Big Data Information Reconstruction on an Infinite Tree for a $4 \times 4$-state Asymmetric Model with Community Effects

Wenjian Liu · Ning Ning

Abstract The information reconstruction problem on an infinite tree, is to collect and analyze massive data samples at the $n$th level of the tree to identify whether there is non-vanishing information of the root, as $n$ goes to infinity. This problem has wide applications in various fields such as biology, information theory and statistical physics, and its close connections to cluster learning, data mining and deep learning have been well established in recent years. Although it has been studied in numerous contexts, the existing literatures with rigorous reconstruction thresholds established are very limited. In this paper, motivated by a classical deoxyribonucleic acid (DNA) evolution model, the F81 model, and taking into consideration of the Chargaff’s parity rule by allowing the existence of a guanine-cytosine content bias, we study the noise channel in terms of a $4 \times 4$-state asymmetric probability transition matrix with community effects, for four nucleobases of DNA. The corresponding information reconstruction problem in molecular phylogenetics is explored, by means of refined analyses of moment recursion, in-depth concentration estimates, and thorough investigations on an asymptotic 4-dimensional nonlinear second order dynamical system. We rigorously show that the reconstruction bound is not tight when the sum of the base frequencies of adenine and thymine falls in the interval $(0, 1/2 - \sqrt{3}/6) \cup (1/2 + \sqrt{3}/6, 1)$, which is the first rigorous result on asymmetric noisy channels with community effects.

Keywords Kesten-Stigum reconstruction bound · Markov random fields on trees · Distributional recursion · Nonlinear dynamical system

Mathematics Subject Classification (2010) 60K35 · 82B26 · 82B20

1 Introduction

1.1 Big data and the reconstruction problem

“Big Data”, like the name implies, refers to massive data sets. Finding information about the source of massive data as time evolves, is one of the toughest big data challenges. In this paper, we consider the following broadcasting process that can be considered as signals transmitting on an infinite communication tree network, as a model for propagation of a genetic property, or as a tree-indexed Markov chain. It has two building blocks: the first is an irreducible aperiodic Markov chain on a finite characters set $\mathcal{C}$; the second is a $d$-ary tree, which is a rooted tree with every vertex having exactly $d$ offspring, denoted as $T = (\mathbb{V}, \mathbb{E}, \rho)$ with nodes $\mathbb{V}$, edges $\mathbb{E}$, and root $\rho \in \mathbb{V}$. A configuration on $T$ is an element of $\mathcal{C}^\mathbb{V}$, which is an assignment of a state in $\mathcal{C}$ to each vertex. The state of the root $\rho$, denoted by $\sigma_\rho$, is chosen according to an initial distribution $\pi$ on $\mathcal{C}$. This symbol is then propagated on the tree according to a probability transition matrix $M = (M_{ij})_{i,j \in \mathcal{C}}$, which functions as a noisy communication channel.
on each edge. In other words, for each vertex $v$ having $u$ as its parent, the spin at $v$ is defined according to the probabilities

$$P(\sigma_v = j | \sigma_u = i) = M_{ij}, \quad i, j \in \mathcal{C}.$$ 

The problem of reconstruction is to analyze whether there exists a non-vanishing information on the letter transmitted by the root, given all the symbols received at the vertices of the $n$th generation, as $n$ goes to infinity. Denote $\sigma(n)$ as the spins at distance $n$ from the root and $\sigma'(n)$ as $\sigma(n)$ conditioned on $\sigma_r = i$. In this paper, we use the following definition to mathematical formulate reconstructibility and we remark that more equivalent formulations can be seen in Mossel [2001] and Mossel [2004b].

**Definition 1** We say that a model is **reconstructible** on an infinite tree $T$, if for some $i, j \in \mathcal{C}$

$$\limsup_{n \to \infty} d_{TV}(\sigma(n), \sigma'(n)) > 0,$$

where $d_{TV}$ is the total variation distance. When the limsup is 0, we say that the model is **non-reconstructible** on $T$.

### 1.2 Existing results

The reconstruction problem arises naturally in statistical physics, where the reconstruction threshold corresponds to the threshold for extremality of the infinite-volume Gibbs measure with free boundary conditions (see Georgii [2011]). The reconstruction bound is known to have a crucial determination effect on the efficiency of the Glauber dynamics on trees and random graphs (see Berger et al. [2005], Martinelli et al. [2007], Tetali et al. [2012]). It is especially worth mentioning that one of the classical techniques in tackling the reconstruction problem was initiated in Chayes et al. [1986] in the context of spin-glasses. The reconstructability is believed to play an important role in a variety of other contexts including, but not limited to, the following: phylogenetic reconstruction in evolutionary biology (see Mossel [2004a], Daskalakis et al. [2006], Roch [2006]), communication theory in the study of noisy computation (see Evans et al. [2000]), analogous investigations in the realm of network tomography (see Bhamidi et al. [2010]), reconstructability and distinguishability regarding the clustering problem in the stochastic block model (see Mossel et al. [2013, 2014], Neeman and Netrapalli [2014]), analogous phase transition analysis using the cavity method in Bayesian inference (see Ricci-Tersenghi et al. [2018]), classification and labeling in a semi-supervised learning setting using deep algorithms on a closely related and well defined family of hierarchical generative models (see Mossel [2016]). More applications can be seen in Section 1.4 of Sly [2011] and Section 1.3 of Liu et al. [2018], and the references therein.

Clearly, for any channel, the reconstruction problem is closely related to $\lambda$, the second largest eigenvalue in absolute value of the transition probability matrix $M$. Kesten and Stigum [1966, 1967] showed that the reconstruction problem is solvable if $d\lambda^2 > 1$, which is known as the Kesten-Stigum bound. However, when it comes to the case of larger noise, that is $d\lambda^2 < 1$, retrieving root information from the symbols received at the $n$th generation, is a challenging problem whose solvability highly depends on the channel. One important case is the binary channel, that is, the Ising model in statistical mechanics terminology, with the transition matrix

$$M = \frac{1}{2} \begin{pmatrix} 1 + \theta & 1 - \theta \\ 1 - \theta & 1 + \theta \end{pmatrix} + \frac{\Delta}{2} \begin{pmatrix} -1 & 1 \\ -1 & 1 \end{pmatrix},$$

where $|\theta| + |\Delta| \leq 1$ and $\Delta$ is used to describe the deviation of $M$ from the symmetric channel. For the binary symmetric model, i.e. $\Delta = 0$, it was shown in Bleher et al. [1995] that the reconstruction problem is solvable if and only if $d\lambda^2 > 1$. For the binary asymmetric model, i.e. $\Delta \neq 0$, Mossel [2001, 2004b] showed that the Kesten-Stigum bound is not the bound for reconstruction with sufficiently large asymmetry. The breakthrough result in Borgs et al. [2006] established the first tightness result of Keston-Stigum reconstruction bound in roughly a decade for a binary asymmetric channel on the $d$-ary tree, provided that the asymmetry is sufficiently small. Liu and Ning [2018] gave a complete answer to the question of how small it needs to be to establish the tightness of the reconstruction threshold.

Another important case is the $q$-state symmetric channel, that is, the Potts model in statistical mechanics terminology, with the transition matrix

$$M = \begin{pmatrix} p_0 & p_1 & \cdots & p_1 \\ p_1 & p_0 & \cdots & p_1 \\ \vdots & \vdots & \ddots & \vdots \\ p_1 & p_1 & \cdots & p_0 \end{pmatrix}_{q \times q}.$$
Mossel [2001, 2004b] showed that the Kesten-Stigum bound is not sharp in the \( q \)-state symmetric channel with sufficiently many characters. Sly [2011] proved the first exact reconstruction threshold in a nonbinary model, established the Kesten-Stigum bound for the 3-state Potts model on regular trees of large degree, and further showed that the Kesten-Stigum bound is not tight when \( q \geq 5 \), which confirms much of the picture conjectured earlier by Mézard and Montanari [2006]. Inspired by a popular Markov model of DNA sequence evolution, the K80 model (Kimura [1980]), to distinguish between transitions and transversions, Liu et al. [2018] analyzed the case that the transition matrix has two mutation classes and \( q \) states in each class, with the transition matrix

\[
M = \begin{pmatrix}
\begin{array}{cccc}
p_0 & p_1 & \cdots & p_1 \\
p_1 & p_0 & \cdots & p_2 \\
\vdots & \vdots & \ddots & \vdots \\
p_1 & p_1 & \cdots & p_0 \\
p_2 & \cdots & \cdots & p_2 \\
p_2 & \cdots & \cdots & p_2 \\
p_2 & \cdots & \cdots & p_2 \\
\end{array}
\end{pmatrix}
\]

Liu et al. [2018] showed that when \( q \geq 4 \), for every \( d \) the Kesten-Stigum bound is not tight, i.e. the reconstruction is solvable for some \( \lambda \) even if \( d \lambda^2 < 1 \).

1.3 Motivation and the Main Theorem

Deoxyribonucleic acid (DNA) is a molecule composed of two chains coiling around each other to form a double helix, which carries the genetic instructions used in the growth, development, functioning and reproduction of all known living organisms and many viruses. These two DNA strands are composed of simpler monomeric units called nucleotides, each of which is further composed of one of four nitrogen-containing nucleobases (guanine [G], cytosine [C], adenine [A] or thymine [T]). A number of different Markov models of DNA sequence evolution were proposed and frequently used in molecular phylogenetics, a branch of phylogeny to analyze the genetic or hereditary molecular differences in order to gain information on an organism’s evolutionary relationships. Specifically, they are used during the calculation of likelihood of a tree and used to estimate the evolutionary distance from the observed differences between sequences. For detailed descriptions of phylogenetic reconstruction and Markov models of DNA evolution, we refer to Felsenstein [2004].

In this paper, the objective model taken into account is based on the F81 model (Felsenstein [1981]), a classical DNA evolution model, whose base frequencies are allowed to vary from 0.25 that is enforced in several other classical DNA evolution models. We allow a guanine-cytosine (G-C) content bias, where the G-C content refers to the percentage of nitrogenous bases on a DNA molecule that are either guanine or cytosine. We further follow the Chargaff’s parity rule, that is, pairing nucleotides do have the same frequency on a single DNA strand, with G and C on the one hand while A and T on the other hand. Therefore, the four base frequencies can be expressed as a function of \( \theta \in (0, 1) \): \( \pi_A = \pi_T = \frac{\theta}{2} \) and \( \pi_G = \pi_C = \frac{1-\theta}{2} \). In other word, we consider the rate matrix of \( \{A, T, G, C\} \), or the configuration set \( \{1, 2, 3, 4\} \), of the form

\[
Q = \frac{1}{2} \begin{pmatrix}
-2 + \theta & \theta & 1 - \theta & 1 - \theta \\
\theta & -2 + \theta & 1 - \theta & 1 - \theta \\
\theta & \theta & -1 - \theta & 1 - \theta \\
\theta & \theta & 1 - \theta & -1 - \theta
\end{pmatrix},
\]

Set \( \lambda = \exp\left(-\frac{\nu}{2 + \theta(1-\theta)}\right) \), where \( \nu \) is defined as the branch length measured in the expected number of substitutions per site, and then the corresponding probability transition matrix can be written as

\[
P = \frac{1}{2} \begin{pmatrix}
2 \lambda + \theta(1-\lambda) & \theta(1-\lambda) & (1-\theta)(1-\lambda) & (1-\theta)(1-\lambda) \\
\theta(1-\lambda) & 2 \lambda + \theta(1-\lambda) & (1-\theta)(1-\lambda) & (1-\theta)(1-\lambda) \\
\theta(1-\lambda) & \theta(1-\lambda) & 2 \lambda + (1-\theta)(1-\lambda) & (1-\theta)(1-\lambda) \\
\theta(1-\lambda) & \theta(1-\lambda) & (1-\theta)(1-\lambda) & 2 \lambda + (1-\theta)(1-\lambda)
\end{pmatrix}.\]
It is clear that the second eigenvalue of the channel $P$ is $\lambda$. Because non-reconstruction always holds for $d\lambda^2 > 1$, without loss of generality, it would be convenient to presume $d\lambda^2 \leq 1$ in the following context.

We focus on the $4 \times 4$-state probability transition matrix allowing $\theta \neq 1/2$. Note that, when $\theta = 1/2$, the model degenerates to the symmetric $q$-state Potts model with $q = 4$, whose reconstructibility is an open problem, while Sly [2011] gave perfect answers on the reconstructibility with $q = 3$ and $q \geq 5$. The asymmetry and community effects can be seen in the following: (1) One state stays unchanged with probability $\lambda + \frac{1}{2}(1 - \lambda)$ if that state is in $[1, 2]$, but with probability $\lambda + \frac{1}{2}(1 - \lambda)$ if that state is in $[3, 4]$; (2) For the community $[1, 2]$ the probability to transfer from one state to the other is $\theta(1 - \lambda)$, while for the community $[3, 4]$ the probability to transfer from one state to the other is $\frac{1}{2}(1 - \lambda)$ for the case in $[3, 4]$; (3) The transition probabilities from $[1, 2]$ to $[3, 4]$ are given by $\frac{1}{2}(\lambda - 1)$, while the transition probabilities from $[3, 4]$ to $[1, 2]$ are given by $\frac{\theta(1 - \lambda)}{2}$.

The main result is given below and its rigorous proof is provided in Section 5. **Main Theorem** When $\theta \in \left(0, \frac{3 - \sqrt{3}}{6}\right) \cup \left(\frac{3 + \sqrt{3}}{6}, 1\right)$, for every $d$ the Kesten-Stigum bound is not sharp.

1.4 Proof sketch

The ideas and techniques used to prove the Main Theorem can be seen in the following. One standard to classify reconstruction and non-reconstruction is to analyze the difference, between the probability of giving a correct guess of the root given the spins $\sigma(n)$ at distance $n$ from the root, and the probability of guessing the root according to the stationary initial distribution. Unlike the symmetric models, the model under investigation has two distinct base frequencies $\theta/2$ and $(1 - \theta/2)$, therefore we need to analyze two different quantities $x_{n, \theta}$ (probability of giving a correct guess of the root in $[1, 2]$ minus $\theta/2$) and $x_{n, 1 - \theta}$ (probability of giving a correct guess of the root in $[3, 4]$ minus $(1 - \theta)/2$). Since non-reconstruction means that the mutual information between the root and the spins at distance $n$ goes to 0 as $n$ tends to infinity, here it can be established that non-reconstruction is equivalent to

$$\lim_{n \to \infty} x_{n, \theta} = \lim_{n \to \infty} x_{n, 1 - \theta} = 0.$$ 

Furthermore, it is necessary to consider the quantity $y_{n, \theta}$ (resp. $y_{n, 1 - \theta}$) which corresponds to the probability of giving a wrong guess but right group $[1, 2]$ (resp. $[3, 4]$), and the quantity $z_{n, 1 - \theta}$ (resp. $z_{n, \theta}$) which corresponds to the probability of giving a wrong guess and even wrong group $[3, 4]$ (resp. $[1, 2]$). Together with their second moment forms, we have to take care of much more objective quantities than Sly [2011], Liu and Ning [2018] and Liu et al. [2018].

Through carefully analyzing the relation between the $n$th and the $(n + 1)$th level, and fully taking advantage of the Markov random field property and the symmetries incorporated in the model and the tree structure, we establish the distributional recursion and moment recursion. We show that the interactions between spins become very weak if they are sufficiently far away from each other, and prove that $x_{n, \theta}$ is small and the decrease from $x_{n, \theta}$ to $x_{n + 1, \theta}$ is never too large. Consequently, we are able to derive the concentration estimates and achieve the following asymptotic 4-dimensional second order nonlinear dynamical system:

$$\begin{align*}
x_{n+1, \theta} &\approx d \lambda^2 x_{n, \theta} + \frac{d(d-1)}{2} \lambda^4 \left[ -6 + \frac{2(1-\theta)}{\theta} + 2\theta \right] x_{n, \theta}^2 + \left[ -\frac{4\theta}{1-\theta} - 16 \right] (z_{n, 1-\theta})^2 - 4(\theta - 1)(\theta - 16) x_{n, \theta} (z_{n, 1-\theta}) + \theta \left( x_{n, 1-\theta}^2 + y_{n, 1-\theta}^2 \right), \\
-z_{n+1, 1-\theta} &\approx d \lambda^2 (-z_{n, 1-\theta}) + \frac{d(d-1)}{2} \lambda^4 \left[ \frac{2(1-\theta)^2}{\theta} x_{n, \theta}^2 + \left( -8 - \frac{4\theta}{1-\theta} + 4(1-\theta) \right) (z_{n, 1-\theta})^2 - 4 \frac{(1-\theta)^2}{\theta} x_{n, \theta} (z_{n, 1-\theta}) + \theta \left( x_{n, 1-\theta}^2 + y_{n, 1-\theta}^2 \right). \right]
\end{align*}$$

Methods in the previous works are to analyze the stability of fixed points of the $x$-dynamics alone, by showing that if the quadratic terms are strictly positive, given the fact that $x_{n, \theta}$ is nonnegative, it will not goes to zero as $n$ goes to infinity. However, a closer look at the above quadratic terms in the $x$-dynamics reveals that it is hard to obtain the desired results. Therefore, we turn to focus on analyzing the $z$-dynamics while fully take into consideration of the coupled relationship between $x_{n, \theta}$ and $z_{n, 1-\theta}$, and conclude that when $\theta \in \left[0, \frac{3 - \sqrt{3}}{6}\right] \cup \left(\frac{3 + \sqrt{3}}{6}, 1\right)$, even if $d \lambda^2 < 1$ for some $\lambda$, $z_{n, 1-\theta}$ does not converge to 0 and then $x_{n, \theta} (z_{n, 1-\theta} \geq 0)$ does not converge to 0. That is, there is reconstruction beyond the Kesten-Stigum bound.
1.5 Organization of the paper

The rest of the paper is organized as follows. In Section 2, we give detailed definitions and interpretations, conduct preliminary analyses, and then provide an equivalent condition for non-reconstruction. In Section 3, we develop the second order recursive relations associated with \( x_{n+1,\theta} \) and \( z_{n+1,1-\theta} \). In-depth concentration results are established in Section 4. A complete proof of the Main Theorem is given in Section 5.

2 Preparation

2.1 Notations

Recall that the stationary distribution \( \pi = (\pi_1, \pi_2, \pi_3, \pi_4) \) of \( P \) is given by

\[
\pi_1 = \pi_2 = \frac{\theta}{2} \quad \text{and} \quad \pi_3 = \pi_4 = \frac{1-\theta}{2}.
\]

Let \( u_1, \ldots, u_d \) be the children of root \( \rho \) and \( T_v \) be the subtree of descendants of \( v \in T \). Denote the \( n \)th level of the tree by \( L_n = \{ v \in V : d(\rho, v) = n \} \) with \( d(\cdot, \cdot) \) being the graph distance on \( T \). Let \( \sigma(n) \) and \( \sigma_{j}(n) \) denote the spins on \( L_n \) and \( L_n \cap \bigcup_{u_j} \) respectively. For a configuration \( A \) on the spins of \( L_n \), define the posterior function by

\[
f_n(i, A) = P(\sigma_\rho = i \mid \sigma(n) = A), \quad i = 1, 2, 3, 4.
\]

By the recursive nature of the tree, for a configuration \( A \) on spins in \( L(n+1) \cap \bigcup_{u_j} \), we can give an equivalent form of the previous one

\[
f_n(i, A) = P(\sigma_{u_j} = i \mid \sigma_{j}(n+1) = A), \quad i = 1, 2, 3, 4, \quad j = 1, \ldots, d.
\]

Define \( X_i \) as the posterior probability that \( \sigma_\rho = i \) given the random configuration \( \sigma(n) \) on the spins in \( L_n \), that is

\[
X_i = X_i(n) = f_n(i, \sigma(n)), \quad i = 1, 2, 3, 4.
\]

Then one has

\[
X_1(n) + X_2(n) + X_3(n) + X_4(n) = 1
\]

and

\[
E(X_1) = E(X_2) = \frac{\theta}{2}, \quad E(X_3) = E(X_4) = \frac{1-\theta}{2}.
\]

Define \( Y_{i,j} \) as the posterior probability that \( \sigma_{u_j} = i \) given the random configuration \( \sigma_{j}(n+1) \) on spins in \( L(n+1) \cap \bigcup_{u_j} \), that is

\[
Y_{i,j} = Y_{i,j}(n) = f_n(i, \sigma_{j}(n+1)), \quad i = 1, 2, 3, 4, \quad j = 1, \ldots, d,
\]

where the random variables \( \{Y_{i,j}\} \) are independent and identically distributed apparently. At last, we define the following moment variables, which will serve as the main quantities under investigation,

\[
x_{n,\theta} = E \left( f_n(1, \sigma^1(n)) - \frac{\theta}{2} \right), \quad y_{n,\theta} = E \left( f_n(2, \sigma^1(n)) - \frac{\theta}{2} \right), \quad z_{n,\theta} = E \left( f_n(1, \sigma^3(n)) - \frac{\theta}{2} \right),
\]

\[
u_{n,\theta} = E \left( f_n(1, \sigma^1(n)) - \frac{\theta}{2} \right)^2, \quad v_{n,\theta} = E \left( f_n(2, \sigma^1(n)) - \frac{\theta}{2} \right)^2, \quad w_{n,\theta} = E \left( f_n(1, \sigma^3(n)) - \frac{\theta}{2} \right)^2,
\]

and

\[
x_{n,1-\theta} = E \left( f_n(3, \sigma^3(n)) - \frac{1-\theta}{2} \right), \quad y_{n,1-\theta} = E \left( f_n(4, \sigma^3(n)) - \frac{1-\theta}{2} \right), \quad z_{n,1-\theta} = E \left( f_n(3, \sigma^1(n)) - \frac{1-\theta}{2} \right),
\]

\[
u_{n,1-\theta} = E \left( f_n(3, \sigma^3(n)) - \frac{1-\theta}{2} \right)^2, \quad v_{n,1-\theta} = E \left( f_n(4, \sigma^3(n)) - \frac{1-\theta}{2} \right)^2, \quad w_{n,1-\theta} = E \left( f_n(3, \sigma^1(n)) - \frac{1-\theta}{2} \right)^2.
\]
2.2 Preliminary analyses

Let us firstly establish some important lemmas which will be used frequently in the sequel.

**Lemma 1** For any \( n \in \mathbb{N} \cup \{0\} \), we have

(a) \( x_{n,\theta} = \frac{\theta}{2} \mathbb{E} \left( X_1(n) - \frac{\theta}{2} \right)^2 = u_{n,\theta} + v_{n,\theta} + 2 \frac{1 - \theta}{\theta} w_{n,\theta} \geq 0 \).

(b) \( z_{n,\theta} = -\frac{X_{n,\theta} + y_{n,\theta}}{2} \leq 0 \), \( x_{n,\theta} \geq z_{n,1-\theta} \); \( z_{n,\theta} = -\frac{X_{n,\theta} + y_{n,\theta}}{2} \leq 0 \), \( x_{n,1-\theta} \geq z_{n,\theta} \).

(c) \( \theta z_{n,1-\theta} = (1 - \theta) z_{n,\theta} \).

**Proof.** (a) By the law of total probability and Bayes’ theorem, we have

\[
\mathbb{E} f_n(1, \sigma^1(n)) = \sum_A f_n(1, A) \mathbb{P}(\sigma(n) = A \mid \sigma_\theta = 1)
= 2 \frac{\theta}{2} \sum_A \mathbb{P}(\sigma_\theta = 1 \mid \sigma(n) = A) \mathbb{P}(\sigma(n) = A) f_n(1, A)
= 2 \frac{\theta}{2} \sum_A f_n^2(1, A) \mathbb{P}(\sigma(n) = A)
= 2 \frac{\theta}{2} \mathbb{E}(X_1^2).
\]

Then it follows from the fact \( \mathbb{E}(X_1) = \frac{\theta}{2} \) that

\[
x_{n,\theta} = 2 \frac{\theta}{2} \left( \mathbb{E}(X_1^2) - \left( \frac{\theta}{2} \right)^2 \right) = 2 \frac{\theta}{2} \left( X_1 - \frac{\theta}{2} \right)^2.
\]

Furthermore, by the law of total expectation, we have

\[
x_{n,\theta} = 2 \frac{\theta}{2} \mathbb{E} \left( X_1 - \frac{\theta}{2} \right)^2
= 2 \frac{\theta}{2} \sum_{i=1}^4 \mathbb{E} \left( \left( X_1 - \frac{\theta}{2} \right)^2 \right) \mathbb{P}(\sigma_\theta = i)
= 2 \frac{\theta}{2} \left[ \mathbb{P}(\sigma_\theta = 1) \mathbb{E} \left( f_n(1, \sigma^1(n)) - \frac{\theta}{2} \right)^2 + \mathbb{P}(\sigma_\theta = 2) \mathbb{E} \left( f_n(1, \sigma^2(n)) - \frac{\theta}{2} \right)^2 \right.
+ \left. \mathbb{P}(\sigma_\theta = 3) \mathbb{E} \left( f_n(1, \sigma^3(n)) - \frac{\theta}{2} \right)^2 + \mathbb{P}(\sigma_\theta = 4) \mathbb{E} \left( f_n(1, \sigma^4(n)) - \frac{\theta}{2} \right)^2 \right]
= u_{n,\theta} + v_{n,\theta} + 2 \frac{1 - \theta}{\theta} w_{n,\theta}.
\]

(b) Similarly, we have

\[
y_{n,\theta} + \frac{\theta}{2} = \sum_A f_n(2, A) \mathbb{P}(\sigma(n) = A \mid \sigma_\theta = 1)
= 2 \frac{\theta}{2} \sum_A f_n(1, A) f_n(2, A) \mathbb{P}(\sigma(n) = A)
= 2 \frac{\theta}{2} \mathbb{E}(X_1 X_2),
\]

and then

\[
y_{n,\theta} = 2 \frac{\theta}{2} \mathbb{E} \left( X_1 - \frac{\theta}{2} \right) \left( X_2 - \frac{\theta}{2} \right),
\]

(2.1)

It follows from the Cauchy-Schwarz inequality that

\[
\left[ \mathbb{E} \left( X_1 - \frac{\theta}{2} \right) \left( X_2 - \frac{\theta}{2} \right) \right]^2 \leq \mathbb{E} \left( X_1 - \frac{\theta}{2} \right)^2 \mathbb{E} \left( X_2 - \frac{\theta}{2} \right)^2,
\]
which implies
\[
\left( \frac{\theta}{2} y_{n;\theta} \right)^2 \leq \left( \frac{\theta}{2} x_{n;\theta} \right)^2, \quad \text{i.e.} \quad -x_{n;\theta} \leq y_{n;\theta} \leq x_{n;\theta}.
\] (2.2)

By the definitions of \(x_{n;\theta}, y_{n;\theta}\) and \(z_{n;1-\theta}\), we know that \(z_{n;1-\theta} = -\frac{x_{n;\theta} + y_{n;\theta}}{2}\), and thus (2.2) implies
\[
x_{n;\theta} + z_{n;1-\theta} = x_{n;\theta} - \frac{x_{n;\theta} + y_{n;\theta}}{2} = \frac{x_{n;\theta} - y_{n;\theta}}{2} \geq 0 \quad \text{and} \quad z_{n;1-\theta} \leq 0.
\]

An analogous proof of
\[
z_{n;\theta} = -\frac{x_{n;1-\theta} + y_{n;1-\theta}}{2} \leq 0 \quad \text{and} \quad x_{n;1-\theta} + z_{n;\theta} \geq 0
\]
can be easily carried out.

(c) Similarly, we have
\[
z_{n;1-\theta} = E\left( f_n(3, \sigma^1(n)) - \frac{1-\theta}{2} \right) = \frac{2}{\theta} E\left( X_1 X_3 \right) - \frac{1-\theta}{2} = \frac{2}{\theta} E\left( X_1 - \frac{\theta}{2} \right) \left( X_3 - \frac{1-\theta}{2} \right)
\] (2.3)
and then
\[
\theta z_{n;1-\theta} = 2E\left( X_1 - \frac{\theta}{2} \right) \left( X_3 - \frac{1-\theta}{2} \right) = (1-\theta)z_{n;\theta}.
\]

**Lemma 2** For any \(n \in \mathbb{N} \cup \{0\}\), we have

(a) \(E\left( f_n(1, \sigma^1(n)) - \frac{\theta}{2} \right) f_n(2, \sigma^1(n)) - \frac{\theta}{2} = \frac{\theta}{2} y_{n;\theta} + \left( v_{n;\theta} - \frac{\theta}{2} x_{n;\theta} \right)\).

(b) \(E\left( f_n(1, \sigma^1(n)) - \frac{\theta}{2} \right) f_n(3, \sigma^1(n)) - \frac{1-\theta}{2} = \frac{\theta}{2} z_{n;1-\theta} - \frac{1}{2} \left( u_{n;\theta} - \frac{\theta}{2} x_{n;\theta} \right) - \frac{1}{2} \left( v_{n;\theta} - \frac{\theta}{2} x_{n;\theta} \right)\).

(c) \(E\left( f_n(2, \sigma^1(n)) - \frac{\theta}{2} \right) f_n(3, \sigma^1(n)) - \frac{1-\theta}{2} = \frac{\theta}{2} z_{n;1-\theta} - \frac{1}{2} \left( u_{n;\theta} - \frac{\theta}{2} x_{n;\theta} \right)\).

(d) \(E\left( f_n(3, \sigma^1(n)) - \frac{1-\theta}{2} \right) f_n(4, \sigma^1(n)) - \frac{1-\theta}{2} = \frac{1-\theta}{2} \left( y_{n;1-\theta} + \frac{\theta}{2} u_{n;\theta} \right) + \frac{1}{2} \left( v_{n;\theta} - \frac{\theta}{2} x_{n;\theta} \right) - \frac{1}{2} \left( u_{n;1-\theta} - \frac{1-\theta}{2} z_{n;1-\theta} \right)\).

(e) \(E\left( f_n(1, \sigma^3(n)) - \frac{\theta}{2} \right) f_n(2, \sigma^3(n)) - \frac{\theta}{2} = \frac{\theta}{2} y_{n;\theta} - \frac{\theta}{1-\theta} \left( v_{n;\theta} - \frac{\theta}{2} x_{n;\theta} \right)\).

**Proof.** (a) By the law of total probability, one has
\[
E f_n(1, \sigma^1(n)) f_n(2, \sigma^1(n)) = \sum_A P(\sigma_\rho = 1 \mid \sigma(n) = A) P(\sigma_\rho = 2 \mid \sigma(n) = A) P(\sigma(n) = A \mid \sigma_\rho = 1)
\]
\[
= \sum_A \left[ P(\sigma_\rho = 2 \mid \sigma(n) = A) \right]^2 P(\sigma(n) = A \mid \sigma_\rho = 1)
\]
\[
= E \left( f_n(2, \sigma^1(n)) \right)^2,
\]
and therefore,
\[
E \left( f_n(1, \sigma^1(n)) - \frac{\theta}{2} \right) \left( f_n(2, \sigma^1(n)) - \frac{\theta}{2} \right) = E \left( f_n(2, \sigma^1(n)) - \frac{\theta}{2} \right)^2 + \frac{\theta (y_{n;\theta} - x_{n;\theta})}{2}
\]
\[
= v_{n;\theta} + \frac{1}{2} (y_{n;\theta} - x_{n;\theta})
\]
\[
= \frac{\theta}{2} y_{n;\theta} + \left( v_{n;\theta} - \frac{\theta}{2} x_{n;\theta} \right).
\]

(b) Similarly, we can achieve that
\[
E \left( f_n(1, \sigma^1(n)) - \frac{\theta}{2} \right) \left( f_n(3, \sigma^1(n)) - \frac{1-\theta}{2} \right) = \frac{\theta}{2} \left( z_{n;1-\theta} - \frac{1}{2} \left( u_{n;\theta} - \frac{\theta}{2} x_{n;\theta} \right) \right) - \frac{1}{2} \left( v_{n;\theta} - \frac{\theta}{2} x_{n;\theta} \right)
\]
\[
= \frac{\theta}{2} (x_{n;\theta} + z_{n;1-\theta}) - \frac{1}{2} \left( u_{n;\theta} - \frac{\theta}{2} x_{n;\theta} \right) - \frac{1}{2} \left( v_{n;\theta} - \frac{\theta}{2} x_{n;\theta} \right).
\]
(c) Note that
\[
\left( f_n(2, \sigma^1(n)) - \frac{\theta}{2} \right) + \left( f_n(3, \sigma^1(n)) - \frac{1-\theta}{2} \right) = - \left( f_n(1, \sigma^1(n)) - \frac{\theta}{2} \right) - \left( f_n(4, \sigma^1(n)) - \frac{1-\theta}{2} \right),
\]
and by taking square of both sides and then expectation, one has
\[
\mathbb{E} \left( f_n(2, \sigma^1(n)) - \frac{\theta}{2} \right) \left( f_n(3, \sigma^1(n)) - \frac{1-\theta}{2} \right) = \mathbb{E} \left( f_n(1, \sigma^1(n)) - \frac{\theta}{2} \right) \mathbb{E} \left( f_n(4, \sigma^1(n)) - \frac{1-\theta}{2} \right)
\]
\[
= \frac{u_{n,\theta} - v_{n,\theta}}{2} + \frac{1}{2} \mathbb{E} \left( f_n(3, \sigma^1(n)) - \frac{1-\theta}{2} \right) \left( f_n(2, \sigma^1(n)) - \frac{1-\theta}{2} \right) - \mathbb{E} \left( f_n(3, \sigma^1(n)) - \frac{1-\theta}{2} \right)^2
\]
\[
= \frac{\theta z_{n,1-\theta}}{2} + \frac{1}{2} \left( u_{n,\theta} - \frac{\theta x_{n,\theta}}{2} \right) + \frac{1}{2} \left( v_{n,\theta} - \frac{\theta x_{n,\theta}}{2} \right) - \frac{\theta z_{n,1-\theta}}{2} - \frac{1}{2} \left( u_{n,\theta} - \frac{\theta x_{n,\theta}}{2} \right)
\]
\[
= \frac{(1-\theta) y_{n,1-\theta}}{2} + \frac{1}{2} \left( u_{n,\theta} - \frac{\theta x_{n,\theta}}{2} \right) + \frac{1}{2} \left( v_{n,\theta} - \frac{\theta x_{n,\theta}}{2} \right) - \frac{1}{2} \left( u_{n,\theta} - \frac{\theta x_{n,\theta}}{2} \right).
\]

(d) Similarly, we have
\[
\mathbb{E} \left( f_n(3, \sigma^1(n)) - \frac{1-\theta}{2} \right) \left( f_n(4, \sigma^1(n)) - \frac{1-\theta}{2} \right) = \theta \mathbb{E} \left( f_n(1, \sigma^1(n)) - \frac{\theta}{2} \right) \left( f_n(2, \sigma^1(n)) - \frac{\theta}{2} \right) + (1-\theta) \mathbb{E} \left( f_n(3, \sigma^1(n)) - \frac{1-\theta}{2} \right) \left( f_n(2, \sigma^1(n)) - \frac{1-\theta}{2} \right)
\]
\[
= \frac{\theta^2}{2} y_{n,\theta} + \theta \left( v_{n,\theta} - \frac{\theta x_{n,\theta}}{2} \right) + (1-\theta) \mathbb{E} \left( f_n(1, \sigma^1(n)) - \frac{\theta}{2} \right) \left( f_n(2, \sigma^1(n)) - \frac{\theta}{2} \right),
\]
which implies
\[
\mathbb{E} \left( f_n(1, \sigma^1(n)) - \frac{\theta}{2} \right) \left( f_n(2, \sigma^1(n)) - \frac{\theta}{2} \right) = \frac{\theta}{2} y_{n,\theta} - \frac{\theta}{1-\theta} \left( v_{n,\theta} - \frac{\theta x_{n,\theta}}{2} \right).
\]

Recall that \( Y_{ij}(n) = f_n\left[i, \sigma^1(n) + 1\right] \) is simply the posterior probability that \( \sigma_{u_i} = i \) given the random configuration \( \sigma^1(n+1) \) on spins in \( L(n+1) \cap T_{u_i} \). By the symmetry of the model, the random vectors \( (Y_{ij})_{i=1}^4 \) are independent. The central moments of \( Y_{ij} \) would play a key role in further analysis, and therefore it is necessary to figure them out in the first place.

**Lemma 3** For each \( 1 \leq j \leq d \), we have

(a) \( \mathbb{E} \left( Y_{1j}(n) - \frac{\theta}{2} \right) = \lambda x_{n,\theta} \).

(b) \( \mathbb{E} \left( Y_{2j}(n) - \frac{\theta}{2} \right) = \lambda y_{n,\theta} \).

(c) \( \mathbb{E} \left( Y_{ij}(n) - \frac{1-\theta}{2} \right) = \lambda z_{n,1-\theta} \), \( i = 3, 4 \).

(d) \( \mathbb{E} \left( Y_{1j}(n) - \frac{\theta}{2} \right)^2 = \frac{\theta}{2} x_{n,\theta} + \lambda \left( u_{n,\theta} - \frac{\theta}{2} x_{n,\theta} \right) \).

(e) \( \mathbb{E} \left( Y_{2j}(n) - \frac{\theta}{2} \right)^2 = \frac{\theta}{2} x_{n,\theta} + \lambda \left( v_{n,\theta} - \frac{\theta}{2} x_{n,\theta} \right) \).
implies and the rest can be shown analogously.

\[ E \left( Y_{ij}(n) - \frac{1-\theta}{2} \right)^2 = \frac{1-\theta}{2} x_{n;1-\theta} + \lambda \left( u_{n;1-\theta} - \frac{1-\theta}{2} x_{n;1-\theta} \right), \quad i = 3, 4. \]

\[ E \left( Y_{ij}(n) - \frac{\theta}{2} \right)^2 \leq \frac{\theta}{2} y_{n;1-\theta} + \lambda \left( u_{n;1-\theta} - \frac{\theta}{2} x_{n;1-\theta} \right). \]

\[ E \left( Y_{ij}(n) - \frac{1-\theta}{2} \right) = \frac{1-\theta}{2} y_{n;1-\theta} + \lambda \left( u_{n;1-\theta} - \frac{1-\theta}{2} x_{n;1-\theta} \right) - \lambda \left( u_{n;1-\theta} - \frac{1-\theta}{2} x_{n;1-\theta} \right). \]

**Proof.** In the following, we only prove some results in Lemma 3 and the rest can be shown analogously.

(a) \[ E \left( Y_{ij}(n) - \frac{\theta}{2} \right) = p_{11} E \left( f_{n}(1, \sigma^1(n)) - \frac{\theta}{2} \right) + p_{12} E \left( f_{n}(1, \sigma^2(n)) - \frac{\theta}{2} \right) + p_{13} E \left( f_{n}(1, \sigma^3(n)) - \frac{\theta}{2} \right) + p_{14} E \left( f_{n}(1, \sigma^4(n)) - \frac{\theta}{2} \right) \]

\[ = \left( \frac{1+\theta(1-\lambda)}{2} \right) x_{n;1-\theta} + \frac{\theta}{2} y_{n;1-\theta} + (1-\theta)(1-\lambda) z_{n;1-\theta} \]

\[ = \lambda x_{n;1-\theta}. \]

(b) \[ E \left( Y_{ij}(n) - \frac{\theta}{2} \right) = p_{11} E \left( f_{n}(2, \sigma^1(n)) - \frac{\theta}{2} \right) + p_{12} E \left( f_{n}(2, \sigma^2(n)) - \frac{\theta}{2} \right) + p_{13} E \left( f_{n}(2, \sigma^3(n)) - \frac{\theta}{2} \right) + p_{14} E \left( f_{n}(2, \sigma^4(n)) - \frac{\theta}{2} \right) \]

\[ = \left( \frac{1+\theta(1-\lambda)}{2} \right) y_{n;1-\theta} + \frac{\theta}{2} x_{n;1-\theta} + (1-\theta)(1-\lambda) z_{n;1-\theta} \]

\[ = \lambda y_{n;1-\theta}. \]

(c) It follows immediately from the identity \[ \sum_{i=1}^{4} Y_{ij}(n) = 1 \] that, for \( i = 3, 4 \)

\[ E \left( Y_{ij}(n) - \frac{1-\theta}{2} \right) = -\frac{1}{2} \sum_{i=1}^{2} E \left( Y_{ij}(n) - \frac{\theta}{2} \right) = \lambda z_{n;1-\theta}. \]

(h) By equation \((2.3)\), we have

\[ \frac{\theta}{2} z_{n;1-\theta} = E \left[ X_1 - \frac{\theta}{2} \right] \left[ X_3 - \frac{1-\theta}{2} \right] \]

\[ = \frac{\theta}{2} \left[ f_{n}(1, \sigma^1(n)) - \frac{\theta}{2} \right] \left( f_{n}(3, \sigma^1(n)) - \frac{1-\theta}{2} \right) + \frac{\theta}{2} \left[ f_{n}(1, \sigma^2(n)) - \frac{\theta}{2} \right] \left( f_{n}(3, \sigma^2(n)) - \frac{1-\theta}{2} \right) \]

\[ + \frac{1-\theta}{2} \left[ f_{n}(1, \sigma^3(n)) - \frac{\theta}{2} \right] \left( f_{n}(3, \sigma^3(n)) - \frac{1-\theta}{2} \right) + \frac{1-\theta}{2} \left[ f_{n}(1, \sigma^4(n)) - \frac{\theta}{2} \right] \left( f_{n}(3, \sigma^4(n)) - \frac{1-\theta}{2} \right) \]

and then

\[ E \left( f_{n}(1, \sigma^3(n)) - \frac{\theta}{2} \right) \left( f_{n}(3, \sigma^3(n)) - \frac{1-\theta}{2} \right) + E \left( f_{n}(1, \sigma^4(n)) - \frac{\theta}{2} \right) \left( f_{n}(3, \sigma^4(n)) - \frac{1-\theta}{2} \right) \]

\[ = \theta z_{n;1-\theta} + \frac{\theta}{2(1-\theta)} \left[ u_{n;1-\theta} - \frac{\theta}{2} x_{n;1-\theta} \right] + 3 \left( v_{n;1-\theta} - \frac{\theta}{2} x_{n;1-\theta} \right). \]

Therefore, for \( i = 3, 4 \), Lemma 2 implies

\[ E \left( Y_{ij}(n) - \frac{\theta}{2} \right) \left( Y_{ij}(n) - \frac{1-\theta}{2} \right) \]

\[ = p_{11} E \left( f_{n}(1, \sigma^1(n)) - \frac{\theta}{2} \right) \left( f_{n}(3, \sigma^1(n)) - \frac{1-\theta}{2} \right) + p_{12} E \left( f_{n}(1, \sigma^2(n)) - \frac{\theta}{2} \right) \left( f_{n}(3, \sigma^2(n)) - \frac{1-\theta}{2} \right) \]

\[ + p_{13} E \left( f_{n}(1, \sigma^3(n)) - \frac{\theta}{2} \right) \left( f_{n}(3, \sigma^3(n)) - \frac{1-\theta}{2} \right) + p_{14} E \left( f_{n}(1, \sigma^4(n)) - \frac{\theta}{2} \right) \left( f_{n}(3, \sigma^4(n)) - \frac{1-\theta}{2} \right) \]

\[ = \left[ \lambda + \frac{\theta(1-\lambda)}{2} \right] \left( \frac{\theta}{2} x_{n;1-\theta} + z_{n;1-\theta} - u_{n;1-\theta} + v_{n;1-\theta} \right) + \frac{\theta(1-\lambda)}{2} \left( \frac{\theta}{2} x_{n;1-\theta} + z_{n;1-\theta} - v_{n;1-\theta} \right) \]

\[ + \frac{(1-\theta)(1-\lambda)}{2} \left( \theta z_{n;1-\theta} + \frac{\theta}{2(1-\theta)} \left[ u_{n;1-\theta} - \frac{\theta}{2} x_{n;1-\theta} - v_{n;1-\theta} - \frac{\theta}{2} x_{n;1-\theta} \right] + 3 \left( v_{n;1-\theta} - \frac{\theta}{2} x_{n;1-\theta} \right) \right). \]
(j) Lemma 2 gives
\[
E \left[ f_n(1, \sigma^3(n)) - \frac{\theta}{2} \left( f_n(3, \sigma^3(n)) - \frac{1 - \theta}{2} \right) \right] = \frac{\theta z_{n,1-\theta}}{2} + \frac{\theta}{1-\theta} \left( w_{n,1-\theta} - \frac{1 - \theta}{2} x_{n,1-\theta} \right),
\]
and then we have
\[
E \left[ Y_{3j}(n) - \frac{1 - \theta}{2} \right] = \frac{1 - \theta}{2} \left( Y_{4j}(n) - \frac{1 - \theta}{2} \right) + \frac{\lambda}{2} \left( u_{n,\theta} - \frac{\theta}{2} x_{n,\theta} \right) + \frac{3 \lambda}{2} \left( v_{n,\theta} - \frac{\theta}{2} x_{n,\theta} \right) - \lambda \left( w_{n,1-\theta} - \frac{1 - \theta}{2} x_{n,1-\theta} \right).
\]

2.3 An equivalent condition for non-reconstruction

If the reconstruction problem is solvable, \( \sigma(n) \) contains significant information of the root variable. This can be expressed in several equivalent ways (Mossel [2001, 2004b]).

**Lemma 4** The non-reconstruction is equivalent to
\[
\lim_{n \to \infty} x_{n,\theta} = \lim_{n \to \infty} x_{n,1-\theta} = 0.
\]

3 Recursive formulas

3.1 Distributional recursion

In this section, we will explore the asymptotic behavior of \( x_{n,\theta} \) as \( n \) goes to infinity, which plays a crucial role in determining the reconstructibility. However, it is extremely challenging to obtain an explicit expression for \( x_{n,\theta} \). Therefore, we analyze the recursive relation between \( x_{n,\theta} \) and \( x_{n,1-\theta} \) on the tree structure instead. Consider \( A \) as a configuration on \( L(n+1) \) and let \( A_j \) be its restriction to \( \tau_{u_j} \cap L(n+1) \). Then from the Markov random field property, we have
\[
f_{n+1}(1, A) = \frac{N_1}{N_1 + N_2 + N_3 + N_4}, \tag{3.1}
\]
where
\[
N_1 = \frac{\theta}{2} \prod_{j=1}^d \left[ \sum_{i=1}^4 P_{i1} P(\sigma_j(n+1) = A_j | \sigma_{u_j} = i) \right] = \frac{\theta}{2} \prod_{j=1}^d \left[ 1 + \frac{2 \lambda}{\theta} \left( f_{n+1}(1, A_j) - \frac{\theta}{2} \right) \right] P(\sigma_j(n+1) = A_j); \\
N_2 = \frac{\theta}{2} \prod_{j=1}^d \left[ \sum_{i=1}^4 P_{i2} P(\sigma_j(n+1) = A_j | \sigma_{u_j} = i) \right] = \frac{\theta}{2} \prod_{j=1}^d \left[ 1 + \frac{2 \lambda}{\theta} \left( f_{n+1}(2, A_j) - \frac{\theta}{2} \right) \right] P(\sigma_j(n+1) = A_j); \\
N_3 = \frac{1 - \theta}{2} \prod_{j=1}^d \left[ \sum_{i=1}^4 P_{3i} P(\sigma_j(n+1) = A_j | \sigma_{u_j} = i) \right] = \frac{1 - \theta}{2} \prod_{j=1}^d \left[ 1 + \frac{2 \lambda}{1 - \theta} \left( f_{n+1}(3, A_j) - \frac{1 - \theta}{2} \right) \right] P(\sigma_j(n+1) = A_j); \\
N_4 = \frac{1 - \theta}{2} \prod_{j=1}^d \left[ \sum_{i=1}^4 P_{4i} P(\sigma_j(n+1) = A_j | \sigma_{u_j} = i) \right] = \frac{1 - \theta}{2} \prod_{j=1}^d \left[ 1 + \frac{2 \lambda}{1 - \theta} \left( f_{n+1}(4, A_j) - \frac{1 - \theta}{2} \right) \right] P(\sigma_j(n+1) = A_j).
\]

Setting \( A = \sigma^1(n+1) \), we have
\[
f_{n+1}(1, \sigma^1(n+1)) = \frac{\theta}{2} Z_1 \\
\frac{Z_1}{Z_2} = \frac{Z_3}{Z_4} = \frac{1}{\theta} \frac{Z_1}{Z_2} + \frac{1 + \theta}{\theta} \frac{Z_3}{Z_4} + \frac{1 + \theta}{\theta} \frac{Z_1}{Z_2} + \frac{1 + \theta}{\theta} \frac{Z_3}{Z_4},
\]
where
\[
Z_i = \left\{ \begin{array}{ll}
\prod_{j=1}^d \left[ 1 + \frac{2 \lambda}{\theta} \left( Y_{ij}(n) - \frac{\theta}{2} \right) \right] & \text{for } i = 1, 2, \\
\prod_{j=1}^d \left[ 1 + \frac{2 \lambda}{\theta} \left( Y_{ij}(n) - \frac{1 - \theta}{2} \right) \right] & \text{for } i = 3, 4.
\end{array} \right.
\]

In the next two lemmas, we provide some important identities regarding \( Z_i(n) \).
Lemma 5. For any nonnegative \( n \in \mathbb{Z}^+ \), we have
\[
E(Z_1(n)Z_2(n)) = EZ_2^2(n).
\]

Proof. For any configuration \( A = (A_1, \ldots, A_d) \) on the \((n+1)\)th level, with \( A_j \) denoting the spins on \( L_{n+1} \cap T_{u_j} \), we have
\[
Z_i = \frac{2}{\theta} \frac{P(\sigma(n+1) = A)}{\prod_{j=1}^{d} P(\sigma_j(n+1) = A_j)} P(\sigma_\phi = i | \sigma(n+1) = A), \quad \text{for } i = 1, 2.
\]

By the symmetry of the tree, we have
\[
E(Z_1(n)Z_2(n)) = \left(\frac{2}{\theta}\right)^2 \sum_A \left(\frac{P(\sigma(n+1) = A)}{\prod_{j=1}^{d} P(\sigma_j(n+1) = A_j)}\right)^2 P(\sigma_\phi = 1 | \sigma(n+1) = A) P(\sigma_\phi = 2 | \sigma(n+1) = A) P(\sigma(n+1) = A | \sigma_\phi = 1)
\]
\[
= \left(\frac{2}{\theta}\right)^2 \sum_A \left(\frac{P(\sigma(n+1) = A)}{\prod_{j=1}^{d} P(\sigma_j(n+1) = A_j)}\right)^2 \mathbb{P}(\sigma_\phi = 1 | \sigma(n+1) = A) \mathbb{P}(\sigma(n+1) = A | \sigma_\phi = 2)
\]
\[
= \left(\frac{2}{\theta}\right)^2 \sum_A \left(\frac{P(\sigma(n+1) = A)}{\prod_{j=1}^{d} P(\sigma_j(n+1) = A_j)}\right)^2 \mathbb{P}(\sigma_\phi = 2 | \sigma(n+1) = A) \mathbb{P}(\sigma(n+1) = A | \sigma_\phi = 1)
\]
\[
= EZ_2^2.
\]

The means and variances of monomials of \( Z_i \) can be approximated, using the notation \( O_\theta \) to emphasize that the constant associated with the \( O \)-term depends on \( \theta \) only, as follows:

Lemma 6. One has:

(i) \( EZ_1 = 1 + d\lambda^2 \frac{2}{\theta} x_{n, \theta} + \frac{d(d-1)}{2} \lambda^4 \frac{4}{\theta^2} x_{n, \theta}^2 + O_\theta(x_{n, \theta}^3) \),

(ii) \( EZ_2 = 1 + d\lambda^2 \frac{2}{\theta} y_{n, \theta} + \frac{d(d-1)}{2} \lambda^4 \frac{4}{\theta^2} y_{n, \theta}^2 + O_\theta(x_{n, \theta}^3) \),

(iii) \( EZ_i = 1 + d\lambda^2 \frac{2}{1-\theta} z_{n, 1-\theta} + \frac{d(d-1)}{2} \lambda^4 4 \frac{1}{1-\theta} z_{n, 1-\theta}^2 + O_\theta(x_{n, \theta}^3), \quad i = 3, 4. \)

(iv) \( EZ_1^2 = 1 + d\Pi_1 + \frac{d(d-1)}{2} \Pi_1^2 + O_\theta(x_{n, \theta}^3) \), where
\[
\Pi_1 = E \left[ 1 + \frac{2\lambda}{\theta} \left( Y_j - \frac{\theta}{2} \right) \right]^2 - 1 = \frac{6\lambda^2}{\theta} x_{n, \theta} + \frac{4\lambda^3}{\theta^2} \left( u_{n, \theta} - \frac{\theta}{2} x_{n, \theta} \right).
\]

(v) \( EZ_2^2 = EZ_1 Z_2 = 1 + d\Pi_2 + \frac{d(d-1)}{2} \Pi_2^2 + O_\theta(x_{n, \theta}^3) \), where
\[
\Pi_2 = E \left[ 1 + \frac{2\lambda}{\theta} \left( Y_j - \frac{\theta}{2} \right) \right]^2 - 1 = \frac{2\lambda^2}{\theta} x_{n, \theta} + \frac{4\lambda^2}{\theta} y_{n, \theta} + \frac{4\lambda^3}{\theta^2} \left( v_{n, \theta} - \frac{\theta}{2} x_{n, \theta} \right).
\]

(vi) \( EZ_i^2 = 1 + d\Pi_3 + \frac{d(d-1)}{2} \Pi_3^2 + O_\theta(x_{n, \theta}^3) \), for \( i = 3, 4 \), where
\[
\Pi_3 = E \left[ 1 + \frac{2\lambda}{1-\theta} \left( Y_j - \frac{1-\theta}{2} \right) \right]^2 - 1 = \frac{4\lambda^2}{1-\theta} z_{n, 1-\theta} + \frac{2\lambda^2}{1-\theta} x_{n, 1-\theta} + \frac{4\lambda^3}{1-\theta} \left( w_{n, 1-\theta} - \frac{1-\theta}{2} x_{n, 1-\theta} \right).
\]

(vii) \( EZ_1 Z_i = 1 + d\Pi_4 + \frac{d(d-1)}{2} \Pi_4^2 + O_\theta(x_{n, \theta}^3) \), for \( i = 3, 4 \), where
\[
\Pi_4 = E \left[ 1 + \frac{2\lambda}{1-\theta} \left( Y_j - \frac{1-\theta}{2} \right) \right]^2 - 1 = \frac{2\lambda^2}{1-\theta} x_{n, 1-\theta} + \frac{4\lambda^2}{1-\theta} z_{n, 1-\theta} - \frac{2\lambda^3}{1-\theta} \left( u_{n, \theta} - \frac{\theta}{2} x_{n, \theta} \right) + \frac{v_{n, \theta} - \theta x_{n, \theta}}{2}.
\]
(viii) \( \mathbb{E}Z_2Z_i = 1 + d \Pi_5 + \frac{d(d-1)}{2} \Pi_5^2 + O(\varepsilon^3) \), for \( i = 3, 4 \), where

\[
\Pi_5 = \mathbb{E}\left[1 + \frac{2\lambda}{\theta} \left( \frac{Y_{2j} - \theta}{2} \right) \right] \left[1 + \frac{2\lambda}{1 - \theta} \left( \frac{Y_{2j} - 1 - \theta}{2} \right) \right] - 1 = \frac{2\lambda^2}{\theta} y_{n\theta} + \frac{4\lambda^2}{1 - \theta} z_{n1 - \theta} - \frac{4\lambda^3}{\theta(1 - \theta)} \left( v_{n\theta} - \frac{\theta}{2} x_{n\theta} \right). 
\]

(ix) \( \mathbb{E}Z_3Z_4 = 1 + d \Pi_6 + \frac{d(d-1)}{2} \Pi_6^2 + O(\varepsilon^3) \), where

\[
\Pi_6 = \mathbb{E}\left[1 + \frac{2\lambda}{1 - \theta} \left( \frac{Y_{4j} - 1 - \theta}{2} \right) \right] \left[1 + \frac{2\lambda}{1 - \theta} \left( \frac{Y_{4j} - 1 - \theta}{2} \right) \right] - 1 = \frac{4\lambda^2}{1 - \theta} z_{n1 - \theta} + \frac{2\lambda^2}{1 - \theta} y_{n1 - \theta} + \frac{3\lambda^2}{(1 - \theta)^2} \left( u_{n\theta} - \frac{\theta}{2} x_{n\theta} \right) + 3 \left( v_{n\theta} - \frac{\theta}{2} x_{n\theta} \right) - \frac{4\lambda^3}{(1 - \theta)^2} \left( w_{n1 - \theta} - \frac{1 - \theta}{2} x_{n1 - \theta} \right). 
\]

3.2 Main expansions of \( x_{n1;\theta} \) and \( z_{n1;1 - \theta} \)

In this section, we investigate the second order recursive relations associated with \( x_{n1;\theta} \) and \( z_{n1;1 - \theta} \), with the assistance of the following identity

\[
\frac{a}{s + r} = \frac{a}{s} - \frac{ar}{s^2} + \frac{r^2}{s^2} \frac{a}{s + r}. \tag{3.2}
\]

Plugging \( a = \frac{1 - \theta}{2} Z_3 \) in equation (3.2), we have

\[
z_{n1;1 - \theta} + \frac{1 - \theta}{2} = \mathbb{E}\left[ \frac{1 - \theta}{2} Z_3 \right] - \mathbb{E}\left[ \frac{1 - \theta}{2} Z_3 \left( \frac{\theta}{2} Z_1 + \frac{\theta}{2} Z_2 + \frac{1 - \theta}{2} Z_3 + \frac{1 - \theta}{2} Z_4 \right) \right]
\]

\[
= \mathbb{E}\left[ \frac{1 - \theta}{2} Z_3 \right] - \mathbb{E}\left[ \frac{1 - \theta}{2} Z_3 \left( \frac{\theta}{2} Z_1 + \frac{\theta}{2} Z_2 + \frac{1 - \theta}{2} Z_3 + \frac{1 - \theta}{2} Z_4 - 1 \right) \right]
\]

\[
+ \mathbb{E}\left[ \left( \frac{\theta}{2} Z_1 + \frac{\theta}{2} Z_2 + \frac{1 - \theta}{2} Z_3 + \frac{1 - \theta}{2} Z_4 - 1 \right)^2 \frac{1 - \theta}{2} Z_3 \right]. \tag{3.3}
\]

Next by the results in Section 3.1 and taking \( Z_{n\theta} = -z_{n1; -\theta} \), we have

\[
Z_{n1;\theta} = d \lambda^2 Z_{n\theta} + \frac{d(d-1)}{2} \lambda^4 \left( \frac{2(1 - \theta)^2}{\theta} x_{n\theta} - 4(1 - \theta)^2 x_{n\theta} + 4 \left( \frac{4}{\theta} - \frac{4}{1 - \theta} - 8 \right) \right) x_{n\theta}
\]

\[
+ \theta \left( x_{n1;\theta} + y_{n1;1 - \theta} \right) + R_z + V_z, \tag{3.4}
\]

where

\[
R_z = \mathbb{E}\left( \frac{\theta}{2} Z_1 + \frac{\theta}{2} Z_2 + \frac{1 - \theta}{2} Z_3 + \frac{1 - \theta}{2} Z_4 - 1 \right)^2 \frac{1 - \theta}{2} Z_3 \left( \frac{\theta}{2} Z_1 + \frac{\theta}{2} Z_2 + \frac{1 - \theta}{2} Z_3 + \frac{1 - \theta}{2} Z_4 - 1 \right)
\]

\[
\frac{1 - \theta}{2} Z_3 \right) \left( \frac{\theta}{2} Z_1 + \frac{\theta}{2} Z_2 + \frac{1 - \theta}{2} Z_3 + \frac{1 - \theta}{2} Z_4 \right). \tag{3.5}
\]

and

\[
|V_z| \leq C_{V'} x_{n\theta}^2 \left( \left| \frac{u_{n\theta}}{x_{n\theta}^2} - \frac{\theta}{2} \right| + \left| \frac{v_{n\theta}}{x_{n\theta}^2} - \frac{\theta}{2} \right| + x_{n\theta} \right) + C_{V'} x_{n1;1 - \theta} \left( \left| \frac{w_{n1;1 - \theta}}{x_{n1;1 - \theta}} - \frac{1 - \theta}{2} \right| + x_{n1;1 - \theta} \right) \tag{3.6}
\]

with \( C_{V'} \) a constant depending on \( \theta \) only.

Similarly, there is the recursive relation for \( x_{n1;\theta} \):

\[
x_{n1;\theta} = \mathbb{E}\left( \frac{\theta}{2} Z_1 - \frac{\theta}{2} Z_1 + \frac{1 - \theta}{2} Z_3 + \frac{1 - \theta}{2} Z_4 - 1 \right)
\]

\[
+ \mathbb{E}\left( \frac{\theta}{2} Z_1 + \frac{\theta}{2} Z_2 + \frac{1 - \theta}{2} Z_3 + \frac{1 - \theta}{2} Z_4 - 1 \right)^2 \frac{\theta}{2} Z_1 + \frac{\theta}{2} Z_2 + \frac{1 - \theta}{2} Z_3 + \frac{1 - \theta}{2} Z_4 \right) \right)
\]

\[
= \lambda^2 x_{n\theta} + \frac{d(d-1)}{2} \lambda^4 \left( \left( -6 + \frac{2(1 - \theta)^2}{\theta} \right) + 20 \right) x_{n\theta} + \left( -4 \Theta - 16 \right) x_{n1;1 - \theta}
\]

\[
+ (4 \Theta - 16) x_{n\theta} z_{n1;1 - \theta} + \theta \left( x_{n1;1 - \theta} + y_{n1;1 - \theta} \right) + R_x + V_x. \tag{3.7}
\]
where
\[ R_x = \mathbb{E} \left( \frac{\theta}{2} Z_1 + \frac{\theta}{2} Z_3 + \frac{1-\theta}{2} Z_2 + \frac{1-\theta}{2} Z_4 - 1 \right) \left( \frac{\theta}{2} Z_1 + \frac{\theta}{2} Z_3 + \frac{1-\theta}{2} Z_2 + \frac{1-\theta}{2} Z_4 - 1 \right) \]
and
\[ |V_x| \leq C_\gamma x_{n,\theta}^2 \left( \frac{\|u_{n,\theta}\|}{x_{n,\theta}} \frac{\theta}{2} + \frac{\|v_{n,\theta}\|}{x_{n,\theta}} \frac{\theta}{2} + x_{n,\theta} \right) + C_\gamma x_{n,1-\theta}^2 \left( \frac{\|u_{n,1-\theta}\|}{x_{n,1-\theta}} \frac{1-\theta}{2} + x_{n,1-\theta} \right). \]

4 Concentration analysis

In order to study the stability of the dynamical system in equation (3.4), we expect \( R_x \) and \( V_x \) to be just small perturbations in the order of \( o \left( x_{n,\theta}^2 + x_{n,1-\theta}^2 \right) \). The following lemma ensures that \( x_n \) does not drop too fast.

**Lemma 7** For any \( \phi > 0 \), there exists a constant \( \gamma = \gamma(\phi, \phi) > 0 \), such that for all \( n \) when \( |\lambda| > \phi \)
\[ x_{n+1,\theta} \geq \gamma x_{n,\theta}. \]

**Proof.** Define \( \Delta_n = \mathbb{E} \max(X_I(n), X_H(n)) \), where \( I = (1, 2) \) and \( H = (3, 4) \). Then the following inequality holds
\[ 2 \left( x_{n,\theta} + \frac{\theta}{2} \right) \leq \mathbb{E} f_n(I, \sigma^I(n)) \leq \Delta_n, \]
where \( f_n(I, \sigma^I(n)) \) satisfies
\[ f_n(1, \sigma^1(n)) + f_n(2, \sigma^2(n)) \leq f_n(I, \sigma^I(n)). \]
Furthermore, Lemma 1 indicates that \( x_{n,\theta} = \frac{2}{\theta} \mathbb{E} \left( X_1(n) - \frac{\theta}{2} \right)^2 \). In the sequel, we consider \( \theta \geq \frac{1}{2} \), and the case of \( \theta < \frac{1}{2} \)
can be handled similarly. By the Cauchy-Schwarz inequality and using the fact that
\[ (X_I - \theta) + (X_H - (1-\theta)) = 0, \]
we have
\[ \Delta_n \leq \theta + \mathbb{E} \max(X_I(n) - \theta, X_H(n) - (1-\theta)) \]
\[ = \theta + \mathbb{E}|X_I(n) - \theta| \]
\[ = \theta + \mathbb{E}|X_I(n) + X_2(n) - \theta| \]
\[ \leq \theta + 2\mathbb{E} \left| X_I(n) - \frac{\theta}{2} \right| \]
\[ \leq \theta + 2 \left( \mathbb{E} \left( X_I(n) - \frac{\theta}{2} \right)^2 \right)^{1/2} \]
\[ = \theta + \sqrt{2\theta x_{n,\theta}}. \quad (4.1) \]

For a configuration \( A = (A_1, \ldots, A_d) \) on \( L(n+1) \) with \( A_j \) on \( \mathbb{T}_{u_j} \cap L(n+1) \), define
\[ f_{n+1}^*(1, A) = \mathbb{P}(\sigma_\rho = 1 | \sigma_1(n+1) = A_1) \]
\[ = \frac{\theta}{2} \mathbb{P}(\sigma_1(n+1) = A | \sigma_\rho = 1) \]
\[ = \frac{\theta}{2} \sum_{i=1}^4 P_{i1} \mathbb{P}(\sigma_1(n+1) = A | \sigma_{u_i} = i) \]
\[ = \frac{\theta}{2} \left[ \frac{2P_{11}}{\theta} f_n(1,A) + \frac{2P_{12}}{\theta} f_n(2,A) + \frac{2P_{13}}{1-\theta} f_n(3,A) + \frac{2P_{14}}{1-\theta} f_n(4,A) \right] \]
\[ = \frac{\theta}{2} \left[ 1 + \frac{2\lambda}{\theta} \left( f_n(1,A) - \frac{\theta}{2} \right) \right]. \]

and hence
\[ \mathbb{E} f_{n+1}^*(1, \sigma_1^1(n+1)) = \frac{\theta}{2} + \lambda^2 x_{n,\theta}. \]
Therefore, it follows from equation (4.1) that
\[ \frac{\theta}{2} + \lambda^2 x_{n;\theta} \leq \Delta_{n+1} \leq \frac{\theta}{2} + \sqrt{2\theta x_{n+1;\theta}} \leq \frac{\theta}{2} + \sqrt{2x_{n+1;\theta}}, \]

namely,
\[ x_{n+1;\theta} \geq \frac{\lambda^2}{2} x_{n;\theta} \geq \frac{\theta^2}{2} x_{n;\theta}, \quad (4.2) \]

Noting that \( Z_i \geq 0 \) and then \( 0 \leq \frac{\theta Z_i}{2 + \frac{\lambda^2}{2} Z_i + \frac{\lambda^2}{2} Z_i + \frac{\lambda^2}{2} Z_i} \leq 1 \), it is concluded from equation (3.7) that
\[ x_{n+1;\theta} - d\lambda^2 x_{n;\theta} \geq \frac{\theta}{2} Z_1 - \frac{\theta}{2} Z_1 \left( \frac{\theta}{2} Z_1 + \frac{\theta}{2} Z_2 + \frac{1-\theta}{2} Z_3 + \frac{1-\theta}{2} Z_4 - 1 \right) - d\lambda^2 x_{n;\theta} \]
\[ \geq -C_0 x_{n;\theta}^2. \quad (4.3) \]

Thus there exists a \( \delta = \delta(\theta, \rho) = \rho^2/C_0 > 0 \) such that if \( x_{n;\theta} < \delta \) then
\[ x_{n+1;\theta} \geq (d\lambda^2 - \rho^2) x_{n;\theta} \geq (d-1)\rho^2 x_{n;\theta} \geq \rho^2 x_{n;\theta}. \]

Also noting that if \( x_{n;\theta} \geq \delta \), equation (4.2) becomes \( x_{n+1;\theta} \geq \frac{\theta^2}{2} \delta x_{n;\theta}. \)

Finally taking \( \gamma = \min(\rho^2, \rho^4 \delta/2) \) completes the proof. \( \square \)

It is known that fixed finite different vertices far away from the root carry little information of the root, based on which, in-depth concentration results could be established.

**Lemma 8** For any \( \epsilon > 0 \) and a positive integer \( k \), there exists \( M = M(\theta, \epsilon, k) \) such that for any collection of vertices \( v_1, \ldots, v_k \in L(M) \), when \( i = 1, 2 \)
\[
\sup_{i_1, \ldots, i_k \notin \epsilon} \left| P(\sigma_{i_1} = i_1, \sigma_{v_j} = i_j, 1 \leq j \leq k) - \frac{\theta}{2} \right| \leq \epsilon,
\]

while for \( i = 3, 4 \)
\[
\sup_{i_1, \ldots, i_k \notin \epsilon} \left| P(\sigma_{i_1} = i_1, \sigma_{v_j} = i_j, 1 \leq j \leq k) - \frac{1-\theta}{2} \right| \leq \epsilon.
\]

**Proof.** Denote the transition probability from state \( i \) to state \( j \) at distance \( s \) by \( P_{i,j}^s \). Applying Bayes’ theorem, we have
\[
P_{1,1}^{s+1} = P_{11}^{s+1} + P_{12}^{s+1} P_{2,1}^{s+1} + P_{13}^{s+1} P_{3,1}^{s+1} + P_{14}^{s+1},
\]
\[
= \left( \lambda + \frac{\theta(1-\lambda)}{2} \right) \left( \frac{\theta}{2} + \frac{1-\theta}{2} \lambda^s \right) + \frac{\theta(1-\lambda)}{2} \left( \frac{\theta}{2} + \frac{1-\theta}{2} \lambda^s \right) + \left( \lambda(1-\lambda) \left( \frac{\theta}{2} + \frac{1-\theta}{2} \lambda^s \right) \right)
\]
\[
= \frac{\theta}{2} + \frac{1-\theta}{2} \lambda^{s+1}.
\]

Similarly, we can achieve the following results by induction:
\[
P_{1,1}^s = P_{2,2}^s = \frac{\theta}{2} + \frac{1-\theta}{2} \lambda^s,
\]
\[
P_{1,2}^s = P_{2,1}^s = \frac{\theta}{2} - \frac{1-\theta}{2} \lambda^s,
\]
\[
P_{3,3}^s = P_{4,4}^s = \frac{1-\theta}{2} + \frac{1-\theta}{2} \lambda^s,
\]
\[
P_{3,4}^s = P_{4,3}^s = \frac{1-\theta}{2} - \frac{1-\theta}{2} \lambda^s,
\]
\[
P_{3,1}^s = P_{4,1}^s = P_{3,2}^s = P_{4,2}^s = \frac{\theta}{2} - \frac{1-\theta}{2} \lambda^s,
\]
\[
P_{1,3}^s = P_{4,1}^s = P_{2,3}^s = P_{4,2}^s = \frac{1-\theta}{2} + \frac{1-\theta}{2} \lambda^s.
\]

Consequently, under the condition that \( d\lambda^2 \leq 1 \), when \( i = 1, 2 \) one has
\[
\frac{\theta}{2} - \frac{d^{-s/2}}{2} \leq P_{i,l}^s \leq \frac{\theta}{2} + d^{-s/2},
\]
and when \( j = 3, 4 \) one has

\[
\frac{1 - \theta}{2} - d^{-s/2} \leq P_{\ell, j} \leq \frac{1 - \theta}{2} + d^{-s/2}.
\]

For fixed \( \theta, d \) and \( k \), define

\[
B(s) = \max \left\{ \frac{\theta}{2} + d^{-s/2}, \frac{1 - \theta}{2} + d^{-s/2}, \frac{\theta}{2} - d^{-s/2}, \frac{1 - \theta}{2} - d^{-s/2} \right\},
\]

and let \( N = N(\theta, k, \varepsilon) \) be a sufficiently large integer such that

\[
B^k(N) \leq 1 + \varepsilon,
\]

which holds for the reason that \( d^{-\frac{s}{2}} \leq 2^{-\frac{s}{2}} \to 0 \) as \( s \to \infty \) and then \( B(s) \to 1 \) uniformly for all \( d \).

Fix an integer \( M \) such that \( M > kN \) and choose any \( v_1, \ldots, v_k \in L(M) \). For \( 0 \leq \ell \leq M \), define \( n_{\ell} \) as the number of vertices in distance \( \ell \) from the root with a descendent in the set \( \{v_1, \ldots, v_k\} \), that is,

\[
n_{\ell} = \{ v \in L(\ell) : |v_\ell \cap \{v_1, \ldots, v_k\}| > 0 \}.
\]

Apparently, \( n_0 = 1 \) and \( n_M = k \). Also, we can easily see that \( n_{\ell} \) is an increasing integer valued function and there must exist some \( \ell \) such that \( n_{\ell} = n_{\ell+N} \). Let \( \overline{w_1, \ldots, w_{n_{\ell}}} \) be the vertices in the set \( \{ v \in L(\ell) : |v_\ell \cap \{v_1, \ldots, v_k\}| > 0 \} \), and \( \{w_1, \ldots, w_{n_{\ell}}\} \) be the vertices in the set \( \{ v \in L(\ell+N) : |v_{\ell+N} \cap \{v_1, \ldots, v_k\}| > 0 \} \) such that \( w_j \) is the descendent of \( \overline{w_j} \). By the Markov random field property, \( \{\sigma_{w_j}\}_{j=1, \ldots, n_{\ell}} \) are conditionally independent given \( \sigma_{\overline{w_j}} \). The distribution of \( \sigma_{w_j} \) given \( \sigma_{\overline{w_j}} \) is given by

\[
P(\sigma_{w_j} = i_2 | \sigma_{\overline{w_j}} = i_1) = P_{i_1, i_2}^N.
\]

By Bayes’ theorem and the Markov random field property, for any \( i_1, \ldots, i_{n_{\ell}} \in \mathcal{C} \), we have

\[
P(\sigma_{\rho} = 1 | \sigma_{w_j} = i_j, 1 \leq j \leq n_{\ell}) = \frac{P(\sigma_{\rho} = i_1, 1 \leq j \leq n_{\ell} | \sigma_{\rho} = 1)}{P(\sigma_{\rho} = 1, \sigma_{w_j} = i_j, 1 \leq j \leq n_{\ell})} \leq 1 + \varepsilon,
\]

and

\[
P(\sigma_{\rho} = m | \sigma_{w_j} = i_j, 1 \leq j \leq n_{\ell}) \leq \frac{\theta}{1 - \theta}(1 + \varepsilon), \quad m = 3, 4.
\]

Therefore, we obtain

\[
\frac{\theta}{2} - \varepsilon \leq P(\sigma_{\rho} = 1 | \sigma_{w_j} = i_j, 1 \leq j \leq n_{\ell}) \leq \frac{\theta}{2} + \varepsilon.
\]

Finally, since \( \sigma_{\rho} \) is conditionally independent of the collection \( \{\sigma_{v_1}, \ldots, \sigma_{v_k}\} \) given \( \{\sigma_{w_1}, \ldots, \sigma_{w_{n_{\ell}}}\} \), it is concluded that

\[
sup_{i_1, \ldots, i_k} \left| P(\sigma_{\rho} = 1 | \sigma_{v_j} = i_j, 1 \leq j \leq k) - \frac{\theta}{2} \right| \leq \sup_{i_1, \ldots, i_{n_{\ell}}} \left| P(\sigma_{\rho} = 1 | \sigma_{w_j} = i_j, 1 \leq j \leq n_{\ell}) - \frac{\theta}{2} \right| \leq \varepsilon.
\]

The rest follows similarly. \( \square \)
Now, we are able to bound the remainders $R_z$ and $V_z$ in equation (3.5) and (3.6), using the preceding concentration results.

**Lemma 9** Assume $|\lambda| > \varphi$ for some $\varphi > 0$. For any $\epsilon > 0$, there exist $N = N(\theta, \epsilon)$ and $\delta = \delta(\theta, \epsilon, \varphi) > 0$, such that if $n \geq N$ and $x_{n, \theta}, x_{n, 1-\theta} \leq \delta$ then

$$|R_z| \leq \epsilon \left( x_{n, \theta}^2 + x_{n, 1-\theta}^2 \right).$$

**Proof.** Fix $k$ an integer such that $k > 6$. Choose $M$ such that the conclusions of Lemma 8 hold with bound $\xi$. Denote $v_1, \ldots, v_{|L(M)|}$ as the vertices in $L(M)$, define $\sigma_\varphi(n+1)$ as the spins of vertices in $T_v \cap L(n+1)$ conditioned on $\sigma_\varphi = 1$, and for $v \in \{v_1, \ldots, v_{|L(M)|}\}$ let

$$W(v) = f_{n+1-M}(1, \sigma_\varphi^1(n+1)).$$

Then, for $i \in \{1, 2, 3, 4\}$, $W(v)$ would be distributed as

$$W(v) \sim f_{n+1-M}(1, \sigma_i(n+1-M)), \quad \text{if } \sigma_\varphi^1 = i. \quad (4.4)$$

Using the recursive formula in equation (3.1), the posterior probability of a vertex can be written as a function of the posterior probabilities of its children, so there exists a function $H(W_1, \ldots, W_{|L(M)|})$ such that

$$H(W_1, \ldots, W_{|L(M)|}) = f_{n+1}(1, \sigma_\varphi^1(n+1)) = \frac{\frac{\theta}{2} Z_1}{\frac{\theta}{2} Z_1 + \frac{A}{Z_2} + \frac{1-A}{Z_3} + \frac{A}{Z_4}},$$

where $W_i = W(v_i)$ for $1 \leq i \leq |L(M)|$. We can see that $H$ is a continuous function, and when $W_i = \theta/2$ for all $i$

$$H(W_1, \ldots, W_{|L(M)|}) = \theta/2.$$

Therefore, by Lemma 8, if there are at most $k$ vertices in $L(M)$ such that $W(v) \neq \theta/2$ then

$$\left| H(W_1, \ldots, W_{|L(M)|}) - \frac{\theta}{2} \right| < \frac{\epsilon}{2},$$

and thus there exists some $\delta = \delta(\epsilon) > 0$ such that if

$$\# \{ v \in L(M) : \left| W(v) - \frac{\theta}{2} \right| > \delta \} \leq k$$

then

$$\left| H(W_1, \ldots, W_{|L(M)|}) - \frac{\theta}{2} \right| < \epsilon.$$

Next by the Chebyshev’s inequality and equation (4.4), the following holds

$$P \left( \left| W(v) - \frac{\theta}{2} \right| > \delta \right) \leq \delta^{-2} \sum_{i=1}^{A} E \left( f_{n+1-M}(1, \sigma_i^1(n+1-M)) - \frac{\theta}{2} \right)^2$$

$$= \delta^{-2} (u_{n+1-M, \theta} + v_{n+1-M, 1-\theta} + 2 w_{n+1-M})$$

$$\leq \frac{\delta^{-2}}{1-\theta} x_{n+1-M}.\theta.$$

For the reason that random variables $||W(v) - \theta/2||_{v \in \{v_1, \ldots, v_{|L(M)|}\}}$ are conditionally independent given $\sigma(M)$, there exist suitable constants $C(\theta, \epsilon, \varphi)$ and $N(\theta, \epsilon)$, such that whenever $n \geq N$,

$$P \left( \left| \frac{\theta}{2} Z_1 + \frac{A}{Z_2} + \frac{1-A}{Z_3} + \frac{A}{Z_4} - \frac{\theta}{2} \right| > \epsilon \right) \leq P \left( \# \{ v \in L(M) : \left| W(v) - \frac{\theta}{2} \right| > \delta \} > k \right)$$

$$= \sum_{A} P \left( \# \{ v \in L(M) : \left| W(v) - \frac{\theta}{2} \right| > k \} \sigma(M) = A \right) P(\sigma(M) = A)$$

$$\leq \sum_{A} \binom{\delta^{-2}}{1-\theta} x_{n+1-M}^6 \left| k \right| P(\sigma(M) = A)$$

$$\leq C x_{n, \theta}.\theta^6$$
where $\text{Bin}(\cdot, \cdot)$ denotes the Binomial distribution and the last inequality follows from Lemma 7. Similarly, we can show that, when $j = 3, 4,$

$$P\left(\left| \frac{1}{2} Z_1 + \frac{1}{2} Z_2 + \frac{1}{2} Z_3 + \frac{1}{2} Z_4 - \frac{1 - \theta}{2} \right| > \varepsilon \right) \leq C \alpha_{n,1-\theta}^6.$$  

For any $\eta > 0,$ it follows from the Cauchy-Schwarz inequality that

$$|R_2| = \left| \mathbb{E}\left(\frac{\theta}{2} Z_1 + \frac{\theta}{2} Z_2 + \frac{1 - \theta}{2} Z_3 + \frac{1 - \theta}{2} Z_4 - 1\right) \right|^2 \left| \mathbb{E}\left(\frac{1 - \theta}{2} Z_3 - \frac{1 - \theta}{2} \right) > \eta \right| \leq \frac{\mathbb{E}(\theta Z_1 + \theta Z_2 + \frac{1 - \theta}{2} Z_3 + \frac{1 - \theta}{2} Z_4 - 1)^2}{\mathbb{E}(\frac{1 - \theta}{2} Z_3 - \frac{1 - \theta}{2} > \eta)} \leq C \alpha_{n,1-\theta}^6.$$  

$$\left(\frac{1 - \theta}{2} Z_3 - \frac{1 - \theta}{2} \right) > \eta \right| \leq \frac{\mathbb{E}(\theta Z_1 + \theta Z_2 + \frac{1 - \theta}{2} Z_3 + \frac{1 - \theta}{2} Z_4 - 1)^2}{\mathbb{E}(\frac{1 - \theta}{2} Z_3 - \frac{1 - \theta}{2} > \eta)} \leq C \alpha_{n,1-\theta}^6.$$  

Note that the calculations in Section 3.1 imply that

$$\mathbb{E}\left(\frac{\theta}{2} Z_1 + \frac{\theta}{2} Z_2 + \frac{1 - \theta}{2} Z_3 + \frac{1 - \theta}{2} Z_4 - 1\right)^2 \leq C_1 \alpha_{n,1-\theta}^2 + \alpha_{n,1-\theta}^2 \leq C_2.$$  

Thus, there exist $C_3 = C_3(\theta, \eta, \theta)$ and $N = N(\theta, \eta),$ such that if $n > N$ then

$$\mathbb{E}\left(\frac{\theta}{2} Z_1 + \frac{\theta}{2} Z_2 + \frac{1 - \theta}{2} Z_3 + \frac{1 - \theta}{2} Z_4 - 1\right)^4 \leq C_2.$$  

Finally, taking $\eta = \frac{C_3}{2C_1}$ and $\delta = \frac{\varepsilon}{2 \sqrt{C_2 C_3}},$ we conclude that if $n \geq N$ and $x_{n,\theta}, x_{n,1-\theta} \leq \delta$ then

$$|R_2| \leq \eta C_1 \alpha_{n,\theta}^2 + C_2 C_3 \alpha_{n,1-\theta}^3 \leq \varepsilon \left(x_{n,\theta}^2 + x_{n,1-\theta}^2\right).$$  

\textbf{Lemma 10} Assume $|\lambda| > \rho$ for some $\rho > 0.$ For any $\varepsilon > 0,$ there exist $N = N(\theta, \varepsilon)$ and $\delta = \delta(\varepsilon, \rho), $ such that if $n \geq N$ and $x_{n,\theta}, x_{n,1-\theta} \leq \delta$ then

$$|V_2| \leq \varepsilon \left(x_{n,\theta}^2 + x_{n,1-\theta}^2\right).$$  

\textbf{Proof.} Plugging

$$a = \left(\frac{\theta}{2} Z_1 + \frac{\theta}{2} Z_2 + \frac{1 - \theta}{2} Z_3 + \frac{1 - \theta}{2} Z_4\right)^2, \quad r = \left(\frac{\theta}{2} Z_1 + \frac{\theta}{2} Z_2 + \frac{1 - \theta}{2} Z_3 + \frac{1 - \theta}{2} Z_4\right)^2 - 1, \quad s = 1.$$
in equation (3.2), we have
\[ u_{n+1,\theta} = E \left( f_{n+1}(1, \sigma^1(n+1)) - \frac{\theta}{2} \right)^2 \]
\[ = E \left( \frac{\theta}{2} Z_1 \left( \frac{\theta}{2} Z_1 + \frac{\theta}{2} Z_2 + \frac{1-\theta}{2} Z_3 + \frac{1-\theta}{2} Z_4 \right) \right)^2 \]
\[ = \frac{\theta^2}{4} E \left( Z_1 \left( \frac{\theta}{2} Z_1 + \frac{\theta}{2} Z_2 + \frac{1-\theta}{2} Z_3 + \frac{1-\theta}{2} Z_4 \right) \right)^2 + \frac{\theta^2}{4} E \left( Z_1 \left( \frac{\theta}{2} Z_1 + \frac{1-\theta}{2} Z_3 + \frac{1-\theta}{2} Z_4 \right) \right)^2 \]
\[ + \frac{\theta^2}{4} E \left( Z_1 \left( \frac{\theta}{2} Z_1 + \frac{1-\theta}{2} Z_3 + \frac{1-\theta}{2} Z_4 \right) \right)^2 \]
\[ = \frac{\theta}{2} x_{n+1,\theta} + d\lambda^3 \left( u_{n,\theta} - \frac{x_{n,\theta}}{2} \right) + O_\theta(x_{n,\theta}). \quad (4.5) \]

By equation (4.3), there exist constants \( N_1 = N_1(\theta) \) and \( \delta_1 = \delta_1(\theta) \), such that if \( n \geq N_1 \) and \( x_{n,\theta} \leq \delta_1 \) then
\[ \frac{x_{n,\theta}}{x_{n+1,\theta}} \leq \frac{1}{\left(1 - \frac{4}{3}\right) d\lambda^2 x_{n,\theta}} = \frac{4}{3} d\lambda^2. \]

For fixed \( k \), it is known that there exists a \( \delta_2 = \delta_2(\theta, \varphi, k) < \delta_1 \), such that if \( x_{n,\theta} < \delta_2 \) then for any \( 1 \leq \ell \leq k \) one has \( x_{n+k,\theta} < 2\delta_2 \). Therefore, for any positive integer \( k \), equation (4.5) yields
\[ \frac{u_{n+k,\theta}}{x_{n+k,\theta}} - \frac{\theta}{2} = d\lambda^3 \left( \frac{k}{x_{n+k,\theta}} \right) \left( u_{n,\theta} - \frac{x_{n,\theta}}{2} \right) + O_\theta(x_{n,\theta}). \]
\[ = (d\lambda^3)^k \left( \prod_{\ell=1}^{k} \frac{x_{n+k,\theta}}{x_{n+k,\theta}} \right) \left( u_{n,\theta} - \frac{x_{n,\theta}}{2} \right) + R, \]

where
\[ \left| (d\lambda^3)^k \left( \prod_{\ell=1}^{k} \frac{x_{n+k,\theta}}{x_{n+k,\theta}} \right) \right| \leq |d\lambda^3|^k \left( \frac{4}{3} |\lambda| \right)^k = \left( \frac{4}{3} |\lambda| \right)^k \]

and
\[ |R| \leq 2C\delta_3 \left( \sum_{i=0}^{k} \frac{4}{3d\lambda^2} \right) \frac{1}{d\lambda^3} \left( \frac{1}{1 - \frac{1}{3}|\lambda|} \right)^{k-1} \]
\[ \leq 2C\delta_3 \frac{1 - \frac{4}{3}|\lambda|}{1 - \frac{4}{3}|\lambda|} \frac{4}{3d\lambda^2} \]

with \( C \) denoting the \( O_\theta \) constant in equation (4.5). By Lemma 1, it is easy to obtain \( 0 \leq \frac{u_{n,\theta}}{x_{n,\theta}} \leq 1 \), which implies
\[ \left| \frac{u_{n,\theta}}{x_{n,\theta}} - \frac{\theta}{2} \right| \leq 1. \]

Noticing the fact that \( |\lambda| \leq \varphi^{-1/2} \leq 1/\sqrt{3}, \) we achieve \( \frac{4}{3}|\lambda| < 1 \). Taking \( k = k(\varepsilon) \) sufficiently large and \( \delta_3 = \delta_3(\theta, k, \varepsilon) = \delta_3(\theta, \varepsilon) \) sufficiently small, we have
\[ \left| \frac{u_{n+k,\theta}}{x_{n+k,\theta}} - \frac{\theta}{2} \right| < \varepsilon. \]

Finally, in view of \( |\lambda| > \varphi \) and by Lemma 7, there exists \( \gamma = \gamma(\theta, \varphi) \) such that \( x_{n-k,\theta} \leq \gamma^{-k} x_{n,\theta} \), and then by choosing \( N = N(\theta, \varepsilon, k) = N(\theta, \varepsilon) > N_1 + k \) and \( \delta = \gamma^k \delta_3 \), if \( x_{n,\theta} \leq \delta \) and \( n \geq N \) one has
\[ \left| \frac{u_{n,\theta}}{x_{n,\theta}} - \frac{\theta}{2} \right| < \varepsilon. \quad (4.6) \]

Similar discussions yield
\[ \left| \frac{u_{n+1,\theta}}{x_{n+1,\theta}} - \frac{\theta}{2} \right| < \varepsilon \quad \text{and} \quad \left| \frac{u_{n+1,\theta}}{x_{n,\theta}} - \frac{1-\theta}{2} \right| < \varepsilon \]
and thus we complete the proof. \( \square \)
5 Proof of the Main Theorem

It follows from Lemma 1 that
\[ x_{n;1-\theta}^2 + y_{n;1-\theta}^2 \geq \frac{1}{2} (x_{n;1-\theta} + y_{n;1-\theta})^2 = \frac{2\theta^2}{(1-\theta)^2} x_{n;1-\theta}^2 \]
and equation (3.4) implies
\[ \mathcal{I}_{n;1,\theta} \geq d\lambda^2 \mathcal{I}_{n,\theta} + \frac{d(d-1)}{2} \lambda^4 \left[ \frac{2(1-\theta)}{\theta} (x_{n,\theta} - \mathcal{I}_{n,\theta})^2 + \frac{4}{1-\theta} - \frac{2(1-\theta)^2}{\theta} + \frac{2\theta^3}{(1-\theta)^2} \right] \mathcal{I}_{n,\theta}^2 + R_z + V_z. \]
Solving \( \left\{ \frac{4}{\theta} - \frac{4}{1-\theta} - 8 - \frac{2(1-\theta)^2}{\theta} + \frac{2\theta^3}{(1-\theta)^2} \right\} > 0 \), one has
\[ \frac{3+\sqrt{3}}{6} \leq \theta \leq 1 \quad \text{or} \quad 0 \leq \theta < \frac{3-\sqrt{3}}{6}. \]
For fixed \( \theta \in \left( \frac{3+\sqrt{3}}{6}, 1 \right] \), there exists \( \frac{1}{2} < \zeta < 1 \) such that
\[ \frac{4}{\theta} - \frac{4}{1-\theta} - 8 - \frac{2(1-\theta)^2}{\zeta \theta} + \frac{2\theta^3}{(1-\theta)^2} > 0. \]
To investigate the non-tightness, it would be convenient to assume that \( d\lambda^2 \geq \frac{1}{2} \), say, \( |\lambda| \geq \frac{1}{\sqrt{2d}} \). We take \( \varphi = \frac{1}{\sqrt{2d}} \) in Lemma 7 and then get \( \gamma = \gamma(\theta, d) \). By Lemma 9 and Lemma 10, there exist \( N = N(\theta, \zeta) \) and \( \delta = \delta(\theta, d, \zeta) > 0 \), such that if \( n \geq N \) and \( x_{n,\theta} \leq \delta \) then
\[ |R_z + V_z| \leq \frac{1}{16} \left( \frac{2(1-\theta)^2}{\theta} - 1 \right) x_{n,\theta}^2 + \frac{\theta}{16} (1-\zeta) x_{n,1-\theta}^2. \]
(5.1)
To accomplish the proof, it suffices to show that when \( d\lambda^2 \) is close enough to 1, at least one of \( x_{n,\theta} \) and \( x_{n;1-\theta} \) does not converge to 0. We apply reductio ad absurdum, by assuming that
\[ \lim_{n \to \infty} x_{n,\theta} = \lim_{n \to \infty} x_{n;1-\theta} = 0. \]
(5.2)
Thus, there exists \( \mathcal{N} = \mathcal{N}(\theta, d, \zeta) > N(\theta, \zeta) \), such that whenever \( n > \mathcal{N} \), we have \( x_{n,\theta} < \delta \). Denote
\[ \Gamma = \min \left\{ \frac{2(1-\theta)^2}{\theta}, \frac{4}{\theta} - \frac{4}{1-\theta} - 8 - \frac{2(1-\theta)^2}{\zeta \theta} + \frac{2\theta^3}{(1-\theta)^2} \right\} > 0 \]
and note that
\[ \frac{d(d-1)}{2} \lambda^4 \geq \left( \frac{d\lambda^2}{2} \right)^2 \geq \frac{1}{16}. \]
Consequently, by equations (3.4) and (5.1), it is concluded that if \( n > \mathcal{N} \)
\[ \mathcal{I}_{n+1,\theta} \geq d\lambda^2 \mathcal{I}_{n,\theta} + \frac{d(d-1)}{2} \lambda^4 \left[ \frac{2(1-\theta)^2}{\theta} x_{n,\theta}^2 - \frac{4(1-\theta)^2}{\theta} x_{n,\theta} \mathcal{I}_{n,\theta} + \left( \frac{4}{\theta} - \frac{4}{1-\theta} - 8 \right) \mathcal{I}_{n,\theta}^2 + \zeta \left( x_{n;1-\theta}^2 + y_{n;1-\theta}^2 \right) \right] \]
\[ \geq d\lambda^2 \mathcal{I}_{n,\theta} + \frac{d(d-1)}{2} \lambda^4 \left[ \frac{2(1-\theta)^2}{\theta} (x_{n,\theta} - \mathcal{I}_{n,\theta})^2 + \zeta \left( \frac{4}{\theta} - \frac{4}{1-\theta} - 8 - \frac{2(1-\theta)^2}{\zeta \theta} + \frac{2\theta^3}{(1-\theta)^2} \right) \left( \frac{\mathcal{I}_{n,\theta}}{\zeta} \right)^2 \right] \]
\[ \geq d\lambda^2 \mathcal{I}_{n,\theta} + \frac{d(d-1)}{4} \lambda^4 \left( x_{n,\theta} - \frac{\mathcal{I}_{n,\theta}}{\zeta} + \frac{\mathcal{I}_{n,\theta}}{\zeta} \right)^2 \]
\[ \geq \mathcal{I}_{n,\theta} \left( d\lambda^2 + \frac{d(d-1)}{4} \lambda^4 \right) x_{n,\theta}. \]
(5.3)
From the second inequality of equation (5.3), we can also conclude
\[ \mathcal{I}_{n+1,\theta} \geq d\lambda^2 \mathcal{I}_{n,\theta}^2 + \left( \frac{d\lambda^2}{2} \right)^2 \left( \frac{2(1-\theta)^2}{\theta} x_{n,\theta} - \frac{\mathcal{I}_{n,\theta}}{\zeta} \right)^2 \geq \frac{1}{8} \left( \frac{\mathcal{I}_{n,\theta}}{\zeta} \right)^2 + \frac{(1-\theta)^2}{16 \theta} \left( x_{n,\theta} - \frac{\mathcal{I}_{n,\theta}}{\zeta} \right)^2 \geq \zeta x_{n,\theta}^2, \]
(5.4)
where $\xi$ depends only on $\theta$.

Considering the initial point $x_0 = 1 - \frac{\theta}{2} > 0$ and by Lemma 7, we have $x_n, \theta \geq x_0 \gamma^n$. Define $\varepsilon = \varepsilon(\theta, d, \xi) = \xi \left( x_0 \gamma^{-\gamma} \right)^2 > 0$, choose suitable $|\lambda| < d^{-\frac{1}{4}}$, and then we have

$$d \lambda^2 + \frac{d(d-1)}{4} \lambda^4 \varepsilon > 1.$$ (5.5)

It is easy to see that equation (5.4) implies that $Z_{n+1, \theta} \geq \xi x_n^2 \geq \varepsilon$. Suppose $Z_n \geq \varepsilon$ for some $n > N$, and then it follows from equations (5.3) and (5.5) that

$$x_{n+1, \theta} \geq Z_{n+1, \theta} \geq Z_{n, \theta} \left[ d \lambda^2 + \frac{d(d-1)}{4} \lambda^4 \varepsilon \right] > Z_{n, \theta} \geq \varepsilon.$$

Therefore, by induction we have $x_{n, \theta} \geq Z_{n, \theta} \geq \varepsilon$ for all $n > N$, which contradicts the assumption imposed in equation (5.2). Thus, the proof of Theorem 1.3 is completed.

Acknowledgements We would like to thank Joseph Felsenstein and Sébastien Roch for conversations on this subject.

References

Noam Berger, Claire Kenyon, Elchanan Mossel, and Yuval Peres. Glauber dynamics on trees and hyperbolic graphs. *Probability Theory and Related Fields*, 131(3):311–340, 2005.

Shankar Bhamidi, Ram Rajagopal, and Sébastien Roch. Network delay inference from additive metrics. *Random Structures & Algorithms*, 37(2):176–203, 2010.

Pavel M Bleher, Jean Ruiz, and Valentin A Zagrebnov. On the purity of the limiting Gibbs state for the Ising model on the bethe lattice. *Journal of Statistical Physics*, 79(1-2):473–482, 1995.

Christian Borgs, Jennifer Chayes, Elchanan Mossel, and Sébastien Roch. The Kesten-Stigum reconstruction bound is tight for roughly symmetric binary channels. In *Foundations of Computer Science, 2006. FOCS’06. 47th Annual IEEE Symposium on*, pages 518–530. IEEE, 2006.

JT Chayes, L Chayes, James P Sethna, and DJ Thouless. A mean field spin glass with short-range interactions. *Communications in Mathematical Physics*, 106(1):41–89, 1986.

Constantinos Daskalakis, Elchanan Mossel, and Sébastien Roch. Optimal phylogenetic reconstruction. In *Proceedings of the thirty-eighth annual ACM symposium on Theory of computing*, pages 159–168. ACM, 2006.

William Evans, Claire Kenyon, Yuval Peres, and Leonard J Schulman. Broadcasting on trees and the Ising model. *Annals of Applied Probability*, pages 410–433, 2000.

Joseph Felsenstein. Evolutionary trees from DNA sequences: a maximum likelihood approach. *Journal of molecular evolution*, 17(6):368–376, 1981.

Joseph Felsenstein. *Inferring phylogenies*, volume 2. Sinauer associates Sunderland, MA, 2004.

Hans-Otto Georgii. *Gibbs measures and phase transitions*, volume 9. Walter de Gruyter, 2011.

Harry Kesten and Bert P Stigum. Additional limit theorems for indecomposable multidimensional galton-watson processes. *The Annals of Mathematical Statistics*, 37(6):1463–1481, 1966.

Harry Kesten and Bernt P Stigum. Limit theorems for decomposable multi-dimensional galton-watson processes. *Journal of Mathematical Analysis and Applications*, 17(2):309–338, 1967.

Motoo Kimura. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *Journal of molecular evolution*, 16(2):111–120, 1980.

Wenjian Liu and Ning Ning. Large degree asymptotics and the reconstruction threshold of the asymmetric binary channels. *https://arxiv.org/abs/1812.06039*, 2018.

Wenjian Liu, Sreenivasa Rao Jammalamadaka, and Ning Ning. The tightness of the Kesten-Stigum reconstruction bound of symmetric model with multiple mutations. *Journal of Statistical Physics*, 170(3):617–641, 2018.

Fabio Martinelli, Alistair Sinclair, and Dror Weitz. Fast mixing for independent sets, colorings, and other models on trees. *Random Structures & Algorithms*, 31(2):134–172, 2007.

Marc Mézard and Andrea Montanari. Reconstruction on trees and spin glass transition. *Journal of statistical physics*, 124(6):1317–1350, 2006.

Elchanan Mossel. Reconstruction on trees: beating the second eigenvalue. *Annals of Applied Probability*, pages 285–300, 2001.
Elchanan Mossel. Phase transitions in phylogeny. *Transactions of the American Mathematical Society*, 356(6):2379–2404, 2004a.

Elchanan Mossel. Survey: information flow on trees. *DIMACS series in discrete mathematics and theoretical computer science*, 63:155–170, 2004b.

Elchanan Mossel. Deep learning and hierarchal generative models. *arXiv preprint arXiv:1612.09057*, 2016.

Elchanan Mossel, Joe Neeman, and Allan Sly. A proof of the block model threshold conjecture. *Combinatorica*, pages 1–44, 2013.

Elchanan Mossel, Joe Neeman, and Allan Sly. Belief propagation, robust reconstruction and optimal recovery of block models. In *Conference on Learning Theory*, pages 356–370, 2014.

Joe Neeman and Praneeth Netrapalli. Non-reconstructability in the stochastic block model. *arXiv preprint arXiv:1404.6304*, 2014.

Federico Ricci-Tersenghi, Guilhem Semerjian, and Lenka Zdeborova. Typology of phase transitions in Bayesian inference problems. *arXiv preprint arXiv:1806.11013*, 2018.

Sebastien Roch. A short proof that phylogenetic tree reconstruction by maximum likelihood is hard. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 3(1):92–94, 2006.

Allan Sly. Reconstruction for the Potts model. *The Annals of Probability*, 39:1365–1406, 2011.

Prasad Tetali, Juan C Vera, Eric Vigoda, and Linji Yang. Phase transition for the mixing time of the glauber dynamics for coloring regular trees. *Ann. Appl. Probab.*, 22:2210–2239., 2012.