The Annals of Probability
2011, Vol. 39, No. 4, 1365–1406
DOI: 10.1214/10-AOP584
© Institute of Mathematical Statistics, 2011

RECONSTRUCTION FOR THE POTTS MODEL[1]

BY ALLAN SLY

Microsoft Research

The reconstruction problem on the tree has been studied in numerous contexts including statistical physics, information theory and computational biology. However, rigorous reconstruction thresholds have only been established in a small number of models. We prove the first exact reconstruction threshold in a nonbinary model establishing the Kesten–Stigum bound for the 3-state Potts model on regular trees of large degree. We further establish that the Kesten–Stigum bound is not tight for the q-state Potts model when $q \geq 5$. Moreover, we determine asymptotics for these reconstruction thresholds.

1. Introduction.

1.1. *Preliminaries.* We begin by giving a general description of broadcast (or Markov) models on trees and the reconstruction problem. The broadcast model on a tree T is a model in which information is sent from the root ρ across the edges, which act as noisy channels, to the leaves of T. For some given finite set of states $\mathcal{C}$, a configuration on T is an element of $\mathcal{C}^T$, that is an assignment of a state $\mathcal{C}$ to each vertex. We will denote the elements of $\mathcal{C}$ as $\{1, \dots, q\}$ and $q = |\mathcal{C}|$ as the number of states. The broadcast model is a probability distribution on configurations defined as follows. Some $|\mathcal{C}| \times |\mathcal{C}|$ probability transition matrix M is chosen as the noisy channel on each edge. The spin σ_ρ is chosen from $\mathcal{C}$ according to some initial distribution and is then propagated along the edges of the tree according to the transition matrix M. That is, if vertex u is the parent of v in the tree then the spin at v is defined according to the probabilities

$$P(\sigma_v = j | \sigma_u = i) = M_{i,j}.$$

The focus of this paper is on the symmetric channels which are given by transition matrices of the form

$$M_{i,j} = \begin{cases} 1 - p, & \text{if } i = j, \\ \dfrac{p}{q-1}, & \text{otherwise,} \end{cases}$$

where $0 < p \leq 1$. The state of the root is chosen according to the uniform distribution on $\mathcal{C}$.

Received July 2009; revised July 2010.

[1]Supported by NSF Grants DMS-05-28488 and DMS-05-48249 and ONR Grant N0014-07-1-05-06.

MSC2010 subject classification. 60K35.

Key words and phrases. Reconstruction problem on the tree, Potts model.

The symmetric channel corresponds to the statistical physics model known as the q-state Potts model on the tree. The Potts model weights configurations according to the Hamiltonian $H(\sigma) = \sum_{(u,v)\in E} 1_{\{\sigma_u=\sigma_v\}}$ which counts the number of edges in which the states on each side are equal. On a finite tree, the probability distribution is given by

$$P(\sigma) = \frac{1}{Z}\exp\biggl(\beta \sum_{(u,v)\in E} 1_{\{\sigma_u=\sigma_v\}}\biggr),$$

where Z is a normalising constant. On an infinite tree, more than one Gibbs measure may exist, the symmetric channel corresponds to the free Gibbs measure. The two models coincide when $1 - p = \frac{e^\beta}{e^\beta+q-1}$. It will be convenient to parameterize the symmetric channel by its second largest eigenvalue by absolute value (i.e., either the second eigenvalue or the last eigenvalue, whichever is larger). It is given by

$$\lambda = \lambda(M) = 1 - \frac{pq}{q-1} = \frac{e^\beta - 1}{e^\beta + q - 1}$$

and takes values in the interval $[-\frac{1}{q-1}, 1)$. The special case of proper colorings corresponds to $\lambda = -\frac{1}{q-1}$. In line with the terminology for the Potts model, we will say the channel is ferromagnetic when $\lambda > 0$ and anti-ferromagnetic when $\lambda < 0$.

We will restrict our attention to d-ary trees, that is the infinite rooted tree where every vertex has d offspring. Let $\sigma(n)$ denote the spins at distance n from the root and let $\sigma^i(n)$ denote $\sigma(n)$ conditioned on $\sigma_\rho = i$.

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

$$\limsup_n d_{\mathrm{TV}}(\sigma^i(n), \sigma^j(n)) > 0,$$

where d_{TV} is the total variation distance. When the limsup is 0, we will say the model has *nonreconstruction* on T.

Nonreconstruction is equivalent to the mutual information between $\sigma_\rho = \sigma(0)$ and $\sigma(n)$ going to 0 as n goes to infinity and also to $\{\sigma(n)\}_{n=1}^\infty$ having a trivial tail sigma-field. In the statistical physics, nomenclature nonreconstruction is equivalent to the free measure being extremal, that is not a convex combination of two other Gibbs measures. More equivalent formulations are given in [20], Proposition 2.1. In contrast, consider the uniqueness property of a Gibbs measure.

DEFINITION 2. We say that a model has *uniqueness* on a tree T if

$$\limsup_n \sup_{A,B} d_{\mathrm{TV}}\bigl(P(\sigma_\rho = \cdot|\sigma(n) = A), P(\sigma_\rho = \cdot|\sigma(n) = B)\bigr) > 0,$$

where the supremum is over all configurations A, B on the vertices at distance n from the root.

Reconstruction implies nonuniqueness and is a strictly stronger condition. Essentially uniqueness says that there is some configuration on the leaves which provides information on the root while reconstruction says that a typical configuration on the leaves provides information on the root.

1.2. *Background.* For a given parameterized collection of models, the key question in studying reconstruction is finding which models have reconstruction, which typically involves finding a threshold. The reconstruction problem naturally arises in biology, information theory and statistical physics and involves the trade off between increasing numbers of leaves with increasingly noisy information as the distance from the root to the leaves increases. In the case of the Potts model, this is the question of for which λ is there reconstruction for each choice of q and d. Proposition 12 of [18] implies that for each q and d there exist $\lambda^- < 0 < \lambda^+$ such that there is nonreconstruction when $\lambda \in (-\lambda^-, \lambda^+)$ and reconstruction when $\lambda \in [-\frac{1}{q-1}, \lambda^-) \cup (\lambda^+, 1)$. The result does not say what happens when $\lambda \in \{\lambda^-, \lambda^+\}$.

The most general result on reconstruction is the Kesten–Stigum bound [13] which says that reconstruction holds when $\lambda^2 d > 1$ which in our parameterization says that $\lambda^+ \le d^{-1/2}$ and $\lambda^- \ge -d^{-1/2}$. When $d\lambda^2 > 1$ it is possible to asymptotically reconstruct information on the root using only enumerations of each type of spin at the leaves (census reconstruction) and without using the information on their positions on the leaves.

The simplest collection of models is the binary (2-state) symmetric channel which is defined on two states and corresponds to the Ising model on the tree with no external field. It was shown in [4] and later in [9, 12] that this channel has reconstruction if and only if $d\lambda^2 > 1$, that is, the Kesten–Stigum bound is sharp. Before this paper, exact reconstruction thresholds had only been calculated in the binary symmetric channel and binary asymmetric channels with sufficiently small asymmetry [5] where the Kesten–Stigum is also sharp. While it was once conjectured that the Kesten–Stigum bound was tight for all channels, it was shown [18, 20] that the Kesten–Stigum bound is not the bound for reconstruction in the binary-asymmetric model with sufficiently large asymmetry or in the ferromagnetic Potts model with $q \ge 18$. For general Potts models, [21] showed nonreconstruction when

$$\frac{qd\lambda^2}{2 + (q-2)\lambda} \le 1$$

and these bounds were improved in [15] and recently for $q \in \{3, 4, 5\}$ in [3, 10]. None of these results are tight. Several recent results deal with the special case of proper colorings which is now known to good accuracy. By analyzing a simple reconstruction algorithm, reconstruction was shown to hold when $d \ge q[\log q +$

$\log\log q + 1 + o(1)]$ see [21, 24]. The tightest bounds for nonreconstruction are $d \le q[\log q + \log\log q + 1 - \log 2 + o(1)]$ established by [25], the difference between the upper and lower bounds is just $q \log 2$.

After the results of [18, 20], it was for a while thought that the Kesten–Stigum bound may be tight only in 2-state channels. This intuition was overturned by deep insights coming from statistical physics. Leading experts in the study of random constraint satisfaction problems Mézard and Montanari [16] analyzed the reconstruction model for symmetric channels. The reconstruction problem plays a key role in the analysis of replica-symmetry breaking transitions for distributions such as random k-SAT and random colorings of random graphs and the efficiency of belief-propagation algorithms. Using techniques developed to analyze transitions in glassy systems, they made a series of conjectures for the symmetric channels.

CONJECTURE 1 ([16]). *The Kesten–Stigum bound is tight for the ferromagnetic symmetric channel when $q \le 4$ and is not tight when $q \ge 5$. In the antiferromagnetic model, the Kesten–Stigum bound is tight when $q \le 3$ and not tight when $q \ge 4$.*

As this conjecture was based partially on numerical evidence, they qualified it by stating that it might not hold for large d. This paper confirms much of the predicted picture.

1.3. *Main results.* Our results rigorously confirm most of the picture predicted by Mezard and Montanari [16]. We determine the asymptotic values of the thresholds calculating the limit of $d^{1/2}\lambda^{\pm}$ for large d. We give a complete answer to the question of whether the Kesten–Stigum bound is tight for large d except in the case of $q = 4$. When $q = 3$, we show that for large enough d the Kesten–Stigum bound is tight. This is the first time a reconstruction threshold has been rigorously established in a nonbinary model. Conversely, when $q \ge 5$, the Kesten–Stigum bound is never sharp. Our proof also gives significant new insight into the reasons why this a transition occurs between $q = 3$ and $q = 5$.

THEOREM 1.1. *When $q = 3$, there exists a $d_{\min}$ such that for $d \ge d_{\min}$ the Kesten–Stigum bound is sharp for both the ferromagnetic and antiferrmagnetic channels, that is, $\lambda^+(d) = d^{-1/2}$ and $\lambda^-(d) = -d^{-1/2}$. Furthermore, there is nonreconstruction at the Kesten–Stigum bound, when $\lambda = \lambda^+$ or $\lambda = \lambda^-$.*

Conversely, when $q \ge 5$, the Kesten–Stigum bound is never sharp.

THEOREM 1.2. *When $q \ge 5$ for every d, the Kesten–Stigum bound is not sharp, that is, $\lambda^+ < d^{-1/2}$ and $\lambda^- > -d^{-1/2}$.*

1.3.1. *Asymptotic results.* When the Kesten–Stigum bound is not sharp, we are not able to exactly compute the threshold, doing so involves finding a nontrivial fixed point of an equation of vector-valued distributions. Nonetheless, we are able to give precise asymptotics for the thresholds for fixed q as d goes to infintiy. In light of the Kesten–Stigum bound, it makes sense to consider $d^{1/2}\lambda^{\pm}$. When $q \geq 5$, however, the limit is strictly bounded away from from ± 1.

THEOREM 1.3. *When* $q \geq 5$,

$$\lim_{d\to\infty} d^{1/2}\lambda^{+} = C_q,$$

$$\lim_{d\to\infty} d^{1/2}\lambda^{-} = -C_q,$$

where C_q *is a constant strictly less than* 1.

Of course when $q = 3$, we have that $d^{1/2}\lambda^{\pm} = \pm 1$ for large d. The value C_q is given by the maximum of a function which we shall describe in the proof.

1.4. *Applications of reconstruction.* Our results add to the knowledge of the reconstruction problem which has numerous applications from various fields. In computational biology, the broadcast model is the main model for the evolution of base pairs of DNA. In phylogenetic reconstruction, the goal is to reconstruct the ancestry tree of a collection of species given the genetic data of the present-day species. The reconstruction threshold here determines the possibility of reconstructing ancestral DNA sequences from a known phylogenetic tree. Establishing a conjecture of Mike Steel, it was shown that the number of samples required for phylogenetic reconstruction undergoes a phase transition at the reconstruction threshold for the binary symmetric channel [7, 19].

The reconstruction threshold on trees is believed to play a critical role in the dynamic phase transitions in certain glassy systems given by random constraint satisfaction problems such as random k-SAT and the anti-ferromagnetic Potts model on random graphs. We will briefly describe the broad picture conjectured by physicists about such systems [14, 26], generally without rigorous proof. Recently, much of this picture has been established by [1]. The theory relates to the structure and connectivity of the set of satisfying configurations of the distribution, with the topology given by the hamming distance on the space of configurations. At "high temperatures" or low densities of constraints, the Gibbs measure places all but an exponentially small fraction of its weight in a single "connected cluster." As the temperature decreases, there is a threshold called the "dynamical replica symmetry breaking threshold" at which the set supporting most of the measure splits into exponentially many smaller clusters. The clusters are each well separated from each other and each contain an exponentially small amount of the measure. This threshold is believed to exactly correspond to the reconstruction threshold on the corresponding tree model but this has not yet been established.

Recently, [1] proved that for random colorings on Erdős–Rényi random graphs with average connectivity d when $(1+o(1))q\log q \le d \le (2-o(1))q\log q$ the space of solutions indeed breaks into exponentially many small clusters while when $(1-o(1))q\log q \ge d$ most of the mass is contained in a giant component. Similar results were also proved for K-SAT and other models. This phase transition from the connected to the clustered regimes corresponds to known bounds on the reconstruction threshold for proper colorings on trees [21, 24, 25]. It is believed that message passing algorithms such as belief-propagation work up to this threshold. Our results, therefore, have direct implications for phase transitions and message passing algorithms on anti-ferromagnetic Potts models.

One can also ask whether the reconstruction threshold on trees corresponds to the reconstruction threshold on locally treelike graphs. This problem was investigated by Gerschenfeld and Montanari [11] who showed that reconstruction on trees is equivalent to reconstruction on random graphs when a certain "replica condition" holds. In a large number of "frustrated" systems, this was shown to hold including for the antiferromagnetic Potts model at nonzero temperature on random graphs and in [17] it was established for colorings.

The reconstruction threshold is known to play an important role in the efficiency of the Glauber dynamics on trees and random graphs. In [2], it was shown that the mixing time for the the Glauber dynamics on trees is $n^{1+\Theta(1)}$ when the model has reconstruction which is slower than at higher temperatures when the mixing time is $O(n\log n)$. In the case of the Ising model this is tight, the mixing time is $O(n\log n)$ when $d\lambda^2 < 1$.

Local MCMC algorithms are conjectured to be efficient up to the reconstruction threshold for sampling random colorings on random graphs but experience an exponential slowdown beyond it [14]. This is to be expected since a local MCMC algorithm can not move between clusters each of which has exponentially small probability. Rigorous proofs of rapid mixing of MCMC algorithms, such as the Glauber dynamics, fall a long way behind. For colorings of random regular graphs, results of [8] imply rapid mixing when $q \ge 1.49d$, well below the reconstruction threshold and even the uniqueness threshold. Even less is known for Erdős–Rényi random graphs as almost all MCMC results are given in terms of the maximum degree which in this case grows with n. Polynomial time mixing of the Glauber dynamics has been shown [22] for a constant number of colors in terms of d, the average connectivity.

1.5. *Proof sketch.* The proof analyzes a quantity denoted by x_n. One interpretation of x_n is that if we guess the value of σ_ρ according to its postier distribution given $\sigma(n)$ then x_n is the probability of being correct minus $\frac{1}{q}$, which is the chance of being correct by simply guessing randomly. More formally, if Z is a $\mathcal{C}$-valued random variable with distribution given by $P(Z=i|\sigma(n)) = P(\sigma_\rho = i|\sigma(n))$ then $x_n = P(Z=\sigma_\rho) - \frac{1}{q}$. Our analysis is similar to the expansion of [5] but with more precise estimates derived by establishing certain concentration results. Such expansions go back to [6] in the context of spin-glasses.

We show that x_n is always positive and that nonreconstruction is equivalent to

$$\lim_{n\to\infty} x_n = 0.$$

In general, finding the reconstruction threshold requires understanding recursive equations of vector-valued posterior distributions cf. [16]. However, when x_n is small, these recursive relationships become approximately too linear. Using Taylor series expansions and concentration estimates, we establish that for small x_n

$$x_{n+1} = d\lambda^2 x_n + \bigl(1 + o(1)\bigr)\frac{d(d-1)}{2}\frac{q(q-4)}{q-1}\lambda^4 x_n^2. \tag{1}$$

A key role is played by the sign of $q - 4$. When $q \geq 5$, it is positive and this allows us to show that if $d\lambda^2$ is sufficiently close to 1 then x_n does not converge to 0 and hence there is reconstruction beyond the Kesten–Stigum bound.

However, when $q = 3$ the second order term is negative. Suppose we could establish that x_n is eventually small when $d\lambda^2 \leq 1$. Then equation (1) implies that x_n converges to 0 which establishes nonreconstruction. Unfortunately for small d, we are not able to show that x_n becomes sufficiently small to apply this argument.

A key new ingredient to our analysis is studying the problem when d is large and the interactions between spins become very weak but there are many of them. Using the central limit theorem, we approximate this collection of small independent interactions to show that

$$x_{n+1} \approx g_q(d\lambda^2 x_n)$$

for some increasing function g_q. When $q = 3$ for all $0 < s < 1$, the function satisfies $g_3(s) < s$. Using this estimate for large enough d, it is established that x_n becomes arbitrarily small. Crucially combining our analysis of what happens when d is large and x_n is small, we prove nonreconstruction for large enough d. When $q = 4$ for all $0 < s < 1$, the function also satisfies $g_4(s) < s$ while when $q \geq 5$ the equation $g_5(s) = s$ has nonzero solutions. The function $g_q(s)$ determines the limiting value of x_n, a consequence of which is Theorem 1.3.

2. Proofs. We introduce the notation we use in the proofs. Denote the colors by $\mathcal{C} = \{1, \dots, q\}$ and let T be the infinite d-ary tree rooted at ρ. Let $u_1, \dots, u_d$ be the children of ρ and for a vertex $v \in T$ let T_v denote the subtree of descendants of v (including v). We will use the convention that i will denote an element of $\mathcal{C}$ and j will be an element of $\{1, \dots, d\}$ corresponding to a child of ρ. Let σ denote a random configuration given by the symmetric channel with transition matrix given by

$$M_{i,i'} = \begin{cases} 1 - p, & \text{if } i = i', \\ \dfrac{p}{q-1}, & \text{otherwise,} \end{cases}$$

where $0 < p \le 1$. Rather than looking at the unconditioned configurations σ we will work mainly with configurations where the spin at the root is conditioned; we let σ^i denote a random configuration according to the symmetric channel conditioned on $\sigma^i_\rho = i$. Let λ denote the second eigenvalue of M which is given by

$$\lambda = \lambda(M) = 1 - \frac{pq}{q-1}. \tag{2}$$

In light of the Kesten–Stigum bound, we will always assume that $d\lambda^2 \le 1$.

Let $S(n)$ denote the vertices on level n, $\{v \in T : d(v, \rho) = n\}$, let $\sigma(n) := \sigma_{S(n)}$ denote the spins on $S(n)$ and let $\sigma_j(n)$ denote the spins in $S(n) \cap T_{u_j}$. For a configuration A on $S(n)$, define the posterior function f_n as

$$f_n(i, A) = P\bigl(\sigma_\rho = i | \sigma(n) = A\bigr).$$

By the recursive nature of the tree for a configuration A on $S(n+1) \cap T_{u_j}$, we also have (with a slight abuse of notation) that

$$f_n(i, A) = P\bigl(\sigma_{u_j} = i | \sigma_j(n+1) = A\bigr).$$

Now define $X_i(n) = X_i$ by

$$X_i(n) = f_n(i, \sigma(n)).$$

These random variables are a deterministic function of the random configuration $\sigma(n)$ of the leaves which gives the posterior probability that the root is in state i. Recall that a collection of random variables are exchangeable if their distribution is invariant under permutations. By symmetry, the X_i are exchangeable. Now we define two random variables

$$X^+ = X^+(n) = f_n(1, \sigma^1(n))$$

and

$$X^- = X^-(n) = f_n(2, \sigma^1(n)).$$

We will establish nonreconstruction (resp., reconstruction) by showing that X^+ and X^- both converge (resp., do not converge) to $\frac{1}{q}$ in probability as n goes to infinity. By symmetry, we have

$$f_n(i_2, \sigma^{i_1}(n)) \stackrel{d}{=} \begin{cases} X^+, & i_1 = i_2, \\ X^-, & \text{otherwise}, \end{cases}$$

and the set $\{f_n(i, \sigma^1(n)) : 2 \le i \le q\}$ is exchangeable. Moreover, they are conditionally exchangeable given $f_n(1, \sigma^1(n))$.

Now define

$$Y_{ij} = Y_{ij}(n) = f_n\bigl(i, \sigma^1_j(n+1)\bigr).$$

This is none other than the posterior probability that $\sigma_{u_j} = i$ given the random configuration $\sigma^1_j(n+1)$ on the spins in $S(n+1) \cap T_{u_j}$. Conditional on the spin at the root, the spins in the subtrees T_{u_j} are conditionally independent for $j = 1, \dots, d$. Taking advantage of this and the symmetries of the model, the following proposition is immediate.

PROPOSITION 2.1. *The Y_{ij} satisfy the following properties*:

- *The random vectors $Y_j = (Y_{1j}, \dots, Y_{qj})$ are independent for $j = 1, \dots, d$.*
- *Conditional on $\sigma^1_{u_j}$ the random variable $Y_{\sigma^1_{u_j} j}$ is equal in distribution to $X^+(n)$ while for $i \neq \sigma^1_{u_j}$ the random variables Y_{ij} are equal in distribution to $X^-(n)$.*
- *Further given $\sigma^1_{u_j}$ and $Y_{\sigma^1_{u_j} j}$ the collection of random variables $\{Y_{ij}\}_{i \neq \sigma^1_{u_j}}$ are conditionally exchangeable.*

The key method of this paper will be to analyze the relation between the distributions $X^+(n)$ and $X^+(n+1)$ using the recursive structure of the tree. The recursive analysis of posterior distributions is a central tool in analyzing reconstruction problems, our approach is similar to [5] but we will make more precise estimates by deriving concentration results. Suppose A is a configuration on $S(n+1)$ and let A_j be its restriction to $T_{u_j} \cap S(n+1)$. Applying Bayes theorem, we have that

$$
\begin{aligned}
f_{n+1}(1, A) &= \frac{\prod_{j=1}^d (M_{11} f_n(1, A_j) + \sum_{l \neq 1} M_{1l} f_n(l, A_j))}{\sum_{i=1}^q \prod_{j=1}^d (M_{ii} f_n(i, A_j) + \sum_{l \neq i} M_{il} f_n(l, A_j))} \\
&= \frac{\prod_{j=1}^d (M_{12} + (M_{11} - M_{12}) f_n(1, A_j))}{\sum_{i=1}^q \prod_{j=1}^d (M_{12} + (M_{11} - M_{12}) f_n(i, A_j))} \qquad (3) \\
&= \frac{\prod_{j=1}^d (1 + \lambda q (f_n(1, A_j) - 1/q))}{\sum_{i=1}^q \prod_{j=1}^d (1 + \lambda q (f_n(i, A_j) - 1/q))},
\end{aligned}
$$

where the second equality is a consequence of the fact that $\sum_{i=1}^q f_n(i, A_j) = 1$ and the symmetry of M and the final equality follows from equation (2) since

$$
M_{12} + \frac{1}{q}(M_{11} - M_{12}) = M_{12} + \frac{1}{q}\bigl(1 - (q-1)M_{12} - M_{12}\bigr) = \frac{1}{q}
$$

and

$$
M_{11} - M_{12} = 1 - q M_{12} = \lambda.
$$

Conditioning the root to be 1 and letting $A = \sigma^1(n+1)$, we have that

$$
X^+(n+1) = \frac{Z_1}{\sum_{i=1}^k Z_i}, \qquad (4)
$$

where

$$
Z_i = Z_i(n) = \prod_{j=1}^d \left(1 + \lambda q \left(Y_{ij}(n) - \frac{1}{q}\right)\right). \qquad (5)
$$

Equation (4) will be our major tool for recursively analyzing the reconstruction problem.

2.1. *Basic identities*. Denote

$$x_n = E\left(X^+(n) - \frac{1}{q}\right) = Ef_n(1, \sigma^1(n)) - \frac{1}{q}$$

and

$$z_n = E\left(X^+(n) - \frac{1}{q}\right)^2 = E\left(f_n(1, \sigma^1(n)) - \frac{1}{q}\right)^2.$$

As discussed in the Introduction, the main proof relies on analyzing recursions of x_n. One interpretation of x_n is that if you were to choose a state randomly according to the posterior distribution, then the probability of correctly guessing the root is $\frac{1}{q} + x_n$ [16]. This quantity (up to a constant factor) was also studied in [5] for the binary asymmetric channel. The following lemma, which can be viewed as the analogue of Lemma 1 of [5], allows us to relate the first and second moments of X^+.

LEMMA 2.2. *The following relations hold*:

$$x_n + \frac{1}{q} = EX^+ = E\sum_{i=1}^{q}(X_i(n))^2 = E(X^+(n))^2 + (q-1)E(X^-(n))^2$$

and

$$x_n = E\sum_{i=1}^{q}\left(X_i(n) - \frac{1}{q}\right)^2 = E\left(X^+(n) - \frac{1}{q}\right)^2 + (q-1)E\left(X^-(n) - \frac{1}{q}\right)^2 \geq z_n.$$

PROOF. From the definition of conditional probabilities and of f_n and the fact that $P(\sigma_\rho = 1) = \frac{1}{q}$, we have that

$$\begin{aligned}
EX^+(n) &= Ef_n(1, \sigma^1(n)) \\
&= \sum_A f_n(1, A)P\big(\sigma(n) = A|\sigma_\rho = 1\big) \\
&= \sum_A \frac{P(\sigma(n) = A, \sigma_\rho = 1)}{P(\sigma_\rho = 1)} f_n(1, A) \\
&= q\sum_A P\big(\sigma(n) = A\big) f_n(1, A)^2 \\
&= qE(X_1(n))^2 \\
&= E\sum_{i=1}^{q}(X_i(n))^2
\end{aligned}$$

and

$$E\sum_{i=1}^{k}\left(X_i(n)-\frac{1}{q}\right)^2 = E\sum_{i=1}^{q}(X_i(n))^2 - \frac{2}{q}E\sum_{i=1}^{q}X_i(n) + \frac{1}{q} = EX^+ - \frac{1}{q}.$$

Conditional on σ_ρ, we have that $X_{\sigma_\rho}(n)$ is distributed as $X^+(n)$ and for $i \neq \sigma_\rho$ we have that $X_i(n)$ is distributed as $X^-(n)$. It follows that

$$E\sum_{i=1}^{q}(X_i(n))^2 = E(X^+(n))^2 + (q-1)E(X^-(n))^2$$

and

$$E\sum_{i=1}^{q}\left(X_i(n)-\frac{1}{q}\right)^2 = E\left(X^+(n)-\frac{1}{q}\right)^2 + (q-1)E\left(X^-(n)-\frac{1}{q}\right)^2$$

which completes the result. $\square$

Define $\hat{\sigma}_\rho(n)$ to be the maximum likelihood estimator of σ_ρ given $\sigma(n)$ which is given by

$$\hat{\sigma}_\rho(n) := \arg\max_i X_i(n),$$

where in the case that multiple states maximize the likelihood the estimator chooses randomly between these states. This estimator maximizes the probability of correctly reconstructing the root. Define the probability of correct reconstruction as

$$p_n := P\big(\sigma_\rho = \hat{\sigma}_\rho(n)\big) = E\max_{1\le i\le q} X_i(n).$$

This represents the probability of correctly reconstructing the spin at the root using the maximum likelihood estimator which maximizes the probability of correctly determining the root. Since $\sigma(n)$ is a Markov process, p_n is clearly decreasing.

LEMMA 2.3. *We have that*

$$x_n \le p_n - \frac{1}{q} \le x_n^{1/2}.$$

PROOF. The inequality $x_n + \frac{1}{q} \le p_n$ was shown in [16] by noting that the algorithm that chooses σ_ρ randomly according to probabilities X_i is correct with probability $x_n + \frac{1}{q}$. By the Cauchy–Schwarz inequality and Lemma 2.2,

$$\begin{aligned} p_n = E\max_i X_i &\le \frac{1}{q} + E\max_i\left|X_i - \frac{1}{q}\right| \le \frac{1}{q} + \left(E\max_i\left(X_i-\frac{1}{q}\right)^2\right)^{1/2} \\ &\le \frac{1}{q} + \left(E\sum_{i=1}^{q}\left(X_i-\frac{1}{q}\right)^2\right)^{1/2} = \frac{1}{q} + x_n^{1/2} \end{aligned}$$

as required. $\square$

The following corollary of Lemmas 2.2 and 2.3 justifies our focus on x_n.

COROLLARY 2.4. *We have that $x_n \geq 0$ and the condition*

$$\lim_n x_n = 0$$

is equivalent to nonreconstruction.

PROOF. Lemma 2.2 implies that $x_n \geq z_n \geq 0$. By Lemma 2.2, x_n converging to 0 is equivalent to

$$\sum_{i=1}^{k} E\left(X_i(n) - \frac{1}{q}\right)^2 \to 0$$

which is equivalent to the posteriors converging to the stationary distribution which is in turn equivalent to reconstruction [20]. $\square$

Using the identities from Lemma 2.2, we calculate the means and covariances of the Y_{ij}.

LEMMA 2.5. *For each $1 \leq j \leq q$, the following hold:*

$$E\left(Y_{1j} - \frac{1}{q}\right) = \lambda x_n, \qquad E\left(Y_{1j} - \frac{1}{q}\right)^2 = \lambda z_n + \frac{1}{q}(1-\lambda)x_n. \tag{6}$$

For $i \neq 1$ we have that

$$E\left(Y_{ij} - \frac{1}{q}\right) = -\frac{\lambda x_n}{q-1}, \qquad E\left(Y_{ij} - \frac{1}{q}\right)^2 = \frac{1}{q}\left(1 + \frac{\lambda}{q-1}\right)x_n - \frac{\lambda}{q-1}z_n \tag{7}$$

and

$$E\left(Y_{1j} - \frac{1}{q}\right)\left(Y_{ij} - \frac{1}{q}\right) = -\frac{\lambda}{q-1}z_n - \frac{1-\lambda}{q(q-1)}x_n. \tag{8}$$

When $1 < i_1 < i_2 \leq q$,

$$E\left(Y_{i_1 j} - \frac{1}{q}\right)\left(Y_{i_2 j} - \frac{1}{q}\right) = \frac{1}{(q-1)(q-2)}\left[2\lambda z_n - \frac{1}{q}(q-2+2\lambda)x_n\right]. \tag{9}$$

PROOF. By Proposition 2.1, if $\sigma_{u_j}^1 = 1$ then Y_{1j} is distributed according to $X^+(n)$ otherwise it is distributed according to $X^-(n)$. By equation (2), we have that

$$P(\sigma_{u_j}^1 = 1) = \frac{1 + \lambda(q-1)}{q}.$$

Noting that $\sum_{i=1}^{q} Y_{ij} = 1$ it follows that $EX^+(n) + (q-1)EX^-(n) = 1$ and so $E(X^-(n) - \frac{1}{q}) = -\frac{x_n}{q-1}$. It follows that

$$\begin{aligned}
E\left(Y_{1j} - \frac{1}{q}\right) &= P(\sigma^1_{u_j} = 1)E\left(X^+(n) - \frac{1}{q}\right) + \left(1 - P(\sigma^1_{u_j} = 1)\right)E\left(X^-(n) - \frac{1}{q}\right) \\
&= \frac{1+\lambda(q-1)}{q}x_n + \left(1 - \frac{1+\lambda(q-1)}{q}\right)\frac{-x_n}{q-1} \\
&= \lambda x_n.
\end{aligned}$$

Using Lemma 2.2 and Proposition 2.1 we have that

$$\begin{aligned}
E\left(Y_{1j} - \frac{1}{q}\right)^2 &= P(\sigma^1_{u_j} = 1)E\left(X^+(n) - \frac{1}{q}\right)^2 \\
&\quad + \left(1 - P(\sigma^1_{u_j} = 1)\right)E\left(X^-(n) - \frac{1}{q}\right)^2 \\
&= \frac{1+\lambda(q-1)}{q}z_n + \left(1 - \frac{1+\lambda(q-1)}{q}\right) \\
&\qquad \times \frac{1}{q-1}\left[E\left(X^+(n) - \frac{1}{q}\right) - E\left(X^+ - \frac{1}{q}\right)^2\right] \\
&= \lambda z_n + \frac{1}{q}(1-\lambda)x_n
\end{aligned} \tag{10}$$

which establishes equation (6). Now since $\sum_{l=1}^{q} Y_{lj} = 1$ and since by Proposition 2.1 we have that $Y_{2j}, \ldots, Y_{qj}$ are exchangeable, for $i \neq 1$ we have that

$$\begin{aligned}
E\left(Y_{ij} - \frac{1}{q}\right) &= \frac{1}{q-1}\sum_{l=2}^{q} E\left(Y_{lj} - \frac{1}{q}\right) \\
&= -\frac{1}{q-1}E\left(Y_{1j} - \frac{1}{q}\right) \\
&= -\frac{\lambda x_n}{q-1}.
\end{aligned}$$

Again using Lemma 2.2 and the exchangeability of $Y_{2j}, \ldots, Y_{qj}$, we have that

$$\begin{aligned}
E\left(Y_{ij} - \frac{1}{q}\right)^2 &= \frac{1}{q-1}\left[-E\left(Y_{1j} - \frac{1}{q}\right)^2 + \sum_{l=1}^{q} E\left(Y_{lj} - \frac{1}{q}\right)^2\right] \\
&= \frac{1}{q-1}\left[-\left(\lambda z_n + \frac{1}{q}(1-\lambda)x_n\right) + x_n\right] \\
&= \frac{1}{q}\left(1 + \frac{\lambda}{q-1}\right)x_n - \frac{\lambda}{q-1}z_n.
\end{aligned}$$

By the fact that $\sum_{l=2}^{q}(Y_{lj}-\frac{1}{q})=-(Y_{1j}-\frac{1}{q})$,

$$E\Big(Y_{1j}-\frac{1}{q}\Big)\Big(Y_{ij}-\frac{1}{q}\Big)=\frac{1}{q-1}\sum_{l=2}^{q}E\Big(Y_{1j}-\frac{1}{q}\Big)\Big(Y_{lj}-\frac{1}{q}\Big)$$

$$=-\frac{1}{q-1}E\Big(Y_{1j}-\frac{1}{q}\Big)^2=-\frac{\lambda}{q-1}z_n-\frac{1-\lambda}{q(q-1)}x_n,$$

where the third equality follows from equation (10). Finally,

$$\begin{aligned}
&E\Big(Y_{i_1 j}-\frac{1}{q}\Big)\Big(Y_{i_2 j}-\frac{1}{q}\Big)\\
&\quad=\frac{1}{(q-1)(q-2)}E\bigg[\Big(Y_{1j}-\frac{1}{q}\Big)^2-\sum_{l=2}^{q}\Big(Y_{lj}-\frac{1}{q}\Big)^2\bigg]\\
&\quad=\frac{1}{(q-1)(q-2)}\bigg[\Big(\lambda z_n+\frac{1}{q}(1-\lambda)x_n\Big)\\
&\qquad\qquad-(q-1)\Big(\frac{1}{q}\Big(1+\frac{\lambda}{q-1}\Big)x_n-\frac{\lambda}{q-1}z_n\Big)\bigg]\\
&\quad=\frac{1}{(q-1)(q-2)}\bigg[2\lambda z_n-\frac{1}{q}(q-2+2\lambda)x_n\bigg]. \qquad\square
\end{aligned}$$

2.2. *Taylor series bounds*. In the following lemma, we calculate precise estimates of the expected values of monomials of the Z_i by expanding them using Taylor series approximations. This allows us to make estimates using an expansion of equation (4).

LEMMA 2.6. *For each positive integer k, there exists a $C=C(q,k)$ not depending on λ or d such that for each $0\le k_1,\dots,k_q\le k$,*

$$E\prod_{i=1}^{q}Z_i^{k_i}\le C$$

and

$$\left|E\prod_{i=1}^{q}Z_i^{k_i}-1-d\left(E\prod_{i=1}^{q}\Big(1+\lambda q\Big(Y_{i1}-\frac{1}{q}\Big)\Big)^{k_i}-1\right)\right|\le Cx_n^2$$

and

$$\begin{aligned}
&\left|E\prod_{i=1}^{q}Z_i^{k_i}-1-d\left(E\prod_{i=1}^{q}\Big(1+\lambda q\Big(Y_{i1}-\frac{1}{q}\Big)\Big)^{k_i}-1\right)\right.\\
&\qquad\left.-\frac{d(d-1)}{2}\left(E\prod_{i=1}^{q}\Big(1+\lambda q\Big(Y_{i1}-\frac{1}{q}\Big)\Big)^{k_i}-1\right)^2\right|\le Cx_n^3.
\end{aligned}$$

PROOF. Recall that

$$Z_i = Z_i(n) = \prod_{j=1}^{d}\left(1 + \lambda q\left(Y_{ij}(n) - \frac{1}{q}\right)\right)$$

so each Z_i is a product of independent and identically distributed terms and that

$$E\prod_{i=1}^{q} Z_i^{k_i} = \left(E\prod_{i=1}^{q}\left(1 + \lambda q\left(Y_{i1}(n) - \frac{1}{q}\right)\right)^{k_i}\right)^{d}.$$

As such, we begin with a simple bound on $(1+y)^d$ using Taylor series. Suppose that $d|y| \le C'$ for some constant $C' > 0$. Then we have that

$$\begin{aligned}\left|(1+y)^d - \sum_{i=0}^{\ell}\binom{d}{i} y^i\right| &\le \sum_{i=\ell+1}^{d}\binom{d}{i}|y|^i \\ &\le \sum_{i=\ell+1}^{\infty}\frac{d^i}{i!}|y|^i \\ &= e^{d|y|} - \sum_{i=0}^{\ell}\frac{(d|y|)^i}{i!} \\ &\le e^{C'}|dy|^{\ell+1},\end{aligned} \tag{11}$$

where the third inequality follows by Taylor's theorem since $\max_{x\le C'} \frac{d^{\ell+1}}{dx^{\ell+1}} e^x = e^{C'}$.

Suppose that $s_1, \dots, s_q$ are nonnegative integers. If for some ℓ, $s_\ell \ge 2$, then since by definition $0 \le Y_{ij} \le 1$, by Lemma 2.2,

$$\left|E\prod_{i=1}^{q}\left(Y_{i1} - \frac{1}{q}\right)^{s_i}\right| \le E\left(Y_{\ell 1} - \frac{1}{q}\right)^2 \le x_n. \tag{12}$$

If for distinct integers ℓ, ℓ', $s_\ell = s_{\ell'} = 1$, then again by Lemma 2.2,

$$\begin{aligned}\left|E\prod_{i=1}^{q}\left(Y_{i1} - \frac{1}{q}\right)^{s_i}\right| &\le E\left|\left(Y_{\ell 1} - \frac{1}{q}\right)\left(Y_{\ell' 1} - \frac{1}{q}\right)\right| \\ &\le E\left[\left(Y_{\ell 1} - \frac{1}{q}\right)^2 + \left(Y_{\ell' 1} - \frac{1}{q}\right)^2\right] \le x_n.\end{aligned} \tag{13}$$

Finally if $s_\ell = 1$ and $s_i = 0$ for all $i \ne \ell$, then by Lemma 2.5,

$$\left|E\prod_{i=1}^{q}\left(Y_{i1} - \frac{1}{q}\right)^{s_i}\right| = \left|EY_{\ell 1} - \frac{1}{q}\right| \le |\lambda| x_n. \tag{14}$$

Then applying equations (12)–(14),

$$\begin{aligned}
&\left| E\prod_{i=1}^{q}\left(1+\lambda q\left(Y_{i1}-\frac{1}{q}\right)\right)^{k_i}-1\right| \\
&\quad =\left|\sum_{(s_1,\dots,s_q)} E\prod_{i=1}^{q}\binom{k_i}{s_i}\lambda^{s_i}q^{s_i}\left(Y_{i1}-\frac{1}{q}\right)^{s_i}-1\right| \\
&\quad =\left|E\sum_{i=1}^{q}k_i\lambda q\left(Y_{i1}-\frac{1}{q}\right)+\sum_{(s_1,\dots,s_q),\sum s_i\ge 2} E\prod_{i=1}^{q}\binom{k_i}{s_i}\lambda^{s_i}q^{s_i}\left(Y_{i1}-\frac{1}{q}\right)^{s_i}\right| \\
&\quad \le C'\lambda^2 x_n,
\end{aligned}$$

where the sum runs over all q-tuples of nonegative integers $(s_1,\dots,s_q)$ with $s_i\le k_i$ for all i and the constant C' depends only on q and $k_1,\dots,k_q$. The final inequality in the last equation follows from equations (12)–(14) since every term is bounded by $C''\lambda^2 x_n$ where C'' depends only on q and k. Since $0\le x_n\le 1$ and $\lambda^2 d\le 1$, applying equation (11) with

$$y=E\prod_{i=1}^{q}\left(1+\lambda q\left(Y_{i1}-\frac{1}{q}\right)\right)^{k_i}-1$$

completes the result. $\square$

2.3. *Main expansion.* In order to evaluate the expected value of $EX^+(n+1)$ using equation (4), we expand it out using the identity

$$\frac{a}{s+r}=\frac{a}{s}-\frac{ar}{s^2}+\frac{r^2}{s^2}\frac{a}{s+r}. \tag{15}$$

With this expansion and $a=Z_1$, $s=q$ and $r=(\sum_{i=1}^{q}Z_i)-q$ clearly,

$$\begin{aligned}
x_{n+1}&=E\frac{Z_1}{\sum_{i=1}^{q}Z_i}-\frac{1}{q}\\
&=E\frac{Z_1}{q}-E\frac{Z_1((\sum_{i=1}^{q}Z_i)-q)}{q^2}+E\frac{Z_1}{\sum_{i=1}^{q}Z_i}\frac{((\sum_{i=1}^{q}Z_i)-q)^2}{q^2}-\frac{1}{q}.
\end{aligned}\tag{16}$$

We estimate the expected value of each of the terms in the preceding equation. First,

$$\begin{aligned}
EZ_1&=1+d\lambda qE\left(Y_{11}-\frac{1}{q}\right)+\frac{d(d-1)}{2}\left(\lambda qE\left(Y_{11}-\frac{1}{q}\right)\right)^2+R_1\\
&=1+d\lambda^2 qx_n+\frac{d(d-1)}{2}\lambda^4q^2x_n^2+R_1,
\end{aligned}\tag{17}$$

where by Lemma 2.6 the error term satisfies $|R_1| \leq C_1 x_n^3$ where C_1 does not depend on λ, d or x_n. Next, applying Lemma 2.6 and Lemma 2.5 and canceling terms

$$
\begin{aligned}
EZ_1\left(\sum_{i=1}^q Z_i - q\right) &= EZ_1^2 + \sum_{i=2}^q EZ_1 Z_i - qEZ_1 \\
&= \frac{d(d-1)}{2}\lambda^4 q^2\biggl[((3-\lambda)x_n + \lambda q z_n)^2 \\
&\qquad + \frac{1}{q-1}((q-3+\lambda)x_n - \lambda q z_n)^2 - q x_n^2\biggr] \\
&\quad + R_2,
\end{aligned}
\tag{18}
$$

where by Lemma 2.6 $|R_2| \leq C_2 x_n^3$ and C_2 does not depend on λ, d or x_n. Finally, again using Lemma 2.6 and Lemma 2.5,

$$
\begin{aligned}
&E\left(\left(\sum_{i=1}^q Z_i\right) - q\right)^2 \\
&\quad = EZ_1^2 + \sum_{i=2}^q EZ_i^2 + 2\sum_{i=2}^q EZ_1 Z_i + 2\sum_{i_1=2}^q \sum_{i_2=i_1+1}^q EZ_{i_1} Z_{i_2} \\
&\qquad - 2qEZ_1 - 2q\sum_{i=2}^q EZ_i + q^2 \\
&\quad = \frac{d(d-1)}{2}\lambda^4 q^2\biggl[((3-\lambda)x_n + \lambda q z_n)^2 \\
&\qquad + \frac{3}{q-1}((q-3+\lambda)x_n - \lambda q z_n)^2 - 2q x_n^2 - \frac{2q x_n^2}{q-1} \\
&\qquad + \frac{1}{(q-1)(q-2)}((3q-6+2\lambda)x_n - 2\lambda q z_n)^2\biggr] + R_3,
\end{aligned}
\tag{19}
$$

where by Lemma 2.6 $|R_3| \leq C_3 x_n^3$ and C_3 does not depend on λ, d or x_n. By Lemma 2.2, we have that $0 \leq z_n \leq x_n$ and since $|\lambda| \leq 1$ the expressions in equations (18) and (19) are both bounded by $C\frac{d(d-1)}{2}\lambda^4 x_n^2$ where C depends only on q. Now using the fact that $0 \leq \frac{Z_1}{\sum Z_i} \leq 1$ and substituting equations (17)–(19) into equation (16), we have that

$$
|x_{n+1} - d\lambda^2 x_n| \leq C_q \lambda^4 \frac{d(d-1)}{2} x_n^2 \leq C_q x_n^2, \tag{20}
$$

where C_q depends only on q since $\lambda^2 d \leq 1$. In order to complete the proof, we will need a more precise bound. To motivate the rest of the proof, suppose that we could establish the following condition.

CONDITION 2.7. Suppose the following hold:

- that $z_n = (\frac{1}{q} + o(1))x_n$,
- that $\frac{Z_1}{\sum_{i=1}^q Z_i}$ is sufficiently concentrated around $\frac{1}{q}$ so that

$$E\frac{Z_1}{\sum_{i=1}^q Z_i}\frac{((\sum_{i=1}^q Z_i) - q)^2}{q^2} = \left(\frac{1}{q} + o(1)\right)E\frac{((\sum_{i=1}^q Z_i) - q)^2}{q^2}.$$

If we established Condition 2.7 then by substituting equations (17)–(19) into equation (16), we would have that

$$x_{n+1} = d\lambda^2 x_n + \bigl(1 + o(1)\bigr)\frac{q(q-4)}{q-1}\frac{d(d-1)}{2}\lambda^4 x_n^2. \tag{21}$$

Proving Condition 2.7 is one of the main technical challenges in this paper.

2.4. *Concentration lemmas.* In this section, we establish a number of lemmas in order to establish the Condition 2.7. The following lemma follows immediately from equation (20).

LEMMA 2.8. *For any* $\varepsilon > 0$, *there exists a constant* $\delta = \delta(q, \varepsilon)$ *such that for all* n, *if* $x_n < \delta$ *then*

$$|x_{n+1} - d\lambda^2 x_n| \le \varepsilon x_n.$$

The following lemma ensures that the decrease from x_n to x_{n+1} is never too large.

LEMMA 2.9. *For any* $\kappa > 0$ *there exists a constant* $\gamma = \gamma(q, \kappa, d) > 0$ *such that for all* n *when* $\kappa < |\lambda|$,

$$x_{n+1} \ge \gamma x_n.$$

PROOF. For a configuration A on $T_{u_1} \cap S(n+1)$ define

$$f_{n+1}^*(i, A) = P\bigl(\sigma_\rho = i | \sigma_1(n+1) = A\bigr);$$

that is, the probability the root is in state i given the configuration on the leaves in $T_{u_1} \cap S(n+1)$. Now

$$f_{n+1}^*(i, A) = \frac{(e^\beta f_n(1, A) + \sum_{l \ne 1} f_n(l, A))}{\sum_{i=1}^q (e^\beta f_n(i, A) + \sum_{l \ne i} f_n(l, A))} = \frac{(1 + \lambda q(f_n(1, A) - 1/q))}{q}$$

and so

$$Ef_{n+1}^*\bigl(i, \sigma_1^1(n+1)\bigr) = \frac{1}{q} + \lambda^2 x_n.$$

The estimator that chooses a state with probability $f_{n+1}^*(i, \sigma_1(n+1))$ correctly reconstructs the root with probability $\frac{1}{q} + \lambda^2 x_n$. Since this probability must be less than the MLE, it follows that

$$\lambda^2 x_n + \frac{1}{q} \le p_{n+1} \le x_{n+1}^{1/2} + \frac{1}{q}$$

and so $x_{n+1} \ge \lambda^4 x_n^2 \ge \kappa^4 x_n^2$ for an value of x_n. Now when $x_n < \delta$ by Lemma 2.8 it follows that

$$x_{n+1} \ge (d\lambda^2 - \varepsilon)x_n.$$

Combining these results completes the proof. □

2.4.1. *Concentration.* We will establish some concentration results which will be required in order to make the approximation

$$\frac{Z_1}{\sum_{i=1}^q Z_i} \approx \frac{1}{q}.$$

The first lemma establishes a technical uniqueness result where the set of vertices which can be conditioned is limited to a set of k vertices.

LEMMA 2.10. *For any $\varepsilon > 0$ and positive integer k there exists $\Lambda = \Lambda(q, d, \varepsilon, k)$ not depending on λ such that for any collection of vertices $v_1, \dots, v_k \in S(\Lambda)$,*

$$\sup_{i, i_1, \dots, i_k \in \mathcal{C}} \left| P(\sigma_\rho = i | \sigma_{v_t} = i_t, 1 \le t \le k) - \frac{1}{q} \right| < \varepsilon.$$

PROOF. This lemma simply says that fixing the spins at k distant vertices a long way from the root has only a small effect on the root. We note that

$$M^s_{i_1,i_2} = \begin{cases} \frac{1}{q} + \left(1 - \frac{1}{q}\right)\lambda^s, & i_1 = i_2, \\ \frac{1}{q} - \frac{1}{q}\lambda^s, & \text{otherwise}, \end{cases}$$

and so since $\lambda^2 d \le 1$,

$$\frac{1}{q} - d^{-s/2} \le M^s_{i_1,i_2} \le \frac{1}{q} + d^{-s/2}.$$

Let γ be an integer sufficiently large such that

$$\left(\frac{1/q + d^{-\gamma/2}}{1/q - d^{-\gamma/2}}\right)^k < 1 + \varepsilon.$$

Fix an integer Λ such that $\Lambda > k\gamma$. Now choose any $v_1, \dots, v_k \in S(\Lambda)$ with $d(v_i, \rho) = \Lambda$. For $0 \le \ell \le \Lambda$ define a_ℓ to be the number of vertices distance ℓ from the root with a descendant in the set $\{v_1, \dots, v_k\}$, that is, $a_\ell = \#\{v \in$

$S(\ell):|T_v\cap\{v_1,\dots,v_k\}|>0\}$. Then $a_0=1, a_\Lambda=k$ and the a_ℓ are increasing and integer valued. Therefore, there must be some ℓ such that $a_\ell=a_{\ell+\gamma}$. Let $\overline{w}_1,\dots,\overline{w}_{a_\ell}$ denote the vertices in the set $\{v\in S(\ell):|T_v\cap\{v_1,\dots,v_k\}|>0\}$ and $w_1,\dots,w_{a_\ell}$ denote the vertices in the set $\{v\in S(\ell+\gamma):|T_v\cap\{v_1,\dots,v_k\}|>0\}$ such that w_t is the descendant of $\overline{w}_t$. By the Markov random field property, the σ_{w_t} are conditionally independent given the $\sigma_{\overline{w}_t}$. The distribution of σ_{w_t} given $\sigma_{\overline{w}_t}$ is

$$P(\sigma_{w_t}=i_2|\sigma_{\overline{w}_t}=i_1)=M^{\gamma}_{i_1,i_2}.$$

By Bayes Rule and the Markov random field property, we have that for any $i,i',i_1,\dots,i_{a_\ell}\in\mathcal{C}$,

$$\begin{aligned}
&\frac{P(\sigma_\rho=i|\sigma_{w_t}=i_t,1\le t\le a_\ell)}{P(\sigma_\rho=i'|\sigma_{w_t}=i_t,1\le t\le a_\ell)}\\
&\quad=\frac{P(\sigma_{w_t}=i_t,1\le t\le a_\ell|\sigma_\rho=i)}{P(\sigma_{w_t}=i_t,1\le t\le a_\ell|\sigma_\rho=i')}\\
&\quad=\frac{\sum_{h_1,\dots,h_{a_\ell}\in\mathcal{C}}P(\forall t\sigma_{w_t}=i_t|\forall t\sigma_{\overline{w}_t}=h_t)P(\forall t\sigma_{\overline{w}_t}=h_t|\sigma_\rho=i)}{\sum_{h_1,\dots,h_{a_\ell}\in\mathcal{C}}P(\forall t\sigma_{w_t}=i_t|\forall t\sigma_{\overline{w}_t}=h_t)P(\forall t\sigma_{\overline{w}_t}=h_t|\sigma_\rho=i')}\\
&\quad=\frac{\sum_{h_1,\dots,h_{a_\ell}\in\mathcal{C}}P(\sigma_{\overline{w}_t}=h_t,1\le t\le a_\ell|\sigma_\rho=i)\prod_{t=1}^{a_\ell}M^{\gamma}_{h_t,i_t}}{\sum_{h_1,\dots,h_{a_\ell}\in\mathcal{C}}P(\sigma_{\overline{w}_t}=h_t,1\le t\le a_\ell|\sigma_\rho=i')\prod_{t=1}^{a_\ell}M^{\gamma}_{h_t,i_t}}\\
&\quad\le\frac{\sum_{h_1,\dots,h_{a_\ell}\in\mathcal{C}}P(\sigma_{\overline{w}_t}=h_t,1\le t\le a_\ell|\sigma_\rho=i)(1/q+d^{-\gamma/2})^{a_\ell}}{\sum_{h_1,\dots,h_{a_\ell}\in\mathcal{C}}P(\sigma_{\overline{w}_t}=h_t,1\le t\le a_\ell|\sigma_\rho=i')(1/q-d^{-\gamma/2})^{a_\ell}}\\
&\quad\le\frac{(1/q+d^{-\gamma/2})^{a_\ell}}{(1/q-d^{-\gamma/2})^{a_\ell}}\\
&\quad\le 1+\varepsilon
\end{aligned}$$

so it follows that

$$P(\sigma_\rho=i|\sigma_{w_t}=i_t,1\le t\le a_\ell)\le\frac{1}{q}(1+\varepsilon)$$

and

$$P(\sigma_\rho=i|\sigma_{w_t}=i_t,1\le t\le a_\ell)\ge\frac{1}{q}\frac{1}{1+\varepsilon}\ge\frac{1}{q}(1-\varepsilon).$$

By the Markov random field property since σ_ρ is conditionally independent of the collection $\sigma_{v_1},\dots,\sigma_{v_k}$ given the spins $\sigma_{w_1},\dots,\sigma_{w_{a_\ell}}$ it follows that

$$\begin{aligned}
&\sup_{i,i_1,\dots,i_k\in\mathcal{C}}\left|P(\sigma_\rho=i|\sigma_{v_t}=i_t,1\le t\le k)-\frac{1}{q}\right|\\
&\qquad\le\sup_{i,i_1,\dots,i_{a_\ell}\in\mathcal{C}}\left|P(\sigma_\rho=i|\sigma_{w_t}=i_t,1\le t\le a_\ell)-\frac{1}{q}\right|<\varepsilon
\end{aligned}$$

which completes the result. □

The next lemma establishes concentration of the posterior distributions when x_n is small.

LEMMA 2.11. *For any $\varepsilon, \alpha, \kappa > 0$, there exists $C = C(q, d, \varepsilon, \alpha, \kappa)$ and $N = N(q, d, \varepsilon, \alpha, \kappa)$ such that for any λ with $\kappa < |\lambda| \le d^{-1/2}$ and for $n > N$,*

$$P\left(\left|\frac{Z_1}{\sum_{i=1}^q Z_i} - \frac{1}{q}\right| > \varepsilon\right) \le C x_n^{\alpha}.$$

PROOF. The conclusion is trivially true if both C and x_n are large so we will suppose that x_n is small. Fix k an integer such that $k > \alpha$. Choose Λ large enough so that the conclusion of Lemma 2.10 holds with bound $\varepsilon/2$ and set $N = \Lambda$. Let $v_1, \dots, v_{|S(\Lambda)|}$ denote the vertices in $S(\Lambda)$. For $v \in S(\Lambda)$, let $\sigma_v^1(n+1)$ denote the spins of the vertices in $T_v \cap S(n+1)$ and define

$$W(i, v) = f_{n-\Lambda+1}\big(i, \sigma_v^1(n+1)\big)$$

which is the conditional probability that σ_v is in state i given the boundary condition $\sigma_v^1(n+1)$. Conditional on $\sigma^1(\Lambda)$, the spins of $S(\Lambda)$, the $W(i, v)$ are distributed as

$$W(i, v) \sim \begin{cases} X^+(n+1-\Lambda), & \sigma_v^1 = i, \\ X^-(n+1-\Lambda), & \sigma_v^1 \ne i. \end{cases}$$

Also conditional on $\sigma^1(\Lambda)$ the vectors $(W(1, v), \dots, W(q, v))$ are conditionally independent for different $v \in S(\Lambda)$. Using the recursion of equation (3), a posterior probability of a vertex can be written as a function of the posterior probabilities of its children so there exists a function $g_\lambda(\mathcal{W})$ such that

$$\frac{Z_1}{\sum_{i=1}^q Z_i} = f_{n+1}\big(1, \sigma^1(n+1)\big) = g_\lambda(\mathcal{W}),$$

where $\mathcal{W}$ denotes the vector

$$\mathcal{W} = \big(W(1, v_1), \dots, W\big(1, v_{|S(\Lambda)|}\big), W(2, v_1), \dots, W\big(q, v_{|S(\Lambda)|}\big)\big).$$

When x_n is small, we expect most of the $W(i, v)$ to be close to $\frac{1}{q}$. If all the entries in $\mathcal{W}$ are identically $\frac{1}{q}$, then $g_\lambda(\mathcal{W}) = \frac{1}{q}$. It follows by Lemma 2.10 that if there are at most k vertices $v \in S(\Lambda)$ such that for some $1 \le i \le q$, $W(i, v) \ne \frac{1}{q}$ then

$$\left|g_\lambda(\mathcal{W}) - \frac{1}{q}\right| < \varepsilon/2.$$

Observe that g_λ is a continuous function of each of the elements of the vector $\mathcal{W}$ and of λ. It follows that there exists a $\delta > 0$ such that if $\mathcal{W}$ satisfies

$$\#\left\{v \in S(\Lambda) : \max_{1 \le i \le q}\left|W(i, v) - \frac{1}{q}\right| > \delta\right\} \le k$$

then

$$\left|g_\lambda(\mathcal{W}) - \frac{1}{q}\right| < \varepsilon.$$

As the random variables $\max_{1\le i\le q} |W(i,v) - \frac{1}{q}|$ are independent since they are conditionally independent given $\sigma(\Lambda)$ and by the symmetry of the model they do not in fact depend on the spins in $S(\Lambda)$. By Chebyshev's inequality and Lemma 2.2, we have that

$$\begin{aligned}
&P\left(\max_{1\le i\le q}\left|W(i,v) - \frac{1}{q}\right| > \delta\right) \\
&\quad\le P\left(\left|X^+(n+1-\Lambda) - \frac{1}{q}\right| > \delta\right) + (q-1)P\left(\left|X^-(n+1-\Lambda) - \frac{1}{q}\right| > \delta\right) \\
&\quad\le \delta^{-2}\left[E\left(X^+(n+1-\Lambda) - \frac{1}{q}\right)^2 + (q-1)E\left(X^-(n+1-\Lambda) - \frac{1}{q}\right)^2\right] \\
&\quad= \frac{x_{n+1-\Lambda}}{\delta^2}.
\end{aligned}$$

As noted above, we may suppose that x_n is very small so these events are rare. In particular, we have that

$$\begin{aligned}
P\left(\left|\frac{Z_1}{\sum_{i=1}^q Z_i} - \frac{1}{q}\right| > \varepsilon\right) &\le P\left(\#\left\{\max_{1\le i\le q}\left|W(i,v) - \frac{1}{q}\right| > \delta\right\} > k\right) \\
&\le P\left(\mathrm{Binom}\left(|S(\Lambda)|, \frac{x_{n+1-\Lambda}}{\delta^2}\right) > k\right) \\
&\le C' x_{n+1-\Lambda}^\alpha \\
&\le C x_n^\alpha,
\end{aligned}$$

where the third inequality holds for large enough C' since $k > \alpha$ and the final inequality follows by Lemma 2.9 which completes the proof. Only in this final inequality do we use the assumption that $\kappa < |\lambda|$. $\square$

To establish the necessary concentration results, we will make use of Bennett's inequality which is stated below (see, e.g., [23], Appendix B, Lemma 4).

LEMMA 2.12. *For independent mean* 0 *random variables*, $W_1, \ldots, W_n$ *satisfying* $W_i \le M$, $b_n^2 = \sum_{i=1}^n E(W_i^2)$. *Then for any* $\eta \ge 0$,

$$P\left(\sum_{i=1}^n W_i \ge \eta\right) \le \exp\left(-\frac{b_n^2}{M^2}\theta\left(\frac{\eta M}{b_n^2}\right)\right), \tag{22}$$

where $\theta(x) = (1+x)\log(1+x) - x$.

The following concentration result holds uniformly provided λ is small enough. It is necessary in taking limits for large d.

LEMMA 2.13. *For any* $0 < \varepsilon < 1$ *and* $\alpha > 1$ *there exists* $C = C(q, \varepsilon, \alpha)$ *and* $N = N(q, \varepsilon, \alpha)$ *depending only on* q, α *and* ε *such that whenever* $|\lambda| q \le \frac{1}{2}$ *and*

$$|\lambda| q + \lambda^2 q^2 \le \frac{\max\{-\log(1-\varepsilon), \log(1+\varepsilon)\}}{4\alpha}$$

then for $1 \le i \le q$ *and* $n > N$,

$$P\big(|Z_i(n) - 1| > \varepsilon\big) \le C x_n^{\alpha}.$$

PROOF. Observe that the hypothesis only holds when $|\lambda|$ is small, that is, the interactions are weak enough. Let

$$M = \frac{\max\{-\log(1-\varepsilon), \log(1+\varepsilon)\}}{4\alpha}.$$

By taking C large enough, we can assume that

$$x_n < \frac{q^2}{2} \min\{-\log(1-\varepsilon), \log(1+\varepsilon)\},$$

since otherwise the conclusion is trivial.

Since $1 - 2y \le \frac{1}{1+y} \le 1$ when $0 \le y \le \frac{1}{2}$ and $1 - 2y \ge \frac{1}{1+y} \ge 1$ when $-\frac{1}{2} \le y \le 0$ by integrating it follows that when $|y| \le \frac{1}{2}$,

$$y - y^2 \le \log(1+y) \le y. \tag{23}$$

Taking $y = \lambda q(Y_{ij} - \frac{1}{q})$, then

$$-M \le -|\lambda q| - \lambda^2 q^2 \le \lambda q\left(Y_{ij} - \frac{1}{q}\right) - \lambda^2 q^2 \left(Y_{ij} - \frac{1}{q}\right)^2 \le \log\left(1 + \lambda q\left(Y_{ij} - \frac{1}{q}\right)\right)$$

and

$$\log\left(1 + \lambda q\left(Y_{ij} - \frac{1}{q}\right)\right) \le \lambda q\left(Y_{ij} - \frac{1}{q}\right) \le |\lambda q| \le M.$$

Let

$$W_j = \lambda q\left(Y_{1j} - \frac{1}{q}\right) - \lambda^2 q^2 \left(Y_{1j} - \frac{1}{q}\right)^2$$

and so by Lemma 2.5,

$$EW_j = \lambda^3 q x_n - \lambda^2 q^3 z_n \le |\lambda|^3 q x_n$$

and $-(W_j - EW_j) \le M + |\lambda|^3 q \le 2M$. Also $EW_j = \lambda^3 q x_n - \lambda^3 q^2 z_n \ge -|\lambda|^3 q^2 x_n$ so $dEW_j \ge -q^2 x_n$. Since by definition, $0 \le Y_{ij} \le 1$, our assumption that $|\lambda| q < \frac{1}{2}$ implies that $|\lambda q(Y_{1j} - \frac{1}{q})| < \frac{1}{2}$. From the inequality $(a+b)^2 \le 2a^2 + 2b^2$ and Lemma 2.5, it follows that

$$\begin{aligned} E(W_j - EW_j)^2 \le EW_j^2 &\le 2E\left(\lambda q\left(Y_{1j} - \frac{1}{q}\right)\right)^2 + 2E\left(\lambda q\left(Y_{1j} - \frac{1}{q}\right)\right)^4 \\ &\le 4\lambda^2 q^2 x_n \end{aligned}$$

and so if $B = \sum_{j=1}^d E(W_j - EW_j)^2$ then $B \le 4d\lambda^2 q^2 x_n \le 4q^2 x_n$ since $d\lambda^2 \le 1$. Now

$$\begin{aligned} P(Z_1 \le 1 - \varepsilon) &= P\left(\sum_{j=1}^d \log\left(1 + \lambda q\left(Y_{1j} - \frac{1}{q}\right)\right) \le \log(1-\varepsilon)\right) \\ &\le P\left(\sum_{j=1}^d W_j \le \log(1-\varepsilon)\right) \\ &\le P\left(\sum_{j=1}^d -(W_j - EW_j) \ge -\log(1-\varepsilon) - q^2 x_n\right) \\ &\le P\left(\sum_{j=1}^d -(W_j - EW_j) \ge -\frac{1}{2}\log(1-\varepsilon)\right) \\ &\le \exp\left(-\frac{B}{4M^2}\theta\left(\frac{(-(1/2)\log(1-\varepsilon))2M}{B}\right)\right), \end{aligned} \tag{24}$$

where the first inequality follows from equation (23), the second from the fact that $dEW_j \ge -q^2 x_n$, the third from our assumption that $x_n < \frac{q^2}{2}\min\{-\log(1-\varepsilon), \log(1+\varepsilon)\}$ and the final inequality by applying Lemma 2.12.

Since $\frac{1}{x}\theta(x)$ is increasing in x, the right-hand side of equation (24) is increasing in B and hence substituting $B \le 4q^2 x_n$ gives

$$\begin{aligned} P(Z_1 \le 1 - \varepsilon) &\le \exp\left(-\frac{4q^2 x_n}{4M^2}\theta\left(\frac{-\log(1-\varepsilon)M}{4q^2 x_n}\right)\right) \\ &\le \exp\left[-\frac{-\log(1-\varepsilon)}{4M}\left(\log\left(\frac{-\log(1-\varepsilon)M}{4q^2 x_n}\right) - 1\right)\right] \\ &\le \exp\left[\frac{\log(1-\varepsilon)}{4M}\left(\log\left(\frac{-\log(1-\varepsilon)M}{4q^2}\right) - 1\right)\right] x_n^{-\log(1-\varepsilon)/(4M)} \\ &\le C x_n^{\alpha}, \end{aligned} \tag{25}$$

where the second inequality uses the fact that $\theta(x) < x(\log(x) - 1)$. With essentially the same argument, we have $P(Z_1 \geq 1 + \varepsilon) < Cx_n^\alpha$. Furthermore, the result holds similarly for the other Z_i as well which completes the result. $\square$

Combining the results of this section, the following corollary gives us the concentration result we need.

COROLLARY 2.14. *For any $0 < \varepsilon < 1$ and $\alpha > 1$ there exists $C = C(q, \varepsilon, \alpha)$ and $N = N(q, \varepsilon, \alpha)$ depending only on q, α and ε such that for $1 \leq i \leq q$ and $n > N$,*

$$P\left(\left|\frac{Z_1}{\sum_{i=1}^q Z_i} - \frac{1}{q}\right| > \varepsilon\right) \leq Cx_n^\alpha. \tag{26}$$

PROOF. In light of Lemmas 2.11 and Lemma 2.13, we split the result into two cases, when $|\lambda|$ is big and small. Let $\varepsilon'(q) > 0$ be small enough so that if for all i, $|Z_i - 1| < \varepsilon'$ then

$$\left|\frac{Z_1}{\sum_{i=1}^q Z_i} - \frac{1}{q}\right| < \varepsilon,$$

and let

$$M = \frac{\max\{-\log(1-\varepsilon'), \log(1+\varepsilon')\}}{4\alpha}.$$

For each fixed d, define

$$\mathcal{K}_d = \{\lambda : |\lambda|q < \tfrac{1}{2}, |\lambda|q + \lambda^2 q^2 < M\},$$

an open set which includes 0. Let $\mathcal{J}_d = [-d^{-1/2}, d^{1/2}] \setminus \mathcal{K}_d$.

By Lemma 2.13, equation (26) holds with a bound $C' = C'(q, \varepsilon, \alpha)$ not depending on λ or d, provided $\lambda \in \mathcal{K}_d$. For each fixed d, Lemma 2.11 implies that equation (26) holds with a bound $C_d'' = C_d''(q, \varepsilon, \alpha)$ not depending on λ, provided $\lambda \in \mathcal{J}_d$. Since $\lambda^2 d \leq 1$, for large enough d so that $d \geq 4q^2$ and $d^{-1/2}q + d^{-1}q^2 \leq M$ the set $\mathcal{J}_d$ is empty. It follows that equation (26) holds with a bound

$$C = \max\Big\{C', \max_{d' : \mathcal{J}_{d'} \neq \phi} C_{d'}''\Big\}$$

that is independent of λ and d. $\square$

2.5. *Bound on $z_n - \frac{1}{q}x_n$.* In this section, we bound the term $z_n - \frac{1}{q}x_n$ when x_n is small.

LEMMA 2.15. *For any $\varepsilon, \kappa > 0$ there exists a $\delta = \delta(q, \kappa, d)$ and $k = k(q, \kappa, d)$ such that if $x_n < \delta$ and $|\lambda| \geq \kappa$ then*

$$\left|\frac{z_{n+k}}{x_{n+k}} - \frac{1}{q}\right| \leq \varepsilon.$$

PROOF. Using the identity (15), we have

$$
\begin{aligned}
z_{n+1} &= E\frac{(Z_1-(1/q)\sum_{i=1}^q Z_i)^2}{(\sum_{i=1}^q Z_i)^2} \\
&= E\frac{1}{q^2}\left(Z_1-\frac{1}{q}\sum_{i=1}^q Z_i\right)^2-\frac{1}{q^4}\left(Z_1-\frac{1}{q}\sum_{i=1}^q Z_i\right)^2\left(\left(\sum_{i=1}^q Z_i\right)^2-q^2\right) \qquad (27)\\
&\quad+\frac{1}{q^4}\frac{(Z_1-(1/q)\sum_{i=1}^q Z_i)^2}{(\sum_{i=1}^q Z_i)^2}\left(\left(\sum_{i=1}^q Z_i\right)^2-q^2\right)^2.
\end{aligned}
$$

Expanding and using Lemma 2.6 and Lemma 2.5, we get that

$$\left|E\frac{1}{q^2}\left(Z_1-\frac{1}{q}\sum_{i=1}^q Z_i\right)^2-d\lambda^2\left((1-\lambda)\frac{1}{q}x_n+\lambda z_n\right)\right|\le C_q x_n^2.$$

Similarly,

$$\left|E\frac{1}{q^4}\left(Z_1-\frac{1}{q}\sum_{i=1}^q Z_i\right)^2\left(\left(\sum_{i=1}^q Z_i\right)^2-q^2\right)\right|\le C_q x_n^2$$

and

$$E\left(\left(\sum_{i=1}^q Z_i\right)^2-q^2\right)^2\le C_q x_n^2.$$

Substituting these bounds into equation (27) and noting that

$$\left|\frac{(Z_1-(1/q)\sum_{i=1}^q Z_i)^2}{(\sum_{i=1}^q Z_i)^2}\right|\le 1$$

so we have that

$$\left|z_{n+1}-d\lambda^2\left((1-\lambda)\frac{1}{q}x_n+\lambda z_n\right)\right|\le C_q' x_n^2.$$

Dividing by x_{n+1}, we get

$$\left|\frac{z_{n+1}}{x_{n+1}}-\frac{d\lambda^2 x_n}{x_{n+1}}\left((1-\lambda)\frac{1}{q}+\lambda\frac{z_n}{x_n}\right)\right|\le C_q'\frac{x_n^2}{x_{n+1}}.$$

By Lemma 2.9, we have that $\frac{x_n}{x_{n+1}}\le\gamma^{-1}$ and by equation (20) $|\frac{d\lambda^2 x_n}{x_{n+1}}-1|\le C_q'''\frac{x_n^2}{x_{n+1}}$. It follows that

$$\left|\frac{z_{n+1}}{x_{n+1}}-\left((1-\lambda)\frac{1}{q}+\lambda\frac{z_n}{x_n}\right)\right|\le C_q'' x_{n+1}. \qquad (28)$$

Iterating this equation, we get that

$$\begin{aligned}
&\left|\frac{z_{n+k}}{x_{n+k}}-(1-\lambda^k)\frac{1}{q}+\lambda^k\frac{z_n}{x_n}\right| \\
&\quad\le\sum_{\ell=1}^{k}\left|(1-\lambda^{k-\ell})\frac{1}{q}+\lambda^{k-\ell}\frac{z_{n+\ell}}{x_{n+\ell}}-(1-\lambda^{k-\ell+1})\frac{1}{q}-\lambda^{k-\ell+1}\frac{z_{n+\ell-1}}{x_{n+\ell-1}}\right| \\
&\quad\le\sum_{\ell=1}^{k}|\lambda|^{k-\ell}\left|\frac{z_{n+\ell}}{x_{n+\ell}}-\left((1-\lambda)\frac{1}{q}+\lambda\frac{z_{n+\ell-1}}{x_{n+\ell-1}}\right)\right| \\
&\quad\le C''_q\sum_{\ell=1}^{k}|\lambda|^{k-\ell}x_{n+\ell-1}.
\end{aligned}\tag{29}$$

Iteratively applying Lemma 2.8 implies that if $\delta>0$ is small enough and $x_n<\delta$ then for $0\le\ell\le k$, $x_{n+\ell}\le 2\delta$. Since $0\le z_n\le x_n$ it follows from equation (29) that

$$\left|\frac{z_{n+k}}{x_{n+k}}-\frac{1}{q}\right|\le\lambda^k+2\delta C''_q\sum_{\ell=1}^{k}\lambda^{k-\ell}.$$

By taking k sufficiently large and δ sufficiently small, we complete the result. □

COROLLARY 2.16. *For any $\varepsilon,\kappa>0$ there exists a $\delta=\delta(q,\kappa,d)$ and $k=k(q,\kappa,d)$ such that if $x_n<\delta$, $n>k$ and $|\lambda|\ge\kappa$ then*

$$\left|\frac{z_n}{x_n}-\frac{1}{q}\right|\le\varepsilon.$$

PROOF. By Lemma 2.9, if $x_n<\delta$ then $x_{n-k}<\gamma^{-k}x_n$ and so the result follows by Lemma 2.15. □

3. Reconstruction for $q\ge 5$. The lemmas proved in Sections 2.4 and 2.5 establish Condition 2.7. We now use these results to establish the change from x_n to x_{n+1} when x_n is small.

LEMMA 3.1. *There exists a $\delta=\delta(q)>0$ and $N=N(q)$ such that if $x_n\le\delta$ and $n>N$ then*

$$x_{n+1}\ge d\lambda^2x_n+\frac{1}{2}\frac{d(d-1)}{2}\frac{q(q-4)}{q-1}\lambda^4x_n^2.$$

PROOF. Let $\varepsilon > 0$. Then

$$
\begin{aligned}
&\left| E\frac{Z_1}{\sum_{i=1}^q Z_i}\frac{((\sum_{i=1}^q Z_i)-q)^2}{q^2} - E\frac{1}{q}\frac{((\sum_{i=1}^q Z_i)-q)^2}{q^2}\right| \\
&\quad \le \varepsilon E\frac{1}{q}\frac{((\sum_{i=1}^q Z_i)-q)^2}{q^2} + EI\left(\left|\frac{Z_1}{\sum_{i=1}^q Z_i} - \frac{1}{q}\right| > \varepsilon\right)\frac{((\sum_{i=1}^q Z_i)-q)^2}{q^2} \\
&\quad \le \varepsilon E\frac{1}{q}\frac{((\sum_{i=1}^q Z_i)-q)^2}{q^2} \\
&\qquad + P\left(\left|\frac{Z_1}{\sum_{i=1}^q Z_i} - \frac{1}{q}\right| > \varepsilon\right)^{1/2}\left(E\left(\frac{((\sum_{i=1}^q Z_i)-q)^2}{q^2}\right)^2\right)^{1/2} \\
&\quad \le \varepsilon E\frac{1}{q}\frac{((\sum_{i=1}^q Z_i)-q)^2}{q^2} + C'x_n^3\left(E\left(\frac{((\sum_{i=1}^q Z_i)-q)^2}{q^2}\right)^2\right)^{1/2} \\
&\quad \le \varepsilon E\frac{1}{q}\frac{((\sum_{i=1}^q Z_i)-q)^2}{q^2} + Cx_n^3,
\end{aligned} \tag{30}
$$

where the second inequality comes from the Cauchy–Schwarz inequality and the third follows by Corollary 2.14 provided that n is sufficiently large while the fourth inequality follows by Lemma 2.6.

Now by substituting equations (17)–(19), we have that

$$
\begin{aligned}
&E\frac{Z_1}{q} - E\frac{Z_1((\sum_{i=1}^q Z_i)-q)}{q^2} + E\frac{1}{q}\frac{((\sum_{i=1}^q Z_i)-q)^2}{q^2} \\
&\quad = \frac{1}{q} + d\lambda^2 x_n \\
&\qquad + \frac{d(d-1)}{2}\lambda^4\biggl[\frac{2q(q-2)}{q-1}x_n^2 - \frac{q-1}{q}\bigl((3-\lambda)x_n + \lambda q z_n\bigr)^2 \\
&\qquad\qquad - \frac{q-3}{q(q-1)}\bigl((q-3+\lambda)x_n - \lambda q z_n\bigr)^2 \\
&\qquad\qquad + \frac{1}{q(q-1)(q-2)}\bigl((3q-6+2\lambda)x_n - 2\lambda q z_n\bigr)^2\biggr] + R \\
&\quad \ge \frac{1}{q} + d\lambda^2 x_n + \frac{d(d-1)}{2}\frac{q(q-4)}{q-1}\lambda^4 x_n^2 \\
&\qquad - C'\frac{d(d-1)}{2}\lambda^5\left|\frac{z_n}{x_n} - \frac{1}{q}\right| x_n^2 - R,
\end{aligned} \tag{31}
$$

where $|R| \le C x_n^3$ and C and C' depend only on q. Let $\kappa = \frac{q(q-4)}{3C'(q-1)}$ then if $|\lambda| \le \kappa$ then since $0 \le z_n \le x_n$,

$$\begin{aligned} C' \frac{d(d-1)}{2} \lambda^5 \left| \frac{z_n}{x_n} - \frac{1}{q} \right| x_n^2 &\le C' \kappa \lambda^4 \left| \frac{z_n}{x_n} - \frac{1}{q} \right| x_n^2 \\ &\le \frac{1}{3} \frac{d(d-1)}{2} \frac{q(q-4)}{q-1} \lambda^4 x_n^2. \end{aligned} \tag{32}$$

When $d > \kappa^{-2}$ then we always have $|\lambda| < \kappa$ because $d\lambda^2 \le 1$. For the finite number of cases when $d \le \kappa^2$ by taking δ to be sufficiently small and N to be sufficiently large, we may assume by Corollary 2.16 that when $|\lambda| > \kappa$ and $n > N$ then

$$\left| \frac{z_n}{x_n} - \frac{1}{q} \right| < \kappa.$$

It follows that we may take equation (32) to hold for all d and λ.

Now combining equations (16), (30), (31) and (32) and taking δ and ε to be sufficiently small and N sufficiently large, we complete the result. $\square$

PROOF OF THEOREM 1.2. We will prove the result for the ferromagnetic case, the anti-ferromagnetic case will follow similarly. We will establish that when λ is close enough to $d^{-1/2}$ then x_n does not converge to 0. First, we will verify that x_n does not drop from a very large value to a very small one. Fix some $\kappa < d^{-1/2}$. By Lemma 2.9, there exists $0 < \gamma < 1$ such that if $\kappa < \lambda \le d^{-1/2}$ then $x_{n+1} \ge \gamma x_n$. Now we use Lemma 3.1. We can take $\delta > 0$ and N so that if $n \ge N$ and $x_n < \delta$ then

$$x_{n+1} \ge d\lambda^2 x_n + \frac{1}{2} \frac{d(d-1)}{2} \frac{q(q-4)}{q-1} \lambda^4 x_n^2. \tag{33}$$

Let $\varepsilon = \min\{\frac{1}{2}\gamma^{N+1}, \delta\gamma\} > 0$. Since $q - 4 > 0$, we can choose $\kappa < \lambda < d^{-1/2}$ such that

$$1 \le d\lambda^2 + \frac{1}{2} \frac{d(d-1)}{2} \frac{q(q-4)}{q-1} \lambda^4 \varepsilon. \tag{34}$$

We now show by induction that for all n that $x_n \ge \varepsilon$. Since $x_0 = 1 - \frac{1}{q} > \frac{1}{2}$, then $x_n \ge \frac{1}{2}\gamma^n \ge \varepsilon$ when $n \le N$ so suppose that $n > N$. Now if $x_n \ge \varepsilon\gamma^{-1}$ then $x_{n+1} \ge \gamma x_n \ge \varepsilon$. If $\varepsilon \le x_n \le \gamma^{-1}\varepsilon \le \delta$, then by Lemma 3.1 and equation (34) we have that

$$\begin{aligned} x_{n+1} &\ge d\lambda^2 x_n + \frac{1}{2} \frac{d(d-1)}{2} \frac{q(q-4)}{q-1} \lambda^4 x_n^2 \\ &\ge x_n \left(d\lambda^2 + \frac{1}{2} \frac{d(d-1)}{2} \frac{q(q-4)}{q-1} \lambda^4 \varepsilon \right) \\ &\ge x_n. \end{aligned}$$

It follows by induction that for all n, $x_n \geq \varepsilon$ which implies that $\lambda^+ \leq \lambda < d^{-1/2}$ which establishes that the Kesten–Stigum bound is not tight. $\square$

4. Large degree asymptotics. In this section, we will analyze what happens as we let d grow. As d increases, the interactions become weaker and λ decreases. We will parameterize the interaction strengths with $\hat{\lambda}$ defined by $\hat{\lambda} = \hat{\lambda}(d) = \lambda d^{1/2}$. With this parameterization, $\hat{\lambda} = 1$ corresponds to the Kesten–Stigum bound in the ferromagnetic case while $\hat{\lambda} = -1$ corresponds to the Kesten–Stigum bound in the antiferromagnetic case. We will, therefore, restrict our attention to $|\hat{\lambda}| \leq 1$. We define

$$U_{ij} = \log\left(1 + \lambda q\left(Y_{ij} - \frac{1}{q}\right)\right)$$

and denote $U_j = (U_{1j}, \dots, U_{qj}) \in \mathbb{R}^q$. We have the following estimates on the means and covariances of the U_{ij}.

LEMMA 4.1. *There exists constants C and d' depending only on q such that when $d > d'$,*

$$\left|dEU_{1j} - \tfrac{1}{2}\hat{\lambda}^2 q x_n\right| \leq C d^{-1/2}, \tag{35}$$

and for $i \geq 2$,

$$\left|dEU_{ij} + \left(\frac{1}{2} + \frac{1}{q-1}\right)\hat{\lambda}^2 q x_n\right| \leq C d^{-1/2}. \tag{36}$$

For any $1 \leq i \leq q$,

$$|d \operatorname{Var}(U_i) - \hat{\lambda}^2 q x_n| \leq C d^{-1/2} \tag{37}$$

and for and $1 \leq i_1 < i_2 \leq q$,

$$\left|d \operatorname{Cov}(U_{i_1 j}, U_{i_2 j}) + \frac{1}{q-1}\hat{\lambda}^2 q x_n\right| \leq C d^{-1/2}. \tag{38}$$

PROOF. Using the Taylor series expansion of $\log(1 + w)$, there exists a constant $W > 0$ such that when $|w| < W$ then $|\log(1 + w) - w + \frac{1}{2}w^2| \leq |w|^3$. Since by definition $0 \leq Y_{ij} \leq 1$ by taking d' to be sufficiently large we may assume that $|\lambda q(Y_{ij} - \frac{1}{q})| \leq |\lambda| q \leq W$ since $|\lambda| \leq d^{-1/2}$. Then by Lemma 2.5,

$$\begin{aligned} E\left|U_{1j} - \lambda q\left(Y_{ij} - \frac{1}{q}\right) + \frac{1}{2}\lambda^2 q^2\left(Y_{ij} - \frac{1}{q}\right)^2\right| &\leq E|\lambda|^3 q^3\left|Y_{ij} - \frac{1}{q}\right|^3 \\ &\leq d^{-3/2} q^3 E\left|Y_{ij} - \frac{1}{q}\right|^3 \\ &\leq q^3 d^{-3/2}. \end{aligned} \tag{39}$$

Now since by Lemma 2.2, $0 \le z_n \le x_n \le 1$ and applying the identities of Lemma 2.5,

$$\begin{aligned}
&\left| E\lambda q\left(Y_{ij} - \frac{1}{q}\right) - E\frac{1}{2}\lambda^2 q^2\left(Y_{ij} - \frac{1}{q}\right)^2 - \frac{1}{2}\lambda^2 q x_n\right| \\
&\quad = \left|\lambda^2 q x_n - \frac{1}{2}\lambda^2 q^2\left(\lambda z_n + \frac{1}{q}(1-\lambda)x_n\right) - \frac{1}{2}\lambda^2 q x_n\right| \\
&\quad = \frac{1}{2}|\lambda|^3 q^2\left|z_n - \frac{1}{q}x_n\right| \\
&\quad \le \frac{1}{2}q^2 d^{-3/2}.
\end{aligned} \tag{40}$$

Combining equation (39) and (40) establishes equation (35). Equations (36)–(38) follow similarly. $\square$

Since the random vectors $Y_j = (Y_{1j}, \dots, Y_{qj})$ are independent and identically distributed so are the $U_j = (U_{1j}, \dots, U_{qj})$ for $j = 1, \dots, d$. Also each U_{ij} satisfies

$$|U_{ij}| \le \max\{\log(1 + d^{-1/2}q), |\log(1 - d^{-1/2}q)|\} \to 0$$

as $d \to \infty$. Such a collection of random vectors suggests the use of a central limit theorem.

The following standard proposition can be established using the central limit theorem and Gaussian approximation.

PROPOSITION 4.2. *Let $\psi : \mathbb{R}^q \mapsto \mathbb{R}$ be a differentiable bounded function and let $\varepsilon > 0$. Let $V_1, \dots, V_D$ be a sequence of i.i.d. q-dimensional vectors denoted $V_j = (V_{1j}, \dots, V_{qj})$. Let $\mu \in \mathbb{R}^q$ be a vector and let $\Sigma \in \mathbb{R}^{q\times q}$ be a positive semi-definite symmetric $q \times q$-matrix. Let $(W_1, \dots, W_q)$ be distributed according to the q-dimensional Gaussian vector $N(\mu, \Sigma)$.*

Suppose there exists some $C > 0$ such that for $1 \le i < j \le q$ the following holds: $\|\mu_i\|_\infty \le C$, $\|\Sigma_{ij}\|_\infty \le C$, $\|\mu - DEV_1\|_\infty \le CD^{-1/2}$ and $\|\Sigma - D\,\mathrm{Cov}(V_1)\|_\infty \le CD^{-1/2}$ and $\|\cdot\|_\infty$ denotes the standard L^∞ norm. Then there exists a D' depending only on q, C and ψ such that if $D > D'$ then

$$\left|\psi\left(\sum_{i=1}^q V_{1j}, \dots, \sum_{i=1}^q V_{qj}\right) - \psi(W_1, \dots, W_q)\right| \le \varepsilon.$$

Let μ be the q-dimensional vector given by

$$\mu_i = \begin{cases} \dfrac{q}{2}, & i = 1, \\ -q\left(\dfrac{1}{2} + \dfrac{1}{q-1}\right), & i \ne 2, \end{cases}$$

and let Σ is the $q \times q$-covariance matrix given by

$$\Sigma_{ij} = \begin{cases} q, & i = j, \\ -\dfrac{q}{q-1}, & i \neq j. \end{cases}$$

Define

$$\psi(w_1, \dots, w_q) = \frac{e^{w_1}}{\sum_{i=1}^q e^{w_i}}.$$

The function ψ is positive, analytic and bounded by 1. Now if $(W_1, \dots, W_q)$ is a Gaussian vector distributed according to $N(0, \Sigma)$ then $(s\mu_1 + \sqrt{s}W_1, \dots, s\mu_q + \sqrt{s}W_q)$ is distributed according to $N(s\mu, s\Sigma)$. We define

$$\begin{aligned} g(s) = g_q(s) &= E\psi\bigl(s\mu_1 + \sqrt{s}W_1, \dots, s\mu_q + \sqrt{s}W_q\bigr) - \frac{1}{q} \\ &= \frac{e^{s\mu_1 + \sqrt{s}W_1}}{\sum_{i=1}^q e^{s\mu_i + \sqrt{s}W_i}} - \frac{1}{q}. \end{aligned} \tag{41}$$

Since $Z_i = \exp(\sum_{i=1}^q U_{ij})$, we have that

$$x_{n+1} = E\frac{Z_1}{\sum_{i=1}^q Z_i} - \frac{1}{q} = E\psi\left(\sum_{j=1}^d U_{1j}, \dots, \sum_{j=1}^d U_{qj}\right) - \frac{1}{q}.$$

Then Proposition 4.2 and Lemma 4.1 immediately imply the following lemma.

LEMMA 4.3. *For each $\varepsilon > 0$ there exists a d' such that when $d > d'$,*

$$|x_{n+1} - g(\hat{\lambda}^2 x_n)| \leq \varepsilon.$$

Understanding the function $g_q(s)$, and in particular the solutions to the equation $g_q(s) = s$, provides key information into the reconstruction problem when d is large. Since $0 < x_n \leq \frac{q-1}{q}$, we will restrict our attention on g to this interval.

LEMMA 4.4. *For each q, the function g_q is continuously differentiable on the interval $(0, \frac{q-1}{q}]$ and increasing.*

PROOF. Since

$$\sup_x \left|\frac{d}{dx}\frac{e^x}{1+e^x}\right| = \sup_x \left|\frac{e^x}{(1+e^x)^2}\right| = \frac{1}{4} \tag{42}$$

we have that when $s > 0$,

$$E\left|\frac{d}{ds}\psi\bigl(s\mu_1 + \sqrt{s}W_1, \dots, s\mu_q + \sqrt{s}W_q\bigr)\right| \leq \frac{1}{4}E\sum_{i=1}^q \left|\frac{d}{ds}s\mu_i + \sqrt{s}W_i\right| < \infty$$

which establishes that g_q is differentiable. Now let $(\widetilde{W}_1, \widetilde{W}_2, \dots, \widetilde{W}_q)$ be an independent copy of $(W_1, \dots, W_q)$. Then when $0 \le s' < s$ the following equality in distribution holds:

$$\sqrt{s}(W_1, \dots, W_q) \stackrel{d}{=} \sqrt{s'}(W_1, \dots, W_q) + \sqrt{s - s'}(\widetilde{W}_1, \widetilde{W}_2, \dots, \widetilde{W}_q).$$

Recall that if W is distributed as $N(\mu, s^2)$ then $Ee^W = e^{\mu+(1/2)s^2}$. For $2 \le i \le q$, since $\widetilde{W}_i - \widetilde{W}_1$ is distributed as $N(0, 2q + \frac{2q}{q-1})$,

$$\begin{aligned} &E\big[\exp(\sqrt{s'}(W_i - W_1) + \sqrt{s-s'}(\widetilde{W}_i - \widetilde{W}_1))|\{W\}_{j=1}^q\big] \\ &\quad = \exp\biggl(\sqrt{s'}(W_i - W_1) + (s - s')\Bigl(q + \frac{q}{q-1}\Bigr)\biggr). \end{aligned}$$

Noting that $\frac{1}{1+u}$ is convex, by Jensen's inequality

$$\begin{aligned} g_q(s) &= E\psi(s\mu_1 + \sqrt{s}W_1, \dots, s\mu_q + \sqrt{s}W_q) - \frac{1}{q} \\ &= E\Biggl(1\Big/\Biggl(1 + \sum_{i=2}^q \exp\biggl(-s\Bigl(q + \frac{q}{q-1}\Bigr) + \sqrt{s'}(W_i - W_1) \\ &\qquad\qquad + \sqrt{s-s'}(\widetilde{W}_i - \widetilde{W}_1)\biggr)\Biggr)\Biggr) - \frac{1}{q} \\ &\ge E\Biggl(1\Big/\Biggl(1 + E\Biggl[\sum_{i=2}^q \exp\biggl(-s\Bigl(q + \frac{q}{q-1}\Bigr) + \sqrt{s'}(W_i - W_1) \\ &\qquad\qquad + \sqrt{s-s'}(\widetilde{W}_i - \widetilde{W}_1)\biggr)\Big| \\ &\qquad\qquad \{W\}_{j=1}^q\Biggr]\Biggr)\Biggr) - \frac{1}{q} \\ &= E\frac{1}{1 + \sum_{i=2}^q \exp(-s'(q + q/(q-1)) + \sqrt{s'}(W_i - W_1))} - \frac{1}{q} \\ &= g_q(s') \end{aligned}$$

which establishes that $g_q(s)$ is increasing. $\square$

LEMMA 4.5. *For all q and small s, we have that*

$$g_q(s) = s + \frac{1}{2}\frac{(q-4)q}{q-1}s^2 + \frac{1}{6}\frac{(q^2 - 18q + 42)q^2}{(q-1)^2}s^3 + O(s^4) \tag{43}$$

and so when $q \ge 5$ there is a root $0 < s^* < \frac{q-1}{q}$ to the equation $g(s^*) = s^*$.

PROOF. Using the identity

$$\frac{a}{r+s} = \left(\sum_{i=1}^{m}(-1)^{i-1}\frac{ar^{i-1}}{s^i}\right) + (-1)^m \frac{r^m}{s^m}\frac{a}{r+s}$$

and taking $a = \exp(s\mu_1 + \sqrt{s}W_1)$, $s = q$ and $r = (\sum_{i=1}^q \exp(s\mu_i + \sqrt{s}W_i) - q)$, we have that

$$\begin{aligned} g_q(s) &= E\psi(s\mu_1 + \sqrt{s}W_1, \dots, s\mu_q + \sqrt{s}W_q) - \frac{1}{q} \\ &= E\sum_{i=1}^{4}(-1)^{i-1}\frac{(\sum_{i=1}^q \exp(s\mu_i + \sqrt{s}W_i) - q)^{i-1}\exp(s\mu_1 + \sqrt{s}W_1)}{q^i} \qquad (44) \\ &\quad + E\frac{(\sum_{i=1}^q \exp(s\mu_i + \sqrt{s}W_i) - q)^4}{q^4}\frac{\exp(s\mu_1 + \sqrt{s}W_1)}{\sum_{i=1}^q \exp(s\mu_i + \sqrt{s}W_i)} - \frac{1}{q}. \end{aligned}$$

Now again using the fact that if W is distributed as $N(\mu, s^2)$ then $Ee^W = e^{\mu + s^2/2}$ and doing Taylor series expansions with the help of Mathematica we have that

$$\begin{aligned} &E\sum_{i=1}^{4}(-1)^{i-1}\frac{(\sum_{i=1}^q \exp(s\mu_i + \sqrt{s}W_i) - q)^{i-1}\exp(s\mu_1 + \sqrt{s}W_1)}{q^i} \\ &\quad = (4qe^{6qs} + 6e^{qs(q-10)/(q-1)} - e^{10qs} + 8e^{3(q-2)sq/(q-1)}q^2 \\ &\qquad - 3e^{2qs(3q-5)/(q-1)}q + 3e^{2qs(3q-5)/(q-1)} - 6e^{2qs(q-5)/(q-1)} \\ &\qquad - q^3 - 6q^2e^{3qs} + 4e^{2qs(2q-5)/(q-1)} - 11e^{qs(q-10)/(q-1)}q \\ &\qquad - 12e^{qs(q-6)/(q-1)}q^2 - e^{qs(q-10)/(q-1)}q^3 + 4e^{2qs(q-3)/(q-1)}q^2 \\ &\qquad - 4e^{2qs(q-3)/(q-1)}q + 4e^{qs(q-6)/(q-1)}q^3 + 8e^{qs(q-6)/(q-1)}q \\ &\qquad - 4e^{2qs(2q-5)/(q-1)}q - 3e^{qs(-10+3q)/(q-1)}q^2 - 3e^{2qs(q-5)/(q-1)}q^2 \\ &\qquad + 9e^{2qs(q-5)/(q-1)}q + 6e^{qs(q-3)/(q-1)}q^2 - 6q^3e^{qs(q-3)/(q-1)} \\ &\qquad - 6e^{qs(-10+3q)/(q-1)} + 4q^3e^{qs} - 8e^{3(q-2)sq/(q-1)}q \\ &\qquad\qquad + 6e^{qs(q-10)/(q-1)}q^2 + 9e^{qs(-10+3q)/(q-1)}q)q^{-4} \\ &\quad = \frac{1}{q} + s + \frac{1}{2}\frac{(q-4)q}{q-1}s^2 + \frac{1}{6}\frac{(q^2 - 18q + 42)q^2}{(q-1)^2}s^3 + O(s^4) \end{aligned}$$

and

$$\begin{aligned} &E\frac{(\sum_{i=1}^q \exp(s\mu_i + \sqrt{s}W_i) - q)^4}{q^4} \\ &\quad = -(4qe^{6qs} + 60e^{qs(q-10)/(q-1)} - e^{10qs} + 16e^{3(q-2)sq/(q-1)}q^2 \end{aligned}$$

$$
\begin{aligned}
&- 5e^{2qs(3q-5)/(q-1)}q + 4e^{-6qs/(q-1)}q^4 - 6q^4e^{-3qs/(q-1)}\\
&- 12e^{-3qs/(q-1)}q^2 - e^{-10qs/(q-1)}q^4 - 35e^{-10qs/(q-1)}q^2\\
&- 24e^{-6qs/(q-1)}q + 50e^{-10qs/(q-1)}q + 44e^{-6qs/(q-1)}q^2\\
&+ 5e^{2qs(3q-5)/(q-1)} + 10e^{-10qs/(q-1)}q^3 - 30e^{2qs(q-5)/(q-1)}\\
&- q^4 - 6q^2e^{3qs} + 4q^4e^{-qs/(q-1)} + 10e^{2qs(2q-5)/(q-1)}\\
&- 110e^{qs(q-10)/(q-1)}q - 72e^{qs(q-6)/(q-1)}q^2 - 10e^{qs(q-10)/(q-1)}q^3\\
&+ 12e^{2qs(q-3)/(q-1)}q^2 - 12e^{2qs(q-3)/(q-1)}q + 24e^{qs(q-6)/(q-1)}q^3\\
&+ 48e^{qs(q-6)/(q-1)}q - 10e^{2qs(2q-5)/(q-1)}q - 10e^{qs(-10+3q)/(q-1)}q^2\\
&- 15e^{2qs(q-5)/(q-1)}q^2 + 45e^{2qs(q-5)/(q-1)}q + 18e^{qs(q-3)/(q-1)}q^2\\
&- 18q^3e^{qs(q-3)/(q-1)} - 24e^{-10qs/(q-1)} - 24e^{-6qs/(q-1)}q^3\\
&+ 18q^3e^{-3qs/(q-1)} - 20e^{qs(-10+3q)/(q-1)} - 4q^3e^{-qs/(q-1)} + 4q^3e^{qs}\\
&- 16e^{3(q-2)sq/(q-1)}q + 60e^{qs(q-10)/(q-1)}q^2 + 30e^{qs(-10+3q)/(q-1)}q)q^{-4}\\
=\ & O(s^4).
\end{aligned}
$$

Since $0 \le \frac{(\sum_{i=1}^q \exp(s\mu_i+\sqrt{s}W_i)-q)^4}{q^4}$ and $0 \le \frac{\exp(s\mu_1+\sqrt{s}W_1)}{\sum_{i=1}^q \exp(s\mu_i+\sqrt{s}W_i)} \le 1$ combining these estimates establishes equation (43).

Since $q-4>0$ when $q \ge 5$ for small $s>0$ we have that $g_q(s) > s$. Since

$$
g_q\left(1-\frac{1}{q}\right) = E\psi(s\mu_1+\sqrt{s}W_1,\dots,s\mu_q+\sqrt{s}W_q) - \frac{1}{q} < 1-\frac{1}{q}
$$

by the Intermediate Value theorem, there must be some $0 < s^* < \frac{q-1}{q}$ such that $g(s^*) = s^*$. $\square$

THEOREM 4.6. *When $q \ge 5$, define*

$$
w^* = \inf\left\{w : \exists 0 < s^* < \frac{q-1}{q}, g(ws^*) = s^*\right\}.
$$

Then $0 < w^* < 1$ and for each $\delta > 0$ there exists a $d'(q,\delta)$ such that if $d > d'$ then the model has reconstruction when $\hat\lambda^2 \ge w^* + \delta$ but does not have reconstruction when $\hat\lambda^2 \le w^* - \delta$.

PROOF. The key idea of this result is that when $\hat\lambda^2 > w^*$, $g_q(\hat\lambda s)$ has a nonzero attractive fixed point as a function of s while if $\hat\lambda < w^*$ then $g_q(\hat\lambda s) < s$ for $s > 0$. By Lemma 4.5, we have the expansion $g_q(s) = s + \frac{1}{2}\frac{(q-4)q}{q-1}s^2 + o(s^2)$ so for small s, $g_q(s) > s$. It also implies that for any $0 < w < 1$, the set $\{0 < s <$

$\frac{q-1}{q} : g_q(ws) \geq s\}$ is a compact set bounded away from 0. By the continuity of g_q,

$$\left\{0 < s < \frac{q-1}{q} : g(w^*s) = s\right\} = \bigcap_{w^*<w<1} \left\{0 < s < \frac{q-1}{q} : g(ws) \geq s\right\}$$

and by the Finite Intersection Property of compact sets it is nonempty and compact so let $s^* \in \{0 < s < \frac{q-1}{q} : g(w^*s) = s\}$.

Now set $\hat{\lambda}^2 = w^* + \delta$ and so

$$g_q\left((w^*+\delta)\left(s^*\frac{w^*}{w^*+\delta}\right)\right) = g_q(s^*w^*) = s^* > s^*\frac{w^*}{w^*+\delta}.$$

Take d large enough so that Lemma 4.3 holds with $0 < \varepsilon < s^* - s^*\frac{w^*}{w^*+\delta}$. Then when $x_n > s^*\frac{w^*}{w^*+\delta}$ since g_q is monotone it follows that

$$\begin{aligned} x_{n+1} &\geq g_q\big((w^*+\delta)x_n\big) - \varepsilon \\ &> g_q\left((w^*+\delta)\left(s^*\frac{w^*}{w^*+\delta}\right)\right) - \left(s^* - s^*\frac{w^*}{w^*+\delta}\right) \\ &= s^*\frac{w^*}{w^*+\delta} \end{aligned}$$

and hence $\inf x_n \geq s^*\frac{w^*}{w^*+\delta}$ which establishes reconstruction.

By equation (20)

$$|x_{n+1} - \hat{\lambda}^2 x_n| \leq C_q \lambda^4 \frac{d(d-1)}{2} x_n^2 \leq C_q x_n^2,$$

where C_q does not depend on d or $\hat{\lambda}$. So when $|\hat{\lambda}| < 1$ and if $x_n < \frac{1-\hat{\lambda}^2}{2C_q}$ then

$$x_{n+1} \leq \hat{\lambda}^2 x_n + C_q x_n^2 \leq \hat{\lambda}^2 x_n + \frac{1-\hat{\lambda}^2}{2} x_n < \frac{1+\hat{\lambda}^2}{2} x_n.$$

When $\hat{\lambda}^2 < w^*$ then $g(\hat{\lambda}^2 s) \leq \frac{\hat{\lambda}^2}{w^*} s$ and so by Lemma 4.3 for large enough d, we have that for some n, $x_n < \frac{1-\hat{\lambda}^2}{2C_q}$. It follows then that x_n converges to 0 which proves nonreconstruction for large enough d. $\square$

4.1. *Nonreconstruction for $q = 3$.*

LEMMA 4.7. *When $q = 3$ for all $0 \leq s \leq \frac{q-1}{q}$, then $g_q(s) < s$.*

We defer this proof to the Appendix.

LEMMA 4.8. *When $q = 3$ there exists a $\delta > 0$ and N not depending on d or λ such that if $x_n \leq \delta$ and $n > N$ then*

$$x_{n+1} \leq d\lambda^2 x_n - \frac{3}{4}\frac{d(d-1)}{2}\lambda^4 x_n^2.$$

The proof is essentially identical to the proof of Lemma 3.1 and so we omit it.

PROOF OF THEOREM 1.1. At the Kesten–Stigum bound, we have that $|\hat{\lambda}| = 1$. Since $g(s) < s$ for all $s > 0$ by Lemma 4.3 there exists a d' such that when $d > d'$ and m is sufficiently large then $x_m < \delta$ where δ is the constant in Lemma 4.8. It follows from Lemma 4.8 that if for some m, $x_m < \delta$ then $\lim_n x_n = 0$ and hence nonreconstruction. □

APPENDIX: DEFERRED PROOF

PROOF OF LEMMA 4.7. Recall that μ is the q-dimensional vector given by

$$\mu_i = \begin{cases} \frac{q}{2}, & i = 1, \\ -q\left(\frac{1}{2} + \frac{1}{q-1}\right), & i \neq 2, \end{cases}$$

and that Σ is the $q \times q$-covariance matrix given by

$$\Sigma_{ij} = \begin{cases} q, & i = j, \\ -\frac{q}{q-1}, & i \neq j. \end{cases}$$

With $(W_1, \ldots, W_q)$ a Gaussian vector distributed according to $N(0, \Sigma)$ the function $g_q(s)$ is defined as

$$g_q(s) = E\psi(s\mu_1 + \sqrt{s}W_1, \ldots, s\mu_q + \sqrt{s}W_q) - \frac{1}{q},$$

where

$$\psi(w_1, \ldots, w_q) = \frac{e^{w_1}}{\sum_{i=1}^{q} e^{w_i}}.$$

In this lemma we consider the case of $q = 3$. By equation (42) we have that for any x, y,

$$\left|\frac{e^x}{1+e^x} - \frac{e^y}{1+e^y}\right| \le \frac{1}{4}|x-y|, \qquad \left|\frac{1}{1+e^x} - \frac{1}{1+e^y}\right| \le \frac{1}{4}|x-y|.$$

Using this estimate and the fact that $E|W_i| = \sqrt{\frac{6}{\pi}}$ it follows that

$$\begin{aligned} |g_3(s_1) - g_3(s_2)| &\le \frac{1}{4}\sum_{i=1}^{3} |\mu_i(s_1 - s_2)| + |\sqrt{s_1} - \sqrt{s_2}| E|W_i| \\ &= \frac{15}{8}|s_1 - s_2| + \sqrt{\frac{27}{8\pi}}|\sqrt{s_1} - \sqrt{s_2}|. \end{aligned}$$

Now $\max_{x\in[0.1,2/3]} \frac{d}{dx} x^{1/2} = \frac{1}{2}\sqrt{10}$. Hence if we take $0.1 \le s_1 < s_2 \le \frac{2}{3}$ then

$$|g_3(s_1) - g_3(s_2)| \le \left(\frac{15}{8} + \sqrt{\frac{135}{16\pi}}\right)|s_1 - s_2| \le 3|s_1 - s_2|. \tag{45}$$

Let

$$\mathcal{S} = \left\{\frac{100}{1000}, \frac{101}{1000}, \dots, \frac{667}{1000}\right\}$$

and suppose that

$$\forall s^* \in \mathcal{S} \qquad g_3(s^*) - s^* < -\frac{5}{1000}. \tag{46}$$

Now fix some $s \in [0.1, \frac{2}{3}]$. Then for some $s^* \in \mathcal{S}$, $|s - s^*| < \frac{1}{1000}$ which implies that

$$\begin{aligned} g_3(s) - s &\le g_3(s^*) - s^* + |g_3(s) - g_3(s^*)| + |s - s^*| \\ &< -\frac{5}{1000} + 4|s - s^*| + |s - s^*| < 0, \end{aligned}$$

where the second inequality follows from equation (45). So proving equation (46) would imply that $g_3(s) < s$ for all $0.1 \le s \le \frac{2}{3}$. We do this by a rigorous method of numerical integration.

Let U_1, U_2 be independent standard Gaussians. The random vectors $(W_2 - W_1, W_3 - W_1)$ and $(3U_1, \frac{3}{2}U_1 + \frac{3\sqrt{3}}{2}U_2)$ have the same covariance matrix and therefore are equal in distribution. Hence

$$\begin{aligned} g_3(s) &= E\frac{1}{1 + \sum_{i=2}^3 \exp(-(9s)/2 + \sqrt{s}(\widetilde{W}_i - \widetilde{W}_1))} - \frac{1}{3} \\ &= E\biggl(1\Big/\biggl(1 + \exp\biggl(-\frac{9s}{2} + 3\sqrt{s}U_1\biggr) \\ &\qquad + \exp\biggl(-\frac{9s}{2} + \frac{3}{2}\sqrt{s}U_1 + \frac{3\sqrt{3}}{2}\sqrt{s}U_2\biggr)\biggr)\biggr) - \frac{1}{3} \\ &= \int_{\mathbb{R}^2}\biggl(1\Big/\biggl(1 + \exp\biggl(-\frac{9s}{2} + 3\sqrt{s}x\biggr) + \exp\biggl(-\frac{9s}{2} + \frac{3}{2}\sqrt{s}x + \frac{3\sqrt{3}}{2}\sqrt{s}y\biggr)\biggr)\biggr) \\ &\qquad \cdot \frac{\exp(-x^2/2 - y^2/2)}{2\pi}\,dx\,dy - \frac{1}{3} \qquad (47) \\ &\le \int_{-5}^{5}\int_{-5}^{5}\biggl(1\Big/\biggl(1 + \exp\biggl(-\frac{9s}{2} + 3\sqrt{s}x\biggr) \\ &\qquad + \exp\biggl(-\frac{9s}{2} + \frac{3}{2}\sqrt{s}x + \frac{3\sqrt{3}}{2}\sqrt{s}y\biggr)\biggr)\biggr) \\ &\qquad \cdot \frac{\exp(-x^2/2 - y^2/2)}{2\pi}\,dx\,dy - \frac{1}{3} + 10^{-5}, \end{aligned}$$

where the inequality uses the standard inequality that

$$\int_x^\infty \frac{\exp(-x^2/2)}{\sqrt{2\pi}}\,dx \le \frac{\exp(-x^2/2)}{x\sqrt{2\pi}}$$

which implies that

$$\iint_{\mathbb{R}^2\setminus[-5,5]^2} \frac{\exp(-x^2/2-y^2/2)}{2\pi} \le 4\frac{\exp(-5^2/2)}{5\sqrt{2\pi}} \le 10^{-5}.$$

Define the function $\phi(i)=\min\{|i|,|i+1|\}$. Then for integers i and j,

$$\begin{aligned}
&\int_{i/200}^{(i+1)/200}\int_{j/200}^{(j+1)/200}\Bigl((\exp(-x^2/2-y^2/2)\,dx\,dy)\\
&\qquad\Big/\Bigl(1+\exp\Bigl(-\frac{9s}{2}+3\sqrt{s}x\Bigr)\\
&\qquad\qquad+\exp\Bigl(-\frac{9s}{2}+\frac{3}{2}\sqrt{s}x+\frac{3\sqrt{3}}{2}\sqrt{s}y\Bigr)\Bigr)2\pi\Bigr)\\
&\le\Bigl(\exp\Bigl(-\Bigl(\frac{\phi(i)}{200}\Bigr)^2/2-\Bigl(\frac{\phi(j)}{200}\Bigr)^2/2\Bigr)40{,}000^{-1}\Bigr)\\
&\quad\Big/\Bigl(1+\exp\Bigl(-\frac{9s}{2}+3\sqrt{s}\frac{i}{200}\Bigr)\\
&\qquad+\exp\Bigl(-\frac{9s}{2}+\frac{3}{2}\sqrt{s}\frac{i}{200}+\Bigl(\frac{3\sqrt{3}}{2}\Bigr)\sqrt{s}\frac{j}{200}\Bigr)\Bigr)2\pi.
\end{aligned}\tag{48}$$

Let $\psi(i,j)$ denote the right-hand side of equation (48). Substituting this bound in (47) we have that

$$g_3(s) \le -\frac{1}{3}+10^{-5}+\sum_{i=-1000}^{999}\sum_{j=-1000}^{999}\psi(i,j).\tag{49}$$

The right-hand side of equation (49) is merely a combination of basic arithmetic operations and exponentials and so can be rigorously computed to arbitrarily high precision (e.g., in Mathematica). Evaluating this expression for each $s^*\in\mathcal{S}$ establishes equation (46). As noted above this implies that $g(s)<s$ when $s\in[0.1,\frac{2}{3}]$.

It remains to show that $g_3(s)<s$ when $0<s\le 0.1$. Using equation (44) and noting that

$$\frac{\exp(s\mu_1+\sqrt{s}W_1)}{\sum_{i=1}^3\exp(s\mu_i+\sqrt{s}W_i)}\le 1$$

we have that

$$g_3(s) \le E\sum_{i=1}^{4}(-1)^{i-1}\frac{(\sum_{i=1}^{3}\exp(s\mu_i+\sqrt{s}W_i)-3)^{i-1}\exp(s\mu_1+\sqrt{s}W_1)}{3^i} + E\frac{(\sum_{i=1}^{3}\exp(s\mu_i+\sqrt{s}W_i)-3)^4}{81}-\frac{1}{3}.$$

Using the fact that if W is distributed as $N(\mu,\sigma^2)$ then $Ee^W = e^{\mu+\sigma^2/2}$ we have after simplifying that

$$g_3(s) \le \frac{74}{27}-\frac{4}{27}e^{-9s/2}+\frac{4}{27}e^{3s}-\frac{202}{81}e^{-3s/2} + \frac{8}{27}e^{-6s}+\frac{4}{81}e^{12s}-\frac{16}{27}e^{9s/2}. \tag{50}$$

By Taylor's theorem we have that if $|x| \le 1.2$ then

$$\left|\exp(x)-\sum_{i=0}^{5}\frac{x^i}{i!}\right| \le \frac{x^6}{6!}\max_{y\in[-1.2,1.2]}\left|\frac{d^6e^y}{dy^6}\right| \le 2\frac{x^6}{6!}.$$

Applying this to equation (50) we get that when $0 \le s \le 0.1$ that

$$g_3(s)-s \le \frac{1}{1280}s^2h(s),$$

where

$$h(s) = -960-1440s+58{,}860s^2+98{,}334s^3+595{,}795s^4.$$

Now $h(s)$ is convex and $h(0)<0$ and $h(0.1)<0$ which imples that $h(s)<0$ for all $0 \le s \le 0.1$. It follows that $g_3(s)<s$ for all $0<s\le 0.1$ which completes the proof. □

Acknowledgments. The author would like to thank Elchanan Mossel for his encouragement, insightful discussions and careful reading of a draft of this paper. Thanks also to Elitza Maneva and an anonymous referee for their careful reading of this paper. Most of the work was carried out while the author was a student at UC Berkeley.

REFERENCES

[1] ACHLIOPTAS, D. and COJA-OGHLAN, A. (2008). Algorithmic barriers from phase transition. In *Proceedinds of FOCS'08* 793–802.

[2] BERGER, N., KENYON, C., MOSSEL, E. and PERES, Y. (2005). Glauber dynamics on trees and hyperbolic graphs. *Probab. Theory Related Fields* **131** 311–340. MR2123248

[3] BHATNAGAR, N. and MANEVA, E. (2010). A computational method for bounding the probability of reconstruction on trees. *SIAM J. Discrete Math.* To appear.

[4] BLEHER, P. M., RUIZ, J. and ZAGREBNOV, V. A. (1995). On the purity of the limiting Gibbs state for the Ising model on the Bethe lattice. *J. Statist. Phys.* **79** 473–482. MR1325591

[5] BORGS, C., CHAYES, J. T., MOSSEL, E. and ROCH, S. (2006). The Kesten–Stigum reconstruction bound is tight for roughly symmetric binary channels. In *FOCS* 518–530. IEEE Comput. Soc., Berkeley, CA.

[6] CHAYES, J. T., CHAYES, L., SETHNA, J. P. and THOULESS, D. J. (1986). A mean field spin glass with short-range interactions. *Comm. Math. Phys.* **106** 41–89. MR0853978

[7] DASKALAKIS, C., MOSSEL, E. and ROCH, S. (2006). Optimal phylogenetic reconstruction. In *STOC'*06: *Proceedings of the* 38*th Annual ACM Symposium on Theory of Computing* 159–168. ACM, New York. MR2277141

[8] DYER, M., FRIEZE, A., HAYES, T. P. and VIGODA, E. (2004). Randomly coloring constant degree graphs. In *FOCS* '04: *Proceedings of the* 45*th Annual IEEE Symposium on Foundations of Computer Science* (*FOCS'*04) 582–589. IEEE Computer Society, Washington, DC.

[9] EVANS, W., KENYON, C., PERES, Y. and SCHULMAN, L. J. (2000). Broadcasting on trees and the Ising model. *Ann. Appl. Probab.* **10** 410–433. MR1768240

[10] FORMENTIN, M. and KÜLSKE, C. (2009). On the purity of the free boundary condition Potts measure on random trees. *Stochastic Process. Appl.* **119** 2992–3005. MR2554036

[11] GERSCHENFELD, A. and MONTANARI, A. (2007). Reconstruction for models on random graphs. In 48*th Annual Symposium on Foundations of Computer Science*. Providence, RI.

[12] IOFFE, D. (1996). On the extremality of the disordered state for the Ising model on the Bethe lattice. *Lett. Math. Phys.* **37** 137–143. MR1391195

[13] KESTEN, H. and STIGUM, B. P. (1966). Additional limit theorems for indecomposable multidimensional Galton–Watson processes. *Ann. Math. Statist.* **37** 1463–1481. MR0200979

[14] KRZĄKAŁA, F., MONTANARI, A., RICCI-TERSENGHI, F., SEMERJIAN, G. and ZDEBOROVÁ, L. (2007). Gibbs states and the set of solutions of random constraint satisfaction problems. *Proc. Natl. Acad. Sci. USA* **104** 10318–10323 (electronic). MR2317690

[15] MARTINELLI, F., SINCLAIR, A. and WEITZ, D. (2007). Fast mixing for independent sets, colorings, and other models on trees. *Random Structures Algorithms* **31** 134–172. MR2343716

[16] MÉZARD, M. and MONTANARI, A. (2006). Reconstruction on trees and spin glass transition. *J. Stat. Phys.* **124** 1317–1350. MR2266446

[17] MONTANARI, A., RESTREPO, R. and TETALI, P. (2009). Reconstruction and clustering in random constraint satisfaction problems. Available at http://arxiv.org/abs/0904.2751.

[18] MOSSEL, E. (2001). Reconstruction on trees: Beating the second eigenvalue. *Ann. Appl. Probab.* **11** 285–300. MR1825467

[19] MOSSEL, E. (2004). Phase transitions in phylogeny. *Trans. Amer. Math. Soc.* **356** 2379–2404 (electronic). MR2048522

[20] MOSSEL, E. (2004). Survey: Information flow on trees. In *Graphs, Morphisms and Statistical Physics. DIMACS Ser. Discrete Math. Theoret. Comput. Sci.* **63** 155–170. Amer. Math. Soc., Providence, RI. MR2056226

[21] MOSSEL, E. and PERES, Y. (2003). Information flow on trees. *Ann. Appl. Probab.* **13** 817–844. MR1994038

[22] MOSSEL, E. and SLY, A. (2009). Gibbs rapidly samples colorings of $g(n, d/n)$. *Probab. Theory Related Fields* **149** 37–69.

[23] POLLARD, D. (1984). *Convergence of Stochastic Processes*. Springer, New York. MR0762984

[24] SEMERJIAN, G. (2008). On the freezing of variables in random constraint satisfaction problems. *J. Stat. Phys.* **130** 251–293. MR2375745

[25] SLY, A. (2009). Reconstruction of random colourings. *Comm. Math. Phys.* **288** 943–961. MR2504861

[26] ZDEBOROVÁ, L. and KRZĄKAŁA, F. (2007). Phase transitions in the coloring of random graphs. *Phys. Rev. E* (3) **76** 031131, 29. MR2365566

MICROSOFT RESEARCH
ONE MICROSOFT WAY
REDMOND, WASHINGTON 98052-6399
USA
E-MAIL: sly@stat.berkeley.edu