Illinois Journal of Mathematics
Volume 54, Number 1, Spring 2010, Pages 327–370
S 0019-2082

GROWTH IN FREE GROUPS (AND OTHER STORIES)—TWELVE YEARS LATER

IGOR RIVIN

To Paul Schupp, with the greatest affection

Abstract. We start by studying the distribution of (cyclically reduced) elements of the free groups F_n with respect to their Abelianization (or equivalently, their class in $H_1(F_n, \mathbf{Z})$). We derive an explicit generating function, and a limiting distribution, by means of certain results (of independent interest) on Chebyshev polynomials; we also prove that the reductions $\bmod p$ (p—an arbitrary prime) of these classes are asymptotically equidistributed, and we study the deviation from equidistribution. We extend our techniques to a more general setting and use them to study the statistical properties of long cycles (and paths) on regular (directed and undirected) graphs. We return to the free group to study some growth functions of the number of conjugacy classes as a function of their cyclically reduced length.

Introduction—2010

The paper "Growth in free groups (and other stories)," has been around in preprint form ([45]) since the late nineties (the arXiv version cited dates to 1999, but this was preceded by a 1997 IHES preprint). Since the paper has had a fair amount of influence (and parts of it have since become separate papers), it seems a good idea to publish it at last—this version is not very different from the preprint, except for this introduction, which gives a bit of background on how and why it was written together with a survey (necessarily incomplete and subjective) of what has happened since the arXiv preprint appeared in 1999.

Received March 25, 2010; received in final form November 18, 2010.
2000 *Mathematics Subject Classification.* Primary 05C25, 05C20, 05C38, 60J10, 60F05, 42A05. Secondary 22E27.

Why? The work described in the paper was initially motivated by the author's (continuing to this day) interest in the counting questions on geodesics on hyperbolic surface, stemming from some conversations with Peter Sarnak in the early 1990s. More precisely, Sarnak had asked about the asymptotics of the number of *simple* geodesics on the punctured torus, where the only result appeared to be the one in the paper of Beardon, Lehner, and Sheingorn [3], where the authors had shown that the number of simple geodesics of length bounded by L grew somewhere between quadratically and quartically in L. This did not seem to be very sharp, and indeed, Greg McShane and I improved it to an asymptotic result (with quadratic growth) in a pair of short papers [38, 39], using purely geometric methods (showing that the length of the *unique* shortest geodesic (which can be showed to be simple) in a primitive integral homology class extends to a norm on real homology (which is the Gromov, or the stable norm, though at the time McShane and I had no knowledge of the connection). The fact that there is at most one simple closed geodesic in a homology class is specific to the punctured torus, and while other methods can be used to compute the asymptotics of the number of simple closed geodesics of bounded length on a surface of finite type (the order of growth was computed by the author in [46], while asymptotics were computed by Maryam Mirzakhani in [40]—see also [49]), the following question is still wide open:

How many simple curves of length bounded by L are there in a fixed homology class h on a hyperbolic surface? Mirzakhani's work implies that a constant proportion of all simple geodesics are separating, but for a nontrivial homology class nothing seems known to-date.

Geodesics in homology classes. Given the interest in geodesics and homology, it was natural to investigate a similar question for *all* closed geodesics, not necessarily simple. It is a well-known result of Huber (for hyperbolic surfaces—Huber uses the Selberg Trace Formula)—[18]–[20] and Margulis [35, 36] for arbitrary negatively curved surfaces, using ergodic theory) that the number of closed geodesics of length bounded by L *without* homological restrictions is asymptotic to $\exp hL/(hL)$, where h is the topological entropy of the geodesic flow ($h = 1$ for a hyperbolic surface). The methods used by Huber and Margulis (Selberg Trace Formula and ergodic dynamics, respectively) are the two principal tools used in the vast majority of the paper discussed below (generally either one technique or the other, but not both, generally because the Trace Formula gets sharp results but only works in the constant curvature setting, while dynamical methods are softer, so give weaker results in a wider setting).

The first result on geodescis in homology classes is due to Parry and Pollicott—in their paper [41] they show that when the homology group $H_1(S, \mathbf{Z})$ is *finite*, then closed geodesics are equidistributed among homology classes. Parry and Pollicott use the machinery of thermodynamic formalism

and dynamical zeta functions, and their argument mimics the proof of the Chebotarev density theorem. Parry and Pollicott's methods work in variable negative curvature, and they also analyze the lifting of geodesics in a homology class to (finite) Galois covers. Roughly concurrently, Katsuda and Sunada showed in [28] that for homology with coefficients in a finite group, every homology class contains an infinite number of closed geodesics (but no estimate of the growth of their number as a function of length).

The next result is due to Adachi and Sunada—in the paper [1], they show that the exponential growth rate of the number curves in any homology class is equal to h (just like for homologically unrestricted geodesics)—they use Markov partitions as introduced by Bowen in [5] and use results on paths in finite graphs to get the result (which is rather weak, since they do not actually get an asymptotic result). They point out that getting such a result (via the usual Tauberian machinery) would require an understanding of the singularity of the L-functions involved greater than they could produce at the time. They conjecture that the the number of geodesics of length bounded by L in a homology class should grow like $\exp(hL)/(L^{b+1})$, where b is the first Betti number of the manifold.

This conjecture turns out to be false—in the paper [44], published almost simultaneously with [1], Phillips and Sarnak give an asymptotic expansion valid for a *hyperbolic* surface: the number of closed geodesics in a fixed homology class, of length bounded by L grows as

$$\frac{e^L}{L^{g+1}}(1 + c_1/L + c_2/L^2 + \cdots),$$

where $c_1, \ldots, c_k, \ldots$ depend on the homology class. This sort of expansion appeared (at the time) to be possible only because the manifold had *constant* negative curvature. The work of Phillips and Sarnak was extended (again, approximately at the same time) by Epstein to *cusped* surfaces in [8], again using the Selberg Trace Formula. As often with these kinds of extensions, the result is a lot harder technically than the Phillips–Sarnak result.

At roughly the same time, Katsuda and Sunada extended the dynamical methods of [1] first to surfaces of constant negtive curvature in [29] (by observing that the complicated L-function that could not be dealt with in [1] became much simpler in constant curvature), and then for general negatively curved surfaces in [30].

Last, but not least, Lalley uses the thermodynamical formalism and some fairly intricate harmonic analysis in [33] to recover the results of Katsuda-Sunada, and more: He shows a central limit theorem for the distribution of homology classes of closed geodesics, and also a "large deviation result." Lalley's result is closest in spirit to the current paper, but the methods are completely different (and I had no knowledge of the paper's existence until this writing).

Some motivation. All of the results mentioned in the survey above are technically quite involved, and it was not clear what was really going on. This is what gave birth to the current paper. One observation was that it is a lot easier to work with groups (especially free groups) than with surfaces, and secondly, since fundamental groups are often quasi-isometric to the spaces they are fundamental groups of, one has the hope of obtaining "universal" results (that is, a result for a surface group implies a result (usually somewhat weaker) for every surface of the appropriate type.

One particular insight (on which much of the paper is based) is the observation that for graphs, the Selberg Trace Formula (quite pervasive in the work surveyed above) is a triviality: the number of closed (based) cycles of length N in the graph is the trace of the Nth power of the adjacency matrix, and thus the sum of the Nth powers of eigenvalues of the adjacency matrix of the graph. In the particular case where the graph is undirected, the adjacency matrix is symmetric, and analysis becomes easy. Technically simpler methods (based in large part on perturbation theory for eigenvalues) have helped to get results of much wider scope than previously. Let us now review the results and their follow-up in subsequent years.

Then what happened? Free groups and related subjects. In Section 1 we have set up the basic model, and used it to count cyclically reduced words in a free group. The basic method works for any automatic group, and if the structure is bi-automatic, we similarly get an undirected graph. Somewhat surprisingly, the count of cyclically reduced words has been used in a number of papers (see, e.g., [7, 25]), and in the paper [31] by Koganov it is shown that the formula is equivalent to Whitney's formula for the chromatic polynomial of the cycle graph. Koganov had apparently published two other papers (in 2002 and 2004) deriving the enumeration of cyclically reduced words—see references [1] and [2] in [31].

A related question is considered in Sections 13–15, where we study the number of *conjugacy classes* of fixed minimal length in the free group (and elsewhere). We construct an ordinary generating function (in the form of a Lambert Series, see [16] for definition), which turns out to be horribly irrational (this result has gone on to have a life of its own in [47]), and the zeta function enumerating *primitive* conjugacy classes, which turns out to be an Ihara-type zeta function of the defining graph (see also the papers of Stark and Terras [59]–[61]). The conjecture that the (standard) generating function is irrational for all nonvirtually-cyclic Gromov-hyperbolic groups is still open. The Ihara zeta function immediately gives asymptotic growth rates for primitive classes, however this is computed again by Coornaert in [7].

In Section 2, we write down explicit generating functions for the number of elements in the free group with a given Abelianization. These formulas can be expressed as Chebyshev polynomials—this is so, because the adjacency matrix of the "recognizing automaton" graph has only two nontrivial eigenvalues, and

this is special to free groups. It would be interesting to write down formulas of this type for example, surface groups, and see what special functions arise.

The fact that certain variations on Chebyshev polynomials arise as generating functions give previously unknown positivity result on combinations of their coefficients and shows that the functions $T_n(c\cos x)$ and $U_n(c\cos x)$, where T and U are Chebyshev polynomials of first and second kind respectively, and $c > 1$ are positive semi-definite in the sense of Bochner. This, and the Central Limit theorem for the coefficients of "Symmetrized Chebyshev Polynomials" appear in the author's paper [48].

The Central Limit theorem for distribution of elements of the free group F_n is proved in Section 3, but the methods actually go through without much change to prove a "Local Limit Theorem." Such a theorem was also shown by Sharp, using much more heavy lifting in his paper [57]. The Central Limit theorem was reproved, together with some variants of results of Phillips–Sarnak, Adachi–Sunada, and Katsuda–Sunada in Petridis and Risager's papers [42, 43]. The methods of [42] involve perturbation theory, and so are similar to those of the current paper. Results of [42] are closely related to those of [24]—in that paper we show (using the ergodicity of the $\mathrm{SL}:(n, \mathbf{Z})$ action on $\mathbf{R}^n$ and the Central Limit theorem for free groups in the current paper) that some probabilistic phenomena in the free group F_n can be studied by descending to the Abelian quotient.

The Central Limit theorem has been extended in other ways as well: Calegari and Fujiwara proved a Central Limit theorem for the values of *bicombable* functions on word-hyperbolic groups in [6], using Markov chain methods, while Horsham and Sharp extended the results to *quasi-morphisms* of free groups by using the usual symbolic dynamics and thermodynamic formalism in [17].

Lest one think that every function of interest on free (or word-hyperbolic) group satisfies a Central Limit theorem, we should note the results of Guivarc'h–LeJan ([14, 15]) and Vardi ([62]), which show that the the distribution of lengths of geodesics on the modular surface satisfies a *stable law* of Cauchy type.

Then what happened? Walks on graphs. In Sections 6 and 6.1, we look at homology modulo a prime p and derive the expected equidistribution results (and also the analogue of *Chebyshev bias*, see [54], which in this case is completely explicit). More importantly, however, a study of the argument showed that instead of a finite abelian group we can take any *compact* (in particular, any *finite*) group—the harmonic analysis goes through, although with some more work. The arguments in this paper are a little sketchy, but are presented in full detail in my papers [50, 51]. These papers, together with [52] are devoted to proving that certain phenomena in algebraic groups, as well as "geometric" groups, like the mapping class group and the outer automorphism group of the free group (and a large class of subgroups) are generic

(which means that in large subsets of the groups in question, the vast majority of elements have a certain property—see [25] for other examples). The way the results of the current paper are used is essentially through a "Chinese remaindering" argument—if a certain property does *not* hold for some fraction of the elements in the projection of an algebraic group (scheme) over $\mathbf{Z}/p\mathbf{Z}$, then it does not hold generically in the group over $\mathbf{Z}$. Using property T and a more refined analysis (as in [51]) give estimates of convergence speed. The appearance of the paper [50] is responsible the subsequent appearance of Kowalski's book [32], where these rather simple ideas are couched in a rather formidable apparatus.

Then what happened? Topological entropy. In the mid-to-late 1990s, the spectacular results of Besson, Courtois, and Gallot on "volume rigidity" of locally symmetric spaces (see [4]) were generating a lot of excitement. The result was that among all the metrics of a given volume on a hyperbolic manifold, the metric of constant sectional curvature minimizes volume entropy—this answered a conjecture of Gromov stated in [13], and previously known only in dimension two (thanks to Katok's result [27]). Any time a function has a single minimum, there is a suspicion that some sort of convexity is afoot, and entropy in the simplest setting (see, for example, [56]) is a convex function of the probabilities, and this pushed the author to analyze topological entropy for walks on graphs as a function of weights on the vertices in Section 11. The methods are again those of perturbation theory. Later, the result was extended to *edge* weightings by Lim in [34]. Lim does *not* prove convexity, but does write down the unique metric of minimal entropy. A related minimality result is proved by Kapovich and Nagnibeda in their paper [22] for *regular* graphs (their work has its roots in the study of Outer Space). In a different direction, the convexity of entropy was used by Kapovich and myself in [23] to show that there is no analogue to McShane's identity in Outer Space.

Introduction

In this paper, we begin by studying certain growth functions of the free group F_r, related to well-studied questions on the growth functions of geodesics on manifolds. The free group is a relatively simple combinatorial object, and this allows us to get fairly complete answers to our questions. Our techniques, which are quite elementary, allow us to get precise results on the distribution of elements in F_r as a function of their Abelianization and in terms of their Abelianization mod p. Our techniques turn out to be easily extensible to the study of paths in graphs with coefficients in compact groups.

Here is an outline of the paper: In Section 1, we set up an equivalence between counting cyclically reduced words on the free group F_r and counting circuits on an associated graph $\mathcal{G}_r$, which, in turn, involves understanding the spectrum of the adjacency matrix of $\mathcal{G}_r$ (of course the answer is easily

obtained, and is well known; for convenience we state it as Theorem 1.1). We use this framework to obtain a generating function for the number of elements of a fixed cyclically reduced length with prescribed Abelianization (or homology class). This turns out to be essentially a Chebyshev polynomial of the first kind; see Definition 2.2 of the function R_r and Theorem 2.3 (a very brief introduction to Chebyshev polynomials is given in Section 3). The fact that the function $R_r(c;\mathbf{x})$ (at least for some special values of the parameter c) is a combinatorial generating function implies a previously unnoticed positivity result on Chebyshev polynomials; this result is generalized in Section 4 in Theorems 4.1 and 4.2. Theorem 2.3 is used in Section 5 to derive a limiting distribution (as n tends to infinity) of cyclically reduced words length n among the possible homology classes. From the analytic standpoint, this is also a qualitative result about Chebyshev polynomials, complementing the positivity Theorems 4.1 and 4.2. In Section 6, we show that if we study homology $\bmod p$, then the cyclically reduced words in F_r are asymptotically equidistributed among the p^r classes in $H_1(F_r, \mathbf{Z}/p\mathbf{Z})$. We also succeed in estimating the extent to which the cyclically reduced words in F_r are *not* equidistributed mod p (Section 6.1).

While the results in Sections 5 and 6 seem to depend on the explicit generating function that we have obtained, in Section 7 we show that our techniques are more general, and use them to study the equidistribution properties of long walks on regular graphs—we obtain a complete answer (Theorem 7.1)—and, without any change, closed orbits of irreducible primitive Markov processes (with a finite number of states). The arguments use elementary perturbation theory and the necessary technical results are contained in Section 10.

In Section 8, we extend our methods to study the functions defined on the *edges* of a graph, and as an application we derive the statistical properties of long walks *without backtracking* on the edges of an undirected graph.

We apply our methods to derive equidistribution results for long walks with coefficients in compact groups in Sections 7.1 and 9. Our results are completely explicit, in that knowing the irreducible representations of the group in question allows us to obtain complete asymptotics for the convergence to uniformity. Our results also apply, via the construction of a directed edge graph to the statistics of "geodesic," that is, backtrackless paths (Section 8). This, in turn, implies a result on the statistical properties of "primitive" orbits of Markov processes as above.

In Section 12 we point out real and philosophical applications of the above mentioned result to group theory (where this all started) and geometry.

Finally, in Sections 13–14.1, we derive a relationship between the number of cyclically reduced words and the number of conjugacy classes of bounded length. While the generating function of the first is a rational function, the generating function of the second is the integral of a Lambert series with an infinite number of poles. These results are then extended to a slightly more

general case than that of free groups. We then (in Section 15) compute a zeta function for primitive conjugacy classes, and show that this *is* a rational function.

1. A model and a generating function

Let G be the free group $F_r = \langle a_1, \ldots, a_r \rangle$, and let $g \in G$ be an element. The defining property of G is that g is *uniquely* represented by a *reduced* word in $a_1, \ldots, a_r$, that is, a word where a_i is never adjacent to a_i^{-1} (Notation: in the sequel we shall write W for w^{-1}). We observe that such words over the alphabet $a_1, A_1, \ldots, a_n, A_n$ are, in turn, be generated by walks on the graph $\mathcal{G}_r$, constructed as follows: $\mathcal{G}_n$ has $2r$ vertices, labelled with the symbols $a_1, \ldots, a_r, A_r, \ldots, A_1$—this peculiar order will simplify notation later. The vertex corresponding to a_i is connected by an edge to every vertex *except* A_i. In particular, there is a loop joining a_i to itself (so that $\mathcal{G}_r$ is not a *simple* graph). A walk $v_1 v_2 \cdots v_k$ gives the word $v_2 \cdots v_k$, so the correspondence between walks and words is a $2r-1$-to-1 mapping. Note, however, that if we restrict our attention to closed walks (circuits with basepoint) on $\mathcal{G}_r$, then those are in bijective correspondence with *cyclically reduced* words in G. In the sequel, we will be interested exclusively with cyclically reduced words.

1.1. Counting cyclically reduced words. To count cyclically reduced words, then, we need to count circuits in $\mathcal{G}_r$. This is a well-understood problem: If $\mathcal{A}_r$ is the adjacency matrix of $\mathcal{G}_r$, then the number of circuits of length k is equal to the trace of $\mathcal{A}_r^k$. To compute this trace, we must compute the spectrum of $\mathcal{A}_r$, and to do this, it is better to write $\mathcal{A}_r = J_{2r} - P_r$, where J_N is an $N \times N$ matrix all of whose elements are 1 and P_r is the $2r \times 2r$ matrix such that

$$(P_r)_{ij} = \begin{cases} 1, & \text{if } i + j = 2r; \\ 0, & \text{otherwise.} \end{cases}$$

In order to compute the spectrum of $\mathcal{A}_r$, we note first that the matrix J_{2r} has rank 1. The kernel of J_{2r} is

$$\ker J_{2r} = \left\{ (v_1, \ldots, v_{2r}) \Big| \sum_{i=1}^{2r} v_i = 0 \right\},$$

while the vector $\mathbf{1} = (1, \ldots, 1)$ is the eigenvector of eigenvalue $2r$.

The spectrum of P_r is not much more difficult to compute: The vector $\mathbf{1}$ is the eigenvector of P_r as well as of J_{2r}, this time with eigenvalue 1. To compute the rest of the spectral decomposition, let $\mathbf{x}$ be an eigenvector of P_r orthogonal to $\mathbf{1}$, and let λ be the corresponding eigenvalue. Then we have the

following set of equations:

$$\sum_{j=1}^{2r} x_j = 0,$$
$$x_j = \lambda x_{2r-j+1}, \quad j = 1, \ldots, 2r.$$

Since at least one of the x_j is not equal to zero, we see that $\lambda^2 = 1$, so $\lambda = \pm 1$. The orthogonality condition equation (1.1) can be rewritten as $\sum_{j=1}^{r}(1 + \lambda)x_j = 0$. Suppose $\lambda = -1$. Then, equation (1.1) holds a forteriori, and so the eigenspace of of -1 is r-dimensional. On the other hand, if $\lambda = 1$, then we have the additional constraint that $\sum_{j=1}^{r} x_j = 0$, so the eigenspace of 1 is $(n-1)$-dimensional. Putting this all together, we see that the spectrum of the adjacency matrix $\mathcal{A}_r$ is $(2r-1, \underbrace{1, \ldots, 1}_{r}, \underbrace{-1, \ldots, -1}_{r-1})$. We see therefore the following theorem.

THEOREM 1.1. *The number of cyclically reduced words of length m in F_r is equal to $(2r-1)^m + 1 + (r-1)[1 + (-1)^m]$.*

2. Counting cyclically reduced words in homology classes

Recall that the Abelianization of F_r is $\mathbf{Z}^r$, generated by the classes of $[a_1], \ldots, [a_r]$ of $a_1, \ldots, a_r$ respectively. To compute the homology class of a word w in F_r, we simply count the total exponents $e_1(w), \ldots, e_r(w)$ of the generators used to write w. Then, $[w] = e_1(w)[a_1] + \cdots + e_r(w)[a_r]$. In this section, we will compute the following generating function:

$$\mathcal{H}_r^{(k)}(x_1, \ldots, x_r) = \sum_{w \in W_k} \prod_{i=1}^{r} x_i^{e_i(w)},$$

where the sum is taken over the set W_k of all cyclically reduced words w in $a_1, \ldots, a_r, A_1, \ldots, A_r$ of length k.

To compute $\mathcal{H}_r^{(k)}$, we return to circuits in $\mathcal{G}_r$. Given a circuit $c = v_1, \ldots, v_k, v_{k+1} = v_1$, the contribution of c to $\mathcal{H}_r^{(k)}$ is the monomial m_c given by the following iterative procedure: we start with 1, every time we see the vertex a_i, we multiply m_c by x_i, and every time we see A_i, we multiply m_c by $1/x_i$. From this, it follows the following theorem.

THEOREM 2.1. *The Laurent polynomial* $\mathcal{H}_r^{(k)}$ *is given by* $\operatorname{tr} B_r^k$, *where* $B_r = D_r \mathcal{A}_r$, *where, in turn,*

$$D_r = \begin{pmatrix} x_1 & & & & & & \\ & \ddots & & & & & \\ & & x_n & & & & \\ & & & 1/x_n & & & \\ & & & & \ddots & & \\ & & & & & 1/x_1 & \end{pmatrix}.$$

Computing the trace of B_r^k seems daunting at first, but one can use the approach we have used to prove Theorem 1.1.

First, note that

$$B_r = D_r \mathcal{A}_r = D_r J_{2r} - D_r P_r.$$

Evidently, the rank of $D_r J_{2r}$ is still equal to 1, and

$$\ker D_r J_{2r} = \left\{ \mathbf{v} = (v_1, \ldots, v_{2r}) \Big| \sum_{j=1}^{2r} v_j = 0 \right\}.$$

Note further that an eigenvector $\mathbf{v}$ of $D_r P_r$, such that $\mathbf{v} \in \ker D_r J_{2r}$, with associated eigenvalue λ, is also an eigenvector of B_r, with associated eigenvalue $-\lambda$. To find such an eigenvector, we must solve the system of equations:

$$\begin{aligned} &\sum_{j=1}^{2r} v_j = 0, \\ &\lambda v_j = v_{2r-j+1}/x_j, \quad j \le r, \\ &\lambda v_j = v_{2r-j+1} x_j, \quad j > r. \end{aligned}$$

We find, as before, that $\lambda = \pm 1$. The first equation reduces (almost as before) to

$$\sum_{j=1}^{r} v_j (1 + \lambda x_j) = 0,$$

so that the eigenspaces of both 1 and -1 are $(r-1)$-dimensional. What are the two remaining eigenvalues μ_1 and μ_2 of B_r? Note that since $\det D_r = 1$, we know that $\det B_r = \det \mathcal{A}_r$. Note now that $\det B_r = \mu_1 \mu_2 (-1)^{r-1}$, while $\det \mathcal{A}_r = (2r-1)(-1)^{r-1}$. So

$$\mu_1 \mu_2 = 2r - 1. \tag{2.1}$$

On the other hand,

$$\mu_1 + \mu_2 = \operatorname{tr} B_r = \sum_{j=1}^{r} \left(x_j + \frac{1}{x_j} \right). \tag{2.2}$$

Denoting $y_r = \frac{1}{2}\sum_{j=1}^{n}(x_j + 1/x_j)$, we see that μ_1, μ_2 are the two roots of the equation $z^2 - 2y_r z + (2r-1) = 0$, so that:

$$\mu_1 = y_r - \sqrt{y_r^2 - (2r-1)},$$
$$\mu_2 = y_r + \sqrt{y_r^2 - (2r-1)}.$$

The trace of B_r^k is then equal to $\mu_1^k + \mu_2^k + (r-1)[1 + (-1)^k]$. This can be expressed in terms of well-known special functions, if we make the substitution $y_r = \sqrt{2r-1}\,y_r'$. Then,

$$\mu_1^k = (2r-1)^{k/2}\big(y_r' - \sqrt{{y_r'}^2 - 1}\big)^k,$$
$$\mu_2^k = (2r-1)^{k/2}\big(y_r' + \sqrt{{y_r'}^2 - 1}\big)^k,$$

and so

$$\begin{aligned}\mu_1^k + \mu_2^k &= (2r-1)^{k/2}\big\{\big(y_r' - \sqrt{{y_r'}^2 - 1}\big)^k + \big(y_r' + \sqrt{{y_r'}^2 - 1}\big)^k\big\}\\ &= 2\big(\sqrt{2r-1}\big)^k T_k(y_r'),\end{aligned}$$

where $T_k(x)$ is the kth Chebyshev polynomial of the first kind. To simplify notation in the sequel, we define the following.

Definition 2.2.

$$R_n(c; x_1, \dots, x_k) = T_n\left(\frac{c}{2k}\sum_{i=1}^{k}\left(x_i + \frac{1}{x_i}\right)\right),$$
$$S_n(c; x_1, \dots, x_k) = U_n\left(\frac{c}{2k}\sum_{i=1}^{k}\left(x_i + \frac{1}{x_i}\right)\right).$$

And to summarize the following theorem.

Theorem 2.3. *The number of cyclically reduced words of length k in F_r homologous to $e_1[a_1] + \cdots + e_r[a_r]$ is equal to the coefficient of $x_1^{e_1}\cdots x_r^{e_r}$ in*

$$2\big(\sqrt{2r-1}\big)^k R_k\left(\frac{r}{\sqrt{2r-1}}; x_1, \dots, x_r\right) + (r-1)[1 + (-1)^k]. \tag{2.3}$$

Remark 2.4. *The rescaled Chebyshev polynomial $T_k(ax)/a^k$ is called the kth Dickson polynomial $T_k(x, a)$ (see* [55]).

3. Some facts about Chebyshev polynomials

The literature on Chebyshev polynomials is enormous; [53] is a good to start. Here, we shall supply the barest essentials in an effort to keep this paper self-contained.

There are a number of ways to define Chebyshev polynomials (almost as many as there are of spelling their inventor's name). A standard definition of the *Chebyshev polynomial of the first kind* $T_n(x)$ is:

$$T_n(x) = \cos n \arccos x. \tag{3.1}$$

In particular, $T_0(x) = 1$, $T_1(x) = x$. Using the identity

$$\cos(x+y) + \cos(x-y) = 2\cos x \cos y \tag{3.2}$$

we immediately find the three-term recurrence for Chebyshev polynomials:

$$T_{n+1}(x) = 2xT_n(x) - T_{n-1}(x). \tag{3.3}$$

The definition of equation (3.1) can be used to give a "closed form" used in Section 2:

$$T_n(x) = \frac{1}{2}\left[\left(x - \sqrt{x^2-1}\right)^n + \left(x + \sqrt{x^2-1}\right)^n\right]. \tag{3.4}$$

Indeed, let $x = \cos\theta$, then $(x-\sqrt{x^2-1})^n = \exp(-in\theta)$, while $(x+\sqrt{x^2-1})^n = \exp(in\theta)$, so $\frac{1}{2}(x - \sqrt{x^2-1})^n + (x + \sqrt{x^2-1})^n = \Re\exp(in\theta) = \cos n\theta$.

Though we will not have too many occasions to use them, we also define Chebyshev polynomials of the second kind $U_n(x)$, which can again be defined in a number of ways, one of which is:

$$U_n(x) = \frac{1}{n+1}T'_{n+1}(x). \tag{3.5}$$

A simple manipulation shows that if we set $x = \cos\theta$, as before, then

$$U_n(x) = \frac{\sin(n+1)\theta}{\sin\theta}. \tag{3.6}$$

In some ways, Schur's notation $\mathcal{U}_n = U_{n-1}$ is preferable. In any case, we have $U_0(x) = 1$, $U_1(x) = 2x$, and otherwise the U_n satisfy the same recurrence as the T_n, to wit,

$$U_{n+1}(x) = 2xU_n(x) - U_{n-1}(x). \tag{3.7}$$

From the recurrences, it is clear that for $f = T, U$, $f_n(-x) = (-1)^n f(x)$, or, in other words, every second coefficient of $T_n(x)$ and $U_n(x)$ vanishes. The remaining coefficients alternate in sign; here is the explicit formula for the coefficient $c^{(n)}_{n-2m}$ of x^{n-2m} of $T_n(x)$:

$$c^{(n)}_{n-2m} = (-1)^m \frac{n}{n-m}\binom{n-m}{m}2^{n-2m-1}, \quad m = 0, 1, \ldots, \left[\frac{n}{2}\right]. \tag{3.8}$$

This can be proved easily using equation (3.3).

4. Analysis of the functions R_n and S_n

In view of the alternation of the coefficients, the appearance of the Chebyshev polynomials as generating functions in Section 2 seems a bit surprising, since combinatorial generating functions have nonnegative coefficients. Below we state and prove a generalization. Remarkably, Theorems 4.1 and 4.2 do not seem to have been previously noted.

THEOREM 4.1. *Let $c>1$. Then all the coefficients of $R_n(c;x)$ are nonnegative. Indeed the coefficients of $x^n, x^{n-2}, \dots, x^{-n+2}, x^{-n}$ are positive, while the other coefficients are zero. The same is true of S_n in place of R_n.*

Proof. Let a_n^k be the coefficient of x^k in $U_n((c/2)(x+1/x))$. The recurrence gives the following recurrence for the a_n^k:

$$a_{n+1}^k = c(a_n^{k-1} + a_n^{k+1}) - a_{n-1}^k. \tag{4.1}$$

Now we shall show that the following always holds:

(a) $a_n^k \geq 0$ (inequality being strict if and only if $n-k$ is even).

(b) $a_n^k \geq \max(a_{n-1}^{k-1}, a_{n-1}^{k+1})$, the inequality strict, again, if and only if $n-k$ is even.

(c) $a_n^k \geq a_{n-2}^k$ (strictness as above).

The proof proceeds routinely by induction; first the induction step (we assume throughout that $n-k$ is even; all the quantities involved are obviously 0 otherwise):

By induction $a_{n-1}^k < \min(a_n^{k-1}, a_n^{k+1})$, so by the recurrence (4.1) it follows that $a_{n+1}^k > \max(a_n^{k-1}, a_n^{k+1})$. (a) and (c) follow immediately.

For the base case, we note that $a_0^0 = 1$, while $a_1^1 = a_1^{-1} = c > 1$, and so the result for U_n follows. Notice that the above proof does not work for T_n, since the base case fails. Indeed, if b_n^k is the coefficient of x^k in $T_n((c/2)(x+1/x))$, then $b_0^0 = 1$, while $b_1^1 = c/2$, not necessarily bigger than one. However, we can use the result for U_n, together with the observation (which follows easily from the addition formula for sin) that

$$T_n(x) = \frac{U_n(x) - U_{n-2}(x)}{2}. \tag{4.2}$$

Equation (4.2) implies that $b_n^k = a_n^k - a_n^{k-2} > 0$, by (c) above. □

The proof above goes through almost verbatim to show the following theorem.

THEOREM 4.2. *Let $c>1$. Then all the coefficients of R_n are nonnegative. The same is true of S_n in place of R_n.*

To complete the picture, we note the following theorem.

THEOREM 4.3.

$$R_n(1;x) = \frac{1}{2}\left(x^n + \frac{1}{x^n}\right).$$

Proof. Let $x = \exp i\theta$. Then $1/2(x+1/x) = \cos\theta$, and $R_n(1;x) = T_n(1/2 \times (x+1/x)) = \cos n\theta = 1/2(x^n + 1/x^n)$. □

REMARK 4.4. *For $c < -1$ it is true that all the coefficients of $R_n(c;\cdot)$ and $S_n(c;\cdot)$ have the same sign, but the sign is $(-1)^n$. For $|c| < 1$, the result is completely false. For c imaginary, the result is true. I am not sure what happens for general complex c.*

By the formula (3.8), we can write

$$T_n\left(\frac{c}{2}\left(x+\frac{1}{x}\right)\right) = \frac{1}{2}\sum_{m=0}^{\left[\frac{n}{2}\right]} (-1)^m \frac{n}{n-m}\binom{n-m}{m} c^{n-2m}\left(x+\frac{1}{x}\right)^{n-2m}. \tag{4.3}$$

Noting that

$$\left(x+\frac{1}{x}\right)^k = \sum_{i=0}^{k}\binom{k}{i}x^{k-2i} \tag{4.4}$$

we obtain the expansion

$$R_n(c;x) = c^n \sum_{k=-n}^{n} x^k \sum_{m=0}^{\left[\frac{n}{2}\right]} \left(-\frac{1}{c^2}\right)^m \frac{n}{n-m}\binom{n-m}{m}\binom{n-2m}{(n-2m-k)/2}, \tag{4.5}$$

where it is understood that $\binom{a}{b}$ is 0 if $b < 0$, or $b > a$, or $b \notin \mathbf{Z}$. We shall denote the coefficient of x^k by $t(n,k,c)$.

5. Limiting distribution of coefficients

While the formula (4.5) is completely explicit, and a similar (though somewhat more cumbersome) expression could be obtained for $R_n(c;x_1,\ldots,x_k)$, for many purposes it is more useful to have a limiting distribution formula as given by Theorem 5.1 below. To set up the framework, we note that since all the coefficients of $R_n(c;x_1,\ldots,x_k)$ are nonnegative (according to Theorem 4.2), they can be thought of defining a probability distribution on the integer lattice $\mathbf{Z}^k$, defined by $p(l_1,\ldots,l_k) = [x_1^{l_1}x_2^{l_2}\cdots x_k^{l_k}]R_n(c;x_1,\ldots,x_k)/R_n(c;1,\ldots,1)$ (where the square brackets mean that we are extracting the coefficients of the bracketed monomial). Call the resulting probability distribution $\mathcal{P}_n(c;\mathbf{z})$, where $\mathbf{z}$ now denotes a k-dimensional vector.

THEOREM 5.1. *When $c>1$, the probability distributions $\mathcal{P}_n(c;\mathbf{z}/\sqrt{n})$ converge to a normal distribution on $\mathbf{R}^k$, whose mean is $\mathbf{0}$, and whose covariance matrix C is diagonal, with entries*

$$\sigma^2 = \frac{c}{k}\left[1+\left(\frac{c+1}{c-1}\right)^{1/2}\right].$$

To prove Theorem 5.1, we will use the method of characteristic functions (Fourier transforms), and more specifically at first the *Continuity Theorem* ([10, Chapter XV.3, Theorem 2]).

THEOREM 5.2. *In order that a sequence $\{F_n\}$ of probability distributions converges properly to a probability distribution F, it is necessary and sufficient that the sequence $\{\phi_n\}$ of their characteristic functions converges pointwise to a limit ϕ, and that ϕ is continuous in some neighborhood of the origin.*

In this case, ϕ is the characteristic function of F. (Hence, ϕ is continuous everywhere and the convergence $\phi_n \to \phi$ is uniform on compact sets.)

The characteristic function ϕ_n of $\mathcal{P}_n(c;\mathbf{z})$ is simply

$$R_n(c;\exp(i\theta_1),\dots,\exp(i\theta_k))/R_n(c;1,\dots,1),$$

since the characteristic function is just the generating function evaluated on the unit circle.

By definition of R_n,

$$R_n(c;\exp(i\theta_1),\dots,\exp(i\theta_k)) = T_n\left(\frac{c}{k}\sum_{j=1}^{k}\cos\theta_j\right),$$

$$R_n(c;1,\dots,1) = T_n\left(\frac{c}{k}\sum_{j=1}^{k}\cos 0\right) = T_n(c).$$

We now use the form of equation (3.4):

$$T_n(x) = \frac{1}{2}\left(\left(x-\sqrt{x^2-1}\right)^n + \left(x+\sqrt{x^2-1}\right)^n\right),$$

setting

$$u = \sum_{j=1}^{k}\cos\frac{\theta_j}{\sqrt{n}}, \quad \boldsymbol{\theta} = (\theta_1,\dots,\theta_k),$$

we get

$$\phi_n\left(\boldsymbol{\theta}/\sqrt{n}\right) = \frac{1}{T_n(c)}\left\{\frac{1}{2}\left(\frac{c}{k}u+\sqrt{\frac{c^2}{k^2}u^2-1}\right)^n + \frac{1}{2}\left(\frac{c}{k}u-\sqrt{\frac{c^2}{k^2}u^2-1}\right)^n\right\}. \tag{5.1}$$

Notice, however, that for $c > 1$, the ratio of the second term in braces to the first is exponentially small as $n \to \infty$, since the first term grows like $(c+\sqrt{c^2-1})^n$, while the second as $(c-\sqrt{c^2-1})^n$ (since $\cos\frac{\theta_j}{\sqrt{n}} \to 1$). Since, for the same reason, $2T_n(c) = (c+\sqrt{c^2-1})^n[1+o(1)]$, we can write:

$$\phi_n\left(\frac{\boldsymbol{\theta}}{\sqrt{n}}\right) = \left[\frac{(c/k)u + \sqrt{(c^2/k^2)u^2-1}}{c+\sqrt{c^2-1}}\right]^n + o(1).$$

Substituting the Taylor expansions for the cosine terms (hidden in u for typesetting reasons), we get:

$$u = k + \frac{1}{2n}\langle\boldsymbol{\theta},\boldsymbol{\theta}\rangle + o(1/n), \tag{5.2}$$

so

$$\frac{c}{k}u = c + \frac{c}{2kn}\langle\boldsymbol{\theta},\boldsymbol{\theta}\rangle + o(1/n). \tag{5.3}$$

A similar computation gives

$$\frac{c^2}{k^2}u^2 = c^2 + \frac{c^2}{kn}\langle\boldsymbol{\theta},\boldsymbol{\theta}\rangle + o(1/n). \tag{5.4}$$

Substituting the last expansion into the square root, we see that

$$\begin{aligned}\sqrt{\frac{c^2}{k^2}u^2-1} &= \sqrt{c^2-1}\sqrt{1+\frac{1}{n}\left[\frac{c^2}{(c^2-1)k}\langle\boldsymbol{\theta},\boldsymbol{\theta}\rangle + o\left(\frac{1}{n}\right)\right]}\\ &= \sqrt{c^2-1}\left[1+\frac{1}{2n}\frac{c^2}{(c^2-1)k}\langle\boldsymbol{\theta},\boldsymbol{\theta}\rangle\right] + o\left(\frac{1}{n}\right).\end{aligned}$$

Adding equation (5.3) and collecting terms, get

$$\begin{aligned}&\frac{(c/k)u + \sqrt{(c^2/k^2)u^2-1}}{c+\sqrt{c^2-1}}\\ &\quad= 1 + \frac{1}{2n}\left(1+\frac{1}{c+\sqrt{c^2-1}}\right)\left(\frac{c}{k}+\frac{c^2}{(c^2-1)^{1/2}k}\right)\langle\boldsymbol{\theta},\boldsymbol{\theta}\rangle + o\left(\frac{1}{n}\right).\end{aligned} \tag{5.5}$$

Performing some further simplifications, we see that

$$\phi_n\left(\frac{\boldsymbol{\theta}}{\sqrt{n}}\right) = \exp\left(-\frac{1}{2}\boldsymbol{\theta}^t C\boldsymbol{\theta}\right) + o(1),$$

where C is the covariance matrix described in the statement of Theorem 5.1, and Theorem 5.1 follows immediately.

REMARK 5.3. *The speed of convergence in Theorem 5.1 can be estimated using standard technology* (*see* [10, *Chapter* XVI], [58, *Chapter* III.11]), *but the speed of convergence in practice* (*as checked by numerical experiments*) *seems to be much better than the general estimates. Indeed the* L^1 *difference between* $\mathcal{P}_n$ *and the normal distribution appears to decrease almost exactly linearly in* n.

6. Distribution mod p

The explicit generating functions derived above can be used to study the distribution of cyclically reduced words in F_r with respect to their mod p-homology class (this is the analogue, in this setting, of the work of [44]).

THEOREM 6.1. *Let h_1 and h_2 be two elements of $H_1(F_r, \mathbf{Z}/p\mathbf{Z}) = \mathbf{Z}/p\mathbf{Z}^r$, and let W_{r,n,h_1} and W_{r,n,h_2} be the numbers of cyclically reduced words in F_r homologous to h_1 and h_2, respectively. Then*

$$\lim_{n\to\infty} \frac{W_{r,n,h_2}}{W_{r,n,h_1}} = 1. \tag{6.1}$$

Proof. By elementary algebra (in one dimension, formula (6.3), the statement of theorem is equivalent to the statement that

$$\lim_{n\to\infty} \frac{\phi_n(\boldsymbol{\theta})}{\phi_n(\mathbf{0})} = 0 \tag{6.2}$$

for $\boldsymbol{\theta} = (2n_1\pi/p, \dots, 2n_r\pi/p)$, with not all n_j equal to $0 \mod p$, where ϕ_n is the characteristic function defined in the previous section.

The estimate of equation (6.2), however, follows immediately from the explicit formula (5.1): indeed, in the current context,

$$u(\boldsymbol{\theta}) = \sum_{j=1}^{k} \cos(2n_j\pi/p),$$

which is strictly smaller than $u(\mathbf{0})$, so the ratio of $\phi_n(\boldsymbol{\theta})$ to $\phi_n(\mathbf{0})$ goes to zero exponentially fast in n. □

REMARK 6.2. *Another way to see the equivalence of statements (6.1) and (6.2) is though the well-known fact that the Fourier transform is an isometry (of the corresponding L^2 spaces). For a probability density to be close to uniform, its Fourier transform has to be close to that of the uniform distribution, which is a delta function centered at the origin, which is precisely the statement we need.*

6.1. Deviation from uniformity. Although the distribution of homology mod p approaches uniformity, it turns out that there is a persistent *bias* in favor of certain homology classes. This is very much akin to the Chebyshev bias, analyzed in [54]. To simplify the discussion, we project one more time: for each cyclically reduced word in F_r homologous to $a_1^{k_1}a_2^{k_2}\cdots a_r^{k_r}$ we consider $k_1 + \cdots + k_r \mod p$. In this case, we have a univariate distribution, whose generating function is given by $\psi_n(x) = R_n(c; x, \dots, x)$, with $c = \frac{r}{\sqrt{2r-1}}$ (as per formula (2.3); we leave in the general c, to underline that our results apply to general question on distribution of coefficients of the Laurent polynomials R_n).

The number of elements congruent to $q \mod p$ is given by

$$\mathcal{N}_{n,q} = \frac{1}{p}\sum_{j=0}^{p-1} \chi^{-qj}\psi_n(\chi^j), \tag{6.3}$$

where $\chi = \exp(2\pi i/p)$ is a primitive pth root of unity. Let us recall that

$$\psi_n(e^{ix}) = \frac{1}{T_n(c)}\left\{\frac{1}{2}\left(c\cos x + \sqrt{c^2\cos^2 x - 1}\right)^n + \frac{1}{2}\left(c\cos x - \sqrt{c^2\cos^2 x - 1}\right)^n\right\}. \tag{6.4}$$

Note the following properties of the function ψ_n:

$$\psi_n(1/x) = \psi_n(x), \tag{6.5a}$$

$$\text{if } c\cos x < 1, \text{ then } |\psi_n(\exp(ix))|T_n(c) \le 1. \tag{6.5b}$$

$$\psi_n\{\exp(i(\pi - x))\} = (-1)^n\psi_n\{\exp(ix)\}, \tag{6.5c}$$

$$\text{if } c\cos x \ge 1, \text{ then } \psi_n(\exp(ix)) > 0. \tag{6.5d}$$

$$\text{If } x \in [0, \arccos 1/c], n \gg 1, \text{ then } \frac{|\psi_n(\exp(ix))|T_n(c)}{[c + \sqrt{c^2\cos^2 x - 1}]^n} = 1 + o(1), \tag{6.5e}$$

$$\psi_n(\exp(ix_1)) = o(\psi_n(\exp(ix_2))) \quad \text{for } 0 \le x_2 < \arccos 1/c, x_2 < x_1 < \pi - x_2. \tag{6.5f}$$

Using property (6.5a), we can write

$$\mathcal{N}_{n,q} = \frac{1}{p}\left[\psi_n(1) + 2\sum_{j=1}^{\frac{p-1}{2}} \cos\frac{2\pi qj}{p}\psi_n(\chi^j)\right]. \tag{6.6}$$

Since $\cos\frac{2\pi m}{p} < 1$ is monotonically decreasing as a function of m for $0 \le m \le \frac{p-1}{2}$, we see the following theorem.

THEOREM 6.3. *For sufficiently large even n, $\mathcal{N}_{n,q} < \mathcal{N}_{n,0}$.*

Proof. This is an immediate consequence of the monotonicity of cos, equation (6.6) and properties (6.5a), (6.5c), (6.5d) and (6.5f) above. □

For $q \neq 0 \mod p$, the term largest in absolute value in the sum (aside the $\psi_n(1)$ term) on the right-hand side of equation (6.6) is the $\psi_{(\chi^{\frac{p-1}{2}})}$ term, so if we assume that n is even, then the next largest (after $\mathcal{N}_{n,0}$) term will be $\mathcal{N}_{n,p-2}$ (since $(p-2)[(p-1)/2] = 1 \mod p$), then $\mathcal{N}_{n,p-4}$, and so on. For n odd, the ordering is reversed.

7. An extension and limiting distributions for graphs

An inspection of the proof of Theorem 5.1 reveals that in order to show that for a sequence of probability distributions $\{P_n(x)\}$ on $\mathbf{Z}$, the distributions $\{P_n(x/\sqrt{n})\}$ converged to a limiting normal distribution with mean 0, we used the following conditions (we will state them in a univariate setting for simplicity; the multivariate case is the same).

CONDITION 1. The characteristic function of $\{P_n\}$ has the form

$$\chi(P_n) = f^n(\theta) + o(1),$$

where $f_j(\theta)$ is twice continuously differentiable at 0, so that $f_j(\theta) = a_j + b_j\theta + c_j\theta^2 + o(\theta^2)$.

CONDITION 2.

$$a_1 = 1, \qquad b_2 = 0, \qquad c_2 < 0.$$

Suppose now we generalize the setting of Section 1 as follows:

Let $\mathcal{G}$ be a connected r-regular nonbipartite graph, directed or not (possibly with self-loops and multiple edges), on k vertices. Let v_1 and v_2 be two vertices of $\mathcal{G}$. Consider now the set W_N of all closed walks (circuits) of length N on $\mathcal{G}$. Let $\mathbf{f} : V(\mathcal{G}) \to R$ be a function assigning a weight to each vertex of $\mathcal{G}$, and define a random variable $X_{\mathbf{f}}$ to be $\sum_{l=1}^{N} \mathbf{f}(v_l)$ for $w = v_1, \ldots, v_N \in W_N$. What can we say about the distribution of $X_{\mathbf{f}}$? It turns out that asymptotically we can say a lot. First, however, define

$$\mu(\mathbf{f}) = \frac{1}{k} \sum_{j=1}^{k} \mathbf{f}(v_j),$$

and $\mathbf{f}_0 = \mathbf{f} - \mu(\mathbf{f})\mathbf{1}$. Define further the *Laplacian* $\Delta(\mathcal{G})$ of $\mathcal{G}$ to be $\Delta(\mathcal{G}) = r\mathbf{I} - A(\mathcal{G})$, and define $\Delta_0(\mathcal{G})$ to be $\Delta(\mathcal{G})$ viewed as an operator on the orthogonal complement to $\mathbf{1}$ (that is, vectors with 0 sum). Let $P_N(x)$ be the distribution of $X_{\mathbf{f}}$ on W_N.

THEOREM 7.1. *The distributions* $P_N((x - N\mu(\mathbf{f}))/\sqrt{N})$ *converge to a balanced* (*that is, mean* 0) *normal distribution with variance*

$$\sigma^2(\mathbf{f}) = \frac{1}{k}[-\|\mathbf{f}_0\|^2 + 2r\mathbf{f}_0^t\Delta_0^{-1}(\mathcal{G})\mathbf{f}_0] = \frac{1}{k}\big[\mathbf{f}_0^t\big(-\mathbf{I}_0 + 2r\Delta_0^{-1}(\mathcal{G})\big)\mathbf{f}_0\big]. \tag{7.1}$$

Proof. Exactly as in Section 1 we construct a generating function g_N for $X_{\mathbf{f}}$ on W_N. To do this, let A be the adjacency matrix of $\mathcal{G}$, and let

$$D_k(x) = \begin{pmatrix} x^{\mathbf{f}(v_1)} & & & & \\ & x^{\mathbf{f}(v_2)} & & & \\ & & x^{\mathbf{f}(v_3)} & & \\ & & & \ddots & \\ & & & & x^{\mathbf{f}(v_k)} \end{pmatrix}.$$

Then

$$g_N(x) = \operatorname{tr}(D_k(x)A)^N = \sum_{j=1}^{k} \lambda_j^N (D_k(x)A),$$

where $\lambda_1, \ldots, \lambda_j$ are eigenvalues, and, just as in Section 5, we have $\chi(P_N)(\theta) = g_N(\exp(i\theta))/c_N$, where

$$c_N = |W_N| = \sum_{j=1}^{k} \lambda_j^N (A).$$

Since $\mathcal{G}$ is an *r-regular*, *nonbipartite* graph, it has a unique eigenvalue of maximal modulus, and that eigenvalue is $\lambda_1 = r$.

Now, we can directly apply Conditions 1 and 2 (and accompanying comments) above, and the results of Section 10 (noting that Assumptions 1–4 hold) to obtain the desired result (in particular, the estimate needed in Condition 2 is precisely Theorem 10.8). We replaced the resolvent in formula (10.9) by the equivalent (by the discussion in the beginning of Section 10) Laplacian form, since that is more common in graph theory. □

REMARK 7.2. *If the vector* $\mathbf{f}$ *is an eigenvector of* A^tA *with eigenvalue* r^2, *the corresponding variance is equal to zero. By Remark 10.9 this will not happen, for example, if* G *is a connected* nonbipartite undirected *graph, but it does happen for general directed graphs*; *see the discussion of the directed line graph in Section 8.*

The above remark leads to the following question:

QUESTION 7.3. What combinatorial property of an r-regular directed graph G is reflected in the algebraic statement that the operator norm of $A_0(G)$ is equal to r?

A slight change in notation transforms Theorem 7.1 into a Central Limit theorem for distributions over closed orbits of primitive irreducible Markov processes over a finite number of states—the irreducibilty is exactly equivalent to the connectivity of the graph $\mathcal{G}$ above. For ease of reference, we state this as a separate theorem. The notation for $\mathbf{f}$, μ, etc., is as before; the space W_N is now a probability space with the obvious probability measure; $\mathbf{P} = \mathbf{P}^t$ is the transition matrix (note that Remark 7.2 remains valid in this setting as well).

REMARK 7.4. *Let* $P_N(x)$ *be the distribution of* $X_{\mathbf{f}}$ *on* W_N. *Then* $P_N((x - N\mu(\mathbf{f}))/\sqrt{N})$ *converge to a balanced* (*that is, mean* 0) *normal distribution with variance*

$$\sigma^2(\mathbf{f}) = \frac{1}{k}[-\|\mathbf{f}_0\|^2 + 2\mathbf{f}_0^t(\mathbf{I}_0 - \mathbf{P}_0)^{-1}\mathbf{f}_0] \tag{7.2}$$

$$= \frac{1}{k}\left[\mathbf{f}_0^t\left(-\mathbf{I}_0 + 2r(\mathbf{I}_0 - \mathbf{P}_0)^{-1}\right)\mathbf{f}_0\right].$$

REMARK 7.5. *We have actually shown a slightly stronger result*: *instead of the trace* (*distribution over cycles*), *we could have considered the* ij*th element of* $\mathbf{P}$. *Since the principal eigenvector varies continuously under perturbations* (*see* [26, *Chapter* II.4.1]), *we could have replaced our sample space* W_N *as above by the space* $\mathcal{C}_N$ *of paths of length* N *joining the* i*th to the* j*th vertex. An easy computation shows that the covariance is the covariance given in equation* (7.2), *divided by a further factor of* k. *The same remark applies to Theorem* 7.1.

7.1. Distribution modulo a prime. Theorems 7.1 and 7.4 have particularly simple analogues if the function f we are studying is integer valued, and we are interested in the distribution of the $\mathbf{Z}/p\mathbf{Z}$-valued random variable $Y_f(n)$ which assigns to each cycle of length n the sum of the values of f modulo p. In that case, under the assumption that the adjacency matrix A (in the context of Theorem 7.1) or the transition matrix A (in the context of Theorem 7.4) is irreducible and primitive (the last two $A(\mathcal{L}_u(G))$ conditions guarantee that A has a single eigenvalue λ_0 of maximal modulus, the eigenspace of λ_0 is one-dimensional, and the orthogonal subspace is invariant under A), then we see that the distributions $\mathcal{P}_n$ of $Y_f(n)$ approach the uniform distribution (on $\mathbf{Z}/p\mathbf{Z}$) exponentially fast in n (though a more reasonable measure of the speed of convergence is the size of W_n, in which case the convergence is polynomial). This statement follows from the following lemma.

LEMMA 7.6. *If* A *is a matrix satisfying the conditions above, then the spectral radius* r_{UA} *of* UA, *for* U *any nontrivial unitary matrix such that the top eigenvector of* A *is not also an eigenvector of* U, *is strictly smaller than that of* A (r_A).

The proof of the lemma is immediate.

In our case, the matrix U is the diagonal matrix $U(\chi)$ with $u_{jj} = \chi_p^{f_j}$, with χ_p a nontrivial pth root of unity. The speed of convergence to the uniform distribution is given by $(\max_{\chi^p=1} r(U(\chi)A))/r(A)$.

8. Functions on edges and distributions over paths without backtracking

In this section, we consider two kinds of questions, which are seen to be intimately related. The first is:

QUESTION 8.1. Let f be a function on the *edges* of G. How are the averages of f over long cycles or paths in G distributed?

The second question is:

QUESTION 8.2. Let f be a function on the *vertices* of G. How are the averages of f distributed over long cycles in G *without backtracking*—such

cycles are more closely related to, for example, geodesics on surfaces, then arbitrary cycles.

Both questions can be answered at the same time by constructing the *directed line graph* (or *line digraph*) of G. This construction can be performed for either a directed or undirected graph G; In Section 8.1 we will derive the results for undirected graphs in detail, whilst in Section 8.3 we will discuss the directed case somewhat more briefly (since the technical details are essentially identical).

8.1. The directed line graph of an undirected graph. The *directed line graph* of G, denoted by $\mathcal{L}(G)$, is constructed as follows: The vertices of $\mathcal{L}(G)$ are edges of G labelled with a $+$ or a $-$; that is, to each edge e of G there correspond vertices e_- and e_+ of $\mathcal{L}(G)$. These correspond to the two possible orientations of e: if the vertices of e are v and w, then we say that v is the head of e_-, and w the tail (and write $v = h(e_-)$, $w = t(e_-)$), while for e_+ this nomenclature is reversed. Two vertices v_1 and v_2 of $\mathcal{L}(G)$ are joined by a (directed) edge if the head of v_1 is the same as the tail of v_2, except that e_- is never joined to e_+, and vice versa. We now make some observations and definitions.

Definition 8.3. Let f be a function defined on the vertices of a graph G. We say that a function g defined on the vertices of $\mathcal{L}(G)$ is the *gradient* of f, and write $g = \nabla f$ if $g(e) = f(h(e)) - f(t(e))$.

Definition 8.4. We can identify functions on the vertices of G with (a subset of) functions on the the vertices of $\mathcal{L}(G)$. To wit, if a f is a function on the vertices of G, we let $\mathcal{L}f(e) = f(t(e))$.

Observation 8.5. There is a natural correspondence between walks on $\mathcal{L}(G)$ and walks on G without backtracking. Indeed, passing through a vertex e of $\mathcal{L}(G)$ corresponds to going from $t(e)$ to $h(e)$. Since e_+ is not connected to e_- for any $e \in E(G)$, any such walk is automatically without backtracking. Similarly, a cycle on $\mathcal{L}(G)$ corresponds to a tailless cycle without backtracking on G.

If G is an r-regular graph, then $\mathcal{L}(G)$ is $r-1$-regular, in the strong sense: each vertex of $\mathcal{L}(G)$ has in-degree and out-degree equal to $r-1$ (thus, the total degree is $2r-2$), and from the above Observation 8.5, $\mathcal{L}(G)$ is connected if and only if G is. It follows that the adjacency matrix $A(\mathcal{L}(G))$ of $\mathcal{L}(G)$ is an irreducible nonnegative matrix, all of whose row and column sums are equal to $r-1$. It follows that the space of functions on the vertices of $\mathcal{L}(G)$ orthogonal to the vector $\mathbf{1}$ is an invariant subspace of $A(\mathcal{L}(G))$ and of $A^t(\mathcal{L}(G))$—we will, as before, denote the two matrices restricted to this subspace by A_0 and A_0^t, respectively; the algebraic and geometric multiplicities of the eigenvalue $r-1$ is equal to 1, by standard Perron–Frobenius theory. Despite this, it turns

out that A^tA is spectacularly degenerate. Indeed, the ijth entry of A^tA is equal to the number of vertices of $\mathcal{L}(G)$ adjacent simultaneously to the ith and the jth vertex. It follows that the iith entry of A^tA is equal to $r-1$, while the ijth entry is equal to $r-2$ if the corresponding directed edges of G have the same tail, and is 0 otherwise. It follows that

$$A^tA = I_{2E(G)} + (r-2)\begin{pmatrix} J_1 & & & \\ & J_2 & & \\ & & \ddots & \\ & & & J_{V(G)} \end{pmatrix}, \tag{8.1}$$

where the last term contains $V(G)$ $r \times r$ blocks, each of which is the matrix of all 1s. We thus have the following observation.

OBSERVATION 8.6. The spectrum of A^tA has the following form: The eigenvalue $(r-1)^2$ occurs $V(G)$ times, and the corresponding eigenvectors are given precisely by $\mathcal{L}f$ for arbitrary functions f on G (the Perron eigenvector corresponding to the constant function), while the eigenvalue 1 occurs $2E(G) - V(G)$ times. The eigenvectors are those functions on the directed edges of G, for which, for all vertices v of G, the sum of values on all the edges *leaving* v is equal to 0.

COROLLARY 8.7. *The operator norm of A_0 is equal to $r-1$.*

Consider now the Laplace operator on $\mathcal{L}(G)$: $\Delta_{\mathcal{L}(G)} = (r-1)I - A(\mathcal{L}(G))$. We will need the following in the sequel.

THEOREM 8.8. *Let E_{r-1} be the eigenspace of $(r-1)^2$ for A^tA. If $V^*(G)$ is the space of functions on the vertices of G, then*

(a) $E_{r-1} = \mathcal{L}(V^*(G))$,
(b) $\Delta_{\mathcal{L}(G)}(E_{r-1}) = \nabla(V^*(G))$,
(c) $\nabla(V^*(G)) \cap E_{r-1} \cap \mathbf{1}^\perp = \emptyset$, *unless G is bipartite.*

Proof. Part (a) is the content of Observation 8.6. Part (b) is a corollary of part (a). Indeed, $\Delta_{\mathcal{L}(G)}(f)(x) = (r-1)f(x) - \sum_{h(x)=t(y)} f(y)$. If $f = \mathcal{L}g$, then

$$\Delta_{\mathcal{L}(G)}(f)(x) = (r-1)\big(g(t(x)) - g(h(x))\big), \tag{8.2}$$

since all the y adjacent to x have the same tail, equal to the head of x.

To show part(c), suppose $\nabla(V^*(G)) \cap E_{r-1} \neq \emptyset$. Let g be in the intersection, and k be such that $\nabla(k) = g$. It follows that for any x, y such that $t(x) = t(y)$, $g(x) = g(y)$. We see that $k(h(x)) - k(t(x)) = k(h(y)) - k(t(y))$, which implies in turn that $k(h(x)) = k(h(y))$. So, k is the eigenvector of the 0 eigenvalue of the Laplace operator on G, and hence is constant, unless G is bipartite. □

We end this section with a remark necessary to compute distributions, as done in the following Section 8.2.

REMARK 8.9. *The adjacency matrix of the line graph of a nonbipartite graph G is primitive. That is, there is only one eigenvalue on the circle of radius $r-1$ in the complex plane, and that is $r-1$. Its geometric multiplicity is* 1.

Proof. Doubtlessly there are simpler arguments, but we choose to use the results (described in [59]) on the Ihara zeta function Z of G, which can be expressed as a determinant in two ways:

The first way (original theorem of Ihara [21]) is:

$$Z^{-1}(u) = (1-u^2)^{\mathcal{R}-1} \det\big(\big(1+(r-1)u^2\big)\mathbf{I} - uA\big), \tag{8.3}$$

with A the adjacency matrix of G, and $\mathcal{R}$ the rank of the fundamental group of G.

The second way (due to Hyman Bass [2]) is:

$$Z^{-1}(u) = \det(\mathbf{I} - uM), \tag{8.4}$$

where M is the adjacency matrix of the directed line graph of G.

The equality of the two expressions implies that v is an eigenvalue of M if and only if $v + (r-1)/v$ is an eigenvalue of A (we are ignoring the eigenvalues ± 1, which occur with large multiplicity in the spectrum of M). Suppose that v has modulus $r-1$, so that $v = (r-1)\exp(i\theta)$, for some θ. It follows that $w = \exp(i\theta) + (r-1)\exp(-i\theta)$ is an eigenvalue of A, and since A is symmetric, $\theta \in \{0, \pi\}$. If $\theta = 0$, $v = r-1$, while if $\theta = \pi$, $v = -(r-1)$, but then $w = -r$ is an eigenvalue of A, and so G is bipartite.

The statement about the multiplicity of the eigenvalue $r-1$ is immediate, since $\mathcal{L}(G)$ is clearly strongly connected. □

We include the following observations both for the sake of completeness, and in view of Lemma 8.14 below.

LEMMA 8.10.

$$\Delta\mathcal{L} = (r-1)\nabla.$$

Proof. Indeed, $\mathcal{L}(f)(x) = f(t(x))$. Further,

$$\begin{aligned} \Delta\mathcal{L}(f)(x) &= \sum_{t(y)=h(x)} f\big(t(x) - f(t(y)\big) = (r-1)\big(f(t(x)) - f(h(x))\big) \\ &= \nabla(f)(x). \end{aligned} \tag{8.5}$$

□

LEMMA 8.11. *For any $f, g \in V^*(G)$, we have*

$$(\mathcal{L}f)^t \nabla g = f^t \Delta g.$$

Proof. Indeed,

$$(\mathcal{L}f)^t\nabla g = \sum_x \big(f(t(x))(g(t(x)) - g(h(x)))\big) \tag{8.6}$$
$$= \sum_{v\in V(G)} \sum_{\text{w adjacent to v}} f(v)g(v) - f(v)g(w)$$
$$= \sum_{v\in V(G)} f(v)\Delta(g)(v)$$
$$= f^t\Delta g. \qquad \square$$

Consider now a function g on the directed edges of G. How do we decompose it into a gradient and a function orthogonal to gradients? First, we note that a basis of the gradients is formed by the gradients of δ functions:

$$\delta_v(x) = \begin{cases} 1, & x = v, \\ 0, & \text{otherwise.} \end{cases} \tag{8.7}$$

So that

$$\nabla\delta_v(x) = \begin{cases} 1, & t(x) = v, \\ -1, & h(x) = v, \\ 0, & \text{otherwise.} \end{cases} \tag{8.8}$$

The functions $\nabla\delta_v$ form a basis of $\nabla(V^*(G))$, though not an orthonormal one. Now, note that

$$g^t\nabla\delta_v = \sum_{t(x)=v} g(x) - \sum_{h(y)=v} g(y).$$

In other words, we have the following lemma.

LEMMA 8.12. *g is orthogonal to the gradients, if and only if the sum of g over the edges coming into any vertex v is equal to the sum of g over the edges leaving v. An equivalent condition is that $\nabla^t g = 0$.*

One may ask: what is the orthogonal projection of a given $\mathcal{L}f$ onto the gradients? The following comes out of an easy computation:

OBSERVATION 8.13. The orthogonal projection of $\mathcal{L}f$ onto the set of gradients is $\nabla\Delta f$.

8.2. Applications to distribution. We can use the results of the previous section to understand the limiting distribution of functions defined on (directed) edges of G. Indeed, we can use Theorem 7.1 in the form corresponding to equation (10.10) to observe that

$$\sigma^2(\mathbf{f}) = \frac{1}{2rk}\mathbf{f}^t(\Delta_0^{-1})^t(r-1)^2\mathbf{I} - A(\mathcal{L}(G))^t A(\mathcal{L}(G))\Delta_0^{-1}\mathbf{f} \tag{8.9}$$

for $\mathbf{f}$ any function on the directed edges of G, and Δ_0 the restriction of the Laplace operator on $\mathcal{L}(G)$ to the subspace of 0-sum vectors.

LEMMA 8.14. *The right-hand side of equation* (8.9) *vanishes precisely when* $\mathbf{f}$ *is the gradient of a function on the vertices of* G.

Proof. Let $\mathbf{f} = \Delta u$. By Observation 8.6, we see that the right-hand side of equation (8.9) vanishes precisely if $u \in \mathcal{L}(V^*(G))$. By part (b) of Theorem 8.8 it follows that this is so if and only if $\mathbf{f} \in \nabla(V^*(G))$. □

One direction of the above lemma is just common sense, since the sum over any cycle of a gradient is equal to 0.

Keeping the above in mind, we note that a simpler form of the covariance is given by Theorem 7.1:

$$\sigma^2(\mathbf{f}) = \frac{r-1}{2rk}\left[\mathbf{f}^t\left(\mathbf{I} - 2(r-1)\Delta_0^{-1}\right)\mathbf{f}\right]. \tag{8.10}$$

For functions on the vertices of G, the above assumes the form:

$$\sigma^2(\mathbf{f}) = \frac{r-1}{2rk}\left[\mathbf{f}^t\mathcal{L}^t\left(\mathbf{I} - 2(r-1)\Delta_0^{-1}\right)\mathcal{L}\mathbf{f}\right]. \tag{8.11}$$

8.3. The line graph of a directed graph. The construction of the line graph of a directed graph G is essentially the same as that of an undirected graph. This time, the vertices of $\mathcal{L}(G)$ without labels (so $\mathcal{L}(G)$ has $E(G)$ vertices). The operators ∇ and $\mathcal{L}$ are defined as in Section 8.1. We have an observation even simpler than Observation 8.5.

OBSERVATION 8.15. There is a natural bijective correspondence between walks on $\mathcal{L}(G)$ and walks on G.

If G is an r-regular directed graph (by this we mean that both the in- and out-degree of each vertex is equal to r), then so is $\mathcal{L}(G)$; by Observation 8.15 $\mathcal{L}(G)$ is connected whenever G is. As before, $A(\mathcal{L}(G))$ is the adjacency matrix of $\mathcal{L}(G)$. We can compute:

$$A^tA = r\begin{pmatrix} J_1 & & & \\ & J_2 & & \\ & & \ddots & \\ & & & J_{V(G)} \end{pmatrix}, \tag{8.12}$$

where each block corresponds to the set of edges of G emanating from a given vertex. From this we have the following observation.

OBSERVATION 8.16. The spectrum of $A^t(\mathcal{L}(G))A(\mathcal{L}(G))$ has the following form: The eigenvalue r^2 occurs $V(G)$ times, and the corresponding eigenvectors are given by $\mathcal{L}f$ for arbitrary functions f on G (The Perron eigenvector corresponding to the constant function) while the eigenvalue 0 occurs $E(G) - V(G)$ times. The eigenvectors are those functions on the edges of G

for which the sums of the values over all edges leaving a vertex v is equal to 0 (for all v).

COROLLARY 8.17. *The operator norm of $A_0(\mathcal{L}(G))$ is equal to r.*

The Laplace operator on $\mathcal{L}(G)$ is defined as: $\Delta_{\mathcal{L}(G)} = rI - A(\mathcal{L}(G))$.
We have

THEOREM 8.18. *Let E_r be the eigenspace of r^2 for A^tA. If $V^*(G)$ is the space of functions on the vertices of G, then*

(a) $E_r = \mathcal{L}(V^*(G))$,
(b) $\Delta_{\mathcal{L}(G)}(E_r) = \nabla(V^*(G))$.

We also include the following remark.

REMARK 8.19. *The adjacency matrix of the line graph of G is primitive if the adjacency matrix of G is.*

Proof. We use Observation 8.15 and Theorem 15.1 to note that the nonzero eigenvalues of G are exactly the same as those of $\mathcal{L}(G)$, since $\det(I - uA(G)) = \det(I - uA(\mathcal{L}(G)))$. □

LEMMA 8.20.

$$\Delta\mathcal{L} = r\nabla.$$

LEMMA 8.21. *For any $f, g \in V^*(G)$, we have*

$$(\mathcal{L}f)^t \nabla g = f^t \Delta g.$$

Lemma 8.12 and Observation 8.13 go through without change.

The results of Section 8.2 go through essentially without change. Since some constants change, we restate them here. First, let f be a function defined on the edges of $\mathcal{L}(G)$. We see that:

$$\sigma^2(\mathbf{f}) = \frac{1}{2rk}\mathbf{f}^t(\Delta_0^{-1})^t r^2\mathbf{I} - A(\mathcal{L}(G))^t A(\mathcal{L}(G))\Delta_0^{-1}\mathbf{f}. \tag{8.13}$$

Lemma 8.14 holds as well, and this gives us the following useful corollary (a homological condition) about distribution on G itself.

THEOREM 8.22. *The variance of a function f on the vertices of G vanishes, precisely when there exists a function g, such that $\mathcal{L}f = \nabla g$.*

Finally, we have a version of formula (8.10):

$$\sigma^2(\mathbf{f}) = \frac{1}{2k}[\mathbf{f}^t(\mathbf{I} - 2r\Delta_0^{-1})\mathbf{f}]. \tag{8.14}$$

9. Distribution in compact groups

The methods of the Section 7.1 can be adapted to the following setting: Let G is a graph, and T be a compact topological group. Label the ith vertex of G with $t_i \in T$. Now, associate to each cycle $c = v_1, \ldots, v_k$ on G the element $t_c = t_k \cdot \cdots \cdot t_1 \in T$. We ask: as c varies over the cycle space W_N, how are the elements t_c distributed in T (with respect to the Haar measure). The answer is given by the following theorem.

THEOREM 9.1. *If the graph G is as before (connected, nonbipartite), the closed subgroup generated by the t_i $(i = 1, \ldots, k)$ is equal to T, and the elements t_i do not all lie in the same coset with respect to a one-dimensional representation of T, then the elements t_c become equidistributed, as $N \to \infty$.*

Proof. As before, the equidistribution is equivalent to the assertion that for a *nontrivial* irreducible unitary representation ρ,

$$\sum_{c \in W_n} \operatorname{tr}(\rho(t_c)) = o(|W_n|). \tag{9.1}$$

This follows from the Fourier transform formula for compact groups; see [11] for the finite case, [63] for the general compact topological group case. See also [37] Now, let $U(\rho)$ be the $k \deg \rho \times k \deg \rho$ block-diagonal matrix whose jth block is just $\rho(t_j)$. Further more, as before, let $A(G)$ be the adjacency matrix of G, and $A_l(G) = A(G) \otimes \mathbf{I}_l$ (where $\mathbf{I}_l$ is the $l \times l$ diagonal matrix): in other words, $A_l(G)$ is a $kl \times kl$ matrix, obtained from $A(G)$ by replacing each element a_{ij} by a $k \times k$ matrix M_{ij}, all of whose elements are equal to a_{ij}. It is not hard to see that the left-hand side of equation (9.1) is equal to $\operatorname{tr}(U(\rho)A_{\deg \rho}(G))^N$, and so it suffices to show that the spectral radius of $M_\rho = U(\rho)A_{\deg \rho}(G)$ is strictly smaller than the spectral radius of $A(G)$ (which we normalize to be equal to 1 by scaling) under the hypotheses of the theorem. Suppose not. Since $(U(\rho))$ is unitary, the worst that can happen is that there exists a unit vector v, such that $\|M_\rho(v)\| = 1$. If that is so, v is contained in the eigenspace of eigenvalue 1 of $A_{\deg \rho}$. In such a case, $v = v_1 \otimes u$, where $u \in V(\rho)$, and v_1 is an eigenvector of $A(G)$ with eigenvalue 1. If $v_1 = (v_1^1, \ldots, v_1^n)$, then $v_1^i u$ must be an eigenvector of $\rho(t_i)$, for all i. Since $v_1^i \neq 0$ $\forall i$, this implies that u is an eigenvector $\rho(t_i)$, $\forall i$. Since ρ is irreducible, this implies that *either* the elements $t_1, \ldots, t_k$ do not generate all of T, or ρ is 1-dimensional, in which case clearly $\rho(t_i) = \rho(t_j)$, $\forall i, j$, which proves the theorem. □

REMARK 9.2. *As in Remark 7.5, the above argument also works if we pick all paths between the ith and the jth vertex of G, instead of all cycles.*

10. Some perturbations and estimates

Consider an analytic family of linear operators $M(x)$, acting on $\mathbf{R}^k$, with $M(0) = M$, and let λ be a simple eigenvalue of M. Then if

$$M(x) = M + M^{(1)}x + M^{(2)}x^2 + \cdots,$$

perturbation theory (see [26, page 79, formula (2.33)]) tells us that

$$\lambda(x) = \lambda + \lambda^{(1)}x + \lambda^{(2)}x^2 + \cdots,$$

where

$$\lambda^{(1)} = \operatorname{tr} M^{(1)}P_\lambda, \tag{10.1}$$

$$\lambda^{(2)} = \operatorname{tr}\bigl[M^{(2)}P_\lambda - M^{(1)}S_\lambda M^{(1)}P_\lambda\bigr], \tag{10.2}$$

where P_λ is the projection onto the eigenspace of λ, while S_λ is the *reduced resolvent* of M at λ, which is the holomorphic part of the resolvent of M at λ, defined by the properties

$$S_\lambda P_\lambda = P_\lambda S_\lambda = 0; \qquad (M - \lambda\mathbf{I})S_\lambda = S_\lambda(M - \lambda\mathbf{I}) = \mathbf{I} - P_\lambda \tag{10.3}$$

(in other words, S_λ is the inverse of $M - \lambda\mathbf{I}$ restricted to the orthogonal complement of the eigenspace of λ), and thus

$$MS_\lambda = \mathbf{I} - P_\lambda + \lambda S_\lambda. \tag{10.4}$$

Now we will specialize a bit:

Assumption 1. The eigenvalue λ is such that the constant vector $\mathbf{1}$ spans the eigenspace of λ.

In this case, $P_\lambda = J_k/k$, where we recall that J_k is the $k \times k$ matrix of all 1s.

In addition, the following assumption.

Assumption 2. We will assume that $M(x) = D(x)M$, where $D(x)$ is an analytically varying diagonal matrix, $D(x) = D + D^{(1)}x + D^{(2)}x^2 + \cdots$, where we say that the diagonal elements of $D^{(l)}$ are $\mathbf{d^{(l)}} = (d_1^{(l)}, \ldots, d_k^{(l)})$.

Lemma 10.1. *Let $A = (A_{ij})$ be an $n \times n$ matrix. Then*

$$\operatorname{tr} AJ_n = \sum_{1 \le i,j \le n} A_{ij}.$$

Lemma 10.2. *Let $A = (A_{ij})$ be an $n \times n$ matrix, and let X be an $n \times n$ diagonal matrix. Then*

$$(XA)_{ij} = A_{ij}X_{ii},$$
$$(XAX)_{ij} = A_{ij}X_{ii}X_{jj}.$$

LEMMA 10.3. *Let D be a diagonal matrix, with diagonal elements $d_1, \ldots, d_n$. Then*

$$v^t D v = \sum_{i=1}^{n} d_i v_i^2.$$

The proofs of the above lemmas are immediate.

LEMMA 10.4. *Let P_v is the projection operator on the subspace generated by v (a unit vector). Then*

$$\operatorname{tr} M P_v = v^t M v.$$

In particular, if v is an eigenvector of M with eigenvalue λ, then $\operatorname{tr} M P_v = \lambda \|v\|$.

Proof. This follows by a direct computation, since when v is a unit vector, $(P_v)_{ij} = v_i v_j$. □

LEMMA 10.5. *If v is an eigenvector of M with eigenvalue λ, then $M P_v = \lambda P_v$.*

LEMMA 10.6. *Suppose that λ has multiplicity 1, and $v(\lambda)$ is a unit vector generating the eigenspace of λ, and $M(t) = D(t)M$, where $D(t)$ is a diagonal matrix. Then*

$$\lambda'(M) = \lambda v^t(\lambda) D' v.$$

Proof. By formula (10.1), we have

$$\lambda'(M) = \operatorname{tr} M' P_{v(\lambda)} = v^t(\lambda) M' v(\lambda) = \lambda v^t(\lambda) D' v.$$ □

COROLLARY 10.7. *In the case when $v(\lambda) = \frac{1}{\sqrt{k}}\mathbf{1}$, we have*:

$$\lambda^{(1)} = \frac{\lambda}{k} \sum_{j=1}^{k} d_j^{(1)}. \tag{10.5}$$

To compute the second derivative of λ, we use the formula (10.2) (we are assuming that λ is an isolated eigenvalue with eigenvector $v(\lambda)$, and $M(t) = D(t)M$, as before):

$$\begin{aligned}
\lambda'' &= \operatorname{tr}[M'' P_{v(\lambda)} - M' S_\lambda M' P_\lambda] \\
&= \lambda v^t D'' v - \operatorname{tr}[M' S_\lambda M' P_\lambda] \\
&= \lambda v^t D'' v - \lambda \operatorname{tr}[D' M S_\lambda D' P_\lambda] \\
&= \lambda v^t [D'' - D' M S_\lambda D'] v.
\end{aligned}$$

We can now use the formula (10.4) to get:

$$\lambda'' = \lambda v^t [D'' - D'(\mathbf{I} - P_\lambda) D' - \lambda D' S_\lambda D'] v. \tag{10.6}$$

In the special case where the eigenvector v is proportional to $\mathbf{1}$, we can rewrite the formula in coordinates in a simple way. To wit, any diagonal

matrix D can be written (uniquely) as $D_0 + d\mathbf{I}$, where D_0 is such that $\operatorname{tr} D_0 = 0$. A simple computation then shows that

$$\lambda'' = \frac{\lambda}{k}\left[\sum_{j=1}^{n} d_j'' - \sum_{j=1}^{n}(d_0')^2 - \lambda \mathbf{d}'^t S_\lambda \mathbf{d}'\right]. \tag{10.7}$$

The case we are interested in is still more special, and that is where we get the following assumption.

Assumption 3.

$$D(x) = \begin{pmatrix} \exp(if_1 x) & & & \\ & \exp(if_2 x) & & \\ & & \ddots & \\ & & & \exp(if_k x) \end{pmatrix}.$$

Here, $\mathbf{d}^{(1)} = (if_1, if_2, \ldots, if_k)$, while $\mathbf{d}^{(2)} = -\frac{1}{2}(f_1^2, f_2^2, \ldots, f_k^2)$, and so, letting $\mathbf{f} = (f_1, \ldots, f_k)$,

$$\lambda^{(2)} = \frac{\lambda}{k}\left[-\frac{1}{2}\|\mathbf{f}\|^2 + \|\mathbf{f}_0\|^2 + \lambda \mathbf{f}^t S_\lambda \mathbf{f}\right], \tag{10.8}$$

where, as before, $\mathbf{f}_0$ is the component of $\mathbf{f}$ orthogonal to constants.

To show our final estimates we shall need the following assumption.

Assumption 4. The matrix M is $\lambda > 0$ times a doubly stochastic matrix (this implies that the operator norm and the spectral radius of M are both equal to λ).

Theorem 10.8. *With assumptions as above, and, in addition,* $\mathbf{f} = \mathbf{f}_0$ (*that is,* $\sum_{j=1}^{k} f_j = 0$), *then* $\lambda^{(2)}$ *is nonpositive.*

Proof. Since $\mathbf{f} = \mathbf{f}_0$, equation (10.8) can be rewritten as

$$\lambda^{(2)} = -\frac{\lambda}{2k}[-\|\mathbf{f}_0\|^2 - 2\lambda \mathbf{f}_0^{\perp} S_\lambda \mathbf{f}_0] = -\frac{\lambda}{2k}[\mathbf{f}_0^t(-\mathbf{I} - 2\lambda S_\lambda)\mathbf{f}_0]. \tag{10.9}$$

If we regard S_λ as an operator on the orthogonal complement to $\mathbf{1}$, then by equations (10.3) and (10.4), $S_\lambda(\lambda \mathbf{I}_0 - M_0) = -\mathbf{I}_0$. Let $v = -S_\lambda \mathbf{f}_0$. Then the term in square brackets in equation (10.9) can be rewritten as:

$$v^t(\lambda \mathbf{I}_0 - M_0)^t(-\mathbf{I} - 2\lambda S_\lambda)(\lambda \mathbf{I}_0 - M_0)v = v_t(\lambda^2 \mathbf{I}_0 - M_0^t M_0)v, \tag{10.10}$$

where we have used the fact that for any matrix A and any vector v, $v^t A v = v^t A^t v$. The quadratic form $\lambda^2 \mathbf{I}_0 - M_0^t M_0$ is positive semi-definite, since the biggest eigenvalue of the symmetric matrix $M_0^t M_0$ is equal to the square of the operator norm of M_0, which, in turn, is no greater then λ, by Assumption 4 (since $M^t M$ is λ^2 times a doubly stochastic matrix). □

REMARK 10.9. *In the statement of Theorem 10.8, the word "nonpositive" can be improved to "negative" under the further assumption that M is irreducible, primitive, and normal.*

Proof. Since the orthogonal complement to the subspace generated by the vector $\mathbf{1}$ is invariant under M, it follows that M_0 is also normal, and so its operator norm is equal to its spectral radius μ. Under the assumptions of irreducibility and primitivity, Perron–Frobenius theory tells us that $|\mu| < \lambda$. □

11. Topological entropy

Consider a graph G, and consider a positive function f on its vertices. For each cycle c we let $F(c)$ to be the sum of values of f over c, and we want to know how many c are there for which $F(c) \leq L$. We denote that number by $N(f, L)$, and we ask ourselves how $N(f, L)$ behaves asymptotically as L tends to infinity. To understand $N(L, f)$, we consider first the matrix $U(f) = D(u^{f_1}, \ldots, u^{f_n})A(G)$. As before, we observe that the coefficient of u^r in $\operatorname{tr} U^n(f)$ is the number of cycles of (combinatorial) length n, for which $F(c) = r$. Write a formal series

$$L(f, u) = \sum_n \operatorname{tr} U^n(f).$$

This series converges for sufficiently small u, and can there be written in closed form as $L(f, u) = \operatorname{tr}(\mathbf{I} - U(f))^{-1}$, from which it follows that the exponential rate of growth of $N(c)$ is equal to negative logarithm of the radius of convergence of $L(f, u)$—we call this the *entropy of* G, f—which, in turn, is equal to the smallest positive real value of u, such that the spectral radius of $U(f)$ is equal to 1. Since it is more convenient to deal with analytic functions (which $L(f, u)$ is not), for arbitrary real values of f_i, so we write $u = \exp(-s)$, and now ask for the abscissa of convergence of $L(f, \exp(-s))$. This will give us the entropy. In this section, we use perturbation methods in a rather straightforward way to get explicit information on the entropy.

Let A be an $n \times n$ nonnegative primitive irreducible matrix. Let $f_1, \ldots, f_n$ be a collection of weights. We then define the matrix $E(s, \mathbf{f})$ to be the diagonal matrix whose iith element is equal to $\exp(-sf_i)$. Define $M(s, \mathbf{f})$ to be $M(s, \mathbf{f}) = E(s, \mathbf{f})A$. We are interested in $\rho(s, \mathbf{f})$: the spectral radius of $M(s, \mathbf{f})$. By Perron–Frobenius theory, we know that there is a real eigenvalue of $M(s, \mathbf{s})$ equal to $\rho(s, \mathbf{f})$, and the eigenvector v_ρ of this eigenvalue is positive.

LEMMA 11.1.

$$\frac{\partial \rho}{\partial s} = -\rho v_\rho^t D(f_1, \ldots, f_n) v. \tag{11.1}$$

For positive $\mathbf{f}$, $\frac{\partial \rho}{\partial s} < 0$.

Proof. This follows immediately from Lemma 10.4 and the positivity of ρ and v_ρ. □

LEMMA 11.2. *We have the following expression for the gradient of ρ with respect to* $\mathbf{f}$:

$$\nabla_{\mathbf{f}}\rho = -s\rho(v_1^2, \dots, v_n^2), \tag{11.2}$$

where $v_\rho = (v_1, \dots, v_n)$.

Proof. We note that

$$\frac{\partial M}{\partial f_i} = -sD(0, \dots, 1, \dots, 0)M,$$

where the 1 is in the ith place. Thus, by formula (10.1), we have

$$\frac{\partial \rho}{\partial f_i} = -sv_\rho^t D(0, \dots, 1, \dots, 0)Mv = -s\rho v_i^2. \qquad \square$$

This can be restated as saying that the derivative of ρ in the direction of a vector $\mathbf{g}$ is equal to $-\rho s v_\rho^t D(\mathbf{g})v$.

This gives us the following important corollary.

COROLLARY 11.3. *Consider deformations g keeping the sum of f_i fixed. Then the critical points of ρ occur precisely for those ρ for which $|v_i| = |v_j|$, for any i, j.*

We can also compute the second directional derivative of ρ. Indeed, let $\mathbf{g} = (g_1, \dots, g_n)$ be the direction vector, so that we want to compute the second derivative with respect to t of $\rho(s, \mathbf{f} + t\mathbf{g})$ at $t = 0$. To do this, we use the formula (10.2):

$$\rho'' = \operatorname{tr}\big[M'' P_{v(\rho)} - M' S_\rho M' P_\rho\big]. \tag{11.3}$$

Note that (as in the proof of Lemma 11.2)

$$M' = -sD(g_1, \dots, g_n)M, \tag{11.4}$$

while

$$M'' = s^2 D(g_1^2, \dots, g_n^2)M,$$

and so

$$\begin{aligned} \operatorname{tr} M'' P_{v(\rho)} &= s^2 \rho v^t D(g_1^2, \dots, g_n^2)v \\ &= s^2 \rho \{D(g_1, \dots, g_n)v\}^t \{D(g_1, \dots, g_n)v\}. \end{aligned} \tag{11.5}$$

To understand the second term of the right-hand side of equation (11.3), first note that (by equation (11.4))

$$\begin{aligned} M' S_\rho M' P_\rho &= D(g_1, \dots, g_n)MP_\rho \\ &= \rho s^2 D(g_1, \dots, g_n)MS_\rho D(g_1, \dots, g_n)P_\rho, \end{aligned}$$

where the second equality is by Lemma 10.5. Now

$$\operatorname{tr} M' S_\rho M' P_\rho = \rho s^2 v(\rho)^t D(g_1, \dots, g_n) M S_\rho D(g_1, \dots, g_n) v \tag{11.6}$$

$$= \rho s^2 \{D(g_1, \dots, g_n)v\}^t M S_\rho \{D(g_1, \dots, g_n)v\}. \tag{11.7}$$

Putting together equations (11.5) and (11.6), we see that

$$\rho'' = \rho s^2 \{D(g_1, \dots, g_n)v\}^t (\mathbf{I} - M S_\rho)\{D(g_1, \dots, g_n)v\}. \tag{11.8}$$

Using the formula (10.4), equation (11.8) simplifies further to:

$$\rho'' = \rho s^2 \{D(g_1, \dots, g_n)v\}^t (P_{v(\rho)} - \rho S_\rho)\{D(g_1, \dots, g_n)v\}. \tag{11.9}$$

The following lemma is not surprising.

LEMMA 11.4. *The quadratic form given $P_v - \rho S_\rho$ is positive-definite.*

Proof. On the span of v, the projection operator P_v is equal to the identity, whilst the reduced resolvent S_ρ vanishes. On the orthogonal complement, the projection operator vanishes, so since the Perron–Frobenius eigenvalue ρ is positive, we need to show that S_ρ is negative-definite. Consider a vector w, in the orthogonal complement of v. Such a w is equal to $(\rho \mathbf{I} - M)z$, for some z orthogonal to v. So,

$$w^t S_\rho w = z^t (\rho I - M) z.$$

So, it will suffice to show that $(\rho I - M)$ is negative-definite. Suppose not. Then there exists a z_0, such that $z_0^t M z_0 \geq \rho \|z_0\|^2$. By the argument in the proof of Theorem 10.8, we see that $\|M z_0\| \leq \rho \|z_0\|$. So, $z_0^t M z_0 \geq \rho \|z_0\|^2$ implies that $\langle z_0, M z_0 \rangle \geq \rho \|z_0\|^2$, and hence that z_0 is an eigenvactor of M with eigenvalue ρ, which is impossible by assumption that M is irreducible and primitive. □

We finish with the following theorem.

THEOREM 11.5. *Let $s_0(\mathbf{f})$ be the unique s such that $\rho(s_0, \mathbf{f})$ is equal to* 1. *Then s_0 is a convex function of $\mathbf{f}$, and hence assumes a unique minimum on each linear subspace of values of $\mathbf{f}$. In particular, if we restrict to the the subspace F_0, where the sum of the values of of $\mathbf{f}$ is equal to* 1, *then the minimum is achieved at the point where*

$$f_i = \frac{\log(A\mathbf{1})_i}{\sum_i \log(A\mathbf{1})_i}, \tag{11.10}$$

in which case the entropy is equal to $\sum \log(A\mathbf{1})_i$.

Proof. The convexity of s_0 follows from Lemmas 11.4 and 11.1. The point at which the minimum is achieved is computed easily using Corollary 11.3, as is the value of entropy. □

12. Applications to groups and other objects

The asymptotic results in the previous sections apply directly to the question of the growth of homology classes in the free groups, and give in some sense complete information:

Observation 12.1. We see that the asymptotic order of growth of any two *fixed* homology classes is the same.

Observation 12.2. Theorem 7.1 shows in particular that a random long cycle is equidistributed among the vertices of a regular graph.

Observation 12.3. We see that the order of growth the number of words length n in any fixed homology class in F_k is asymptotic to $c_k(2k-1)^n/n^{k/2}$, where c_k is easily computed using the expression for σ in the statement of Theorem 5.1, keeping in mind that

$$c_{F_k} = \frac{k}{\sqrt{2k-1}},$$

where c is the parameter in the statements of theorems of the last two sections. Alternately, Theorem 7.1 can be used.

We can compute other growth functions. For example, let $h : F_n \to \mathbf{Z}$ be the "total exponent" homomorphism, that is, if $F_n = \langle a_1, \ldots, a_n \rangle$, then $h(a_i) = 1$. We see that the generating function for the preimages of $j \in Z$ is given by

$$\left(2\sqrt{2n-1}\right)^k R_k\left(\frac{n}{\sqrt{2n-1}}; x, \ldots, x\right) = \left(2\sqrt{2n-1}\right)^k R_k\left(\frac{n}{\sqrt{2n-1}}; x\right).$$

Observation 12.4. Instead of cyclically reduced words, it is perhaps more natural to study conjugacy classes (ordered by their cyclically reduced length). It seems futile to seek any enumeration as neat as Theorem 2.3, however, since the relationship between the number $\mathcal{C}_k$ of conjugacy classes of words of length k and the number of cyclically reduced words $\mathcal{W}_k$ is:

$$\mathcal{C}_k = \frac{\mathcal{W}_k}{k} + O\left(\sqrt{\mathcal{W}_k}\right), \tag{12.1}$$

it is clear that the asymptotic results are the same for the two problems. For more on this subject, see Section 13 and the sequel.

Observation 12.5. Counting conjugacy classes is a problem closely related to that of counting closed geodesics on manifold. In the context of compact hyperbolic surfaces, it was observed by Sarnak (see, for example, [54]) that among all geodesics shorter than L, null-homologous geodesics are more numerous than those in any other prescribed homology class (that is, while the ratio of the two quantities approaches 1, the difference is asymptotically positive). The results of the current note provide a certain justification for this, since any limiting distribution likely to arise in this context is, for reasons of

symmetry, likely to be unimodal, with the mode at **0**. Certainly this is true of the normal distribution, though even in this case, a careful analysis of the error terms is required.

13. Counting conjugacy classes

Consider a finitely presented group G. Let g be an element of G. We define the *reduced length* of g—denoted by $|g|$—to be the length of the shortest word in the generators of G representing g. We define the *length up to conjugacy* of g—denoted by $|g|_c$—to be the minimum of $|h|$, the minimum being taken over all group elements h conjugate to g. Length up to conjugacy is obviously invariant under conjugation, and we will also use the term to apply to conjugacy classes.

$$\begin{aligned}\mathcal{N}_G(r) &= \big|\{g \in G \mid |g| = r\}\big|,\\ \mathcal{C}_G(r) &= \big|\{g \in \mathcal{N}_G(r) \mid |g|_c = r\}\big|,\\ \mathcal{CC}_G(r) &= \big|\{C \in G/\text{conjugacy} \mid |C|_c = r\}\big|.\end{aligned}$$

The subscript G will be omitted whenever the group G is obvious from context.

Given a sequence $A = a_0, \ldots, a_i, \ldots$, we can define a *generating function* $\mathcal{F}[A]$, by

$$\mathcal{F}[A](z) = \sum_{i=0}^{\infty} a_i z^i.$$

There is frequently confusion as to whether the generating function is a holomorphic function or an element of the ring of formal power series. In this section, "generating function" will mean a function analytic at $0 \in \mathbf{C}$.

The three counting functions above give rise to corresponding generating functions $\mathcal{F}[\mathcal{N}_G]$, $\mathcal{F}[\mathcal{C}_G]$, $\mathcal{F}[\mathcal{CC}_G]$. Our real interest will lie in the last of these; the first one has been the most extensively studied, and the result most relevant to us is:

Fact 1. If G is an *automatic* group, then the generating function $\mathcal{F}[\mathcal{N}_G]$ is a rational function.

For definitions and properties of automatic groups, see [9].

Fact 2 (Gromov, Epstein). If G is an automatic group, then the generating function $\mathcal{F}[\mathcal{C}_G]$ is a rational function.

Facts 1 and 2 might lead us to expect that $\mathcal{F}[\mathcal{CC}_G]$ is, likewise, rational, but in fact the opposite seems to be the case, and we are led to the following.

Conjecture 13.1. Let G be a word-hyperbolic group. The $\mathcal{F}[\mathcal{CC}_G]$ is rational if and only if G is virtually cyclic (*elementary* in the terminology of [12]).

In the sequel, this conjecture is supported by the complete analysis of the case where G is F_k—the free group on k generators.

14. Growth functions for free groups

Let F_k be the free group on k generators. The following is obvious:

FACT 3. $\mathcal{N}_{F_k}(r) = 2k(2k-1)^{r-1}$.

Theorem 1.1 says that

$$\mathcal{C}_{F_k}(r) = (2k-1)^r + 1 + (k-1)[1 + (-1)^r].$$

COROLLARY 14.1.

$$\mathcal{F}[\mathcal{C}_{F_k}](z) = \frac{1}{1-(2k-1)z} + \frac{1}{1-z} + \frac{2(k-1)}{1-z^2} - 2k.$$

In order to compute $\mathcal{CC}_{F_k}(r)$, it is enough to notice the following theorem.

THEOREM 14.2.

$$r\mathcal{CC}(r) = \sum_{d|r} \phi(d)\mathcal{C}(r/d),$$

where ϕ denotes the Euler totient function.

Proof. The theorem is a trivial consequence of Burnside's lemma, stated below as Theorem 14.3 for convenience, applied to the action of the cyclic group $\mathbf{Z}/(r\mathbf{Z})$ on the set of cyclically reduced words of length r. □

THEOREM 14.3. *Let G be a finite group acting on a finite set X. For $g \in G$ let $\psi(g)$ denote the number of $x \in X$, such that $g(x) = x$. Then the number of orbits of X under the G-action is*

$$\frac{1}{|G|} \sum_{g \in G} \psi(g).$$

We now have the following general observation.

THEOREM 14.4. *Suppose we have three sequences $A = \{a_i\}$, $B = \{b_j\}$, and $C = \{c_k\}$, satisfying*

$$a_n = \sum_{d|n} c_d b_{\frac{n}{d}}.$$

Then

$$\mathcal{F}[A](z) = \sum_{d=1}^{\infty} c_d \mathcal{F}[B](x^d).$$

Proof. On the level of formal power series, the statement is clear by expanding the left-hand side. Otherwise, if the radius of convergence of $\mathcal{F}[A]$ is r_a, then the radius of convergence of $G_d[A]$, defined as $G_d[A](z) = \mathcal{F}[A](z^d)$ is, by Hadamard's criterion, equal to $r_a^{1/d}$, so all of $G_d[A]$ converge on the disk of radius $R_a = \min(r_a, 1)$ around the origin. Since the series on the right-hand side converges at 0 (since all the terms vanish), it converges uniformly on compact subsets of the disk of radius R_a around the origin. □

COROLLARY 14.5. *Let $\mathcal{H}$ be the generating function of the sequence $h_r = r\mathcal{CC}(r)$. Then*

$$\mathcal{H}(z) = 1 + \sum_{d=1}^{\infty} \phi(d)\mathcal{F}[\mathcal{C}](z^d).$$

We can combine all of the above results into the following conclusion.

THEOREM 14.6. *The generating function $\mathcal{H}$ as in the statement of Corollary 14.5 can be expanded as*:

$$\mathcal{H} = 1 + (k-1)\frac{x^2}{(1-x^2)^2} + \sum_{d=1}^{\infty} \phi(d)\left(\frac{1}{1-(2k-1)x^d} - 1\right).$$

In particular, $\mathcal{H}$ has an infinite number of poles, and is not a rational function for any $k > 1$. The generating function $\mathcal{F}[\mathcal{CC}_{F_k}]$ can be written as

$$\mathcal{F}[\mathcal{CC}_{F_k}](z) = \int_0^z \frac{\mathcal{H}(t)}{t}\, dt$$

and so is not a rational function either.

Proof. The expression for $\mathcal{H}$ is fairly obvious, with the comment that the second summand is a consequence of the fact that

$$\sum_{d|n} \phi(d) = n.$$

That $\mathcal{H}$ has an infinite number of poles follows from the observation that the dth term in the third summand has its d poles on the circle $|z| = (2k-1)^{-1/d}$, while the first two summands are analytic in the open unit disk. The expression for $\mathcal{F}[\mathcal{CC}_{F_k}]$ is immediate. □

REMARK 14.7. *For $k = 1$, it is not hard to see that*

$$\mathcal{H} = 1 + \frac{x}{(x-1)^2}.$$

REMARK. Various people, when shown Theorem 14.6, appeared to believe that it contradicts [12, Theorem 5.2D]. In fact (as pointed out by Greg McShane), Gromov's function $[N]_k$ is *not* (as the common misunderstanding has it) the same as $\mathcal{CC}_G(r)$ in the case of a free group, but *is* the same as $\mathcal{C}_G(r)$.

14.1. Some further comments. The following observation is quite obvious:

OBSERVATION 14.8. Let G_1 and G_2 be two groups. Then

$$\mathcal{F}[\mathcal{CC}_{G_1\times G_2}](z) = \mathcal{F}[\mathcal{CC}_{G_1}](z)\mathcal{F}[\mathcal{CC}_{G_2}](z).$$

It would be interesting to find other relationships (for example, what happens for HNN extensions)?

Observation 14.8 has some consequences.

THEOREM 14.9. *Let G_1 and G_2 be two groups, then if $\mathcal{F}[\mathcal{CC}_{G_1}]$ is rational, while $\mathcal{F}[\mathcal{CC}_{G_2}]$ is not, then $\mathcal{F}[\mathcal{CC}_{G_1\times G_2}]$ is not rational. If both $\mathcal{F}[\mathcal{CC}_{G_1}]$ and $\mathcal{F}[\mathcal{CC}_{G_2}]$ are rational, then so is $\mathcal{F}[\mathcal{CC}_{G_1\times G_2}]$.*

COROLLARY 14.10. *If $G_1 = \mathbf{Z}^n$ and G_2 is a finite group, then $\mathcal{F}[\mathcal{CC}_{G_1\times G_2}]$ is rational.*

REMARK. It is not clear whether $\mathcal{F}[\mathcal{CC}_G]$ is rational when G is a Bieberbach group—most likely this depends on the choice of the generating set, as conjectured by Epstein.

COROLLARY 14.11. *If $G_1 = F_k$ and G_2 is a direct product of finite groups and infinite cyclic groups, then $\mathcal{F}[\mathcal{CC}_{G_1\times G_2}]$ is irrational.*

THEOREM 14.12. *If $G = F_{k_1} \times F_{k_2} \times \cdots \times F_{k_n}$, then $\mathcal{F}[\mathcal{CC}_G]$ is irrational (with respect to the "obvious" generating set).*

Proof. This is an immediate consequence of Theorem 14.6. □

15. Primitive conjugacy class zeta function

One can compute a zeta-function analogous to that of Ihara for the numbers of *primitive* conjugacy classes of a given length (a primitive class is one which is not the power of a smaller class), using, essentially, the elementary method described by Stark and Terras, [59], as applied to the graph constructed in Section 1. This function turns out to be rational (in fact, there is a simple formula for it, see Theorem 15.1). More precisely, consider

$$\zeta(G)^{-1} = \prod_{[c]} \bigl(1 + u^{l}(c)\bigr), \tag{15.1}$$

where $[c]$ denotes the equivalences classes of primitive cycles, where two cycles are considered equivalent if one can be obtained from the other by a rotation.

A computation then shows that

$$\zeta(F_r) = (1-u^2)^{r-1}(1-u)\bigl(1-(2r-1)u\bigr). \tag{15.2}$$

The computation goes as follows: first, note that

$$\log \zeta(G) = \sum_{[c]} \sum_{i=1}^{\infty} \frac{1}{i} u^{il(c)},$$

and thus

$$u \frac{d \log \zeta(G)}{du} = \sum_{[c]} \sum_{i=1}^{\infty} l(c) u^{il(c)}.$$

The above can be rewritten (note that the sum is now over primitive cycles, and not equivalence classes thereof):

$$u \frac{d \log \zeta(G)}{du} = \sum_{c} \sum_{i=1}^{\infty} u^{il(c)}.$$

But note that the right-hand side is simply the ordinary generating function for *all* cycles:

$$u \frac{d \log \zeta(G)}{du} = \sum_{i=1}^{\infty} N_i u^i, \tag{15.3}$$

where N_i is the number of cycles of length i in G, and this generating function was computed in Section 1:

$$\sum_{i=i}^{\infty} N_i u^i = \frac{1}{1+(2r-1)u} + \frac{r}{1-u} + \frac{r-1}{1+u}.$$

The formula (15.2) now follows by a straightforward integration.

An quick examination of the above argument shows that the formula (15.2) is a special case of the following result.

THEOREM 15.1. *Let G be a finite graph, and let ζ_G be the zeta-function defined by formula (15.1). Let $A(G)$ be the adjacency matrix of G. Then*

$$\zeta_G(u) = \det\bigl(I - uA(G)\bigr). \tag{15.4}$$

In other words, the zeta-function is essentially the characteristic polynomial of $A(G)$.

Proof. The argument above up to equation (15.3) is completely general. On the other hand, the right-hand side of equation (15.3) can be rewritten as:

$$\begin{aligned}
\sum_{i=1}^{\infty} N_i u^i &= \sum_{i=1}^{\infty} \operatorname{tr} A(G)^i u^i \\
&= \operatorname{tr}\left[-I + \sum_{i=0}^{\infty} [A(G)u]^i\right] \\
&= \operatorname{tr}\bigl[-I + \bigl(I - uA(G)\bigr)^{-1}\bigr] \\
&= \operatorname{tr}\bigl(uA(G)\bigl(I - uA(G)\bigr)^{-1}\bigr).
\end{aligned}$$

Thus,

$$\frac{d\log\zeta(G)}{du} = \operatorname{tr}\bigl(A(G)\bigl(I - uA(G)\bigr)^{-1}\bigr),$$

and so it follows that

$$\zeta(G) = C\det\bigl(I - uA(G)\bigr),$$

where C is a constant of integration, seen to be equal to 1 by computing both sides at $u = 0$. □

Acknowledgments. I would like to thank those who had comments on earlier versions of this paper (released as an IHES preprint in September 1997). The irrationality of the growth function of the number of conjugacy classes was independently shown by D. B. A. Epstein and Murray Macbeath. I would like to thank the anonymous referees for their comments on the current version.

References

[1] T. Adachi and T. Sunada, *Homology of closed geodesics in a negatively curved manifold*, J. Differential Geom. **26** (1987), 81–99. MR 0892032

[2] H. Bass, *The Ihara–Selberg zeta function of a tree lattice*, Internat. J. Math. **3** (1992), 717–797. MR 1194071

[3] A. F. Beardon, J. Lehner and M. Sheingorn, *Closed geodesics on a Riemann surface with application to the Markov spectrum*, Trans. Amer. Math. Soc. **295** (1986), 635–647. MR 0833700

[4] G. Besson, G. Courtois and S. Gallot, *Entropies et rigidités des espaces localement symétriques de courbure strictement négative*, Geom. Funct. Anal. **5** (1995), 731–799. MR 1354289

[5] R. Bowen, *Symbolic dynamics for hyperbolic flows*, Amer. J. Math. **95** (1973), 429–460. MR 0339281

[6] D. Calegari and K. Fujiwara, *Combable functions, quasimorphisms, and the central limit theorem*, Ergodic Theory Dynam. Systems **30** (2010), 1343–1369. MR 2718897

[7] M. Coornaert, *Asymptotic growth of conjugacy classes in finitely-generated free groups*, Internat. J. Algebra Comput. **15** (2005), 887–892. MR 2197812

[8] C. L. Epstein, *Asymptotics for closed geodesics in a homology class, the finite volume case*, Duke Math. J. **55** (1987), 717–757. MR 0916117

[9] D. B. A. Epstein, J. W. Cannon, D. F. Holt, S. V. F. Levy, M. S. Paterson and W. P. Thurston, *Word processing in groups*, Jones and Bartlett Publishers, Boston, MA, 1992. MR 1161694

[10] W. Feller, *An introduction to probability theory and its applications. Vol. II*, 2nd ed., Wiley, New York, 1971.

[11] W. Fulton and J. Harris, *Representation theory*, Graduate Texts in Mathematics, vol. 129, Springer, New York, 1991. MR 1153249

[12] M. Gromov, *Hyperbolic groups*, Essays in group theory, Math. Sci. Res. Inst. Publ., vol. 8, Springer, New York, 1987, pp. 75–263. MR 0919829

[13] M. Gromov, *Volume and bounded cohomology*, Inst. Hautes Études Sci. Publ. Math. **56** (1983), 5–99. 1982. MR 0686042

[14] Y. Guivarc'h and Y. Le Jan, *Asymptotic winding of the geodesic flow on modular surfaces and continued fractions*, Ann. Sci. École Norm. Sup. (4) **26** (1993), 23–50. MR 1209912

[15] Y. Guivarch and Y. Le Jan, *Note rectificative: "Asymptotic winding of the geodesic flow on modular surfaces and continued fractions"* [Ann. Sci. École Norm. Sup. (4) **26** (1993), 23–50]. Ann. Sci. École Norm. Sup. (4) **29** (1996), 811–814. MR 1209912
[16] G. H. Hardy and E. M. Wright, *An introduction to the theory of numbers*, 5th ed., The Clarendon Press, Oxford University Press, New York, 1979. MR 0568909
[17] M. Horsham and R. Sharp, *Lengths, quasi-morphisms and statistics for free groups*, Spectral analysis in geometry and number theory, Contemp. Math., vol. 484, Amer. Math. Soc., Providence, RI, 2009, pp. 219–237. MR 1500150
[18] H. Huber, *Zur analytischen theorie hyperbolischen raumformen und bewegungsgruppen*, Math. Ann. **138** (1959), 1–26. MR 0109212
[19] H. Huber, *Zur analytischen theorie hyperbolischer raumformen und bewegungsgruppen. II*, Math. Ann. **142** (1960/1961), 385–398. MR 0126549
[20] H. Huber, *Zur analytischen theorie hyperbolischer raumformen und bewegungsgruppen. II*, Math. Ann. **143** (1961), 463–464. MR 0154980
[21] Y. Ihara, *On discrete subgroups of the two by two projective linear group over p-adic fields*, J. Math. Soc. Japan **18** (1966), 219–235. MR 0223463
[22] I. Kapovich and T. Nagnibeda, *The Patterson–Sullivan embedding and minimal volume entropy for outer space*, Geom. Funct. Anal. **17** (2007), 1201–1236. MR 2373015
[23] I. Kapovich and I. Rivin, *On the absence of McShane-type identities for the outer space*, J. Algebra **320** (2008), 3659–3670. MR 2457714
[24] I. Kapovich, I. Rivin, P. Schupp and V. Shpilrain, *Densities in free groups and $\mathbb{Z}^k$, visible points and test elements*, Math. Res. Lett. **14** (2007), 263–284. MR 2318624
[25] I. Kapovich and P. Schupp, *On group-theoretic models of randomness and genericity*, Groups Geom. Dyn. **2** (2008), 383–404. MR 2415305
[26] T. Kato, *Perturbation theory for linear operators*, Classics in Mathematics, Springer, Berlin, 1995. MR 1335452
[27] A. Katok, *Entropy and closed geodesics*, Ergodic Theory Dynam. Systems **2** (1983), 339–365. MR 0721728
[28] A. Katsuda and T. Sunada, *Homology of closed geodesics in certain Riemannian manifolds*, Proc. Amer. Math. Soc. **96** (1986), 657–660. MR 0826498
[29] A. Katsuda and T. Sunada, *Homology and closed geodesics in a compact Riemann surface*, Amer. J. Math. **110** (1988), 145–155. MR 0926741
[30] A. Katsuda and T. Sunada, *Closed orbits in homology classes*, Inst. Hautes Études Sci. Publ. Math. **71** (1990), 5–32. MR 1079641
[31] L. M. Koganov, *The number of cyclically irreducible words in the alphabet of a free group of finite rank*, Kibernet. Sistem. Anal. **43** (2007), 39–48, 189. MR 2374427
[32] E. Kowalski, *The large sieve and its applications*, Cambridge Tracts in Mathematics, vol. 175, Cambridge Univ. Press, Cambridge, 2008. MR 2426239
[33] S. P. Lalley, *Closed geodesics in homology classes on surfaces of variable negative curvature*, Duke Math. J. **58** (1989), 795–821. MR 1016446
[34] S. Lim, *Minimal volume entropy for graphs*, Trans. Amer. Math. Soc. **360** (2008), 5089–5100. MR 2415065
[35] G. A. Margulis, *Certain applications of ergodic theory to the investigation of manifolds of negative curvature*, Funkcional. Anal. i Priložen. **3** (1969), 89–90. MR 0257933
[36] G. A. Margulis, *On some aspects of the theory of Anosov systems*, Springer Monographs in Mathematics, Springer, Berlin, 2004. MR 2035655
[37] D. K. Maslen, *Efficient computation of Fourier transforms on compact groups*, J. Fourier Anal. Appl. **4** (1998), 19–52. MR 1650948
[38] G. McShane and I. Rivin, *Geometry of geodesics and a norm on homology*, Int. Math. Res. Not. **2** (1995), 61–69. MR 1317643
[39] G. McShane and I. Rivin, *Simple curves on hyperbolic tori*, Comptes Rendus Acad. Sci. Paris Sér. I. Math. **320** (1995), 1523–1528. MR 1340065

[40] M. Mirzakhani, *Growth of the number of simple closed geodesics on hyperbolic surfaces*, Ann. of Math. (2) **168** (2008), 97–125. MR 2415399

[41] W. Parry and M. Pollicott, *The Chebotarov theorem for Galois coverings of axiom a flows*, Ergodic Theory Dynam. Systems **6** (1986), 133–148. MR 0837980

[42] Y. N. Petridis and M. S. Risager, *Discrete logarithms in free groups*, Proc. Amer. Math. Soc. **134** (2006), 1003–1012 (electronic). MR 2196031

[43] Y. N. Petridis and M. S. Risager, *Equidistribution of geodesics on homology classes and analogues for free groups*, Forum Math. **20** (2008), 783–815. MR 2445118

[44] R. Phillips and P. Sarnak, *Geodesics in homology classes*, Duke Math. J. **55** (1987), 287–297. MR 0894581

[45] I. Rivin, *Growth in free groups (and other stories)*, preprint, available at `arXiv:math/9911076v2`, 1999.

[46] I. Rivin, *Simple curves on surfaces*, Geometriae Dedicata **87** (2001), 345–360. MR 1866856

[47] I. Rivin, *Some properties of the conjugacy class growth function*, Group theory, statistics, and cryptography, Contempory Mathematics, vol. 360, Amer. Math. Soc., Providence, RI, 2004, pp. 113–117. MR 2105439

[48] I. Rivin, *Symmetrized Chebyshev polynomials*, Proc. Amer. Math. Soc. **133** (2005), 1299–1305 (electronic). MR 2111935

[49] I. Rivin, *A simpler proof of Mirzakhani's simple curve asymptotics*, Geom. Dedicata **114** (2005), 229–235. MR 2174101

[50] I. Rivin, *Walks on groups, counting reducible matrices, polynomials, and surface and free group automorphisms*, Duke Math. J. **142** (2008), 353–379. MR 2401624

[51] I. Rivin, *Walks on graphs and lattices—effective bounds and applications*, Forum Math. **21** (2009), 673–685. MR 2541479

[52] I. Rivin, *Zariski density and genericity*, Int. Math. Res. Not. **19** (2010), 3649–3657. MR 2725508

[53] T. J. Rivlin, *Chebyshev polynomials*, 2nd ed., Pure and Applied Mathematics (New York). Wiley, New York, 1990. MR 1060735

[54] M. Rubinstein and P. Sarnak, *Chebyshev's bias*, Experiment. Math. **3** (1994), 173–197. MR 1329368

[55] I. Schur, *Gesammelte Abhandlungen. Band III*, Springer, Berlin, 1973, pp. 422–453.

[56] C. E. Shannon, *A mathematical theory of communication*, Bell System Tech. J. **27** (1948), 379–423, 623–656. MR 0026286

[57] R. Sharp, *Local limit theorems for free groups*, Math. Ann. **321** (2001), 889–904. MR 1872533

[58] A. N. Shiryaev, *Probability*, 2nd ed., Graduate Texts in Mathematics, vol. 95, Springer, New York, 1996. MR 1368405

[59] H. M. Stark and A. A. Terras, *Zeta functions of finite graphs and coverings*, Adv. Math. **121** (1996), 124–165. MR 1399606

[60] H. M. Stark and A. A. Terras, *Zeta functions of finite graphs and coverings. II*, Adv. Math. **154** (2000), 132–195. MR 1780097

[61] A. A. Terras and H. M. Stark, *Zeta functions of finite graphs and coverings. III*, Adv. Math. **208** (2007), 467–489. MR 2304325

[62] I. Vardi, *Dedekind sums have a limiting distribution*, Int. Math. Res. Not. **1** (1993), 1–12. MR 1201746

[63] A. Weil, *L'intégration dans les groupes topologiques et ses applications*, Actual. Sci. Ind., vol. 869, Hermann et Cie., Paris, 1940. MR 0005741

Igor Rivin, Mathematics Department, Temple University, Philadelphia, PA 19122, USA and School of Mathematics, Institute for Advanced Study, Princeton, NJ 08540, USA

E-mail address: rivin@temple.edu; rivin@ias.edu