

ENTROPY SENSITIVITY OF LANGUAGES DEFINED BY INFINITE AUTOMATA, VIA MARKOV CHAINS WITH FORBIDDEN TRANSITIONS

WILFRIED HUSS, ECATERINA SAVA, WOLFGANG WOESS

ABSTRACT. A language L over a finite alphabet Σ is growth-sensitive (or entropy sensitive) if forbidding any set of subwords F yields a sub-language L^F whose exponential growth rate (entropy) is smaller than that of L . Let (X, E, ℓ) be an infinite, oriented, labelled graph with label alphabet Σ . Considering the graph as an (infinite) automaton, We associate with any pair of vertices $x, y \in X$ the language $L_{x,y}$ consisting of all words that can be read as the labels along some path from x to y . Under suitable, general assumptions we prove that these languages are growth-sensitive. This is based on using Markov chains with forbidden transitions.

1. INTRODUCTION

Let Σ be a finite alphabet and Σ^* the set of all finite words over Σ , including the empty word ϵ . A language L over Σ is a subset of Σ^* . We denote by $|w|$ the length of the word w . The *growth* or *entropy* of L is

$$h(L) = \limsup_{n \rightarrow \infty} \frac{1}{n} \log |\{w \in L : |w| = n\}|.$$

All our languages will be infinite. For a finite, non-empty set $F \subset \Sigma^+ = \Sigma^* \setminus \{\epsilon\}$ consisting of subwords of elements of L , we let

$$L^F = \{w \in L : \text{no } v \in F \text{ is a subword of } w\}.$$

The issue addressed here is to provide conditions under which, for a class of languages associated with infinite graphs, $h(L^F) < h(L)$ strictly. If this holds for *any* set F of *forbidden subwords*, then the language L is called *growth sensitive* (or *entropy sensitive*).

Questions related with growth sensitivity have been considered in different context.

Date: October 20, 2009.

2000 Mathematics Subject Classification. 05C63, 37A35, 60J10, 68Q45.

Key words and phrases. Formal language, oriented graph, infinite sofic system, growth sensitivity, entropy, irreducible Markov chain, spectral radius.

W. Huss and W. Woess were supported by the Austrian Science Fund project FWF-P19115-N18. E. Sava was supported by the NAWI Graz project.

In *group theory*, in relation with regular normal forms of finitely generated groups, the study of growth-sensitivity has been proposed by GRIGORCHUK AND DE LA HARPE [8] as a tool for proving Hopfianity of a given group or class of groups, see also ARZHANTSEVA AND LYSENOK [1] and CECCHERINI-SILBERSTEIN AND SCARABOTTI [3].

In *symbolic dynamics*, the number $h(L)$ associated with a regular language accepted by a finite automaton with suitable properties appears as the *topological entropy* of a *sofic system*, see LIND AND MARCUS [10, Chapters 3 & 4]. Entropy sensitivity appears as the strict inequality between the entropies of an irreducible sofic shift and a proper subshift [10, Cor. 4.4.9].

Motivated by these bodies of work, CECCHERINI-SILBERSTEIN AND WOESS [5], [6], [4] have elaborated practicable criteria that guarantee growth-sensitivity of *context-free* languages.

The main result of the present note can be seen as a direct extension of [10, Cor. 4.4.9] to the entropies of infinite sofic systems; see below for further comments and references.

Our basic object is an infinite automaton, that is, an oriented, labelled graph (X, E, ℓ) whose edges are labelled by elements of a finite alphabet Σ . Each edge has the form $e = (x, a, y)$, where $e^- = x$ and $e^+ = y \in X$ are the initial and the terminal vertex of e and $\ell(e) = a \in \Sigma$ is its label. We will also write $x \xrightarrow{a} y$ for the edge $e = (x, a, y)$, or just $x \rightarrow y$ in situations where we do not care about the label. Multiple edges and loops are allowed, but two edges with the same end vertices must have distinct labels. Often, we shall refer to the graph just as X .

A *path* of length n in X is a sequence $\pi = e_1 e_2 \dots e_n$ of edges such that $e_i^+ = e_{i+1}^-$, for $i = 1, 2, \dots, n-1$. We say that it is a path from x to y , if $e_1^- = x$ and $e_n^+ = y$. The label $\ell(\pi)$ of π is the word $\ell(\pi) = \ell(e_1)\ell(e_2)\dots\ell(e_n) \in \Sigma^*$ that we read along the path. We also allow the empty path from x to x , whose label is the *empty word* $\epsilon \in \Sigma^*$. For $x, y \in V$, denote by $\Pi_{x,y}$ the set of all paths π from x to y in X .

The languages which we consider here are

$$L_{x,y} = \{\ell(\pi) \in \Sigma^* : \pi \in \Pi_{x,y}\}, \text{ where } x, y \in X.$$

That is, we consider the graph X as an infinite automaton (labelled digraph) with initial state x and terminal state y , so that $L_{x,y}$ is the language accepted by the automaton.

We say that (X, E, ℓ) is *deterministic*, if for every vertex x and every $a \in \Sigma$, there is at most one edge with initial point x and label a . A finite automaton can always be transformed into a deterministic one that accepts the same language, but in the infinite case, we shall need this as an assumption.

As in the finite case, we need an irreducibility assumption. The graph (X, E) is called *strongly connected*, if for every pair of vertices x, y , there is an (oriented !) path from x to y . Furthermore, we say that it is *uniformly connected*, if in addition the following holds.

- There is a constant K such that for every edge $x \rightarrow y$ there is a path from y to x with length at most K .

In the finite case, the two notions coincide. The *forward distance* $d^+(x, y)$ of $x, y \in X$ is the minimum length of a path from x to y . We write

$$h(X) = h(X, E, \ell) = \sup_{x, y \in X} h(L_{x, y})$$

and call this the entropy of our oriented, labelled graph. It is a well known and easy to prove fact that for a strongly connected graph, $h(L_{x, y}) = h(X)$ for all $x, y \in X$.

We also need a reasonable assumption on the set of forbidden subwords.

We say that a finite set $F \subset \Sigma^+$ is *relatively dense* in the graph X , if there is a constant D such that for every $x \in X$ there are $y \in X$ and $w \in F$ such that $d^+(x, y) \leq D$ and there is a path starting at y which has label w .

Theorem 1.1. *Suppose that (X, E, ℓ) is uniformly connected and deterministic with label alphabet Σ . Let $F \subset \Sigma^+$ be a finite, non-empty set which is relatively dense in X . Then*

$$\sup_{x, y \in X} h(L_{x, y}^F) < h(X) \quad \text{strictly.}$$

We say that X is *fully deterministic*, if for every $x \in X$ and $a \in \Sigma$, there is precisely one edge with initial point x and label a .

Corollary 1.2. *If (X, E, ℓ) is uniformly connected and fully deterministic then $L_{x, y}$ is growth-sensitive for all $x, y \in X$.*

Indeed, in this case, for every $x \in X$ and every $w \in \Sigma^*$, there is precisely one path with label w starting at x .

With our graph, we can consider the shift space which consists of all bi-infinite words over Σ that can be read along the edges of a bi-infinite path in X . The associated (topological) entropy is $h(L_{x, y})$, and this number is independent of x and y when the graph is strongly connected. See e.g. GUREVIČ [9], PETERSEN [11] or BOYLE, GUZZI AND GÓMEZ [2] for a selection of related work and references, and also the discussion in [10, §13.9].

Then we can consider the shift space consisting of all those bi-infinite words as above that do not contain any subword in F . Then the interpretation of Corollary 1.2 is that the associated entropy is strictly smaller than $h(X)$.

The theorem, once approached in the right way, is not hard to prove. It is based on the version of the Perron-Frobenius theorem for infinite non-negative matrices; see e.g. SENETA [13]. We shall first reformulate things in terms of Markov chains (“random walks”) and forbidden transitions.

2. MARKOV CHAINS AND FORBIDDEN TRANSITIONS

We now equip the oriented, labelled graph (X, E, ℓ) with additional data: with each edge $e = (x, a, y)$, we associate a probability $p(e) = p(x, a, y) \geq \alpha > 0$, where α is a fixed constant, such that

$$(1) \quad \sum_{e \in E: e^- = x} p(e) \leq 1 \quad \text{for every } x \in X.$$

Our assumption to have the uniform lower bound $p(e) \geq \alpha$ for each edge implies that the outdegree (number of outgoing edges) of each vertex is bounded by $1/\alpha$. We interpret $p(e)$ as the probability that a particle with current position $x = e^-$ moves in one (discrete) time unit along e to its end vertex $y = e^+$. Observing the successive random positions of the particle at the time instants $0, 1, 2, \dots$, we obtain a Markov chain with state space X whose one-step transition probabilities are

$$p(x, y) = \sum_{a \in \Sigma: (x, a, y) \in E} p(x, a, y).$$

We shall also want to record the edges, resp. their labels used in each step, which means to consider a Markov chain on a somewhat larger state space, but we will not need to formalise this in detail. In (1), we admit the possibility that $1 - \sum_y p(x, y) > 0$ for some x . This number is then interpreted as the probability that a particle positioned at x dies at the next step.

We write $p^{(n)}(x, y)$ for the probability that the particle starting at x is at position y after n steps. This is the (x, y) -element of the n -power P^n of the transition matrix $P = (p(x, y))_{x, y \in X}$. If X is strongly connected, then P is irreducible, and it is well-known that the number

$$\rho(P) = \limsup_{n \rightarrow \infty} p^{(n)}(x, y)^{1/n}$$

is independent of x and y . See once more [13]. Often, $\rho(P)$ is called the spectral radius of P . It is the parameter of exponential decay of the transition probabilities.

Let once more $F \subset \Sigma^+$ be finite. We interpret the elements of F as sequences of *forbidden transitions*. That is, we restrict the motion of the particle: at no time, it is allowed to traverse any path π with $\ell(\pi) \in F$ in k successive steps, where k is the length of π . We write $p_F^{(n)}(x, y)$ for the probability that the particle starting at x is at position y after n steps, without having made any such sequence of forbidden transitions. Let

$$\rho_{x,y}(P_F) = \limsup_{n \rightarrow \infty} p_F^{(n)}(x, y)^{1/n}, \quad x, y \in X.$$

These numbers are not necessarily independent of x and y , and they are not the elements of the n -matrix power of some substochastic matrix. In order to give an upper bound for the restricted transition probabilities $p_F^{(n)}(x, y)$, we first show the following.

Lemma 2.1. *Suppose that (X, E, l) is strongly connected with label alphabet Σ and equipped with transition probabilities $p(e) \geq \alpha > 0$, $e \in E$. Let $F \subset \Sigma^+$ be a finite, non-empty set which is relatively dense in X . Then there are $k \in \mathbb{N}$ and $\varepsilon_0 > 0$ such that*

$$\sum_{y \in X} p_F^{(k)}(x, y) \leq 1 - \varepsilon_0 \quad \text{for all } x \in X.$$

In other words, the transition matrix $Q = (p_F^{(k)}(x, y))_{x, y \in X}$ is strictly substochastic, with all row sums bounded by $1 - \varepsilon_0$.

Proof. Let $R = \max_{w \in F} |w|$, and let $D \in \mathbb{N}$ be the constant from the definition of relative denseness of F . Set $k = D + R$. For each $x \in X$, we can find a path π_1 from x to some $y \in X$ with length $d \leq D$ and a path π_2 starting at y which has label w . Let z be the endpoint of π_2 , and choose any path π_3 that starts at z and has length $k - d - |w|$. (Such a path exists by uniform connectedness.) Then let π be the path obtained by concatenating π_1 , π_2 and π_3 .

The probability that the Markov chain starting at x makes its first k steps along the edges of π is

$$\mathbb{P}(\pi) \geq \alpha^k = \varepsilon_0 > 0.$$

Hence

$$\sum_{y \in X} p_F^{(k)}(x, y) \leq \sum_{y \in X} p^{(k)}(x, y) - \mathbb{P}(\pi) \leq 1 - \varepsilon_0,$$

and this upper bound holds for every x . □

We prove the following result on sensitivity of the Markov chain with respect to forbidding the transitions in F .

Theorem 2.2. *Suppose that (X, E, ℓ) is strongly connected with label alphabet Σ and equipped with transition probabilities $p(e) \geq \alpha > 0$, $e \in E$. Let $F \subset \Sigma^+$ be a finite, non-empty set which is relatively dense in X . Then*

$$\sup_{x, y \in X} \rho_{x, y}(P_F) < \rho(P) \quad \text{strictly.}$$

Proof. We shall proceed in two steps.

Step 1. We assume that $P = (p(x, y))_{x, y \in X}$ is stochastic and that $\rho(P) = 1$.

Consider the matrix Q of Lemma 2.1. Let $Q^n = (q^{(n)}(x, y))_{x, y \in X}$ be its n -th matrix power. $q^{(n)}(x, y)$ is the probability that the Markov chain starting at x is in y at time nk and does not make any forbidden sequence of transitions in each of the discrete time intervals $[(j-1)k, jk]$ for $j \in \{1, \dots, n\}$. Therefore

$$p_F^{(nk)}(x, y) \leq q^{(n)}(x, y),$$

and also, by the same reasoning, for $i = 0, \dots, k-1$,

$$p_F^{(nk+i)}(x, y) \leq \sum_{z \in X} q^{(n)}(x, z) p_F^{(i)}(z, y), \quad i = 0, \dots, k-1.$$

Therefore, for every $x \in X$ and $i = 0, \dots, k-1$,

$$\sum_{y \in X} p_F^{(nk+i)}(x, y) \leq \sum_{z \in X} q^{(n)}(x, z) \underbrace{\sum_{y \in X} p_F^{(i)}(z, y)}_{\leq 1} \leq (1 - \varepsilon_0)^n,$$

since Lemma 2.1 implies that the row sums of the matrix power Q^n are bounded above by $(1 - \varepsilon_0)^n$. We conclude that

$$\limsup_{n \rightarrow \infty} p_F^{(nk+i)}(x, y)^{1/(nk+i)} \leq (1 - \varepsilon_0)^{1/k},$$

so that $\rho_{x,y}(P_F) \leq (1 - \varepsilon_0)^{1/k} = 1 - \varepsilon$, where $\varepsilon > 0$.

Step 2. General case.

We reduce this case to the previous one.

The matrix P acts on functions $h : X \rightarrow \mathbb{R}$ by $Ph(x) = \sum_y p(x, y)h(y)$. Also, P is irreducible, and every row of P has only finitely many non-zero entries. Given these properties, old results of PRUITT [12, Lemma 1 & Corollary to Theorem 2] imply that there exists a strictly positive function $h : X \rightarrow \mathbb{R}$ such that

$$Ph = \rho(P) \cdot h,$$

that is, h is $\rho(P)$ -harmonic. Consider now the h -transform of the transition probabilities $p(e)$ of P , $e = (x, a, y) \in E$, given by

$$p^h(e) = p^h(x, a, y) = \frac{p(x, a, y)h(y)}{\rho(P)h(x)}$$

and the associated transition matrix P^h with entries

$$p^h(x, y) = \sum_{a : (x, a, y) \in E} p^h(x, a, y).$$

The Markov chain associated with P^h is called the h -process.

Then $\rho(P^h) = 1$. Using uniform connectedness, we show that there is a constant $\bar{\alpha} > 0$ such that $p^h(e) \geq \bar{\alpha}$ for each $e = (x, a, y) \in E$. Indeed, for such an edge, there is $k \leq K$ such that $d^+(y, x) = k$, whence

$$\rho(P)^k h(y) = \sum_{z \in X} p^{(k)}(y, z) h(z) \geq \alpha^k h(x),$$

so that

$$p^h(x, a, y) \geq (\alpha / \rho(P))^{k+1}.$$

That is, we can choose $\bar{\alpha} = (\alpha/\rho(P))^{K+1}$. We see that with P^h we are now in the situation of Step 1. Thus, forbidding the transitions of F for the Markov chain with transition matrix P^h , we get $\rho_{x,y}(P_F^h) \leq 1 - \varepsilon$ for all $x, y \in X$, where $\varepsilon > 0$.

We now show that $\rho_{x,y}(P_F^h) = \rho_{x,y}(P)/\rho(P)$, which will conclude the proof.

For a path $\pi = e_1 \dots e_n$ from x to y , let (as above) $\mathbb{P}(\pi)$ be the probability that the original Markov chain traverses the edges of π in n successive steps, and let $\mathbb{P}^h(\pi)$ be the analogous probability with respect to the h -process. Then

$$\mathbb{P}^h(\pi) = \frac{\mathbb{P}(\pi)h(y)}{\rho(P)^n h(x)}.$$

Let us write $\Pi_{x,y}^n(\neg F)$ for the set of all paths π from x to y with length n for which $\ell(\pi)$ does not contain a subword in F . Then the n -step transition probabilities of the h -process with the transitions in F forbidden are

$$p_F^{h(n)}(x, y) = \sum_{\pi \in \Pi_{x,y}^n(\neg F)} \mathbb{P}^h(\pi) = \sum_{\pi \in \Pi_{x,y}^n(\neg F)} \frac{\mathbb{P}(\pi)h(y)}{\rho(P)^n h(x)} = \frac{p_F^{(n)}(x, y)h(y)}{\rho(P)^n h(x)}$$

Taking n -th roots and passing to the upper limit, we obtain the required identity. \square

With this result, is now easy to deduce Theorem 1.1.

Proof of Theorem 1.1. Since (X, E, l) is deterministic with label alphabet Σ , the outdegree of every $x \in X$ is at most $|\Sigma|$. Equip the edges of X with the transition probabilities $p(x, a, y) = 1/|\Sigma|$, when $(x, a, y) \in E$. Then the n -step transition probabilities of the resulting Markov chain are given by

$$p^{(n)}(x, y) = \frac{|\{w \in L_{x,y} : |w| = n\}|}{|\Sigma|^n}.$$

Therefore (because X is uniformly connected)

$$h(X) = h(L_{x,y}) = \limsup_{n \rightarrow \infty} \frac{1}{n} \log(p^n(x, y)|\Sigma|^n) = \log(\rho(P) \cdot |\Sigma|).$$

Analogously,

$$h(L_{x,y}^F) = \log(\rho_{x,y}(P_F) \cdot |\Sigma|).$$

By Theorem 2.2

$$\sup_{x,y \in X} \rho_{x,y}(P_F) < \rho(P),$$

and this implies that

$$\sup_{x,y \in X} h(L_{x,y}^F) < h(X)$$

strictly. \square

Application to pairs of groups and their Schreier graphs

Let G be a finitely generated group and K a (not necessary finitely generated) subgroup. Let also Σ be a finite alphabet and $\psi : \Sigma \rightarrow G$ be such that the set $\psi(\Sigma)$ generates G as a semigroup. We extend ψ to a monoid homomorphism from Σ^* to G by $\psi(w) = \psi(a_1) \cdots \psi(a_n)$, if $w = a_1 \dots a_n$ with $a_i \in \Sigma$ (and $\psi(\epsilon) = 1_G$). The mapping ψ is called a *semigroup presentation* of G in [7].

The *Schreier graph* $X = X(G, K, \psi)$ has vertex set

$$X = \{Kg : g \in G\},$$

the set of all right K -cosets in G , and the set of all labelled, directed edges E is given by

$$E = \{e = (x, a, y) : x = Kg, y = Kg\psi(a), \text{ where } g \in G, a \in \Sigma\}.$$

Note that the graph X is fully deterministic and uniformly connected.

The *word problem* of (G, K) with respect to ψ is the language

$$L(G, K, \psi) = \{w \in \Sigma^* : \psi(w) \in K\}.$$

If we consider the “root” vertex $o = K$ of the Schreier graph, then in the notation of the introduction, we have $L(G, K, \psi) = L_{o,o}$, compare with [7, Lemma 2.4].

We can therefore apply Theorem 1.1 and Corollary 1.2 to the graph $X(G, K, \psi)$ in order to deduce that

Corollary 2.3. *The word problem of the pair (G, K) with respect to any semigroup presentation ψ is growth sensitive (with respect to forbidding an arbitrary non-empty subset $F \subset \Sigma^*$).*

REFERENCES

- [1] G. N. Arzhantseva, I. G. Lysenok, *Growth tightness for word hyperbolic groups*. Math. Z. **241** (2002) 597–611.
- [2] M. Boyle, Mike, J. Buzzi, R. Gómez, *Almost isomorphism for countable state Markov shifts*. J. Reine Angew. Math. **592** (2006) 23–47.
- [3] T. Ceccherini-Silberstein, F. Scarabotti, *Random walks, entropy and hopfianity of free groups*. In *Random Walks and Geometry* (V. A. Kaimanovich, ed.) pp. 413–419, de Gruyter, Berlin, 2004.
- [4] T. Ceccherini-Silberstein, *Growth and ergodicity of context-free languages. II. The linear case*. Trans. Amer. Math. Soc. **359** (2007) 605–618.
- [5] T. Ceccherini-Silberstein, W. Woess, *Growth and ergodicity of context-free languages*. Trans. Amer. Math. Soc. **354** (2002) 4597–4625.
- [6] T. Ceccherini-Silberstein, W. Woess, *Growth-sensitivity of context-free languages*. Theoret. Comput. Sci. **307** (2003) 103–116.
- [7] T. Ceccherini-Silberstein, W. Woess, *Context-free pairs of groups I - Context-free pairs and graphs*. in preparation.
- [8] R. I. Grigorchuk, P. de la Harpe, *On problems related to growth, entropy, and spectrum in group theory*. J. Dynam. Control Systems **3** (1997) 51–89.

- [9] B. M. Gurevič, *Topological entropy of a countable Markov chain*. (Russian) Dokl. Akad. Nauk SSSR **187** (1969) 715–718; Engl. transl. in Soviet Math. Dokl. **10** (1969) 911–915.
- [10] D. Lind, B. Marcus, *An Introduction to Symbolic Dynamics and Coding*. Cambridge University Press, Cambridge, 1995.
- [11] K. Petersen, *Chains, entropy, coding*. Ergodic Theory Dynam. Systems **6** (1986) 415–448.
- [12] W. E. Pruitt, *Eigenvalues of Non-Negative Matrices*. Ann. Math. Statist. Volume 35, Number 4 (1964), 1797–1800.
- [13] E. Seneta, *Non-negative Matrices and Markov Chains*. Revised reprint of the second edition, Springer, New York, 2006.

INSTITUT FÜR MATHEMATISCHE STRUKTURTHEORIE
TECHNISCHE UNIVERSITÄT GRAZ
STEYRERGASSE 30, 8010 GRAZ, AUSTRIA

E-mail address: `huss@finanz.math.tu-graz.ac.at`, `sava@TUGraz.at`, `woess@TUGraz.at`