University of Ottawa

Department of Mathematics and Statistics

Faculty of Science

Limits of Rauzy graphs of low-complexity words

Blair Drummond

Supervised by

Dr. Vadim Kaimanovich

Thesis submitted to the Faculty of Graduate and Postdoctoral Studies in partial fulfillment of the requirements for the degree of Master of Science in Mathematics[1]

[1] The Master's program is a joint program with Carleton University, administered by the Ottawa-Carleton Institute of Mathematics and Statistics

© Blair Drummond, Ottawa, Canada, 2019

Abstract

We consider Benjamini-Schramm limits of Rauzy Graphs of low-complexity words. Low-complexity words are infinite words (over a finite alphabet), for which the number of subwords of length n is bounded by some Kn — examples of such a word include the Thue-Morse word $01101001\ldots$ and the Fibonacci word. Rauzy graphs $\vec{\mathcal{R}}^n(\omega)$ have the length n subwords of ω as vertices, and the oriented edges between vertices indicate that two words appear immediately adjacent to each other in ω (with overlap); the edges are also equipped with labels, which indicate what "new letter" was appended to the end of the terminal vertex of an edge. In a natural way, the labels of consecutive edges in a Rauzy graph encode subwords of ω. The Benjamini-Schramm limit of a sequence of graphs is a distribution on (possibly infinite) rooted graphs governed by the convergence in distribution of random neighborhoods of the sequence of finite graphs.

In the case of Rauzy graphs without edge-labelings, we establish that the Rauzy graphs of aperiodic low-complexity words converge to the line graph in the Benjamini-Schramm sense. In the same case, but for edge-labelled Rauzy graphs, we also prove that that the limit exists when the frequencies of all subwords in the infinite word, ω, are well defined (when the subshift of ω is uniquely ergodic), and we show that the limit can be identified with the unique ergodic measure associated to the subshift generated by the word. The eventually periodic (i.e. finite) cases are also shown. Finally, we show that for non-uniquely ergodic systems, the Benjamini-Schramm limit need not exist —though it can in some instances— and we provide examples to demonstrate the variety of possible behaviors.

Notation & Conventions

$\mathbb{N}$	$\{0, 1, 2, 3, \dots\}$. (0 is a natural number.)
$\mathcal{A}$	A finite alphabet, such as $\{a, b, c\}$ or $\{0, \dots, n-1\}$.
$\mathcal{A}^*$	The free monoid on $\mathcal{A}$
$\omega \in \mathcal{A}^{\mathbb{N}}$	A singly infinite word with alphabet $\mathcal{A}$. Sometimes ω will be bi-infinite, in which case $\omega \in \mathcal{A}^{\mathbb{Z}}$
ω_n	The n^{th} letter in the infinite word ω
$[u]$	u is a finite word, and $[u]$ is the cylinder set based at u consisting of all infinite words beginning with u ($= u\mathcal{A}^{\mathbb{N}}$), though sometimes it will refer to simply any open set in the product topology over discrete $\mathcal{A}$, and $[u] = \prod_i \mathcal{A} \times \{u_1\} \times \cdots \times \{u_n\} \times \prod_j \mathcal{A}$. The latter definition is necessary in $\mathcal{A}^{\mathbb{Z}}$.
$\lvert u \rvert$	u is a finite word, and $\lvert u \rvert$ is its length.
$S : X \circlearrowleft$	S will always denote the shift map on a symbolic system.
$\mathcal{G}_\bullet$	Space of (isomorphism classes of) rooted connected and locally finite oriented graphs.
$\overrightarrow{\mathcal{G}}_\bullet$	Space of (isomorphism classes of) rooted connected and locally finite oriented graphs with labelled edges.
$\mathcal{R}^n_{\mathcal{A}}(\omega)$	The n^{th} Rauzy Graph of ω with edges labelled by $\mathcal{A}$. Usually ω will be implicit. We will also consider unlabelled Rauzy graphs, and $\mathcal{A}$ will then be omitted.
$\overrightarrow{\mathcal{R}}^n_{\mathcal{A}}(\omega)$	The nth directed Rauzy Graph of ω with edges labelled by $\mathcal{A}$. We will also consider unlabelled Rauzy graphs, and $\mathcal{A}$ will then be omitted.
$\mathfrak{t}$	The Thue-Morse word 0110100110010110...
$\mathrm{freq}_u(\omega)$	The limit frequency of u in ω (if this is defined)
$D_r(G, v)$	The closed subgraph of (G, v) of all vertices and edges within r of v.
$B_r([(G, o)])$	The closed ball in $\mathcal{G}_\bullet$ or $\overrightarrow{\mathcal{G}}_\bullet$ of graphs which agree with (G, o) on an r-neighborhood of o.

Table of Contents

1. Introduction

We consider Benjamini-Schramm limits of Rauzy Graphs of low-complexity words. Each of these terms are defined in detail in the text, but we will also provide a brief definition of each here.

Low-complexity words

Given an infinite word or bi-infinite word ω, we can define the *language of* ω, $L(\omega)$, to be the collection of all finite length subwords which occur in ω, and we can denote by $L_n(\omega) \subset L(\omega)$ the subwords of ω of length $n \in \mathbb{N}$ (taking $\mathbb{N} = \{0, 1, 2, \dots\}$). A word is called *low-complexity* (Definition 4.0.1) if $|L_n(\omega)| < Kn$ for some $K > 0, K \in \mathbb{R}$. Examples of such words include the Thue-Morse word $\mathfrak{t} = 01101001\dots$ (Example 2.4.8) and the Fibonacci word, $\mathfrak{f}$ (Example 4.0.6). (Note, by contrast, that most[2] infinite words would have exponentially many subwords of a given length, not linearly many.)

Rauzy graphs

The nth Rauzy graph of ω, $\vec{\mathcal{R}}^n(\omega)$, has $L_n(\omega)$ as its vertex set, and there is an (oriented) edge between two length n words $u, v \in L_n(\omega)$, if they appear adjacent to one another (with overlap) in ω (Definition 5.1.1). So,

$$u \to v \ \text{ in } \vec{\mathcal{R}}^n(\omega) \quad \Leftrightarrow \quad \overbrace{w_1\, \underbrace{w_2 \dots w_n}_{v}}^{u} w_{n+1},\ w \in L_{n+1}(\omega)$$

The edges $u \to v$ are also labelled by the last letter of v because the last letter of v is the only letter in v which is not shared by u. These edges also have a description in terms of *windows* [Kai18]: One can place a window of length n on ω, and then one can apply the shift, S, to the word in order to slide the word visible under the window. With this interpretation, the edge $u \to v$ is in the Rauzy graph if u lies under the window for $S^k\omega$, and if v lies under the window for $S^{k+1}\omega$ — the edge is also labelled by the last letter to enter the window. For example, the label of the edge $011 \to 110$ from $\boxed{011}01... \to\ \boxed{110}10...$ would be 0, because of the rightmost 0 which just entered the window.

The Benjamini-Schramm limit

The *Benjamini-Schramm limit* of a sequence of finite graphs is a distribution on (possibly infinite) rooted graphs governed by the convergence in distribution of random neighborhoods of the sequence of finite graphs. The full definition is given in Definition 5.3.2, but in brief, one defines a space $\vec{\mathcal{G}}_\bullet$ of isomorphism classes of labelled, rooted, connected, and locally finite graphs — isomorphism classes are required for capturing graphs in a way that is independent of the choice of vertex set, but also to handle symmetries in the graphs, and the rootedness, local finiteness, and connectedness requirements are necessary for ensuring that $\vec{\mathcal{G}}_\bullet$ can be given a reasonable topology. The Benjamini-Schramm limit then works as follows: take a sequence of finite (unrooted) graphs (G_i), and for each

[2] In a sense that will be defined, Lebesgue almost-all

G_i consider the *random rooting* of the graph; that is, the measure $\mu_i = \frac{1}{|G_i|}\sum_{v \in G_i} \delta_{[(G_i,v)]}$ on $\overrightarrow{\mathcal{G}}_\bullet$ which captures every possible rooting of G_i. Through this, we get a sequence of measures (μ_i) on $\overrightarrow{\mathcal{G}}_\bullet$, and we say the Benjamini-Schramm limit exists if the sequence of measures converges to a measure μ.

(*Convergence in the space of measures is defined via identification of the Borel probability measures with* $C(\mathcal{G}_\bullet, \mathbb{R})^*$ *endowed with the weak-** *topology*)

Symbolic dynamics and ergodic theory

Our goal is ultimately to identify the Benjamini-Schramm limit (or subsequential limits) of Rauzy graphs of low-complexity words. Using only the linear growth of low-complexity words, a result of Frid [Fri01] effectively gives a proof that bifurcations in large Rauzy graphs of low-complexity words are sparse, and that, when one passes to the limit, the Rauzy Graphs converge to the line graph (in the unlabelled sense). When one introduces edge-labelings, then the line-segments of the Rauzy graphs carry subwords of ω, and so the Benjamini-Schramm limit of the Rauzy graphs (in the labelled sense) becomes intimately related to the "frequency" of subwords of ω.

To this end, we provide a reasonably self-contained discussion of the necessary ideas from symbolic dynamics and ergodic theory to define the notion of *unique-ergodicity*, and we show that when a low-complexity word has a unique shift-invariant probability measure, then the unique measure simply captures the frequency of subwords (Proposition 3.2.6). We also develop the tools for treatment of the non-uniquely-ergodic case.

Limits of Rauzy graphs of low-complexity words

Using this theory from symbolic dynamics and ergodic theory, we show that when a word ω has well-defined subword frequencies, that the frequencies of neighborhoods in large Rauzy graphs converge (recall that the neighborhoods themselves also re-encode subwords of ω). Using this, we get that the Benjamini-Schramm limit exists in this case, and the limit measure can be identified with the unique shift invariant measure on the symbolic system.

However, this is no longer true (in general) when considering words which do not have well-defined subword frequencies (in other words, when the subshift is not uniquely ergodic). We provide some results that help describe the subsequential limits, and we provide examples showing that the limit may exist, it may not, and there may be no subsequence converging to a measure that resembles an ergodic measure from the shift space of ω (the smallest shift space containing all shifts of ω; Definition 2.4.3).

Our results are contained in Chapter 6, and Chapters 1-5 introduce the theorems and definitions required by our proofs, plus some fun examples and motivating ideas.

2. Infinite words and symbolic dynamics

2.1. The one-sided shift

Let $\mathcal{A}^{\mathbb{N}}$ to be the space of sequences over some finite set, $\mathcal{A}$. We call the sequences $\omega \in \mathcal{A}^{\mathbb{N}}$ (infinite) *words*, and we refer to $\mathcal{A}$ as the *alphabet*. We will usually write $\mathcal{A}$ with either letters or numbers, so that $\mathcal{A} \subseteq \{0, \ldots, 9\}$ or $\mathcal{A} \subseteq \{a, b, \ldots, z\}$; an exception to this general rule is when we make use of *higher block encodings* [LM95, §1.4], where we take $\mathcal{A} \subseteq \{0, \ldots, 9\}^n$ or $\mathcal{A} \subseteq \{a, b, \ldots, z\}^n$ (we do this later, when we define *Rauzy graphs*, Definition 5.1.1). The alphabet will always be finite and we will require that $|\mathcal{A}| \geq 2$. We equip the space with the *projective topology* (or *product topology*) on finite words. It has a number of consistent metrics, one of which is

$$d(\omega, \omega') = 2^{-k}, \text{ where } k = \inf\{\, r : \omega_r \neq \omega'_r \,\}.$$

This space is compact, and we can define a continuous self-map $S : \mathcal{A} \circlearrowleft$ on the space by

$$S(\omega) = S(\omega_0\omega_1\omega_2\ldots) = \omega_1\omega_2\ldots.$$

We can view the infinite words ω as functions $\omega : \mathbb{N} \to \mathcal{A}$, and with this interpretation, $S(\omega)(n) = \omega(n+1)$. This is *the left-shift*. With this setup, we define the full one-sided shift space:

Definition 2.1.1: The full one-sided shift space
For any alphabet $\mathcal{A}$, define the *full one-sided shift associated to* $\mathcal{A}$ to be the pair $(\mathcal{A}^{\mathbb{N}}, S)$. That is, $\mathcal{A}^{\mathbb{N}}$ viewed as a topological dynamical system with the (non-invertible) continuous self-map $S : \omega_0\omega_1\omega_2\cdots \mapsto \omega_1\omega_2\ldots$

Remark 2.1.2: Topological properties of the shift space
This topology has a countable basis consisting of the *cylinder sets* $[u] := \{\omega : \omega_i = u_i, 0 \leq i < |u|\}$, where u is a finite word in $\mathcal{A}^*$ (= the *free monoid on* $\mathcal{A}$), and $|u|$ is the length of u. The topology is also separable, and complete, making $\mathcal{A}^{\mathbb{N}}$ a *Polish space* [Kec95]. In addition, this space is totally disconnected, making it isomorphic to the Cantor set.

Subshifts

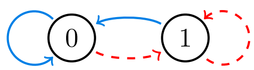
We also have a notion of subsystems, which are called *subshifts*:

Definition 2.1.3: Subshifts
We define a *subshift* of $\mathcal{A}^{\mathbb{N}}$ to be a closed subset $X \subseteq \mathcal{A}^{\mathbb{N}}$ which satisfies that $SX \subseteq X$; that is, it is *closed under the shift.*

Example 2.1.4: Periodic words
A simple example of a subshift is the set $X = \{abcabcabc\ldots, bcabcabc\ldots, cabcabc\ldots\}$. That is, X consists of the three distinct shifts of the periodic word $abcabcabc\ldots$.

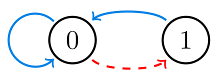
Example 2.1.5: Vertex walks

Many classic examples of shift spaces come from the path spaces of finite graphs. Taking the vertex set of the graph as the alphabet $\mathcal{A}$, *the vertex shift of a graph* is the set of all sequences of vertices where two vertices appear beside one another in an infinite word when there is an (oriented) edge connecting them. Below are a few examples; the edge colors will be explained when we discuss *labelled graphs* (Definition 5.0.2).



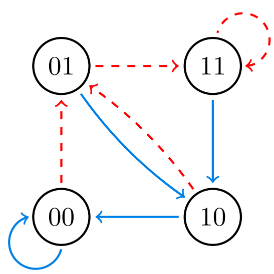
The full shift

Every possible sequence of 0s and 1s is permissible, so the associated subshift is the full shift $\{0,1\}^{\mathbb{N}}$



The golden mean shift

$1 \to 1$ is a forbidden transition, so the associated subshift is a subset of $\{0,1\}^{\mathbb{N}}$ consisting of infinite words which do not contain 11 as a subword.



The second de Bruijn Graph, $\vec{\mathcal{B}}^2$

This is a *Higher Block Encoding* [LM95] of the full shift on $\{0,1\}$. Place a 2 letter window at the start of a word on $\{0,1\}^{\mathbb{N}}$. The alphabet is what can be seen through the window, and the transitions represent words sliding through the window. For example, with the word $\boxed{01}$11010... you can see that a window over 01 would be followed by 11 if you slide the word left. It is clear that the following subwords cannot occur

$$\left\{ \begin{matrix} (00)(11), (10)(11), (00)(10), (10)(10), \\ (01)(00), (01)(01), (11)(00), (11)(01) \end{matrix} \right\}$$

The first example provides a model for the full shift space $\{0,1\}^{\mathbb{N}}$; the second is a proper subshift of $\{0,1\}^{\mathbb{N}}$. The third example is a subshift of $\{00,01,10,11\}^{\mathbb{N}}$.

2.2. A detailed example: symbolic encoding of a dynamical system

Shift spaces are well understood; because of this, it is useful to study more general dynamical systems by relating them to some shift space. In fact, shift spaces originally arose as encodings of more complicated dynamical systems [CN08]. Given a dynamical system (X, T), and some subshift $Y \subseteq \mathcal{A}^{\mathbb{N}}$, we are interested in finding *conjugacies* or *semi-conjugacies* of dynamical systems — that is, isomorphisms (for conjugacies) or homomorphisms (for semi-conjugacies), θ, in the appropriate category[3], with $\theta : Y \twoheadrightarrow X$ such that $\theta \circ S = T \circ \theta$. That is, we require that the following diagram commutes in the chosen category:

$$\begin{array}{ccc} Y & \xrightarrow{S} & Y \\ \theta \big\downarrow & & \big\downarrow \theta \\ X & \xrightarrow[T]{} & X \end{array}$$

By the nature of mapping from a Cantor set[4] into another topological space, it cannot be expected that the map θ be one-to-one (such a restriction would force X to be also be a Cantor set). We do want θ to be onto though, so that we can study dynamics of X through Y, and we can ask that θ be one-to-one on a set of full measure. For now, it suffices to say that we want a topological semi-conjugacy, and in the measure category we may get a *bona fide* conjugacy.

The following example, borrowing from [LR10], is meant to provide a fun but non-trivial example of a (semi-)conjugacy.

Example 2.2.1: An example: non-integer base expansions of real numbers

It is common to think about real numbers via their decimal expansion $x = \sum_i \omega_i 10^{-i}$, however an interesting question is: can we do the same with a more interesting base, such as $\varphi = \frac{1+\sqrt{5}}{2}$, the *golden ratio*? The answer is yes. We can show that every real number $x \in [0, 1]$ has a base φ expansion within the subshift $Y \subset \{0, 1\}^{\mathbb{N}}$ containing all infinite words which do not have the subword 11 (this is the definition of the *the golden mean shift*), and that on this subshift, the base φ expansion is unique except for on the φ-rationals.

RESTRICTING TO A SUBSHIFT

Take $X = \mathbb{R}/\mathbb{Z} \cong [0, 1)$ as a topological group, so that 0 is identified with 1, and take T to be the continuous self-map $Tx = \varphi x \mod 1$. Take $\mathcal{A} = \{0, 1\}$. We make a couple of observations about φ-expansions. Defining, $\theta : \{0, 1\}^{\mathbb{N}} \to \mathbb{R}/\mathbb{Z}$, $\theta : \omega \mapsto \sum_i \omega_i \varphi^{-i-1}$, we have that

$$\theta(101010\ldots) = \frac{1}{\varphi}\left(1 + \frac{1}{\varphi^2} + \frac{1}{\varphi^4} + \ldots\right) = \frac{1}{\varphi}\frac{1}{1 - 1/\varphi^2} = \frac{1}{\varphi}\frac{1}{1/\varphi} = 1.$$

We also have that $\varphi^{-2} + \varphi^{-1} = 1 = \theta(110000\ldots)$. So any word $110^m 1*$ would have that $\theta(\omega) > 1$, and this will cause a large class of numbers to have multiple representations mod 1. We therefore remove all such words that contain 11 as a subword. Also, every word $u110^\infty$ is equivalent (for our purposes) to a word not including 11, since, where u is any word of length k,

$$\theta(u110^\infty) = \sum_{i=0}^{k-1} \frac{u_i}{\varphi^{i+1}} + \frac{1}{\varphi^k}\theta(110^\infty) = \sum_{i=0}^{k-1} \frac{u_i}{\varphi^{i+1}} + \frac{1}{\varphi^k}\theta((10)^\infty) = \theta(u(10)^\infty).$$

[3] For our purposes, either the topological or measured/measurable category

[4] recall that $\mathcal{A}^{\mathbb{N}}$ is homeomorphic to a Cantor set

So words containing 11 either have that $\theta(S^k\omega)$ exits the interval $[0,1]$, or it has another representation. Therefore we restrict to the subshift

$$Y = \{\, \omega \in \mathcal{A}^{\mathbb{N}} : 11 \text{ is not a subword of } \omega \,\}.$$

Within Y, $\theta(101010\dots) = 1$ is the maximum value attained. Also, observing that $\theta(S\omega) = T\theta(\omega)$, we have that the following diagram commutes:

$$\begin{tikzcd} Y \arrow[r, "S"] \arrow[d, "\theta"'] & Y \arrow[d, "\theta"] \\ X \arrow[r, "T"'] & X \end{tikzcd}$$

Next we show that once we restrict to Y, every number $x \in X = \mathbb{R}/\mathbb{Z}$ has a φ-expansion in Y, and expansions are unique except for on the φ-rationals.

θ is onto and nearly one-to-one

We will show that θ is onto by showing that every x has an associated φ-expansion in Y. We must handle two cases separately:

The rational case: Now, as in the case of decimal numbers, where we have that $1.000\cdots = 0.9999\dots$, symbolic representations of (φ-)rational numbers are not unique. If x is φ-rational, then $x = \sum_{i=0}^{k} u_i \varphi^{-i-1}$, $u_i \in \{0,1\}$, and we can show that $u = (u_0, u_1, \dots)$ can be selected so as to not contain 11 as a subword. The case of $x = 0$ and $x = 1$ are already shown, so suppose that $0 < x < 1$. This can be approached algorithmically.

Noting that $\varphi^{-(k+1)} + \varphi^{-(k+2)} = \varphi^{-k}$, the words 100 and 011 are equivalent when replaced in u, and since u has finite length, there is a greatest index for which we witness an occurrence of 11. Because $x < 1$, we know that u cannot begin with 11, so it begins with 01, 10, or 00, in every case, if there is a first occurrence of 11, it must be preceded by a 0 because of this. We read the word from left to right, applying our substitution $011 \to 100$ on every occurrence of the word as we cross it, possibly creating new occurrences of 11 behind us. At the end of one iteration, we have reduced the index of the **last** occurrence of 11 by at least 2. Further, we have created an equivalent word to u, so our claim that the new word does not begin with 11 still holds. We can therefore iterate this substitution again, each time decreasing the index last occurrence of 11, until there are no remaining occurrences.

Irrational case: We can approximate every φ-irrational x with a finite sequence not including 11 to an arbitrary precision — it follows that the infinite expansion does not contain 11 as a subword, as it would have to appear in some finite approximation. Suppose that $0 < x < 1$ and that x is φ-irrational:

The algorithm: Start with (x, ω_0)

$x < 1/\varphi$	If $x < 1/\varphi$, then set $\omega_i := 0$ (and note that $Tx < 1$). Recursively compute later terms of the expansion by applying this algorithm to (Tx, ω_{i+1}).

$x > 1/\varphi$ | If $1/\varphi < x < 1$, set $\omega_i := 1$. Then, note that $1/\varphi = \varphi - 1$ and also that since we are working modulo 1, that $Tx \equiv Tx - 1$. From these two facts we get that

$$\frac{1}{\varphi} < x < 1 \Rightarrow 1 < Tx < \varphi \Rightarrow 0 < Tx - 1 < \varphi - 1 \Rightarrow 0 < Tx < \frac{1}{\varphi}$$

So that every $\omega_i = 1$ implies that $\omega_{i+1} = 0$ in the φ-expansion of x. This shows that the expansion lives in Y, since 11 does not occur in it. We can now recursively compute ω_{i+2} by application of this algorithm to (T^2x, ω_{i+2}).

The expansion is unique, because $\theta^{-1}[\{0\}] = \{000000\ldots, 101010\ldots\}$, and so the **only** points with non-unique expansions are the φ-rationals.

Revisiting our old diagram, we get that with the exception of the φ-rationals, θ is now onto *and almost one-to-one*[5]

$$\begin{tikzcd} Y \arrow[r, "S"] \arrow[d, "\theta"] & Y \arrow[d, "\theta"] \\ X \arrow[r, "T"] & X \end{tikzcd}$$

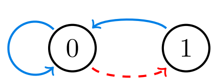
It fails to be one-to-one on the dense subset of φ-rationals, but this is the price we pay for mapping a totally disconnected space into a connected one. However, in the measure category, the φ-rationals constitute almost none of the space, so that as measure theoretic dynamical systems (with Lebesgue measure), this is actually a conjugacy.

The Markov partition

It can be checked that the algorithm described above is equivalent to defining the following "partition"

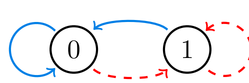
$$P_0 := \left[0, \frac{1}{\varphi}\right], \; P_1 := \left[\frac{1}{\varphi}, 1\right], \; l(x) = \begin{cases} 0 & x \in P_0 \\ 1 & x \in P_1 \end{cases}$$

and (discarding the φ-rationals) taking the sequence $(l \circ T^k x)_{k \in \mathbb{N}}$. That is, Y is the supporting subshift of the *Markov partition* [BKS91] formed by $\{P_0, P_1\}$. Also, it is not an accident that Y, also known as *the golden mean shift*, arises as the path space of



with the transition matrix $\begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}$

which has φ as its positive eigenvalue. It is interesting to compare and contrast this with the matrix $\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$, with the eigenvalue 2, with the associated graph



which has every binary sequence as a permissible path.

[5] it is one-to-one on a set of full measure when we equip Y with the measure $\mu(A) = \lambda(\theta[A])$, where λ is the Lebesgue measure on $\mathbb{R}/\mathbb{Z} \cong [0, 1)$. Since θ is onto, and since θ is a (measurable) one-to-one and onto map between subsets of full measure, it is actually a measure isomorphism $\theta : (Y, \mu) \to ([0, 1), \lambda)$.

2.3. Two-sided symbolic systems

The example of multiplication in $[0,1)$ by φ of the previous section was an example of a *topological dynamical system* — that is, a topological space X paired with a continuous transformation $T : X \circlearrowleft$. The Markov Partition [BKS91] gives an example of the use of the symbolic encodings of topological dynamical systems. It is more generally true that all non-invertible "topologically irreducible" topological dynamical systems (that is, *minimal systems* — Definition 2.4.1) on compact metric spaces arise as factors of a symbolic system $(\mathcal{A}^{\mathbb{N}}, S)$, and many more systems also have a symbolic realization.

It is natural then to ask about symbolic models for invertible systems.

Definition 2.3.1: The full two-sided shift space
Let $\mathcal{A}^{\mathbb{Z}}$ be the space of bi-infinite sequences, which we will also call infinite (or bi-infinite) words. It is again equipped with the projective topology (= product of discrete topology). •

As before, this space is compact and has many compatible metrics, one of which being $d(\omega, \omega') = 2^{-k}$, where $k = \inf\{\ |r| : \omega_r \neq \omega'_r \text{ or } \omega_{-r} \neq \omega'_{-r}\ \}$. The space is compact, Polish, and totally disconnected. However, the (left) shift transformation S is now a homeomorphism. Subshifts of this space are defined analogously in the one-sided case; we define a **(two-sided) subshift** of $\mathcal{A}^{\mathbb{Z}}$ to be a subset $X \subseteq \mathcal{A}^{\mathbb{Z}}$ which is closed and which satisfies that $SX = X$. This makes the two-sided orbit of every $x \in X$ remain within X. It is worth noting that $\mathcal{A}^{\mathbb{Z}}$ is **marked** in a natural way, by the zero index. However, the two-sided system that we are building towards is a space of *paths emanating away from, but not including a mark.* So, instead of marking the zero index

$$\dots \omega_{-2}\omega_{-1}\overset{\downarrow}{\omega_0}\omega_1\omega_2 \dots$$

we consider marked bi-infinite words, $\mathcal{A}^{-\mathbb{N}} \times \mathcal{A}^{\mathbb{N}}$.

$$\dots \omega_{-2}\omega_{-1}\omega_{-0} \overset{\downarrow}{\times} \omega_0\omega_1\omega_2 \dots$$

This space is not different in any interesting way, but it will be convenient for us. We will call this the space of *marked bi-infinite sequences.*

Definition 2.3.2: Marked bi-infinite sequences
We define the space of marked bi-infinite sequences to be the set $\mathcal{A}^{-\mathbb{N}} \times \mathcal{A}^{\mathbb{N}}$ (we think of $\mathbb{N} =: +\mathbb{N}$ now as $\{(+,0),(+,1),\dots\}$ and $-\mathbb{N} = \{(-,0),(-,1),(-,2),\dots\}$, so that $\mathbb{N}$ and $-\mathbb{N}$ are disjoint) equipped with the metric $d(\omega,\omega') = 2^{-k}$, where $k = \inf\{\ |r| : \omega_r \neq \omega'_r \text{ or } \omega_{-r} \neq \omega'_{-r}\ \}$ and with the shift operator defined by $S(\omega) = \omega_{-2}\omega_{-1}\omega_{-0}\omega_0 \times \omega_1\omega_2 s$. S is again a homeomorphism of the space. This dynamical system $\left(\mathcal{A}^{-\mathbb{N}} \times \mathcal{A}^{\mathbb{N}}, S\right)$ is conjugate to $(\mathcal{A}^{\mathbb{Z}}, S)$. •

2.4. Minimality and systems associated to a word

Most mathematical objects tend to have a class of irreducible examples which form building blocks for larger objects, which is why one studies objects such as simple (Lie) groups or irreducible representations. For topological dynamical systems the appropriate definition of irreducibility is the absence of a (non-empty) proper subsystem, and the name of this property is *minimality.* For our purposes we only consider the case where X is a compact metric space, and we then have the following definition:

Definition 2.4.1: Minimality
Let X be a compact metric space. A topological dynamical system (X, T) is *minimal* if the only closed subsets $Y \subseteq X$ which are closed under T (that is $TY \subseteq Y$) are the empty set and X.

Remark 2.4.2: Every compact (X, T) has a minimal subset
A priori one might imagine that there could exist decreasing chains of smaller systems with no smallest subset, but the Cantor intersection theorem gives that this cannot be the case. Let $\{A_i\}_{i \in I}$ be any set of subsystems, then the A_i are closed (and hence compact), and for any decreasing chain $A_i \supset A_{i+1}$, $\bigcap_i A_i$ is non-empty, closed, and T-invariant. We get that every decreasing chain has a lower bound, so Zorn's lemma gives that there is a minimal set. This proof is well known, and can be found, for instance, in [BS02, Prop 2.1.2].

Minimality and orbits

A useful observation is that *the orbit of* x, $\mathcal{O}(x) = \left\{T^k x : k \in \mathbb{N}\right\}$, is closed under T and non-empty, so by the definition of minimality the orbit must be dense in X [FM10], else it admits a proper non-empty subsystem. In fact, it is equivalent to minimality that the orbit of every point is dense, so that $X = \overline{\mathcal{O}(x)}$ for every $x \in X$.

Since every orbit of a given point in a minimal system is dense, every minimal dynamical system can be realized as $X = \overline{\mathcal{O}(x)}$. Because of this, in the context of symbolic dynamics, it makes sense to discuss the system associated to a word.

Definition 2.4.3: The shift space associated to a word
Let ω be an element of $\mathcal{A}^{\mathbb{N}}$, $\mathcal{A}^{\mathbb{Z}}$, or $\mathcal{A}^{-\mathbb{N}} \times \mathcal{A}^{\mathbb{N}}$. We call the *shift space of* ω or the *symbolic system associated to* ω the system (X_ω, S), where $X_\omega := \overline{\mathcal{O}(\omega)}$, where the orbit is one-sided if $\omega \in \mathcal{A}^{\mathbb{N}}$ and two-sided otherwise.

Words and non-minimal subshifts

Note that, while all minimal systems can be realized $(\overline{\mathcal{O}(x)}, T)$, the converse is not true — not every $(\overline{\mathcal{O}(x)}, T)$ is minimal, as the following examples show:

Example 2.4.4: Word generating the full-shift
We can construct a word which generates the full shift. For instance, take the following word, known as the *Champernowne word* [CN10, Example 4.2.3], (the product $\prod$ is concatenation of finite words)

$$\omega_{\mathcal{A}} := \prod_n \left(\prod_{u \in \mathcal{A}^n} u \right)$$

For example, $\omega_{\{0,1\}} = (0)(1)(00)(01)(10)(11)(000)\ldots$. That is, take all subwords of finite length and concatenate them all in some order, making the words longer on each iteration. For every $x \in \mathcal{A}^{\mathbb{N}}$ there is a sequence (n_k) such that $S^{n_k}\omega \to x$, simply because you can choose n_k to select the arbitrarily long *prefixes* of x (a *prefix* is a word $u = x_1 x_2 \dots x_k$). So $X_\omega = \mathcal{A}^{\mathbb{N}}$. This system contains many subsystems, such as periodic words, and so it is not minimal. Amusingly, uniformly-almost-all x from $\mathcal{A}^{\mathbb{N}}$ will also have that $X_x = \mathcal{A}^{\mathbb{N}}$, because every finite word will occur in x eventually with probability 1. •

Example 2.4.5: A countable X_ω
This next example from (adapted from [LM95, Example 1.2.6]) is closer to being minimal, but it still fails. Take the infinite word

$$\omega = \prod_{n\in\mathbb{N}} 0^n 1 = 101001000100001\dots$$

Since arbitrarily long stretches of zeros occur in the ω, the infinite-zero word 0^∞ is in the closure of the orbit. But $\{0^\infty\}$ by itself is closed and closed under T, so X_ω has a proper non-empty subshift, and consequently is not minimal. However, X_ω has exactly one minimal subshift, which is just $\{0^\infty\}$ — as mentioned, if X is minimal then $X = \overline{\mathcal{O}(x)} =: X_x$ for any $x \in X$. But, for any x other than 0^∞, $X_{Sx} \subset X_x$ is easily shown to be a proper, closed, and non-empty subset. So $X_x \subseteq X_\omega$ is only minimal for $x = 0^\infty$, and $X_{0^\infty} = \{0^\infty\}$. (An alternative and more interesting proof can be found with Proposition 3.2.6 and [Oxt52, (5.2)], but this requires tools from ergodic theory, which we have to wait until the next chapter for.) •

A criterion for minimality

While not all words induce minimal systems, there is a simple condition on ω which is equivalent to minimality of X_ω called *uniform recurrence*. First we introduce a notation for *the language of a collection of words*

Definition 2.4.6: The language of a collection of words
The language of a set of words X, denoted $L(X)$, is simply the collection of all finite words which occur in any infinite word in X. That is

$$\mathcal{A}^* \supseteq L(X) := \{u \,:\, \exists_{\omega\in X}.\ u = \omega_k \dots \omega_{k+n-1},\ k \in \mathbb{Z},\ n \in \mathbb{N}\}$$

and $L_n(X) \subset L(X)$ simply refers to the subwords of length n (Another name for finite-length subwords is *factors* [CN10]). For subshifts generated by an infinite word, $L(X_\omega)$ is simply equal to the set of subwords of ω, so we will often refer to the language of a word, taking $L(\omega) := L(X_\omega)$. •

With that little bit of notation, we have that the following condition on ω is equivalent to X_ω being minimal. [FM10]

Definition 2.4.7: Uniform recurrence
An infinite word ω is called *uniformly recurrent* if for every subword u of ω, there is some length n such that u occurs in every word in $L_n(\omega)$. Equivalently, every word u occurs infinitely often in ω, and the gaps between occurrences of u are bounded. If a subword occurs infinitely often, but without bounded gaps, then it is simply *recurrent*. •

Trivial examples of such words are periodic words. A more interesting example is the Thue-Morse word.

Example 2.4.8: The Thue-Morse word
The Thue-Morse word[6], $\mathfrak{t}$, the fixed point of the *substitution* (defined in Definition 4.0.4) $\sigma : \{0,1\}^* \circlearrowleft$ beginning with 0 (we extend the definition of σ to $\{0,1\}^{\mathbb{N}}$).

$$\sigma : \begin{cases} 0 \mapsto 01 \\ 1 \mapsto 10 \end{cases}, \; \mathfrak{t}_0 := 0, \; \sigma(\mathfrak{t}) = \mathfrak{t}.$$

Alternatively, if we recall that $\mathcal{A}^{\mathbb{N}}$ was given the projective topology, we can the infinite Thue-Morse word as a (projective) limit of finite words, where a sequence of finite words converges if it is eventually constant on prefixes of every length. The limit word is then the infinite word which agrees with all of these limit prefixes. With this interpretation of limit, the Thue-Morse word can be defined as:

$$\mathfrak{t} := \lim_n \sigma^n(0) = \lim(0, 01, 0110, 01101001, \dots) = 0110100110010110\dots$$

It is a famous fact that neither 000 nor 111 occur in $\mathfrak{t}$. To see this, take all three extensions of 111, $\{0111, 1110, 1111\}$, and observe that none of these are the image of any two letter word under σ (ditto for 000). Using this, note that every word $u \in L_n(\omega)$ occurs in some initial block B of length 2^k, and $B = \sigma^k(0)$. Letting $\overline{B} = \sigma^k(1)$, we have that

$$\mathfrak{t} = 01101001 \cdots = \sigma^k(\mathfrak{t}) = \sigma^k(0)\sigma^k(1)\sigma^k(1)\sigma^k(0) \cdots = B\overline{B}\overline{B}B\overline{B}BB\overline{B}\dots$$

Since 111 never occurs, every block of length $4 \cdot 2^k$ must contain one of $B\overline{B}B, \overline{B}BB, BB\overline{B}, \overline{B}\overline{B}B, \overline{B}B\overline{B}$, or $B\overline{B}\overline{B}$; that is, every word of this length contains a copy of B, and thus a copy of u. So $\mathfrak{t}$ is uniformly recurrent.

So far we have focused on topological properties of these symbolic systems, but as we mentioned in the example of representations of $x \in [0,1)$ in base φ, it is common that the symbolic representations of systems is better behaved in the measure category than in the topological category. This leads us to study some results from ergodic theory.

[6] This word was actually discovered at least three times; the earliest known discovery was by Prouhet in 1851 [LM95]. However it was popularized by Morse.

3. Ergodic theory of compact metric spaces

3.1. The theory of Kryloff-Bogolioúboff

Ergodic theory is the study of dynamical systems in the category of measure spaces, which is often a more pleasant category, as isomorphism is only considered modulo sets of measure zero [Bil78, Chapter 2]. In the measure category every compact polish space with a continuous probability measure is measure isomorphic to the Cantor set, which is itself isomorphic to $[0,1]$ with Lebesgue measure [Kec95, Thm. 17.41].[7]

While the general considerations of ergodic theory are broad, we are interested in the case of dynamics of a compact metric space X with a continuous self-map $T : X \circlearrowleft$, lifted to the measure category. That is, we are interested in X equipped with its *Borel σ-algebra*, $\mathcal{B}$ — the smallest σ-algebra which makes all open sets measurable. The continuous self-map $T : X \circlearrowleft$ is measurable with respect to this σ-algebra, since it pulls-back open sets to open sets, which generate the σ-algebra. We are interested in the different T-invariant probability measures that we can equip this space (X, T) with. That is, we require μ to satisfy that for any measurable set $A \in \mathcal{B}$, that $\mu(T^{-n}A) = \mu(\{\, x \mid T^n x \in A \,\}) = \mu(A)$ (this can be thought of as forward time-invariance of the distribution [BKS91, §2.2.1]). For our purposes, μ will always be a probability measure, so that $\mu(X) = 1$.

The theorem of Kryloff & Bogolioúboff

An obvious first question, given the above setup, is whether or not a system (X, T) necessarily has any T-invariant measure. This was studied and answered by Kryloff and Bogolioúboff [KB37], and the theory is summarized concisely in an monograph of Oxtoby [Oxt52].

We first define $\mathcal{M}(X)$, the space of Borel probability measures on X. That is

$$\mathcal{M}(X) = \{\, \mu : \mathcal{B} \to [0,1] \, : \, \mu(X) = 1 \}$$

It was shown by Riesz that $\mathcal{M}(X)$ can be identified with the positive part[8] of the unit sphere of $C(X, \mathbb{R})^*$, and $\mathcal{M}(X)$ is then equipped with the weak-* topology via this identification. This means that $\mu_n \to \mu$ iff for every continuous f we have that $\int f \, d\mu_n \to \int f \, d\mu$.

In 1937 Kryloff and Bogolioúboff showed that this space is sequentially compact. That is, for any sequence of measures (μ_n), there is a subsequence $\mu_{n_k} \to \mu$ [KB37, Théorème I]. They proved this directly using detailed measure-theoretic arguments, but a more common argument now uses Banach's theorem (which may not have been known to Kryloff and Bogolioúboff at the time) that if X is separable[9], then the unit ball in $C(X, \mathbb{R})^*$ is compact [Oxt52, §2]. Kryloff and Bogolioúboff showed sequential compactness directly, but since $C(X, \mathbb{R})^*$ is weak-* metrizable (when X is compact metric), sequential compactness follows from compactness [Kec95, Prop. 4.2].

This result culminates in what is known as *the Kryloff-Bogolioúboff theorem.*

[7] For this reason, it is sometimes said informally in ergodic theory that there is only one measure space.
[8] In the sense that $\varphi \in C(X, \mathbb{R})^*$ is *positive* if $f \geq 0 \Rightarrow \varphi(f) \geq 0$
[9] Alaoglu showed that the separability assumption can be dropped.

Theorem 3.1.1: The Kryloff-Bogoliouboff theorem
If X is a compact metric space and $T : X \circlearrowleft$ is continuous, then X has a T-invariant probability measure.

Proof: Let X be a compact metric space. Since $\mathcal{M}(X)$ is sequentially compact, take any $x \in X$, and the sequence

$$\left(\frac{1}{n}\sum_{i=0}^{n-1}\delta_{T^i(x)}\right)_n$$

has a weak-* limit point in $\mathcal{M}(X)$ and the limit point is T-invariant. (Here δ_x is the point mass at x.) □

THE SPACE OF INVARIANT MEASURES

Knowing that every system (X, T) has at least one invariant measure, it is natural to study the collection of invariant measures.

Definition 3.1.2: The T-invariant Borel measures, $\mathcal{M}(X, T)$
We define $\mathcal{M}(X, T) \subset \mathcal{M}(X)$ as

$$\mathcal{M}(X,T) := \left\{\mu \in \mathcal{M}(X) \,:\, \forall A \in \mathcal{B}(X).\, \mu(A) = \mu(T^{-1}A)\right\}$$

We claim that $\mathcal{M}(X, T)$ is also non-empty, convex, and closed (and since $\mathcal{M}(X)$ is compact, $\mathcal{M}(X, T)$ is compact). Non-emptiness is simply a restatement of Kryloff-Bogoliouboff, and convexity is obvious. For showing that the subset of T-invariant measures is closed, note that if μ_n is a convergent sequence of T-invariant measures, then

$$\begin{array}{ccc} \int f\, d\mu_n & \longrightarrow & \int f\, d\mu \\ = & & \\ \int (f \circ T) d\mu_n & \longrightarrow & \int (f \circ T) d\mu \end{array}$$

So the limit measure μ is also T-invariant.

Among the invariant measures are a special class, known as ergodic measures, which —as will be seen shortly— are in a sense "irreducible" measures.

Definition 3.1.3: Ergodicity
For a system (X, T), where X is any measure space and $T : X \circlearrowleft$ is measurable, a measure μ is called *ergodic* or *ergodic with respect to T* if the Borel sets A which satisfy

$$\mu(A \Delta T^{-1}A) = 0, \text{ where } A\Delta B = (A \setminus B) \cup (B \setminus A)$$

have either either measure zero or full measure. This condition says that if A is *essentially T-invariant*, ($A = T^{-1}A$ except for on a set of μ-measure 0) then either A is essentially the whole space or is a null set [Pet83, §2.4] [BS02, §4.3][10].

[10] Sometimes another definition of T-invariance is used [FM10, §7.2.1], which is that $T^{-1}A = A$, but that definition is less robust in handling measure 0 issues. The above definition is morally the same, but is preferable for these technical reasons. This is explained in [Bil78, page 8].

If μ is ergodic with respect to $T : X \circlearrowleft$, then we say that the system (X, T, μ) is ergodic; when this is the case, ergodicity gives that there are no non-trivial T-invariant up to measure zero subsets, and hence there are no $Y \subsetneq X$ with $\mu(X \setminus Y) > 0$ for which $\frac{\mu|_Y}{\mu(Y)}$ is a an invariant measure on (Y, T). This is related to the fact that distinct ergodic measures are mutually singular, which is a sense in which the ergodic measures can be thought of as "irreducible".

Ergodic measures as extreme points

Since $\mathcal{M}(X, T)$ is compact and convex, (and since $C(X, \mathbb{R})^* \supset \mathcal{M}(X, T)$ is a locally-convex topological vector space) the Krein-Milman theorem gives that $\mathcal{M}(X, T)$ is equal to the closed convex hull of its extreme points [KM40, Thm 1]. Moreover, the extreme points of $\mathcal{M}(X, T)$ are exactly the ergodic measures, as was observed by Choquet. Also, denoting by $\mathcal{E}(X, T) \subseteq \mathcal{M}(X, T)$ the ergodic measures, the Choquet theorem [Cho58, §9] [Phe01, Prop 1.2] gives that every T-invariant measure can be given as a integral over $\mathcal{E}(X, T)$ with respect to some probability measure on $\mathcal{E}(X, T)$. That is,

$$\forall \mu \in \mathcal{M}(X, T).\, \mu = \int_{\nu \in \mathcal{E}(X,T)} \nu \, dP$$

This *effectively* says that every T-invariant measure is a convex combination of ergodic measures — if $|\mathcal{E}(X, T)|$ is finite, then this is literally true, and in the infinite case a measure might be a limit of convex combinations of ergodic measures (which, after working through details, is an integral with respect to a probability measure on $\mathcal{E}(X, T)$).

3.2. Ergodic theorems

Ergodic systems are exactly the systems which satisfy "the ergodic hypothesis" from thermodynamics, which says that for these systems "*time averages converge to space averages*" [BK32, §1] [Wie39, §1]. A famous theorem which makes this idea more precise is the Birkhoff ergodic theorem [Bir31] [Hal56, §Pointwise Convergence].

Theorem 3.2.1: The Birkhoff ergodic theorem
Let (X, μ) be a probability space and let $T : X \circlearrowleft$ be measure-preserving. Then if f is integrable, the following limit exists μ-almost-everywhere,

$$f^*(x) := \lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} f\left(T^k(x)\right),$$

and

$$\int_X f^* \, d\mu = \int_X f \, d\mu;$$

finally, if (X, T, μ) is ergodic, then μ-almost-all orbit-averages are equal, so that $f^*(x)$ is constant (except possibly on a null set) and equal to $\int f \, d\mu$. •

To illustrate the use of this theorem, we give a simple application of the Birkhoff ergodic theorem and give a short proof of the Borel normal number theorem [Bil78, p. 15].

Example 3.2.2: The Borel normal number theorem
A real number $x \in [0,1]$ is called *normal in base* 2 if all base 2 expansions of x have an equal asymptotic-relative-frequency[11] of 0s and 1s. Of course, the dyadic-rationals do not have unique expansions, but the rationals are of Lebesgue measure 0 anyway, and with them removed the expansion of x is unique. To show this, we first pass from $[0,1]$ to the symbolic realizations in $\{0,1\}^{\mathbb{N}}$. As mentioned earlier, $[0,1] \subset \mathbb{R}$ with Lebesgue measure is isomorphic to $\{0,1\}^{\mathbb{N}}$ —the Cantor set— equipped with the S-invariant measure determined by $\mu([0]) = \mu([1]) = 1/2$, and in fact $\left(2^{\mathbb{N}}, \mu, S\right)$ is ergodic[12]. As a result of this, by the Birkhoff ergodic theorem we get that for μ-almost-every sequence ω in $2^{\mathbb{N}}$, that

$$\lim_n \frac{1}{n} \sum_{k=0}^{\infty} I_{[0]}(S^k(\omega)) = \int I_{[0]} d\mu = \mu([0]) = \frac{1}{2}$$

which says that the frequency of 0s in μ-almost-every ω is $\frac{1}{2}$. Applying the measure isomorphism between $([0,1], \lambda)$ and $(2^{\mathbb{N}}, \mu)$, we conclude that Lebesgue-almost-every real number is normal in base 2.

The above proof works *mutatis mutandis* for any integer $b \geq 2$, and so one gets that for every base b, the set A_b of numbers in $[0,1]$ normal in base b have full Lebesgue measure in $[0,1]$. Since the countable intersection of sets of full measure is again of full measure, we get that almost every real number is normal with respect to every base *simultaneously*. This is the Borel normal number theorem. •

Ergodic sets

The Birkhoff ergodic theorem can also give some insight into the structure of $\mathcal{M}(X,T)$. While $\mathcal{M}(X,T)$ is compact and convex in the weak-* topology, it can be unruly in the sense that it can have uncountably many extremal points (= ergodic measures). For instance, if we consider the full-shift $\{0,1\}^{\mathbb{N}}$, it turns out that the S-invariant measure determined by $\mu([0]) = p, \mu([1]) = 1-p$, is ergodic with respect to S for any p — that is, every *Bernoulli measure* is ergodic. One immediately has uncountably many ergodic measures, and they are concentrated on disjoint (dense) S-closed subsystems of $\{0,1\}^{\mathbb{N}}$. To see this, take μ, ν to be distinct Bernoulli measures μ and ν and notice that by the Birkhoff ergodic theorem we have that

$$\text{for } \mu\text{-almost-all } \omega.\ \lim_n \frac{1}{n+1} \sum_{i=0}^{n} I_{[0]}\left(S^i \omega\right) = \int I_{[0]} d\mu = p$$

$$\text{for } \nu\text{-almost-all } \omega.\ \lim_n \frac{1}{n+1} \sum_{i=0}^{n} I_{[0]}\left(S^i \omega\right) = \int I_{[0]} d\nu = q$$

Seeing the left side of these equations, it is immediately obvious then that these ergodic measures cannot share any generic points ω, and moreover it is clear that these measures μ and ν are concentrated on the S-closed set of points with an asymptotic-relative-frequency of 0s being equal to p or q, respectively (these Borel sets are neither open nor closed, and they are dense, but they are Borel [Oxt52, (2.2)]). It is true more generally that the ergodic measures are concentrated on disjoint S-closed subspaces (it is true for any

[11] The asymptotic-relative-frequency is the limit of finite ratios, which after examination is just an ergodic average (which is a limit of finite averages).

[12] It is easily shown to be *mixing* which is stronger than ergodicity

measured dynamical system). Because of this, all distinct ergodic measures are mutually singular, and the ergodic measures μ are actually characterized in $\mathcal{M}(X,T)$ by having no measures such that $\nu \ll \mu$ other than $\mu = \nu$ [DGS76, Prop 5.4].

Unique ergodicity

While in the generic case there may be many ergodic measures, there is a special situation, which is similar to minimality (**but not the same**. See [NS60, §VI.9, p. 511] or [Oxt52, p. 134]), in which there is only a single ergodic measure that can be placed on the topological system (X,T). Such a system is called *uniquely ergodic*.

Definition 3.2.3: Uniquely ergodic
When X is a compact metric space, (X,T) is called *uniquely ergodic* if $|\mathcal{M}(X,T)| = 1$, or equivalently if there is a single measure with which to equip $(X, \mathcal{B}(X), \cdot, T)$. The measure is necessarily ergodic, since the single measure is extremal in $\mathcal{M}(X,T)$. •

A criterion for unique ergodicity

Returning to our main focus, which is effectively dynamics of compact metric spaces, we have another useful ergodic theorem due to Oxtoby [Oxt52, (5.3)]

Theorem 3.2.4: The Uniform ergodic theorem
If X is a compact metric space and if $T : X \circlearrowleft$ is a continuous self-map, then the following are equivalent.

(1) (X,T) is uniquely ergodic.

(2) For every continuous map $f \in C(X,\mathbb{R})$, there exists a constant $c(f)$ such that the ergodic averages converge to $c(f)$ *uniformly in x*. That is

$$\lim_{n\to\infty} \sup_{x\in X} \left| \frac{1}{n+1} \sum_{i=0}^{n} f(T^i x) - c(f) \right| = 0.$$

A posteriori, if 2. holds, $c(f) = \int f d\mu$, where μ is the unique ergodic measure on (X,T). •

When we restrict ourselves to symbolic systems associated to a word, a reformulation of the above gives us the following:

Definition 3.2.5: Uniform frequencies
Let ω be a singly infinite word; we denote by $|\omega_k\omega_{k+1}\dots\omega_{k+n}|_u$ the number of occurrences of u within a finite stretch of ω. That is,

$$|\omega_k\omega_{k+1}\dots\omega_{k+n}|_u := \sum_{i=0}^{n-|u|} I_u(S^{i+k}\omega)$$

a word ω is said to have *uniform frequencies* if for every u there exists a constant $\mathrm{freq}_u(\omega)$ such that

$$\lim_n \sup_k \left| \frac{|\omega_k\omega_{k+1}\dots\omega_{k+n}|_u}{n+1} - \mathrm{freq}_u(\omega) \right| = 0$$

An analogous definition holds for two sided words, but as is typical, one has to be careful not to take a symmetrically expanding neighborhood, and must consider all possible limits in both m and n

$$\lim_{m,n} \sup_k \left| \frac{|\alpha_{-m}\dots\alpha_{-0}\times\alpha_0\dots\alpha_n|_u}{m+n+2} - \mathrm{freq}_u(\omega) \right| = 0,\ \ \alpha = S^k\omega$$

But if this limit exists independently of the sequence (m_i, n_i), then we say that the doubly infinite word has uniform frequencies.

Proposition 3.2.6: Uniform frequencies imply unique ergodicity

A word ω has uniform frequencies if and only if (X_ω, S) is uniquely ergodic.

Proof: Note here that $\left\{S^k\omega \,\middle|\, k \in \mathbb{N}\right\}$ is dense in X_ω, so taking a supremum over the orbit of ω is equivalent to taking a supremum over X_ω.

If this limit exists, unpacking the definition shows that this is just an ergodic average

$$\begin{aligned}
\mathrm{freq}_u(S^k\omega) &:= \lim_n \frac{|\omega_k\omega_{k+1}\dots\omega_{k+n}|_u}{n+1} \\
&= \lim_n \frac{1}{n+1}\sum_{i=0}^{n-|u|} I_u(S^{i+k}\omega) \\
&= \lim_n \left(\frac{1}{n+1}\sum_{i=0}^{n} I_u(S^{i+k}\omega) - \frac{1}{n+1}\sum_{i=n-|u|+1}^{n} I_u(S^{i+k}\omega) \right) \\
&= \lim_n \frac{1}{n+1}\sum_{i=0}^{n} I_u(S^{i+k}\omega)
\end{aligned}$$

The collection $\{I_u\}_{u\in\mathcal{A}^*} \subset C(X,\mathbb{R})$ generates a subalgebra which contains all the constant functions and separates points, so by the Stone-Weierstraß theorem it forms a dense subalgebra — as a result, the existence of uniform frequencies is sufficient to conclude condition 2. of the uniform ergodic Theorem. Applying the theorem, we see that if this uniform limit exists, it is equal to $\int I_u d\mu = \mu([u])$, where μ is the unique ergodic probability measure. In the other direction, if the system is uniquely ergodic then the word has uniform frequencies. □

Unique ergodicity is topological

Remark 3.2.7: The Jewett-Krieger theorem

The Jewett-Krieger theorem shows that any ergodic measure-theoretic system $(X, \mathcal{F}, \mu, T)$ has some uniquely ergodic topological model. That is, there is some topological space Y with a Borel measure ν and $S : Y \circlearrowleft$ such that $(Y, \mathcal{B}, \nu, S) \cong (X, \mathcal{F}, \mu, T)$, and such that (Y, S, ν) is uniquely ergodic **and** there are also choices of topological spaces such that (Y, S, ν) is not uniquely ergodic.

The consequence of this, is that unique-ergodicity is not something intrinsic to a measure-theoretic system, but to a topological system.[13]

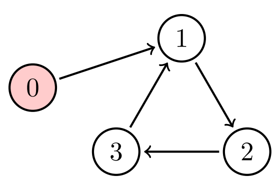
Unique ergodicity almost implies minimality

If a system (X, T) is uniquely ergodic, then (X, T) has only one minimal subsystem; this is simply a consequence of Kryloff-Bogolioubоff Theorem 3.1.1, as if $Y_1, Y_2 \subset X$ were disjoint closed T-invariant subsets, then we would have at least one measure μ_1 with $\mu_1(Y_1) = 1,\ \mu_1(Y_2) = 0$, and similarly such a μ_2 supported on Y_2 which assigns no mass to Y_1. So we would have that $|\mathcal{M}(X, T)| \geq 2$, which would contradict unique ergodicity. We will restate this as a proposition.

Proposition 3.2.8: Unique ergodicity almost implies minimality

If (X, T) is uniquely ergodic, then (X, T) has exactly one minimal subsystem Y (it may be that $Y = X$), and for the unique ergodic measure μ, $\mu(Y) = 1$.

However it is possible for this minimal subsystem to be proper, as can be seen with an example as basic as a small Topological Markov chain:



$\{0123123123\ldots, 123123\ldots, 23123123\ldots, 3123123\ldots\}$

This system is not minimal because it fails to even be recurrent (0 only appears once). However it has a single minimal subsystem, which is the support of the unique ergodic measure on it. The ergodic measure in this case is the uniform measure on $\{123\ldots, 231\ldots, 312\ldots\}$.

Another example is the word $\prod_n 0^n 1$ — the ergodic measure on the infinite subshift associated to this word is concentrated on the perodic word 0^∞ in the orbit closure of the word. A final example, which is less trivial, is the following

[13] Jewett showed this for weakly mixing systems in [Jew70] and Krieger generalized the result to ergodic sytems in [Kri72]. Some of this history can be found in [DGS76].

Example 3.2.9: Modified Thue-Morse word

An example extracted from [FM10, §7.4.2] simply takes an infinite word known to have uniform frequencies, such as the Thue-Morse word (the explicit measure has been constructed by Dekking [Dek92]), and insert a word which does not otherwise occur in it with unbounded gaps. For instance, embed the Thue-Morse word on $\{a, b\}$ into $\{a, b, c\}^{\mathbb{N}}$, and take

$$\mathfrak{t}' := \prod_{n} (\mathfrak{t}_{2^n-1} \dots \mathfrak{t}_{2^{n+1}})\, c.$$

Inserting c in this way (along a sequence with density zero) ensures that the frequency of c (or any word containing c) is zero. For words not containing u, $\mathrm{freq}_u(\mathfrak{t}') = \mathrm{freq}_u(\mathfrak{t})$, so frequencies are well defined, but $\mathfrak{t}'$ is no longer uniformly recurrent. However, $X_{\mathfrak{t}}$ appears within $X_{\mathfrak{t}'}$ (it is easily checked that $\mathfrak{t}$ arises as a limit point in $X_{\mathfrak{t}'}$), and it is the only minimal subset.

This leads to slight variation of uniquely ergodic systems,

Definition 3.2.10: Strict ergodicity

A system (X, T) is called *strictly ergodic* if it is uniquely ergodic and minimal.

3.3. Minimal non-uniquely ergodic systems

We have already seen that uniform frequencies *nearly* imply uniform recurrence, but what about the converse? It turns out that despite the vast majority of minimal systems that are encountered “in the wild” being uniquely ergodic, there exist minimal systems with arbitrarily many ergodic measures. Though such *minimal non-uniquely ergodic* systems are hard to come by, the earliest construction of such a system is due to A. A. Markov [Oxt52, p. 134], which was unpublished but included in [NS60, §VI.9; p. 511]. In the symbolic setting, an example was also provided by Oxtoby [Oxt52] in 1952, and the very famous examples of Furstenburg on higer dimensional tori which were given in 1961 [Fur61]. There are also a number of examples of minimal non-uniquely ergodic transformations in the literature of interval exchange maps, which we discuss further soon, and which can be seen in more detail in [FM10, §7.5].

3.4. Entropy

Classifying and differentiating between different mathematical objects is a favorite past-time of mathematicians, and it is natural to ask what invariants can be used to differentiate between topological dynamical systems or systems with an invariant measure.

Invariants of dynamical systems

Cardinality: An obvious question is "what are the possible the cardinalities of a minimal dynamical system"; however, all minimal spaces are either finite (and periodic) or else uncountable. Because any infinite minimal X is compact, complete, and has no isolated points — an application of the Baire Catgeory theorem gives that the space has to be uncountable [LM95]. To see this, note that if (X, T) is not periodic, then every point in X is a limit point, and so $X \setminus \{x_1, \ldots, x_k\}$ is always open (singletons are closed) and dense. Supposing that $\{x_i\}$ is an enumeration of X, we would conclude that $\bigcap_k X \setminus \{x_i\}_{i<k}$ is dense by the Baire category theorem, but obviously this is empty. We are forced to conclude that X has no enumeration, and is therefore uncountable. So cardinality only differentiates the interesting case from the (dull) periodic case.

Spectrum: The spectrum is another invariant of dynamical systems, and the study of the spectral properties of systems has been fruitful. Given a transformation $T : X \circlearrowleft$ and an invariant measure μ, T corresponds to a unitary operator on $L^2(X, \mu)$ by $Uf = f \circ T$, and it is in this sense that we can talk about the spectrum of a system in terms of the spectrum of the operator U. This is an invariant, and for systems with discrete spectrum it is a complete invariant, and all systems with discrete spectrum are conjugate to rotations on an Abelian group [Hal56, §Discrete Spectrum] (for systems with continuous spectrum, this is not true, as was shown by Kolmogorov [Rok60, p. 2]). We are less interested in this invariant here, but it would be negligent not to mention it.

Entropy: Entropy has perhaps been one of the most powerful invariants discovered in ergodic theory, and it created a renaissance in ergodic theory. Borrowing ideas from Shannon's entropy [Sha48], Kolmogorov gave a new invariant for dynamical systems that could differentiate systems with the same spectrum (Kolmogorov's original work on this is in Russian, but Rokhlin [Rok60] summarizes the progress that came immediately after Kolmogorov, *circa 1960*). For subshifts, the entropy is particularly easy to compute, and in fact the first formulation of measure theoretic entropy (sometimes called *metric entropy*, especially in the Russian literature) was for *Bernoulli shifts* (where μ on $\mathcal{A}^{\mathbb{Z}}$ is a product measure of ν on $\mathcal{A}$) by Kolmogorov [BS02].

Metric entropy

As originally defined by Kolmogorov, the entropy of a Bernoulli shift is simply

$$h(\mathcal{A}^{\mathbb{Z}}, \mu) = -\sum_{x \in \mathcal{A}} \mu([x]) \log(\mu([x]))$$

In this restricted class, it turns out that entropy is a complete invariant, as was proven in the celebrated Ornstein isomorphism theorem [Orn70].

The general definition of metric entropy can be seen as a slight generalization of the above. For Bernoulli shifts, the cylinders $\{[x] : x \in \mathcal{A}\}$ form a partition of $\mathcal{A}^{\mathbb{Z}}$ which fully capture the information contained in the measure μ. However, more complicated measures might require partitions of finer granularity, and so instead of summing over $\mathcal{A} \cong L_1(X)$, one takes a limit over finer partitions. In the symbolic case, these partitions can be taken to be the cylinders of length n, $L_n(X)$ (there is no problem of ambiguity between a word and the choice of cylinder because of shift-invariance of the measure). This leads us to the following definition:

Definition 3.4.1: Metric entropy
The *metric entropy of a subshift* (X, S, μ) of $\mathcal{A}^{\mathbb{Z}}$ is given by

$$h_X(\mu) = \lim_n - \sum_{u \in L_n(X)} \mu([u]) \log(\mu[u])$$

This is easily shown to be bound by $\log |\mathcal{A}|$, and the sequence is monotone, so that this limit always exists. [BCD+89]

Maximum entropy

An interesting question that can be asked is how h_X varies over $\mathcal{M}(X, S)$. Using the ergodic decomposition theorem [DGS76, §13] (the fact that every invariant measure μ is equal to an integral against a probability measure P on $\mathcal{E}(X, S)$) , it can be shown that if $\mu = \int_{\nu \in \mathcal{E}(X,S)} \nu \, dP$, then $h_X(\mu) = \int_{\nu \in \mathcal{E}(X,S)} h_X(\nu) \, dP$. This also shows that if there is a unique measure of maximal entropy, then it is ergodic, and more generally the set of measures of maximal entropy form a sub-simplex of $\mathcal{M}(X, S)$ whose extreme points are still ergodic measures (See 3.5 in [EL10]).

But there are more tools available. In 1964 Parry considered an "intrinsic" entropy which did not rely on any fixed measure, but simply captures the growth of the system [Par64]. This intrinsic entropy is *the topological entropy*, which was defined in its proper generality by Adler, Konheim and McAndrew [AKM65] and and later a simpler (and more transparent) definition which applies to compact metric spaces was given by Bowen [Bow71]. But for the symbolic case, which is our main interest, the definition of Parry suffices.

Definition 3.4.2: Topological entropy [Par64]
When (X, S) is a symbolic dynamical system, the *topological entropy of* (X, S) is given by

$$h(X) = \lim_n \frac{\log |L_n(X)|}{n}$$

For symbolic systems with alphabet $\mathcal{A}$, $|L_n| \leq |\mathcal{A}|^n$, so the entropy is bounded above by $\log |\mathcal{A}|$.

Finally, Goodwyn proved that metric entropy is bounded by topological entropy [Goo69], and Dinaburg showed that there is always some measure which attains topological entropy [Din70].

Entropy and minimality

All examples of minimal systems that we have provided so far —including the subshifts of periodic words and the Thue-Morse word— have topological entropy 0. In fact, the vast majority of minimal systems "found in the wild" have zero entropy, and it was conjectured by Parry that when X is a compact metric space and (X, T) is minimal that (X, T) must have topological entropy[14] of zero. This was independently answered by Furstenberg and Hahn & Katznelson ([Fur61], [HK67]) in the negative, the latter showing that there exist minimal (and even uniquely ergodic) systems (X, T) of arbitrarily high entropy. The constructions are very complicated, and so we cannot include them here, but suffice to say that **not all minimal systems have zero entropy**.

[14] The system need not be symbolic here, so one is forced to use Bowen's definition of entropy [Bow71].

4. Low-complexity words

Our main interest here is in a very special class of infinite words, called *low-complexity words*, and the subshifts associated to them.

Definition 4.0.1: Low-complexity word
We define a word to be of *low-complexity* if

$$\limsup_n \frac{|L_n(\omega)|}{n} < \infty$$

Equivalently, a word ω is of low complexity if $p(n) = \frac{|L_n(\omega)|}{n}$ is dominated by some linear function. •

Remark 4.0.2: The Lowest Possible Complexity is Linear
A theorem of Morse and Hedlund [MH38, Thm 7.4] gives that if $p(n) = \frac{|L_n(\omega)|}{n}$ is dominated by a linear function, then either $p(n)$ also has a lower bounding function $0 < n < p(n) < Kn$, or else ω is eventually periodic so that $\lim_n L(n) < \infty$. As a result of this, aperiodic low-complexity words can be thought of as the slowest growing class of words, with approximately linear growth. •

An important result about low-complexity words for us will be the following:

Proposition 4.0.3: $|L_{n+1}| - |L_n|$ is bounded [Cas96, Thm. 1]
The difference $|L_{n+1}(\omega)| - |L_n(\omega)|$ is bounded by a constant if and only if the word ω is of low-complexity. That is

$$\exists k. \forall n.\ |L_{n+1}(\omega)| - |L_n(\omega)| \leq k \qquad \Leftrightarrow \qquad \exists K.\ \limsup_n \frac{|L_n(\omega)|}{n} < K$$

•

Substitutions

As mentioned in Example 2.4.4, (Lebesgue) generic words would have that $L_n = \mathcal{A}^n$ — that is, exponential rather than linear growth. Nonetheless, there are many examples of low-complexity words, one class being words arising as fixed points of *substitutions* (= D0L words).

Definition 4.0.4: Substitutions and D0L words
A *substitution* is a monoid morphism $\zeta : \mathcal{A}^* \circlearrowleft$ which does not delete any word. That is, for no $a \in \mathcal{A}$ does $\zeta(a) = \epsilon$ (ϵ being the empty word, the identity of $\mathcal{A}^*$). Being a morphism on the free monoid $\mathcal{A}^*$, ζ is fully determined by its values on $\mathcal{A}$, and so when defining a substitution ζ one usually only specifies its values on the alphabet.[15]

[15] Sometimes one also considers a variant of the standard substitution, which applies to two-sided words $u.v$ or $u \times v \in \mathcal{A}^* \times \mathcal{A}^*$. The application is the same, but one has to preserve the mark "." or $\times$. So $\zeta(u \times v) = \zeta(u) \times \zeta(v)$.

Low-complexity words

The term "D0L" stands for Deterministic Lindenmayer system with 0 symbols of context [AS03]. A D0L word ω is an infinite word corresponding to a limit $\omega = \lim_n \zeta^n(a)$. Alternatively, one can extend the definition of a substitution to the boundary[16] of $\mathcal{A}^*$ (where words are close if they agree on long initial subwords), $\mathcal{A}^{\mathbb{N}}$, and then D0L words are fixed points of a substitution.

Remark 4.0.5: Primitive substitutions

Among substitutions there is a special class that is guarenteed to generate D0L words. If a substitution satisfies the following

1. For some $a \in \mathcal{A}$, $\zeta(a)$ starts with a.
2. For all $b \in \mathcal{A}$, $|\zeta^n(b)| \to \infty$.
3. For every $a \in \mathcal{A}$, and for every $b \in \mathcal{A}$, b appears in some $\zeta^k(a)$.

Then ζ generates a D0L word $\omega = \lim_n \zeta^n(a)$, and the associated symbolic system $\left(\overline{\mathcal{O}(\omega)}, S\right)$ is minimal and uniquely ergodic. [Fog02, Thm 1.2.7]

The Thue-Morse word Example 2.4.8 is an example of such a word. Another is the *Fibonacci word* [Fog02, §1.1.4]

Example 4.0.6: The Fibonacci word

The substitution $\zeta : \mathcal{A}^* \circlearrowleft$ determined by $\zeta(a) = ab,\ \zeta(b) = a$, is known as the *Fibonacci substitution.* The corresponding D0L word is

$$\mathfrak{f} = \lim_n \zeta^n(a) = abaababaabaab\ldots$$

This word is also an instance of a Sturmian word, which is an interesting class for other reasons. [Fog02]

Standard references on substitutions are [Fog02] and [Que10].

Tougher examples: adic words

D0L words are easy to define and work with, but it is worthwhile to know about words for which (X_ω, S) is **minimal but non-uniquely ergodic**. These examples are famously hard to come up with.

The examples of low-complexity words for which (X_ω, S) is minimal but not uniquely ergodic are much more recent, but the history is described in [FM10] — the earliest examples are due to Veech, but a nice class of words which can be engineered to yield minimal non-uniquely ergodic systems are a special class of *adic words.* These examples were originally constructed by Keane (see [FM10] and references therein), and adic systems are well enough studied that their invariant measures can be computed, making them an interesting, concrete, class of words for which the results in this thesis can be applied. While these examples are challenging, they provide useful (counter)examples in the study of interval-exchange maps [FM10].

[16] One places a metric on $\mathcal{A}^*$ which makes words close if they agree on a long prefix, and then it can be shown that the boundary of $\mathcal{A}^*$ with this topology can be identified with $\mathcal{A}^{\mathbb{N}}$.

Entropy of low-complexity systems

It is interesting, also, to note that these low-complexity words have entropy zero. For a symbolic system the topological entropy is computed by

$$h(\omega) = \lim_{n} \frac{\log |L_n(\omega)|}{n}$$

However, since low-complexity words have roughly linear growth in $|L_n|$, the topological entropy of the system is always zero. A consequence of this is that for low-complexity words with more than one S-invariant measure, the measures cannot be distinguished from one another by their entropy.

5. Graphs

Our ultimate goal in this text is to demonstrate the convergence properties of a sequence of graphs associated to low-complexity words. In order to do this, we first have to specify what we mean by a graph — and in fact, we will have to juggle a few different definitions.

Definition 5.0.1: Graph
We will define a *graph* to be a vertex set paired with a set of oriented edges. That is, $G = (V, E)$, $E \subseteq V \times V$. We will say that there is an edge from u to v, or write $u \to v$, if $(u, v) \in E$.

We will say that two graphs G and H are isomorphic if there exists some invertible map $f : V_G \to V_H$, such that $(u, v) \in E_G \Leftrightarrow (f \times f)(u, v) = (f(u), f(v)) \in E_H$. (That is, they are equivalent up to relabeling of the vertices).

Note that this makes our definition of a graph correspond to what is normally called an *oriented* or *directed* graph with no multiple edges. •

Definition 5.0.2: Labelled graph
A *labelled graph* $\overrightarrow{G}$ is simply a graph G together with a labeling function ℓ. We can consider both *vertex labelings* as well as *edge labelings*, in which case ℓ either has the type $\ell : V \to \mathcal{A}$ or $\ell : E \to \mathcal{A}$. So we define a labelled graph to be a pair $\overrightarrow{G} = (G, \ell)$, and we can recover an unlabelled graph by "forgetting" ℓ.

We will say that two labelled graphs $\overrightarrow{G}$ and $\overrightarrow{H}$ are isomorphic if G and H are isomorphic as graphs under $f : V_G \to V_H$, and if the labels stay the same after changing the vertex set. That is, if $\overrightarrow{G}$ and $\overrightarrow{H}$ are vertex labelled, then $\ell_H \circ f = \ell_G$, or if $\overrightarrow{G}$ and $\overrightarrow{H}$ are edge labelled, then $\ell_H \circ (f \times f) = ell_G$.

•

Topology of graphs

We will not directly use unoriented graphs themselves, but they are used in the definition of *weak connectedness* [GLN16].

Definition 5.0.3: Unoriented graph
A unoriented graph $\overline{G}$ is a just a graph where every edge has an edge with the opposite orientation. That is, define the symmetrization of E, $\overline{E} = \{(v, u) : (u, v) \in E\}$. Then $\overline{G} = (V, E \cup \overline{E})$. •

Definition 5.0.4: Paths in a graph
We say that there is a *path* between u and v, if there exists a finite sequence of edges $(e_1, e_2, \ldots, e_k) \in E$ beginning at u and ending at v, that all connect.

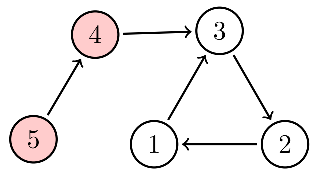
$$\pi_1(e_1) = u,\ \pi_2(e_k) = v,\ \forall_{1 \le i < k}.\, \pi_2(e_i) = \pi_1(e_{i+1})$$

That is, all edges connect properly, $(a, b)(b, d)(d, c) \ldots (y, z)(z, \cdot)$.

Note that paths are oriented, so there may be a path from v_1 to v_2, but not *vice versa*. •

Definition 5.0.5: (Weak) Connectedness
We say that a graph G is (*weakly*) *connected* if there is a path between every pair of vertices (in both directions) in the associated unoriented graph $\overline{G}$. [GLN16]

Note that in this sense, the graph below is connected, despite there being no path from 1 to 5.



Definition 5.0.6: The graph metric for connected graphs
When a graph of any of the above types is (weakly) connected, we can define the *graph metric* $d : V \times V \to [0, \infty)$ to be

$$d(u, v) = \min\{\ k\ :\ (e_1, e_2, \ldots, e_k) \text{ is a path from } u \text{ to } v \text{ in } \overline{G}\ \}.$$

Each connected graph is itself a metric space, and we will soon define a metric space of graphs, and so to distinguish between *neighbourhoods within graphs* (which we can view as subgraphs) versus *neighbourhoods in the space of graphs* we will reserve two different notations for balls in the metric space. We will define for $o \in V$,

$$D_r(o) = (V', E'),\ \text{where } V' = \{v\ :\ d(o, v) \le r\},\ E' = \{(u, v) \in E\ :\ u, v \in V'\}.$$

Note that we are taking closed neighbourhoods.

Definition 5.0.7: Locally finite graphs
A graph is *locally finite* if every vertex has a finite number of incoming and outgoing edges

$$\forall_{v \in V}.\ |\{(e_1, e_2)\ :\ e_1 = v \vee e_2 = v\}| < \infty$$

A graph is said to have *bounded vertex degree* if, stronger than being locally finite, there is an upper bound on the number of incoming or outgoing edges.

$$\exists M. \forall v \in V.\ |\{(e_1, e_2)\ :\ e_1 = v \vee e_2 = v\}| < M$$

The motivation for this is that connected locally finite graphs are locally compact with respect to the graph metric.[17]

[17] Any radius-r neighbourhood of a vertex locally finite graph has finitely many vertices, and therefore any open cover for that neighbourhood can be restricted to a finite subcover (pick one open set per vertex). This fails for non-locally finite graphs.

5.1. Rauzy graphs

Definition 5.1.1: Rauzy graphs

The nth Rauzy graph of a word ω, $\mathcal{R}^n_{\mathcal{A}}(\omega)$, has $L_n(\omega)$ as the vertex set, and there is an edge between two length n factors if the factors appear together in a length $n+1$ factor of ω. That is

$$u \to v \ \text{ in } \mathcal{R}^n_{\mathcal{A}}(\omega) \quad \Leftrightarrow \quad \overbrace{w_1\, \underbrace{w_2 \ldots w_n}}^{u}\underbrace{ w_{n+1}}_{v},\ w \in L_{n+1}(\omega)$$

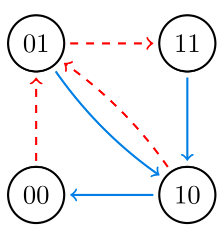
It can be seen by the above that the edge set of $\mathcal{R}^n(\omega)$ can be identified with $L_{n+1}(\omega)$.

One can also think of this in terms of *windows* [Kai18]: if u is the contents of the window of length n, and v is the content of the window after sliding the word (= applying the shift), then $u \to v$ in $\vec{\mathcal{R}}^n$

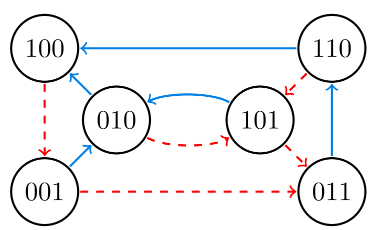
$$\overset{\mathfrak{t}}{\boxed{011}0100110\ldots} \to \overset{S(\mathfrak{t})}{\boxed{110}1001100\ldots} \Rightarrow 011 \to 110 \text{ in } \mathcal{R}^3(\mathfrak{t})$$

An edge in the third Rauzy graoh of the Thue-Morse word, $\mathfrak{t}$.

Then the second and third Rauzy graph of $\mathfrak{t}$ turn out to be[18]



$\vec{\mathcal{R}}^2(\mathfrak{t})$



$\vec{\mathcal{R}}^3(\mathfrak{t})$

Instead of simply considering edges, we can consider *labelled* edges, with labels coming from $\mathcal{A}$. In the graphs above, this is represented by the color (blue-thick for 0, red-dashed for 1).

Definition 5.1.2: Labelled Rauzy graphs

A slight modification of the earlier definition of Rauzy graphs, labelled Rauzy graphs also color the edges with the letter from $\mathcal{A}$ last seen in the window. For example, the label of the edge $\boxed{011}01\ldots \to \boxed{110}10\ldots$ would be labelled with a 0, because of the rightmost 0 which just entered the window. The labelled graph will be denoted by $\vec{\mathcal{R}}^n_{\mathcal{A}}(\omega)$.

[18] How one determines that the list of permissible transitions is comprehensive is not obvious, but for D0L words Definition 4.0.4 there are algorithms; see [BR17] or [Fri01].

Properties of rauzy graphs

We are interested in Rauzy graphs of *low-complexity words* (Definition 4.0.1). Since these words have few factors, the corresponding Rauzy graphs have very few vertices, and since $E \cong L_{n+1}(\omega)$, not many more edges. This will play a very important role when we consider what large Rauzy graphs "look like" on the large scale. We also have two more properties of Rauzy graphs of words:

1. The Rauzy graph of a word is always weakly connected (i.e. connected in the undirected sense).
2. If ω is recurrent, then the graph is *strongly connected* or *irreducible* in the sense often applied to Markov chains — meaning that for every pair of vertices u, v there is an oriented path from u to v.

Of course, these graphs also have bounded vertex degree ($\leq |\mathcal{A}|$); also, each edge in $\vec{\mathcal{R}}^n(\omega)$ corresponds to an element in $L_{n+1}(\omega)$, and by Proposition 4.0.3, this means that a bounded number of vertices in $\mathcal{R}^n(\omega)$ have more than one outgoing edge. We will call these *special vertices*.

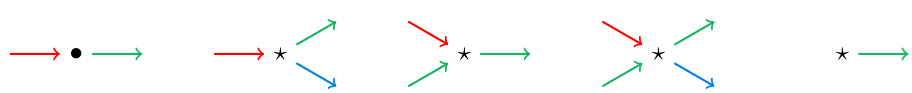
Special vertices

From the definition of Rauzy graphs it is easily seen that the vertices in L_n are connected by edges which are in correspondence with words in L_{n+1} (for example, $01 \to 11$ in $\mathcal{R}^2(\omega)$ iff $011 \in \mathcal{R}^3(\omega)$). Moreover, every vertex has at least one outgoing edge ($|L_{n+1}| \geq |L_n|$), and for every *additional* outgoing edge the size of L_{n+1} increases. Denoting by $\text{out}(v)$ the set of outward edges, $\{e \in E : e = (v, u)\}$, we have that

$$|L_{n+1}(\omega)| = |L_n(\omega)| + \sum_{v \in L_n(\omega)} (|\text{out}(v)| - 1)$$

The same holds when summing over the in-degree $\text{in}(v) = \{e \in E : e = (u, v)\}$, and this motivates us to name these vertices which have $\text{in}(v)$ or $\text{out}(v)$ with 2 or more elements.

Definition 5.1.3: Special Vertices
We call a vertex v of a graph *right-special* if $|\text{out}(v)| = |\{e \in E : e = (v, u)\}| \geq 2$, and we will call the vertex *left-special* if $|\text{in}(v)| = |\{e \in E : e = (u, v)\}| \geq 2$ **or** $|\text{in}(v)| = 0$ — that is, we choose to call an vertex with no predecessors left-special. A vertex will be called *special* if it is left-special or right-special (and it is usually called *bi-special* if it is left and right special). A vertex will be called *regular* if it is not special.



The leftmost vertex (•) is regular; the four on the right are special (⋆).

Now, we have a very useful result for Rauzy graphs of low-complexity words.

Proposition 5.1.4: Bounded number of special vertices [Fri01]
Let ω be a low-complexity word, with k such that $|L_{n+1}(\omega)| - |L_n(\omega)| < k$ (Proposition 4.0.3). Then Rauzy graphs of ω, $\vec{\mathcal{R}}^n(\omega)$, have at most $2k+1$ special vertices (this is independent of n). •

Proof: In Proposition 4.0.3 we saw that the difference $|L_{n+1}(\omega)| - |L_n(\omega)|$ is bounded by a constant if and only if the word ω is of low-complexity. In terms of Rauzy graphs, this gives that there are a bounded (by k) number of words in L_n which extend (on the right or the left) to two or more different words in L_{n+1}, and since the out-degree of every vertex in a Rauzy is at least one, there are at most k right-special vertices. However, the initial length n word $\omega_0 \dots \omega_{n-1}$ in $\vec{\mathcal{R}}^n(\omega)$ *may* not have any incoming edge, and so even if $|L_{n+1}(\omega)| - |L_n(\omega)| = 0$, there may be one vertex v with $\text{in}(v) = \emptyset$, and another with $|\text{in}(u)| \geq 2$. But there can be at most one vertex with in-degree of zero, and so the number of left-special vertices is bound by $k+1$. □

$$
\begin{tikzcd}
 & \bullet \arrow[rr] & & \star \arrow[dr] & \\
\bullet \arrow[ur] & & \bullet \arrow[ur] & & \bullet \arrow[ll]
\end{tikzcd}
$$

Example where $|L_{n+1}| = |L_n|$, but there is one left-special vertex.

This gives us that for large Rauzy graphs, we can remove at most $2k+1$ special vertices and be left with a finite ($\leq |\mathcal{A}|(2k+1)$) number of lines. As a result of this, if the number of vertices goes to infinity, then necessarily the Rauzy graphs contain arbitrarily long *line-like stretches*. This will be very important soon, when we consider random neighbourhoods of large Rauzy graphs.

5.2. Cayley graphs

We want to motivate the dynamics on graphs by describing symbolic dynamics as a special case of dynamics on Cayley graphs. As we mentioned earlier, in the classic setup of symbolic dynamics one often considers bi-infinite sequences of symbols over some (finite) alphabet $\mathcal{A}$, and this space $\mathcal{A}^{\mathbb{Z}}$ is equipped with the projective topology, with a compatible metric

$$a, b \in \mathcal{A}^{\mathbb{Z}}.\ d(a,b) = 2^{-i}, \text{ where } i = \inf\{\ r \mid a_r \neq b_r \text{ or } a_{-r} \neq b_{-r}\ \}$$

It is also equipped with a (continuous) group action $\mathbb{Z} \curvearrowright \mathcal{A}^{\mathbb{Z}}$, the shift

$$S^n(\dots, \omega_{-2}, \omega_{-1}; \omega_0, \omega_1, \omega_2, \dots) = (\dots, \omega_{n-2}, \omega_{n-1}; \omega_n, \omega_{n+1}, \omega_{n+2}, \dots)$$

We can also describe this in more geometrical terms; by thinking about $\mathbb{Z}$ in terms of its *Cayley graph.*

Definition 5.2.1: Cayley graph (of a group)
When Γ is a finitely generated (not necessarily Abelian) group with a finite generating set S (in that $S \cup S^{-1}$ generate the group), we can define the *Cayley graph of* Γ *with respect to* S to be the graph

$$\mathfrak{C}ay\,(\Gamma, S) = (\Gamma, \{(x, x\gamma)\ :\ x \in \Gamma, \gamma \in S\})$$

For example, the Cayley graph of $\mathbb{Z}$ with the generating set $\{1\}$ gives the graph $\mathfrak{C}ay\,(\mathbb{Z}, \{1\})$

$$\cdots \longrightarrow -2 \longrightarrow -1 \longrightarrow 0 \longrightarrow 1 \longrightarrow 2 \longrightarrow \cdots$$

(Note the orientation.) This graph will be called *the line graph*, or *the oriented line graph.* •

Symbolic dynamics on Cayley graphs

Using the Cayley graph, instead of thinking of $\mathcal{A}^{\mathbb{Z}}$ as a space of infinite words, we can think of it as a space of colorings (= vertex-labelings) of the integer line. That is, we can identify a word $\omega \in \mathcal{A}^{\mathbb{Z}} \cong \{\ell : \mathbb{Z} \to \mathcal{A}\}$, with $f = (\mathfrak{C}ay\,(\mathbb{Z}, \{1\}), \ell)$ as a vertex-labelled graph:

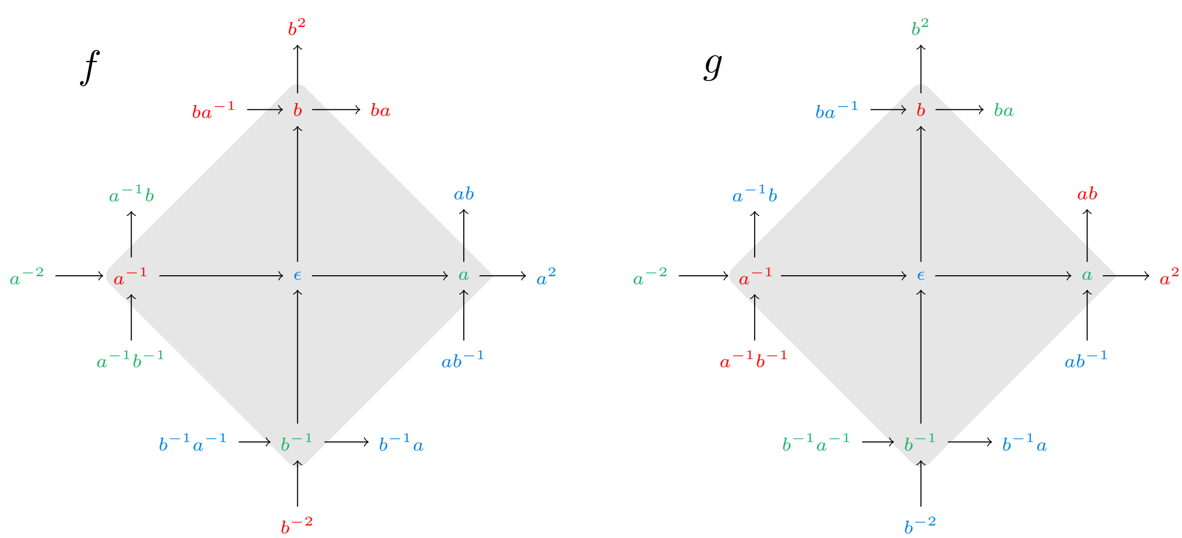
$$f_{-3} \longrightarrow f_{-2} \longrightarrow \overbrace{f_{-1} \longrightarrow f_0 \longrightarrow f_1}^{f\big|_{B_1(0)}} \longrightarrow f_2 \longrightarrow f_3$$

$$g_{-3} \longrightarrow g_{-2} \longrightarrow \underbrace{g_{-1} \longrightarrow g_0 \longrightarrow g_1}_{g\big|_{B_1^n(0)}} \longrightarrow g_2 \longrightarrow g_3$$

$$d(f,g) = 2^{-2}, \text{ since } f\big|_{B_2(0)} \neq g\big|_{B_2(0)}$$

Here we are comparing colored vertices.

For finitely generated infinite groups, the Cayley graph, $\mathfrak{C}ay\,(\Gamma, \{\gamma_i\})$, is connected and has bounded vertex degree, so that as a metric space[19] it is locally compact, and Γ acts on its Cayley graph by isometries [Gro87, §8.4] [Mei08, Proof of Thm 9.2]. In the same way as with $\mathcal{A}^{\mathbb{Z}}$, we can equip $\mathcal{A}^{\Gamma}$ with the projective topology. For example, denoting by F_2 the *free group on two generators*, we could imagine a metric on $\mathcal{A}^{F_2} \cong \{(\mathfrak{C}ay\,(F_2, \{a,b\}), \ell) \mid \ell : F_2 \to \mathcal{A}\}$ like so:



$$d(f,g) = 2^{-2}, \text{ since } f\big|_{B_2(\epsilon)} \neq g\big|_{B_2(\epsilon)} \text{ in } \mathfrak{C}ay\,(F_2, \{a,b\})$$

That is, instead of $0 \in \mathbb{Z}$, we consider the identity of the group $\epsilon \in \Gamma$ as the root, and we measure the distance between colorings of the group (or *configurations*) [Lov12] based on the radius from the root (ϵ)

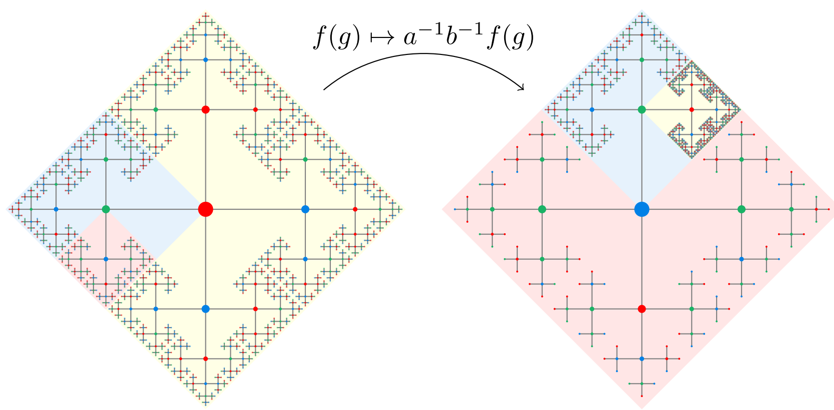
[19] equipped with the graph metric or equivalently with the word metric $d(x,y) = \min\{|u| : xu = y\}$, where u is a word over $\{\gamma_i\} \cup \{\gamma_i^{-1}\}$.

for which the colorings agree. The Cayley graph is locally compact in the word/graph metric when there are finitely many generators, which makes $\mathcal{A}^\Gamma$ compact[20] in the projective topology, and it has a familiar compatible metric

$$d(f,g) = 2^{-i}, \text{ where } i = \inf\{r \,:\, f\big|_{D_r(\epsilon)} \neq g\big|_{D_r(\epsilon)}\}$$

Also, $\Gamma \curvearrowright \mathcal{A}^\Gamma : (\gamma, f(x)) \mapsto f(\gamma x)$ is continuous[21], and for any of its generators γ_i it satisfies

$$\frac{1}{2} d(f,g) \leq d(\gamma_i f, \gamma_i g) \leq 2d(f,g)$$



Re-rooting of $\mathfrak{C}ay\,(F_2, \{a,b\})$ to $a^{-1}b^{-1}$

A last observation: with the shift, one gets an orbit equivalence relation, $f \sim g \Leftrightarrow \exists k. f = S^k g$. That is, f is just a translate of g, like

$$f = (\ldots 1, 1, 1, 0, \overset{\downarrow}{0}, 0, 1, 1, 1 \ldots)$$

$$g = (\ldots 0, 0, 0, 1, \overset{\downarrow}{1}, 1, 0, 0, 0 \ldots)$$

and similarly, the group action induces an equivalence relation [FM77] of translations of colorings of the group, which amounts to just *re-rooting* the Cayley graph at γ [Mei08]. $f(x) \sim g(x) \Leftrightarrow \exists \gamma \in \Gamma. f(x) = g(\gamma x)$.

From this, we can try to imagine what needs to be preserved in a generalization of this setup:

Local finiteness: Local finiteness, and especially bounded degree, are important for the topology of the underlying metric space — in our case $\mathfrak{C}ay\,(\Gamma, \{\gamma_i\})$, as it gives us compactness.

Translation invariance: We have an action on this space, and invariance of a measure under $\Gamma \curvearrowright \mathcal{A}^\Gamma$ is equivalent to invariance with respect to the *orbit-equivalence relation*, which itself is equivalent (in Cayley graphs) to invariance under *re-rooting* of the graphs.

[20] Of course, $\mathcal{A}^\Gamma$ is still compact when given the product-of-discrete topology, but when Γ is countable it is metrizable, and when Γ is finitely generated we can construct the metric using the word metric.

[21] The Cayley graph transitions multiply on the right, so for a continuous action we need to multiply on the left.

Remark 5.2.2: $\mathcal{A}^{\Gamma}$ is usually hard to work with
Of course, the similarities between $\mathcal{A}^{\mathbb{Z}}$ and $\mathcal{A}^{\Gamma}$ have to end at some point. One dissimilarity is that in general one cannot even compute the word metric (= graph metric) on $\mathfrak{C}ay\,(\Gamma, \{\gamma_i\})$ even if Γ is finitely generated (even if it is finitely presented). This is known as *Dehn's word problem* [Deh11], which is known not to be algorithmically solvable in general — the negative solution to the problem is known as the Novikov-Boone theorem [BBN59] [Mei08]. Of course, the word problem is solvable for some restricted class of groups.

5.3. Graphical dynamics as an extension of symbolic dynamics

The example of dynamics on Cayley graphs of finitely generated groups motivates the definition of the following space of graphs, which is meant to preserve the important properties that we needed in the symbolic dynamics setup, while removing the requirement of algebraic structure. We define a *rooted graph* to be a pair (G, o), where o is a vertex in G.

Definition 5.3.1: The space of graphs
We define $\mathcal{G}_\bullet$ to be the space of (isomorphism classes of) rooted, connected and locally finite directed graphs, equipped with the projective topology. This space is locally compact and metrizable [BS11, 1.2]. We have to pass between rooted graphs and isomorphism classes of rooted graphs often, so we will reserve the notation (G, o) (or $(\overrightarrow{G}, o)$) for rooted graphs, and we will denote their isomorphism classes by $[(G, o)]$ (cumbersome though this may be). We will denote neighbourhoods in this space by

$$B_r([(G,o)]) = \{[(H,o')] \mid D_r(G,o) \cong D_r(H,o')\ \}.$$

These neighbourhoods form a basis for the projective topology. We analogously, we can define a space of labelled graphs, modifying only the notion of isomorphism. We will denote this space by $\overrightarrow{\mathcal{G}}_\bullet$. (Recall that D_r is the closed disk in the graph of radius r)

The subspaces of $\mathcal{G}_\bullet^M \subset \mathcal{G}_\bullet$ (or $\overrightarrow{\mathcal{G}}_\bullet^M \subset \overrightarrow{\mathcal{G}}_\bullet$) of graphs of vertex degree bounded by M form compact subspaces of $\mathcal{G}_\bullet$ (or $\overrightarrow{\mathcal{G}}_\bullet$), bounded degree playing the same role as the finite generating set in Cayley graphs of finitely generated groups.

The root of the graphs plays the role that the identity played when we topologized the space, and connectedness takes the place of vertex-transitivity that was previously available from the group structure. So, this can be seen as a natural generalization of the earlier setup. In addition to this, while there is no longer a group action available, $\mathcal{G}_\bullet$ can be endowed with a *root moving equivalence relation*, R [Kai98, Kai15], which is the natural generalization of the *orbit equivalence relation* that was provided by the group.

Benjamini-Schramm convergence

Within $\mathcal{G}_\bullet$ we can study convergence of rooted graphs, but, if we consider unrooted graphs, we have no reasonable topology available to take limits [Kai15]. So, instead, we can consider a *random rooting* of unrooted graphs, to represent finite unrooted graphs as measures on rooted graphs. That is, we can consider a measure μ on $\overrightarrow{\mathcal{G}}_\bullet$ or $\mathcal{G}_\bullet$ which assigns uniform probability $P(\{(G,o)\}) = |G|^{-1}$ to each rooting of G, and then this induces a measure on $\overrightarrow{\mathcal{G}}_\bullet$ or $\mathcal{G}_\bullet$. Note, however, that the measure on $\mathcal{G}_\bullet$ is not necessarily uniform, because some distinct rootings may yield isomorphic graphs in the presence of symmetries (if $\mathrm{Aut}\,(G) \neq \{id\}$).

This leads us to the definition of *Benjamini-Schramm convergence*.

Definition 5.3.2: Benjamini-Schramm convergence [BS11, §1.2]
We say that a sequence of finite graphs (G_i) has a *Benjamini-Schramm limit* if the measures μ_i on $\mathcal{G}_\bullet$ induced by uniform rootings of G_i converge *-weakly to a measure μ on $\mathcal{G}_\bullet$. We call this measure μ a *random graph*.
Elaborating: For the sequence (G_i) of finite graphs, define the measures $\mu_i := \frac{1}{|G_i|}\sum_{v\in G_i}\delta_{[(G_i,v)]}$. We say that the graphs (G_i) converge in the Benjamini-Schramm sense when the μ_i (*-weakly) converge in $\mathcal{M}(\mathcal{G}_\bullet)$ to a measure μ, and this defines a *random graph* $(G,o)\sim\mu$. •

Proposition 5.3.3: Equivalent conditions for convergence
The following are equivalent formulations of Benjamini-Schramm convergence of (G_i), with (μ_i) as defined in Definition 5.3.2:

1. Definition of weak-* limit:

$$\mu_i \to \mu \quad \Leftrightarrow \quad \forall_{f\in C(\mathcal{G}_\bullet,\mathbb{R})}.\ \int f\, d\mu_i \to \int f\, d\mu$$

2. Probabilistic formulation:

For every finite rooted graph (H,o_h), and radius r

$$P(D_r(G_i,o_i)\cong(H,o_h)) \xrightarrow{i} P(D_r(G,o)\cong(H,o_h))$$

3. Convergence of averages:

For every neighbourhood in $\mathcal{G}_\bullet$, $U=B_r(H,o_h)$, the following exists

$$\lim_i \frac{1}{|G_i|}\sum_{v\in G_i} I_U\left([(G_i,v)]\right)$$

The second and third condition are easily seen to be equivalent once one unpacks what $P(D_r(G_i,o_i)\cong(H,o_h))$ means. Both are equivalent to the first after one notices that the indicator functions of neighbourhoods of finite graphs, I_{U_n}, generate a dense subalgebra of $C(X,\mathbb{R})$, and that summing over all rootings is just integration of I_{U_n} with respect to μ_i.

Convergence of labelled graphs is defined in the same way for $\overrightarrow{\mathcal{G}}_\bullet$, after modifying the interpretation of "$\cong$" appropriately. •

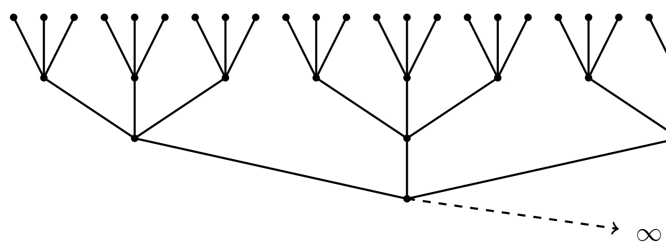
It is a happy fact that if the graphs G_i have bounded degree, then by compactness of $\mathcal{G}_\bullet^M$ one gets that there is always a subsequential limit of the μ_i.

An example: the Bethe lattice and the canopy graph

The Benjamini-Schramm limit captures what the a sequence of graphs tends to look like locally. Sometimes this is more surprising that it sounds, as the next example illustrates.

Example 5.3.4: The Bethe lattice and the canopy graph
The neighbourhoods of ϵ in $\mathfrak{C}ay\left(F_2, \{a, b, a^{-1}, b^{-1}\}\right)$, $D_r(\epsilon)$, converge to the *Canopy graph*, **not** $\mathfrak{C}ay\left(F_2, \{a, b, a^{-1}, b^{-1}\}\right)$ [Bow15][BS11, Remark after Thm 1.1]. Note that we are taking the unoriented Cayley graph in this case, with all generators and inverses. This simplifies some symmetry problems that otherwise arise.

The first intuition is that the neighbourhoods "grow into" $\mathfrak{C}ay\left(F_2, \{a, b\}\right)$, because this is what happens when you take the limit of neighbourhoods rooted at ϵ. However, when the root is randomly sampled from the neighbourhoods the majority of roots ($\approx \frac{2}{3}$) are on the boundary (they have degree 1), whereas all points of $\mathfrak{C}ay\left(F_2, \{a, b\}\right)$ have degree 4. The proportion in $D_n(\epsilon)$ of vertices n-steps from the boundary of the ball roughly $\frac{2}{3} \cdot 3^{-n}$, and passing to the limit this becomes exact. The graph which captures this distribution is, as mentioned, the Canopy graph:

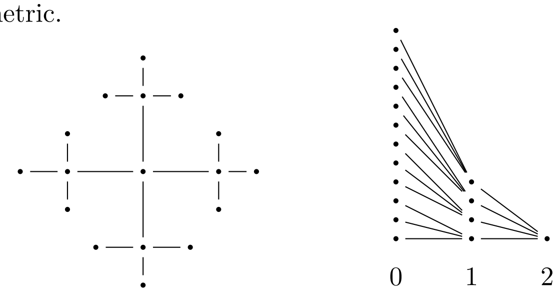


The 3-canopy graph

Elaborating a little, in $D_r(\epsilon)$, the proportion of vertices within k steps of the boundary is

$$P_r(\text{ distance from } o \text{ to boundary } = k) = \begin{cases} \frac{1}{1+2(3^r-1)} & k = r \\ \frac{4 \cdot 3^{n-k-1}}{1+2(3^r-1)} & \text{otherwise} \end{cases}$$

After assigning a uniform random rooting to $D_r(\epsilon)$, quotienting by the rotational symmetry gives that the resulting isomorphism classes are determined entirely by the distance of the root from ϵ. So our proportion function P_r actually turns into a probability distribution on the isomorphism classes indexed by $\overline{\mathbb{N}} = \mathbb{N} \cup \{\infty\}$, and it is approximately geometric.



A randomly selected root is typically near the boundary, so when we pass to the limit it is the center of the neighbourhoods, ϵ, that disappears into the distance. And this can be seen by taking

$$\lim P_r(\{k\}) = \frac{2}{3} \cdot 3^{-k}$$

What gives the Bethe lattice?: To construct the Bethe lattice, $\mathfrak{C}ay\left(F_2, \{a, b\}\right)$, as a Benjamini-Schramm limit, one needs to take a sequence of 4-regular graphs with *girth (= length of the shortest cycle)* going to infinity [Lov12, Example 19.6] [Bow15]. (These graphs are hard to construct, but they exist [HK83])

6. Limits of the Rauzy graphs of low-complexity words

Limits of high-complexity systems

As mentioned, our interest is in studying the Benjamini-Schramm convergence of Rauzy graphs of low-complexity words. The case for Rauzy graphs of *subshifts of finite type*[22] has already been studied by Leemann [Lee16], and his work motivated this project. The tools used by Leemann were much more complex, involving horospheric products of trees [Woe14] [BNW08], whereas the results here use fairly elementary tools from ergodic theory, plus a few results about low-complexity words.

The low-complexity case

The question is, when we consider the sequence of Rauzy graphs $\vec{\mathcal{R}}^n(\omega)$ of a low complexity word ω, what is the limiting random graph?

Big Rauzy graphs are line-like

We have to differentiate between the labelled and unlabelled case, and we handle the eventually-periodic ω case separately. In the infinite case it turns out the low-complexity Rauzy graphs extend into a *bounded* number of very long loops, and locally the graphs look like a line nearly everywhere. The consequence of this is that when passing to the Benjamini-Schramm limit one actually arrives at a random graph that looks exactly like the line graph.

After adding labels, the situation becomes more complicated. In the case where the low-complexity word has uniform frequencies, the frequency of the subwords manifests itself in the Rauzy graph, as stretches of edges in the Rauzy graphs re-encode subwords. However, if the word does not have uniform frequencies then one cannot always recover the all of the measures from the symbolic system. We illustrate this with a few examples.

[22] Shifts consisting of all infinite words that do not contain any forbidden subwords from a finite list F. For instance, the golden mean shift featured earlier is the shift of all words on $\{0, 1\}$ not containing 11 as a subword.

6.1. A lemma on the shape of Rauzy graphs

We claim that the Rauzy graphs of low-complexity words resemble lines, but first we have to define what we mean by a line.

Definition 6.1.1: Lines in a graph

We want to define a class of graphs that resemble $\mathfrak{C}ay\,(\Gamma, \{1\})$ where Γ is a finite cyclic group $(\mathbb{Z}/n\mathbb{Z})$ or the infinite cyclic group $\mathbb{Z}$, because locally these graphs are always lines (even if they have a closed cycle globally). We will define a *line in a graph* to be a finite sequence of joined edges between regular vertices containing no loops except possibly between the first and last vertex. So $(e_i)_{0\leq i\leq p}$ is a line in the graph if

$$(e_i)_{0\leq i\leq p} = ((v_i^1, v_i^2))_{0\leq i\leq p}, \quad i < p \Rightarrow v_i^2 = v_{i+1}^1, \quad v_p^2 = v_i^1 \Rightarrow i = 0, \quad v_i^1, v_i^2 \text{ are all regular}$$

The *length* of such a path will be the number of edges, $p+1$.

For technical reasons, we will also make use of r-*trimmed* lines, which remove the first and last r edges. We will denote this by T_r, and

$$T_r((e_i)_{0\leq i\leq p}) = (e_i)_{r\leq i\leq p-r}$$

The following lemma shows that large Rauzy graphs locally resemble lines — it is basically just Proposition 5.1.4 with inequalities added

Lemma 6.1.2: The Rauzy graphs contain arbitrarily long lines

Let ω be a low-complexity word with $|L_n(\omega)| \to \infty$. Then for n sufficiently large, and arbitrarily large proportion of the vertices are on a finite number of long subgraphs which are lines. That is, for any $M > 0$ and $\epsilon > 0$, there is an N such that for all $n > N$ every Rauzy graph $\overrightarrow{\mathcal{R}}^n(\omega)$ has the following:

1. After removing all special vertices and cutting $\overrightarrow{\mathcal{R}}^n(\omega)$ into lines, at least one line of length greater than or equal to M.
2. More than that, for sufficiently large N, a $1-\epsilon$ proportion of the vertices are on the lines of length exceeding M.
3. This is still true if we trim the edges of the lines by some fixed amount r.

Proof: By Proposition 5.1.4 we know that there are a bounded number of special vertices — we will denote this bound by K. If a Rauzy graph $\overrightarrow{\mathcal{R}}^n(\omega)$ has no special vertices, then it is a finite cycle (this is the only way for every vertex to have in-degree and out-degree constantly 1 and for the graph to be finite), and so just joining every compatible edge in a sequence, starting from an arbitrary vertex, we get a line (by our definition, Definition 6.1.1).

In the general case where there is at least one special vertex, take $\overrightarrow{\mathcal{R}}^n(\omega) = (V, E)$, and take the regular vertices. $V_r = V \setminus \{v \in V \,:\, v \text{ is special }\}$, and then take $E_r \subset E$ to be every edge with vertices in V_r (all edges between regular vertices). Define $G_r = (V_r, E_r)$.

The edges in each connected component of G_r can be ordered to form lines with no loops[23], and note that since the initial vertices of λ_j in $\overrightarrow{\mathcal{R}}^n(\omega)$ are adjacent to a special vertex, and since there are at most $K|\mathcal{A}|$ out-going edges from special vertices, there are at most $K|\mathcal{A}|$ components (which are lines) which we will call λ_j.

Proof of `(1)`

Since there are a bounded number of special vertices, a bounded number of λ_j, and since $|L_n(\omega)| \to \infty$, it is clear that for any M, in a large enough $\overrightarrow{\mathcal{R}}^n(\omega)$ the size of some λ_j must exceed M.

Proof of `(2)`

The proportion of vertices taken up by bounded λ_j and special vertices is at most

$$\frac{K + (K|\mathcal{A}| - 1)M}{|L_n(\omega)|}$$

since there are at most $K|\mathcal{A}|$ lines λ_j, and at least one of them is not bounded by M (since $|L_n \to \infty|$). Clearly the above quantity goes to zero since $|L_n(\omega)| \to \infty$. So we have that the unbounded λ_j eventually account for an arbitrary number of the vertices.

Proof of `(3)`

This is not shocking, but if we "trim" the edges of the finite number of unbounded segments by r (from both sides), then we would still have that

$$\frac{K + (K|\mathcal{A}| - 1)M + K|\mathcal{A}|2r}{|L_n(\omega)|} \longrightarrow 0$$

so the trimmed lines just remove a bounded number of vertices, and so the trimmed unbounded lines still account for an arbitrarily large proportion of the vertices of the graph. $\square$

6.2. The unlabelled case

We first consider the case of convergence in the unlabelled sense, where the proof is an easy consequence of Lemma 6.1.2. For technical reasons we have to extend our definition of lines in the graph from sequences of edges to elements of $\mathcal{G}_\bullet$, and so we begin with a definition.

[23] If there was any loop (a line whose first and last edge share a vertex) then this loop would have been a disconnected component from the special vertex in the original Rauzy graph, but all Rauzy graphs of a word are (weakly) connected. Similarly, the connected components cannot have cycles as a proper subgraph, since this would force at least one vertex in the component to be special. So each connected component of G_r is a tree (since it has no cycles) where every vertex has in-degree and out-degree of 1.

Definition 6.2.1: Line segments
Let

$$\Lambda' := (V, E),\ V = \mathbb{Z},\ E = \{(n, n+1) : n \in \mathbb{Z}\}.$$

and let $\Lambda = [\Lambda'] \in \mathcal{G}_\bullet$. Then a *line segment* is a graph of the form $[\Lambda'_r]$, where $\Lambda'_r := D_r(0) \subset \Lambda'$, and Λ is *the line graph.*

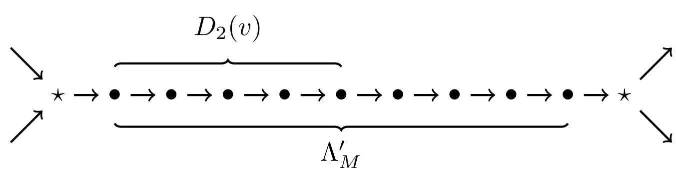
$$\overbrace{-3 \longrightarrow -2 \longrightarrow -1 \longrightarrow 0 \longrightarrow 1 \longrightarrow 2 \longrightarrow 3}^{\Lambda'_3}$$

We analogously define *the oriented line graph* $\vec{\Lambda}$ and *the oriented line segments* $\vec{\Lambda}_n$ to be edge-labelled line graphs in $\vec{\mathcal{G}}_\bullet$.

Now, we prove the infinite unlabelled case.

Theorem 6.2.2: Unlabelled convergence when $|L_n| \to \infty$
Suppose that ω is a low-complexity word and that $|L_n(\omega)| \to \infty$. Then the unlabelled Rauzy graphs $\mathcal{R}^n(\omega)$ converge to the Dirac-measure on Λ in $\mathcal{G}_\bullet$ — the doubly-infinite oriented line graph.

Proof: Fix an $r > 0$ and $M > 2r$. In a line of length M, an r-neighbourhood centered at a vertex at least r away from the edge of the line will lie within the line:



Fixing $\epsilon > 0$, we apply Lemma 6.1.2 part (3) and get that for some N, for all $n > N$, the r-trimmed unbounded lines account for more than a $1-\epsilon$ proportion of the vertices, and so with probability exceeding $1-\epsilon$ a randomly chosen vertex in $\mathcal{R}^n(\omega)$ will have an r-neighbourhood isomorphic to a line. That is,

$$P_n(B_r([(\mathcal{R}^n(\omega), v)]) \cong \Lambda_r) = \mu_n(\{\Lambda_r\}) \to 1$$

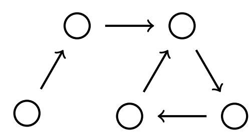
The limit $\mu(U_i)$ is therefore defined for all neighbourhoods of finite graphs, which form a basis for the topology of $\mathcal{G}_\bullet$, and therefore we have a defined limit measure μ. Also, since $B_1(\Lambda) \supset B_2(\Lambda) \supset B_3(\Lambda) \supset$ ..., $\{\Lambda\} = \bigcap_{r>0} B_r(\Lambda)$, we have that

$$\mu\left(\bigcap_{r>0} B_r(\Lambda)\right) = \lim_{r\to\infty} \mu(B_r(\Lambda)) = 1 = \mu(\{\Lambda\})$$

So the Benjamini-Schramm limit is the point mass on the bi-infinite line graph. □

The finite case

For the case where $|L_n(\omega)|$ is bounded, low-complexity words are simply eventually periodic words [CN10, MH38], and so the Rauzy graphs stabilize to the periodic portion (a loop), plus a short path leading into the loop.



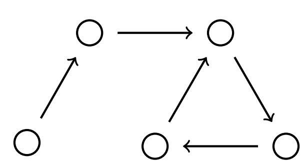
Theorem 6.2.3: Unlabelled convergence when $|L_n| < M$
Suppose that ω is a low-complexity word and that $|L_n(\omega)|$ is bounded. Then ω is either periodic or eventually periodic. The limit graphs are described in two cases.

If ω is periodic with period p (i.e. $p = \min\{\, p' > 0 \mid \omega = S^{p'}(\omega)\}$), then $\vec{\mathcal{R}}^n(\omega)$ are all isomorphic for $n \geq p$, and are all just cycles — which have a single isomorphism class. So in this case, the Benjamini-Schramm limit is a delta function on the oriented cycle graph with p vertices.

If ω is eventually periodic, define $k = \min\{\, k' \mid S^{k'}(\omega)$ is periodic $\}$, and define the period p of S^k as before. For $n \geq \max(k, p)$, the Rauzy graphs form a path of length k leading into a cycle of order p. For all $n \geq \max(p, k)$, the Rauzy graphs have $p + k$ vertices and no symmetries. The point where the path meets the cycle effectively indexes every vertex by its (oriented) distance from the point where the path enters. As a result, the Rauzy graphs stabilize to a rigid oriented graph with $p+k$ vertices and the structure previously described, and the Benjamini-Schramm limit is a uniform distribution on the $p + k$ distinct rootings of (isomorphism classes of) the graph.

Comment: If ω is a two-sided infinite word, it is impossible to be eventually periodic.

Example 6.2.4: An eventually periodic word
For example, if $\omega = cccabcabcabc\ldots$ the graphs stabilize to



Remark 6.2.5: If ω is uniformly recurrent
In either case, whether $|L_n(\omega)| \to \infty$ or $|L_n(\omega)| < k$, if ω is uniformly recurrent, the Benjamini-Schramm limit is a delta function on a cyclic graph — either the infinite-cyclic line graph, or a finite cyclic loop.

Remark 6.2.6: Rauzy graphs of doubly infinite words
Up to this point we have defined the edges in Rauzy graphs to be *oriented forwards*, i.e. $u \to v$ whenever $uc' = cv = vc'$ for some $c, c' \in \mathcal{A}$ (alternatively, one can use the window condition). However, for doubly infinite words it also makes sense to consider an edge $v \to u$, since we can also shift backwards. This does not have much impact on the results, except that eventually periodic words no longer exist for doubly infinite words, and so the Benjamini-Schramm limit is always a delta function on a cyclic or infinite-cyclic graph. However, the cyclic graph is now effectively unoriented, because for every $u \to v$ there is also a $v \to u$.

6.3. The labelled case for words with uniform frequencies

When we consider the limit of Rauzy graphs of singly infinite words, we arrive at a doubly infinite line graph. There is a natural reason for this:

Definition 6.3.1: Natural extension
The *natural extension* of a system (X, T) is an invertible system $(\widetilde{X}, \widetilde{T})$ with the universal property that if (Y, T') is another invertible system and there is a map from (Y, T') into (X, T), then (Y, T') uniquely factors through $(\widetilde{X}, \widetilde{T})$. That is, we have that the diagram below commutes. [DHS99, §9.2]

$$\begin{tikzcd}
(Y, T') \arrow[r, dashed, "\exists! f"] \arrow[dr, "\varphi"'] & (\widetilde{X}, \widetilde{T}) \arrow[d, "\pi"] \\
& (X, T)
\end{tikzcd}$$

And if (X, T) is uniquely ergodic, then so is $(\widetilde{X}, \widetilde{T})$. [BCD+89]

The universal definition is satisfying and informative in its own way, but in the symbolic setting it is also very easy to construct the natural extension. In the case of the unilateral shift, the natural extension is the projective limit of the diagram (all arrows are the left shift)

$$\mathcal{A}^{\mathbb{N}} \xrightarrow{\sim} \mathcal{A}^{-0} \times \mathcal{A}^{\mathbb{N}} \xrightarrow{\sim} \mathcal{A}^{-1} \times \mathcal{A}^{\mathbb{N}} \xrightarrow{\sim} \dots$$

And $\widetilde{X}$ is isomorphic to (or can be taken to be equal to) $\mathcal{A}^{-\mathbb{N}} \times \mathcal{A}^{\mathbb{N}}$.

When we take the Benjamini-Schramm limit, we wind up randomly rooting at some point in a line and taking a large symmetric neighbourhood. This mimics the construction of the natural extension, by effectively replacing a singly infinite line with a limit of lines which extend arbitrarily far to the left of the root.

The frequencies on the natural extension

As mentioned, if a one-sided subshift is uniquely ergodic then the measure is just governed by the frequencies:

$$\mu(u) = \lim_{n} \frac{|\alpha_0 \alpha_1 \dots \alpha_{n-1}|_u}{n}, \quad \alpha = S^k \omega$$

and the unique measure on the natural extension is more-or-less the same. It is a shift-invariant measure fully determined by its values on cylinders, and by Oxtoby's uniform ergodic theorem (Theorem 3.2.4), for $\widetilde{\mu}$-almost-any $\widetilde{\omega}$, the forward frequencies $\frac{1}{n+1}\sum_{i=0}^{n} I_u(S^{i+k}\widetilde{\omega})$ converge to $\widetilde{\mu}([u]) = \mu([u])$, and since S is now invertible, $\mu(A) = \mu(SA)$, so that the backwards frequencies $\frac{1}{n+1}\sum_{i=0}^{n} I_u(S^{-i+k}\widetilde{\omega})$ also converge to $\mu([u])$, and so the symmetric neighbourhoods

$$\mathrm{freq}_u(\widetilde{\omega}) := \lim_n \frac{|\alpha_{-n}\dots\alpha_{-1}\alpha_{-0}\times\alpha_0\alpha_1\dots\alpha_n|_u}{2n+2},\ \ \alpha = S^k(\widetilde{\omega})$$

also converge to $\mu([u])$.

However, these systems can be different topologically. For instance, eventually periodic bi-infinite words cannot exist; so if ω eventually periodic, then it is not always possible to view any $\widetilde{\omega}$ as an extension of ω. To illustrate this, consider the case of $c(ab)^\infty$: It is obvious that with the sole exception of the word "c", c never occurs as the last letter of any subword in $L_n(\omega)$, and so any two-sided infinite word containing ω as an infinite subword necessarily has a larger language than ω (it contains a subword ending in c). The issue is that, measure-theoretically, *non-recurrent subwords* have measure (= frequency) zero, and so they do not really exist from the measure-theoretic view.

But, if ω is recurrent (that is, every subword occurs infinitely often), then $\widetilde{\omega}$ can be selected so that ω is an infinite subword of $\widetilde{\omega}$, as mentioned by [DHS99][24]. So for any ω with uniform frequencies, we can take an $\widetilde{\omega}$ with the same frequencies and with the added requirement that ω is recurrent, we can choose $\widetilde{\omega}$ such that the projection from $\mathcal{A}^{-\mathbb{N}}\times\mathcal{A}^{\mathbb{N}}$ to $\mathcal{A}^{\mathbb{N}}$ of $\widetilde{\omega}$ yields ω.

Proposition 6.3.2: Uniform frequencies and recurrence
If ω is recurrent and has uniform frequencies, then there is an $\widetilde{\omega}$ which generates the natural extension of X_ω and $\widetilde{\omega}$ is an extension of ω.

Proof of the uniquely ergodic case

Theorem 6.3.3: Uniquely ergodic case when $|L_n(\omega)| \to \infty$
If ω is a low-complexity word with uniform frequencies and if $|L_n(\omega)| \to \infty$, then the Benjamini-Schramm limit of the labelled Rauzy graphs $\vec{\mathcal{R}}^n(\omega)$ exists and can be identified with the unique S-invariant measure on $(\widetilde{X}, S)$ from $\mathcal{A}^{\mathbb{Z}}$ to $\vec{\mathcal{G}}_\bullet$.

Sketch: The essence of the argument is that by Lemma 6.1.2 large Rauzy graphs have $1-\epsilon$ of their mass concentrated on very long lines, and when these lines are long enough, we can use an analogue of the uniform frequencies formula over them to get that the frequency of any word u over these long lines, $f_M(u)$, can be made within $\phi > 0$ of $\mu([u])$, i.e. $|f_M(u) - \mu([u])| < \phi$. Since the frequency of u over the whole graph, $f(u)$, is approximately equal to $f_M(u)$, because $(1-\epsilon)f_M(u) \le f(u) \le f_M(u) + \epsilon$, we get that

[24] Durand, Host, and Skau define subshifts with the requirement that $S : X \circlearrowleft$ must be onto; this forces their subshifts to be recurrent.

$$|f(u) - f_M(u)| \leq \epsilon, \quad \text{and} \quad |f_M(u) - \mu([u])| < \phi$$

and so $|f(u) - \mu([u])| < \phi + \epsilon$. From this, we will conclude that the limiting distribution of $\overrightarrow{\mathcal{G}}_\bullet$ determined by neighbourhoods can be identified with the unique measure μ on the shift associated to $\widetilde{\omega}$. •

Proof: First, assume that ω is recurrent; we will show that this assumption is not necessary after demonstrating the core of the proof.

We start by fixing u to be a word of even length (the odd case will be derived from the even case, since for u of odd length $\mu([u]) = \sum_{a\in\mathcal{A}} \mu([ua])$). Take $\widetilde{\omega}$ to be a two-sided extension of ω (or simply ω if ω was two-sided to begin with) and take its unique ergodic measure Definition 6.3.1. Using the definition of uniform frequencies Definition 3.2.5, and fixing an error bound $\phi > 0$, we take $M(u, \phi)$ such that

$$\forall_{2r+2>M(u,\phi)}. \left|\sup_k \left\{ \frac{|\alpha_{-r+1}\alpha_{-r+2}\dots\alpha_{-0} \times \alpha_0\alpha_1\dots\alpha_r|_u}{2r+2} \right\} - \mu([u])\right| < \phi, \text{ where } \alpha = S^k(\widetilde{\omega})$$

Now, by Lemma 6.1.2 part (3) we have that there is some N such that for $n > N$ the $|u|$-trimmed components λ_i' (the lines in the graph viewed as subgraphs) of length exceeding $M + |u|$ can be made to account for $1 - \epsilon$ of the Rauzy graphs. Now, these line subgraphs can be identified with their sequences of edges $\lambda_i = (e_{i,j})_j$, and these edges carry labels corresponding to sequences symbols in $\mathcal{A}$ which appear consecutively in $\widetilde{\omega}$. So the edges can be viewed as subwords of $\widetilde{\omega}$ of length exceeding $M + |u|$. We will denote this word by $\ell(\lambda_i)$, where ℓ is the labeling function on the Rauzy graph. We have for every λ_i with $\text{length}(\lambda_i) > M + |u|$ that

$$\left|\frac{\#\{\text{occurrences of } u \text{ in } \lambda_i\}}{\text{length}(\lambda_i)} - \mu([u])\right| < \phi$$

and denoting by $\{\zeta_i\} = \{\lambda_i : \text{length}(\lambda_i) > M + 2|u|\}$, we have that

$$\left|\frac{\sum_i \#\{\text{occurrences of } u \text{ in } \zeta_i\}}{\sum_i \text{length}(\zeta_i)} - \mu([u])\right| < \phi$$

But occurrences of the word u in $\ell(\lambda_i)$ are exactly neighbourhoods $D_{\frac{|u|}{2}}(u')$ within the line graphs λ_i', and denoting by $T_{\frac{|u|}{2}}(\lambda_i)'$ the $\frac{|u|}{2}$-trimmed subgraph of λ_i, we can rewrite this as

$$\frac{\sum_i \#\{\text{occurrences of } u \text{ in } \zeta_i\}}{\sum_i \text{length}(\zeta_i)} = \frac{\sum_{v\in\bigcup_i T_{\frac{|u|}{2}}(\zeta_i')} I_{B_{|u|/2}(u')}([(\overrightarrow{\mathcal{R}}^n(\omega), v)])}{\sum_i |\zeta_i'|}$$

For convenience, let us denote this average over the lines of length exceeding M by $f_M^n(u)$. We will similarly take the frequency over the whole Rauzy graph to be

$$f^n(u) := \frac{\sum_{v\in\overrightarrow{\mathcal{R}}^n(\omega)} I_{B_{|u|/2}(u')}([(\overrightarrow{\mathcal{R}}^n(\omega), v)])}{|L_n(\omega)|}$$

We have shown that $|f_M^n(u) - \mu([u])| < \phi$, and now we will show that $|f_M^n(u) - f^n(u)| < \epsilon$.

Clearly $(\sum_i |\zeta_i'|)\, f_M^n(u) \le |L_n(\omega)|\, f_M^n$, and since by Lemma 6.1.2 we have $\frac{\sum_i |\zeta_i'|}{|L_n(\omega)|} > 1-\epsilon$, we get that $(1-\epsilon) f_M^n(u) \le f^n(u)$. But we can also get an upper bound on $f^n(u)$ by looking at

$$\sum_{v\in\overrightarrow{\mathcal{R}}^n(\omega)} I_{B_{|u|/2}(u')}([(\overrightarrow{\mathcal{R}}^n(\omega),v)]) \le \sum_{v\in\bigcup_i T_{\frac{|u|}{2}}(\zeta_i')} I_{B_{|u|/2}(u')}([(\overrightarrow{\mathcal{R}}^n(\omega),v)]) + \sum_{v\in\overrightarrow{\mathcal{R}}^n(\omega)\setminus\left(\bigcup_i T_{\frac{|u|}{2}}(\zeta_i')\right)} 1$$

dividing by $|L_n(\omega)|$, we get

$$f^n(u) \le f_M^n(u)\frac{\sum_i |\zeta_i'|}{|L_n(\omega)|} + \frac{|L_n(\omega)| - \left|\bigcup_i T_{\frac{|u|}{2}}(\zeta_i)\right|}{|L_n(\omega)|}$$

but by part (3) of our lemma $\frac{|L_n(\omega)| - \left|\bigcup_i T_{\frac{|u|}{2}}(\zeta_i)\right|}{|L_n(\omega)|} < \epsilon$, and also $\sum_i \frac{|\zeta_i|}{|L_n(\omega)|} \le 1$ so

$$f^n(u) \le f_M^n(u)\frac{\sum_i |\zeta_i'|}{|L_n(\omega)|} + \frac{|L_n(\omega)| - \left|\bigcup_i T_{\frac{|u|}{2}}(\zeta_i)\right|}{|L_n(\omega)|} \le f_M^n(u) + \epsilon.$$

And as such, we get that

$$(1-\epsilon) f_M^n(u) \le f^n(u) \le f_M^n(u) + \epsilon$$

and since $0 \le f_M^n(u) \le 1$, we get that $|f^n(u) - f_M^n(u)| \le \epsilon$.

But we already showed that $|f_M^n(u) - \mu(u)| < \phi$, and so we get that(!)

$$|f^n(u) - \mu([u])| \le \epsilon + \phi$$

Recapping, the fact that $|f^n(u) - \mu([u])| \le \epsilon + \phi$ and ϵ and ϕ were arbitrarily chosen, so we get that for every word u of even length that

$$f^n(u) = \frac{\sum_{v\in\overrightarrow{\mathcal{R}}^n(\omega)} I_{B_{|u|/2}(u')}([(\overrightarrow{\mathcal{R}}^n(\omega),v)])}{|L_n(\omega)|} \to \mu([u])$$

but as mentioned, this determines the measure everywhere, since for w of odd length, $\mu([w]) = \sum_{a\in\mathcal{A}} \mu([wa])$. This concludes the proof and shows that the Benjamini-Schramm limit can be identified with the (unique) ergodic measure μ on the two-sided extension of X_ω, viewing the bi-infinite words as the space of edge-labelings of the line-graph. It is also easily seen that this measure is R-invariant, as this comes from S-invariance of μ.

The same argument holds if μ is not recurrent. As mentioned, for any factor u which occurs only finitely many times in ω, $\mathrm{freq}_u(\omega) = 0$. As a result (any by the definition of uniform frequencies), for any non-recurrent u and for any $\epsilon > 0$ we have that there is some n such that

$$\forall_{\alpha=S^k\omega} \frac{|\alpha_0\alpha_2\ldots\alpha_{n-1}|_u}{n} < \epsilon$$

and since $\mu([u]) = \widetilde{\mu}([u]) = 0$, for any generic $\widetilde{\omega}$, the frequencies of u in $\widetilde{\omega}$ are also 0. The result of this is that, while in the non-recurrent case, $\widetilde{\omega}$ can no-longer be thought of as an extension of ω, it is still a two-sided infinite word with the same statistical properties as ω, carrying the same distribution on finite words, which is what the Benjamini-Schramm limit actually captures. One can see that the difference between the frequency of u in a word in $L_n(\omega)$ and a word in $L_n(\widetilde{\omega})$ can be made within ϵ', so that the argument above is ultimately unaffected (in the proof, an application of the triangle inequality gives that you can replace ϕ with $\phi + \epsilon$) — that is, the Rauzy graphs of $\widetilde{\omega}$ and ω, while different, share the same limit. $\square$

The finite case

With the main case out of the way, we discuss the finite case. The unlabelled convergence was already shown, but in the labelled case there is oscillatory behavior on the labels. Consider the Rauzy graphs of the word $ccc(abc)^\infty = cccabcabcabc\ldots$.

$$\begin{tikzcd}
 & cca \arrow[r, "\text{b}"] & cab \arrow[dr, "\text{c}"] & \\
ccc \arrow[ur, "\text{a}"] & & bca \arrow[u, "\text{b}"] & abc \arrow[l, "\text{a}"]
\end{tikzcd}$$

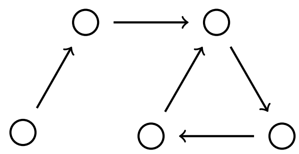
$$\begin{tikzcd}
 & ccab \arrow[r, "\text{c}"] & cabc \arrow[dr, "\text{a}"] & \\
ccca \arrow[ur, "\text{b}"] & & bcab \arrow[u, "\text{c}"] & abca \arrow[l, "\text{b}"]
\end{tikzcd}$$

$$\begin{tikzcd}
 & ccabc \arrow[r, "\text{a}"] & cabca \arrow[dr, "\text{b}"] & \\
cccab \arrow[ur, "\text{c}"] & & bcabc \arrow[u, "\text{a}"] & abcab \arrow[l, "\text{c}"]
\end{tikzcd}$$

In the case of periodic (= finite and *recurrent*) words, the oscillatory behavior is negated by the symmetries of the cyclic graph, so that that the Benjamini-Schramm limit exists and can be identified with the shift-invariant measure on X_ω. However, in the case of eventually periodic words, where the unlabelled graph is rigid, this does not happen, and one gets p distinct subsequential Benjamini-Schramm limits, where p is the period of $S^k\omega$.

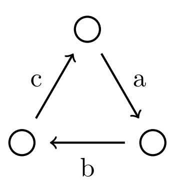
Theorem 6.3.4: Labelled case when $|L_n(\omega)| < M$
When $|L_n(\omega)|$ is bounded ω is either periodic or eventually periodic. In the case where ω is periodic, then the Benjamini-Schramm limit exists and can be identified with the uniform distribution on shifts of ω. In the eventually periodic case, the Benjamini-Schramm limit does not exist, but we do have p subsequential limits (where p is the period of $S^k\omega$), which we can describe. •

Proof: **Non-recurrent case:** We first handle the eventually periodic case. Recall that as unlabelled graphs, we get convergence to a graph that resembles the one below (where the length of the loop and the handle are the two variables)



Let k be the smallest integer such that $S^k\omega$ is periodic, and denote by p the period of $S^k\omega$. For $n > max(p,k)$, we have that all of the non-recurrent words are captured within the first vertex of $\overrightarrow{\mathcal{R}}^n(\omega)$ (the vertex $\omega_0\omega_1\ldots\omega_{n-1}$) so all edges are labelled with transitions sliding into recurrent words from the periodic component of ω. But, since the first edge (oriented out of $\omega_0\ldots\omega_{n-1}$) is labelled with ω_n, all other edges are fully determined by this edge, because the rest of the graph only has a single infinite walk, but as n evolves the labels $\omega_n\omega_{n+1}\ldots\omega_{n+k+p-1} \mid S^k\omega$ cycle through p different configurations (since it is a subword of a word with period p and it has length larger than p). The resulting graphs equipped with these different labels are rigid, so that for each $0 \leq j < p$, the subsequence of graphs (modulo vertex labelings) $\left(\overrightarrow{\mathcal{R}}^{n+ip+j}(\omega)\right)_i$ is constant, and each rooting yields a distinct rooted graph. These are all of the possible configurations of the graphs, and so there cannot be any other subsequential limits. So the Benjamini-Schramm limit does not exist, but there are exactly p subsequential limits, and the subsequential limits are uniform measures on the distinct symmetry-free rootings of a cycle with a path leading into it, with labels coming from the periodic component of ω.

Periodic case: In the periodic case, the answer is simpler. From the unlabelled case, we know the limit graph to be a cycle (of length p), and all unlabelled rootings are isomorphic. After adding labels, for $n > p$, the unrooted graph is a cycle labelled by $\omega_i\omega_{i+1}\ldots\omega_{i+p-1}$ (the word is cyclic, so that i just rotates the graph, but all graphs are isomorphic). Each root of the graph yields a distinct rooted graph from p possible rootings, so one gets that the Benjamini-Schramm limit is a uniform measure on these rooting, which can be identified with the uniform measure on the shift associated to ω.



$\cong$ Unif($\{abcabc\ldots, bcabca\ldots, cabcab\ldots\}$)

$\square$

6.4. The case of multiple minimal subshifts

There is no polished answer for the case where X_ω contains more than one minimal subsystem, however we can give some insights. To this end, there has been fairly recent progress which determined a bound on the number of ergodic measures such a system can have, which turns out to have useful applications.

Theorem 6.4.1: Low-complexity words have finitely many ergodic measures [CK19, Thm 1.1]
Suppose that ω is an infinite word and that there is a $K \in \mathbb{N}$ such that

$$\liminf_n \frac{|L_n(\omega)|}{n} < K$$

then (X_ω, S) has at most $K-1$ distinct non-atomic ergodic measures.

This is a strengthening of a result of Boshernitzan [Bos85] for when uniformly recurrent words had a unique ergodic measure. There were later generalizations to bounding the number of ergodic measures for this class of words [FM10], and the result above, due to Cyr and Kra [CK19], extended this bound to non-minimal systems (and also a more general class of measures than stated above).

We can use this to better understand $\mathcal{M}(X_\omega, S)$.

Factoring the space of invariant measures

From Theorem 6.4.1 we get that the set of ergodic measures on the subshift of a low-complexity word, $\mathcal{E}(X_\omega, S)$, is always finite, and from this we get that there are finitely many minimal subsystems of X_ω. By Krylov-Bogoliouboff each minimal system supports at least one ergodic measure, and so there cannot be infinitely many minimal subsystems. Moreover, denoting by $\{A_1, \ldots, A_p\}$ the set of all (necessarily disjoint) minimal systems in X_ω, we have that $\mathcal{E}(X, S) = \bigcup_i \mathcal{E}(A_i, S)$. That is, every ergodic measure is concentrated within a single minimal subshift.

Proposition 6.4.2: Decomposing $\mathcal{M}(X, T)$
Suppose that X is a compact metric space and that $\mathcal{E}(X, T)$ is finite. Then $\mathcal{E}(X, T) = \bigcup_i \mathcal{E}(Y_i, T)$, where $\{Y_i\}$ is the finite collection of minimal subsets of X.

Proof: Select $\mu \in \mathcal{E}(X, T)$, and suppose that $Y \subseteq X$ is a minimal subset and that $\mu(Y) > 0$. Then we have that the normalized restriction of μ to Y, $\frac{\mu|_Y}{\mu(Y)}$ is also an invariant measure on Y, and since $\mathcal{E}(Y, T) \subseteq \mathcal{E}(X, T)$, and since $\mathcal{E}(X, T)$ is finite by assumption, by Choquet's theorem we can represent $\frac{\mu|_Y}{\mu(Y)}$ as a convex combination of the measures $\{\nu_i\} = \mathcal{E}(Y, T)$

$$\frac{\mu\big|_Y}{\mu(Y)} = \sum_{i=0}^{k-1} \alpha_i \nu_i, \quad \sum_{i=0}^{k-1} \alpha_i = 1, \quad \alpha_i \geq 0$$

But clearly the restricted and normalized measure $\frac{\mu|_Y}{\mu(Y)}$ has more null sets than μ, so that

$$\mu \gg \frac{\mu\big|_Y}{\mu(Y)} = \sum_{i=0}^{k-1} \alpha_i \nu_i$$

and for at least one i, we know that $\alpha_i > 0$, and for that i we get that $\mu \gg \nu_i$. But by assumption μ and ν_i were ergodic, and so $\nu_i \ll \mu$ implies that $\nu_i = \mu$. Therefore, every element of $\mathcal{E}(X,T)$ lives within some $\mathcal{E}(Y,T)$. $\square$

Remark 6.4.3: The finiteness assumption cannot be dropped
Note that the finiteness assumption on $\mathcal{E}(X,T)$ in Proposition 6.4.2 cannot be dropped. For instance, the Bernoulli-$\frac{1}{2}$ measure on $\{0,1\}^{\mathbb{N}}$ is ergodic and assigns positive probability to every non-empty open set — as a result, it cannot be equal to a restriction to a proper compact subset of $\{0,1\}^{\mathbb{N}}$, and $(\{0,1\}^{\mathbb{N}}, S)$ itself is not minimal.

More generally, once $\mathcal{E}(X,T)$ becomes infinite, the topology becomes non-trivial, and it can be the case that the ergodic measures in $\mathcal{E}(X,T)$ which are supported on all of X are a dense G_δ set in $\mathcal{M}(X,T)$. This is the case for $\mathcal{M}(\mathcal{A}^{\mathbb{N}}, S)$. (See, for instance, [DGS76] Proposition 21.9 and 21.11) •

Application to Rauzy graphs

Now, suppose that ω is a word with finitely many ergodic measures. It is immediately obvious that for any S-invariant measure μ that $\mu(X_\omega \setminus \bigcup Y_i) = 0$, where Y_i are all the minimal subsets. Also, since the minimal components Y_i are disjoint family of compact sets, there is some n such that the length n cylinders in $L_n(\omega)$ form a disjoint set of covers for the Y_i. One way to see this, is that if Y_i and Y_j, $i \neq j$ had cylinder sets of arbitrary length which intersected both, without loss of generality this a sequence of cylinders could be taken to be a nested chain, and the Cantor intersection theorem applied to this sequence of cylinders (which are closed) would give that $Y_i \cap Y_j \neq \emptyset$. Therefore, so pairwise there must be some n such that a cylinder set intersects at most one of Y_i or Y_j — we can apply this finitely many times to every pair and take a maximum n. What this says is that there is some N such that for all $n > N$, $L_n(Y_i) \cap L_n(Y_j) = \emptyset$ for $i \neq j$. This has an interpretation in terms of Rauzy graphs.

Proposition 6.4.4: Decomposition of Rauzy graphs
If ω is a low-complexity word, so that $|\mathcal{E}(X_\omega, S)| < K$ by Theorem 6.4.1, then there is some $N > 0$ such that for all $n > N$, the minimal subsets $Y_i \subseteq X$ have disjoint covers in the form of length n cylinders in $L_n(\omega)$, and for every S-invariant measure μ, $\mu([u]) = 0$ for $u \in (L_n(\omega) \setminus \bigcup_i L_n(Y_i))$. And the Rauzy graphs $\vec{\mathcal{R}}^n(\omega)$ have that

$$\vec{\mathcal{R}}^n(\omega) = \left(\dot{\bigcup_i} \vec{\mathcal{R}}^n(y_i)\right) \cup \left(\vec{\mathcal{R}}^n(\omega) \setminus \dot{\bigcup_i} \vec{\mathcal{R}}^n(y_i)\right)$$

The minimal components are each disjoint and support their own ergodic measures, and the left-over words which connect the minimal components together (the vertices themselves, viewed as finite words) have measure zero with respect to every invariant measure on (X_ω, S). However, for every oriented path directed into a $\vec{\mathcal{R}}^n(y_i)$, the labels on the n edges leading into $\vec{\mathcal{R}}^n(y_i)$ encode subwords of y_i, and since every minimal component $\vec{\mathcal{R}}^n(y_i)$ has at most linear (in n) volume, the paths leading into the minimal components contribute a non-negligible amount to any Benjamini-Schramm subsequential limit (if all minimal systems are periodic, then these connecting paths can even dominate the limit). •

We can now provide a few examples to illustrate how this works, and also why the space of subsequential Benjamini-Schramm limits fails to be pleasant in this case.

COMPUTATIONS OF THE LIMIT FOR ILL-BEHAVED WORDS

We begin with an example with several subsequential limits, where no ergodic measure on the shift-system can be achieved. We follow this example with one where the Benjamini-Schramm limit exists, even when the underlying system has two minimal subsystems. We conclude with a modification of the second example, where reversing the orientation causes the Benjamini-Schramm limit (which used to exist) to fail to exist.

Example 6.4.5: A non-minimal word with two ergodic measures
There is an easy way to construct a word with two ergodic measures, which is forming a bi-infinite word from two singly-infinite words where each is known to be uniquely ergodic. In this example, we can take the Thue-Morse word Example 2.4.8 (the unique invariant measure was found by [Dek92]), and we will take one copy of $\mathfrak{t}$ to be over the alphabet $\{0,1\}$ and another copy over $\{a,b\}$. We will then form the word

$$\omega = \dots baababba \times 01101001\dots$$

Alternatively [Dur10, §3.4], one can define

$$\sigma : \{0,1,a,b\}^* \circlearrowleft, \quad \sigma = \left\{ \begin{array}{ll} 0 & \mapsto 01 \\ 1 & \mapsto 10 \\ a & \mapsto ab \\ b & \mapsto ba \end{array} \right\}$$

And then ω is a fixed point of σ^2, with $\omega = \lim_n \sigma^{2n}(a \times 0)$. In fact,

$$\overline{\mathcal{O}(\omega)} = \overline{\mathcal{O}(\widetilde{\mathfrak{t}}_{01})} \,\dot\cup\, \overline{\mathcal{O}(\widetilde{\mathfrak{t}}_{ab})} \,\dot\cup\, \mathcal{O}(\omega)$$

It can be computed [CN10, §4.10.4] that for the Thue-Morse word

$$\liminf_n \frac{|L_n(\mathfrak{t})|}{n} = 3, \quad \limsup_n \frac{|L_n(\mathfrak{t})|}{n} = \frac{10}{3}$$

Using this we can compute two different Benjamini-Schramm subsequential limits for the sequence $(\overrightarrow{\mathcal{R}}^n(\omega))$. Despite the symmetry of the word, the Rauzy graphs are not quite symmetrical, because the orientation of the graph interacts with the left-versus-right relationship between $\mathfrak{t}_{ab}$ and $\mathfrak{t}_{01}$. The Rauzy graph of ω contains $\overrightarrow{\mathcal{R}}^n(\mathfrak{t}_{ab})$ and $\overrightarrow{\mathcal{R}}^n(\mathfrak{t}_{01})$ as subgraphs, *plus* and oriented path connecting these two graphs with labels coming from $\mathfrak{t}_{01}$.

$$\overrightarrow{\mathcal{R}}^n(\mathfrak{t}_{01}) \ni [\dots abba\times] \xrightarrow{0} [\dots abba\times 0] \xrightarrow{1} [\dots abba\times 01] \xrightarrow{1} \cdots \xrightarrow{\mathfrak{t}_{01,n-3}} [ba\times 0110\dots] \xrightarrow{\mathfrak{t}_{01,n-2}} [a\times 0110\dots] \xrightarrow{\mathfrak{t}_{01,n-1}} [\times 0110\dots] \in \overrightarrow{\mathcal{R}}^n(\mathfrak{t}_{01})$$

As in the unlabelled case, the bifurcations in these graphs are sparse, so we can compare most neighbourhoods to words in $\{0,1,a,b\}^*$, and our result for words with uniform frequencies gives that we can approximate the frequencies in each component $\overrightarrow{\mathcal{R}}^n(\mathfrak{t}_{ab})$ and $\overrightarrow{\mathcal{R}}^n(\mathfrak{t}_{01})$ separately. But now, the path joining the components $\overrightarrow{\mathcal{R}}^n(\mathfrak{t}_{ab})$ and $\overrightarrow{\mathcal{R}}^n(\mathfrak{t}_{01})$ is of length n, and since we know that the Rauzy graphs of the components are roughly between $3n$ and $\frac{10}{3}n$, the path joining the graphs contributes to approximately 1/7th of the mass of the graph, and the distribution of words on it is approximately $\mu_{\mathfrak{t}_{01}}$, the unique ergodic measure on the Thue-Morse word, with the alphabet $\{0,1\}$. From this, we get that for a subsequence (n_k) such that $\frac{|L_{n_k}(\mathfrak{t})|}{n_k} \to 3$, that the Benjamini-Schramm limit is can be identified with

$$\frac{3\mu_{ab}}{7} + \frac{\mu_{01}}{7} + \frac{3\mu_{01}}{7}$$

With the left and right terms representing the measure along the Rauzy graphs of the two Thue-Morse words, and the middle term representing the measure on the path between them, which encodes the measure on the Thue-Morse word on $\{0,1\}$.

However, for a sequence (n_k) such that $\frac{|L_{n_k}(\mathfrak{t})|}{n_k} \to \frac{10}{3}$, we get that

$$\frac{(10/3)\mu_{ab}}{23/3} + \frac{3\mu_{01}}{23/3} + \frac{(10/3)\mu_{01}}{23/3}$$

Presumably, one can also get some number of convex combinations of these measures by taking different sub-sequences, perhaps even every convex combination. But regardless, the "symmetry" of the Rauzy graphs ensures that the mass of the $\{a,b\}$ and $\{0,1\}$ Rauzy graphs are the same, and so it would not be possible to have a subsequential Benjamini-Schramm limit supported entirely on μ_{01} or μ_{ab}. This demonstrates that in the non-minimal case, one cannot realize every measure, or even any ergodic measure, in $\mathcal{M}(X_\omega, S)$ as a Benjamini-Schramm limit of the corresponding Rauzy graphs. •

Example 6.4.6: The Benjamini-Schramm limit cannot see every subsystem
In a similar fashion to the argument above, we can also show that the Benjamini-Schramm limit of an infinite word might exist, even if $|\mathcal{E}(X_\omega, S)| \geq 2$. Take

$$\omega = (ab)^\infty \times \mathfrak{t}$$

that is, the bi-infinite word formed from one periodic word and the Thue-Morse word. There are two minimal subsystems of X_ω, which are unsurprisingly $\{(ab)^\infty, (ba)^\infty\}$ and $\overline{\mathcal{O}(\mathfrak{t})}$, but because the periodic system has bounded Rauzy graphs and the Thue-Morse system does not, the Benjamini-Schramm limit exists and can be identified with the unique ergodic measure on the Thue-Morse word. •

Example 6.4.7: The Rauzy graph with backward edges
Amusingly, if we modify the previous example, then the limit no longer exists. Take

$$\omega' = \mathfrak{t} \times (ab)^\infty$$

while the Rauzy graph of the periodic word is bounded, the path from $\vec{\mathcal{R}}^n(\mathfrak{t})$ to $\vec{\mathcal{R}}^n((ab)^\infty)$ now has n edges with labels alternating between a and b, and so by a similar argument to Example 6.4.5, there are subsequential Benjamini-Schramm limits converging to $\frac{3}{4}\mu + \frac{1}{4}\nu$ and $\frac{10/3}{13/3}\mu + \frac{3}{13/3}\nu$, where ν is the ergodic measure on the periodic word $(ab)^\infty$ and μ is the ergodic measure on the Thue-Morse word.

The same problem happens with ω from Example 6.4.6 if one considers the Rauzy graph with arrows oriented by S^{-1} instead of S. That is, if we consider the edges to be given by $u \leftarrow v$ if $S^{-1}v = u$, then the labels of the path connecting the two Rauzy graphs come from ν instead of μ. •

6.5. The case of a single minimal subsystem

As was seen in the previous section, there are serious topological issues if the word ω does not give a minimal system (or, at least, a system containing a single minimal subsystem). I believe that in the case of minimal non-uniquely ergodic systems that more can be done to classify the possible Benjamini-Schramm limits, and it may even be possible to identify the space of subsequential Benjamini-Schramm limits with $\mathcal{M}(X_\omega, S)$, though this has not been proven, and will be left to future work. To this end, I think that a careful reading of the K-disconnectability condition of Monteil [FM10] may provide progress in this direction.

References

[AKM65] R. L. Adler, A. G. Konheim, and M. H. McAndrew. Topological entropy. *Trans. Amer. Math. Soc.*, 114:309–319, 1965.

[AS03] Jean-Paul Allouche and Jeffrey Shallit. *Automatic sequences. Theory, applications, generalizations.* Cambridge: Cambridge University Press, 2003.

[BBN59] Gilbert Baumslag, W. W. Boone, and B. H. Neumann. Some unsolvable problems about elements and subgroups of groups. *Math. Scand.*, 7:191–201, 1959.

[BCD+89] L. A. Bunimovich, I. P. Cornfeld, R. L. Dobrushin, M. V. Jakobson, N. B. Maslova, Ya. B. Pesin, Ya. G. Sinaĭ, Yu. M. Sukhov, and A. M. Vershik. *Dynamical systems. II*, volume 2 of *Encyclopaedia of Mathematical Sciences*. Springer-Verlag, Berlin, 1989. Ergodic theory with applications to dynamical systems and statistical mechanics, Edited and with a preface by Sinaĭ, Translated from the Russian.

[Bil78] Patrick Billingsley. *Ergodic theory and information.* Robert E. Krieger Publishing Co., Huntington, N.Y., 1978. Reprint of the 1965 original.

[Bir31] G. D. Birkhoff. Proof of the ergodic theorem. *Proc. Natl. Acad. Sci. USA*, 17:656–660, 1931.

[BK32] G. D. Birkhoff and B. O. Koopman. Recent contributions to the ergodic theory. *Proc. Natl. Acad. Sci. USA*, 18:279–282, 1932.

[BKS91] Tim Bedford, Michael Keane, and Caroline Series, editors. *Ergodic theory, symbolic dynamics, and hyperbolic spaces.* Oxford Science Publications. The Clarendon Press, Oxford University Press, New York, 1991. Papers from the Workshop on Hyperbolic Geometry and Ergodic Theory held in Trieste, April 17–28, 1989.

[BNW08] Laurent Bartholdi, Markus Neuhauser, and Wolfgang Woess. Horocyclic products of trees. *J. Eur. Math. Soc. (JEMS)*, 10(3):771–816, 2008.

[Bos85] Michael Boshernitzan. A unique ergodicity of minimal symbolic flows with linear block growth. *J. Analyse Math.*, 44:77–96, 1984/85.

[Bow71] Rufus Bowen. Entropy for group endomorphisms and homogeneous spaces. *Trans. Amer. Math. Soc.*, 153:401–414, 1971.

[Bow15] Lewis Bowen. An overview of Benjamini-Schramm convergence in group theory and dynamics, 10 2015. Lecture given at Princeton University.

[BR17] Scott Balchin and Dan Rust. Computations for symbolic substitutions. *J. Integer Seq.*, 20(4):Art. 17.4.1, 36, 2017.

[BS02] Michael Brin and Garrett Stuck. *Introduction to dynamical systems.* Cambridge: Cambridge University Press, 2002.

[BS11] Itai Benjamini and Oded Schramm. Recurrence of distributional limits of finite planar graphs [mr1873300]. In *Selected works of Oded Schramm. Volume 1, 2*, Sel. Works Probab. Stat., pages 533–545. Springer, New York, 2011.

[Cas96] Julien Cassaigne. Special factors of sequences with linear subword complexity. In *Developments in language theory, II (Magdeburg, 1995)*, pages 25–34. World Sci. Publ., River Edge, NJ, 1996.

[Cho58] Gustave Choquet. Existence et unicit des reprsentations intgrales au moyen des points extrmaux dans les cnes convexes. *Sminaire Bourbaki*, 4:33–47, 1956-1958.

[CK19] Van Cyr and Bryna Kra. Counting generic measures for a subshift of linear growth. *J. Eur. Math. Soc. (JEMS)*, 21(2):355–380, 2019.

[CN08] Ethan M. Coven and Zbigniew H. Nitecki. On the genesis of symbolic dynamics as we know it. *Colloq. Math.*, 110(2):227–242, 2008.

[CN10] Julien Cassaigne and François Nicolas. Factor complexity. In *Combinatorics, automata and number theory*, volume 135 of *Encyclopedia Math. Appl.*, pages 163–247. Cambridge Univ. Press, Cambridge, 2010.

[Deh11] M. Dehn. Über unendliche diskontinuierliche Gruppen. *Math. Ann.*, 71(1):116–144, 1911.

[Dek92] F. M. Dekking. On the Thue-Morse measure. *Acta Univ. Carolin. Math. Phys.*, 33(2):35–40, 1992.

[DGS76] Manfred Denker, Christian Grillenberger, and Karl Sigmund. *Ergodic theory on compact spaces.* Lecture Notes in Mathematics, Vol. 527. Springer-Verlag, Berlin-New York, 1976.

[DHS99] F. Durand, B. Host, and C. Skau. Substitutional dynamical systems, Bratteli diagrams and dimension groups. *Ergodic Theory Dynam. Systems*, 19(4):953–993, 1999.

[Din70] E. I. Dinaburg. A correlation between topological entropy and metric entropy. *Dokl. Akad. Nauk SSSR*, 190:19–22, 1970.

[Dur10] Fabien Durand. Combinatorics on Bratteli diagrams and dynamical systems. In *Combinatorics, automata and number theory*, volume 135 of *Encyclopedia Math. Appl.*, pages 324–372. Cambridge Univ. Press, Cambridge, 2010.

[EL10] M. Einsiedler and E. Lindenstrauss. Diagonal actions on locally homogeneous spaces. In *Homogeneous flows, moduli spaces and arithmetic*, volume 10 of *Clay Math. Proc.*, pages 155–241. Amer. Math. Soc., Providence, RI, 2010.

[FM77] Jacob Feldman and Calvin C. Moore. Ergodic equivalence relations, cohomology, and von Neumann algebras. II. *Trans. Amer. Math. Soc.*, 234(2):325–359, 1977.

[FM10] Sébastien Ferenczi and Thierry Monteil. Infinite words with uniform frequencies, and invariant measures. In *Combinatorics, automata and number theory*, volume 135 of *Encyclopedia Math. Appl.*, pages 373–409. Cambridge Univ. Press, Cambridge, 2010.

[Fog02] N. Pytheas Fogg. *Substitutions in dynamics, arithmetics and combinatorics*, volume 1794 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, 2002. Edited by V. Berthé, S. Ferenczi, C. Mauduit and A. Siegel.

[Fri01] A. E. Frid. On factor graphs of DOL words [MR1760731 (2001d:05156)]. *Discrete Appl. Math.*, 114(1-3):121–130, 2001. Discrete analysis and operations research.

[Fur61] H. Furstenberg. Strict ergodicity and transformation of the torus. *Amer. J. Math.*, 83:573–601, 1961.

[GLN16] R. Grigorchuk, P.-H. Leemann, and T. Nagnibeda. Lamplighter groups, de brujin graphs, spider-web graphs and their spectra, May 2016. Preprint.

[Goo69] L. Wayne Goodwyn. Topological entropy bounds measure-theoretic entropy. *Proc. Amer. Math. Soc.*, 23:679–688, 1969.

[Gro87] M. Gromov. Hyperbolic groups. In *Essays in group theory*, volume 8 of *Math. Sci. Res. Inst. Publ.*, pages 75–263. Springer, New York, 1987.

[Hal56] Paul R. Halmos. *Lectures on ergodic theory.* Publications of the Mathematical Society of Japan, no. 3. The Mathematical Society of Japan, 1956.

[HK67] Frank Hahn and Yitzhak Katznelson. On the entropy of uniquely ergodic transformations. *Trans. Amer. Math. Soc.*, 126:335–360, 1967.

[HK83] Frank Harary and Peter Kovács. Regular graphs with given girth pair. *J. Graph Theory*, 7(2):209–218, 1983.

[Jew70] Robert I. Jewett. The prevalence of uniquely ergodic systems. *J. Math. Mech.*, 19:717–729, 1969/1970.

[Kai98] Vadim A. Kaimanovich. Hausdorff dimension of the harmonic measure on trees. *Ergodic Theory Dynam. Systems*, 18(3):631–660, 1998.

[Kai15] V. A. Kaimanovich. Invariance, quasi-invariance and unimodularity for random graphs. *Zap. Nauchn. Sem. S.-Peterburg. Otdel. Mat. Inst. Steklov. (POMI)*, 441(Veroyatnost'i Statistika. 22):210–238, 2015.

[Kai18] Vadim A. Kaimanovich. Circular slider graphs: de Bruijn, Kautz, Rauzy, lamplighters and spiders. In *Unimodularity in randomly generated graphs*, volume 719 of *Contemp. Math.*, pages 129–154. Amer. Math. Soc., Providence, RI, 2018.

[KB37] Nicolas Kryloff and Nicolas Bogoliouboff. La théorie générale de la mesure dans son application à l'étude des systèmes dynamiques de la mécanique non linéaire. *Ann. of Math. (2)*, 38(1):65–113, 1937.

[Kec95] Alexander S. Kechris. *Classical descriptive set theory*, volume 156 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 1995.

[KM40] M. Krein and D. Milman. On extreme points of regular convex sets. *Studia Math.*, 9:133–138, 1940.

[Kri72] Wolfgang Krieger. On unique ergodicity. pages 327–346, 1972.

[Lee16] Paul-Henry Leemann. *On subgroups and Schreier graphs of finitely generated groups.* PhD thesis, University of Geneva, 2016.

[LM95] Douglas Lind and Brian Marcus. *An introduction to symbolic dynamics and coding.* Cambridge: Cambridge University Press, 1995.

[Lov12] László Lovász. *Large networks and graph limits*, volume 60 of *American Mathematical Society Colloquium Publications*. American Mathematical Society, Providence, RI, 2012.

[LR10] Pierre Lecomte and Michel Rigo. Abstract numeration systems. In *Combinatorics, automata and number theory*, volume 135 of *Encyclopedia Math. Appl.*, pages 108–162. Cambridge Univ. Press, Cambridge, 2010.

[Mei08] John Meier. *Groups, graphs and trees*, volume 73 of *London Mathematical Society Student Texts*. Cambridge University Press, Cambridge, 2008. An introduction to the geometry of infinite groups.

[MH38] Marston Morse and Gustav A. Hedlund. Symbolic Dynamics. *Amer. J. Math.*, 60(4):815–866, 1938.

[NS60] V. V. Nemytskii and V. V. Stepanov. Qualitative theory of differential equations. (Princeton Mathematical Series). New Jersey: Princeton University Press. VIII, 523 p. (1960)., 1960.

[Orn70] Donald Ornstein. Bernoulli shifts with the same entropy are isomorphic. *Advances in Math.*, 4:337–352, 1970.

[Oxt52] John C. Oxtoby. Ergodic sets. *Bull. Amer. Math. Soc.*, 58:116–136, 1952.

[Par64] William Parry. Intrinsic Markov chains. *Trans. Amer. Math. Soc.*, 112:55–66, 1964.

[Pet83] Karl E. Petersen. *Ergodic Theory*. Cambridge Studies in Advanced Mathematics. Cambridge University Press, 1983.

[Phe01] Robert R. Phelps. *Lectures on Choquet's theorem*, volume 1757 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, second edition, 2001.

[Que10] Martine Queffélec. *Substitution dynamical systems—spectral analysis*, volume 1294 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, second edition, 2010.

[Rok60] V. A. Rokhlin. New progress in the theory of transformations with invariant measure. *Russ. Math. Surv.*, 15(4):1–22, 1960.

[Sha48] C. E. Shannon. A mathematical theory of communication. *Bell System Tech. J.*, 27:379–423, 623–656, 1948.

[Wie39] Norbert Wiener. The ergodic theorem. *Duke Math. J.*, 5(1):1–18, 1939.

[Woe14] W. Woess. What is a horocyclic product, and how is it related to lamplighters?, January 2014. Preprint.