A Study on the Non-Reconstruction Conjecture in Markov Trees

Yurong Liu

Abstract

Consider a d-ary tree T which simulates the process of broadcasting information from the root to other vertices, where each edge is a copy of an irreducible and aperiodic Markov chain M with reversible transition matrix $\mathbf{M} \in \mathbf{R}^{n \times n}$ on state space Ω, the goal is to reconstruct the value of the root given values of nodes at level n of the tree, where $n \to \infty$. This branching process is useful for modeling complex populations that exhibit dependencies between the states of individuals and their ancestors. It can be used to study a wide range of phenomena, including the spread of diseases in populations, the growth of organisms in ecosystems, and the diffusion of information and ideas. We are going to work on the non-reconstruction conjecture of this problem. The conjecture states that information on root cannot be reconstructed if $|\lambda_2(\mathbf{M})| < \frac{1}{d}$, where $\lambda_2(\mathbf{M})$ is the second largest eigenvalue of $\mathbf{M}$. Our focus is on the scenario where $\mathbf{M}$ is symmetric.

1 Introduction

The study of information propagation has gained significant attention in recent years due to its wide-ranging applications in diverse domains such as epidemiology, ecology, and social network analysis. The ability to model the behavior of these systems, as well as the limitations of information recovery, can provide valuable insights into the underlying mechanisms driving their dynamics.

In this paper, we investigate a specific instance of information broadcasting in a d-ary tree, wherein the edges represent irreducible and aperiodic Markov chains with a symmetric transition matrix. The d-ary tree T serves as a natural model for representing the process of broadcasting information from a root node to the remaining vertices. Each edge in this tree is a copy of an irreducible and aperiodic Markov chain M with a reversible transition matrix $\mathbf{M} \in \mathbf{R}^{n \times n}$ on the state space Ω. Our objective is to reconstruct the value of the root node based on the values of nodes at level n of the tree, as $n \to \infty$. This branching process is particularly relevant in the context of modeling complex populations that exhibit dependencies between the states of individuals and their ancestors.

We dedicate our efforts to understanding the non-reconstruction conjecture associated with this problem. The conjecture posits that it is not possible to reconstruct information about the root node if $|\lambda_2(\mathbf{M})| < \frac{1}{d}$, where $\lambda_2(\mathbf{M})$ represents the second largest eigenvalue of the transition matrix $\mathbf{M}$. This paper contributes to the existing body of knowledge by shedding light on the conditions under which the non-reconstruction conjecture holds by focusing on the case where $\mathbf{M}$ is symmetric, thereby offering a deeper understanding of the limitations of information propagation in d-ary trees with symmetric Markov chain edges. Furthermore, our findings have the potential to inform the design of more efficient strategies for managing the spread of diseases, the growth of organisms, and the diffusion of information and ideas in complex systems.

2 Related Works

2.1 Applications of the non-reconstruction conjecture

One of the prominent applications of the non-reconstruction conjecture lies in the field of phylogeny reconstruction in computational biology [1, 2]. Phylogenetic trees are graphical models that represent the evolutionary relationships among various species, tracing back to their common ancestors. In this context, the d-ary tree with symmetric Markov chain edges can serve as a model for the evolutionary process, where each vertex represents a species and each edge signifies the transmission of genetic information. The non-reconstruction conjecture provides valuable insights into the conditions under which it is possible to infer the ancestral state as well as the entire evolutionary tree based on the genetic information of the extant species, thus contributing to the development of robust phylogenetic inference methods.

Additionally, this study has implications for the understanding of the mixing properties of Markov chains on random graphs, since the measure of how rapidly a Markov chain converges to its stationary distribution, also known as the case where sequences of random graphs converging locally to trees, is closely related to the ability to reconstruct the original state from the chain's current state [3, 4]. In essence, the non-reconstruction conjecture implies that certain Markov chains exhibit rapid mixing, making the reconstruction of the root state increasingly difficult.

The non-reconstruction conjecture also finds relevance in the context of the replica symmetric phase of random constraint satisfaction problems (CSPs), where the solutions to a CSP are organized in a tree-like structure, similar to the d-ary tree considered in our study. The conjecture in this case sheds light on the conditions under which the structure of the solution space can be recovered from partial information. Understanding the non-reconstruction conjecture in the context of CSPs can have implications for the landscape of combinatorial optimization problems. It can also be linked to the reconstruction problem for the Potts model [5], a generalization of the Ising model used in statistical mechanics to describe the behavior of interacting particles in a lattice. In this setting, the d-ary tree with symmetric Markov chain edges can be viewed as a lattice structure, where each node represents a particle with one of the possible discrete states.

2.2 Existing reconstruction methods

Various reconstruction methods have been developed to address the problem of inferring the root state in a d-ary tree. One such method to use **maximum likelihood estimation (MLE)**, which is consistent for inferring the tree topology [6]. In particular, we find the optimal assignment of states to the root node that maximizes the likelihood of the observed data. Another approach is the **census** method, which involves observing whether the census of the configuration at level n contains any significant information on the root variable. Reconstruction (and census) solvability when $d\lambda_2(\mathbf{M}) > 1$ was initially demonstrated in [7], though it was expressed in the context of multi-type branching processes which we will later introduce in §3.3. The proofs of the non-reconstruction result when $d\lambda_2(\mathbf{M}) \leq 1$ are harder as shown in [8], where it's also demonstrated that the asymptotic independence of the root in the census is determined by the spectral properties of $\mathbf{M}$.

3 Preliminaries

3.1 Markov chains

In this section, we introduce the basic concepts and notations related to Markov chains, which will be employed throughout the paper to analyze the non-reconstruction conjecture in information broadcast over d-ary trees.

A Markov chain is a stochastic process that models the transition between states in a system, where the future state depends only on the current state and not on the past states. This property is known as the Markov property.

Definition 1. *(Markov Chain) A* ***Markov chain*** *is a sequence of random variables $X_n, n \in \mathbf{N}$ taking values in a finite or countable state space Ω and satisfying the Markov property: for any $n \in \mathbf{N}$ and any states $x_0, x_1, \ldots, x_{n+1} \in \Omega$,*

$$P(X_{n+1} = x_{n+1} | X_0 = x_0, X_1 = x_1, \ldots, X_n = x_n) = P(X_{n+1} = x_{n+1} | X_n = x_n).$$

In this paper, we are mainly dealing with time homogeneous Markov chains, where the probability of moving from one state to another is constant over time and does not depend on when the transition occurs. Hence we have the transition matrix that captures the probabilities of transitioning between the states of a Markov chain.

Definition 2. *(Transition Matrix) Let M be a Markov chain with state space Ω. The* ***transition matrix*** $\mathbf{M} \in \mathbf{R}^{|\Omega| \times |\Omega|}$ *of M is a matrix such that $\mathbf{M}_{ij}$ is the probability of transitioning from state i to state j:*

$$\mathbf{M}_{ij} = P(X_{n+1} = j | X_n = i), i, j \in \Omega,$$

where $\forall i, j \in \Omega, \mathbf{M}_{ij} \geq 0$, and $\sum_{j=1}^{n} \mathbf{M}_{ij} = 1$ for all $i \in \Omega$.

A stationary distribution is a probability distribution over the state space of a Markov chain that remains invariant under the transition probabilities.

Definition 3. *(Stationary Distribution) Let M be a Markov chain with transition matrix $\mathbf{M}$. A probability distribution π over the state space Ω is a* ***stationary distribution*** *of $\mathbf{M}$ if*

$$\pi \mathbf{M} = \pi.$$

Note that another way to express this is that π is an eigenvector with all its elements being nonnegative, and its associated eigenvalue is 1.

Example 1. *Consider a Markov chain represented by a random walk on the nodes of an n-cycle. At each step, there is a $1/2$ probability of staying at the current node, a $1/4$ probability of moving left, and a $1/4$ probability of moving right. The uniform distribution, which assigns a probability of $1/n$ to each node, acts as a stationary distribution for this chain, because it remains constant after performing a single step in the chain.*

For Markov chains, irreducibility and aperiodicity are essential properties that ensure the existence and uniqueness of a stationary distribution.

Definition 4. *(Irreducibility) A Markov chain with transition matrix $\mathbf{M}$ is* ***irreducible*** *if there exists a sequence of transitions between any pair of states $i, j \in \Omega$ with positive probability*

$$\forall i, j \in \Omega, \exists t \in \mathbf{N} \text{ s.t. } (\mathbf{M}^t)_{ij} > 0.$$

Definition 5. *(Aperiodicity) A Markov chain with transition matrix $\mathbf{M}$ is* ***aperiodic*** *if for all states $i \in \Omega$, the greatest common divisor of the set $\{t \in \mathbf{N} : (\mathbf{M}^t)_{ii} > 0\}$ equals 1.*

Theorem 1. *If a Markov chain M is irreducible then it has a unique stationary distribution π.*

A Markov chain is said to be **ergodic** if it is both irreducible and aperiodic. Hence we derive the definition of ergodicity as follows.

Theorem 2. *(Convergence to stationary distribution) If a Markov chain M is ergodic, then there exists a unique stationary distribution π such that for any given (initial) distribution α, $\lim_{t \to \infty} \alpha \mathbf{M} t = \pi$.*

Definition 6. *(Reversibility) An ergodic Markov chain is* ***reversible*** *if the stationary distribution π satisfies the detailed balance equations: $\forall i, j \in \Omega, \pi_i \mathbf{M}_{ij} = \pi_j \mathbf{M}_{ji}$*

3.2 Coupling

Coupling is a technique used in probability theory to study the convergence of Markov chains. It involves constructing two Markov chains on the same probability space that eventually couple or synchronize their states. We will employ this technique in subsequent proofs. In short, the term coupling in probability refers to creating a joint distribution from two separate distributions, ν and γ, with the resulting joint distribution having ν and γ as its marginals. This coupling can provide valuable insight into the difference between the two distributions, measured by the total variation distance. Suppose ν, γ are two distributions on Ω, we want to define measures that enable us to compare ν and γ.

Definition 7. *(Coupling) A **coupling** ω is a joint distribution on $\Omega \times \Omega$ such that*

$$\forall y, \sum_{x\in\Omega} \omega(x,y) = \gamma(y),$$
$$\forall x, \sum_{y\in\Omega} \omega(x,y) = \nu(x).$$

where ν, γ are two distributions on Ω.

Example 2. *Consider a Markov chain on the state space $\Omega = \{0,1\}$ with the following transition probability matrix $\mathbf{M}$:*

$$\mathbf{M} = \begin{pmatrix} 0.7 & 0.3 \\ 0.6 & 0.4 \end{pmatrix}$$

We want to study the convergence of this Markov chain to its stationary distribution. To do this, we construct two copies of the Markov chain, say X and Y, with initial states x_0 and y_0, respectively, where $x_0 \neq y_0$. Now we define a coupling of these two chains such that:

- *If $X_t = Y_t$: 1) If $X_t = Y_t = 0$, then $X_{t+1} = Y_{t+1}$ with probability 0.7 both chains move to state 0, and with probability 0.3 both chains move to state 1; 2) if $X_t = Y_t = 1$, then $X_{t+1} = Y_{t+1}$ with probability 0.6 both chains move to state 0, and with probability 0.4 both chains move to state 1*
- *If $X_t \neq Y_t$: 1) If $X_t = 0, Y_t = 1$, the chains have a 0.24 chance of moving to state (1, 1), 0.4 chance of moving to state (0, 0), 0.36 chance of staying in the same state, and 0 chance of moving to state (1, 0); 2) if $X_t = 1, Y_t = 0$, the chains have a 0.24 chance of moving to state (0, 0), 0.4 chance of moving to state (1, 1), 0.36 chance of staying in the same state, and 0 chance of moving to state (0, 1).*

Hence the coupling matrix can be written as

$$\begin{pmatrix} 0.7 & 0 & 0 & 0.3 \\ 0.4 & 0.36 & 0 & 0.24 \\ 0.4 & 0 & 0.36 & 0.24 \\ 0.6 & 0 & 0 & 0.4 \end{pmatrix}$$

Note that this is only one possible coupling for the given Markov chain. Coupling works as long as the following conditions are satisfied:

- *If X and Y are in the same state (i.e., $X_t = Y_t$), they stay synchronized (i.e., $X_{t+1} = Y_{t+1}$)*
- *If X and Y are in different states, they may synchronize with some probability*

By constructing the coupled Markov chains X and Y, we can analyze the synchronization time (i.e., the time it takes for the chains to reach the same state) and use this information to study the convergence to the stationary distribution.

We also introduce a measure of the difference between two probability distributions. It is defined as the sum of the absolute differences between the probabilities assigned to each event by the two distributions.

Definition 8. *(Total Variation Distance) The **total variation distance** between probability distributions ν and γ is defined as*

$$d_{TV} := \sup_{A \in \Omega} |\nu(A) - \gamma(A)|, \tag{1}$$

and when Ω is countable, it also holds that

$$d_{TV} := \frac{1}{2}||\nu - \gamma||_1 = \frac{1}{2}\sum_{x \in \Omega} |\nu(x) - \gamma(x)|. \tag{2}$$

There's also an alternative way of defining coupling using random variables instead of distributions. Let X, Y be a pair of random variables with probability distributions ν and γ on Ω, that is

$$\mathbf{P}(X = a) = \nu(a), \text{ for all } a \in \Omega,$$
$$\mathbf{P}(Y = b) = \gamma(b), \text{ for all } b \in \Omega.$$

Then we define a joint distribution ω of (X, Y) on $\Omega \times \Omega$ to be $\omega(x, y) = \mathbf{P}(X = x, Y = y)$. Hence ω is a probability measure on the product space $\Omega \times \Omega$ s.t. the marginals of ω coincide with ν and γ.

Hence we can use coupling to bound the distance between probability measure.

Lemma 1. *Let ν and γ be two probability distributions on Ω, then for any coupling (X, Y) of ν and γ,*

$$d_{TV}(\nu, \gamma) \leq \mathbf{P}(X \neq Y). \tag{3}$$

Proof. For any event $A \subset \Omega$ and coupling (X, Y) for ν and γ,

$$\begin{aligned}\nu(A) - \gamma(A) &= \mathbf{P}[X \in A] - \mathbf{P}[Y \in A] \\ &= \mathbf{P}[X \in A, X = Y] + \mathbf{P}[X \in A, X \neq Y] - \mathbf{P}[Y \in A, X = Y] - \mathbf{P}[Y \in A, X \neq Y] \\ &= \mathbf{P}[X \in A, X \neq Y] - \mathbf{P}[Y \in A, X \neq Y] \\ &\leq \mathbf{P}[X \neq Y]\end{aligned}$$

The intuition is that we want to find a coupling (X, Y) s.t. $X \neq Y$ only if $\nu(x) \neq \gamma(x)$ i.e. x is in the marginals of ω coincide with ν and γ. The second line involves three cases when we randomly select a point x in Ω: 1) $X \in A, Y \in A$; 2) $X \in A, Y \notin A$; 3) $X \notin A, Y \in A$. In case 1), we set $X = Y$; in case 2) and 3), we set $X \neq Y$. Similarly, we can show that

$$\gamma(A) - \nu(A) \leq \mathbf{P}[X \neq Y],$$

and hence

$$d_{TV} = \sup_{A \in \Omega} |\nu(A) - \gamma(A)| \leq \mathbf{P}[X \neq Y],$$

□

3.3 Galton-Watson Branching Process

The Galton-Watson branching process (or GW-process for short) is a mathematical model that describes the evolution of a population over time. Formally, the GW-process can be defined as a discrete-time branching process, where the number of offspring produced by each individual in the population is modeled as a random variable. This random variable is typically assumed to follow a certain probability distribution, such as the Poisson distribution or the geometric distribution, which determines the average number of offspring and the variance in the number of offspring. The size of the population at any given time is given by the sum of the number of offspring produced by each individual in the previous generation.

The GW-process is used to model a variety of real-world systems, including the spread of diseases, the growth of populations, and the evolution of species. By analyzing the behavior of the GW-process, it is possible to obtain information about the long-term behavior of the population, such as the probability of extinction or the average population size over time.

Example 3. *Consider a branching process modeling population growth, where each individual can have 0, 1, or 2 offspring with probabilities* $0.4, 0.4$ *and* 0.2*, respectively. Starting with a single individual (generation 0), the process unfolds in discrete generations. Each individual in generation* n *produces a random number of offspring (0, 1, or 2) according to the given probabilities, forming generation* $n+1$*. This Galton-Watson process models the evolution of the population over time, capturing growth or extinction dynamics.*

3.3.1 Single-type Branching Process

The most common formulation of a branching process is **Galton–Watson** process.

Definition 9. *A Galton-Watson process is a discrete-time Markov chain* $\{M_n = 0, 1, 2, \dots\}$ *on* Ω*, where* M_n *denote the number of individuals on* n^{th} *level, with transition function defined in terms of offspring distribution* $\{p_k\}$*, where* $k = 0, 1, 2, \cdots$*,* $p_k \geq 0$*, and* $\sum p_k = 1$*, by*

$$\mathbf{P}(i, j) = \mathbf{P}\{M_{n+1} = j | M_n = i\} = p_j^{*i},\ i, j \in \mathbf{Z}_{\geq 0}^{\kappa}$$

where $\{p_k^{*i}\}$ *denotes the* i^{th} *convolution power of the* $\{p_k; k = 0, 1, 2, \dots\}$*.*

Denote the mean and variance of offspring distribution $\{p_k\}$ as μ and σ. Note that $\mu, \sigma > 0$ and are finite. Now we can calculate the first moment of M_n. Since given M_n, we have

$$E[M_{n+1} \mid M_n] = \mu M_n \Rightarrow E[M_{n+1}] = \mu E[M_n],$$

then since $E[M_0] = 1$, it follows that

$$E[M_n] = \mu^n. \tag{4}$$

In general, there are two types of branching processes: subcritical branching processes and supercritical branching processes. In subcritical branching processes, the population eventually goes extinct, while in supercritical branching processes, the population grows without bounds. The critical threshold between these two types of processes is determined by the mean number of offspring produced by each individual. Therefore, we introduce the following theorem that expresses the probability of extinction of branching process [9].

Theorem 3. *If* $\mu < 1$*, then with probability 1 the branching process* M *extincts eventually, i.e. there exists* $N \in \mathbf{M}_{\geq 0}$ *s.t.* $M_n = 0$ *for some any* $n \geq N$*.*

Proof.

$$E\left[\sum M_n\right] = \sum \mu^i \leq \frac{1}{1 - \mu}$$

when $\mu < 1$. Since the sum of number of offspring on each level is finite, the process must die out at some point. □

For instance, in Example 3, the mean of offspring $\mu = 0.8 < 1$. Then according to Theorem 3, with probability 1 the branching process will go extinct eventually.

3.3.2 Multi-type Branching Process

In many scenarios, the individuals in a branching process are not identical. Some examples of this include: 1) Population Genetics - where the inheritance of alleles can be modeled by a 3-type branching process that corresponds to the genotypes; 2) Physics - such as cosmic-ray cascades that involve both electrons and photons and can be modeled by a 2-type branching process. **Multi-type branching process** refers to a mathematical model that describes the evolution of a population in which individuals can give rise to offspring of multiple types, and the number and type of offspring is determined by a probability distribution that depends on the current state of the individual and its ancestry. In our case, we can form the multi-type branching process as [10].

Definition 10. *A multi ($\kappa-$type) Galton-Watson process is a Markov chain* $\{M_n^\kappa = 0, 1, 2, \dots\}$ *on* Ω^κ*, where* M_n *is a* κ*-dimensional vector whose* i^{th} *entry gives the number of individuals of type* i *on the* n^{th} *level, with transition function*

$$\mathbf{P}(\mathbf{x}, \mathbf{y}) = \mathbf{P}\{M_{n+1} = \mathbf{y} | M_n = \mathbf{x}\},\ \mathbf{x}, \mathbf{y} \in \Omega^\kappa.$$

Now let m_{ij} denote the expected number of children of type j that node of type i has, Then we have an associated $m \times m$ matrix $\mathbf{M}$ where each entry m_{ij} is defined by

$$m_{ij} = E[M_{1j} \mid M_0 = i]$$

Then the expected number of each type at level n (denoted as M_n) is

$$E[M_n] = E[M_n \mid M_0] = M_0 \mathbf{M}^n \tag{5}$$

Now similarly, we consider the extinction probability for multi-type branching process.

Theorem 4. *If* $\rho(\mathbf{M}) < 1$*, then with probability 1 the multi-type process* M *extincts eventually.*

Proof. Since if M survives at level n, there exists node(s) of some type(s) at this level, which means $M_n \geq 1$. To ensure smooth transitions of the process M towards success, $E[M_n]$ should be greater or equal to the probability of $M_n \geq 1$. Hence

$$\mathbf{P}[M \text{ survives forever}] \leq \mathbf{P}[M_n \geq 1] \leq E[M_n] = M_0 \mathbf{M}^n,$$

and thus we have

$$\lim_{n \to \infty} \mathbf{P}[M \text{ survives forever}] \leq \lim_{n \to \infty} M_0 \mathbf{M}^n = 0,$$

as $\lim_{n \to \infty} \mathbf{M}^n = 0$ when $\rho(\mathbf{M}) < 1$. Hence M extincts with probability 1. $\square$

4 Problem Definition

We now turn to the reconstruction problem. When the distribution of the process on n^{th} level is independent of the root value as n goes to infinity, we say that the root is **non-reconstructible**. In this case, we have no way to reconstruct given this "same" distribution. Following this intuition, we can formally define non-reconstructibility as follows.

Definition 11. *Given Markov chain M with transition matrix $\mathbf{M}$ and two trees generated from random roots that are independent, where distributions of level n are denoted as ν_n and γ_n, then the root is* ***non-reconstructible*** *if*

$$\lim_{n\to\infty} d_{TV}(\nu_n, \gamma_n) = 0 \tag{6}$$

Following Lemma 1, suppose we create random variables X, Y with probability distributions ν_n and γ_n, then we have

$$\lim_{n\to\infty} \mathbf{P}(X \neq Y) = 0$$

if the root is non-reconstructible.

5 Recap on $\mathbf{M}^{2\times 2}$ transition matrix

Mossel [11] has showed that the information of the root can not be reconstructed for the d-ary tree and binary symmetric channel M where transition matrix

$$\mathbf{M} = \begin{pmatrix} 1-\delta_1 & \delta_1 \\ 1-\delta_2 & \delta_2 \end{pmatrix} \tag{7}$$

when $|\lambda_2(\mathbf{M})| = |\delta_2 - \delta_1| \leq \frac{1}{d}$.

Theorem 5. *Let $\mathbf{M}$ be in form (7). Take integer d s.t. $|d\lambda_2(\mathbf{M})| \leq 1$, then the root is non-reconstructible for the d-ary tree.*

5.1 Proof I

We first introduce the random process called *λ-percolation* [11]. Denote the $d-$ary tree as $T = \{V, E\}$, where V represents the set of vertices (nodes) in T, and E represents the set of edges. Consider $\tau : E \to \{0, 1\}$ which maps from the set of edges to $\{0, 1\}$. Given any $e \in E$, we define $P(\tau(e) = 1) = \lambda$.

Now we prove can prove Theorem 5 following Mossel [11].

Proof. Given transition matrix $\mathbf{M}^{k\times k}$, we first show that

$$\mathbf{M}_{i,j} = \lambda \mathbf{N}_{i,j} + (1-\lambda)\mathbf{v}_j \tag{8}$$

for some broadcasting matrix $\mathbf{N}^{k\times k}$, distribution vectors $(v_j)_{j=1}^{k}$, and a number $\lambda \in (0, 1)$. In this way, we separate the original broadcast process into two parts: 1) copying the original distribution; 2) broadcast via matrix $\mathbf{N}$

Consider $\lambda = \lambda_2(\mathbf{M}) = |\delta_1 - \delta_2|$.

If $\delta_1 - \delta_2 < 0$, then $\lambda = \delta_2 - \delta_1$. Let $\mathbf{N} = \mathbf{I}$, where $\mathbf{I}$ is the identity matrix $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$, and $\mathbf{v} = \frac{(1-\delta_2, \delta_2)}{1-\lambda}$. Then

$$\mathbf{M} = (\delta_2 - \delta_1)\mathbf{I} + \begin{pmatrix} 1-\delta_2 & \delta_1 \\ 1-\delta_2 & \delta_1 \end{pmatrix},$$

so for each row vector $\mathbf{M}_{i,*}$ in $\mathbf{M}$, we have

$$\begin{aligned} \mathbf{M}_{i,*} &= \lambda \mathbf{I}_{i,*} + (1-\lambda) \cdot \frac{(1-\delta_2, \delta_1)}{1-\lambda} \\ &= \lambda \mathbf{N}_{i,*} + (1-\lambda)\mathbf{v}. \end{aligned}$$

Then similarly, if $\delta_1 - \delta_2 > 0$, then $\lambda = \delta_1 - \delta_2$. Let $\mathbf{N} = \mathbf{J}$, where $\mathbf{J} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$, and $\mathbf{v} = \frac{(1-\delta_1, \delta_2)}{1-\lambda}$. Then

$$\mathbf{M} = (\delta_1 - \delta_2)\mathbf{J} + \begin{pmatrix} 1-\delta_1 & \delta_2 \\ 1-\delta_1 & \delta_2 \end{pmatrix},$$

so for each row vector $\mathbf{M}_{i,*}$ in $\mathbf{M}$, we also have

$$\begin{aligned} \mathbf{M}_{i,*} &= \lambda \mathbf{J}_{i,*} + (1-\lambda) \cdot \frac{(1-\delta_1, \delta_2)}{1-\lambda} \\ &= \lambda \mathbf{N}_{i,*} + (1-\lambda)\mathbf{v}. \end{aligned}$$

We now show that when $d\lambda \leq 1$, the root is non-reconstructible given transition matrix $\mathbf{M}$. In fact, for any transition matrix that can be written in the form (8), the broadcast process is non-reconstructible.

We simulate the broadcast on $d-$ary tree $T = \{V, E\}$ with root node $\zeta \in \Omega$ as a λ-*percolation* process. Note that we use the transition matrix $\mathbf{M}$, which means that the probability of from i to j is $\mathbf{M}_{i,j}$. Hence we can define $M : V \to V$ as a random function which satisfies $P(M(i) = j) = \mathbf{M}_{i,j}$. Now since $\mathbf{M}_{i,j} = \lambda \mathbf{N}_{i,j} + (1-\lambda)\mathbf{v}_j$, we can write $M = XN + (1-X)Y$, where N is a random function that satisfies $P(N(i) = j) = \mathbf{N}_{i,j}$, Y is a random variable that satisfies $P(Y = j) = v_j$, and X is a $\{0, 1\}$ variable that satisfies $P(X = 1) = \lambda$. Hence we can simulate the broadcast process in the following way. For any node $v \in V$, let N_v be an independent copy of the function N to simulate the transition through $\mathbf{N}$ from v, and Y_v be an independent copy of the variable Y. Then for any child v' of v, there exists edge (v, v'), for which we define an element of the space $\{0, 1\}^E$ as $\tau(v, v')$, and it satisfies that

$$\mathbf{P}(\tau(v, v') = 1) = \lambda. \tag{9}$$

Then according the definition of $\tau(v, v')$, we define the procedure as follows

$$v' = \begin{cases} N_v(\tau(v)) & \text{if } \tau((v, v') = 1 \\ Y_v & \text{if } \tau((v, v') = 0 \end{cases}$$

Therefore, for any node $v \in V$, we have probability λ to perform the transition by M, and probability $1 - \lambda$ by Y, and the two different processes are independent.

In this way, we obtain a coupling of the two distributions on n^{th} level of T. Let the set of vertices that has path to root node ζ that contains only set of edges E' s.t. $\tau(E') = 1$ be L, and let the set of vertices at n^{th} level be S_n. Let the probability distribution given root, say ζ_ν, at n^{th} level be ν_n. Then if $L \cap S_n = \emptyset$, we obtain same distribution on n^{th} level given any value of root ζ. Then since

$$\max_{\zeta_\nu, \zeta_\gamma \in \Omega} \mathbf{P}(\nu_n \neq \gamma_n) \leq \mathbf{P}(L \cap S_n = \emptyset),$$

and since it has been proved in [12] that when $d\lambda \leq 1$,

$$\lim_{n \to \infty} \mathbf{P}(L \cap S_n = \emptyset) = 0,$$

we have

$$\lim_{n \to \infty} \max_{\zeta_\nu, \zeta_\gamma \in \Omega} \mathbf{P}(\nu_n \neq \gamma_n) = 0,$$

which implies that the root is non-reconstructible. □

5.2 Proof II

An alternative proof using coupling is proposed as follows.

Proof. Say ν_n and γ_n are distributions of n^{th} level of trees started with different root values. Let X_n, Y_n be random variables with probability distributions ν_n and γ_n. By Lemma 1, we have

$$d_{TV}(\nu, \gamma) \leq \mathbf{P}(X \neq Y),$$

so $\mathbf{P}(X \neq Y)$ is an upper bound of d_{TV}. Now since $\mathbf{M} = \begin{pmatrix} 1 - \delta_1 & \delta_1 \\ 1 - \delta_2 & \delta_2 \end{pmatrix}$, $\lambda_2(\mathbf{M}) = |\delta_1 - \delta_2| \leq \frac{1}{d}$. Given any node X_n and corresponding Y_n, let $X_{n+1,i}$ and $Y_{n+1,i}$ ($i \in [0, d-1]$) be the i^{th} child nodes transmitted from them through transition matrix $\mathbf{M}$. We are going to couple $X_{n+1,i}$ and $Y_{n+1,i}$ as follows. If $X_n = Y_n$, nodes transmitted from these two nodes will always agree using same randomness. Therefore, when n approaches infinity, all the nodes in distributions ν_n and γ_n will agree and hence root nodes cannot be reconstructed. If $X_n \neq Y_n$, then we have the probability that $X_{n+1,i}$ and $Y_{n+1,i}$ disagrees is

$$\mathbf{P}(X_{n+1,i} \neq Y_{n+1,i}) = |1 - \delta_1 - (1 - \delta_2)| = |\delta_1 - \delta_2|.$$

Thus given that $|\delta_1 - \delta_2| \leq \frac{1}{d}$, we have

$$\lim_{n\to\infty} \mathbf{P}(X_{n,i} \neq Y_{n,i}) = \lim_{n\to\infty} |\delta_1 - \delta_2|^n = 0,$$

which implies that

$$\lim_{n\to\infty} d_{TV}(\nu_n, \gamma_n) \leq \lim_{n\to\infty} \mathbf{P}(X_n \neq Y_n) = 0.$$

Therefore, we've showed that when n goes to infinity, $X_{n+1,i}$ and $Y_{n+1,i}$ always agree. ☐

6 Extend to $\mathbf{M}^{3\times 3}$ transition matrix

Now we extend Theorem 5 to 3×3 transition matrices, simulating the transitions as **multi-type branching processes** with 3 types. We start with the case when $\mathbf{M}$ is positive definite (PSD).

6.1 Symmetric 3×3 transition matrix with 2 variables

In order to apply coupling, we first consider the following case where transition matrix $\mathbf{M}$ is symmetric and reversible with 2 variables.

Corollary 1. *Let*

$$\mathbf{M} = \begin{pmatrix} 1-\delta_1-\delta_2 & \delta_1 & \delta_2 \\ \delta_1 & 1-\delta_1-\delta_2 & \delta_2 \\ \delta_2 & \delta_2 & 1-2\delta_2 \end{pmatrix} \tag{10}$$

be a positive definite symmetric transition matrix. Then if we take integer d s.t. $|d\lambda_2(\mathbf{M})| \leq 1$, the root is non-reconstructible for the d-ary tree.

Given $\mathbf{M}$ in form 6.3, we denote the three states as A, B, and C, corresponds to row 1, 2, and 3. Note that we have a choice for coupling the two broadcast processes. Now since we want to compare the broadcast distributions given two different root nodes, we define the coupled new states **A A**, **B B**, **C C**, **A B**, **A C**, and **B C**. Note that when it reaches state **A A**, **B B**, or **C C**, two coupled distribution "agrees" and hence extinct.

Hence we only look at types **A B**, **A C**, and **B C**, where two distributions disagree. Hence we let the coupling matrix be in the form

$$\begin{pmatrix} \mathbf{P}(\mathbf{A\ B} \to \mathbf{A\ B}) & \mathbf{P}(\mathbf{A\ B} \to \mathbf{A\ C}) & \mathbf{P}(\mathbf{A\ B} \to \mathbf{B\ C}) \\ \mathbf{P}(\mathbf{A\ C} \to \mathbf{A\ B}) & \mathbf{P}(\mathbf{A\ C} \to \mathbf{A\ C}) & \mathbf{P}(\mathbf{A\ C} \to \mathbf{B\ C}) \\ \mathbf{P}(\mathbf{B\ C} \to \mathbf{A\ B}) & \mathbf{P}(\mathbf{B\ C} \to \mathbf{A\ C}) & \mathbf{P}(\mathbf{B\ C} \to \mathbf{B\ C}) \end{pmatrix}$$

We first want to show that there exists a coupling s.t.

$$\lambda_2(\mathbf{M}) = \rho(\text{coupling matrix}) \tag{11}$$

by doing a case analysis. Note that given $\delta_1 + \delta_2 + \delta_3 = 1$, we use the sign of

$$\begin{aligned}
&\delta_1 - \delta_2 \\
&\delta_2 - 1/3 \\
&2\delta_1 + \delta_2 - 1 \\
&2\delta_2 + \delta_1 - 1
\end{aligned}$$

to determine the sign of the entries of $\mathbf{M}$. We consider $\delta_1 - \delta_2$ and $\delta_2 - 1/3$ as major cases, and the other two in sub cases.

Case I: Let $\delta_1 \leq \delta_2, \delta_2 \leq \frac{1}{3}$, then $2\delta_1 + \delta_2 - 1 \leq 2\delta_2 + \delta_1 - 1 \leq 0 \Rightarrow \delta_1 \leq 1 - \delta_1 - \delta_2, \delta_2 \leq 1 - \delta_1 - \delta_2$.

	A B	**A C**	**B C**
A B	$1 - 2\delta_1 - \delta_2$	0	0
A C	$1 - \delta_1 - 2\delta_2$	0	0
B C	$\delta_2 - \delta_1$	0	$1 - 3\delta_2$

$|\lambda_2(\mathbf{M})| = \rho = 1 - 2\delta_1 - \delta_2$

Case II: Let $\delta_1 \leq \delta_2, \delta_2 > \frac{1}{3}$, then $2\delta_1 + \delta_2 - 1 > 0$ only if $2\delta_2 + \delta_1 - 1 > 0$.

i) $\delta_2 \leq 1 - \delta_1 - \delta_2$, $\delta_1 \leq 1 - \delta_1 - \delta_2$

	A B	**A C**	**B C**
A B	$1 - 2\delta_1 - \delta_2$	0	0
A C	$1 - \delta_1 - 2\delta_2$	0	$3\delta_2 - 1$
B C	$1 - \delta_1 - 2\delta_2$	$3\delta_2 - 1$	0

$\lambda_2(\mathbf{M}) = \rho = 1 - 2\delta_1 - \delta_2$

ii) $\delta_2 > 1 - \delta_1 - \delta_2$, $\delta_1 \leq 1 - \delta_1 - \delta_2$

	A B	**A C**	**B C**
A B	$1 - 2\delta_1 - \delta_2$	0	0
A C	0	$2\delta_2 + \delta_1 - 1$	$\delta_2 - \delta_1$
B C	0	$\delta_2 - \delta_1$	$2\delta_2 + \delta_1 - 1$

$\lambda_2(\mathbf{M}) = \rho = 3\delta_2 - 1$

iii) $\delta_2 > 1 - \delta_1 - \delta_2$, $\delta_1 > 1 - \delta_1 - \delta_2$

	A B	**A C**	**B C**
A B	$2\delta_1 + \delta_2 - 1$	0	0
A C	0	$2\delta_2 + \delta_1 - 1$	$\delta_2 - \delta_1$
B C	0	$\delta_2 - \delta_1$	$2\delta_2 + \delta_1 - 1$

$\lambda_2(\mathbf{M}) = \rho = 3\delta_2 - 1$

Case III: Let $\delta_1 > \delta_2, \delta_2 \leq \frac{1}{3}$, then $2\delta_2 + \delta_1 - 1 > 0$ only if $2\delta_1 + \delta_2 - 1 > 0$.

i) $\delta_1 \leq 1 - \delta_1 - \delta_2$, $\delta_2 \leq 1 - \delta_1 - \delta_2$

	A B	**A C**	**B C**
A B	$1 - 2\delta_1 - \delta_2$	0	0
A C	0	$1 - \delta_1 - 2\delta_2$	$\delta_1 - \delta_2$
B C	0	$\delta_1 - \delta_2$	$1 - 2\delta_2 - \delta_1$

$\lambda_2(\mathbf{M}) = \rho = 1 - 3\delta_2$

ii) $\delta_1 > 1 - \delta_1 - \delta_2$, $\delta_2 \leq 1 - \delta_1 - \delta_2$

	A B	**A C**	**B C**
A B	$2\delta_1 + \delta_2 - 1$	0	0
A C	0	$1 - \delta_1 - 2\delta_2$	$\delta_1 - \delta_2$
B C	0	$\delta_1 - \delta_2$	$1 - \delta_1 - 2\delta_2$

$\lambda_2(\mathbf{M}) = \rho = 1 - 3\delta_2$

iii) $\delta_1 > 1 - \delta_1 - \delta_2$, $\delta_2 > 1 - \delta_1 - \delta_2$

	A B	**A C**	**B C**
A B	$2\delta_1 + \delta_2 - 1$	0	0
A C	$2\delta_2 + \delta_1 - 1$	0	$1 - 3\delta_2$
B C	$2\delta_2 + \delta_1 - 1$	$1 - 3\delta_2$	0

$\lambda_2(\mathbf{M}) = \rho = 2\delta_1 + \delta_2 - 1$

Case IV: Let $\delta_1 > \delta_2, \delta_2 > \frac{1}{3}$, then $2\delta_1 + \delta_2 - 1 > 2\delta_2 + \delta_1 - 1 > 0 \Rightarrow \delta_1 > 1 - \delta_1 - \delta_2, \delta_2 > 1 - \delta_2 - \delta_2$.

	A B	**A C**	**B C**
A B	$2\delta_1 + \delta_2 - 1$	0	0
A C	$\delta_1 - \delta_2$	$3\delta_2 - 1$	0
B C	$\delta_1 - \delta_2$	0	$3\delta_2 - 1$

$\lambda_2(\mathbf{M}) = \rho = 2\delta_1 + \delta_2 - 1$

Let the coupling matrix be $\hat{M}$, where by definition $\hat{M}_{ij}$ refers to the expected number of children of type j that a node of type i has. Now since Eq.11 is satisfied in all cases, by Theorem 4, the multitype branching process will go extinct with probability 1 when $\rho(\hat{M}) < 1$.

6.2 Extend to certain distributions

Claim 1. *Given $d-$ary tree formed by broadcast process M and transition matrix $\mathbf{M}$. Let M_n denote the vector of node counts for each type at level n. If $\lambda_2(\mathbf{M}) < \frac{1}{d}$ and $E[M_n] < d$, then $\mathbf{M}$ on d-ary tree is non-reconstructible.*

In this section we want to show how we can use the coupling method introduced in the previous section to prove Claim 1 under certain children distributions for special cases of M.

Given broadcast matrix $\mathbf{M}$, let $\mathbf{C}$ be a "type-coupling matrix for $\mathbf{M}$", where

$$\mathbf{C} = (1-\alpha)\mathbf{M} + \alpha\mathbf{I}.$$

If we broadcast on d-ary tree using $\mathbf{C}$, then the branching process with probability α preserves and makes a copy of the original distribution, in which we call the the edges go down to "ineffective" nodes, and hence with probability $1-\alpha$ generates "effective" degree distribution through transition M. We can model the branching process with $\mathbf{Bin}(d, \alpha)$. Now given a d-ary tree at some state during the branching process with X number of nodes, then the number of effective nodes (edges) in the next generation, i.e. effective degree, is

$$Xd - (X-1),$$

since Xd is the number of edges that go down from the X nodes, and $X-1$ is the number of edges connected to ineffective nodes. In this way, if we remove ineffective nodes, the tree generated by $\mathbf{C}$ is equivalent to the one generated with $\mathbf{M}$, and thus the non-reconstructibility is related

Now intuitively, given $d-$ary tree formed by broadcast process M and transition matrix $\mathbf{M}$, we want to show that for certain child distribution, the $d-$ary tree with broadcast matrix $(1-\alpha)\mathbf{M} + \alpha\mathbf{I}$ behaves the same way as the tree with broadcast matrix $\mathbf{M}$. In this way, if we have non-reconstruction for $(1-\alpha)\mathbf{M}+\alpha\mathbf{I}$ on $d-$ary tree, then we should have non-reconstruction for M for certain child distribution.

Lemma 2. *Given r.v. X, the number of ineffective nodes, let $P = P_0, P_1, \ldots$ be a distribution of X. Let T be a tree with offspring distribution P. Perform percolation of T keeping each edge with probability $\alpha < \frac{1}{d}$. Let Y be the effective degree, then*

$$E[Y] = \frac{(1-\alpha)d}{1-d\alpha}$$

Proof. Since with probability α, the branching process preserves the ineffective nodes, then we have

$$\begin{aligned}
E[X] &= 1 + \sum_{k=1}^{d} \binom{d}{k} (1-\alpha)^{d-k} \alpha^k \cdot (k \cdot E[X]) \\
&= 1 + \alpha \sum_{k=1}^{d} \frac{d}{k} \binom{d-1}{k-1} (1-\alpha)^{d-k} \alpha^{k-1} \cdot k \cdot E[X] \\
&= 1 + \alpha E[X] d \sum_{k=1}^{d} \binom{\alpha-1}{k-1} (1-\alpha)^{d-k} \alpha^{k-1} \\
&= 1 + d\alpha \cdot E[X]
\end{aligned}$$

which implies that

$$E[X] = \frac{1}{1-d\alpha}$$

and thus

$$\begin{aligned}
E[Y] = E[Xd - (X-1)] &= \frac{1}{1-d\alpha} \cdot d - (\frac{1}{1-d\alpha} - 1) \\
&= \frac{d-1+1-d\alpha}{1-d\alpha} \\
&= \frac{d-1}{1-d\alpha} + 1 \\
&= \frac{(1-\alpha)d}{1-d\alpha}
\end{aligned}$$

□

Hence Theorem 6 follows.

Theorem 6. *Given $d-$ary tree T_M formed by transition matrix $\mathbf{M}$ with second eigenvalue λ_2. Let T_C be a d-regular tree formed by coupling matrix $\mathbf{C} = (1-\alpha)\mathbf{M} + \alpha\mathbf{I}$. Let the expected number of children for T_M and T_C be $E[M]$ and $E[C]$ respectively, and let $E[M] < d, E[C] = d$. Then T_M is non-reconstructible if T_C is non-reconstructible.*

Proof. Suppose T_C is non-reconstructible, then T_M is non-reconstructible since the it corresponds to the effective parts of T_C. Since $\mathbf{C}$ is symmetric and the second eigenvalue of $\mathbf{C}$ is $(1-\alpha)\lambda_2 + \alpha$ by construction, if $d|(1-\alpha)\lambda_2 + \alpha| < 1$, then by Corollary 1, tree T_C formed by $\mathbf{C}$ is non-reconstructible. Now we want to show that if $d|\lambda_2| < 1$, then $d|(1-\alpha)\lambda_2 + \alpha| < 1$.

By Lemma 2, we have

$$d = E[Y] = \frac{(1-\alpha)d}{1-d\alpha}.$$

Then

$$\begin{aligned}
& d|\lambda_2| < 1 \\
\Rightarrow\ & E[M]|\lambda_2| < 1 \\
\Rightarrow\ & \frac{(1-\alpha)d}{1-d\alpha}|\lambda_2| < 1 \\
\Rightarrow\ & (1-\alpha)d|\lambda_2| < 1 - d|\lambda_2| \\
\Rightarrow\ & d|(1-\alpha)\lambda_2 + \alpha| < 1
\end{aligned}$$

Therefore, if $d|\lambda_2| < 1$, $d|(1-\alpha)\lambda_2 + \alpha| < 1$, which implies that T_C is non-reconstructible, hence the effective part of T_C is non-reconstructible, and thus T_M is non-reconstructible.

□

Therefore, we are able to show the non-reconstructibility of the tree with broadcast matrix $\mathbf{M}$ when $E[\text{children}] < d$.

6.3 Generalized case for 3×3 matrix with certain distributions

Now since we've proved in § 6.1 that when the transition matrix is 3×3 and is PSD, coupling proves the conjecture that when $E[\text{number of children}] \cdot \lambda_2(M) < 1$, the root non-reconstructible, we want to extend it to trees with general offspring distributions. We try to prove it case by case after obtaining the coupling matrix following what we did in § 6.1. We start with the 3 3×3 transition matrix in following distribution. Given

$$\mathbf{M} = \begin{pmatrix} 1-a-b & a & b \\ a & 1-a-c & c \\ b & c & 1-b-c \end{pmatrix}$$

whose eigenvalues are 1 and $1 - a - b - c \pm \sqrt{a^2 - ab + b^2 - ac - bc + c^2}$, we have 6 combinations of a, b, c that forms the general cases, which are

$$\begin{aligned}
& a \geq b \geq c \\
& a \geq c \geq b \\
& b \geq a \geq c \\
& b \geq c \geq a \\
& c \geq a \geq b \\
& c \geq b \geq a
\end{aligned}$$

Now consider the expressions

$$\begin{aligned}
& 2a + b, 2a + c \\
& 2b + a, 2b + c \\
& 2c + a, 2c + b
\end{aligned}$$

Let ab, ac, ba, bc, ca, cb be there abbreviations. Then WLOG, given any case, say $a \geq b \geq c$, we have 9 sub-cases given any general case. Given $a \geq b \geq c$, then if $ab \leq 1$, all the other expressions are all less or equal to 1. If $ab \geq 1$, either $ac \leq 1$ or $ba \leq 1$ leads to all the other expressions follows all less or equal to 1. Continue this way, we can have the cases listed below:

$$\left\{\begin{array}{l} ab \geq 1 \left\{\begin{array}{l} ac \geq 1, ba \geq 1 \left\{\begin{array}{l} bc \geq 1, ca \geq 1 \left\{\begin{array}{l} cb \geq 1 \\ cb \leq 1 \end{array}\right. \\ bc \geq 1, ca \leq 1 \\ bc \leq 1, ca \geq 1 \\ bc \leq 1, ca \leq 1 \end{array}\right. \\ ac \geq 1, ba \leq 1 \\ ac \leq 1, ba \geq 1 \\ ac \leq 1, ba \leq 1 \end{array}\right. \\ ab \leq 1 \end{array}\right.$$

Imagine it as a tree. Every leaf node means 1 case where all the expressions follows (in the order of ab, ac, ba, bc, ca, cb for case $a \geq b \geq c$) have to be less or equal to 1.

Hence we have in total 54 cases.

Now similar to what we did for 2×2 matrices, for each case, we compare the second eigenvalue of the transition matrix and the spectral radius of the coupling matrix. Then we notice that when

$$2a + b,\ 2a + c,\ 2b + a > 1$$
$$2b + c,\ 2c + a,\ 2c + b < 1$$

the coupling matrix is as follows

	A B	**A C**	**B C**
A B	$2a + c - 1$	$b - c$	0
A C	$2b + a - 1$	0	$1 - 2b - c$
B C	0	$a - b$	$1 - a - 2c$

and we obtain the result

$$\lambda_2(\mathbf{M}) \neq \rho(\text{coupling matrix}) \tag{12}$$

Hence there exists a case where coupling fails. For instance, let

$$\mathbf{M} = \begin{pmatrix} 0 & \frac{2}{3} & \frac{1}{3} \\ \frac{2}{3} & \frac{1}{3} & 0 \\ \frac{1}{3} & 0 & \frac{2}{3} \end{pmatrix}.$$

Then in the bad case where the coupling matrix is as follows

$$\begin{pmatrix} \frac{1}{3} & \frac{1}{3} & 0 \\ \frac{1}{3} & 0 & \frac{1}{3} \\ 0 & \frac{1}{3} & \frac{2}{3} \end{pmatrix},$$

we have

$$\lambda_2(\mathbf{M}) = \frac{1}{\sqrt{3}}$$

but

$$\rho(\text{coupling matrix}) = \frac{2}{3}.$$

In our future work, we plan to investigate the non-symmetric case further and solve the bad case.

7 Acknowledgment

I would like to give my special thanks to Professor Daniel Stefankovic, a truly insightful and supportive advisor, for granting me the invaluable opportunity to participate in such an interesting theoretical computer science project. His guidance and mentorship throughout the process have been exceptional. Moreover, I would like to express my gratitude to the members of my review committee — Professor Daniel Stefankovic, Professor Sevak Mkrtchyan, and Professor Jonathan Pakianathan — for their time and valuable suggestions on my paper. Furthermore, I am deeply appreciative of the professors who have significantly contributed to my university studies, particularly Professor Steve Gonek and Professor Naomi Jochnowitz. Their guidance and support have been instrumental in my continued pursuit of mathematical studies.

References

[1] Elchanan Mossel. On the impossibility of reconstructing ancestral data and phylogenies. *Journal of computational biology*, 10(5):669–676, 2003.

[2] 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, 2006.

[3] Antoine Gerschenfeld and Andrea Montanari. Reconstruction for models on random graphs. In *48th Annual IEEE Symposium on Foundations of Computer Science (FOCS'07)*, pages 194–204. IEEE, 2007.

[4] Nayantara Bhatnagar, Allan Sly, and Prasad Tetali. Reconstruction threshold for the hardcore model. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques: 13th International*

Workshop, APPROX 2010, and 14th International Workshop, RANDOM 2010, Barcelona, Spain, September 1-3, 2010. Proceedings, pages 434–447. Springer, 2010.

[5] Fa-Yueh Wu. The potts model. *Reviews of modern physics*, 54(1):235, 1982.

[6] Mark Pagel. The maximum likelihood approach to reconstructing ancestral character states of discrete characters on phylogenies. *Systematic biology*, 48(3):612–622, 1999.

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

[8] Elchanan Mossel. Survey: Information flow on trees. *arXiv preprint math/0406446*, 2004.

[9] TIMOTHY Csernica. Extinction in single and multi-type branching processes. 2015.

[10] Krishna B. Athreya and Peter E. Ney. *Multi-Type Branching Processes*, pages 181–228. Springer Berlin Heidelberg, Berlin, Heidelberg, 1972. ISBN 978-3-642-65371-1. doi: 10.1007/978-3-642-65371-$1_5.URL$