Robust reconstruction on trees is determined by the second eigenvalue

Svante Janson
Uppsala University

Elchanan Mossel
U.C. Berkeley

October 14, 2002; revised December 3, 2003

Abstract

Consider a Markov chain on an infinite tree $T = (V, E)$ rooted at ρ. In such a chain, once the initial root state $\sigma(\rho)$ is chosen, each vertex iteratively chooses its state from the one of its parent by an application of a Markov transition rule (and all such applications are independent). Let μ_j denote the resulting measure for $\sigma(\rho) = j$. μ_j is defined on configurations $\sigma = (\sigma(x))_{x \in V} \in \mathcal{A}^V$, where $\mathcal{A}$ is some finite set. Let μ_j^n denote the restriction of μ to the sigma-algebra generated by the variables $\sigma(x)$ where x is at distance exactly n from ρ. Letting $\alpha_n = \max_{i,j \in \mathcal{A}} d_{TV}(\mu_i^n, \mu_j^n)$, where d_{TV} denotes total variation distance, we say that the *reconstruction problem is solvable* if $\liminf_{n \to \infty} \alpha_n > 0$. Reconstruction solvability roughly means that the n'th level of the tree contains a non-vanishing amount of information on the root of the tree, as $n \to \infty$.

In this paper we study the problem of *robust reconstruction*. Let ν be a non degenerate distribution on $\mathcal{A}$ and $\epsilon > 0$. Let σ be chosen according to μ_j^n and σ' be obtained from σ by letting for each node independently $\sigma(v) = \sigma'(v)$ with probability $1 - \epsilon$ and $\sigma'(v)$ to be an independent sample from ν otherwise. We denote by $\mu_j^n[\nu, \epsilon]$ the resulting measure on σ'. The measure $\mu_j^n[\nu, \epsilon]$ is a perturbation of the measure μ_j^n. Letting $\alpha_n(\nu, \epsilon) = \max_{i,j \in \mathcal{A}} d_{TV}(\mu_i^n[\nu, \epsilon], \mu_j^n[\nu, \epsilon])$, we say that the reconstruction problem is *ν-robust-solvable* if $\liminf_{n \to \infty} \alpha_n(\nu, \epsilon) > 0$ for all $0 < \epsilon < 1$. Roughly speaking, the reconstruction problem is robust-solvable if for any noise-rate and for all n, the n'th level of the tree contains a non-vanishing amount of information on the root of the tree.

Standard techniques imply that if T is the rooted B-ary tree (where each node has B children) and if $B|\lambda_2(M)|^2 > 1$, where $\lambda_2(M)$ is the second largest eigenvalue of M (in absolute value), then for all non-degenerate ν, the reconstruction problem is ν-robust-solvable. We prove a converse and show that the reconstruction problem is not ν-robust-solvable if $B|\lambda_2(M)|^2 < 1$. This proves a conjecture by the second author and Peres. We also consider other models of noise and general trees.

1 Introduction

In this paper we study the perturbative theory of reconstruction on trees, and show how it depends on the spectrum of the underlying Markov chain. In particular, we show that the threshold for "robust reconstruction" for the B-ary tree is $B|\lambda_2(M)|^2 = 1$, where $\lambda_2(M)$ denotes the eigenvalue of M which is the second largest in absolute value. In Section 3 we prove a similar threshold for general bounded degree trees, where B is replaced by the branching number of the tree $\mathrm{br}(T)$. We refer the reader to Subsection 1.2 and to [7, 3, 22, 21] for background.

1.1 Definitions and main results

We proceed with some formal definitions. Let $T = (V, E, \rho)$ be a tree T with nodes V, edges E and root $\rho \in V$. We direct all edges away from the root, so that if $e = (x, y)$ then x is on the path connecting ρ to y. Let $d(\cdot, \cdot)$ 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. For $x \in V$ and $e = (y, z) \in E$, we denote $|x| = d(\rho, x)$, $d(x, (y, z)) = \max\{d(x, y), d(x, z)\}$ and $|e| = d(\rho, e)$. The B-*ary* tree is the infinite rooted tree, where each vertex has exactly B children.

A Markov chain on the tree is a probability measure whose state space is $\mathcal{A}^V$, where $\mathcal{A}$ is a finite set. Without loss of generality we assume that $\mathcal{A} = \{1, \ldots, q\}$. Assume first that T is finite and let $M = (M_{i,j})_{i,j \in \mathcal{A}}$ be a stochastic matrix. In this case the probability measure defined by M on T is given by

$$\overline{\mu}_\ell(\sigma) = \mathbf{1}_{\{\sigma(\rho)=\ell\}} \prod_{(x,y)\in E} M_{\sigma(x),\sigma(y)}. \tag{1}$$

In other words in $\overline{\mu}_\ell$, the root state $\sigma(\rho)$ satisfies $\sigma(\rho) = \ell$ and then each vertex iteratively chooses its state from the one of its parent by an application of the Markov transition rule given by M (and all such applications are independent). We can define the measure $\overline{\mu}_\ell$ on an infinite tree too, by Kolmogorov's extension theorem, but we will not need chains on infinite trees in this paper (see [7] for basic properties of Markov chains on trees).

Instead, for an infinite tree T, we let $T_n = (V_n, E_n, \rho)$, where $V_n = \{x \in V : d(x, \rho) \le n\}, E_n = \{e \in E : d(e, \rho) \le n\}$ and define $\overline{\mu}_\ell^n$ by (1) for T_n. More explicitly,

$$\overline{\mu}_\ell^n(\sigma) = \mathbf{1}_{\{\sigma(\rho)=\ell\}} \prod_{(x,y)\in E_n} M_{\sigma(x),\sigma(y)}. \tag{2}$$

We are particularly interested in the distribution of the states $\sigma(x)$ for $x \in L_n$, the set of leaves in T_n. This distribution, denoted by μ_k^n, is the projection of $\overline{\mu}_k^n$ on $\mathcal{A}^{L_n}$ given by

$$\mu_k^n(\sigma) = \sum \{\overline{\mu}_k^n(\overline{\sigma}) : \overline{\sigma}|L_n = \sigma\}. \tag{3}$$

In this paper we are interested in perturbative theory of the above process. Below we give three definitions of perturbations of μ_n^k representing three different types of "noise". We call a distribution ν on $\mathcal{A} = \{1, \ldots, q\}$ non-degenerate, if $\nu(i) > 0$, for all $1 \le i \le q$.

In the general setting the perturbation is obtained by observing, for leaves $x \in L_n$, not the state $\sigma(x)$ but a state (in a state space $\mathcal{B}$ possibly different from $\mathcal{A}$) derived from $\sigma(x)$ by another random choice (independently for all leaves). The extra choice can be described by a stochastic matrix $N = (N_{i,j})_{i\in\mathcal{A}, j\in B}$; this defines a probability measure on $\mathcal{A}^{V_n} \times \mathcal{B}^{L_n}$ by

$$\overline{\mu}[N]_\ell^n(\sigma, \tau) = 1_{\{\sigma(\rho)=\ell\}} \prod_{(x,y)\in E_n} M_{\sigma(x),\sigma(y)} \times \prod_{y\in L_n} N_{\sigma(y),\tau(y)}, \tag{4}$$

and the distribution of our observed states is the projection $\mu[N]_\ell^n$ on $\mathcal{B}^{L_n}$ given by

$$\mu[N]_\ell^n(\tau) = \sum_\sigma \overline{\mu}[N]_\ell^n(\sigma, \tau). \tag{5}$$

We will mostly be interested in the following types of noise:

- Given $k \geq 0$, define $N = M^k$. Here, for each leaf independently, k additional steps of the chain are performed. We write $\mu_\ell^n[k]$ for $\mu_\ell^n[N]$.
- Given a distribution ν on $\mathcal{A}$, define $N_{i,j} = (1-\epsilon)\mathbf{1}_{\{i=j\}} + \epsilon\nu_j$. Here, for each leaf independently, with probability $1-\epsilon$, there is no noise; otherwise, the leaf state is chosen independently from anything else according to ν. We will write $\mu_k^n[\nu,\epsilon]$ for $\mu[N]_\ell^n$.
- Given $0 \leq \epsilon \leq 1$, we let N be an $q \times (q+1)$ matrix defined by $N_{i,i} = (1-\epsilon)$, $N_{i,q+1} = \epsilon$ and $N_{i,j} = 0$ otherwise. Here, for each leaf independently, the state at the leaf is deleted with probability ϵ (deletion is marked by $q+1$). We write $\mu_\ell^n[\epsilon]$ for $\mu_\ell^n[N]$.

Recall that for distributions μ and ν on the same space Ω the total variation distance between μ and ν is

$$D_V(\mu,\nu) = \frac{1}{2}\sum_{\sigma\in\Omega}|\mu(\sigma) - \nu(\sigma)|. \tag{6}$$

Definition 1.1.

- The reconstruction problem for the B-ary tree T and M is **solvable** if there exist $i,j \in \mathcal{A}$ for which
$$\liminf_{n\to\infty} D_V(\mu_i^n, \mu_j^n) > 0. \tag{7}$$
- The reconstruction problem for the B-ary tree T and M is **robust-solvable** if for all $k < \infty$ there exist $i,j \in \mathcal{A}$ for which
$$\liminf_{n\to\infty} D_V(\mu_i^n[k], \mu_j^n[k]) > 0. \tag{8}$$
- Let ν be a non-degenerate distribution. The reconstruction problem for the B-ary tree T and M is **ν-robust-solvable** if for all $\epsilon < 1$ there exist $i,j \in \mathcal{A}$ for which
$$\liminf_{n\to\infty} D_V(\mu_i^n[\nu,\epsilon], \mu_j^n[\nu,\epsilon]) > 0. \tag{9}$$
- The reconstruction problem for the B-ary tree T and M is **erasure-robust-solvable** if for all $\epsilon < 1$ there exist $i,j \in \mathcal{A}$ for which
$$\liminf_{n\to\infty} D_V(\mu_i^n[\epsilon], \mu_j^n[\epsilon]) > 0. \tag{10}$$

Note that by taking $\epsilon = 0$ in (9) or (10) we obtain the original reconstruction condition (7). The same is true if $k = 0$ in (8).

Let $\lambda_2(M)$ denote the eigenvalue of M which has the second largest absolute value ($\lambda_2(M)$ may be negative or non-real). In our main result we prove

Theorem 1.2. *Consider an ergodic Markov chain on the B-ary tree such that $B|\lambda_2(M)|^2 < 1$. Then*

- *The reconstruction problem is not robust-solvable. Moreover, there exists k^* such that for all $k > k^*$*
$$\max_{i,j}\lim_{n\to\infty} D_V(\mu_i^n[k], \mu_j^n[k]) = 0. \tag{11}$$

- *For all non-degenerate ν, the reconstruction problem is not ν-robust-solvable. Moreover, for all non-degenerate ν, there exists $\epsilon^* < 1$ such that for all $\epsilon > \epsilon^*$,*

$$\max_{i,j} \lim_{n \to \infty} D_V(\mu_i^n[\nu, \epsilon], \mu_j^n[\nu, \epsilon]) = 0. \tag{12}$$

- *If all the entries of M are non-zero, then the reconstruction problem is not erasure-robust-solvable. Moreover, there exists an $\epsilon^* < 1$ such that for all $\epsilon > \epsilon^*$,*

$$\max_{i,j} \lim_{n \to \infty} D_V(\mu_i^n[\epsilon], \mu_j^n[\epsilon]) = 0. \tag{13}$$

It is easy to see that the total variation distances in (11), (12) and (13) are monotone decreasing in k, ϵ and ϵ respectively

The following proposition follows immediately from [12] or from the proofs in [22]. Together with Theorem 1.2, it shows that the threshold for robust reconstruction is given by $B|\lambda_2(M)|^2 = 1$.

Proposition 1.3. *Consider an ergodic Markov chain on the B-ary tree where $B|\lambda_2(M)|^2 > 1$. Then*

- *The reconstruction problem is robust-solvable,*
- *For all non-degenerate ν the reconstruction problem is ν-robust-solvable and*
- *The reconstruction problem is erasure-robust-solvable.*

1.2 Discussion

The reconstruction problem was first studied in statistical physics [28, 9], where the problem was phrased in terms of extremality of the free measure for the Ising model on the $(B+1)$-regular tree (Bethe lattice). It is not too hard to see (e.g. [3]) that the measure is non-extremal if and only if the reconstruction problem is solvable for the Markov chain on the B-ary tree with transition probabilities given by the binary symmetric Markov chain: $\begin{pmatrix} 1-\delta & \delta \\ \delta & 1-\delta \end{pmatrix}$. δ is related to the "inverse temperature" β by $1 - 2\delta = \tanh(2\beta)$. The equivalence between non-extremality of the free measure of a random field and reconstruction solvability of an associated Markov chain on the same tree holds under mild non-degeneracy conditions (see e.g. [18]).

In the past decade, the reconstruction problem re-appeared in many applications: In communication networks (see [3] and the references there), in noisy computation (a model introduced by von Neumann in [29], see [8, 4]) and in phylogeny (molecular evolution, see [5, 26] for general background) [27]. Most recently, it is shown that the reconstruction problem is of crucial importance to basic questions in phylogeny [20, 19, 23]. In all of these applications the interest is to find when is it possible to reconstruct some information on the root state from states at the leaves of a finite tree. In many of the applications it is natural to consider robust-reconstruction as the observed data goes via additional "noise mechanism".

Solvability of the reconstruction problem is also closely related to mixing rate of Glauber dynamics on the tree. See [1, 16] where it is shown that non-solvability roughly corresponds to rapid mixing dynamics on the tree.

Determining if the reconstruction problem is solvable or not turns out to be very hard. Binary symmetric Markov chains is the only family for which the threshold for reconstruction solvability is known. Even here there is a generation gap between the proof of the lower-bound [9] and proofs of the upper bound [2] (see also [10] for a different proof, [3] for the result on a general trees and [24] for the critical case on general trees). For binary symmetric Markov chains on the B-ary tree the threshold for the reconstruction problem is given by $B(1-2\delta)^2 = 1$, or equivalently $B\lambda_2(M)^2 = 1$. For all other families of Markov chain, including q-ary symmetric Markov chain for $q > 2$ and general 2×2 Markov chains only bounds are known [22, 15, 16].

The threshold $B|\lambda_2(M)|^2 = 1$ is also the threshold for "census-solvability" [22], where different nodes of L_n are indistinguishable (in other words, we only observe the "census" of level n). However, in general, it is not the threshold for reconstruction. Indeed, except for the binary symmetric channel we know of no family of chains for which $B|\lambda_2(M)|^2 = 1$ is the threshold for reconstruction. Moreover, [18] shows that for asymmetric binary Markov chains (general stochastic 2 by 2 matrices) or symmetric Markov chains for $q > 2$ (where $M_{i,j} = (1-\delta)\mathbf{1}_{\{i=j\}} + \frac{\delta}{q-1}\mathbf{1}_{\{i \neq j\}}$ for $i, j \in \{1, \ldots, q\}$) the reconstruction problem is sometimes solvable even when $B|\lambda_2(M)|^2 < 1$. In [18] there is also a construction of M with $\lambda_2(M) = 0$ for which the reconstruction problem is solvable for large B.

Why is determining the threshold for reconstruction hard? From the technical point of view a Markov chain on the tree corresponds to a recursion in some random variables ([2, 24, 16]). A natural way to analyze these recursions is to use a perturbative argument around the stationary distribution of the chain. The main problem is that the random variables we start with are atoms - far from the stationary distribution and that in general the recursions lack any convexity. For "robust-reconstruction" the problem is easier - as the recursions begin close to the stationary distribution.

Our proof is based on a new measure of *discrepancy* for a vector of distributions which is a weighted variant of the χ^2 distance. We show that an application of the chain M contracts the discrepancy by a $|\lambda_2(M)|^2$ factor, and that if the discrepancy is smaller than δ, then tensoring B copies of the distributions increases the discrepancy by a factor of at most $B(1+\epsilon(\delta))$, where $\epsilon(\delta) \to 0$, as $\delta \to 0$.

It is interesting to compare our results with the results of [25]. In [25] Pemantle and Steif study robust phase transition on trees. For a Gibbs measure on a tree we say that a robust phase transition occurs if the boundary conditions on a cutset have a non-vanishing effect on the root even when the interactions along the cutsets are made arbitrarily small but fixed (see [25] for exact definition). The main results of [25] give the exact threshold for robust phase transitions for general (bounded degree) trees for Potts and Heisenberg models in terms of the underlying model and the branching number (see [14]) of the tree.

Both in our result and in the results of [25], it is easier to analyze the "robust" problem than it is the original problem for similar reasons. In both cases the "non-robust" problem is hard to control without some convexity assumption, while the solution of "robust" problem allows the use of "local" arguments.

Moreover, like robust phase transition, robust reconstruction is a geometric property, i.e. for general bounded degree T, the threshold for robust-reconstruction depends only $\mathrm{br}(T)$ and $|\lambda_2(M)|$. Indeed, the proof of Theorem 3.3 combines the analysis of the new discrepancy measure introduced here, with some of the techniques developed in [25] for controlling recursions on general trees.

A natural open problem is to determine the behavior of robust-solvability in the critical case, where $B|\lambda_2(M)|^2 = 1$. Our techniques shed no light on this problem. It is also interesting to try and remove the restriction that the entries of M are positive for (13) – see also Remark 2.10. Finally, in the proof presented for Theorem 3.3, for fixed M and ν, the bounds on ϵ and k are becoming weaker as $\mathrm{br}(T)|\lambda_2(M)|^2$ approaches 1 (i.e., $\epsilon \to 1$ and $k \to \infty$). It is natural to ask if for given M, ν and K, there exist ϵ and k for which the result holds uniformly for all infinite trees T with $\mathrm{br}(T)|\lambda_2(M)|^2 < 1$.

Acknowledgment. This research was carried out during a visit to the Computation, Combinatorics and Probability program at the Isaac Newton Institute for Mathematical Sciences in Cambridge, U.K. The second author is supported by a Miller fellowship at U.C. Berkeley. We thank Yuval Peres for suggesting the *3rd* open problem above.

2 Proof

Recall that we denote by M the transition matrix. In this section we will often multiply M from the right by a vector of functions / from the left by a vector of measures – in which case the resulting vector would also be a vector of functions / vector of measures.

Let $\mathbf{1} = (1, \ldots, 1)^t$, then clearly $M\mathbf{1} = \mathbf{1}$. Let v_i be the stationary probability of state i, and $v = (v_1, \ldots, v_q)$ the stationary distribution, so that $vM = v$. In the remainder of this section we will use $a, b, c, \ldots$ for column vectors, and $u, v, w, \ldots$ for row vectors. v will always denote the stationary distribution.

Note that if b is a column vector such that $vb = 0$, then $vMb = vb = 0$. In other words, the linear space $v^\perp = \{b \in \mathbb{R}^q : vb = 0\}$ is invariant under M.

Lemma 2.1. *Let $b = (b_1, \ldots, b_q)^t$ be a vector such that $b_i > 0$ for all i. Then*

$$\sum_{i=1}^{q} \frac{v_i}{(Mb)_i} \le \sum_{i=1}^{q} \frac{v_i}{b_i}.$$

Proof. By Jensen,

$$\frac{1}{(Mb)_i} = \frac{1}{\sum_j m_{i,j} b_j} \le \sum_j m_{i,j} \frac{1}{b_j}.$$

Hence,

$$\sum_i \frac{v_i}{(Mb)_i} \le \sum_i \sum_j v_i m_{i,j} \frac{1}{b_j} = \sum_j (\sum_i v_i m_{i,j}) \frac{1}{b_j} = \sum_j \frac{v_j}{b_j},$$

as needed. ■

By looking at the Jordan form of M it is easy to see that

Lemma 2.2. *Given $\epsilon > 0$, there exists an Euclidean norm $\| \ \|$ on $v^\perp$ such that $\|Mb\| \le (|\lambda_2(M)| + \epsilon)\|b\|$ for all $b \in v^\perp$.* ■

Let Q be the projection onto $v^\perp$ defined by $Qb = b - (vb)\mathbf{1}$ (note that $vQb = vb - (vb)(v\mathbf{1}) = 0$ for all b).

Definition 2.3. Let $\| \ \|$ be a Euclidean norm on $v^\perp$. Let $\nu = (\nu_1, \dots, \nu_q)$ be a vector of distributions on a common space. Let $f = (f_1, \dots, f_q)$ be the vector of density functions with respect to a σ-finite measure μ, such that $\nu_i \ll \mu$ for every i. In other words, $d\nu_i = f_i d\mu$ for all i. We then define the *discrepancy* of the vector by

$$D^\mu_{\| \, \|}(f) = \int \|Qf\|^2 \sum_{i=1}^q \frac{v_i}{f_i} \, d\mu.$$

We also write $D(\nu)$ and $D(f)$ for the discrepancy, without explicitly indicating the norm and reference measure.

Note the similarity between the discrepancy and the χ^2-distance. The χ^2-distance is known to be well behaved with respect to ℓ_2 norms. Note that if $f_1 = \dots = f_q$, then $Qf = fQ\mathbf{1} = 0$. Thus, Q projects into the orthogonal complement of the space where the discrepancy should be 0.

Lemma 2.4. *$D(\nu)$ is independent of the reference measure μ; i.e., if $f = (f_1, \dots, f_q)$ and $g = (g_1, \dots, g_q)$ are such that $d\nu_i = f_i \, d\mu = g_i \, d\tilde{\mu}$ for all i, then $D^\mu_{\| \, \|}(f) = D^{\tilde{\mu}}_{\| \, \|}(g)$.*

Proof. Assume that $\tilde{\mu} \ll \mu$. The general case then follows by considering the three reference measures μ, $\mu + \tilde{\mu}$, $\tilde{\mu}$.

Since Q is linear and $\| \ \|$ is Euclidean, we may write $\|Qb\|^2 = \sum_{i,j=1}^q t_{i,j} b_i b_j$ for some $t_{i,j} \in \mathbb{R}$. Now

$$D^\mu(f) = \sum_{r,i,j=1}^q v_r t_{i,j} \int \frac{f_i f_j}{f_r} d\mu = \sum_{r,i,j=1}^q v_r t_{i,j} \int \frac{g_i g_j \left(\frac{d\tilde{\mu}}{d\mu}\right)^2}{g_r \frac{d\tilde{\mu}}{d\mu}} d\mu = \sum_{r,i,j=1}^q v_r t_{i,j} \int \frac{g_i g_j}{g_r} d\tilde{\mu} = D^{\tilde{\mu}}(g).$$

■

Lemma 2.5. *Let the norm $\| \ \|$ on $v^\perp$ satisfy $\|Mb\| \le \alpha \|b\|$ for all $b \in v^\perp$ and some constant α. Then $D(Mf) \le \alpha^2 D(f)$ for all f as in Definition 2.3.*

Proof. For all b,

$$MQb = M(b - (vb)\mathbf{1}) = Mb - (vb)M\mathbf{1} = Mb - (vb)\mathbf{1} = Mb - (vMb)\mathbf{1} = QMb.$$

Therefore we have pointwise that

$$\|QMf\|^2 = \|MQf\|^2 \le \alpha^2 \|Qf\|^2. \tag{14}$$

Now

$$\begin{aligned} D(Mf) &= \int \|QMf\|^2 \sum_{i=1}^q \frac{v_i}{(Mf)_i} \, d\mu \\ &\le \alpha^2 \int \|Qf\|^2 \sum_{i=1}^q \frac{v_i}{(Mf)_i} \, d\mu \qquad (15) \\ &\le \alpha^2 \int \|Qf\|^2 \sum_{i=1}^q \frac{v_i}{f_i} \, d\mu = \alpha^2 D(f), \qquad (16) \end{aligned}$$

where (15) follows from (14), and (16) follows by Lemma 2.1. ■

Lemma 2.6. *For every Euclidean norm $\| \ \|$ on $v^\perp$, there exists a constant $C = C(\| \ \|)$, such that*

$$\left| \int \frac{f_i f_j}{f_k}\, d\mu - 1 \right| \le CD(f), \tag{17}$$

for all i, j, k, where f and μ are as in Definition 2.3.

Proof. By Cauchy–Schwartz,

$$\left| \int \frac{f_i f_j}{f_k}\, d\mu - 1 \right| = \left| \int \frac{(f_i - f_k)(f_j - f_k)}{f_k}\, d\mu \right| \le \sqrt{\int \frac{|f_i - f_k|^2}{f_k}\, d\mu} \sqrt{\int \frac{|f_j - f_k|^2}{f_k}\, d\mu}.$$

Therefore in order to prove the lemma, it suffices to prove that there exists a constant C such that for all i, j, k it holds that

$$\int \frac{|f_i - f_j|^2}{f_k}\, d\mu \le CD(f). \tag{18}$$

Note that

$$(Qb)_i - (Qb)_j = (b - (vb)\mathbf{1})_i - (b - (vb)\mathbf{1})_j = b_i - b_j.$$

Therefore for all b, it holds that, for some constant C_{ij},

$$|b_i - b_j| = |(Qb)_i - (Qb)_j| \le C_{ij} \|Qb\|.$$

Hence

$$\int \frac{|f_i - f_j|^2}{f_k}\, d\mu \le C_{ij}^2 \int \frac{\|Qf\|^2}{f_k} \le \frac{C_{ij}^2}{v_k} D(f).$$

Now (18) follows by taking $C = \sup_{i,j,k} \frac{C_{ij}^2}{v_k}$. ■

Given a σ-finite measure μ on a space X, we denote by $\mu^{\otimes B}$ the product measure on X^B with marginals μ. Similarly, if f_i is a density of ν_i with respect to μ, write $f_i^{\otimes B}$ for the density of $\nu_i^{\otimes B}$ with respect to $\mu^{\otimes B}$. Finally for $f = (f_1, \ldots, f_q)$, write $f^{\otimes B} = (f_1^{\otimes B}, \ldots, f_q^{\otimes B})$, and for $\nu = (\nu_1, \ldots, \nu_q)$, write $\nu^{\otimes B} = (\nu_1^{\otimes B}, \ldots, \nu_q^{\otimes B})$. We similarly use $\bigotimes_{r=1}^{B} \nu^r$ for the componentwise product of several vectors ν^r of measures.

Lemma 2.7. *Let $\| \ \|$ be an Euclidean norm on $v^\perp$, $B \ge 1$ an integer and $\epsilon > 0$. Then there exists a $\delta > 0$ such that if $\nu^1 = (\nu_1^1, \ldots, \nu_q^1)$, $\ldots$, $\nu^B = (\nu_1^B, \ldots, \nu_q^B)$, satisfy $D(\nu^i) \le \delta$ for $1 \le i \le B$, then*

$$D\big(\textstyle\bigotimes_{r=1}^{B} \nu^r\big) \le (1 + \epsilon)\big(D(\nu^1) + \cdots + D(\nu^B)\big).$$

In particular, given $\epsilon > 0$, there exists a $\delta > 0$ such that if $\nu = (\nu_1, \ldots, \nu_q)$ satisfies $D(\nu) \le \delta$, then $D(\nu^{\otimes B}) \le (1 + \epsilon) B D(\nu)$.

Proof. The second part of the lemma immediately follows from the first part. Choose a reference measure μ with $\nu_i^r \ll \mu$ for every i and r, and let f_i^r be the density $d\nu_i^r / d\mu$.

As in Lemma 2.4, we may write $\|Qb\|^2 = \sum_{i,j} t_{i,j} b_i b_j$. Moreover, since $\|Q\mathbf{1}\|^2 = \|0\|^2 = 0$, it follows that $\sum_{i,j} t_{i,j} = 0$. Hence

$$D(f) = \int \|Qf\|^2 \sum_{k=1}^{q} \frac{v_k}{f_k}\, d\mu = \sum_{i,j,k} v_k t_{i,j} \int \frac{f_i f_j}{f_k}\, d\mu = \sum_{i,j,k} v_k t_{i,j} \left(\int \frac{f_i f_j}{f_k}\, d\mu - 1 \right). \tag{19}$$

Substituting $\bigotimes_{r=1}^{B} f^r$ in (19) we obtain, using the reference measure $\mu^{\otimes B}$,

$$\begin{aligned} D\left(\bigotimes_{r=1}^{B} f^r\right) &= \sum_{i,j,k} v_k t_{i,j} \left(\int \frac{\bigotimes_{r=1}^{B} f_i^r \bigotimes_{r=1}^{B} f_j^r}{\bigotimes_{r=1}^{B} f_k^r}\, d\mu^{\otimes B} - 1 \right) \qquad (20) \\ &= \sum_{i,j,k} v_k t_{i,j} \left(\prod_{r=1}^{B} \left(\int \frac{f_i^r f_j^r}{f_k^r}\, d\mu \right) - 1 \right). \end{aligned}$$

Let C be chosen to satisfy (17) in Lemma 2.6, and $\tilde{C} = C \sum_{i,j} |t_{i,j}|$. Let δ be chosen such that for all $(x_1, \ldots, x_B) \in [1 - C\delta, 1 + C\delta]^B$ it holds that

$$\left| \prod_{r=1}^{B} x_r - 1 - \sum_{r=1}^{B} (x_r - 1) \right| \leq \frac{\epsilon \sum_{r=1}^{B} |x_r - 1|}{\tilde{C}}. \tag{21}$$

By Lemma 2.6, it follows that if $D(f) \leq \delta$, then for all i, j, k,

$$\left| \int \frac{f_i f_j}{f_k}\, d\mu - 1 \right| \leq C D(f) \leq C\delta.$$

Therefore, it follows from (20) and (21) that

$$\begin{aligned} D\left(\bigotimes_{r=1}^{B} f^r\right) &\leq \sum_{i,j,k,r} v_k t_{i,j} \left(\int \frac{f_i^r f_j^r}{f_k^r}\, d\mu - 1 \right) + \frac{\epsilon}{\tilde{C}} \sum_{i,j,k,r} v_k |t_{i,j}| \left| \int \frac{f_i^r f_j^r}{f_k^r}\, d\mu - 1 \right| \\ &= \sum_r D(f^r) + \frac{\epsilon}{\tilde{C}} \sum_{i,j,k,r} v_k |t_{i,j}| \left| \int \frac{f_i^r f_j^r}{f_k^r}\, d\mu - 1 \right| \\ &\leq \sum_r D(f^r) + \frac{\epsilon}{\tilde{C}} \sum_{i,j,k,r} v_k |t_{i,j}| C D(f^r) \qquad (22) \\ &= (1 + \epsilon) \sum_r D(f^r), \end{aligned}$$

where inequality (22) follows from Lemma 2.6. ■

Lemma 2.8. *Given a Euclidean norm $\| \ \|$ on $v^{\perp}$, there exists a constant $C(\| \ \|) < \infty$ such that for any vector $\nu = (\nu_i)_{i=1}^{q} = (f_i d\mu)_{i=1}^{q}$ of distributions we have*

$$\sup_{i,j} d_{TV}(\nu_i, \nu_j) = \sup_{i,j} \int |f_i - f_j|\, d\mu \leq C \sqrt{D(f)}, \tag{23}$$

where $f = (f_1, \ldots, f_q)$.

Proof. By Cauchy–Schwartz,

$$\int |f_i - f_j|\, d\mu \le \sqrt{\int \frac{|f_i - f_j|^2}{f_i}\, d\mu} \sqrt{\int f_i\, d\mu} = \sqrt{\int \frac{|f_i - f_j|^2}{f_i}\, d\mu},$$

and (23) follows from Lemma 2.6. ■

Lemma 2.9. • *Let $\| \ \|$ be a Euclidean norm on $v^{\perp}$. Let μ be a probability distribution on $1, \ldots, q$ such that $\mu(i) > 0$ for all i. Then for all $\delta > 0$, there exists an $\epsilon^* = \epsilon^*(\delta) < 1$ such that for any vector $\nu' = ({\nu'}_1, \ldots, {\nu'}_q)$ of probability distributions on $1, \ldots, q$ and for all $\epsilon > \epsilon^*$, if*

$$\nu = (1-\epsilon)({\nu'}_1, \ldots, {\nu'}_q) + \epsilon(\mu, \ldots, \mu),$$

then $D(\nu) \le \delta$.

- *Let $\| \ \|$ be a Euclidean norm on $v^{\perp}$. Let $\nu_\ell^{(r)}$ be denote the ℓ'th row of M^r, i.e., the distribution of the chain given by M^r, after r steps starting at ℓ, and $\nu^{(r)} = (\nu_1^{(r)}, \ldots, \nu_q^{(r)})$. Then for all $\delta > 0$, there exists an $r^* = r^*(\delta)$ such that if $r \ge r^*$ then $D(\nu^{(r)}) \le \delta$.*

- *Let $\| \ \|$ be an Euclidean norm on $v^{\perp}$. Let $\nu = (\nu_1, \ldots, \nu_q)$ be a vector of probability measures such that for all $1 \le i, j \le q$ it holds that $\nu_i(j) > 0$. Let $\nu^\epsilon = (\nu_1^\epsilon, \ldots, \nu_q^\epsilon)$ be a collection of probability measures on $\{1, \ldots, q+1\}$ such that $\nu_i^\epsilon = (1-\epsilon)\nu_i + \epsilon\nu'$, where ν' is the delta measure on $q+1$. Then for all $\delta > 0$, there exists an $\epsilon^* = \epsilon^*(\delta) < 1$ such that if $\epsilon > \epsilon^*$, then $D(\nu^\epsilon) \le \delta$.*

Proof. For the first part of the lemma we use the representation of $D(\nu)$ as in (19) with respect to the measure μ. Let $m = \min\{\mu(1), \ldots, \mu(q)\}$, and observe that if $d\nu_i = f_i\, d\mu$, then

$$f_i = \frac{d\nu_i}{d\mu} = \epsilon + (1-\epsilon)\frac{d{\nu'}_i}{d\mu},$$

so

$$\epsilon \le f_i \le \epsilon + (1-\epsilon)/m,$$

and for all i, j, k,

$$\frac{f_i f_j}{f_k} \le \frac{(\epsilon + (1-\epsilon)/m)^2}{\epsilon}.$$

Hence by (19)

$$D(\nu) \le \left(\frac{(\epsilon + (1-\epsilon)/m)^2}{\epsilon} - 1 \right) \sum_{i,j} |t_{i,j}|, \tag{24}$$

and the right hand side of (24) converges to 0 as $\epsilon \to 1$.

The second part of the lemma follows from the first one as the ergodicity of M implies that for all i, $\nu_i^{(m)}$ converges to the stationary distribution of the chain as $m \to \infty$.

The third part of the lemma is proven similarly to the first part. Let $m = \min_{i,j \in \mathcal{A}} \nu_i(j)$ and

$$\mu^\epsilon = \epsilon\nu' + \frac{(1-\epsilon)}{q} \sum_{i=1}^{q} \nu_i.$$

Note that if $d\nu_i^\epsilon = f_i\, d\mu^\epsilon$, then $m \le f_i(\ell) \le q$ for $1 \le \ell \le q$ and all i, and $f_i(q+1) = 1$ for all i. It follows that for all i, j, k and $1 \le \ell \le q$,

$$\tfrac{m^2}{q} \le \tfrac{f_i(\ell) f_j(\ell)}{f_k(\ell)} \le \tfrac{q^2}{m} \quad \text{and} \quad \tfrac{f_i(q+1) f_j(q+1)}{f_k(q+1)} = 1.$$

Moreover $\mu^\epsilon(q+1) = \epsilon$ and $\mu^\epsilon(\{1, \ldots, q\}) = 1 - \epsilon$, so for all i, j and k,

$$\int \frac{f_i f_j}{f_k}\, d\mu - 1 = \int \left(\frac{f_i f_j}{f_k} - 1 \right) d\mu \le (\frac{q^2}{m} - 1)(1 - \epsilon).$$

Hence by (19)

$$D(\nu^\epsilon) = \sum_{i,j,k} v_k t_{i,j} \left(\int \frac{f_i f_j}{f_k}\, d\mu - 1 \right) \le (\frac{q^2}{m} - 1)(1 - \epsilon) \sum_{i,j} |t_{i,j}| \tag{25}$$

and the right hand side of (25) converges to 0 as $\epsilon \to 1$. ■

Proof of Theorem 1.2. The basic idea of the proof is that if $\mu[N]^n$ is the vector of probability measures $(\mu[N]_1^n, \ldots, \mu[N]_q^n)$ defined in (5), then we may write $\mu[N]^{n+1}$ in terms of $\mu[N]^n$ using the operator M and tensoring. This will allow us to bound discrepancies recursively. Let $\rho_1, \ldots, \rho_B$ be the B children of ρ in the B-ary tree. Write $E_{n+1}(s)$ for the edges in E_{n+1} that are on the subtree rooted in ρ_s (formally, these are the edges of E_{n+1} that are connected to ρ only by paths going via ρ_s). Define $L_{n+1}(s)$ similarly. Finally, for a configuration σ of the vertices at the first n level of the tree, let $\sigma_1, \ldots, \sigma_B$ denote the configurations restricted to the subtrees rooted at $\rho_1, \ldots, \rho_B$. Then by (4) and (5)

$$\begin{aligned}
\mu[N]_\ell^{n+1}(\tau) &= \sum_\sigma 1_{\{\sigma(\rho)=\ell\}} \prod_{(x,y)\in E_{n+1}} M_{\sigma(x),\sigma(y)} \prod_{y \in L_{n+1}} N_{\sigma(y),\tau(y)} \\
&= \prod_{s=1}^B \left(\sum_{\ell'=1}^q M_{\ell,\ell'} \sum_{\sigma_s} 1_{\{\sigma_s(\rho_s)=\ell'\}} \prod_{(x,y)\in E_{n+1}(s)} M_{\sigma_s(x),\sigma_s(y)} \prod_{y\in L_{n+1}(s)} N_{\sigma_s(y),\tau_s(y)} \right) \\
&= \prod_{s=1}^B \left(\sum_{\ell'=1}^q M_{\ell,\ell'} \mu[N]_{\ell'}^n (\tau | L_{n+1}(s)) \right).
\end{aligned} \tag{26}$$

Note that the expression in the parenthesis in (26) is given by $(M(\mu[N]^n))_\ell(\tau|L_{n+1}(s))$, the ℓ'th coordinate of the vector $M(\mu[N]^n)$. It is now easy to see that

$$\mu[N]^{n+1} = (M(\mu[N]^n))^{\otimes B}. \tag{27}$$

We use (27) in order to bound discrepancies recursively.

The assumption $B|\lambda_2(M)|^2 < 1$, implies by Lemma 2.2 that there exists an $\epsilon > 0$ and a norm $\| \ \|$ on $v^\perp$ such that for all $b \in v^\perp$, it holds that $\|Mb\| \le \alpha \|b\|$, where $B(1+\epsilon)\alpha^2 \le 1 - \epsilon$. Let δ be chosen as to satisfy Lemma 2.7, so that if $D(f) \le \delta$, then $D(f^{\otimes B}) \le B(1+\epsilon) D(f)$.

By Lemma 2.9 it follows that there exist a k^* such that $\mu^0 = (\mu_1^0[k], \ldots, \mu_q^0[k])$ satisfies $D(\mu^0) \le \delta$ for all $k \ge k^*$. Write μ^n for $(\mu_1^n[k], \ldots, \mu_q^n[k])$. Thus (27) implies that $\mu^{n+1} = (M\mu^n)^{\otimes B}$. It now follows by Lemmas 2.7 and 2.5 that

$$D(\mu^{n+1}) \le B(1+\epsilon)\alpha^2 D(\mu^n) \le (1-\epsilon) D(\mu^n).$$

Hence $\lim_{n\to\infty} D(\mu^n) = 0$. We therefore conclude from Lemma 2.8 that

$$\lim_{n\to\infty} \max_{i,j} d_{TV}(\mu_i^n[k], \mu_j^n[k]) = 0,$$

and (11) follows.

In order to prove (12), let ν be a non-degenerate measure and note that by Lemma 2.9 it follows that there exist a $\epsilon^* < 1$ such that $\mu^0 = (\mu_1^0[\nu,\epsilon], \ldots, \mu_q^0[\nu,\epsilon])$ satisfies $D(\mu^0) \leq \delta$ for all $\epsilon > \epsilon^*$. Now (12) follows similarly to (11).

The proof of (13) is similar. We look at $\mu^1 = (\mu_1^1[\epsilon], \ldots, \mu_q^1[\epsilon])$. Note that (27) implies that $\mu^1 = (M\mu^0)^{\otimes B}$. Let $\nu = M\mu^0$. Note that the vector ν satisfies for all i that $\nu_i(q+1) = \epsilon$ and $\nu_i(j) = (1-\epsilon)M_{i,j}$ otherwise.

Since all the entries of M are positive, it follows from Lemma 2.9 that for every $\delta' > 0$ there exists an $\epsilon^* < 1$ such that if $\epsilon > \epsilon^*$ then $D(\nu) \leq \delta'$. We may now apply Lemma 2.7 and choose $\delta' > 0$ in such a way that

$$D(\mu^1) = D(\nu^{\otimes B}) \leq \delta.$$

The rest of the proof is identical. ■

Remark 2.10. It is an interesting goal to extend (13) to general ergodic chains (where some of the entries of M may be zero). Above we proved this for the case where all the entries of M are positive.

The proof of Lemma 2.9 can easily be extended to the case when there exists an n such that the measures $\mu_1^n, \ldots, \mu_q^n$ all have the same support (in such a case one can prove that there exists a value of $\epsilon < 1$ such that $D(\mu^n) \leq \delta$ by showing that for ϵ sufficiently large the measures μ_i^n have most of their mass on the atom $(q+1, \ldots, q+1)$ and bounded relative densities elsewhere). However, we do not know any simple characterizations of the matrices M for which this holds (it evidently depends only on the set of zero entries of M); nor we do believe that this property is necessary for non-erasure-robust-solvability.

3 General trees

Our result readily extend to general infinite bounded degree trees, where B is replaced by $\text{br}(T)$, the branching number of the tree. In [6], Furstenberg introduced the notion of the Hausdorff dimension of a tree. Later, Lyons [13, 14] showed how many of the probabilistic properties of the tree are determined by this number which he named the branching number.

For our purposes it is best to define the branching number via cutsets. A *cutset* S for a tree T rooted at ρ, is a finite set of vertices separating ρ from ∞. In other words, a finite set S is a cutset, if every infinite self avoiding path from ρ intersects S. An *antichain* is a cutset that does not have any proper subset which is also a cutset.

We follow the notation of [25] and for a cutset S, write $\text{Ins}(S)$ for the inside of S (the finite component of $T \setminus S$, containing the root ρ), $\text{Ins}_E(S)$ for edges inside S (those edges (x, y) having $x \in \text{Ins}(S)$) and $\text{Out}(S)$ for the outside of S ($\text{Out}(S) = T \setminus (S \cup \text{Ins}(S))$).

Definition 3.1. The branching number of T is defined as

$$\mathrm{br}(T) = \inf\{\lambda > 0 : \inf_{\text{cutsets } S} \sum_{x \in S} \lambda^{-|x|} = 0\}.$$

Note that

$$\mathrm{br}(T) = \inf\{\lambda > 0 : \inf_{\text{antichains } S} \sum_{x \in S} \lambda^{-|x|} = 0\}.$$

By Min-Cut-Max-Flow, $\mathrm{br}(T)$ is also the supremum of the real numbers $\lambda > 0$, such that T admits a positive flow from the root to infinity, where on every edge e of T, the flow is bounded by $\lambda^{-|e|}$. It is shown in [14] that $\mathrm{br}(T)^{-1}$ is the critical probability for Bernoulli percolation on T. See [14] and [3] for equivalent definitions of $\mathrm{br}(T)$ in terms of percolation, cutset sums and electrical conductance. We note that $\mathrm{br}(T_B) = B$ for the B-ary tree T_B.

As in Section 1, the Markov chain on T is described by an $|\mathcal{A}| \times |\mathcal{A}|$ stochastic matrix M and the perturbations by an $|\mathcal{A}| \times |\mathcal{B}|$ stochastic matrix N. For B-ary trees we observed the process on the special antichains L_n; for general trees it seems more natural to consider arbitrary antichains. The distribution $\mu[N]_\ell^S$ of the observed (perturbed) states on an antichain S in T is given by, extending (4) and (5),

$$\mu[N]_\ell^S(\tau) = \sum_\sigma 1_{\{\sigma(\rho)=\ell\}} \prod_{(x,y)\in \mathrm{Ins}_E(S)} M_{\sigma(x),\sigma(y)} \times \prod_{y \in S} N_{\sigma(y),\tau(y)}. \tag{28}$$

We proceed by defining μ_ℓ^S, the measure $\mu_\ell^S[k]$ for $k \geq 0$, the measure $\mu_\ell^S[\nu, \epsilon]$ for $\epsilon > 0$ and non-degenerate distribution ν on $\mathcal{A}$ and $\mu_\ell^S[\epsilon]$. This is done in exactly the same way as in the case of the B-ary tree, by choosing appropriate N's in (28).

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

$$\inf_{S \text{ antichain}} D_V(\mu_i^S, \mu_j^S) > 0,$$

where μ_ℓ^S denotes the conditional distribution on σ_S given that $\sigma(\rho) = \ell$. We similarly define the notions of robust-solvable, ν-robust-solvable and erasure-robust-solvable.

Remark 3.2. The definitions of solvability for general trees and B-ary tree are *not* compatible. If T is the B-ary tree, then solvability by Definition 1.1 involves only cutsets $S = L_n$ and is therefore a weaker condition than solvability defined here which involves all antichains (same for robust-solvable etc.). However we will obtain the same threshold for robust-reconstruction under both definitions.

The proof of our main result extends to show that

Theorem 3.3. *Consider an ergodic Markov chain on a rooted tree T such that* $\mathrm{br}(T)|\lambda_2(M)|^2 < 1$. *Then*

- *The reconstruction problem is not robust-solvable.*
- *For all non-degenerate ν, the reconstruction problem is not ν-robust-solvable.*

- *If all the entries of M are non-zero, then the reconstruction problem is not erasure-robust-solvable.*

This proves that the threshold for robust reconstruction is given by $\mathrm{br}(T)|\lambda_2(M)|^2 = 1$ as the proof of Theorem 1.4 in [22] immediately generalizes to show that

Proposition 3.4. *Consider an ergodic Markov chain on a tree T such that $\mathrm{br}(T)|\lambda_2(M)|^2 > 1$. Then the reconstruction problem is robust-solvable, for all non-degenerate ν the reconstruction problem is ν-robust-solvable and the reconstruction problem is erasure-robust-solvable.*

We now turn to the proof of Theorem 3.3. The proof generalizes the proof of Theorem 1.2. For a vertex x of the rooted tree T we write $T(x)$ for the subtree rooted at x, i.e. the subtree consisting of x and all of its descendents. We will use the following lemma from Pemantle and Steif [25].

Lemma 3.5 ([25, Lemma 3.3]). *Assume that $\mathrm{br}(T) < g$. Then for all $\epsilon > 0$ there exists an anitchain S such that*

$$\sum_{x \in S} \left(\frac{1}{g}\right)^{|x|} \leq \epsilon, \tag{29}$$

and for all $y \in S \cup \mathrm{Ins}(S)$,

$$\sum_{x \in S \cap T(v)} \left(\frac{1}{g}\right)^{|x|-|y|} \leq 1. \tag{30}$$

Proof of Theorem 3.3. We will show that under the conditions of Theorem 3.3 the reconstruction problem is not robust-solvable.

Let S be an antichain and $y \in S \cup \mathrm{Ins}(S)$. Consider the Markov chain on the subtree $T(y)$, starting with state ℓ at y, and let $\mu[N]_\ell^{y,S}$ be the distribution of the observed states on $T(y) \cap S$. Thus,

$$\mu[N]_\ell^{y,S}(\tau) = \sum_\sigma \mathbf{1}_{\{\sigma(y)=\ell\}} \prod_{(x,z) \in \mathrm{Ins}_E(S), x \in T(y), z \in T(y)} M_{\sigma(x),\sigma(z)} \times \prod_{z \in T(y) \cap S} N_{\sigma(z),\tau(z)}. \tag{31}$$

We write $\mu[N]^{y,S}$ for $(\mu[N]_1^{y,S}, \ldots, \mu[N]_q^{y,S})$. Note that $\mu[N]^S = \mu[N]^{\rho,S}$ for any antichain S.

The proof of Theorem 1.2 easily extends to show that if $y \in \mathrm{Ins}(S)$ and $z_1, \ldots, z_B$ are the children of y then

$$\mu^{y,S}[N] = \bigotimes_{r=1}^{B} (M\mu^{z_r,S}[N]). \tag{32}$$

We will prove the theorem by recursively analyzing discrepancies via (32). We will prove the result for robust-solvability and indicate the modifications needed for other cases at the end of the proof.

Below, we will write $\mu^{y,S}[k]$ for the measure $\mu^{y,S}[N]$, where $N = M^k$. Note that if S is an antichain and $v \in S$, then the measure $\mu^{y,S}$ is a measure on a single node. We may therefore apply Lemma 2.9 and conclude that for all $\delta > 0$ there exists a k^* such that for $k \geq k^*$, for all antichains S and $y \in S$ it holds that

$$D(\mu^{y,S}[k]) \leq \delta. \tag{33}$$

Since $\mathrm{br}(T)|\lambda_2(M)|^2 < 1$, there exists by Lemma 2.2 an $\epsilon > 0$ and a norm $\| \ \|$ on $v^\perp$ such that for all $b \in v^\perp$, it holds that $\|Mb\| \le \alpha\|b\|$, where $(1+\epsilon)\mathrm{br}(T)\alpha^2 \le 1-\epsilon$. Recall that there is a uniform bound K on the number of children of vertices of T. Let δ be chosen as to satisfy Lemma 2.7 for every $B \le K$, so that if $D(f^r) \le \delta$ for $r = 1, \ldots, B$, and $B \le K$, then $D\big(\bigotimes_{r=1}^B f^r\big) \le (1+\epsilon)\sum_{r=1}^B D(f^r)$.

Lemma 3.5 implies that there exists a sequence of antichains S_n such that

$$\lim_{n\to\infty} \sum_{x \in S_n} [(1+\epsilon)\alpha^2]^{|x|} = 0, \tag{34}$$

and that for all n and $y \in S_n \cup \mathrm{Ins}(S_n)$,

$$\sum_{x \in S_n \cap T(y)} [(1+\epsilon)\alpha^2]^{|x|-|y|} \le 1. \tag{35}$$

We will now show by induction (on $s - |y|$, where $s = \max_{x\in S}|x|$), that for all antichains $S = S_n$ and all $y \in S \cup \mathrm{Ins}(S)$, for $k \ge k^*$,

$$D(\mu^{y,S}[k]) \le \delta \sum_{x \in S \cap T(y)} [(1+\epsilon)\alpha^2]^{|x|-|y|}. \tag{36}$$

The case where $y \in S$ follows from (33). This also proves the base of the induction. For the induction step, it therefore suffices to consider $y \in \mathrm{Ins}(S)$ such that the children of v denoted $z_1, \ldots, z_B$ satisfy the induction hypothesis. By Lemma 2.5 and the induction hypothesis, for all r,

$$D(M\mu^{z_r,S}[k]) \le \alpha^2 D(\mu^{z_r,S}[k]) \le \delta\alpha^2 \sum_{x \in S \cap T(z_r)} [(1+\epsilon)\alpha^2]^{|x|-|z_r|}. \tag{37}$$

The right hand side of (37) is bounded by δ by (35), since $\alpha < 1$. Therefore we may apply Lemma 2.7 with (32) and (37) to obtain

$$\begin{aligned} D(\mu^{y,S}[k]) &\le (1+\epsilon)\sum_{r=1}^B D(M\mu^{z_r,S}) \\ &\le (1+\epsilon)\delta\alpha^2 \sum_{r=1}^B \sum_{x \in S\cap T(z_r)} [(1+\epsilon)\alpha^2]^{|x|-|z_r|} \\ &= \delta \sum_{x \in S \cap T(y)} [(1+\epsilon)\alpha^2]^{|x|-|y|} \end{aligned} \tag{38}$$

proving (36).

Applying (36) for the root ρ and S_n we get by (34), taking $n \to \infty$,

$$\limsup_{n\to\infty} D((\mu_1^{\rho,S_n}[k], \ldots, \mu_q^{\rho,S_n}[k])) \le \delta \lim_{n\to\infty} \sum_{x\in S_n} [(1+\epsilon)\alpha^2]^{|x|} = 0,$$

which implies by Lemma 2.8 that

$$\max_{i,j} \lim_{n\to\infty} D_{TV}(\mu_i^{\rho,S_n}[k], \mu_j^{\rho,S_n}[k]) = 0,$$

as needed.

The proofs for ν-robust-solvability is exactly the same. The proof for erasure-robust solvability requires the following modification.

First, as in the proof of Theorem 1.2, we may find $\epsilon^* < 1$ such that if $\epsilon > \epsilon^*$ and $B \le K$, then for all y which have B children $z_1, \ldots, z_B$ in a cutset S, it holds that $D(\mu^{y,S}[\epsilon]) < \delta$ (with δ as above).

If S is an antichain, let S' denote the set of children of S. Note that S'_n is an antichain for all n. We prove by induction that for all antichains $S = S_n$ and all $y \in S \cup \mathrm{Ins}(S)$, for $\epsilon > \epsilon^*$,

$$D(\mu^{y,S'}[\epsilon]) \le \delta \sum_{x \in S \cap T(y)} [(1+\epsilon)\alpha^2]^{|x|-|y|}. \tag{39}$$

The proof is again by induction on $s - |y|$, where $s = \max_{x \in S} |x|$. The only difference is in that for $y \in S$, we use the estimate $D(\mu^{y,S'}[\epsilon]) < \delta$. The remainder of the proof is the same. ■

References

[1] N. Berger, C. Kenyon, E. Mossel, and Y. Peres. Glauber dynamics on trees and hyperbolic graphs. Submitted. Extended abstract by Kenyon, Mossel and Peres appeared in [11], 2003.

[2] P. M. Bleher, J. Ruiz, and V. A. Zagrebnov. On the purity of the limiting Gibbs state for the Ising model on the Bethe lattice. *J. Statist. Phys.*, 79(1-2):473–482, 1995.

[3] W. S. Evans, C. Kenyon, Y. Peres, and L. J. Schulman. Broadcasting on trees and the Ising model. *Ann. Appl. Probab.*, 10(2):410–433, 2000.

[4] W. S. Evans and L. J. Schulman. Signal propagation and noisy circuits. *IEEE Trans. Inform. Theory*, 45(7):2367–2373, 1999.

[5] J. Felsenstein. *Inferring Phylogenies.* Sinauer, New York, New York, 2003.

[6] H. Furstenberg. Intersections of Cantor sets and transversality of semigroups. In *Problems in analysis (Sympos. Salomon Bochner, Princeton Univ., Princeton, N.J., 1969)*, pages 41–59. Princeton Univ. Press, Princeton, N.J., 1970.

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

[8] B. Hajek and T. Weller. On the maximum tolerable noise for reliable computation by formulas. *IEEE Transactions on Information theory*, 37:388–291, 1991.

[9] Y. Higuchi. Remarks on the limiting Gibbs states on a $(d+1)$-tree. *Publ. Res. Inst. Math. Sci.*, 13(2):335–348, 1977.

[10] D. Ioffe. On the extremality of the disordered state for the Ising model on the Bethe lattice. *Lett. Math. Phys.*, 37(2):137–143, 1996.

[11] C. Kenyon, E. Mossel, and Y. Peres. Glauber dynamics on trees and hyperbolic graphs. In *Proceedings of the Forty-Second Annual Symposium on on Foundations of Computer Science*, pages 568–578, 2001.

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

[13] R. Lyons. The Ising model and percolation on trees and tree-like graphs. *Comm. Math. Phys.*, 125(2):337–353, 1989.

[14] R. Lyons. Random walks and percolation on trees. *Ann. Probab.*, 18(3):931–958, 1990.

[15] J. Martin. Reconstruction thresholds on regular trees. In C. Banderier and C. Krattenthaler, editors, *Discrete Random Walks*, Discrete Math. Theoret. Comput. Sci., pages 191–204. 2003. Availible at `http://dmtcs.loria.fr/proceedings/dmACind.html`.

[16] F. Martinelli, A. Sinclair, and D. Weitz. The ising model on trees: Boundary conditions and mixing time. Submitted to communication in mathematical physics. Extended abstarct appeared in [17], 2003.

[17] F. Martinelli, A. Sinclair, and D. Weitz. The ising model on trees: Boundary conditions and mixing time. In *Proceedings of the Forty Fourth Annual Symposium on Foundations of Computer Science*, pages 628–639, 2003.

[18] E. Mossel. Reconstruction on trees: beating the second eigenvalue. *Ann. Appl. Probab.*, 11(1):285–300, 2001.

[19] E. Mossel. On the impossibility of reconstructing ancestral data and phylogenies. *Jour. Comput. Bio.*, 10(5):669–678, 2003.

[20] E. Mossel. Phase transitions in phylogeny. To appear in Trans. of AMS., 2003.

[21] E. Mossel. Survey: Information flow on trees. To appear, 2003.

[22] E. Mossel and Y. Peres. Information flow on trees. *Ann. Appl. Probab.*, 13(3):817–844, 2003.

[23] E. Mossel and M. Steel. A phase transition for a random cluster model on phylogenetic trees. Submitted to Mathematical Biosciences, 2003.

[24] R. Pemantale and Y. Peres. Recursions on trees and the ising model at critical temperatures. Unpublished manuscript, 1995.

[25] R. Pemantle and J. E. Steif. Robust phase transitions for Heisenberg and other models on general trees. *Ann. Probab.*, 27(2):876–912, 1999.

[26] C. Semple and M. Steel. *Phylogenetics*, volume 22 of *Mathematics and its Applications series.* Oxford University Press, 2003.

[27] E. Sober and M. A. Steel. Testing the hypothesis of common ancestry. *Jour. Theor. Bio.*, 218:395–408, 2002.

[28] F. Spitzer. Markov random fields on an infinite tree. *Ann. Probability*, 3(3):387–398, 1975.

[29] J. von Neumann. Probabilistic logics and the synthesis of reliable organisms from unreliable components. In *Automata studies*, Annals of mathematics studies, no. 34, pages 43–98. Princeton University Press, Princeton, N. J., 1956.

Svante Janson
Department of Mathematics, Uppsala University
P.O. Box 480
SE-751 06 Uppsala, Sweden
svante.janson@math.uu.se
http://www.math.uu.se/~svante

Elchanan Mossel
Department of Statistics
Evans Hall, University of California
Berkeley, CA 94720-1776, U.S.A
mossel@stat.berkeley.edu
http://www.stat.berkeley.edu/~mossel