A symmetric entropy bound on the non-reconstruction regime of Markov chains on Galton-Watson trees

Marco Formentin[*] and Christof Külske [†]

March 17, 2009

Abstract

We give a criterion of the form $\mathbb{Q}(d)c(M) < 1$ for the non-reconstructability of tree-indexed q-state Markov chains obtained by broadcasting a signal from the root with a given transition matrix M. Here $c(M)$ is an explicit constant defined in terms of a $q-1$-dimensional variational problem over symmetric entropies, and $\mathbb{Q}(d)$ is the expected number of offspring on the Galton-Watson tree.

This result is equivalent to proving the extremality of the free boundary condition Gibbs measure within the corresponding Gibbs-simplex.

Our theorem holds for possibly non-reversible M and its proof is based on a general "Magic Recursion Formula" for expectations of a symmetrized relative entropy function, which invites their use as a Lyapunov function. In the case of the Potts model, the present theorem reproduces earlier results of the authors, with a simplified proof. In the case of the Ising model (where the method produces the correct reconstruction threshold) the argument becomes similar to the approach of Pemantle and Peres.

AMS 2000 subject classification: 60K35, 82B20, 82B44

Keywords: Broadcasting on trees, Gibbs measures, random tree, Galton-Watson tree, reconstruction problem, free boundary condition.

1 Introduction

The problem of reconstruction of Markov chains on d-ary trees has enjoyed much interest in recent years. There are multiple reasons for this, one of them being that it is a topic where people from information theory, pure probabilists, people dealing with Gibbs measures from a mathematical statistical mechanics point of view, and people from the theoretical physics side of statistical mechanics can meet and make contributions.

Indeed, starting with the symmetric Ising channel for which the reconstruction threshold was settled in [3, 7, 8], using different methods and increased generality w.r.t. the underlying tree, there have been publications by a.o. Borgs, Chayes, Janson, Mossel, Peres from the mathematics side [11, 12, 10, 2], deriving upper and lower bounds on reconstruction threshold for certain models of finite state tree-indexed Markov chains.

---
[*]Università degli Studi di Padova, Dipartimento di Matematica Pura ed Applicata, Via Trieste, 63 - 35121 Padova, Italia `formen@math.unipd.it`

[†]University of Groningen, Department of Mathematics and Computing Sciences, Nijenborgh 9, 9747 AC Groningen, The Netherlands `kuelske@math.rug.nl`, `http://www.math.rug.nl/∼kuelske/`

From the theoretical physics side let us highlight [15] on trees (see also [6] on graphs) which contains a discussion of the potential relevance of the reconstruction problem also to the glass problem. That paper also provides numerical values for the Potts model on the basis of extensive simulation results. The Potts model is interesting because, unlike the Ising model, the true reconstruction threshold behaves (respectively is expected to behave) non-trivially as a function of the degree d of the underlying d-ary tree and the number of states q. For a discussion of this see the conjectures in [15], the rigorous bounds in [4], the numerical estimates in [1], and in particular the proof in [16] showing that the Kesten-Stigum bound is not sharp if $q \geq 5$, and sharp if $q = 3$, for large enough d.

Now, our treatment is motivated in the generality of its setup by the questions raised and type of results given in [14], and technically somewhat inspired by [13, 4]. Indeed, for the Potts model the present paper reproduces the result of [4] (where moreover we provide numerical estimates on the reconstruction inverse temperature, also in the small q, small d regime.) However, in the present paper the focus is on generality, that is the universality of the type of estimate, and the structural clarity of the proof. It should be clear that the condition we provide can be easily implemented in any given model to produce numerical estimates on reconstruction thresholds.

The remainder of the paper is organized as follows. Section 2 contains the definition of the model and the statement of the theorem. Section 3 contains the proof.

2 Model and result

Consider an infinite rooted tree T having no leaves. For $v, w \in T$ we write $v \to w$, if w is the child of v, and we denote by $|v|$ the distance of a vertex v to the root. We write T^N for the subtree of all vertices with distance $\leq N$ to the root.

To each vertex v there is associated a (spin-) variable $\sigma(v)$ taking values in a finite space which, without loss of generality, will be denoted by $\{1, 2, \ldots, q\}$. Our model will be defined in terms of the stochastic matrix with non-zero entries

$$M = (M(v, w))_{1 \leq v, w \leq q} \,. \tag{1}$$

By the Perron-Frobenius theorem there is a unique single-site measure $\alpha = (\alpha(j))_{j=1,\ldots,q}$ which is invariant under the application of the transition matrix M, meaning that $\sum_{i=1}^{q} \alpha(i) M(i, j) = \alpha(j)$.

The object our study is the corresponding *tree-indexed Markov chain in equilibrium*. This is the probability distribution $\mathbb{P}$ on $\{1, \ldots, q\}^T$ whose restrictions $\mathbb{P}_{T^N}$ to the state spaces of finite trees $\{1, \ldots, q\}^{T^N}$ are given by

$$\mathbb{P}_{T^N}(\sigma_{T^N}) = \alpha(\sigma(0)) \prod_{\substack{v, w: \\ v \to w}} M(\sigma(v), \sigma(w)) \,. \tag{2}$$

The notion equilibrium refers to the fact that all single-site marginals are given by the invariant measure α.

A probability measure μ on $\{1, 2, \ldots, q\}^T$ is called a *Gibbs measure* if it has the same finite-volume conditional probabilities as $\mathbb{P}$ has. This means that, for all finite subsets

$V \subset T$, we have for all N sufficiently large

$$\mu(\sigma_V|\sigma_{V^c}) = \mathbb{P}_{T^N}(\sigma_V|\sigma_{\partial V}) \tag{3}$$

μ-almost surely. The Gibbs measures, being defined in terms of a linear equation, form a simplex, and we would like to understand its structure, and exhibit its extremal elements [5]. Multiple Gibbs measures (phase transitions) may occur if the loss of memory in the transition described by M is small enough compared to the proliferation of offspring along the tree T. Uniqueness of the Gibbs measure trivially implies extremality of the measure $\mathbb{P}$, but interestingly the converse is not true. Parametrizing M by a temperature-like parameter may lead to two different transition temperatures, one where $\mathbb{P}$ becomes extremal and one where the Gibbs measure becomes unique. Broadly speaking, statistical mechanics models with two transition temperatures are peculiar to trees (and more generally to models indexed by non-amenable graphs [9]). This is one of the reasons for our interest in models on trees.

Now, our present aim is to provide a general criterion, depending on the model only in a local (finite-dimensional) way, which implies the extremality of $\mathbb{P}$, and which works also in regimes of non-uniqueness. People with statistical mechanics background may think of it as an analogy to Dobrushin's theorem saying that $c(\gamma) < 1$ implies the uniqueness of the Gibbs measure of a local specification γ where $c(\gamma)$ is determined in terms of local (single-site) quantities.

In fact, Martinelli et al. [14] (see Theorem 9.3., see also Theorem 9.3.' and Theorem 9.3") give such a theorem. Their criterion for non-reconstruction of Markov chains on d-ary trees has the form $d\lambda_2\kappa < 1$ where κ is the Dobrushin constant [5] of the system of conditional probabilities described by $\mathbb{P}$. Further λ_2 is the second eigenvalue of M. Our theorem takes a different form. Now, to formulate our result we need the following notation.

We write for the simplex of length-q probability vectors

$$P = \{(p(i))_{i=1,\dots,q}, p(i) \geq 0 \; \forall i, \; \sum_{i=1}^{q} p(i) = 1\} \tag{4}$$

and we denote the relative entropy between probability vectors $p, \alpha \in P$ by $S(p|\alpha) = \sum_{i=1}^{q} p(i) \log \frac{p(i)}{\alpha(i)}$. We introduce the *symmetrized entropy* between p and α and write

$$L(p) = S(p|\alpha) + S(\alpha|p) = (p - \alpha) \log \frac{dp}{d\alpha} \, . \tag{5}$$

While the symmetrized entropy is not a metric (since the triangle inequality fails) it serves us as a "distance" to the invariant measure α.

Let us define the constant, depending solely on the transition matrix M, in terms of the following supremum over probability vectors

$$c(M) = \sup_{p \in P} \frac{L(pM^{\text{rev}})}{L(p)}, \tag{6}$$

where $M^{\text{rev}}(i,j) = \frac{\alpha(j)M(j,i)}{\alpha(i)}$ is the transition matrix of the reversed chain. Note that numerator and denominator vanish when we take for p the invariant distribution α. Consider a Galton-Watson tree with i.i.d. offspring distribution concentrated on $\{1, 2, \dots\}$ and denote the corresponding expected number of offspring by $\mathbb{Q}(d)$.

Here is our main result.

3

Theorem 2.1 *If $\mathbb{Q}(d)c(M) < 1$ then the tree-indexed Markov chain $\mathbb{P}$ on the Galton-Watson tree T is extremal for $\mathbb{Q}$-almost every tree T. Equivalently, in information theoretic language, there is no reconstruction.*

Remark 1. The computation of the constant $c(M)$ for a given transition matrix M is a simple numerical task. Note that fast mixing of the Markov chain corresponds to small $c(M)$. In this sense $c(M)$ is an effective quantity depending on the interaction M that parallels the definition the Dobrushin constant $c_D(\gamma)$ in the theory of Gibbs measures measuring the degree of dependence in a local specification. While the latter depends on the structure of the interaction graph, this is *not* the case for $c(M)$.

Remark 2. Non-uniqueness of the Gibbs measures corresponds to the existence of boundary conditions which will cause the corresponding finite-volume conditional probabilities to converge to different limits. Extremality of the measure $\mathbb{P}$ means that conditioning the measure $\mathbb{P}$ to acquire a configuration ξ at a distance larger than N will cease to have an influence on the state at the root if ξ is chosen according to the measure $\mathbb{P}$ itself and N is tending to infinity. In the language of information theory this is called non-reconstructability (of the state at the origin on the basis of noisy observations far away).

Example (Potts and Ising model). The Potts model with q states at inverse temperature β is defined by the transition matrix

$$
M_\beta = \frac{1}{e^{2\beta} + q - 1}
\begin{pmatrix}
e^{2\beta} & 1 & 1 & \ldots & 1 \\
1 & e^{2\beta} & 1 & \ldots & 1 \\
1 & 1 & e^{2\beta} & \ldots & 1 \\
& & \ldots &
\end{pmatrix}. \tag{7}
$$

This Markov chain is reversible for the equidistribution. In the case $q = 2$, the Ising model, one computes $c(M_\beta) = (\tanh \beta)^2$ which yields the correct reconstruction threshold.

Theorem 2.1 is a generalization of the main result given in our paper [4] for the specific case of the Potts model. That paper also contains comparisons of numerical values to the (presumed) exact transition values. Our discussion of the cases of $q = 3, 4, 5$ shows closeness up to a few percent, and for $q = 3, 4$ and small d these are the best rigorous bounds as of today. To see this connection between the present paper and [4] we rewrite $c(M_\beta) = \frac{e^{2\beta} - 1}{e^{2\beta} + q - 1} \bar{c}(\beta, q)$ and note that the main Theorem of [4] was formulated in terms of the quantity

$$
\bar{c}(\beta, q) = \sup_{p \in P} \frac{\sum_{i=1}^q (qp(i) - 1) \log(1 + (e^{2\beta} - 1)p(i))}{\sum_{i=1}^q (qp(i) - 1) \log qp(i)}. \tag{8}
$$

3 Proof

We denote by T^N the tree rooted at 0 of depth N. The notation T_v^N indicates the sub-tree of T^N rooted at v obtained from "looking to the outside" on the tree T^N. We denote by $\mathbb{P}_v^N$ the measure on T_v^N with free boundary conditions, or, equivalently the Markov chain obtained from broadcasting on the subtree with the root v with the same transition kernel, starting in α. We denote by $\mathbb{P}_v^{N,\xi}$ the correponding measure on

T_v^N with boundary condition on ∂T_v^N given by $\xi = (\xi_i)_{i \in \partial T_v^N}$. Obviously it is obtained by conditioning the free boundary condition measure $\mathbb{P}_v^{N,\xi}$ to take the value ξ on the boundary.

We write

$$\pi_v^N = \pi_v^{N,\xi} = \left(\mathbb{P}^{N,\xi}(\eta(v) = s) \right)_{s=1,\ldots,q}. \tag{9}$$

To control a recursion for these quantities along the tree we find it useful to make explicit the following notion.

Definition 3.1 *We call a real-valued function $\mathcal{L}$ on P a linear stochastic Lyapunov function with center p^* if there is a constant c such that*

- $\mathcal{L}(p) \geq 0 \; \forall p \in P$ *with equality if and only if $p = p^*$;*

- $\mathbb{E}\mathcal{L}(\pi_v^N) \leq c \sum_{w:v \mapsto w} \mathbb{E}\mathcal{L}(\pi_w^N)$.

Proposition 3.2 *Consider a tree-indexed Markov chain $\mathbb{P}$, with transition kernel $M(i,j)$ and invariant measure $\alpha(i)$.*

Then the function

$$L(p) = S(p|\alpha) + S(\alpha|p) = (p - \alpha) \log \frac{dp}{d\alpha} \tag{10}$$

is a linear stochastic Lyapunov function with center α w.r.t. the measure $\mathbb{P}$ for the constant (6).

Proposition 3.2 immediately follows from the following invariance property of the recursion which is the main result of our paper.

Proposition 3.3 *Magic Recursion Formula for expected symmetrized entropy.*

$$\int \mathbb{P}(d\xi) L(\pi_v^{N,\xi}) = \sum_{w:v \to w} \int \mathbb{P}(d\xi) L(\pi_w^{N,\xi} M^{rev}). \tag{11}$$

Warning: Pointwise, that is for fixed boundary condition, things fail and one has

$$L(\pi_v^{N,\xi}) \neq \sum_{w:v \to w} L(\pi_w^{N,\xi} M^{\mathrm{rev}}) \tag{12}$$

in general. In this sense the proposition should be seen as an invariance property which limits the possible behavior of the recursion.

Proof of Proposition 3.3. We need the *measure on boundary configurations* at distance N from the root on the tree emerging from v which is obtained by conditioning the spin in the site v to take the value to be j, namely

$$Q_v^{N,j}(\xi) := \mathbb{P}_v^N(\sigma : \sigma_{|\partial T_v^N} = \xi | \sigma(v) = j). \tag{13}$$

Then the double expected value w.r.t. to the a priori measure α between boundary relative entropies can be written as an expected value w.r.t. $\mathbb{P}$ over boundary conditions w.r.t. to the open b.c. measure of the symmetrized entropy between the distributions at v and α in the following form.

Lemma 3.4

$$\int \mathbb{P}(d\xi) \underbrace{L(\pi_v^{N,\xi})}_{symmetric\ entropy\ at\ v} = \int \alpha(dx_1) \int \alpha(dx_2) \underbrace{S(Q_v^{N,x_2}|Q_v^{N,x_1})}_{boundary\ entropy}. \tag{14}$$

Proof of Lemma 3.4: In the first step we express the relative entropy as an expected value

$$S(Q_v^{N,x_2}|Q_v^{N,x_1}) = \int \mathbb{P}(d\xi) \frac{d\pi_v^N}{d\alpha}(x_2) \Big(\log \frac{d\pi_v^N}{d\alpha}(x_2) - \log \frac{d\pi_v^N}{d\alpha}(x_1) \Big). \tag{15}$$

Here we have used that, with obvious notations,

$$\frac{dQ_v^{N,x_2}}{d\mathbb{P}_v^N}(\xi) = \frac{\mathbb{P}_v(\sigma(v) = x_2, \xi)}{\mathbb{P}_v(\sigma(v) = x_2)\mathbb{P}_v(\xi)} = \frac{d\pi_v^N}{d\alpha}(x_2). \tag{16}$$

Further we have used that

$$\log \frac{dQ_v^{N,x_2}}{dQ_v^{N,x_1}} = \log \frac{d\pi_v^N}{d\alpha}(x_2) - \log \frac{d\pi_v^N}{d\alpha}(x_1), \tag{17}$$

for $x_1, x_2 \in \{1, \ldots, q\}$. This gives

$$\begin{aligned}
&\int \alpha(dx_1) \int \alpha(dx_2) S(Q_v^{N,x_2}|Q_v^{N,x_1}) \\
&= \int \mathbb{P}(d\xi) \int \alpha(dx_2) \frac{d\pi_v^N}{d\alpha}(x_2) \log \frac{d\pi_v^N}{d\alpha}(x_2) \\
&\quad - \int \mathbb{P}(d\xi) \int \alpha(dx_1) \underbrace{\int \alpha(dx_2) \frac{d\pi_v^N}{d\alpha}(x_2)}_{1} \log \frac{d\pi_v^N}{d\alpha}(x_1) \\
&= \int \mathbb{P}(d\xi) S(\pi_v^{N,\xi}|\alpha) + \int \mathbb{P}(d\xi) S(\alpha|\pi_v^{N,\xi})
\end{aligned} \tag{18}$$

and finishes the proof of Lemma 3.4. □

Let us continue with the proof of the Magic Recursion Formula. We need two more ingredients formulated in the next two lemmas. The first gives the recursion of the probability vectors π_v^N in terms of the values π_w^N of their children w, which is valid for any fixed choice of the boundary condition ξ.

Lemma 3.5 *Deterministic recursion.*

$$\pi_v^N(j) = \frac{\alpha(j) \prod_{w:v \to w} \sum_i \frac{M(j,i)}{\alpha(i)} \pi_w^N(i)}{\sum_k \alpha(k) \prod_{w:v \to w} \sum_i \frac{M(k,i)}{\alpha(i)} \pi_w^N(i)}, \tag{19}$$

or, equivalently: for all pairs of values j, k we have

$$\log \frac{d\pi_v^N}{d\alpha}(j) - \log \frac{d\pi_v^N}{d\alpha}(k) = \sum_{w:v \to w} \log \frac{\sum_i \frac{M(j,i)}{\alpha(i)} \pi_w^N(i)}{\sum_i \frac{M(k,i)}{\alpha(i)} \pi_w^N(i)}. \tag{20}$$

The proof of this Lemma follows from an elementary computation with conditional probabilities and will be omitted here.

We also need to take into account the *forward propagation* of the distribution of boundary conditions from the parents to the children, formulated in the next lemma.

Lemma 3.6 *Propagation of the boundary measure.*

$$Q_v^{N,j} = \prod_{w:v \to w} \sum_i M(j,i) Q_w^{N,i}. \tag{21}$$

This statement follows from the definition of the model. Now we are ready to head for the Magic Recursion Formula.

We use the second form of the statement of the deterministic recursion Lemma 3.5 equation (17) to write the boundary entropy in the form

$$S(Q_v^{N,j}|Q_v^{N,k}) = Q_v^{N,j} \sum_{w:v \to w} \log \frac{\sum_i \frac{M(j,i)}{\alpha(i)} \pi_w^N(i)}{\sum_i \frac{M(k,i)}{\alpha(i)} \pi_w^N(i)}. \tag{22}$$

Next, substituting the Propagation-of-the-boundary-measure-Lemma 3.6 and (16) we write

$$
\begin{aligned}
S(Q_v^{N,j}|Q_v^{N,k}) &= Q_v^{N,j} \sum_{w:v \to w} \log \frac{\sum_i \frac{M(j,i)}{\alpha(i)} \pi_w^N(i)}{\sum_i \frac{M(k,i)}{\alpha(i)} \pi_w^N(i)} \\
&= \sum_{w:v \to w} \sum_l M(j,l) Q_w^{N,l} \log \frac{\sum_i \frac{M(j,i)}{\alpha(i)} \pi_w^N(i)}{\sum_i \frac{M(k,i)}{\alpha(i)} \pi_w^N(i)} \\
&= \sum_{w:v \to w} \int d\mathbb{P}(\xi) \sum_l M(j,l) \frac{\pi_w^N(l)}{\alpha(l)} \log \frac{\sum_i \frac{M(j,i)}{\alpha(i)} \pi_w^N(i)}{\underbrace{\sum_i \frac{M(k,i)}{\alpha(i)} \pi_w^N(i)}_{\frac{\pi_w^N M^{\mathrm{rev}}(k)}{\alpha(k)}}} \\
&= \sum_{w:v \to w} \int d\mathbb{P}(\xi) \frac{\pi_w^N M^{\mathrm{rev}}(j)}{\alpha(j)} \log \frac{\frac{\pi_w^N M^{\mathrm{rev}}(j)}{\alpha(j)}}{\frac{\pi_w^N M^{\mathrm{rev}}(k)}{\alpha(k)}},
\end{aligned}
\tag{23}
$$

using in the last step the definition of the reversed Markov chain. Finally applying the sum $\sum_{j,k} \alpha(j)\alpha(k) \cdots$ to both sides of (23) we get the Magic Recursion Formula. To see this, note that the l.h.s. of (23) together with this sum becomes the r.h.s. of the equation in Lemma 3.4. For the r.h.s. of (23) we note that

$$\sum_{j,k} \alpha(j)\alpha(k) \frac{\pi_w^N M^{\mathrm{rev}}(j)}{\alpha(j)} \log \frac{\frac{\pi_w^N M^{\mathrm{rev}}(j)}{\alpha(j)}}{\frac{\pi_w^N M^{\mathrm{rev}}(k)}{\alpha(k)}} = L(\pi_w^N M^{\mathrm{rev}}). \tag{24}$$

This finishes the proof of the Magic Recursion Formula Proposition 3.3. $\qquad\square$

Finally, Theorem 2.1 follows from Proposition 3.2 with the aid of the Wald equality with respect to the expectation over Galton-Watson trees since the contraction of the recursion and the Lyapunov function properties yield

$$\lim_{N \uparrow \infty} \mathbb{P}\left(\xi : \left|\pi^{N,\xi}(s) - \alpha(s)\right| \geq \varepsilon\right) \to 0, \tag{25}$$

for all s, for all $\varepsilon > 0$, and this implies the extremality of the measure $\mathbb{P}$. This ends the proof of Theorem 2.1. $\qquad\square$

References

[1] N. Bhatnagar, E. Maneva, *A computational method for bounding the probability of reconstruction on trees*, preprint 2009.

[2] C. Borgs, J.T. Chayes, E. Mossel, S.Roch, *The Kesten–Stigum reconstruction bound is tight for roughly symmetric binary channels*, 47th Annual IEEE Symposium on Foundations of Computer Science (FOCS), 518-530 (2006).

[3] P. M. Bleher, J. Ruiz, V.A. Zagrebnov, *On the purity of limiting Gibbs state for the Ising model on the Bethe lattice,* J. Stat. Phys. **79**, 473-482 (1995).

[4] M. Formentin, C. Külske, *On the Purity of the free boundary condition Potts measure on random trees*, to be published in Stochastic Processes and their Applications, available at arXiv:0810.0677.

[5] H. O. Georgii, *Gibbs measures and phase transitions,* de Gruyter Studies in Mathematics **9**, Walter de Gruyter & Co., Berlin (1988).

[6] A. Gerschenfeld, A. Montanari, *Reconstruction for models on random graphs*, 48th Annual IEEE Symposium on Foundations of Computer Science (FOCS), 194–204 (2007).

[7] D. Ioffe, *A note on the extremality of the disordered state for the Ising model on the Bethe lattice*, Lett. Math. Phys. **37**, 137-143 (1996).

[8] D. Ioffe, *Extremality of the disordered state for the Ising model on general trees*, Trees (Versailles), 3-14, Progr. Probab. **40**, Birkhäuser, Basel (1996).

[9] R. Lyons, *Phase transitions on nonamenable graphs. In: Probabilistic techniques in equilibrium and nonequilibrium statistical physics,* J. Math. Phys. **41** no. 3, 1099–1126 (2000).

[10] S. Janson, E. Mossel, *Robust reconstruction on trees is determined by the second eigenvalue,* Ann. Probab. **32** no. 3B, 2630–2649 (2004).

[11] E. Mossel, *Reconstruction on trees: Beating the second eigenvalue*, Ann. Appl. Prob. **11** 285-300 (2001).

[12] E. Mossel, Y.Peres, *Information flow on trees*, The Annals of Applied Probability **13** no. 3, 817844 (2003).

[13] R. Pemantle, Y. Peres, *The critical Ising model on trees, concave recursions and nonlinear capacity*, preprint, arXiv:math/0503137v2 [math.PR] (2006).

[14] F. Martinelli, A. Sinclair, D. Weitz, *Fast mixing for independents sets, colorings and other model on trees*, Comm. Math. Phys. **250** no. 2, 301–334 (2004).

[15] M. Mezard, A. Montanari, *Reconstruction on trees and spin glass transition*, J. Stat. Phys. **124** no. 6, 1317–1350 (2006).

[16] A. Sly, *Reconstruction of symmetric Potts models*, preprint, arXiv:0811.1208 [math.PR] (2008).