Reconstruction on trees: Beating the second eigenvalue

Elchanan Mossel
Hebrew University of Jerusalem and INRIA Rocquencourt

September 17, 2002

Abstract

We consider a process in which information is transmitted from a given root node on a noisy d-ary tree network T. We start with a uniform symbol taken from an alphabet $\mathcal{A}$. Each edge of the tree is an independent copy of some channel (Markov chain) M, where M is irreducible and aperiodic on $\mathcal{A}$. The goal is to reconstruct the symbol at the root from the symbols at the nth level of the tree. This model has been studied in information theory, genetics, and statistical physics. The basic question is: Is it possible to reconstruct (some information on) the root? In other words, does the probability of correct reconstruction tend to $1/|\mathcal{A}|$ as $n \to \infty$?

It is known that reconstruction is possible if $d\lambda_2^2(M) > 1$, where $\lambda_2(M)$ is the second eigen-value of M. Moreover, in this case it is possible to reconstruct using a majority algorithm which ignores the location of the data at the boundary of the tree. When M is a symmetric binary channel, this threshold is sharp. In this paper we show, that both for the binary asymmetric channel and for the symmetric channel on many symbols it is sometimes possible to reconstruct even when $d\lambda_2^2(M) < 1$. This result indicates that for many (maybe most) tree indexed Markov chains the location of the data on the boundary plays a crucial role in reconstruction problems.

AMS 1991 subject classifications: Primary 60K35; Secondary 90B15, 68R99.
Keywords: Tree, Markov chain, Percolation, Ising model, Potts model, Coupling.
Abbreviated title: Reconstruction on trees

1 Introduction

1.1 Definitions

We consider the following broadcasting process. The first building block of the process is an irreducible aperiodic Markov chain (or channel) on a finite alphabet $\mathcal{A} = \{1, \ldots, k\}$. We will denote by $\mathbf{M}_{i,j}$ the transition probability from i to j; by M the random function which satisfies $\mathbf{P}[M(i) = j] = \mathbf{M}_{i,j}$; and by $\lambda_2(M)$ the eigen-value of $\mathbf{M}$ which has the second largest absolute value ($\lambda_2(M)$ may be negative). The second building block is a d-ary tree $T = T_d = (V_d, E_d)$ rooted at ρ. At the root ρ one of the symbols of $\mathcal{A}$ is chosen according to an initial distribution $\pi = (\pi_1, \ldots, \pi_k)$. We denote this (random) symbol by σ_ρ. This symbol is then propagated in the tree in the following way. For each vertex v having as a parent v', we let $\sigma_v = M_{v',v}(\sigma_{v'})$, where the $\{M_{v',v}\}$ are independent copies of M. Equivalently, for a vertex v let v' be the parent of v, and let A_v be the set of all vertices which are connected to ρ through paths which do not contain v. Then we have:

$$\mathbf{P}[\sigma_v = j | (\sigma_w)_{w \in A_v}] = \mathbf{P}[\sigma_v = j | \sigma_{v'}] = \mathbf{M}_{\sigma_{v'},j}.$$

This model can be considered as a communication network on T; as a model for propagation of a genetic property; or as a tree-indexed Markov chain - using the terminology of Information theory, Genetics and Statistical Physics respectively. We refer the reader to [4] and the references there for more background.

Let $d(,)$ denote the graph-metric distance on T, and $L_n = \{v \in V : d(\rho, v) = n\}$ be the n'th level of the tree. We denote by $\sigma_{L_n} = (\sigma(v))_{v \in L_n}$ the symbols at the n'th level of the tree. We let $c_{L_n} = (c_{L_n}(1), \ldots, c_{L_n}(k))$ where

$$c_{L_n}(i) = \#\{v \in L_n : \sigma(v) = i\}.$$

That is, c_{L_n} is the count of the n'th level. Note that both $(\sigma_{L_n})_{n=1}^{\infty}$ and $(c_{L_n})_{n=1}^{\infty}$ are Markov chains. We want to know if the data on the boundary gives some information on the root.

Definition 1 *We say that the reconstruction problem is* **solvable** *if there exists $i, j \in \mathcal{A}$ for which*

$$\lim_{n \to \infty} |\, \mathbf{P}_n^i - \mathbf{P}_n^j \,| > 0, \tag{1}$$

where $|\,|$ denotes the total variation norm, and $\mathbf{P}_n^l$ denotes the conditional distribution of σ_{L_n} given that $\sigma_\rho = l$.

Definition 2 *We say that the reconstruction problem is* **count-solvable** *if there exists $i, j \in \mathcal{A}$ for which*

$$\lim_{n \to \infty} |\, \mathbf{P}_n^{(c),i} - \mathbf{P}_n^{(c),j} \,| > 0, \tag{2}$$

where $|\,|$ denotes the total variation norm, and $\mathbf{P}_n^{(c),l}$ denotes the conditional distribution of c_{L_n} given that $\sigma_\rho = l$.

We refer the reader to Section 4 for equivalent definitions of solvability and count-solvability.

1.2 Count reconstruction

Using a theorem of Kesten and Stigum from 1966 [5], and coupling we can give an exact threshold for count-reconstruction. The following theorem was proved jointly with Yuval Peres [8]:

Theorem A *The reconstruction problem is count solvable if $d\lambda_2^2(M) > 1$ and is not count solvable if $d\lambda_2^2(M) < 1$.*

We give a sketch of the proof in Section 5. The details and some generalizations can be found in [8].

1.3 Reconstruction

It turns out that the criteria in Theorem A is tight for solvability for binary symmetric channels:

Theorem B (Bleher, Ruiz and Zagrebnov (1995) [2]) *If M is the binary symmetric channel:*

$$\mathbf{M} = \begin{pmatrix} 1-\delta & \delta \\ \delta & 1-\delta \end{pmatrix},$$

then the reconstruction problem is solvable if and only if $d\lambda_2^2(M) = d(1-2\delta)^2 > 1$.

This result is generalized to general trees in [4].

Theorem A together with Theorem B imply that for binary symmetric channels count-solvability and solvability of the reconstruction problem have the same critical value. It was believed that this kind of phenomena should hold in general. We show that this is not the case. The main results of this paper are the following:

Theorem 1 *Consider the asymmetric binary chains:*

$$\mathbf{M} = \begin{pmatrix} 1-\delta_1 & \delta_1 \\ 1-\delta_2 & \delta_2 \end{pmatrix}, \tag{3}$$

(Note that $\lambda_2(M) = \delta_2 - \delta_1$). Suppose that $0 \le \lambda \le 1$ and that $d\lambda > 1$, then there exists a $\delta > 0$ s.t. if $\lambda_2(M) = \lambda$ and $\delta_1 < \delta$, then the reconstruction problem is solvable for the d-ary tree and the chain (3).

Theorem 2 *Consider the symmetric chains on q symbols:*

$$\mathbf{M} = \begin{pmatrix} 1-(q-1)\delta & \delta & \dots & \delta \\ \delta & 1-(q-1)\delta & \delta & \dots \\ \vdots & \dots & \ddots & \vdots \\ \delta & \dots & \delta & 1-(q-1)\delta \end{pmatrix} \tag{4}$$

(Note that $\lambda_2(M) = 1 - q\delta$). Let $0 < \lambda < 1$, and d s.t. $d\lambda > 1$. Then there exists a Q s.t. if $q > Q$ and $\lambda = 1 - q\delta$, then the reconstruction problem is solvable for the d-ary tree and the chain (4).

The process of broadcasting on a tree with the channels (3) corresponds to the ferromagnetic Ising model with external field on the tree. The broadcasting processes on a tree with the channels (4) corresponds to the Potts model with no external field on the tree.

We remark that theorems A and 1 (2) imply that if $d\lambda_2(M) > 1$ and $d\lambda_2^2(M) < 1$, then for $\delta < \delta_1$ ($q > Q$), the reconstruction problem is solvable, but is not count-solvable.

In [7] we study recursive schemes for reconstruction. We take l to be a fixed number of levels, and consider a reconstruction algorithm for the l-level d-ary tree. Then starting at the boundary, we use this algorithm recursively in order to reconstruct the root of the tree (see [7] for more details).

In [7] it is shown that for the binary symmetric channel, any recursive scheme is inferior to the Majority scheme in the sense that there are binary symmetric channels for which the recursive scheme fails to solve the reconstruction problem while the Majority scheme does solve it. The proofs of Theorems 1 and 2 show that for the channels (3) and (4) this is not generally the case. That is, for these channels, it is sometimes possible to reconstruct using recursive schemes while all majority schemes fail to reconstruct.

Theorems 1 and 2 are sharp in the following sense:

Proposition 3 *Let M be of the form (3) and and d an integer s.t. $|d\lambda_2(M)| \le 1$. Then the reconstruction problem is unsolvable for the d-ary tree and M.*

Proposition 4 *Let $\lambda = 1 - q\delta$. Suppose that $0 \le d\lambda \le 1$. Then, the reconstruction problem is unsolvable for the d-ary tree and the chain (4).*

The proofs of Theorems 1, 2 and Propositions 3 and 4 are given in Section 2.

Is it true for general Markov chains that it is impossible to reconstruct when $|d\lambda_2(M)| \leq 1$? It turns out that a Markov chain constructed in [7] provides a counterexample. In Section 3 we prove:

Proposition 5 *There exists a channel M s.t. $\lambda_2(M) = 0$ and such that the reconstruction problem is solvable for M and all $d \geq 1000$.*

2 Random Cluster methods

The proof Theorems 1 and 2, and of Propositions 3 and 4 all use "random-cluster" arguments. We start with some notations and definitions which we will apply in these proofs. Recall that we denote by $T = T_d = (V_d, E_d)$ the d-ary tree rooted at ρ. We consider the space $\{0,1\}^{E_d}$. We denote an element of this space by $(\tau(e))_{e \in E_d}$. By **λ-percolation** on T we mean the random process which has the state space $\{0,1\}^{E_d}$ and for which $\mathbf{P}[\tau(e) = 1] = \lambda$ independently for all $e \in E_d$. An edge e with $\tau(e) = 1$ is called an **open** edge. More generally we say that a subtree $T' = (V', E')$ of T is **open** if all the edges $e \in E'$ are open. For $v \in V$, the **component of** v which we denote by $\mathcal{C}(v)$, consists of all the vertices in V_d which are connected to v by a path of open edges.

In order to prove Theorem 1 and Theorem 2 we will use the following definition. Let T' be a subtree of the tree T which is rooted at ρ. We say that T' is a **l-diluted b-regular tree** if for all i, all the vertices of T' at level il have exactly b descendents at level $(i+1)l$.

Lemma 6 *Let T_d be the infinite rooted d-ary tree, $0 \leq \lambda \leq 1$ a number such that $d\lambda > 1$. There exists a positive $\epsilon = \epsilon(d, \lambda)$ s.t. for all $b \geq 1$ there exists $l \geq 1$ s.t. if one performs percolation with parameter $\lambda' \geq \lambda$ on T, then*

$$\mathbf{P}[\rho \textit{ is the root of an open } l\textit{-diluted } b\textit{-regular tree}] \geq \epsilon(d, \lambda). \tag{5}$$

In order to prove Lemma 6 we are going to use the following standard fact.

Lemma 7 *Let T_d be the d-ary tree, and $0 \leq \lambda \leq 1$ a number such that $d\lambda > 1$. There exists a number $\epsilon > 0$ such that for all b there exists a number $l = l(b)$ s.t.*

$$\mathbf{P}[|\mathcal{C}(\rho) \cap L_l| \geq b] \geq \epsilon. \tag{6}$$

Proof: Let $Z_l = |\mathcal{C}(\rho) \cap L_l|$. It is clear that $W_l = (d\lambda)^{-l} Z_l$ is a positive martingale. Therefore $W_l \to W$ a.s. Moreover, it is known ([1] p.9) that since $d\lambda > 1$ we also have $\mathbf{P}[W \neq 0] = \lim_l \mathbf{P}[Z_l \neq 0] > 0$. Therefore it follows that there exist positive numbers ϵ_1, ϵ s.t. $\mathbf{P}[Z_l > \epsilon_1 (d\lambda)^l] \geq \epsilon$ for all l. Now the claim follows. □

Proof of Lemma 6: It is clearly enough to prove that (5) holds for λ since the probability on the left of (5) is monotone in λ'. Take ϵ to satisfy (6), take B to be a number such that $\mathbf{P}[Bin(\epsilon/2, B) \geq b] \geq 1/2$, and take l to be a number such that $\mathbf{P}[|\mathcal{C}(\rho) \cap L_l| \geq B] \geq \epsilon$. Let A_r be the event that ρ is the root of rl levels of a l-diluted b-regular open tree. Let $p_r = \mathbf{P}[A_r]$. We have $p_0 = 1$, and

$$p_{r+1} \geq \mathbf{P}[|\mathcal{C}(\rho) \cap L_l| \geq B] \, \mathbf{P}[A_{r+1} | \, |\mathcal{C}(\rho) \cap L_l| \geq B] \geq \epsilon \mathbf{P}[Bin(p_r, B) \geq b].$$

Therefore it follows by induction that for all r, we have $p_r \geq \epsilon/2$ and the lemma follows. $\square$

We also need a complementary result for λ close to 1.

Lemma 8 *Let T_d be the infinite rooted d-ary tree, $l \geq 1$ and $\epsilon > 0$. There exists $\lambda < 1$ such that if one performs percolation with parameter $\lambda' \geq \lambda$ on T, then*

$$\mathbf{P}[\rho \textit{ is the root of an open } l\textit{-diluted } (d^l - 1)\textit{-regular tree}] \geq 1 - \epsilon.$$

Proof: Again by monotonicity it suffices to prove the claim for λ. Let

$$f(p) = \mathbf{P}[Bin(p, d^l) \geq d^l - 1] = p^{d^l} + d^l(1-p)p^{d^l - 1}.$$

Then $f(1) = 1$ and $f'(1) = 0$. Therefore there exists $1 > p^* > 1 - \epsilon$ s.t. $f(p^*) > p^*$. We now take $\lambda < 1$ s.t.

$$\mathbf{P}[|\mathcal{C}(\rho) \cap L_l| = d^l] \geq \frac{p^*}{f(p^*)}.$$

We denote by p_r the probability that ρ is the root of a l-diluted $(d^l - 1)$-regular open tree of rl levels. Then $p_0 = 1 \geq p^*$, and using induction and the monotonicity of f:

$$p_{r+1} \geq \mathbf{P}[|\mathcal{C}(\rho) \cap L_l| = d^l] f(p_r) \geq \frac{p^*}{f(p^*)} f(p_r) \geq p^*.$$

and the lemma follows. $\square$

The last Lemma we need is a simple combinatorial fact:

Lemma 9 *Let r and s be numbers s.t. $r + s > d^l$ and $n \geq 0$. Let T_d be the d-ary tree rooted at ρ. Let $T' = (V', E')$ be a l-diluted r-regular tree rooted at ρ and $T'' = (V'', E'')$ be a l-diluted s-regular tree rooted at ρ. Then there it is impossible that $T' \cup T'' \subset T$ with $V' \cap V'' \cap L_{ln} = \emptyset$.*

Proof: By induction on n. When $n = 0$ there is nothing to prove. Suppose $n > 0$ and $V' \cap V'' \cap L_{ln} = \emptyset$. We will show that $V' \cap V'' \cap L_{l(n-1)} = \emptyset$, so the proof would follow by induction. Let $v \in L_{l(n-1)}$ and look at the d^l descendents of v at L_{ln}. If we have $v \in V' \cap V''$ then v has r descendents in $L_{ln} \cap V'$ and s descendents in $L_{ln} \cap V''$, which is a contradiction to the assumption $V' \cap V'' \cap L_{ln} = \emptyset$. $\square$

Proof of Theorem 1: We will denote the states of the chain by 0 and 1. Using the notation (3) we note that if v is a parent of w in the tree T_d, then $\mathbf{P}[\sigma(w) = 1 | \sigma(v) = 1] = \delta_2$. We may therefore perform percolation with parameter δ_2 in such a way that

$$\mathbf{P}[\mathcal{C}(\rho) \subset \{w : \sigma(w) = 1\} | \sigma(\rho) = 1] = 1. \tag{7}$$

Similarly we may perform percolation with parameter $1 - \delta_1$ in such a way that

$$\mathbf{P}[\mathcal{C}(\rho) \subset \{w : \sigma(w) = 0\} | \sigma(\rho) = 0] = 1. \tag{8}$$

Let ϵ and l be chosen to satisfy (5) for $b = 2$ and λ. Let A_{rl} be the event that there exists a l-diluted binary tree T' rooted at ρ s.t. all of the vertices of T' at the rl level are labeled by 1. Lemma 6 together with (7) imply that for $\delta_2 \geq \lambda$,

$$\mathbf{P}^1_{rl}[A_{rl}] \geq \epsilon \tag{9}$$

for all r. On the other hand it follows from Lemma 8, (8) and Lemma 9 that there exists $\delta > 0$ s.t. if $0 < \delta_1 < \delta$, we have:

$$\mathbf{P}^0_{rl}[A_{rl}] \le \epsilon/2 \tag{10}$$

for all r. We now choose δ_1 s.t. (10) holds and $\delta_2 = \lambda + \delta_1$ so that (9) holds. We get that for all r

$$|\mathbf{P}^0_{rl} - \mathbf{P}^1_{rl}| \ge \epsilon/2.$$

It follows that the reconstruction problem is solvable. □

Proof of Theorem 2: The proof is similar to the proof of Theorem 1. We fix $i \ne j \in \mathcal{A}$. Using the notation (4) we note that if v is a parent of w in the tree T_d, then $\mathbf{P}[\sigma(w) = i|\sigma(v) = i] = 1 - (q-1)\delta$, and for all $i' \ne i$, $\mathbf{P}[\sigma(w) \ne i|\sigma(v) = i'] = 1 - \delta$. We may therefore perform percolation with parameter $1 - (q-1)\delta$ in such a way that

$$\mathbf{P}[\mathcal{C}(\rho) \subset \{w : \sigma(w) = i\}|\sigma(\rho) = i] = 1. \tag{11}$$

Similarly we may perform percolation with parameter $1 - \delta$ in such a way that

$$\mathbf{P}[\mathcal{C}(\rho) \subset \{w : \sigma(w) \ne i\}|\sigma(\rho) = j] = 1. \tag{12}$$

Let ϵ and l be chosen to satisfy (5) for $b = 2$ and λ. Let A_{rl} be the event that there exists a l-diluted binary tree T' rooted at ρ s.t. all of the vertices of T' at the rl level are labeled by i. Lemma 6 together with (11) imply that when $1 - (q-1)\delta \ge \lambda$,

$$\mathbf{P}^i_{rl}[A_{rl}] \ge \epsilon \tag{13}$$

for all r. On the other hand it follows from Lemma 8, (8) and Lemma 9 that when $\delta > 0$ is small enough we get:

$$\mathbf{P}^j_{rl}[A_{rl}] \le \epsilon/2 \tag{14}$$

for all r. We now choose Q large enough s.t. when $\lambda = 1 - q\delta$ (14) holds for $q \ge Q$. We automatically have $1 - (q-1)\delta \ge \lambda$ so that (13) holds. We get that for all r

$$|\mathbf{P}^i_{rl} - \mathbf{P}^j_{rl}| \ge \epsilon/2.$$

It follows that the reconstruction problem is solvable. □

The proof of Propositions 3 and 4 uses another type of random-cluster argument. The channels for which we can use this kind of arguments are channels M which have matrices $(\mathbf{M}_{i,j})_{i,j=1}^k$ which satisfy:

$$\mathbf{M}_{i,j} = \lambda \mathbf{N}_{i,j} + (1-\lambda)\nu_j. \tag{15}$$

for some channel N which has the matrix $(\mathbf{N}_{i,j})_{i,j=1}^k$, a distribution vector $(\nu_j)_{j=1}^k$ and a number $0 \le \lambda \le 1$. The proof of Propositions 4 and 3 follows immediately from the following propositions.

Proposition 10 *Suppose that* $\mathbf{M}$ *has the form (15). Then the reconstruction problem for* M *is unsolvable whenever* $d\lambda \le 1$.

Proposition 11 *All binary channels (3) have the form (15) with* $\lambda = \lambda_2(M)$. *All symmetric channels (4) with* $\lambda = 1 - q\delta \ge 0$ *have the form (15) with* $\lambda = \lambda_2(M) = 1 - q\delta$.

Proof of Proposition 10: Since the matrix $\mathbf{M}$ satisfies (15) we may write the random function M as $M = XN + (1 - X)Y$ where N is a random function which satisfies $\mathbf{P}[N(i) = j] = \mathbf{N}_{i,j}$, Y is a r.v. which satisfies $\mathbf{P}[Y = j] = \nu_j$, X is a $\{0,1\}$ variable which satisfies $\mathbf{P}[X = 1] = \lambda$ and all these variables are independent. Thus the broadcasting on the tree T can be implemented in the following way.

- For each vertex v, let N_v be an independent copy of the function N, and Y_v an independent copy of Y (these variables are all independent).
- Perform percolation with parameter λ on T. The percolation process is independent of the variables Y_v and N_v.
- Fix $\sigma(\rho) = i$.
- Denote $\Gamma = (\tau(e))_{e \in E_d}$.
- In order to produce $(\sigma(v))_{v \in V_d}$ we use the following procedure: Assume that we have produced $\sigma(v)$ and that w is a child of v. If the edge $\{v, w\}$ is open, set $\sigma(w) = N_w(\sigma(v))$; otherwise set $\sigma(w) = Y_w$.

Note that we may use this process simultaneously with $\sigma(\rho) = i$ for all $i \in \mathcal{A}$ with the **same** random variables Γ, Y_v and N_v. In this way we obtain a coupling of the distributions $\{\mathbf{P}_n^i\}_{i=1}^k$. The key observation here is that if the root component of Γ does not intersect L_n, then we would obtain the same labeling of L_n for all root values $i \in \mathcal{A}$. Thus we obtain:

$$max_{i,j \in \mathcal{A}} |\mathbf{P}_n^i - \mathbf{P}_n^j| \le \mathbf{P}[\mathcal{C}(\rho) \cap L_n \neq \emptyset].$$

However, classical results on Branching processes (see e.g. [1]) imply that when $d\lambda \le 1$, we have $\lim_{n\to\infty} \mathbf{P}[\mathcal{C}(v) \cap L_n \neq \emptyset] = 0$. We have thus proved that the reconstruction problem is unsolvable. $\square$

Proof of Proposition 11: For the symmetric channel on q symbols and $\lambda \ge 0$ we take $\mathbf{N} = \mathbf{I}$ the identity matrix, ν the uniform distribution on $\mathcal{A}$, and $\lambda = \lambda_2(M) = 1 - q\delta$. For the general binary channel:

$$\mathbf{M} = \begin{pmatrix} m_{0\to 0} & m_{0\to 1} \\ m_{1\to 0} & m_{1\to 1} \end{pmatrix}, \tag{16}$$

we note that $m_{0\to 0} + m_{0\to 1} = 1 = m_{1\to 0} + m_{1\to 1}$, and therefore there exists a number λ such that $\lambda = m_{0\to 0} - m_{1\to 0} = m_{1\to 1} - m_{0\to 1}$. We may now write:

$$\mathbf{M} = \lambda \mathbf{I} + \begin{pmatrix} m_{1\to 0} & m_{0\to 1} \\ m_{1\to 0} & m_{0\to 1} \end{pmatrix} = -\lambda \mathbf{J} + \begin{pmatrix} m_{0\to 0} & m_{1\to 1} \\ m_{0\to 0} & m_{1\to 1} \end{pmatrix},$$

where $\mathbf{J} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$.

Thus $\lambda = \lambda_2(M)$. Moreover, if $\lambda \ge 0$ we may take $\mathbf{N} = \mathbf{I}$ and $\nu = \frac{(m_{1\to 0}, m_{0\to 1})}{1-\lambda}$, and if $\lambda \le 0$ we may take $\mathbf{N} = \mathbf{J}$ and $\nu = \frac{(m_{0\to 0}, m_{1\to 1})}{1+\lambda}$. $\square$

Another demonstration of the importance of random-cluster representations is given in the following proposition.

Proposition 12 *Suppose that $M^{(1)}$ and $M^{(2)}$ are channels s.t.*

$$\mathbf{M}^{(1)}_{i,j} = \lambda_1 \mathbf{N}_{i,j} + (1-\lambda_1)\nu_j.$$

$$\mathbf{M}^{(2)}_{i,j} = \lambda_2 \mathbf{N}_{i,j} + (1-\lambda_2)\nu_j.$$

with the same $\mathbf{N}$ and ν and $\lambda_1 \le \lambda_2$. Then, if the reconstruction problem is solvable for $M^{(1)}$ and the d-ary tree it is also solvable for $M^{(2)}$ and the d-ary tree.

Proof: Let λ_3 be a number such that $\lambda_1 = \lambda_2\lambda_3$. We then have

$$\mathbf{M}^{(1)}_{i,j} = \lambda_3 \mathbf{M}^{(2)}_{i,j} + (1-\lambda_3)\nu_j.$$

For simplicity we denote the random symbols at the n'th level for the d-ary tree and the chain $M^{(1)}$ given that $\sigma(\rho) = i$ by σ^i_n; for v a vertex of the tree we denote by σ^i_v the symbol at v given that the root is i. Similarly, we denote random symbols at the n'th level for the d-ary tree and the chain $M^{(2)}$ given that $\sigma(\rho) = i$ by τ^i_n, and the symbol at v by τ^i_v.

We assume that the reconstruction problem is unsolvable for $M^{(2)}$ and we will show it is unsolvable for $M^{(1)}$. In other words, we will show that if for all $i, j \in \mathcal{A}$ there exist couplings of τ^i_n and τ^j_n s.t. $\lim_{n\to\infty} \mathbf{P}[\tau^i_n = \tau^j_n] = 1$, then we also have for all $i, j \in \mathcal{A}$ couplings of σ^i_n and σ^j_n such that $\lim_{n\to\infty} \mathbf{P}[\sigma^i_n = \sigma^j_n] = 1$.

We are going to use the same procedure as in Proposition 10 in order to produce σ^i_n using the $M^{(2)}$ Markov chain and λ_3-percolation. Therefore for $v \in L_n$ we can write:

$$\sigma^i_v = \begin{cases} \tau^i_v & \text{if } v \in \mathcal{C}(\rho) \cap L_n, \\ \sigma_v & \text{otherwise }, \end{cases} \tag{17}$$

where σ_v is independent of i. Since we know that for all $i, j \in \mathcal{A}$ there exist couplings of τ^i_n and τ^j_n s.t. $\lim_{n\to\infty} \mathbf{P}[\tau^i_n = \tau^j_n] = 1$, it follows from (17) that we may also couple for all i and j, σ^i_n and σ^j_n so that $\lim_{n\to\infty} \mathbf{P}[\sigma^i_n = \sigma^j_n] = 1$. $\square$

Remark 13 *It is possible to generalize the proofs of propositions 3, 4 and 12 to general trees using the branching number (see [6]) instead of d. We do not believe that such generalizations hold for Theorems 1 and 2.*

3 Proof of proposition 5

Proof: This example is taken from [7]. Take M to have the state space $\mathbf{Z}_2 \times \mathbf{Z}_2$. M can be represented as a random function in the following way:

$$M(x, y) = \begin{cases} (r, x + r \bmod 2) & \text{with probability } 1/2 \\ (r, y + r \bmod 2) & \text{with probability } 1/2 \end{cases}$$

where r takes each of the values 0 and 1 with probability 1/2, and is independent of anything else. Thus M has the following matrix:

$$\mathbf{M} = \begin{pmatrix} 0.50 & 0.00 & 0.00 & 0.50 \\ 0.25 & 0.25 & 0.25 & 0.25 \\ 0.25 & 0.25 & 0.25 & 0.25 \\ 0.00 & 0.50 & 0.50 & 0.00 \end{pmatrix}.$$

It is clear that for all $(x, y), (z, w) \in \mathbf{Z}_2 \times \mathbf{Z}_2$ we have:

$$P[M(M(x, y)) = (z, w)] = 1/4.$$

Therefore $\lambda_2(M) = 0$. In order to show that the reconstruction problem is solvable for M and the 1000-ary tree, we exhibit a recursive algorithm for reconstructing $x \oplus y$ at the root with probability greater than 0.999, where $\oplus$ is the modulo 2 addition (We assume that the initial distribution of the root is uniform; See Proposition 14 to see why this implies that the reconstruction problem is solvable). Let v be a vertex and $w_1, \ldots, w_{1000}$ be it's children. Denote the corresponding labels by (x_v, y_v) etc. When $x_v \oplus y_v = 0$ we have

$$x_{w_1} \oplus y_{w_1} = x_{w_2} \oplus y_{w_2} = \ldots = x_{w_{1000}} \oplus y_{w_{1000}}.$$

If on the other hand, $x_v \oplus y_v = 1$, then $x_{w_i} \oplus y_{w_i}$ are i.i.d variables taking the values $0, 1$ with probability $1/2$ each. This lead us to use the following recursive algorithm: In order to reconstruct $x_v \oplus y_v$, look at the 1000 reconstructed values for $x_{w_1} \oplus y_{w_1}, \ldots, x_{w_{1000}} \oplus y_{w_{1000}}$. If there are at least 700 of them which take the same value, then we reconstruct $x_v \oplus y_v = 0$, otherwise we reconstruct $x_v \oplus y_v = 1$. This reconstruction method leads to the following recursions: Let p_n be the probability that we reconstructed correctly $x_v \oplus y_v$ for the root of the 1000-ary tree of n levels. Then $p_0 = 1$, and

$$p_{n+1} \geq \mathbf{P}[400 \leq Bin(0.5, 1000) \leq 600]\, \mathbf{P}[Bin(p_n, 1000) \geq 901].$$

Therefore we get by induction that $p_n \geq 0.999$ for all n. $\square$.

4 Equivalent definitions for reconstruction solvability

As we noted before both σ_{L_n} and c_{L_n} are Markov chains. Let $(X_n)_{n=0}^{\infty}$ be a Markov chain s.t. X_n has as state space the finite space $\mathcal{A}_n$. We denote $\mathcal{A} = \mathcal{A}_0$ and let $\pi = (\pi_i)_{i \in A}$ be an initial distribution of X_0. We note that given X_n if one uses the optimal reconstruction strategy(maximum likelihood), the probability of reconstructing X_0 given X_n is at least $\max_i \pi_i$.

We denote by $\mathbf{P}^\sigma$ the distribution $\mathbf{P}[X_0 = i | X_n = \sigma]$, by $\mathbf{P}_n$ the distribution $\sum_i \pi_i \mathbf{P}_n^i$, and the reconstruction probability given X_n by $\Delta_n(\pi)$. Since given that $X_n = \sigma$, the optimal algorithm will reconstruct a symbol j s.t. $\mathbf{P}^\sigma[j] = \max_i \mathbf{P}^\sigma[i]$, we have:

$$\Delta_n(\pi) = \sum_\sigma \mathbf{P}_n[\sigma] \max_i \mathbf{P}^\sigma[i]. \tag{18}$$

Let H be the entropy function and let $I(X, Y) = H(X) + H(Y) - H(X, Y)$ be the mutual-information operator (see e.g. [3] for definitions and basic properties). We have the following equivalence:

Proposition 14 *The following conditions are equivalent (where π denotes the initial distribution of X_0):*

1. *There exists a π for which*

$$\lim_{n \to \infty} I(X_0, X_n) > 0. \tag{19}$$

2. If π is the uniform distribution on $\mathcal{A}$ then,

$$\lim_{n\to\infty} I(X_0, X_n) > 0. \tag{20}$$

3. For any distribution π with $\min_i \pi_i > 0$ we have:

$$\lim_{n\to\infty} I(X_0, X_n) > 0. \tag{21}$$

4. There exist $i, j \in \mathcal{A}$ for which

$$\lim_{n\to\infty} |\, \mathbf{P}_n^i - \mathbf{P}_n^j \,| > 0, \tag{22}$$

where $|\,|$ denotes the total variation norm, and $\mathbf{P}_n^l$ denotes the conditional distribution of X_n given that $X_0 = l$.

5. There exist a π for which

$$\liminf_{n\to\infty} \Delta_n(\pi) > \max_i \pi_i. \tag{23}$$

6. If π is the uniform distribution on $\mathcal{A}$, then,

$$\liminf_{n\to\infty} \Delta_n(\pi) > 1/|\mathcal{A}|. \tag{24}$$

Proposition 15 *Suppose that $X_n = \sigma_{L_n}$ or $X_n = c_{L_n}$, and π is some distribution which satisfies $\min_i \pi_i > 0$. Then the conditions in Proposition 14 are all equivalent to the fact that the sequence $\{X_i\}_{i=0}^{\infty}$ has a none trivial tail.*

Proofs We refer the reader to [3] for some standard facts in information theory which we will use in the sequence. By the Data Processing Lemma ([3] page 32) it follows that $I(\sigma_\rho, \sigma_{L_n})$ is a decreasing sequence, so the limits (19), (20) and (21) exist. Similarly, using the coupling between $\mathbf{P}_n^i$ and $\mathbf{P}_n^j$, we see that the sequence in (22) is decreasing, so that the limit in (22) exists.

(22) $\equiv$ (23) $\equiv$ (24)

By (18) we have:

$$\begin{aligned}
\Delta_n(\pi) - \max_i \pi_i &= \sum_\sigma \mathbf{P}_n[\sigma](\max_i \mathbf{P}^\sigma[i] - \max_i \pi_i) \le \sum_\sigma \mathbf{P}_n[\sigma](\sum_i |\mathbf{P}^\sigma[i] - \pi_i|) \\
&= \sum_\sigma \sum_i \pi_i |\mathbf{P}_n^i[\sigma] - \mathbf{P}_n[\sigma]| = \sum_i \pi_i \,|\, \mathbf{P}_n^i - \mathbf{P}_n \,| \\
&\le \sum_i \pi_i \max_{j,j'} |\, \mathbf{P}_n^j - \mathbf{P}_n^{j'} \,| = \max_{i,j} |\, \mathbf{P}_n^i - \mathbf{P}_n^j \,|,
\end{aligned} \tag{25}$$

where the inequality in (25) follows from the fact that $\mathbf{P}_n$ is an average of the $\mathbf{P}_n^i$. Moreover, if π is the uniform distribution, then we have:

$$\begin{aligned}
\Delta_n(\pi) - |\mathcal{A}|^{-1} &= \sum_\sigma \mathbf{P}_n[\sigma](\max_i \mathbf{P}^\sigma[i] - |\mathcal{A}|^{-1}) = |\mathcal{A}|^{-1} \sum_\sigma \max_i(\mathbf{P}_n^i[\sigma] - \mathbf{P}_n[\sigma]) \\
&\ge |\mathcal{A}|^{-2} \sum_\sigma \max_{i,j} |\mathbf{P}_n^i[\sigma] - \mathbf{P}_n^j[\sigma]| \ge |\mathcal{A}|^{-2} \max_{i,j} |\, \mathbf{P}_n^i - \mathbf{P}_n^j \,|,
\end{aligned} \tag{26}$$

where the first inequality follows from the fact that for a sequence $(a_i)_{i=1}^k$ we have:

$$\max_i a_i - \frac{1}{k}\sum_i a_i \geq \frac{1}{k}\max_{i,j}|a_i - a_j|.$$

By (25) we have that (23) implies (22), and by (26) we have that (22) implies (24) (which trivially implies (23)).

(24) $\equiv$ (20)

We are going to use the following known inequalities where $p = (p_1, \ldots, p_k)$ and $q = (q_1, \ldots, q_k)$ are probability distributions, $D(p||q)$ is the relative-entropy (Kullback Leibler distance), $G(x) = (2\ln 2)^{-1}x^2$,$F(x) = -x\log(x/k)$ for $0 \leq x \leq 1/2$ and $F(x) = \log k$ otherwise ([3] pages 488–489):

$$D(p||q) \geq G(|p-q|), \tag{27}$$

and

$$|H(p) - H(q)| \leq F(|p-q|). \tag{28}$$

Let π be any initial distribution, then we have on one hand that:

$$I(X_0, X_n) = H(\pi) - \sum_\sigma \mathbf{P}_n[\sigma]H(\mathbf{P}^\sigma) \leq \sum_\sigma \mathbf{P}_n[\sigma]|H(\pi) - H(\mathbf{P}^\sigma)| \leq \sum_\sigma \mathbf{P}_n[\sigma]F(|\pi - \mathbf{P}^\sigma|), \tag{29}$$

and on the other hand:

$$I(X_0, X_n) = \sum_\sigma \mathbf{P}_n[\sigma]D(\mathbf{P}^\sigma||\pi) \geq \sum_\sigma \mathbf{P}_n[\sigma]G(|\pi - \mathbf{P}^\sigma|). \tag{30}$$

For π the uniform distribution we also have:

$$\Delta_n(\pi) - |A|^{-1} = \sum_\sigma \mathbf{P}_n[\sigma]\max_i(\mathbf{P}^\sigma[i] - |\mathcal{A}|^{-1}) \geq |A|^{-1}\sum_\sigma \mathbf{P}_n[\sigma]|\mathbf{P}^\sigma - \pi|. \tag{31}$$

and

$$\Delta_n(\pi) - |A|^{-1} = \sum_\sigma \mathbf{P}_n[\sigma]\max_i(\mathbf{P}^\sigma[i] - |\mathcal{A}|^{-1}) \leq \sum_\sigma \mathbf{P}_n[\sigma]|\mathbf{P}^\sigma - \pi|. \tag{32}$$

By (29), (30), (31) and (32) it follows that we have $I(X_0, X_n) \to 0$ **iff** $\Delta_n(\pi) - |\mathcal{A}|^{-1} \to 0$.

(19) $\equiv$ (20) $\equiv$ (21)

We note that if we write $p(x) = \mathbf{P}[X = x]$ and $p(y|x) = \mathbf{P}[Y = y|X = x]$ then for fixed $p(y|x)$ the function $I(X,Y)$ is a concave function of $p(x)$ ([3] page 31). Suppose that $\lim_{n\to\infty} I(X_0, X_n) > 0$ where X_0 has density π. If X_0' has the uniform distribution π', we can write

$$\pi' = \alpha\pi + (1-\alpha)\left((1-\alpha)^{-1}(\pi' - \alpha\pi)\right),$$

as a convex sum of distribution vectors where $\alpha = (|\mathcal{A}|\max_i p_i)^{-1}$. Now we obtain that

$$\lim I(X_0', X_n) \geq \alpha\lim(X_0, X_n) \geq |\mathcal{A}|^{-1}\lim(X_0, X_n) > 0. \tag{33}$$

In a similar manner, if X_0' is uniform variable on $\mathcal{A}$, and X_0 has distribution π then we obtain:

$$\lim I(X_0, X_n) \geq |\mathcal{A}|(\min_i \pi_i)\lim I(X_0', X_n). \tag{34}$$

By (33) we have that (19) implies (20), and by (34) we have that (20) implies (21) (which trivially implies (19)).

Proof of Proposition 15: First, if the conditions of Proposition 14 hold, then we have that $\lim I(X_0, X_n) > 0$. In particular, the variable X_0 is not independent of the the tail sigma field.

We now prove the other direction (see [4] for a similar argument). For a set U, let $\sigma_U = (\sigma_v)_{v \in U}$. Fix a level n. For each $m \geq n$ and $w \in L_n$ let $L(w,m)$ be the set of vertices in T which connect to ρ through w. Since the variables $\sigma_{L(w,m)}$ are conditionally independent given σ_{L_n} we have:

$$I(\sigma_{L_n}, \sigma_{L_m}) \leq \sum_{w \in L_n} I(\sigma_{L_n}, \sigma_{L(w,m)}) = \sum_{w \in L_n} I(\sigma_w, \sigma_{L(w,m)}),$$

and the right hand side goes to 0 as $n \to \infty$. It follows that if for the sequence σ_n the conditions in Proposition 14 do not hold the σ-tail is trivial.

We use a similar proof for the sequence c_{L_n}. We let $c_{L(w,m)}$ be the count of the vertices of L_m which connect to ρ through w and we get:

$$\begin{aligned} I(c_{L_n}, c_{L_m}) \leq I(\sigma_{L_n}, c_{L_m}) &\leq I\left(\sigma_{L_n}, (c_{L(w,m)})_{w \in L_n}\right) \\ &\leq \sum_{w \in L_n} I(\sigma_{L_n}, c_{L(w,m)}) = \sum_{w \in L_n} I(\sigma_w, c_{L(w,m)}). \end{aligned}$$

as before. (For the sequence σ_{L_n} it is easy to see that the claim remains true even without the assumption that $\min_i \pi_i > 0$. Indeed, if the assumptions of Proposition 14 hold then we choose a vertex v with $\min_i \mathbf{P}[\sigma_v = i] > 0$. Then the variable σ_v is not independent of the tail σ-field.) $\square$

5 Count reconstruction

Sketch of the Proof of Theorem A [8]: Let $\lambda = \lambda_2(M)$. Assume first that $d\lambda^2 > 1$. In this case let μ' be a left eigenvector of the matrix $\mathbf{M}$ which corresponds to λ (that is, $\mu'\mathbf{M} = \lambda\mu'$). By the Kesten Stigum theorem [5] (or rather by the proof) it follows that there exist $i, j \in \mathcal{A}$ s.t.

$$\lim_{n \to \infty} |\mathbf{P}^i_n[(\mu', c_{L_n}) > 0] - \mathbf{P}^j_n[(\mu', c_{L_n}) > 0]| > 0.$$

where $(,)$ denotes the usual scalar-product. This could be verified more directly using a second moment argument.

We turn to the case $d\lambda^2 < 1$. We denote by $c^j_{L_n}$ the r.v. c_{L_n} conditioned on $\sigma(\rho) = j$. We are going to use the following observation: Given $c^j_{L_n}$ we may write

$$c^j_{L_{n+1}} = \sum_{i=1}^{k} S^i(d c^j_{L_n}(i)), \tag{35}$$

where S^i is a random walk on $\mathbf{Z}^k$ satisfying $S^i(0) = 0$ and

$$\mathbf{P}[S^i(t+1) = S^i(t) + e_s] = \mathbf{M}_{i,s}$$

for $1 \leq s \leq k$. Let $a > 0$ be a constant. Suppose that

$$\min_i c^j_{L_n}(i) > ad^n, \; \min_i c^{j'}_{L_n}(i) > ad^n \tag{36}$$

and

$$|c^j_{L_n} - c^{j'}_{L_n}|_1 < \epsilon d^{n/2}. \tag{37}$$

The local central limit theorem ensures that if (35), (36) and (37) hold, then we may couple $c^j_{L_{n+1}}$ and $c^{j'}_{L_{n+1}}$ in such a way that:

$$\mathbf{P}[c^j_{L_{n+1}} \neq c^{j'}_{L_{n+1}}] \leq f(\epsilon), \tag{38}$$

where $\lim_{\epsilon \to 0} f(\epsilon) = 0$.

When $d\lambda^2 < 1$, the Kesten Stigum theorem implies that if ν is the stationary distribution of M, then for all $j \in \mathcal{A}$,

$$\frac{c^j_{L_n} - d^n \nu}{d^{n/2}} \tag{39}$$

converges to a k-dimensional normal variable $\mathcal{N}$ which does not depend on j. From (39) if follows that for every $\epsilon > 0$ and $i, j \in \mathcal{A}$, we may couple $c^i_{L_n}$ and $c^j_{L_n}$ in such a way that:

$$\lim_{n \to \infty} \mathbf{P}[|c^i_{L_n} - c^j_{L_n}|_1 < \epsilon d^{n/2}] = 1. \tag{40}$$

If we combine the fact (39) with (40) and (38) we obtain that for every $j, j' \in \mathcal{A}$ there exist couplings s.t.

$$\lim_{n \to \infty} \mathbf{P}[c^{j'}_{L_n} = c^j_{L_n}] = 1,$$

as needed □

Acknowledgment: Theorem A is taken from a joint paper with Yuval Peres [8]. I thank Yuval Peres for encouraging me to think about these problems. I thank the referee for many helpful comments.

References

[1] Atherya, K. B. and Ney, P. E. (1972) *Branching Processes*, Springer-Verlag.

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

[3] Cover, T. M. and Thomas, J. A. (1991) *Elements of Information Theory*, John Wiley and Sons.

[4] Evans, W., Kenyon, C., Peres, Y. and Schulman L. J. (1999) Broadcasting on trees and the Ising Model, *Ann. Appl. Prob.*, to appear.

[5] Kesten, H. and Stigum, B. P. (1966) Additional limit theorem for indecomposable multidimensional Galton-Watson processes, *Ann. Math. Statist.* **37**, 1463–1481.

[6] Lyons, R. (1990) Random walks and percolation on trees. *Ann. Probab.* **18**, 931–958.

[7] Mossel, E. (1998) Recursive reconstruction on periodic trees, *Random Structures and algorithms* **13,1** 81–97

[8] Mossel, E. and Peres, Y. (2000) Coupling for tree-indexed Markov chains, in preparation.

Elchanan Mossel
Department of Mathematics, Hebrew University of Jerusalem, 91904 ISRAEL.
mossel@math.huji.ac.il