Growth Functions and Automatic Groups

David B. A. Epstein, Anthony R. Iano-Fletcher, and Uri Zwick

CONTENTS

In this paper we study growth functions of automatic and hyperbolic groups. In addition to standard growth functions, we also want to count the number of finite graphs isomorphic to a given finite graph in the ball of radius n around the identity element in the Cayley graph. This topic was introduced to us by K. Saito [1991]. We report on fast methods to compute the growth function once we know the automatic structure. We prove that for a geodesic automatic structure, the growth function for any fixed finite connected graph is a rational function. For a word-hyperbolic group, we show that one can choose the denominator of the rational function independently of the finite graph.

1. GROWTH FUNCTIONS OF GROUPS

Let G be a group with a finite set of generators. Then there is a finite set Σ of generators such that, if $x \in \Sigma$, then x^{-1} is also in Σ. More formally, Σ is not necessarily a subset of G; instead there is a map $\pi : \Sigma \to G$ and an involution $\iota : \Sigma \to \Sigma$ such that $\pi\iota(x) = (\pi(x))^{-1}$. Moreover, we do not assume that $\pi : \Sigma \to G$ is injective. Let Σ^* be the set of strings over Σ—this can also be thought of as the free monoid generated by Σ, with multiplication given by concatenation. We use the symbol $\pi : \Sigma^* \to G$ to denote the homomorphism already given on $\Sigma \subset \Sigma^*$. Another convenient notation for $\pi(w)$, where $w \in \Sigma^*$, is $\overline{w}$. Where we think there will be no confusion, we sometimes ignore these special notations and denote an element of G by an element of Σ^* that represents it.

Let $b_n = b_n(G, \Sigma)$ be the number of elements of G that can be expressed in terms of words of length at most n in the generating set Σ. Then we can form the formal power series $B(z) = B_{(G,\Sigma)}(z) = \sum b_n z^n$. We call B the *growth series* of (G, Σ). It is well-known that, for any pair (G, Σ), this is

1058-6458/96 $0.50 per page

the power series of a holomorphic function in some neighbourhood of $z = 0$ (see Lemma 1.2), and for this reason B is also known as the *growth function* of (G, Σ). For many interesting examples, the growth function turns out to be a rational function of z with integral coefficients. See [Cannon and Wagreich 1992], where additional references to the origins of this subject can be found.

We will use another, closely related, version of the growth series. Let $c_n = c_n(G, \Sigma)$ be the number of elements of G whose shortest representative in Σ^* has exactly length n. Then we form the formal power series

$$C(z) = C_{(G,\Sigma)}(z) = \sum c_n z^n.$$

We clearly have $B(z)(1 - z) = C(z)$. So studying one of these functions is equivalent to studying the other.

It is often useful to have a geometric view of a group. If G is finitely generated, we can get a geometric view via the *Cayley graph*. The vertices of this graph are the elements of G, and there is a directed edge (g_1, x, g_2) from the vertex $g_1 \in G$ to the vertex $g_2 \in G$ if and only if $x \in \Sigma$ and $g_1 x = g_2$ in G. We refer to x as the *label* on the edge. We denote the Cayley graph by $\Gamma(G, \Sigma)$. Notice that the Cayley graph has an action by G on the left that leaves the label x on a directed edge unchanged. We metrize $\Gamma(G, \Sigma)$ in the obvious way, making the length of each edge equal to one.

The convention is often observed that if a generator x has order two, then the corresponding edge is not directed. Note that this convention is not followed in our work, as it would lead to a less uniform treatment.

Notice that b_n is the number of vertices in the ball of radius n in the Cayley graph centred at the identity element. (Because of the left action of G, a ball of radius n centred at any vertex of Γ is isomorphic to the ball centred at any other vertex.) Also c_n is the number of vertices whose distance from the identity vertex is exactly n.

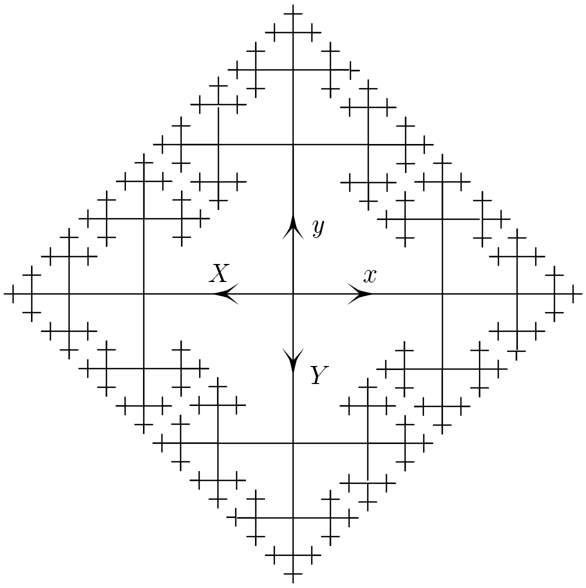

FIGURE 1. Part of the Cayley graph of the free group on two group generators x and y, with inverses X and Y.

Example 1.1 (free group). Let F_n be the free group on n generators. Then Σ, the generating set, has $2n$ elements. We have $c_0 = 1$, $c_1 = 2n$ and $c_i = (2n - 1)c_{i-1}$ if $i \geq 2$ (see Figure 1). It is easy to deduce from this recurrence relation that

$$C(z) = \frac{1 + z}{1 - (2n - 1)z}.$$

Lemma 1.2 (holomorphic growth). *Every finitely generated group with n generators has a growth series whose radius of convergence is at least $1/(2n - 1)$.*

Proof. Let G be our group, and let Σ_+ be the given set of generators. Let Σ_- denote the set of formal inverses of the elements of Σ_+ and let $\Sigma = \Sigma_+ \cup \Sigma_-$. Then Σ has an obvious involution interchanging the elements of Σ_+ and Σ_-. Let F_n be the free group generated by Σ_+.

There is an obvious map $\Sigma \to F_n$ and a surjective homomorphism $F_n \to G$. Any element of G of length i over Σ is the image of an element of F_n of length i. It follows that $0 \leq c_i(G, \Sigma) \leq c_i(F_n, \Sigma)$ for each $i \geq 0$. Therefore the radius of convergence of the growth series for G is no less than that of the growth series for the free group computed in Example 1.1. □

The next result is also well-known.

Theorem 1.3. *Let G be a group with a finite set Σ of generators and a finite (or recursively enumerable) set of relators. The function $n \mapsto c_n(G, \Sigma)$ is computable (by a Turing machine) if and only if G has a solvable word problem.*

Proof. If G has a solvable word problem, then one can test each word w of length n, to check whether another word v whose length is at most n represents the same element of G as w. Therefore c_n is a computable function of n.

Conversely, if c_n is computable and one is given a word w of length n, one can check whether or not w represents the trivial element of G as follows. One systematically checks all products of conjugates of relators until one has found all identities between words of length at most n. If one carries on long enough, all of these will be found. Moreover, since we know $c_0, c_1, \dots c_n$, we know when we can stop the computation. □

2. GROWTH FUNCTION OF AN AUTOMATON

As already remarked, many groups have rational growth functions. Many of these groups can be described in terms of finite state automata; we will see in this section how finite state automata always have rational growth functions. This is a well-known result, included here for the convenience of the reader.

First we recall the definition of a *finite state automaton*. A finite state automaton is a quintuple (S, Σ, s_0, Y, μ), where S is a finite set called the *state set*, Σ is a finite set called the *alphabet*, $s_0 \in S$ is called the *initial state*, $Y \subset S$ is called the set of *accept states*, and $\mu : S \times \Sigma \to S$ is a function called the *transition function*. We denote the finite state automaton by M.

M can be regarded as a labelled directed graph with an edge labelled x from $s \in S$ to $t \in S$ if $\mu(s, x) = t$. We define the *language accepted by* M, denoted by $L(M)$, as a certain set of strings over Σ. A string $w \in \Sigma^*$ is in $L(M)$ (we say that w is *accepted* by M) if, when we follow edges according to their labels so as to trace out w starting at s_0, we end at some element of Y.

Let n be the number of states in M. We number the states from 1 to n, starting with s_0. Let A be the *transition matrix* of M, that is, the $n \times n$-matrix such that a_{ij}, the entry of A in the i-th row and j-th column, is the number of directed edges from the j-th state to the i-th state. Then A^r is the matrix whose (i, j)-entry is the number of edge-paths of length exactly r from the j-th state to the i-th state.

Let $u \in \mathbb{Z}^n$ be the column vector with $u^t = (1, 0, \dots, 0)$, that is, the characteristic function of s_0, and let $v \in \mathbb{Z}^n$ be the characteristic function of Y. Then $c_i(M) = v^t A^i u$ is the number of strings of length i accepted by M, and

$$C(M, z) = \sum_i c_i z^i = v^t \sum_i (zA)^i u = v^t (I - zA)^{-1} u.$$

It follows from the formula for the inverse of a matrix that $C(M, z)$ is a rational function of z. We have proved the following well-known result.

Theorem 2.1 (automaton growth). *Let M be a finite state automaton, with transition matrix A. Then $L(M)$ has a growth function of the form*

$$\frac{P(z)}{\det(I - zA)},$$

where P is a polynomial with integer coefficients.

3. COMPUTING GROWTH FUNCTIONS

In this section we show how to compute growth functions efficiently. In the previous section we were considering a finite state automaton. However, it is clear that the labels on the arrows were irrelevant; we therefore ignore the labels in this section.

Let $\Gamma(V, E)$ be a directed graph, where $V = \{1, \dots, n\}$. Let $X \subseteq V$ be the set of *initial* vertices and let $Y \subseteq V$ be the set of *terminal* vertices.

Let A be the adjacency matrix of Γ, that is A is an $n \times n$ matrix whose element in the i-th row and j-th column is the number of directed edges from vertex j to vertex i in Γ. Let $x, y \in \{0,1\}^n$ be the characteristic vectors of X and Y; thus $x_i = 1$ if and only if $i \in X$.

We are interested in counting the number of directed paths in Γ of length exactly k that begin at some vertex of X and end at some vertex of Y. We denote this number by c_k. It is easy to see that $c_k = y^t A^k x$. As we saw in the previous section, the generating function of the sequence $\{c_k\}$, called the *growth function* of Γ, is

$$\begin{aligned} C(z) &= \sum_{k\geq 0} c_k z^k = y^t \Bigl[\sum_{k\geq 0} (zA)^k\Bigr] x \\ &= y^t (I - zA)^{-1} x = \frac{P(z)}{Q(z)}, \end{aligned}$$

where $Q(z) = \det(I - zA)$ and $P(z)$ is some polynomial with $\deg(P) < n$, as follows from Cramer's rule. Notice that

$$\det(I - zA) = (-z)^n \det(A - z^{-1} I),$$

so $Q(z)$ can be obtained from the characteristic polynomial of the matrix A by reversing the order of the coefficients and possibly negating them. In some cases $P(z)$ and $Q(z)$ will have common factors, which could then be cancelled out. In fact, we will see in Lemma 3.1 that the reduced $Q(z)$ divides the reversed minimum polynomial of A.

We now describe computationally efficient ways for explicitly obtaining the growth function of a directed graph Γ with associated matrix A and characteristic vectors x and y. We are interested in the exact (integral) coefficients of $P(z)$ and $Q(z)$, not in mere approximations for them.

If the matrix A is dense we can begin by computing the characteristic polynomial of A. This can be done using $O(n^3)$ arithmetical operations using a classical method attributed to Danilevski in [Faddeev and Faddeeva 1963]. Keller–Gehrig [1985] has shown that it can also be done using only $O(M(n))$ arithmetical operations, where $M(n)$ is the number of arithmetical operations required for multiplying two $n \times n$ matrices. The best upper bound on $M(n)$ is currently $O(n^{2.376})$ [Coppersmith and Winograd 1990].

We are especially interested in cases in which the matrix A is sparse. We now describe a method for obtaining the reduced $P(z)$ and $Q(z)$ polynomials using only $O(|E|\,|V|) \leq O(dn^2)$ operations where d is the maximal out-degree in Γ. The method used is similar to the method used, over finite fields, by Wiedemann [1986].

The key observation is that the sequence $\{c_k\}$ satisfies the linear recurrence relation specified by the coefficients of $Q(z)$. If $P(z) = \sum_{i=0}^{m-1} p_i z^i$ and $Q(z) = \sum_{i=0}^{m} q_i z^i$, where $m \leq n$, then by extracting the coefficient of z^k in the relation $Q(z)C(z) = P(z)$ we find

$$\sum_{i=0}^{\min\{k,m\}} q_i c_{k-i} = \begin{cases} p_k & \text{if } k < m, \\ 0 & \text{otherwise.} \end{cases}$$

As a consequence we have

$$c_k = -\frac{1}{q_0} \sum_{i=1}^{m} q_i c_{k-i} \qquad \text{for } k \geq m. \tag{3.1}$$

In our case $Q(z)$ is equal to (a factor of) $\det(I - zA)$, so $q_0 = 1$.

Conversely, if we are given the sequence $\{c_k\}$ and the coefficients $q_0, \dots, q_m$ of the recurrence relation 3.1, then we can rapidly compute $p_0, \dots, p_{m-1}$. If we then let $P(z) = \sum_{i=0}^{m-1} p_i z^i$ and $Q(z) = \sum_{i=0}^{m} q_i z^i$ we immediately get $C(z) = P(z)/Q(z)$.

As an aside, we note that, after cancelling common factors with $P(z)$, $Q(z)$ divides the reversed minimum polynomial.

Lemma 3.1 (minimum polynomial). *Let $S(z) = s_p + \dots + s_1 z^{p-1} + s_0 z^p$ be the minimum polynomial of A (so that $s_0 = 1$). Then the reduced $Q(z)$ divides the reversed minimum polynomial $R(z) = s_0 + \dots + s_p z^p$.*

Proof. We have $S(A)A^k = 0$ for each $k \geq 0$. Hence $y^t S(A) A^k x = 0$ and so $\sum_{i=0}^{p} s_i c_{r-i} = 0$ for $r \geq p$. We define $P_1(z) = C(z)R(z)$ and note that P_1 has degree less than p. We have $P_1 Q = CQR = PR$. Since P and Q have no common factors, Q must divide R as claimed. □

Our problem can therefore be solved by finding a recurrence relation of length at most n satisfied by the sequence $\{c_k\}$. In fact, finding the shortest such recurrence will give us the reduced denominator of the growth function. But, how do we find such a recurrence satisfied by the (infinite) sequence $\{c_n\}$? Since we already know that the sequence $\{c_k\}$ satisfies at least one recurrence relation of length at most n, the following simple lemma says that it will be enough to find a recurrence relation satisfied by the first $2n$ elements of the sequence. The recurrence relation is then guaranteed to be satisfied by all subsequent values.

Lemma 3.2. *Let $n > 0$ and let P and P' be polynomials of degree at most $n-1$ and Q and Q' polynomials of degree at most n. Suppose that, when expanded as power series, we have*

$$\frac{P(z)}{Q(z)} \equiv \frac{P'(z)}{Q'(z)} \pmod{x^{2n}}$$

then

$$\frac{P(z)}{Q(z)} = \frac{P'(z)}{Q'(z)}.$$

Proof. Since

$$P(z)Q'(z) - P'(z)Q(z) \equiv 0 \pmod{x^{2n}},$$

the claim follows from the bounds on the degrees. □

We are now able to present our algorithm. The input is the $n \times n$ matrix A and the vectors x and y. Each column of A has at most d non-zero entries. The algorithm is composed of three stages:

1. Compute $c_0, c_1, \ldots, c_{2n-1}$, the first $2n$ elements in the growth sequence.
2. Find the minimal-length recurrence
$$\sum_{i=0}^{m} q_i c_{k-i} = 0$$
satisfied by the sequence $c_0, c_1, \ldots, c_{2n-1}$ and construct $Q(z)$.
3. Compute $p_0, \ldots, p_{m-1}$ and construct the polynomial $P(z)$.

Stage 1 can be implemented naively as follows:

$$\begin{aligned} &v_0 \leftarrow x; && c_0 \leftarrow y^t x; \\ &v_k \leftarrow A v_{k-1}; && c_k \leftarrow y^t v_k \quad \text{for } k = 1, \ldots, 2n-1. \end{aligned}$$

The time complexity of computing Av is $O(|E|) \le O(dn)$, so the overall complexity of this stage is $O(|E|\,|V|) \le O(dn^2)$. This naive approach should be used when d is relatively small compared to n. If d is of the order of n then the complexity of this stage will be $O(n^3)$. This can be reduced (see [Keller-Gehrig 1985], for example) to $O(M(n) \log n)$ using fast matrix multiplications. Although stage 1 seems to be more straightforward than the following stage 2, it turns out to be the dominant stage in terms of the computational complexity of the problem.

The problem we have to solve in stage 2 could be solved using the Berlekamp–Massey algorithm [Massey 1969]. This algorithm was initially proposed by Berlekamp as a decoding algorithm for BCH codes. It was later observed by Massey that Berlekamp's algorithm solves the general *shift register synthesis* problem, which is identical to the problem of finding linear recurrences of minimal length. The Berlekamp–Massey algorithm was designed primarily to work over finite fields, but it can be used over any field. We will use the rationals as the underlying field. In fact, as we may choose $q_0 = 1$, no divisions will be required and all the intermediate results will be integral. The complexity of the Berlekamp–Massey algorithm is $O(n^2)$ and it is fairly simple and easy to program.

Stage 2 can be performed even more efficiently although this does not change the overall $O(dn^2)$ complexity of our algorithm. It is fairly easy to see that the q_i's can be found by finding a non-trivial solution to a homogeneous Toeplitz system of linear equations with the Toeplitz coefficient matrix being simply composed of the elements c_0, c_1, …,

c_{2n-1}. This can be solved in time $O(n \log^2 n)$ using an algorithm by Brent, Gustavson and Yun [Brent et al. 1980]. This algorithm uses a version of the Extended Euclidean Algorithm.

Stage 3 can now be naively performed in $O(n^2)$ operations. The overall complexity is therefore $O(|E||V|) \le O(dn^2)$ integer operations, as promised.

So far we have considered arithmetic operations as basic operations. Note however that in some of the computations huge numbers may be obtained, even if all the coefficients of the polynomials $P(z)$ and $Q(z)$ turn out to be quite small.

As an example, consider the following (randomly chosen) matrix

$$A = \begin{pmatrix} 5 & 7 & 5 & 5 \\ 2 & 8 & 8 & 0 \\ 8 & 5 & 3 & 8 \\ 1 & 2 & 4 & 1 \end{pmatrix}$$

together with the start and stop vectors $x = y = (1,1,1,1)$. The growth function in this case is

$$\begin{aligned} C(z) &= \frac{4 - 4z - 20z^2}{1 - 19z + 2z^2 + 78z^3} \\ &= 4 + 72\,z + 1340\,z^2 + 25004\,z^3 + 466780\,z^4 \\ &\quad + 8714292\,z^5 + 162687676\,z^6 \\ &\quad + 3037228420\,z^7 + \cdots \end{aligned}$$

We see that the coefficients of C can get large. In general these coefficients grow exponentially, and can easily become too large to fit into a computer word in practical problems. On the other hand, in the above case, the characteristic polynomial of A is $156 + 82\,z - 36\,z^2 - 17\,z^3 + z^4$, with all coefficients small.

The usual solution to the blow up of the intermediate results is to work out the solution modulo several, moderate-size, prime numbers, and then combine the solutions obtained to the desired solution over the integers. In practice, each of prime numbers used would fit into one computer word and no multi-precision calculations will be needed (assuming that the final coefficients fit into single computer words).

For each prime number p chosen, the first $2n$ elements in the sequence $\{c_n \bmod p\}$ will be computed, and then two reduced polynomials $P_p(z)$ and $Q_p(z)$ satisfying

$$P_p(z)/Q_p(z) \equiv \textstyle\sum_{i=0}^{2n-1} c_k z^k \pmod p$$

will be found. Unless p is an *unlucky* prime, we would have $P_p(z) \equiv P(z) \pmod p$ and $Q_p(z) \equiv Q(z) \pmod p$. Recall that $P(z)$ and $Q(z)$ are relatively prime over the integers. A prime p is unlucky if and only if $P(z)$ and $Q(z)$ are not relatively prime modulo p. A similar concept of unlucky primes occurs in homomorphic algorithms for the computation of the greatest common divisor of two polynomials over the integers. For more details, see [Lauer 1983]. We just note here that in each specific case the number of unlucky primes is finite (and usually quite small) and that the probability of a prime chosen from a suitably large interval being unlucky is very small. Unlucky primes can be spotted by noticing that the degrees of the polynomials obtained modulo them is smaller than the degrees obtained modulo other (lucky) primes.

If we do the computations therefore modulo distinct primes $p_1, \ldots, p_k$, and if they all turn out to be lucky, we can then construct $P(z)$ and $Q(z)$ modulo the product $p_1 \cdots p_k$, using the Chinese Remainder Theorem. If we have a bound M on the largest coefficient (in absolute value) in $P(z)$ and $Q(z)$ then, using a set of primes whose product is greater than say, $2M+1$, we can reconstruct $P(z)$ and $Q(z)$. A bound M on the coefficients of $P(z)$ and $Q(z)$ may be obtained using Hadamard's inequality [Mignotte 1983]. Alternatively, we may take M to be an upper bound on the absolute values of $c_0, \ldots, c_{2n-1}$. In general, M may be exponentially large. Even so, only the first cn primes, where c is roughly proportional to the logarithm of the largest coefficient in A, have to be used. Note that an upper bound for this largest coefficient can be computed using floating point arithmetic (unless A is too large to handle in any case), since it does not need to be known accurately.

Usually, the coefficients of $P(z)$ and $Q(z)$ will turn out to be much smaller than the coarse bound M obtained for them. This can be exploited using the following randomized version of the preceding algorithm. The algorithm picks primes at random (from some interval). After performing the computations modulo the k-th prime, we calculate the coefficients of $P(z)$ and $Q(z)$ modulo the product of the k primes. If these coefficients are much smaller in absolute value than the product of the primes used, then with large probability the coefficients found are the correct ones. In practice, the use of only a constant number of primes (chosen from a suitable precomputed table) will yield the correct result with overwhelming probability.

The algorithms described here were programmed and used to compute growth functions that helped in the development of this article.

4. COUNTING THE NUMBER OF COPIES OF A FINITE SUBGRAPH

K. Saito [1991] has drawn attention to a modified collection of growth functions related to finite graphs. Saito's work arose from attempts to generalize the Ising model in quantum mechanics to cover more general geometries, such as the universal cover of a surface of higher genus tesselated by fundamental domains. The results in this section are due to Saito. We publish proofs here of the special cases that interest us, in order to give the reader a self-contained account.

Let S and T be two directed graphs with edges labelled by elements of Σ. A *morphism* $f : S \to T$ is defined to be a function that maps each vertex to a vertex and each labelled directed edge to a directed edge with the same label. Moreover the initial and final endpoints of a directed edge in S are required to map to the initial and final endpoints respectively of the image directed edge in T.

As in Section 1, let Σ denote a finite set of generators for a group G and let $\Gamma = \Gamma(G, \Sigma)$ be the corresponding Cayley graph. Let Γ_n be the ball of radius n and centre the identity element of G (or, equivalently, any other choice of centre). We add to Γ_n any edges that connect vertices both of which are already in Γ_n, that is, we turn Γ_n into a full subgraph (sometimes called an induced subgraph) of Γ.

Let S be a finite, connected, directed labelled graph, with labels from Σ. We define $b_n(S, G, \Sigma) = b_n(S)$ to be the number of morphisms $f : S \to \Gamma_n$. Correspondingly we have the formal power series

$$B(S,z) = B_{(G,\Sigma)}(S,z) = \sum_{i=0}^{\infty} b_i(S) z^i.$$

It is of course easy to construct an S so that there are no morphisms from S to Γ. In that case $B(S,z)$ is identically zero. Note that if S has only one vertex, then $B(S,z) = B(z)$, the standard growth function of a group.

Theorem 4.1 (Saito). *Let G be a group and let Σ be a finite set of generators with an involution corresponding to taking the inverse in G. Let S be a finite connected directed graph with edges labelled by the elements of Σ and suppose there is a morphism of S to the Cayley graph $\Gamma(G, \Sigma)$. Then $B(S,z)$ is holomorphic in a neighbourhood of* 0, *and the radius of convergence r is the same as that of $B(z)$. If both $B(z)$ and $B(S,z)$ are meromorphic in a neighbourhood of r, then the order of the pole or zero is the same for the two functions.*

Proof. Let d be an integer greater than the diameter of S. (The diameter of S can be a half unit or one unit greater than the diameter of its 0-skeleton. Consider for example the Cayley graph of the trivial group with either one or two generators.) Fix a vertex $s_0 \in S$ to serve as a basepoint. Then a morphism $S \to \Gamma$ is determined by the image of s_0. It follows that $b_{n-d} \le b_n(S) \le b_n$, and therefore $B(z)$ and $B(S,z)$ have the same radius of convergence r. By Lemma 1.2, we have $r > 0$. Also $z^d B(z) \le B(S,z) \le B(z)$ for $z \in [0, r)$. Taking logarithms, this gives us

$$d \log z + \log B(z) \le \log B(S,z) \le \log B(z)$$

for $z \in (0, r)$. The description when both functions are meromorphic near $z = r$ easily follows by choosing $k \in \mathbb{Z}$ so that $B(z)/(z-r)^k$ is regular and non-zero near $z = r$, and similarly for $B(S, z)$. $\square$

Saito's theorem clearly implies the following corollary.

Corollary 4.2. *If $B(z)$ is a rational function such that r is the nearest pole to the origin and if $B(S, z)$ is also a rational function, then r is also the nearest pole to the origin for $B(S, z)$, and the order of the poles is the same in the two cases.*

If there is an injective morphism of S into the Cayley graph, then all morphisms are injective, because the Cayley graph is homogeneous. In this case, the function $B(S, z)$ can be considered as counting the number of ways of embedding S in the Cayley graph. There is an alternative method of counting, where all morphisms with the same image are identified with each other. This introduces no essential difference in the theory, as we now see.

Let $e_n(S)$ be the number of subgraphs of Γ_n that are isomorphic to S (but the isomorphism is not specified). Since Γ_n is a full (induced) subgraph of Γ, $e_n(S)$ is unchanged if we restrict our attention to the case where S is a full subgraph. That is, given an embedding of S in Γ, we add to S all possible edges between existing vertices of S. We define $E(S, z) = \sum_{n=0}^{\infty} e_n(S) z^n$. Note that $\sigma e_n(S) = b_n(S)$, where σ is the number of automorphisms of S. It follows that $\sigma E(S, z) = B(S, z)$, and so results about B can be transferred to E.

5. EXAMPLES

The main result of this paper (Theorem 8.1) was originally conjectured by Saito for the case of a Fuchsian group with geometric generators. We tested Saito's conjecture for various groups, using programs like those described in Section 3. These experiments led us to other results proved in this paper.

Let G be a group with a fixed set of generators, and let $(g_1, g_2, \ldots, g_n)$ be an ordered list of generators, possibly with repetition. The graph $\Sigma_{g_1, g_2, \ldots, g_k}$ is defined to be a labelling of the graph structure on the interval $[0, n]$, with a vertex at each integer point. The directed edge $[i-1, i]$ is labelled with g_i. Denote the corresponding growth function of morphisms of this graph to the Cayley graph by $B_{g_1, g_2, \ldots, g_n}(z)$, where z is the indeterminate of the power series.

Example 5.1. Consider the $2, 3, 7$ Coxeter group

$$\langle a, b, c : a^2, b^2, c^2, (ab)^2, (bc)^3, (ca)^7 \rangle.$$

This is the group generated by reflections in the sides of a hyperbolic triangle with angles $\pi/2$, $\pi/3$ and $\pi/7$.

The growth function of this group and the subgraphs Σ_g for $g = a, b, c$ are

$$\begin{aligned} B(z) &= \frac{(z^6+z^5+z^4+z^3+z^2+z+1)(z^2+z+1)(z+1)^2}{(1-z)(z^{10}+z^9-z^7-z^6-z^5-z^4-z^3+z+1)} \\ &= 1+4z+9z^2+16z^3+25z^4+37z^5+53z^6+\cdots, \end{aligned}$$

$$\begin{aligned} B_g(z) &= \frac{2z(z+1)(z^6+z^5+z^4+z^3+z^2+z+1)(z^2+z+1)}{(1-z)(z^{10}+z^9-z^7-z^6-z^5-z^4-z^3+z+1)} \\ &= 2z+6z^2+12z^3+20z^4+30z^5+44z^6+62z^7+\cdots. \end{aligned}$$

Notice that the denominators of these functions are all identical and that the ratio of $B(z)$ to $B_g(z)$ is $(z+1) : 2z$ (see Proposition 7.4). The factor $(1-z)$ in the denominator occurs because we are counting everything in a ball of radius n. It would disappear if we were to count only things in the ball of radius n that are not contained in the ball of radius $n-1$.

Example 5.2. A related group is the $2, 3, 7$ triangle group

$$\langle a, b, c : a^2, b^3, c^7, abc \rangle.$$

This is the subgroup of the previous group of index two, consisting only of orientation preserving isometries. Its Cayley graph is shown in Figure 2.

The growth functions of this group and of the subgraphs Σ_g for $g = a, b, c$ and $\Sigma_{a,b}$ are given at the top of the next page. Again, notice that all the denominators are identical.

Example 5.3. Consider the free abelian group on three generators a, b and c. The growth functions for this group, and the subgraphs Σ_g for $g = a, b, c$, $\Sigma_{a,b}$ and $\Sigma_{a,b,c}$ are

$$\begin{aligned} B(z) &= \frac{(z+1)^3}{(z-1)^4} \\ &= 1 + 7z + 25z^2 + 63z^3 + 129z^4 \\ &\quad + 231z^5 + 377z^6 + 575z^7 + \dots, \\ B_g(z) &= \frac{2z}{z+1} B(z) \qquad \text{for } g = a, b, c, \\ B_{a,b}(z) &= \frac{z(z+1)(3z+1)}{(1-z)^4} \\ &= z + 8z^2 + 29z^3 + 72z^4 \\ &\quad + 145z^5 + 256z^6 + 413z^7 + \cdots, \\ B_{a,b,c}(z) &= \frac{4z^2(1+z)}{(1-z)^4} \\ &= 4z^2 + 20z^3 + 56z^4 \\ &\quad + 120z^5 + 220z^6 + 364z^7 + \cdots. \end{aligned}$$

6. AUTOMATIC GROUPS

Let G be a group, and let Σ be a finite set of generators with an involution, as described in Section 1. G is said to be *automatic* if there is a finite state automaton W, called the *word acceptor*, and, for each $x \in \Sigma$, a finite state automaton M_x, called the *multiplier for* x, with the following properties.

1. W is an automaton over Σ.
2. The composite $L(W) \subset \Sigma^* \to G$ is a bijection.
3. Each M_x is an automaton over the two variable alphabet (Σ, Σ). That is, M_x accepts pairs of words (w_1, w_2) with $w_1, w_2 \in \Sigma^*$.
4. The pair of strings (w_1, w_2) is accepted by M_x if and only if w_1 and w_2 are accepted by W and $\overline{w_1 x} = \overline{w_2} \in G$.

We have seen in Theorem 2.1 that the growth function of $L(W)$ is a rational function. However this does not necessarily imply that (G, Σ) has a rational growth function, despite the fact that each element of G has a unique representative in Σ^* accepted by W. This is because the growth function for (G, Σ) is defined using shortest representatives for elements of G and the representatives accepted by W do not need to be shortest. In order to make the two growth functions coincide, we need to assume that $L(W)$ consists entirely of shortest representatives. In fact, the programs written by Epstein, Holt and Rees [Epstein et al. 1991] try to compute an automatic structure of a special type that does satisfy this condition; it follows that these programs can be used to help compute the growth functions of many groups.

Definition 6.1 (geodesic automatic structure). We say that an automatic structure (W, Σ) on a group G is *geodesic*, if, for each element $w \in L(W)$, w is a shortest representative of $\overline{w}$ in Σ^*.

Theorem 6.2 (rational graph growth). *Let G be a group with geodesic automatic structure (W, Σ) and let S be a finite directed labelled graph. Then the growth function $C(S, z)$ counting the number of morphisms of S into Γ is a rational function of z. The same is true if we restrict to injective morphisms.*

Proof. Suppose that S has n vertices. We will construct a finite state automaton M_S that accepts n-tuples of strings over Σ such that there is a one-to-one correspondence between the set of n-variable strings accepted by M_S and the set of morphisms of S to Γ.

Let the vertices of S be $(s_1, \dots, s_n)$. We consider the set of n-tuples $(w_1, \dots, w_n)$ of elements of $L(W)$. If there is an edge in S labelled x from s_i to s_j, then we insist that (w_i, w_j) be accepted by M_x. According to [Epstein et al. 1992, Theorem 1.4.6], there is a finite state automaton M_S that accepts exactly the set of all such n-tuples. It is obvious that there is a one-to-one correspondence between the set of such n-tuples Ω of strings and morphisms $F : S \to \Gamma$. Moreover the length of an element of Ω, as a associated string over $(\Sigma, \dots, \Sigma)$

$$
\begin{aligned}
B(z) &= \frac{(z^8+4z^7+3z^6+2z^5+z^4+2z^3+3z^2+4z+1)(z^2+1)}{(1-z)(z^{10}-z^9-z^7+z^6-z^5+z^4-z^3-z+1)} \\
&= 1+6z+15z^2+31z^3+55z^4+88z^5+136z^6+203z^7+\cdots \\
B_a(z) &= \frac{2z(z^9+2z^8+2z^7+3z^6+2z^5+2z^4+2z^3+3z^2+2z+2)}{(1-z)(z^{10}-z^9-z^7+z^6-z^5+z^4-z^3-z+1)} \\
&= 4z+12z^2+26z^3+48z^4+78z^5+122z^6+184z^7+\cdots \\
B_b(z) &= \frac{z(2z^9+4z^8+5z^7+6z^6+3z^5+4z^4+5z^3+6z^2+3z+4)}{(1-z)(z^{10}-z^9-z^7+z^6-z^5+z^4-z^3-z+1)} \\
&= 4z+11z^2+24z^3+46z^4+75z^5+117z^6+177z^7+\cdots \\
B_c(z) &= \frac{z(2z^9+5z^8+4z^7+6z^6+4z^5+4z^4+4z^3+6z^2+4z+3)}{(1-z)(z^{10}-z^9-z^7+z^6-z^5+z^4-z^3-z+1)} \\
&= 3z+10z^2+23z^3+43z^4+71z^5+112z^6+170z^7+\cdots \\
B_{a,b}(z) &= \frac{z(z+1)(3z^8+z^7+4z^6+z^5+3z^4+z^3+4z^2+z+3)}{(1-z)(z^{10}-z^9-z^7+z^6-z^5+z^4-z^3-z+1)} \\
&= 3z+10z^2+22z^3+42z^4+70z^5+110z^6+167z^7+\cdots
\end{aligned}
$$

Growth functions for Example 5.2.

accepted by M_S, is equal to the minimum radius of a ball in Γ centred at the identity vertex and containing the vertices of the image of F.

In order to restrict to injective morphisms, we can change the automaton to ensure that, for $i \neq j$, $\bar{w}_i \neq \bar{w}_j$. In an automatic group, this condition can be recognized by a finite state automaton. □

7. IDENTITIES FOR MULTIPLIERS

This section is based on suggestions made by M. S. Paterson, to whom we are most grateful.

Let G be a group and let Σ be a finite set of generators with an involution ι that gives the formal inverse of a generator. A directed edge of the Cayley graph $\Gamma(G,\Sigma)$ from a vertex v_1 to a vertex v_2 is called *outward*, *inward* or *tangential*, according to whether $d_\Gamma(v_2,e)$ is greater than, less than, or equal to $d_\Gamma(v_1,e)$, where e is the identity vertex.

Let $x \in \Sigma$. We denote by $i_n(x)$ the number of edges labelled x pointing from a vertex at distance n from the identity to a vertex at distance $n-1$ from the identity. These are *incoming* edges. We denote by $o_n(x)$ the number of edges labelled x pointing from a vertex at distance $n-1$ from the identity to a vertex at distance n from the identity. These are *outgoing* edges. We denote by $t_n(x)$ the number of edges labelled x pointing from a vertex at distance n from the identity to a vertex at distance n from the identity. These are *tangential* edges. Note that $i_0(x)=o_0(x)=0$. Also $t_0(x)=0$ unless x is a trivial element of G. It is clear that for each n, $i_n(x)=o_n(\iota x)$ and $t_n(x)=t_n(\iota x)$.

We define the following functions:

$$
\begin{aligned}
I(x,z) &= \sum_n i_n(x)z^n, \\
T(x,z) &= \sum_n t_n(x)z^n, \\
O(x,z) &= \sum_n o_n(x)z^n.
\end{aligned}
$$

Each power series $I(x,z)$ and $O(x,z)$ is divisible by z.

Example 7.1. Here are three easy examples to illustrate the definitions of I, O and T. For the free group on one generator x we have $T(x,z)=0$,

$$I(x,z)=O(x,z)=\frac{z}{1-z}=z+z^2+z^3+z^4+\cdots.$$

For the free group on two generators x and y we have $T(x,z)=0$,

$$I(x,z)=O(x,z)=\frac{z}{1-3z}=z+3z^2+9z^3+27z^4+\cdots.$$

For the free Abelian group on two generators x and y we have $T(x,z)=0$,

$$I(x,z)=O(x,z)=\frac{z^2+z}{(1-z)^2}=z+3z^2+5z^3+7z^4+\cdots.$$

Example 7.2. Consider the 2,3,7 Coxeter group

$$\langle a,b,c : a^2, b^2, c^2, (ab)^2, (bc)^3, (ca)^7\rangle$$

of Example 5.1.

We have $T(g,z)=0$ and

$$\begin{aligned} I(g,z) & \\ &= O(g,z) \\ &= \frac{z(1+z)(1+z+z^2)(1+z+z^2+z^3+z^4+z^5+z^6)}{(1+z-z^3-z^4-z^5-z^6-z^7+z^9+z^{10})} \\ &= z+2z^2+3z^3+4z^4+5z^5+7z^6+9z^7+\cdots \end{aligned}$$

for each generator $g\in\{a,b,c\}$.

Consider the 2,3,7 triangle group

$$\langle a,b,c : a^2, b^3, c^7, abc\rangle$$

of Example 5.2. The expressions for I, T, and O are shown below. The Cayley graph is shown in Figure 2, from which one can check by hand the correctness of the first few coefficients in these expressions.

$$\begin{aligned}
T(a,z) &= \frac{2z(1+z+2z^2+z^3+z^4+z^5+2z^6+z^7+z^8)}{1-z-z^3+z^4-z^5+z^6-z^7-z^9+z^{10}} \\
&= 2z+4z^2+8z^3+12z^4+16z^5+24z^6+34z^7+\cdots \\
I(a,z)=O(a,z) &= \frac{z(1+z)(1-z+z^2-z^3+z^4)(1+z+z^2+z^3+z^4)}{1-z-z^3+z^4-z^5+z^6-z^7-z^9+z^{10}} \\
&= z+2z^2+3z^3+5z^4+7z^5+10z^6+14z^7+\cdots \\
T(b,z) &= \frac{z(2+z+2z^2+z^3+2z^4+z^5+2z^6+z^7+2z^8)}{1-z-z^3+z^4-z^5+z^6-z^7-z^9+z^{10}} \\
&= 2z+3z^2+5z^3+8z^4+11z^5+16z^6+22z^7+\cdots \\
I(b,z)=O(b,z) &= \frac{z(1+z)(1+2z^2+z^4+2z^6+z^8)}{1-z-z^3+z^4-z^5+z^6-z^7-z^9+z^{10}} \\
&= z+2z^2+4z^3+7z^4+9z^5+13z^6+19z^7+\cdots \\
T(c,z) &= \frac{z(1+2z^2+2z^6+z^8)}{(1-z-z^3+z^4-z^5+z^6-z^7-z^9+z^{10})} \\
&= z+z^2+3z^3+4z^4+4z^5+7z^6+10z^7+\cdots \\
I(c,z)=O(c,z) &= \frac{z(1+z)(1+z+z^2)(1+z^3+z^6)}{1-z-z^3+z^4-z^5+z^6-z^7-z^9+z^{10}} \\
&= z+3\,z^2+5\,z^3+8\,z^4+12\,z^5+17\,z^6+24\,z^7+\cdots
\end{aligned}$$

More growth functions for the group $\langle a,b,c : a^2, b^3, c^7, abc\rangle$ of Example 5.2.

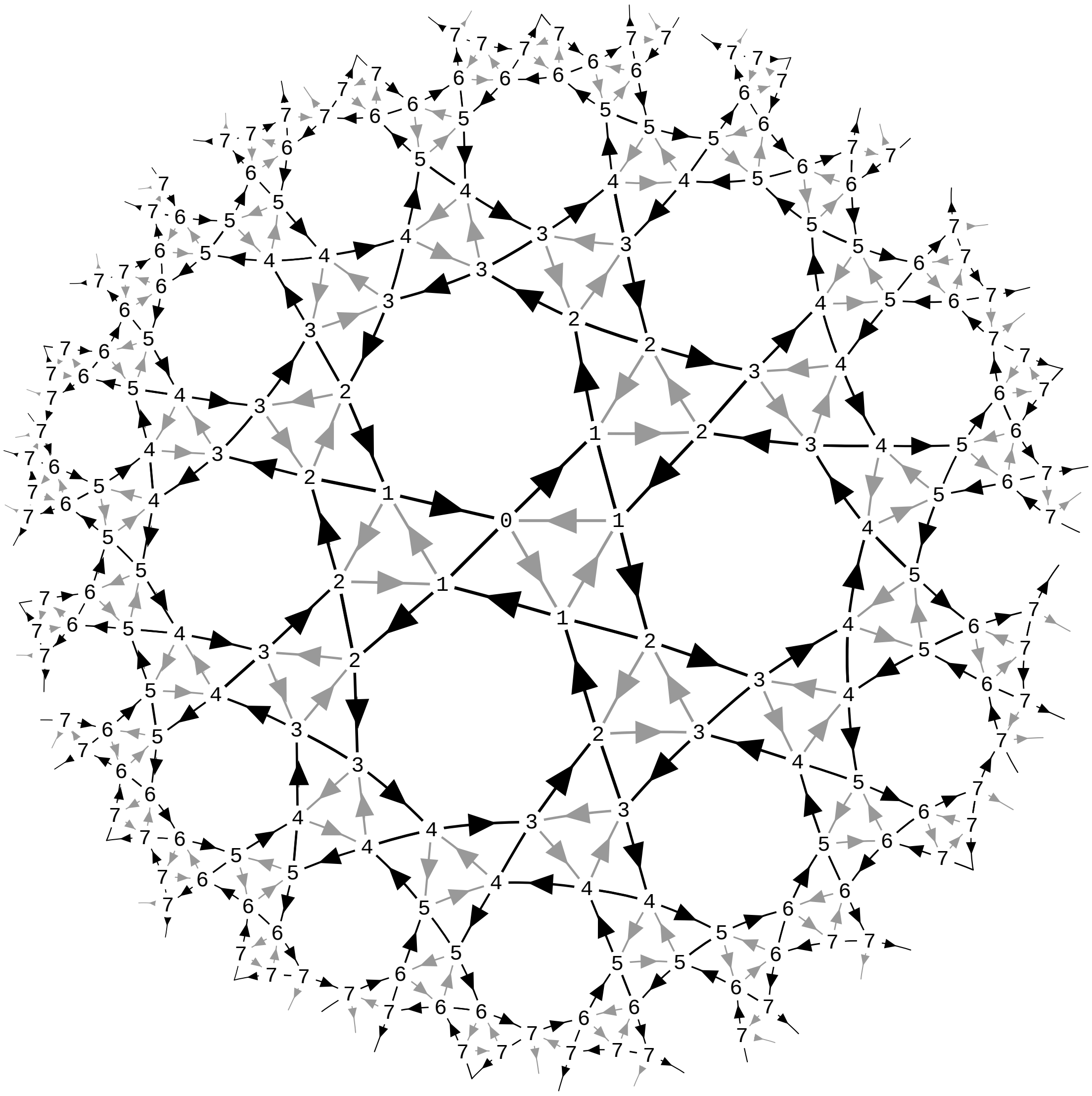

FIGURE 2. This is (part of) the Cayley graph of the orientation preserving $(2,3,7)$ triangle group. The black arrows have label c, the grey arrows have label b and the edges without arrows have label a. The identity element is at the vertex marked 0. All other vertices are marked with their distance from the origin. A picture like this is helpful for doing calculations by hand, and the reader may wish to verify the accuracy of some of the computer calculations by using the picture to determine the coefficients of z^n for small values of n in various growth series. This picture was drawn with the help of the Mathematica package Hyperbolic.m [Goodman and Levy 1993].

Let G and Σ be as above. Let $C(z)$ be the growth function for (G, Σ). Let S_x be the directed labelled graph with a single edge labelled x and let $C(S_x, z)$ be the growth function counting the number of copies of S_x contained in the ball of radius n about the identity element, but not in the ball of radius $n-1$.

Lemma 7.3. *We have the following equations for each $x \in \Sigma$:*

$$\begin{aligned} T(x,z) &= T(\iota x, z), \\ C(z) &= T(x,z) + \frac{(z+1)I(x,z)}{z}, \\ I(x,z) &= I(\iota x, z) = O(x,z) = O(\iota x, z), \\ C(S_x, z) &= 2I(x,z) + T(x,z). \end{aligned}$$

Proof. The first equation follows because each tangential edge labelled x corresponds to a tangential edge labelled ιx in the reverse direction.

For each vertex v of the Cayley graph, there is a single edge labelled x starting at v. Therefore, if c_n is the number of vertices at distance n from e, we have, for each $x \in \Sigma$,

$$\begin{aligned} c_n &= i_n(x) + o_{n+1}(x) + t_n(x) \\ &= i_n(x) + i_{n+1}(\iota x) + t_n(x). \end{aligned}$$

Replacing x by ιx, we see that for each n, $i_n(x) + i_{n+1}(\iota x) = i_n(\iota x) + i_{n+1}(x)$. It follows by induction on n that $i_n(x) = i_n(\iota x)$. The second equality follows from this.

The third equality now follows from the fact (noted above) that $I(x,z) = O(\iota x, z)$.

To prove the fourth equality, note that an edge labelled x that lies in the ball of radius n, but not in the ball of radius $n-1$, is inward or outward or tangential and these "or's" are exclusive. So $C(S_x, z) = I(x,z) + O(x,z) + T(x,z)$. The fourth equality follows. □

The next result is an explanation of results first observed experimentally in tables of growth functions of appropriate automata.

Proposition 7.4 (edge ratio). *If $x \in \Sigma$, let S_x be defined as above. If each relator of the defining relators for G has even length then, for each $x \in \Sigma$,*

$$\frac{C(z)}{C(S_x, z)} = \frac{1+z}{2z}.$$

Proof. Since each relator has even length, there is a homomorphism onto the group with two elements, sending each element of Σ to the non-trivial element. It follows that $T(x,z)$ is identically zero for each $x \in \Sigma$. It also follows that $I(x,z) = O(x,z)$ is not identically zero. □

Proposition 7.5. *In the case of a geodesic automatic group, $I(x,z)$ and $T(x,z)$ are rational functions of z. In fact, if we specify any finite directed graph S, with labels from Σ, and, for each edge of S whether it is to map to an inward pointing, outward pointing, or tangential edge of $\Gamma(G, \Sigma)$, then the associated growth function is a rational function.*

Proof. There is a finite state automaton over (Σ, Σ) that can detect whether, given two strings, the first is exactly one longer than the second. The same is true for the detection of equal length. We combine these finite state automata with the multipliers for the automatic structure to achieve the desired effect. As in the case of Theorem 6.2, we can use [Epstein et al. 1992, Theorem 1.4.6] in order to complete the proof. □

8. GROWTH IN WORD-HYPERBOLIC GROUPS

Let G be a group with a finite set of generators Σ. We assume that Σ has an involution corresponding to taking the inverse. First we recall Gromov's definition [Ghys and de la Harpe 1989] that G is said to be *word-hyperbolic* if there is a positive integer k such that, for any choice of three vertices A, B and C in the Cayley graph $\Gamma = \Gamma(G, \Sigma)$, any choice of geodesic paths AB, BC and CA in the Cayley graph of Γ, and any choice of $x \in AB$, we have $d_\Gamma(x, BC \cup CA) \le k$ (see Figure 3). If G is hyperbolic with respect to one finite set of generators, it is also hyperbolic with respect to any other set, but the constant k, called the *constant of hyperbolicity* of the pair (G, Σ), may change.

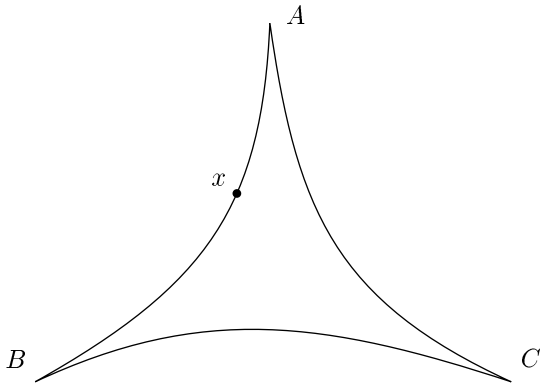

FIGURE 3. The point x is an arbitrary point on AB, and its distance to $BC \cup CA$ is bounded.

For the remainder of this section, we fix Σ and $k > 0$.

This is the main theorem of this paper:

Theorem 8.1. *Let G be a word-hyperbolic group and let Σ be any set of generators with an involution $\iota : \Sigma \to \Sigma$ such that $\overline{\iota x} = \overline{x}^{-1}$ for each $x \in \Sigma$. Then there is a polynomial $Q(z)$ with integral coefficients (depending on G and Σ) with the following property. Let S be any non-empty finite connected labelled directed graph with labels in Σ. Then the growth function $C(S, z)$ is a rational function with denominator $Q(z)$.*

Proof. The idea is to construct a single automaton from which each of the infinite set of growth functions obtained as S varies can be deduced.

We order the elements of Σ, and define L to be the set of strings w over Σ, with the property that, among all representatives in Σ^* for $\overline{w}$, w is shortest and least lexicographically among the shortest representatives. Theorem 3.4.5 and Corollary 2.5.2 of [Epstein et al. 1992] show that (Σ, L) is an automatic structure. Let W be the minimal finite state automaton over Σ such that $L(W) = L$. Since L is prefix-closed, all states of W are accept states except for a single fail state.

Let $\Gamma = \Gamma(G, \Sigma)$ be the Cayley graph. Each vertex v of Γ is labelled by a state of W in the obvious way—we take the unique geodesic path accepted by W from the identity vertex e to v. This traces out a path of arrows in W from the start state to the state of W that labels v.

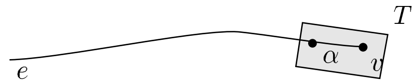

FIGURE 4. An accepted geodesic path from the identity element e to a vertex $v \in G$. The path marked α is the associated $2k$-history, and T is a k-neighbourhood of α.

We use the labelling to define the states of a large automaton H. First we define an *r-history* ($r > 0$ an integer) to be a final segment of an accepted path in Γ starting at the identity vertex e. We define the *initial point* and *final point* of an r-history in the obvious way. The segment must have length exactly r, unless it begins at e, in which case the segment is allowed to have length less than r. Fix for the moment some $2k$-history α (where k is defined at the beginning of this section); and let T be the full subgraph of vertices within a distance k of α. Each of the vertices of T carries its label from the states of W.

Let v be the final point of α. We label each vertex $u \in T$ with the integer $p = d_\Gamma(e, u) - d_\Gamma(e, v)$, so that each vertex of T is labelled by a pair (s, p), where s is a state of W and $-3k \le p \le k$.

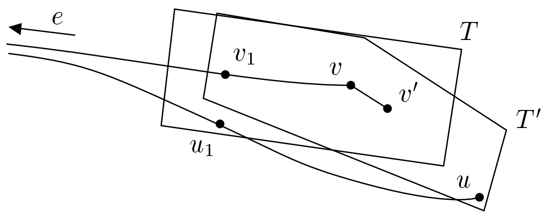

FIGURE 5. Why H is an automaton.

An isomorphism between (T, α) and (T', α') is an isomorphism that makes vertices and edges correspond, sends α to α' and preserves labels (s, p) on vertices and labels in Σ on directed edges. A state of H is an isomorphism class $[T, \alpha]$ of such pairs. H also has a fail state. The initial state of H corresponds to taking for α the path of length zero at e.

We now define the image $s_H x$ of $x \in \Sigma$ acting on a state s_H of H. If s_H is the fail state, then of course $s_H x = s_H$. Otherwise let s_H be represented by (T, α). Let v be the final point of α and let s be the state of W labelling v. If applying x to s in W leads to the fail state of W, we define the action of $s_H x$ to be equal to the fail state of H.

Otherwise we must define $(T', \alpha') = (T, \alpha)x$ and show that $[T', \alpha']$ depends only on the isomorphism class $[T, \alpha]$ and on x. Applying x to v gives a vertex $v' \in \Gamma$ with label s', where s' is the image of s under x. This gives us the $(2k)$-history α' by adding v' at the final end of α and possibly dropping the initial vertex of α. We define T' to be the full subgraph of Γ on the vertices in a k-neighbourhood of α'. Since Γ is a homogeneous graph, T' (without the labelling of vertices) is determined up to isomorphism by x and the isomorphism class of (T, α).

Let u be a vertex of T'. We need to show how to determine the label of u. We already know this label unless $d_\Gamma(u, v) = k + 1$ and $d_\Gamma(u, v') = k$. Let $[e, u]$ and $[e, v']$ denote the accepted geodesics in Γ from the identity element e, and let $[u, v']$ be any geodesic. Then $[e, v']$ contains α and α' and $[u, v']$ has length k. Let v_1 be the vertex on $[e, v']$ a distance $2k$ from v'. (If $d_\Gamma(e, v') < 2k$, we set $v_1 = e$.) Then $v_1 \in \alpha' \cap \alpha$. We now use the fact that G is a hyperbolic group. The definition of k at the beginning of this section shows that there is a vertex $u_1 \in [e, u]$ such that $d_\Gamma(u_1, v_1) \leq k$. Therefore $u_1 \in T$ and we know the label on u_1. The reason we used the $2k$-history of v was precisely to ensure that we could find such a pair (u_1, v_1) within a region that we know about.

We now take the $(8k)$-neighbourhood of v, for example, and trace out in it the portion $[u_1, u] \subset [e, u]$ of the accepted geodesic, using W to accomplish this task. This enables us to assign the correct state of W to u. Using the fact that a fixed size neighbourhood of a vertex is isomorphic to the same size neighbourhood of any other vertex (apart from the labelling of vertices), it is clear that the construction is independent of the isomorphism class of (T, α). This completes the definition of the finite state automaton H over Σ.

Let $c_n(S)$ be the number of morphisms $S \to \Gamma_n$ that do not factor through Γ_{n-1}, and let $C(S, z) = \sum_{n=0}^{\infty} c_n(S) z^n$. From Theorem 6.2 we know that $C(S, z)$ is a rational function, so we only need to find a common denominator for the $C(S, z)$ as S varies over all finite subgraphs of Γ.

Lemma 8.2. *Let A be the transition matrix for the finite state automaton H. Then, for each finite connected subgraph S of Γ, $\det(I - zA)C(S, z)$ is a polynomial function of z.*

Proof. The denominator of a rational function is unaffected by omitting the first finite number of terms in the formal power series, as this changes the rational function by adding a polynomial. We may therefore restrict ourselves to computing $c_n(S, z)$ for large values of n. We fix a basepoint $s_1 \in S$ and let d be the diameter of S. We fix a morphism $f : S \to \Gamma_n$ that does not factor through Γ_{n-1}. Let $d_\Gamma(f s_1, e) = n - r$. Then $0 \leq r \leq d$.

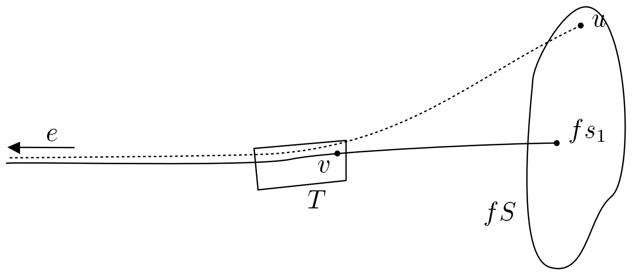

FIGURE 6. Here we see how the structure and position of a gigantic subgraph fS of the Cayley graph can be determined by using the relatively much smaller subset of the Cayley graph corresponding to a state of the finite state automaton H.

Let v be the vertex on the accepted geodesic $[e, f s_1]$ at a distance $d + k$ from $f s_1$, so that

$$d_\Gamma(v, e) = n - r - d - k.$$

There is a unique state $[T, \alpha]$ of H such that v is the final point of α. If u is any vertex of fS, then the accepted geodesic $[e, u]$ passes within a distance k of v, and therefore passes through T. Starting with (T, α), we build up the forward paths from vertices of T for a fixed distance ($2d + 2k$ is big enough), which is independent of n. Such forward paths reach all points of fS.

Let (T, α) represent a state of H and let $0 \leq r \leq d$. We stress here that we are using a particular representative of the state, and not its isomorphism class. There are a finite number of possible morphisms of $f : S \to \Gamma$ such that fs_1 is related to the final point v of α in the manner described in the previous paragraph. Each such morphism gives rise to a particular value of the integer called r above. We define $m(T, \alpha, r)$ be the number of morphisms of S into Γ giving rise to the pair (T, α) and this particular value of r. Clearly $m(T, \alpha, r)$ depends only on r and the isomorphism class $[T, \alpha]$.

Let $H_{[T,\alpha]}$ be the finite state automaton with the same states, arrows and initial state as H and with only one accept state, namely $[T, \alpha]$. For $n \geq d+k$ we have

$$c_n(S) = \sum_{[T,\alpha],r} c_{n-r-d-k}(H_{[T,\alpha]}) m(T, \alpha, r).$$

Let $p(T, \alpha, z)$ be the polynomial in z such that the growth function of $H_{[T,\alpha]}$ is given by $C(H_{[T,\alpha]}, z) = p(T, \alpha, z) / \det(I - zA)$ (see Section 2).

Then $p(T, \alpha, z)$ depends only on the isomorphism class $[T, \alpha]$. We have

$$\begin{aligned} C(S,z) &= \sum_{n=0}^{\infty} c_n(S) z^n \\ &= \sum_{n=0}^{d+k-1} c_n(S) z^n \\ &\quad + \frac{\displaystyle\sum_{[T,\alpha];\, 0 \leq r \leq d} z^{r+d+k} m(T,\alpha,r) p(T,\alpha,r)}{\det(I - zA)}. \end{aligned}$$

This completes the proof of the lemma. □

The theorem follows on putting $Q(z) = \det(I - zA)$ in the notation of Lemma 8.2. □

Corollary 8.3. *Theorem* 8.1 *holds for morphisms of disconnected graphs S as well, provided we allow the denominator to depend on the number of components of S.*

Proof. One approach is to point out that the fact that the Hadamard product of two rational functions is rational [Stanley 1986].

Alternatively, we can extend the proof of Theorem 8.1 as follows. First let L be any regular language over an alphabet Σ and let $E = \{\$\}^*$ be the regular language that consists of any string of padding symbols. Then the concatenation $L.E$ is a regular language over $\Sigma \cup \{\$\}$. This enables us to count strings of length at most n, instead of strings of length exactly n. (Conversely, we may strike out padding symbols from any regular language and still have a regular language.)

If we have several components for S, we form one automaton for each component, as in Theorem 8.1. We then change these automata to allow padding at the end. We can use the product of these automata to obtain the number of morphisms in general.

To be completely watertight, we still need to check what happens when one or more of the components of S is mapped too near to the identity to be controlled by the automaton H defined in the proof of Theorem 8.1. For each component of S, we have to consider a finite number of additional cases. The necessary adjustment to the counting function for all components is to add a finite number of rational functions. Details are left to the reader. □

The dependence in the preceding corollary of the denominator on the number of components is necessary. For example, let G be the cyclic infinite group on a single generator, with Cayley graph Γ. Let S_k be the disjoint union of k vertices. Then the number of morphisms of S_k into the n-ball in Γ is $(2n+1)^k$. It can easily be shown by induction on k that the associated rational function $\sum (2n+1)^k z^n$ has denominator $(1 - z)^{k+1}$. If we wish to restrict to the morphisms that have image in the n-ball, but not in the $(n-1)$-ball, then the denominator is $(1 - z)^k$.

9. COUNTING FINITE SUBGRAPHS THAT ARE NOT LABELLED, DIRECTED AND CONNECTED

Let Γ be the Cayley graph of a group G with generators Σ. In Γ each directed edge labelled with an element $x \in \Sigma$ has a corresponding edge joining the same two vertices, but with the opposite direction and the inverse label x^{-1}. Let S be a finite graph. For each edge we may specify a label, or we may leave the edge unlabelled. For each edge we may specify a direction, or we may leave the edge undirected. If the edge is labelled, we require that it be directed, but an edge of S may be directed without being labelled. We define the notion of a morphism of S into Γ in the obvious way, so that this generalizes the concept of a morphism when S is directed and labelled. If there is no label or direction on an edge e of S, then one is permitted to map e to any edge of Γ, provided the ends of the edge are mapped to the ends of the image of e.

Theorem 9.1. *Let S be as just described. Let $a_n(S)$ be the number of morphisms of S into Γ such that all vertices of S are mapped into the ball of radius n centred at the identity element of Γ. Let $b_n(S)$ be the number of distinct images of injective morphisms of S in Γ with vertices in the same n-ball. If, for each connected, labelled, directed, finite graph S', the growth series for morphisms of S' into Γ is rational, then so are $\sum a_n(S)z^n$ and $\sum b_n(S)z^n$.*

There are variants on these two types of counting, for example just counting distinct images of any morphism, rather than of injective morphisms. These variants can be shown to be rational by similar reasoning.

Proof. The most important case is when S is connected, and we will assume this until otherwise stated. We can complete S to a connected, directed, labelled graph S', by inserting labels and directions on all edges of S where these are missing. For each such completion S', there may or may not exist a morphism into Γ. The set of pairs consisting of a completion S' plus a morphism into Γ is in one-to-one correspondence with the set of morphisms of S into Γ. It follows that the growth series $\sum a_n(S)z^n$ is the sum of the corresponding growth series for the various possible S', and is therefore rational.

Now we work with the b_n's. This paragraph works for S connected or not. Let H be the group of morphisms of S onto itself. Each completion of S to a labelled, directed graph is acted on by $h \in H$, possibly changing the labels and directions on some edges. We take one representative S' in each orbit under H, and we discard representatives which do not have an injective morphism into Γ. Note that if a morphism of S' into Γ is injective, then all its translates under H are also injective. Suppose the stabilizer of S' in H has s' elements. Then the number of morphisms of S' into the ball of radius n about the identity is a multiple of s', and so is the corresponding growth series.

If S is connected, then so is S', and the growth series is rational. We divide this rational function by s'. We then sum over all the representatives S', obtaining the rational function $\sum b_n(S)z^n$. This completes the proof when S is connected.

Now let us consider the case where S is not connected. Recall that the Hadamard product of $\sum \alpha_n z^n$ and $\sum \beta_n z^n$ is defined as $\sum \alpha_n \beta_n z^n$. It is proved in [Stanley 1986] that if two power series are each rational, then their Hadamard product also is. This shows that the rationality of $\sum a_n(S)z^n$ for every connected S implies the rationality of the same series when S is not necessarily connected.

We now prove the rationality of $\sum b_n z^n$ when S is not connected. By the argument given above, we may assume that S is directed and labelled. The argument above also proves that we need only prove the rationality of the growth series for injective morphisms of S into Γ, where S is directed and labelled. Let S be the disjoint union of S_1 and S_2, where S_1 is connected. By induction on the number of components, we may assume that the growth series for S_1 and S_2 are rational. Using the theorem about Hadamard products, it follows that the morphisms from S to Γ which are injective on each of S_1 and S_2 gives a rational growth series. From this we have to subtract the growth function for morphisms where the image of S_1 intersects the image of S_2. Each of the finite number of configurations in which they meet gives rise to a labelled connected graph S_3 which has fewer components than S. By induction, the growth series for injective morphisms of S_3 into Γ is rational. So this completes the proof of the theorem. □

10. HISTORICAL NOTE

The first code to determine automatic structures on groups was written by Epstein, Derek Holt, and Sarah Rees during the period 1986–90. An important step was the realization by Holt that the process could be made feasible for many groups by using the Knuth–Bendix process [Knuth and Bendix 1970]. Around 1988, Saito visited Warwick and told our group of his computations of the growth functions for graphs. Saito was interested in this problem because of the connection with computations he was making of quantum phenomena in hyperbolic geometry. He had worked out by hand some examples of some of the theorems reported here for particular groups, using generators with geometric significance. (In our work, generators do not have to have geometric significance.)

Using automatic group theory, Epstein was immediately able to throw light on some of Saito's conjectures, but solving them completely was much harder. (It is interesting to think how intractable this problem seemed before the introduction of automatic group theory and the associated computer programs.) Iano-Fletcher then undertook a systematic study of a number of groups, using computer algebra packages together with the Warwick automatic groups software, to compute rational functions arising for particular groups and particular finite graphs. Iano-Fletcher's work displayed certain regularities, such as those shown in Lemma 7.3. We managed to prove these regularities without too much difficulty. In 1989, Zwick joined our group and developed highly efficient code based on the algorithms presented in Section 3 of this paper. With this code, it became possible to investigate large numbers of examples very rapidly.

Saito's main conjecture, that a common denominator could be found for growth functions of any finite graph in a hyperbolic group, was investigated first using Iano-Fletcher's code and then Zwick's. So many examples satisfied Saito's conjecture that we soon became convinced that it was true; armed with that conviction, Epstein found the proof given in Section 8.

Whether the conjecture is true for other classes of automatic groups remains an open problem. It would seem to be true, and is probably not hard to prove, for an abelian group with any set of generators.

ACKNOWLEDGEMENTS

This research was supported by SERC. We thank M. S. Paterson for significant contributions to Section 7. We thank P. J. Sanders for programming work in support of this paper.

REFERENCES

[Brent et al. 1980] R. P. Brent, F. G. Gustavson, and D. Y. Y. Yun, "Fast solution of Toeplitz systems of equations and computation of Padé approximants", *J. Algorithms* **1** (1980), 259–295.

[Cannon and Wagreich 1992] J. W. Cannon and P. Wagreich, "Growth functions of surface groups", *Math. Ann.* **293** (1992), 239–257.

[Coppersmith and Winograd 1990] D. Coppersmith and S. Winograd, "Matrix multiplication via arithmetic progressions", *J. Symbolic Comput.* **9** (1990), 251–280.

[Epstein et al. 1992] 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, Boston, 1992.

[Epstein et al. 1991] D. B. A. Epstein, D. F. Holt, and S. E. Rees, "The use of Knuth–Bendix methods to solve the word problem in automatic groups", *J. Symbolic Comput.* **12** (1991), 397–414.

[Faddeev and Faddeeva 1963] D. K. Faddeev and V. N. Faddeeva, *Computational methods in linear algebra*, W. H. Freeman, San Francisco, 1963.

[Ghys and de la Harpe 1989] E. Ghys and P. de la Harpe (editors), *Sur les groupes hyperboliques d'après Mikhael Gromov*, Progress in Math. **83**, Birkhäuser, Boston, 1989.

[Goodman and Levy 1993] O. Goodman and S. Levy, *Hyperbolic.m*, http://www.geom.umn.edu/software/download/hyperbolic.html: Geometry Center, University of Minnesota, 1993.

[Keller-Gehrig 1985] W. Keller-Gehrig, "Fast algorithms for the characteristic polynomial", *Theoret. Comput. Sci.* **36** (1985), 309–317.

[Knuth and Bendix 1970] D. E. Knuth and P. B. Bendix, "Simple word problems in universal algebra", pp. 263–297 in *Computational problems in abstract algebras* (Oxford, 1967), edited by J. Leech, Pergamon Press, Oxford, 1970.

[Lauer 1983] M. Lauer, "Computing by homomorphic images", pp. 139–168 in *Computer algebra: symbolic and algebraic computation* (2nd ed.)., Springer, Wien, 1983.

[Massey 1969] J. L. Massey, "Shift Register Synthesis and BCH Decoding", *IEEE Trans. Information Theory* **IT-15** (1969), 122–127.

[Mignotte 1983] M. Mignotte, "Some useful bounds", pp. 259–263 in *Computer algebra: symbolic and algebraic computation* (2nd ed.)., Springer, Wien, 1983.

[Saito 1991] K. Saito, "The limit element in the configuration algebra for a discrete group", pp. 931–942 in *Proc. Int. Congress Mathematicians* (Kyoto, 1990), Springer, Tokyo, 1991.

[Stanley 1986] R. P. Stanley, *Enumerative combinatorics*, Wadsworth and Brooks/Cole, Monterey, CA, 1986.

[Wiedemann 1986] D. Wiedemann, "Solving sparse linear equations over finite fields", *IEEE Trans. Information Theory* **IT-32** (1986), 54–62.

David B. A. Epstein, Mathematics Institute, University of Warwick, Coventry, CV4 7AL, United Kingdom (D.B.A.Epstein@warwick.ac.uk)

Anthony R. Iano-Fletcher, Mathematics Department, University of Nottingham [AUTHOR: full address?] (Anthony.Iano-Fletcher@maths.nott.ac.uk)

Uri Zwick, Department of Computer Science, Tel Aviv University, Ramat Aviv, Tel Aviv 69978, Israel (zwick@math.tau.ac.il)

Received January 22, 1996; accepted July 19