\mathbb{E}([X_{w,L}^{(n)}]_r)$ converges to $\frac{1}{L^r} |\mathcal{C}_{w,L,r}|$, so we proceed to analyze the set $\mathcal{C}_{w,L,r}$. Let Γ be a graph in $\mathcal{C}_{w,L,r}$. As explained in section 4.3.2, Γ is a disjoint union of cycles, each of which consists of several copies of u . If we let b denote the number of copies of u in some cycle, what are the possible values of b ? To begin with, $b|dL$, as this cycle represents a closed path that consists of dL copies of u . Second, as this cycle represents an L -cycle, $L|b$ and $b \nmid dL'$ for any $1 \leq L' < L$. If we let $h = \frac{b}{L}$, then the constraints on b translate into the following conditions: $h|d$ and $(\frac{d}{h}, L) = 1$, precisely as in the definition of $H(d, L)$.

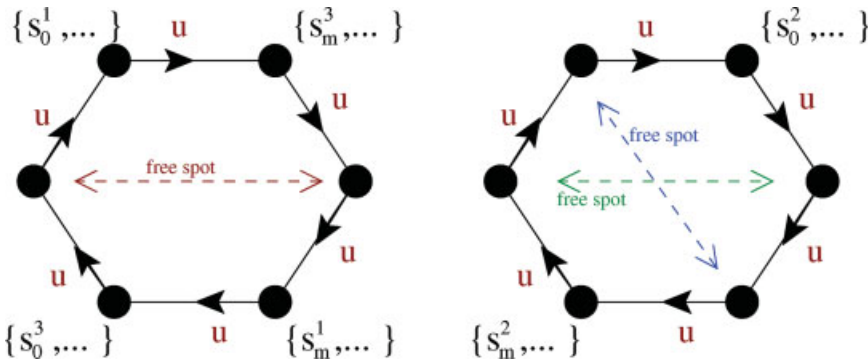


Fig. 10. One of the graphs $\Gamma \in \mathcal{C}_{w,L,r}^h$, where $w = u^3$ for some primitive $u \in \mathbf{F}_k$, $h = 3$, $L = 2$ and $r = 3$. (We use the labels s_i^j introduced in Section 4.1) This Γ contains two cycles, each of which consists of six copies of u . Each of the cycles can correspond to up to three distinct L -cycles (indeed, $h = 3$), so Γ can contain up to six L -cycles. As it already contains three, it has $6 - 3 = 3$ free spots.

Note that a cycle in Γ with $b = Lh$ copies of u can represent up to h distinct L -cycles in $w(\sigma_1, \dots, \sigma_k)$.

As before, let $H(d, L) = \{h_1, \dots, h_p\}$. For $j = 1, \dots, p$ consider those L -cycles which are associated in the quotient Γ to a cycle of length $Lh_j \cdot |u|$. (The i -th L -cycle belongs to the cycle c in Γ if the blocks containing $s_0^i, \dots, s_{L|w|-1}^i$ correspond to vertices in c .) Let r_j be the number of such L -cycles whence $\sum r_j = r$, and there are $\binom{r}{r_1 \dots r_p}$ ways to choose which L -cycles go where (Recall that $\tilde{\Gamma}_{w,L,r}$ consists of an ordered list of r L -cycles). Now let $\mathcal{C}_{w,L,r}^h$ denote the subset of $\mathcal{C}_{w,L,r}$ consisting of all quotient graphs where all disjoint cycles are of equal length of $Lh|u|$ each. Then we have:

$$\begin{aligned} \lim_{n \rightarrow \infty} \mathbb{E}([X_{w,L}^{(n)}]_r) &\stackrel{(26)}{=} \frac{1}{L^r} |\mathcal{C}_{w,L,r}| = \frac{1}{L^r} \sum_{\substack{r_1 + \dots + r_p = r \\ r_j \geq 0}} \binom{r}{r_1 \dots r_p} \prod_{j=1}^p |\mathcal{C}_{w,L,r_j}^{h_j}| \\ &= \sum_{\substack{r_1 + \dots + r_p = r \\ r_j \geq 0}} \binom{r}{r_1 \dots r_p} \prod_{j=1}^p \frac{1}{L^{r_j}} |\mathcal{C}_{w,L,r_j}^{h_j}| \end{aligned} \tag{28}$$

($\mathcal{C}_{w,L,0}^h$ denotes the singleton of the empty graph, and therefore $|\mathcal{C}_{w,L,0}^h| = 1$).

By combining (27) and (28), we conclude that Theorem 25 will follow if we show

$$\frac{1}{L^r} |\mathcal{C}_{w,L,r}^h| = f_{hZ_{1/Lh}}^{(r)}(t) \Big|_{t=1}, \tag{29}$$

for every $L \geq 1, r \geq 0$ and $h \in H(d, L)$.

We begin with the l.h.s. of (29). Recall that by definition, each graph $\Gamma \in \mathcal{C}_{w,L,r}^h$ consists of a disjoint union of cycles of length $Lh|u|$ each. Thus, each cycle represents up to h distinct L -cycles of $w(\sigma_1, \dots, \sigma_k)$, and Γ can represent up to $h \cdot (\# \text{ cycles in } \Gamma)$ distinct L -cycles. But Γ represents only r distinct L -cycles, whence there are $h \cdot (\# \text{ cycles in } \Gamma) - r$ “free spots” in Γ that can contain new L -cycles. We illustrate these notions in Fig. 10.

Now let $\alpha_{w,L,r}^h[j]$ denote the number of graphs in $\mathcal{C}_{w,L,r}^h$ with j free spots. We can define the generating function of the $\alpha_{w,L,r}^h[j]$:

$$g_{w,L,r}^h(t) = \alpha_{w,L,r}^h[0] + \alpha_{w,L,r}^h[1]t + \alpha_{w,L,r}^h[2]t^2 + \dots$$

and obviously $g_{w,L,r}^h(1) = |\mathcal{C}_{w,L,r}^h|$.

Before we derive a recursion formula for this function, we want to illustrate by writing explicit expressions for $r = 0, 1, 2$. Every connected component (=cycle) in every $\Gamma \in \mathcal{C}_{w,L,r}^h$, realizes at least one L -cycle. Thus, when $r = 0$ and there are no L -cycles at all, we have only the empty graph which has no free spots, and so $g_{w,L,0}^h(t) = 1$. When $r = 1$, we have a single graph in $\mathcal{C}_{w,L,1}^h$, with $h - 1$ free spots (a single cycle of Lh copies of u), and therefore $g_{w,L,1}^h(t) = t^{h-1}$. For $r = 2$ there is always a two-cycle graph with one L -cycle in each cycle. This graph has $2(h - 1)$ free spots. In addition, if $h \geq 2$, there are also graphs consisting of a single cycles that corresponds to two L -cycles. There are $L(h - 1)$ ways to place the two L -cycles and such graphs have $(h - 2)$ free spots. Thus, for $h = 1$, $g_{w,L,2}^h(t) = t^{2(h-1)}$ and for $h \geq 2$ $g_{w,L,2}^h(t) = L(h - 1)t^{h-2} + t^{2(h-1)}$.

We now want to derive the functions $g_{w,L,r}^h$ by recursing on r . Let Γ be a graph in $\mathcal{C}_{w,L,r}^h$ with j free spots. In what manners can we add another L -cycle and make it a graph in $\mathcal{C}_{w,L,r+1}^h$? We have two options: we can put the new L -cycle in one of the j free spots, in L possible ways (L possible cyclic shifts), resulting in $j \cdot L$ different graphs in $\mathcal{C}_{w,L,r+1}^h$, each of which has $j - 1$ free spots. Alternatively, we can add one new cycle to Γ and put there our new L -cycle, which yields a single graph in $\mathcal{C}_{w,L,r+1}^h$ with $j + h - 1$ free spots. Thus, we have:

$$g_{w,L,r+1}^h(t) = L \cdot (g_{w,L,r}^h(t))' + t^{h-1} \cdot g_{w,L,r}^h(t).$$

We now go back to the right side of (29). Recall that $f_{hZ_1/Lh}(t) = e^{-\frac{1}{Lh}} e^{\frac{t^h}{Lh}}$. If we write $f_{hZ_1/Lh}^{(r)}(t) = e^{-\frac{1}{Lh}} e^{\frac{t^h}{Lh}} \cdot q_{L,r}^h(t)$ where $q_{L,r}^h(t)$ is the appropriate polynomial, then

$$q_{L,0}^h(t) = 1$$

and

$$q_{L,r+1}^h(t) = (q_{L,r}^h(t))' + \frac{1}{L} \cdot t^{h-1} \cdot q_{L,r}^h(t)$$

Thus $g_{w,L,r}^h(t) = L^r \cdot q_{L,r}^h(t)$, and we can conclude:

$$\frac{1}{L^r} |\mathcal{C}_{w,L,r}^h| = \frac{1}{L^r} \cdot g_{w,L,r}^h(1) = q_{L,r}^h(1) = f_{hZ_1/Lh}^{(r)}(1) \tag{30}$$

when the last equality comes from the fact that $e^{-\frac{1}{Lh}} e^{\frac{t^h}{Lh}} \Big|_{t=1} = 1$.

The proof of Theorem 25 is now complete. Corollary 26 is also proved. For completeness sake, here is a short proof of the last sentence in the corollary, namely, that $\lim_{n \rightarrow \infty} \mathbb{E}(X_{w,L}^{(n)}) = \frac{|H(d,L)|}{L}$.

Proof.

$$\begin{aligned} \lim_{n \rightarrow \infty} \mathbb{E}(X_{w,L}^{(n)}) &= \mathbb{E}(Y_{d,L}) = \mathbb{E}\left(\sum_{h \in H(d,L)} hZ_{1/Lh}\right) = \sum_{h \in H(d,L)} \mathbb{E}(hZ_{1/Lh}) \\ &= \sum_{h \in H(d,L)} h \cdot \frac{1}{Lh} = \frac{|H(d,L)|}{L} \end{aligned}$$

■

Remark 31. In fact, the techniques presented here are likely to yield further results. The method of moment applies as well to random vectors (for distributions that are determined by their moments, see, e.g., [19], Theorem 6.2). The joint moments of $X_{w,L_1}^{(n)}, \dots, X_{w,L_k}^{(n)}$ for some k and positive integers L_1, \dots, L_k can be analyzed similarly to the way we analyzed the moments of $X_{w,L}^{(n)}$ for some L , and the limit joint distribution of these variables is probably determined by its moments.

5. OPEN PROBLEMS

Many interesting questions and conjectures were raised in this article. We collect them here.

- Let u and w be two words such that for any finite group G , the distribution of the two word maps on G are identical. Is it true that $u \sim w$?
- (Conjecture 15) $\beta(w) = \phi(w)$ for every word w .
- (A consequence of Conjecture 15:) For every word w , and sufficiently large n , a random permutation in the image of w in S_n has, on average, at least one fixed point.
- Friedman's Conjecture: For every base graph G , almost surely all new eigenvalues in lifts of G are $\leq \rho + o(1)$.
- Nica's theorem determines the behavior of the number of L -cycles in the S_n -image of any formal word w . There are numerous other parameters of such permutations (e.g., the number of cycles) whose typical behavior is still not understood.

REFERENCES

- [1] N. Linial and E. Rozenman, Random lifts of graphs: Perfect matchings, *Combinatorica* 25 (2005), 407–424.
- [2] A. Amit and N. Linial, Random lifts of graphs II: edge expansion, *Comb Probab Comput* 15 (2006), 317–332.
- [3] A. Amit, N. Linial, and J. Matousek, Random lifts of graphs III: Independence and chromatic number, *Random Struct Algorithms* 20 (2002), 1–22.
- [4] S. Hoory, N. Linial, and A. Wigderson, Expander graphs and their applications, in *Bulletin of the American Mathematical Society*, Vol. 43, Number 4, October 2006, ACM, New York, NY, pp. 439–531.
- [5] J. Friedman, Relative expanders or weakly relatively ramanujan graphs, *Duke Math J* 118 (2003), 19–35.
- [6] M. W. Buck, Expanders and diffusers, *SIAM J Algebraic Discrete Methods* 7 (1986), 282–304.

- [7] Y. Greenberg, On the spectrum of graphs and their universal coverings, PhD thesis, Hebrew University of Jerusalem, Hebrew, 1995.
- [8] P. Cartier, Fonctions harmonique sur un arbre, In *Symposia mathematica*, Vol. IX (Convegno di calcolo delle Probabilita, INDAM, Rome, 1971), Academic Press, London, 1972, pp. 203–270.
- [9] A. Nilli, On the second eigenvalue of a graph, *Discrete Math* 91 (1991), 207–210.
- [10] A. Broder and E. Shamir, On the second eigenvalue of random regular graphs, in *The 28th Annual Symposium on Foundations of Computer Science*, Seattle, WA, 1987, pp. 286–294.
- [11] J. Friedman, A proof of Alon’s second eigenvalue conjecture and related problems, *Mem Amer Math Soc* 195 (2008), no. 910.
- [12] E. Wigner, Characteristic vectors of bordered matrices with infinite dimensions, *Ann Math* 62 (1955), 548–564.
- [13] M. Larsen and A. Shalev, Word maps and Waring type problems, *J Amer Math Soc* 22 (2009), 437–466.
- [14] A. Nica, On the number of cycles of given length of a free word in several random permutations, *Random Struct Algorithms* 5 (1994), 703–730.
- [15] W. Magnus, A. Karrass, and D. Solitar, *Combinatorial group theory*, Dover, New York, 1966.
- [16] M. Abért, On the probability of satisfying a word in a group, *J Group Theory* 9 (2006), 685–694.
- [17] P. Billingsley, *Probability and measure*, 3rd edition, Wiley-Interscience, New York, 1995.
- [18] J. Pitman, Some probabilistic aspects of set partitions, *The American mathematical monthly* 104 (1997), 201–209.
- [19] S. Janson, T. Łuczak, and A. Rucinski, *Random graphs*, Wiley-Interscience, New York, 2000.
- [20] Y. Bilu and N. Linial, Lifts, discrepancy and nearly optimal spectral gaps, *Combinatorica* 26 (2006), 495–519.
- [21] J. Friedman, On the second eigenvalue and random walks in random d -regular graphs, *Combinatorica* 11 (1991), 331–362.