PROBABILITY ON GRAPHS: A COMPARISON OF SAMPLING VIA RANDOM WALKS AND A RESULT FOR THE RECONSTRUCTION PROBLEM

by

BLAIR AHLQUIST

A DISSERTATION

Presented to the Department of Mathematics
and the Graduate School of the University of Oregon
in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy

September 2010

University of Oregon Graduate School

Confirmation of Approval and Acceptance of Dissertation prepared by:

Blair Ahlquist

Title:

"Probability On Graphs: A Comparison Of Sampling Via Random Walks And A Result For The Reconstruction Problem"

This dissertation has been accepted and approved in partial fulfillment of the requirements for the degree in the Department of Mathematics by:

David Levin, Chairperson, Mathematics
Christopher Sinclair, Member, Mathematics
Marcin Bownik, Member, Mathematics
Hao Wang, Member, Mathematics
Van Kolpin, Outside Member, Economics

and Richard Linton, Vice President for Research and Graduate Studies/Dean of the Graduate School for the University of Oregon.

September 4, 2010

Original approval signatures are on file with the Graduate School and the University of Oregon Libraries.

An Abstract of the Dissertation of

Blair Ahlquist for the degree of Doctor of Philosophy

in the Department of Mathematics to be taken September 2010

Title: PROBABILITY ON GRAPHS: A COMPARISON OF SAMPLING VIA RANDOM WALKS AND A RESULT FOR THE RECONSTRUCTION PROBLEM

Approved: ______________________________________

Dr. David Levin

We compare the relaxation times of two random walks - the simple random walk and the metropolis walk - on an arbitrary finite multigraph G. We apply this result to the random graph with n vertices, where each edge is included with probability $p = \frac{\lambda}{n}$ where $\lambda > 1$ is a constant and also to the Newman-Watts small world model. We give a bound for the reconstruction problem for general trees and general 2×2 matrices in terms of the branching number of the tree and some function of the matrix. Specifically, if the transition probabilities between the two states in the state space are a and b, we show that we do not have reconstruction if $\mathrm{Br}(T)\,\theta < 1$, where $\theta = \left(\sqrt{(1-a)(1-b)} - \sqrt{ab}\right)^2$ and $\mathrm{Br}(T)$ is the branching number of the tree in question. This bound agrees with a result obtained by Martin for regular trees and is obtained by more elementary methods. We prove an inequality closely related to this problem.

CURRICULUM VITAE

NAME OF AUTHOR: Blair Ahlquist

GRADUATE AND UNDERGRADUATE SCHOOLS ATTENDED:

University of Oregon, Eugene, OR
University of California Santa Barbara, Santa Barbara, CA

DEGREES AWARDED:

Doctor of Philosophy, University of Oregon, 2010
Master of Science, University of Oregon, 2007
Bachelor of Science, University of California Santa Barbara, 2004

PROFESSIONAL EXPERIENCE:

Graduate Teaching Fellow, University of Oregon, 2004-2010

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER I

INTRODUCTION

This dissertation consists of two parts. In the first part, we do a comparison of relaxation times and mixing times for two types of random walks. We apply what we learn to two types of random graphs. In the second part, we prove a bound for the reconstruction problem for general trees. We also prove an inequality which might lead to better results and is of independent interest.

CHAPTER II

A COMPARISON: LAZY METROPOLIS AND SIMPLE RANDOM WALKS

II.1 Introduction

Much has been written on random walks on random graphs. In particular, results are known for the mixing times of the simple random walk on an Erdős-Rényi random graph. Benjamini et al. show in [9] that the mixing time on the largest component of the random graph on n vertices where each edge is added with probability $p = \frac{\lambda}{n}$, where $\lambda > 1$ is constant, is $\theta(\log^2 n)$ with high probability. Similar results for the lazy random walk, for non-constant λ were shown by Fountoulakis and Reed in [7]. Other models with known results for mixing time include the Newman-Watts small world model for which Durrett shows a bound of $O(\log^3 n)$ bound on the mixing time for the lazy random walk ([4]).

In practice, a random walk can be used to sample from a graph, either to choose a random node or to compute the expectation of some function on the vertex set. Since in a simple random walk, the walker will spend more time at nodes of high degree, this skewing must be dealt with in some way depending on the application. Although it is possible to deal with this skewing in the case of approximating a function on the vertex set by adjusting to compensate for degree, an alternative method uses a version of a random walk whose limiting distribution is uniform over the vertex set. This is the *Metropolis walk*. It applies the Metropolis-Hastings algorithm to the simple random walk with target distribution equal to the uniform distribution ([14], [8]). Since the Metropolis walk uses a censoring process to slow the walk down at vertices of low degree, this walk takes longer to mix. But empirical results by Duffield et al. show that sampling using the Metropolis walk can offer more accurate results when sampling compared to sampling by a regular random walk with correction, at least in the case of a graph that is changing with respect to time [16].

Our comparisons will be for the lazy versions of these walks.

II.2 Definitions and Organization of Chapter

Given a graph $G = (V, E)$, the lazy random walk on G moves from vertex v to vertex w with the following probability:

$$P(v,w) = \begin{cases} \frac{1}{2}\frac{1}{\deg v} & \text{if } v \sim w \\ \frac{1}{2} & \text{if } v = w \\ 0 & \text{else} \end{cases},$$

where $v \sim w$ means that $(v, w) \in E$. We can interpret this as the walker staying still with probability $1/2$, and with probability $1/2$, moving to one of the neighboring vertices chosen uniformly at random. This walk has stationary measure $\pi(x) = \frac{\deg x}{2|E|}$, which is the same as the stationary measure for the simple random walk.

The lazy Metropolis walk is similar, but each possible move is censored with a probability depending on the current and destination vertices:

$$\tilde{P}(v,w) = \begin{cases} \frac{1}{2}\frac{1}{\deg v}\left(1 \wedge \frac{\deg v}{\deg w}\right) & \text{if } w \sim v \\ 1 - \frac{1}{2}\sum_{w \sim v} \frac{1}{\deg v}\left(1 \wedge \frac{\deg v}{\deg w}\right) & \text{if } w = v \\ 0 & \text{else} \end{cases} .$$

Because of this censoring, the Metropolis chain has the uniform measure as its stationary measure (it is reversible with respect to the uniform measure). But, consequently, the chain is slowed down - particularly near vertices of high degree where the censoring occurs with higher probability. We compare the relaxation times of the chains by way of their Dirichlet forms. We use the maximum degree of the graph: $\Delta := \mathrm{maxdeg}(G) = \max_{v \in V}\{\deg v\}$ as the primary quantity in this comparison.

One of the random graph models we consider (the Newman-Watts model) is actually a random multigraph (vertices may have self loops, pairs of vertices may have multiple edges), so we need to define these walks for a multigraph G as well. For the lazy random walk on a multigraph, the walker stays still with probability $1/2$ and with probability $1/2$ we choose uniformly from the edges incident to the current vertex and move to the other vertex incident to it, or stay still in the case that it is a self loop. If we define $\deg v$ in this case to be the number of edges incident to v (counting each self loop only once), and $E^* := \sum_{v \in V} \deg v$, then the stationary measure for the lazy random walk on G is

$$\pi(x) = \frac{\deg x}{E^*}.$$

If G is a simple graph (no self loops or multiple edges) then $E^* = 2|E|$. Formally, if for vertices v and w, we define $\#(v,w)$ to be the number of edges connecting v to w, then the lazy random walk on G is defined by

$$P(v,w) = \begin{cases} \frac{1}{2}\frac{\#(v,w)}{\deg v} & \text{if } v \sim w \\ \frac{1}{2} & \text{if } v = w \\ 0 & \text{else} \end{cases},$$

We can now define the lazy Metropolis walk using this notion of degree:

$$\tilde{P}(v,w) = \begin{cases} \frac{1}{2}\frac{\#(v,w)}{\deg v}\left(1 \wedge \frac{\deg v}{\deg w}\right) & \text{if } w \sim v \\ 1 - \frac{1}{2}\sum_{w\sim v}\frac{\#(v,w)}{\deg v}\left(1 \wedge \frac{\deg v}{\deg w}\right) & \text{if } w = v \\ 0 & \text{else} \end{cases}.$$

We give some standard definitions and results involving Markov chains that can be found in many texts (for example, chapters 12 and 13 of [12]).

Given a transition matrix P Markov chain on state space X, and starting distribution μ, the distribution after t steps is μP^t, and this converges in distribution to the stationary distribution π if the chain is irreducible and aperiodic. The notion of distance we use is the *total variation distance* d_{TV} which is defined by

$$d_{TV}(\mu,\nu) := \frac{1}{2}\sum_{x\in X}|\mu(x) - \nu(x)|.$$

The distance to the stationary distribution after t steps is given by

$$d(t) := \max_{\mu}\left\{d_{TV}\left(\mu P^t, \pi\right)\right\}$$

To measure the rate of convergence, we use the *mixing time* t_{mix}. The mixing time with parameter ε is defined by:

$$t_{mix}(\varepsilon) := \min\{t : d(t) \leq \varepsilon\}$$

We use $\varepsilon = 1/4$ and define $t_{mix} := t_{mix}(1/4)$. We are usually interested in how the mixing time relates to the size of the state space when the state space is parameterized by some n that increases to infinity.

A irreducible transition matrix P is reversible if $\pi(x)P(x,y) = \pi(y)P(y,x)$ for all x and y in the state space. The eigenvalues of a reversible transition matrix P satisfy $1 = \lambda_1 > \lambda_2 \geq \ldots \geq \lambda_N \geq -1$. We define $\lambda^* = \max_{i=2}^{N}\{|\lambda_i|\}$, and the *relaxation time* as $t_{rel} := \frac{1}{1-\lambda^*}$ whenever this makes sense. Using the lazy versions of a Markov chain ensures all eigenvalues are positive so that $t_{rel} = \frac{1}{1-\lambda_2}$. The relaxation time is closely related to the mixing time. In fact, we have that

$$t_{rel} = O\left(t_{mix}\right) \tag{II.1}$$

And

$$t_{mix} = O\left(t_{rel}\log\left(\max_{x\in X}\frac{1}{\pi(x)}\right)\right) \tag{II.2}$$

Where the implicit constants are independent of the size of the state space.

Given a function f on the state space, the *Dirichlet kernel* $\mathscr{E}$ is defined by $\mathscr{E}(f) := \langle (I-P)f, f\rangle_\pi$, where the inner product $\langle\cdot,\cdot\rangle_\pi$ is defined by $\langle f, g\rangle_\pi = \sum_{x\in X} f(x)g(x)\pi(x)$. The Dirichlet kernel can be used to compute the relaxation time. We have that

$$1 - \lambda_2 = \min_{f:\mathbb{E}_\pi f = 0, f\neq 0}\frac{\mathscr{E}(f)}{\langle f, f\rangle_\pi}.$$

It can be shown that

$$\mathscr{E}(f) = \frac{1}{2} \sum_{x,y \in X} (f(x) - f(y))^2 \pi(x) P(x,y).$$

The following lemma (Lemma 13.22 in [12]) is useful in comparing the relaxation times of Markov chains:

Lemma II.1. *Let P and $\tilde{P}$ be the transition matrices for reversible Markov chains on state space X. If there exists a constant C such that $\mathscr{E}(f) \leq C\tilde{\mathscr{E}}(f)$ for all f, then*

$$\frac{1}{1-\tilde{\lambda}_2} \leq C \left(\max_{x \in X} \frac{\tilde{\pi(x)}}{\pi(x)} \right) \frac{1}{1-\lambda_2}.$$

If all eigenvalues are positive then we have:

$$\tilde{t}_{rel} \leq C \left(\max_{x \in X} \frac{\tilde{\pi(x)}}{\pi(x)} \right) t_{rel}.$$

In what follows, P, π, $\mathscr{E}$, t_{rel}, and t_{mix} refers to the transition matrix, stationary measure, dirichlet form, relaxation time and mixing time of the lazy random walk, and $\tilde{P}$, $\tilde{\pi}$, $\tilde{\mathscr{E}}$, $\tilde{t}_{rel}$, and $\tilde{t}_{mix}$ refers to the same for the lazy Metropolis walk. We investigate how this comparison works out for two types of random graphs: the Erdös-Rényi model and the Newman-Watts model.

The comparison is done for an arbitrary multigraph in the next section. This result is then used in conjunction with asymptotic results about the maximum degree for these random graphs in subsequent sections.

II.3 The Comparison: Maximum Degree Bound for an Arbitrary Multigraph

Proposition II.1. *If G is any finite connected multigraph and P and $\tilde{P}$ are the transition matrices for the lazy versions of the simple and Metropolis walks respectively, then*

$$\tilde{t}_{rel} \leq \Delta\, t_{rel}. \tag{II.3}$$

Proof. Let f be an arbitrary function on the vertex set. Then

$$\begin{aligned}
2\tilde{\mathscr{E}}(f) &= \sum_{v,w} (f(v) - f(w))^2 \tilde{\pi}(v) \tilde{P}(v,w) \\
&= \sum_{v,w} (f(v) - f(w))^2 \frac{1}{|V|} \frac{1}{2} \frac{\#(v,w)}{\deg v} \left(1 \wedge \frac{\deg v}{\deg w}\right) \\
&= \frac{1}{|V|} \sum_{v,w} (f(v) - f(w))^2 \frac{1}{2} \frac{\#(v,w)}{\deg v \vee \deg w} \\
&\geq \frac{1}{\Delta\,|V|} \sum_{v,w} (f(v) - f(w))^2 \frac{1}{2} \#(v,w) \\
&= \frac{E^*}{\Delta\,|V|} \sum_{v,w} (f(v) - f(w))^2 \frac{\deg v}{E^*} \cdot \frac{1}{2} \frac{\#(v,w)}{\deg v} \\
&= \frac{E^*}{\Delta\,|V|} \sum_{v,w} (f(v) - f(w))^2 \pi(v) P(v,w) \\
&= \frac{E^*}{\Delta\,|V|} 2\mathscr{E}(f).
\end{aligned}$$

Therefore, $\mathscr{E}(f) \leq \frac{\Delta |V|}{E^*} \tilde{\mathscr{E}}(f)$ and consequently, by Lemma II.1, we have that

$$\tilde{t}_{rel} \leq \frac{\Delta\,|V|}{E^*} \left(\max_{v \in V} \frac{\tilde{\pi(v)}}{\pi(v)} \right) t_{rel} = \Delta \left(\max_{v \in V} \frac{1}{\deg v} \right) t_{rel} \leq \Delta\, t_{rel}$$

$\square$

II.3.1 Tightness of Maximum Degree Bounds

We give an example that shows that (II.3) is tight in the sense that equality is achieved (up to a constant) for some graph. The graph is the star with $n+1$ vertices, that is, the graph with vertex set $\{1, \ldots, n+1\}$ where vertices $1, \ldots, n$ are each connected to vertex $n+1$ by an edge, and there are no other edges. Thus there is one vertex of degree n and all other vertices are degree 1 (see Fig. 1).

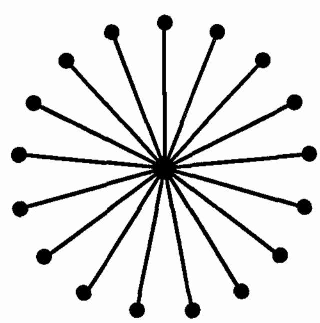

Fig. 1: Star graph

The salient feature of this graph is the one vertex of high degree, which will slow the Metropolis chain considerably. To show tight bounds for (II.3), we find an upper bound for t_{rel} and a lower bound for $\tilde{t}_{rel}$.

For both the upper and lower bounds, we use the bottleneck constant Φ_* defined as follows for a transition matrix P: Given a subset S of the state space, Define

$$Q(S, S^c) := \sum_{v \in S} \sum_{w \in S^c} \pi(v) P(v, w)$$

$$\Phi_* := \min_{S:\pi(S) \leq 1/2} \frac{Q(S, S^c)}{\pi(S)}$$

Jerrum and Sinclair ([10]) show that $\frac{\Phi_*^2}{2} \leq 1 - \lambda_2 \leq 2\Phi_*$. Thus, to get a lower bound on the relaxation time for $\tilde{P}$, we want to get an upper bound on $\tilde{\Phi}_*$. Since we are dealing with reversible chains, we have that $\tilde{Q}(S, S^c) = \tilde{Q}(S^c, S)$, and so without loss of generality we may let S be any subset of the vertex set of the star that contains the center of the star. Then, since the only way to move from S to S^c is through the vertex $n+1$, we have:

$$\begin{aligned}\tilde{Q}(S, S^c) &= \sum_{v \in S} \sum_{w \in S^c} \tilde{\pi}(v)\tilde{P}(v, w) \\ &= \sum_{w \in S^c} \frac{1}{|V|}\tilde{P}(n+1, w) \\ &= \sum_{w \in S^c} \frac{1}{n+1} \cdot \frac{1}{2}\frac{1}{n} \\ &= \frac{n+1-|S|}{2n(n+1)}.\end{aligned}$$

And since $\tilde{\pi}(S) = \frac{|S|}{n+1}$, we have that

$$\frac{\tilde{Q}(S, S^c)}{\tilde{\pi}(S)} = \frac{1 - \tilde{\pi}(S)}{2n}.$$

If we minimize this over sets S with $\tilde{\pi}(S) \leq 1/2$, we get $\tilde{\Phi}_* \leq \frac{1}{2n}\left(1 - \frac{\lfloor n/2 \rfloor}{n}\right) \leq \frac{1}{2n}\left(1 - \frac{n-1}{2n}\right) = \frac{1}{4n}\frac{n+1}{n}$. And this gives us the bound

$$\tilde{t}_{rel} \geq \frac{1}{2\tilde{\Phi}_*} \geq 2n\left(\frac{n}{n+1}\right).$$

For the upper bound, we again let S be any subset of the vertex set that contains the center. Then

$$\begin{aligned} Q(S,S^c) &= \sum_{v\in S}\sum_{w\in S^c} \pi(v)P(v,w) \\ &= \sum_{w\in S^c} \pi(n+1)\cdot P(n+1,w) \\ &= \sum_{w\in S^c} \frac{n}{2n}\cdot\frac{1}{2}\frac{1}{n} \\ &= \frac{n+1-|S|}{4n} \\ &= (1-\pi(S))\,\frac{n+1}{4n}. \end{aligned}$$

Thus,

$$\Phi_* := \min_{S:\pi(S)\le 1/2} \frac{Q(S,S^c)}{\pi(S)} = \min_{S:\pi(S)\le 1/2} \left(\frac{1}{\pi(S)} - 1\right)\frac{n+1}{4n} \ge \frac{n+1}{4n} \ge \frac{1}{4}.$$

This leads to the bound

$$t_{rel} \le \frac{2}{\Phi_*^2} \le 32.$$

Comparing these two bounds, we have

$$\frac{\tilde{t}_{rel}}{t_{rel}} \ge \frac{n}{16}\frac{n}{n+1} = \Omega(\Delta),$$

which shows that our maximum degree bound is asymptotically tight up to a constant.

II.4 Preliminaries for Comparison of Random Walks on Erdös-Rényi Random Graph

The *Erdös-Rényi* random graph is a graph with n vertices where each edge is included with probability p. We denote this probability space by $ER(n, p)$. We take $p = \frac{\lambda}{n}$ where $\lambda > 1$ is a constant. Because $\lambda > 1$, we are in the regime where $ER(n, p)$ has a unique giant component with high probability (see [5]) and we start our walks (somewhere) in this component. To compare our walks on this model we need to find an upper bound for Δ with respect to n for $G \in ER(n, p)$.

II.5 Maximum Degree of an Erdös-Rényi Random Graph

As $\deg v$, for each $v \in V$, is concentrated around λ, there is hope that the maximum of n such random variables will not be too large, but yet should go to infinity with n. An initial observation is that $\deg v \xrightarrow{d} Po(\lambda)$, where $Po(\mu)$ is the Poisson distribution with mean μ. An easier problem is to compute the distribution of the maximum of n independent $Po(\lambda)$ random variables that are independent. In our situation, though, the collection $(\deg v)_{v \in V}$ is not independent. However, there is a very loose dependence, so it will have the same behavior as a collection of independent random variables for a large vertex set. There are some very useful methods for comparing certain types of weakly dependent random variables to a Poisson distribution contained in [1], however, we can get away with something less sophisticated here. This computation is also an exercise (3.5) in [2] (where one is prompted to use the machinery developed there).

Theorem II.1. *For $G \in ER(n,p)$ with $pn = \lambda > 1$ and $\varepsilon > 0$,*

$$\mathbb{P}\left[(1-\varepsilon)\frac{\log n}{\log\log n} \leq \mathrm{maxdeg}(G) \leq (1+\varepsilon)\frac{\log n}{\log\log n}\right] \to 1.$$

The theorem will be proved after a couple of lemmas. An important ingredient will be the Chernoff bounds for a binomial random variable:

Lemma II.2. *Let $X \sim \mathrm{Bin}\left[n-1, \lambda/n\right]$, then $\mathbb{P}\left[X \geq t\right] \leq e^{-t(\log t - log\lambda - 1)-\lambda}$*

Proof. Let $\mu > 0$ be a fixed constant (whose value will be determined.) Using Markov's inequality, we get $\mathbb{P}\left[X \geq t\right] = \mathbb{P}\left[e^{\mu X} \geq e^{\mu t}\right] \leq e^{-\mu t}\,\mathbb{E}e^{\mu X}$. By writing X as a sum of independent Bernoulli random variables we can compute $\mathbb{E}e^{\mu X} = \left(1 + \frac{\lambda}{n}\left(e^{\mu}-1\right)\right)^{n-1} \leq e^{\lambda(e^{\mu}-1)}$. So we get $\mathbb{P}\left[X \geq t\right] \leq e^{-\mu t + \lambda(e^{\mu}-1)}$, which can be minimized by taking $\mu = \log\frac{t}{\lambda}$. □

Lemma II.3. *Let $t_n = \frac{\log n}{\log\log n}$, let γ be a constant, and let $v \in V$. then $n\,\mathbb{P}\left[\deg(v) \geq \gamma t_n\right] \to \infty$ if $\gamma < 1$, and $\to 0$ if $\gamma > 1$.*

Proof. First let $\gamma > 1$. Let $\alpha < 1$ be such that $\gamma\alpha > 1$. For such an α, there exists $N \in \mathbb{N}$ such that $n \geq N$ implies that $\log\left(\gamma\frac{\log n}{\log\log n}\right) - \log\lambda - 1 \geq \alpha\log\log n$. From Lemma II.2, we have

$$\begin{aligned} n\,\mathbb{P}\left[\deg(v) \geq \gamma t_n\right] &\leq n\,e^{-\gamma\frac{\log n}{\log\log n}\left(\log\left(\gamma\frac{\log n}{\log\log n}\right)-\log\lambda-1\right)-\lambda} \\ &\leq n\,e^{-\gamma\alpha\log n - \lambda} \\ &\leq n^{1-\gamma\alpha}, \end{aligned}$$

For $n \geq N$, and this goes to 0 with n. Now let $\gamma \leq 1$.

$$\begin{aligned}
n\,\mathbb{P}\left[\deg(v) \geq \gamma t_n\right] &\geq n\,\mathbb{P}\left[\deg(v) = \lfloor \gamma t_n \rfloor\right] \\
&= n \binom{n-1}{\lfloor \gamma t_n \rfloor} \left(\frac{\lambda}{n}\right)^{\lfloor \gamma t_n \rfloor} \left(1 - \frac{\lambda}{n}\right)^{n-1-\lfloor \gamma t_n \rfloor} \\
&= n \left(1 - \frac{\lambda}{n}\right)^{n-1} \binom{n-1}{\lfloor \gamma t_n \rfloor} \left(\frac{\lambda}{n-\lambda}\right)^{\lfloor \gamma t_n \rfloor} \\
&\geq n \left(1 - \frac{\lambda}{n}\right)^{n-1} \left(\frac{n-1-\lfloor \gamma t_n \rfloor}{n-\lambda}\right)^{\lfloor \gamma t_n \rfloor} \frac{\lambda^{\lfloor \gamma t_n \rfloor}}{\lfloor \gamma t_n \rfloor!}.
\end{aligned}$$

There exists $N \in \mathbb{N}$ such that if $n \geq N$ then $\frac{n-1-\lfloor \gamma t_n \rfloor}{n-\lambda} \geq \frac{1}{\lambda}$ and $\left(1 - \frac{\lambda}{n}\right)^{n-1} \geq e^{-\frac{\lambda}{2}}$. So, for such an n,

$$n\,\mathbb{P}\left[\deg(v) \geq \gamma t_n\right] \geq e^{-\frac{\lambda}{2}} \frac{n}{\lfloor \gamma t_n \rfloor!}.$$

Now, by Stirling's formula, $n! \sim \sqrt{2\pi}\, n^{n+1/2} e^{-n}$, and since $\lfloor n \rfloor \sim n$, we have that:

$$\frac{\sqrt{2\pi}\, n}{\lfloor \gamma t_n \rfloor!} \sim \frac{n\, e^{\gamma t_n}}{(\gamma t_n)^{\gamma t_n + 1/2}} = e^{\log n + \gamma t_n - (\gamma t_n + 1/2) \log \gamma t_n}$$

By replacing n with e^n, we get that this has the same limit as:

$$e^{n + \gamma t(e^n) - (\gamma t(e^n) + 1/2) \log \gamma t(e^n)} = e^{n + \gamma \frac{n}{\log n} - \left(\gamma \frac{n}{\log n} - 1/2\right) \log\left(\gamma \frac{n}{\log n}\right)}.$$

This goes to infinity as the leading term in the exponent is $(1-\gamma)\,n$. Thus $n\,\mathbb{P}\left[\deg(v) \geq \gamma t_n\right] \to \infty$. $\square$

Proof of Thm II.1

Proof. The upper bound follows immediately from Lemma II.3:

$$\begin{aligned}
\mathbb{P}\left[\operatorname{maxdeg}(G) \geq (1+\varepsilon)t_n\right] &= \mathbb{P}\left[\cup_{v\in V}\left\{\deg v \geq (1+\varepsilon)t_n\right\}\right] \\
&\leq \sum_{v\in V} \mathbb{P}\left[\deg v \geq (1+\varepsilon)t_n\right] \\
&= n\mathbb{P}\left[\deg v \geq (1+\varepsilon)t_n\right] \to 0.
\end{aligned}$$

For the lower bound, we use the second moment method. For each $v \in V$, let $I_v = I_v(t) = \mathbf{1}_{\deg v \geq t}$ and let $N = N(t) = \sum_{v\in V} I_v$. N counts the number of vertices whose degree is greater than or equal to t. Since the range of N is in the non-negative integers, $\mathbb{E}N = \mathbb{E}N\mathbf{1}\left(N \geq 1\right) \leq \left(\mathbb{E}N^2\,\mathbb{P}\left(N \geq 1\right)\right)^{1/2}$, which yields

$$\mathbb{P}\left(N \geq 1\right) \geq \frac{\left(\mathbb{E}N\right)^2}{\mathbb{E}N^2}.$$

We work on $\mathbb{E}N^2$ first. Let $\pi_n^t = \mathbb{P}\left[\operatorname{Bin}\left(n-1, \lambda/n\right) \geq t\right]$ and $p_n^t = \mathbb{P}\left[\operatorname{Bin}\left(n-1, \lambda/n\right) = t\right]$, then for $v \neq w$,

$$\begin{aligned}
\mathbb{E}\, I_v I_w &= \frac{\lambda}{n}\left(\pi_{n-1}^{t-1}\right)^2 + \left(1-\frac{\lambda}{n}\right)\left(\pi_{n-1}^t\right)^2 \\
&= \frac{\lambda}{n}\left(p_{n-1}^{t-1} + \pi_{n-1}^t\right)^2 + \left(1-\frac{\lambda}{n}\right)\left(\pi_{n-1}^t\right)^2 \\
&= \frac{\lambda}{n}\left(\left(p_{n-1}^{t-1}\right)^2 + 2p_{n-1}^{t-1}\pi_{n-1}^t\right) + \left(\pi_{n-1}^t\right)^2
\end{aligned}$$

Furthermore, since $N^2 = N + \sum_{v \neq w} I_v I_w$,

$$\mathbb{E}N^2 = \mathbb{E}N + n(n-1)\left(\frac{\lambda}{n}\left(\left(p_{n-1}^{t-1}\right)^2 + 2p_{n-1}^{t-1}\pi_{n-1}^t\right) + \left(\pi_{n-1}^t\right)^2\right)$$
$$= \mathbb{E}N + n(n-1)\left(o(1) + \left(\pi_{n-1}^t\right)^2\right).$$

Also,

$$\mathbb{E}N = n\pi_n^t = n\left(\pi_{n-1}^t + \frac{\lambda}{n}p_{n-1}^{t-1}\right).$$

Thus, letting $t = (1-\varepsilon)t_n = (1-\varepsilon)\frac{\log n}{\log\log n}$, since $n\pi_{n-1}^t \to \infty$ by Lemma II.3, we have that

$$\frac{\mathbb{E}N}{n\pi_{n-1}^t} \to 1$$

and

$$\frac{\mathbb{E}N}{\left(n\pi_{n-1}^t\right)^2} \to 0$$

Thus dividing numerator and denominator of $\frac{(\mathbb{E}N)^2}{\mathbb{E}N^2}$ by $\left(n\pi_{n-1}^t\right)^2$, we get

$$\mathbb{P}\left(N \geq 1\right) \to 1.$$

Since $\mathrm{maxdeg}(G) \leq t$ if and only if $N(t) = 0$,

$$\mathbb{P}\left[\mathrm{maxdeg}(G) \leq (1-\varepsilon)t_n\right] = \mathbb{P}\left(N((1-\varepsilon)t_n) = 0\right) \to 0.$$

□

II.6 Comparison of Walks on Erdös-Rényi Random Graph

We now compare the two random walks on the largest component of $ER(n,p)$. Recall that t_{mix} and t_{rel} are the mixing and relaxation times for the regular random walk and $\tilde{t}_{mix}$ and $\tilde{t}_{rel}$ are the mixing and relaxation times for the Metropolis walk. Proposition II.3 and Theorem II.1 immediately yield:

Corollary II.2. *With high probability for $G \in ER(n,p)$, $pn = \lambda > 1$,*

$$\frac{\tilde{t}_{rel}}{t_{rel}} = O\left(\frac{\log n}{\log\log n}\right).$$

We use the bounds $t_{rel} = O(t_{mix})$ and $\tilde{t}_{mix} = O\left(\tilde{t}_{rel} \log\left(\max_{v\in V} \frac{1}{\tilde{\pi}(v)}\right)\right)$ (see (II.1) and (II.2)), together with the fact that $\max_{v\in V} \frac{1}{\tilde{\pi}(v)} = n$, to get:

Corollary II.3. *With high probability for $G \in ER(n,p)$, $pn = \lambda > 1$,*

$$\frac{\tilde{t}_{mix}}{t_{mix}} = O\left(\frac{\log^2 n}{\log\log n}\right).$$

Using the result $t_{mix} = \Theta(\log^2 n)$ proved in [7], we get:

Corollary II.4. *With high probability for $G \in ER(n,p)$, $pn = \lambda > 1$,*

$$\tilde{t}_{mix} = O\left(\frac{\log^4 n}{\log\log n}\right).$$

II.7 Preliminaries for Comparison of Walks on the Newman-Watts Random Graph

The Newman-Watts random graph is defined with the intent of forcing a small diameter yet a high clustering coefficient (roughly, the clustering coefficient measures how much the neighborhood of of a randomly chosen vertex resembles a complete graph). We use the definition of the model defined on page 10 of Durrett's book ([4]). It takes three parameters: the number of vertices n, a connection coefficient k, and a long distance connection parameter p. We start with n vertices labeled $0, \ldots, n-1$ and attach vertex i to vertices $i \pm 1 \mod n, \ldots, i \pm k \mod n$. Then we add in $Po(\frac{nkp}{2})$ number of additional edges, attaching each end of each edge independently to any of the n vertices (in this way we allow for multiple edges between two vertices and self loops). This creates a fixed ring-like structure with random "shortcuts" and we denote this probability space as $NW(n, k, p)$ (see Fig. 2 for an example). To compare our

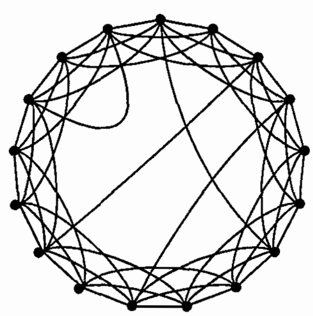

Fig. 2: $G \in NW(n, k, p)$ with $n = 17$, $k = 3$, and 4 shortcuts

walks using Proposition II.3, we need to find an upper bound for Δ for $G \in NW(n, k, p)$ with respect to n.

II.8 Maximum Degree of a Newman-Watts Random Graph

To use the comparison in Proposition II.3, we need to also get a bound on Δ for $G \in NW(n, k, p)$. We first prove a lemma for bounds on a Poisson random variable:

Lemma II.4. *Let $X \sim Po(\mu)$, then for every $c > 0$,*

$$\mathbb{P}(X > c\mu) \leq e^{\mu(c(1-\log c)-1)}.$$

Proof. Let $t > 0$ be some constant (to be chosen later).

$\mathbb{P}(X > c\mu) = \mathbb{P}\left(e^{tX} > e^{tc\mu}\right) \leq e^{-tc\mu}\mathbb{E}e^{tX}$. $\mathbb{E}e^{tX}$ is the moment generating function for the $Po(\mu)$ random variable X and so $\mathbb{E}e^{tX} = e^{\mu(e^t-1)}$. Thus

$$\mathbb{P}(X > c\mu) \leq e^{\mu(e^t-1-tc)}. \tag{II.4}$$

The right hand side of (II.4) is minimized when $t = \log c$. For this value of t we have

$$\mathbb{P}(X > c\mu) \leq e^{\mu(c(1-\log c)-1)}.$$

□

We now give a lower bound for the maximum of a collection of Poisson random variables:

Lemma II.5. *Let $(X_i)_{i=1}^{\infty}$ be a collection of $Po(\mu)$ random variables, not necessarily independent, and let $\varepsilon > 0$, then*

$$\lim_{n\to\infty} \mathbb{P}\left(\max_{i=1}^{n} X_i > (1+\varepsilon)\frac{\log n}{\log\log n}\right) = 0.$$

Proof. Let $f(n) = \frac{1}{\mu}\frac{\log n}{\log\log n}$, then

$$\begin{aligned}\mathbb{P}\left(\max_{i=1}^{n} X_i > (1+\varepsilon)\mu f(n)\right) &= \mathbb{P}\left(\cup_{i=1}^{n}\{X_i > (1+\varepsilon)\mu f(n)\}\right)\\ &\le \sum_{i=1}^{n}\mathbb{P}\left(X_i > (1+\varepsilon)\mu f(n)\right)\\ &= n\mathbb{P}\left(X_0 > (1+\varepsilon)\mu f(n)\right).\end{aligned}$$

By Lemma II.4, we get that this is less than

$$\begin{aligned}ne^{\mu((1+\varepsilon)f(n)(1-\log(1+\varepsilon)f(n))-1)} &= ne^{\mu\left((1+\varepsilon)\frac{1}{\mu}\frac{\log n}{\log\log n}\left(1-\log\left((1+\varepsilon)\frac{1}{\mu}\frac{\log n}{\log\log n}\right)\right)-1\right)}\\ &= e^{\log n}e^{(1+\varepsilon)\frac{\log n}{\log\log n}\left(1-\log\left((1+\varepsilon)\frac{1}{\mu}\frac{\log n}{\log\log n}\right)\right)}e^{-\mu}\\ &= e^{(1+\varepsilon)\frac{\log n}{\log\log n}(1-\log(1+\varepsilon)+\log\mu-\log\log n+\log\log\log n)+\log n}e^{-\mu}\\ &= e^{(1+\varepsilon)\frac{\log n}{\log\log n}(1-\log(1+\varepsilon)+\log\mu+\log\log\log n)-\varepsilon\log n}e^{-\mu}.\end{aligned}$$

Since the largest in the exponent is $-\varepsilon\log n$, this goes to zero as n goes to infinity. □

Proposition II.2. *For $G \in NW(n,k,p)$, we have that $\Delta = O\left(\frac{\log n}{\log\log n}\right)$ with high probability.*

Proof. We first inspect the number of shortcuts attached to an arbitrarily chosen vertex v. Since there are a $Po\left(\frac{nkp}{2}\right)$ number of shortcuts, each of which are attached to v with a probability of $\frac{2}{n}$, the probability that j edges are attached to v given the value of $Po\left(\frac{nkp}{2}\right)$ is

$$\binom{Po\left(\frac{nkp}{2}\right)}{j}\left(\frac{2}{n}\right)^j\left(1-\frac{2}{n}\right)^{Po\left(\frac{nkp}{2}\right)-j}$$

for $j = 0,\dots,Po\left(\frac{nkp}{2}\right)$. Thus, conditioning on the value of $Po\left(\frac{nkp}{2}\right)$, noting that v

has exactly $2k$ non-shortcut attachments, we have that

$$\begin{aligned}\mathbb{P}(\deg v = 2k + j) &= \sum_{i=j}^{\infty} \binom{i}{j} \left(\frac{2}{n}\right)^j \left(1 - \frac{2}{n}\right)^{i-j} \mathbb{P}\left(Po\left(\frac{nkp}{2}\right) = i\right) \\ &= \sum_{i=j}^{\infty} \binom{i}{j} \left(\frac{2}{n}\right)^j \left(1 - \frac{2}{n}\right)^{i-j} \frac{\left(\frac{nkp}{2}\right)^i}{i!} e^{-\frac{nkp}{2}} \\ &= e^{-\frac{nkp}{2}} \frac{\left(\frac{nkp}{2}\right)^j \left(\frac{2}{n}\right)^j}{j!} \sum_{i=j}^{\infty} \left(1 - \frac{2}{n}\right)^{i-j} \frac{\left(\frac{nkp}{2}\right)^{i-j}}{(i-j)!} \\ &= e^{-\frac{nkp}{2}} \frac{(kp)^j}{j!} e^{\left(1-\frac{2}{n}\right)\left(\frac{nkp}{2}\right)} = \frac{(kp)^j}{j!} e^{-kp}\end{aligned}$$

Thus $\deg v \sim 2k + Po(kp)$. Therefore, by Lemma II.5,

$$\lim_{n\to\infty} \mathbb{P}\left(\max_{v=1}^{n} \deg v > (1+\varepsilon)\frac{\log n}{\log\log n}\right) = 0.$$

□

II.9 Comparison of Walks on a Newman-Watts Random Graph

We now compare the two random walks on $NW(n, k, p)$ for $k = 1$. Recall that t_{mix} and t_{rel} are the mixing and relaxation times for the regular random walk and $\tilde{t}_{mix}$ and $\tilde{t}_{rel}$ are the mixing and relaxation times for the Metropolis walk. Proposition II.3 and Theorem II.2 immediately yield:

Corollary II.5. *With high probability for $G \in NW(n, 1, p)$,*

$$\frac{\tilde{t}_{rel}}{t_{rel}} = O\left(\frac{\log n}{\log\log n}\right).$$

Durrett proves in [?] that on $NW(n, 1, p)$, $t_{rel} = O(\log^2 n)$ (Theorem 6.6.1) and this yields the following bound on relaxation time for the Metropolis walk:

Corollary II.6. *With high probability for $G \in NW(n, 1, p)$,*

$$\tilde{t}_{rel} = O\left(\frac{\log^3 n}{\log \log n}\right).$$

We can get bounds on mixing time using $\tilde{t}_{mix} = O\left(\tilde{t}_{rel} \log\left(\max_{v \in V} \frac{1}{\tilde{\pi}(v)}\right)\right)$ (see (II.2)), together with the fact that $\max_{v \in V} \frac{1}{\tilde{\pi}(v)} = n$, to get:

Corollary II.7. *With high probability for $G \in NW(n, 1, p)$,*

$$\tilde{t}_{mix} = O\left(\frac{\log^4 n}{\log \log n}\right).$$

CHAPTER III

THE RECONSTRUCTION PROBLEM

III.1 Introduction

III.1.1 Overview of the Reconstruction Problem

Consider an infinite rooted tree T and a transition matrix M acting on state space A. (Since M determines the size of A, we consider A to be determined by M.) The configuration space is the set of functions $\{\sigma : T \to A\}$, that is, an assignment of an element of the state space to each vertex of the tree. Starting with some distribution μ over A at the root, we generate a distribution over $\{\sigma : T \to A\}$ according to M and the initial distribution μ. The reconstruction problem asks when, asymptotically, information about the distribution at the root can be inferred from a given configuration over some subset of the tree. A sub category of these problems looks at when there is an eventual complete loss of information and when there is always some information preserved and seeks to give an answer in terms of characteristics of both T and M. A *cutset* Π of the tree T is a subset of the vertex set such that every non-intersecting path from the root to infinity passes through exactly one vertex of Π (see Fig. 3). We can reasonably restrict our attention to the case that the information given is a configuration on a cutset. Let X_Π^μ be the distribution on the cutset Π given that the distribution at the root is μ. If we want to know

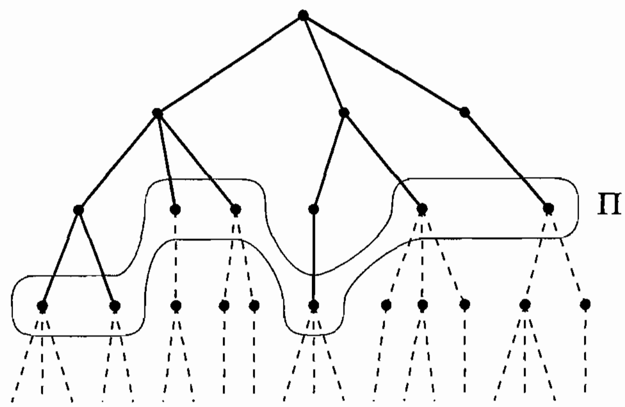

Fig. 3: Infinite tree with cutset Π

how much information about μ is contained in X_{Π}^{μ}, we can answer this question using total variation distance, d_{TV}, for distributions. Specifically, if

$$\sup_{\mu,\nu} \inf_{\Pi} d_{TV}\left(X_{\Pi}^{\mu}, X_{\Pi}^{\nu}\right) = 0, \tag{III.1}$$

where the infimum is over all cutsets of T and the supremum is over all distributions at the root, then there is an eventual loss of all information about the distribution at the root. If this quantity is positive, then some information is always preserved asymptotically. So, given that the information we are provided is a configuration on a cutset, the answer to this problem given a specific T and a specific M is: yes we can infer something about the root asymptotically, or no we cannot. The reconstruction problem would be solved, if we could construct a function F that took as input an infinite rooted tree T and a transition matrix M such that $F(T, M) =$ "yes" if there is reconstruction and $F(T, M) =$"no" if there is no reconstruction. We suspect that F can be broken into two real-valued functions $\mathrm{Br}(T)$ and $\Theta(M)$ such that F outputs "yes" when $\mathrm{Br}(T)\,\Theta(M) > 1$, and "no" when $\mathrm{Br}(T)\,\Theta(M) \leq 1$. The function $\mathrm{Br}(T)$ is the *branching number* of the tree. We do not know what form $\Theta(M)$ would take.The definition of branching number that we use in this paper is:

$$\mathrm{Br}(T) := \sup\left\{\lambda > 0 : \inf_{\Pi} \sum_{w \in \Pi} \lambda^{-|w|} > 0\right\}, \tag{III.2}$$

where the infimum is over all cutsets Π and $|w|$ is the distance from the vertex w to the root.

III.1.2 Previous Work

$\Theta(M)$ could take into account every entry of the matrix or, *a priori*, it could use only information like eigenvalues, trace, or determinant. Prior research by Mossel [15] shows that eigenvalue information is not enough in general. However the only (partial) answer to the reconstruction problem for general matrices bigger than 2×2 is in terms of the second largest eigenvalue, $\lambda_2(M)$, of M. The result of [11] is valid for regular trees and shows that reconstruction is possible for the b-ary tree if $b\,\lambda_2^2(M) > 1$. This result is not tight in general and this is shown in [15]. (But for nearly symmetric 2×2 matrices, this bound is tight [3].) For regular trees, some other results are known for various classes of matrices, but nothing else is known for general matrices larger than 2×2. For the 2×2 case, and for a general tree, [6] has solved the case of a symmetric M, finding the exact threshold for reconstruction. For general 2×2 matrices, the methods of [6] that were used to determine when reconstruction is possible were extended in [3], but these bounds are not tight for all matrices. [13] found bounds for non-reconstruction for general matrices, but only for regular trees. This paper gives a bound for general trees and general matrices that matches bounds of [13] in the case of regular trees, and is found by different methods. [3] shows that this result is not tight.

III.1.3 Main Result

In this paper we assume that T is any rooted tree with root ρ, with paths to infinity, and with the degree of each vertex finite. We analyze the process that starts with a given distribution of $\{0,1\}$ at ρ and proceeds generationally by choosing the value of the child w given the value of the parent v according to

$$\begin{bmatrix} \mathbb{P}(X(w)=0|X(v)=0) & \mathbb{P}(X(w)=1|X(v)=0) \\ \mathbb{P}(X(w)=0|X(v)=1) & \mathbb{P}(X(w)=1|X(v)=1) \end{bmatrix} = \begin{bmatrix} 1-a & a \\ b & 1-b \end{bmatrix}.$$

This generates a tree-indexed Markov chain, where the restriction to any path from the root is a Markov chain with matrix

$$\begin{bmatrix} 1-a & a \\ b & 1-b \end{bmatrix}$$

We want to know under what circumstances the values on the vertices at large depth carry significant information about the value at the root. We answer this question in terms of total variation. Let W be any vertex set, then X_W^i is the (random) sequence of values on W given that the root takes value i. We want to know when

$$\inf_{\Pi} d_{TV}(X_\Pi^0, X_\Pi^1) = 0$$

where the infimum is over all cutsets Π. (The supremum of (III.1) is obtained by these starting distributions.) The left-hand side is zero exactly when the best guess for the value at the root is asymptotically no better than the uniform guess. That is, the left hand side is zero exactly when we can find some cutset Π for which any configuration on that cutset leads to a best guess that is a distribution that is very close to choosing 0 and 1 with equal probability. Let $\theta = \left(\sqrt{(1-a)(1-b)} - \sqrt{ab}\right)^2$. The result we prove is:

Theorem 1. If $\mathrm{Br}(T)\,\theta < 1$, then $\inf_\Pi d_{TV}(X^0_\Pi, X^1_\Pi) = 0$.

This is naturally proved by finding an upperbound for $d_{TV}(X^0_\Pi, X^1_\Pi)$ for each cutset Π in terms of something related to the branching number. The theorem will follow from a couple lemmas. We begin with some definitions.

III.2 Proof of Theorem 1

III.2.1 Definitions and Basic Proof Idea

For a given cutset Π, we let $T(\Pi)$ be the set of vertices and edges between the root and Π (inclusive). (This makes X^0_Π and X^1_Π the corresponding configurations on the leaves.) Our main approach will be to switch to Hellinger distance. For two distributions μ and ν on some set S, the Hellinger distance is defined as:

$$d_H(\mu,\nu) := \sum_{x\in S}\left(\sqrt{\mu(x)} - \sqrt{\nu(x)}\right)^2.$$

This can be related to total variation distance, d_{TV} by using Jensen's inequality with respect to the probability measure $\frac{\mu+\nu}{2}$ as follows:

$$d_{TV}(\mu,\nu) := \frac{1}{2}\sum_{x\in S}|\mu(x)-\nu(x)| = \sum_{x\in S}\frac{|\mu(x)-\nu(x)|}{\mu(x)+\nu(x)}\left(\frac{\mu(x)+\nu(x)}{2}\right)$$

$$\le \sqrt{\sum_{x\in S}\left(\frac{|\mu(x)-\nu(x)|}{\mu(x)+\nu(x)}\right)^2\left(\frac{\mu(x)+\nu(x)}{2}\right)} = \sqrt{\frac{1}{2}\sum_{x\in S}\frac{(\mu(x)-\nu(x))^2}{\mu(x)+\nu(x)}}.$$

Using the relation $\frac{\left(\sqrt{\mu(x)}+\sqrt{\nu(x)}\right)^2}{\mu(x)+\nu(x)} \leq 2$, which can be shown using the concavity of $\sqrt{\cdot}$, we get:

$$d_{TV}(\mu,\nu) \leq \sqrt{d_H(\mu,\nu)}.$$

Therefore we have that

$$d_{TV}(X_\Pi^0, X_\Pi^1)^2 \leq \sum_\sigma \left(\sqrt{\mathbb{P}\left(X_\Pi^0 = \sigma\right)} - \sqrt{\mathbb{P}\left(X_\Pi^1 = \sigma\right)}\right)^2,$$

where the sum is over all configurations on the leaves of $T(\Pi)$. We use this by successively dividing the tree into its independent branches, starting first with those branches issuing from the root and moving to an inequality that regards these descendants as roots of their own trees, and then continuing the same process until the leaves are reached. To get started, we need some additional notation. Given a vertex v, let Ω_v be the set of configurations on the leaves of $T(\Pi)$ below v and let $C(v)$ be the set of those vertices that are immediate descendants of v (the children of v). For two vertices v and w, we write $v \preccurlyeq w$ if the vertex v is on the unique non-intersecting path from the root to w. For a vertex v and a subset of vertices W, we write $v \preccurlyeq W$ if for every vertex $w \in W$, we have $v \preccurlyeq w$. Given a configuration $\sigma \in \Omega_v$ and a vertex w with $v \preccurlyeq w$, let σ_w be the restriction of σ to Ω_w. We suppress the notation Π in further notation, for we consider everything that follows to be for a fixed cutset until further notice. Given a vertex v, let X_v^i be the (random) configuration on Π_v given that the vertex v is labeled as i (for $i = 0, 1$). For any vertex v, let

$$\phi(v) = \sum_{\sigma\in\Omega_v} \left(\sqrt{\mathbb{P}\left(X_v^0 = \sigma\right)} - \sqrt{\mathbb{P}\left(X_v^1 = \sigma\right)}\right)^2$$

We would like to show that

$$\phi(v) \leq \theta \sum_{w \in C(v)} \phi(w),$$

where $\theta = \left(\sqrt{(1-a)(1-b)} - \sqrt{a\,b}\right)^2$. Then, since for $w \in \Pi$, $\phi(w) = 1$, we will have by induction (after proving Lemmas 1 and 2) that $\phi(v) \leq \sum_{w \in \Pi_v} \theta^{|w|-|v|}$, and in particular,

$$d_{TV}(X^0_\Pi, X^1_\Pi)^2 \leq \phi(\rho) \leq \sum_{w \in \Pi} \theta^{|w|}.$$

This is what will allow us to pass to the branching number. Let $X^i_v(w)$ be the configuration X^i_v restricted to those vertices below w. We have then, by the independent evolution on branches, that

$$\mathbb{P}(X^i_v = \sigma_v) = \prod_{w \in C(v)} \mathbb{P}(X^i_v(w) = \sigma_w)$$

With this in mind, we have that

$$\phi(v) = \sum_{\sigma \in \Omega_v} \left(\sqrt{\prod_{w \in C(v)} \mathbb{P}(X^0_v(w) = \sigma_w)} - \sqrt{\prod_{w \in C(v)} \mathbb{P}(X^1_v(w) = \sigma_w)} \right)^2$$

we would like to show that this is bounded above by

$$\sum_{w \in C(v)} \sum_{\sigma \in \Omega_w} \left(\sqrt{\mathbb{P}\left(X^0_v(w) = \sigma\right)} - \sqrt{\mathbb{P}\left(X^1_v(w) = \sigma\right)} \right)^2$$

Since each of the terms $\mathbb{P}\left(X^i_v(w) = \sigma\right)$ sums to 1 over all $\sigma \in \Omega_w$ for each w, this will follow from Lemma III.1. Our inductive step is split into two parts. Lemma III.1 will allow us to split a tree issuing from any vertex into separate trees indexed by the children of that vertex. Lemma III.2, which is just an inequality proved by calculus, will allow us to "move" forward to the next generation, where we can then repeat the process.

Lemma III.1. *Let n be a positive integer. For each $i = 1, \ldots, n$, let f_i and g_i be non-negative, real-valued functions from a finite set A_i with $\sum_{x_i \in A_i} f_i(x_i) = 1 = \sum_{x_i \in A_i} g_i(x_i)$, then*

$$\sum_{(x_1,\ldots,x_n)\in A_1\times\cdot\times A_n} \left(\sqrt{\prod_{i=1}^{n} f_i(x_i)} - \sqrt{\prod_{i=1}^{n} g_i(x_i)} \right)^2 \leq \sum_{i=1}^{n} \sum_{x_i \in A_i} \left(\sqrt{f_i(x_i)} - \sqrt{g_i(x_i)} \right)^2$$

Lemma III.2. *For $x, y \geq 0$ and $0 \leq a, b \leq 1$,*

$$\left(\sqrt{(1-a)x + ay} - \sqrt{bx + (1-b)y} \right)^2 \leq \theta \left(\sqrt{x} - \sqrt{y} \right)^2 \tag{III.3}$$

The proofs of these lemmas is deferred to Section 2. We now apply Lemmas III.1 and III.2 to our situation with probabilities of configurations on a tree. These lemmas will be combined into Proposition III.1, which will be the inductive step of our main theorem.

Proposition III.1. *For any vertex v, with $v \notin \Pi$, we have $\phi(v) \leq \theta \sum_{w \in C(v)} \phi(w)$.*

Proof.

$$\begin{aligned}
\phi(v) &= \sum_{\sigma \in \Omega_v} \left(\sqrt{\mathbb{P}\left(X_v^0 = \sigma\right)} - \sqrt{\mathbb{P}\left(X_v^1 = \sigma\right)} \right)^2 \\
&= \sum_{\sigma \in \Omega_v} \left(\sqrt{\prod_{w \in C(v)} \mathbb{P}(X_v^0(w) = \sigma_w)} - \sqrt{\prod_{w \in C(v)} \mathbb{P}(X_v^1(w) = \sigma_w)} \right)^2
\end{aligned}$$

By Lemma III.1 we get that this is

$$\leq \sum_{w\in C(v)} \sum_{\sigma\in\Omega_w} \left(\sqrt{\mathbb{P}\left(X_v^0(w)=\sigma\right)} - \sqrt{\mathbb{P}\left(X_v^1(w)=\sigma\right)}\right)^2 \tag{III.4}$$

Since for $\sigma \in \Omega_w$,

$$\mathbb{P}\left(X_v^0(w)=\sigma\right) = (1-a)\,\mathbb{P}\left(X_w^0=\sigma\right) + a\,\mathbb{P}\left(X_w^1=\sigma\right)$$

and

$$\mathbb{P}\left(X_v^1(w)=\sigma\right) = b\,\mathbb{P}\left(X_w^0=\sigma\right) + (1-b)\,\mathbb{P}\left(X_w^1=\sigma\right),$$

we have that (III.4) is equal to

$$\sum_{w\in C(v)} \sum_{\sigma\in\Omega_w} \Bigg(\sqrt{(1-a)\,\mathbb{P}\left(X_w^0=\sigma\right) + a\,\mathbb{P}\left(X_w^1=\sigma\right)}$$
$$-\sqrt{b\,\mathbb{P}\left(X_w^0=\sigma\right) + (1-b)\,\mathbb{P}\left(X_w^1=\sigma\right)}\Bigg)^2.$$

By Lemma III.2 this is bounded above by

$$\theta \sum_{w\in C(v)} \sum_{\sigma\in\Omega_w} \left(\sqrt{\mathbb{P}\left(X_w^0=\sigma\right)} - \sqrt{\mathbb{P}\left(X_w^1=\sigma\right)}\right)^2 = \theta \sum_{w\in C(v)} \phi(w).$$

□

Next we apply induction to this result to get a quantity related to the branching number.

Proposition III.2. *For every* $v \in T(\Pi)$, *we have* $\phi(v) \leq \sum_{w\in\Pi_v} \theta^{|w|-|v|}$

Proof. Given $v \in T(\Pi)$, assume that for every vertex $u \in C(v)$ we have that

$\phi(u) \leq \sum_{w\in\Pi_u} \theta^{|w|-|u|}$. From Theorem 1, we have that

$$\phi(v) \leq \theta \sum_{u\in C(v)} \phi(u) \leq \theta \sum_{u\in C(v)} \sum_{w\in\Pi_u} \theta^{|w|-|u|} = \sum_{u\in C(v)} \sum_{w\in\Pi_u} \theta^{|w|-|v|} = \sum_{w\in\Pi_v} \theta^{|w|-|v|}$$

Since it is true for every $v \in \Pi$, and since the inductive step holds, it is true for all $v \in T(\Pi)$. □

We are now ready to prove the main theorem, which we restate for convenience:

Theorem 1. *If* $\mathrm{Br}(T)\,\theta < 1$, *then* $\inf_\Pi d_{TV}(X^0_\Pi, X^1_\Pi) = 0$.

Proof. Applying Proposition III.2 to the root ρ, we have that

$$d_{TV}(X^0_\Pi, X^1_\Pi)^2 \leq \phi(\rho) \leq \sum_{v\in\Pi} \theta^{|v|}$$

and so, by (III.2), if $\mathrm{Br}(T) < \theta^{-1}$, then

$$\inf_\Pi d_{TV}(X^0_\Pi, X^1_\Pi) = 0$$

□

III.3 Proofs of Lemmas III.1 and III.2

We restate the lemmas for convenience.

Lemma 2.1. *Let n be a positive integer. For each $i = 1, \ldots, n$, let f_i and g_i be non-negative, real-valued functions from a finite set A_i with*

$\sum_{x_i \in A_i} f_i(x_i) = 1 = \sum_{x_i \in A_i} g_i(x_i)$, *then*

$$\sum_{(x_1,\ldots,x_n)\in A_1\times\cdot\times A_n} \left(\sqrt{\prod_{i=1}^{n} f_i(x_i)} - \sqrt{\prod_{i=1}^{n} g_i(x_i)} \right)^2 \leq \sum_{i=1}^{n} \sum_{x_i\in A_i} \left(\sqrt{f_i(x_i)} - \sqrt{g_i(x_i)} \right)^2 \tag{III.5}$$

Proof. Since

$$\sum_{(x_1,\ldots,x_n)\in A_1\times\cdot\times A_n} \prod_{i=1}^{n} f_i(x_i) = \prod_{i=1}^{n} \sum_{x_i\in A_i} f_i(x_i) = 1,$$

and

$$\sum_{(x_1,\ldots,x_n)\in A_1\times\cdot\times A_n} \prod_{i=1}^{n} g_i(x_i) = \prod_{i=1}^{n} \sum_{x_i\in A_i} g_i(x_i) = 1,$$

when the summand of the left hand side of (III.5) is multiplied out, the left hand side becomes

$$2 - 2\left(\sum_{(x_1,\ldots,x_n)\in A_1\times\cdot\times A_n} \prod_{i=1}^{n} \sqrt{f_i(x_i)\cdot g_i(x_i)} \right)$$

by interchanging the sum and product we get that this is the same as

$$2 - 2\prod_{i=1}^{n} \sum_{x_i\in A_i} \sqrt{f_i(x_i)\cdot g_i(x_i)}$$

By multiplying out the summand of the right hand side of (III.5) and summing $f_i(x_i)$ and $g_i(x_i)$ over $x_i \in A_i$ the right hand side becomes

$$\sum_{i=1}^{n} \left(2 - 2\sum_{x_i\in A_i} \sqrt{f_i(x_i)\cdot g_i(x_i)} \right)$$

which is the same as

$$2\left(n - \sum_{i=1}^{n} \sum_{x_i \in A_i} \sqrt{f_i(x_i) \cdot g_i(x_i)}\right)$$

So we can verify (III.5) if we show that

$$\sum_{i=1}^{n} \sum_{x_i \in A_i} \sqrt{f_i(x_i) \cdot g_i(x_i)} - \prod_{i=1}^{n} \sum_{x_i \in A_i} \sqrt{f_i(x_i) \cdot g_i(x_i)} \leq n - 1 \tag{III.6}$$

Letting $y_i = \sum_{x_i \in A_i} \sqrt{f_i(x_i) \cdot g_i(x_i)}$, (III.6) becomes

$$\sum_{i=1}^{n} y_i - \prod_{i=1}^{n} y_i \leq n - 1 \tag{III.7}$$

But since $0 \leq y_i \leq \sum_{x_i \in A_i} \frac{f_i(x_i)+g_i(x_i)}{2} = 1$ for each i, (III.7) is true. □

Lemma 2.2. *For $x, y \geq 0$ and $0 \leq a, b \leq 1$,*

$$\left(\sqrt{(1-a)x + ay} - \sqrt{bx + (1-b)y}\right)^2 \leq \theta\left(\sqrt{x} - \sqrt{y}\right)^2 \tag{III.8}$$

Proof. Let $g(x) = \sqrt{(1-a)x + ay} - \sqrt{bx + (1-b)y}$ and let $h(x) = \sqrt{x} - \sqrt{y}$. We seek to optimize $f(x) = {}^{g(x)}/_{h(x)}$. $f(x)$ has a critical point exactly when $h(x)g'(x) = h'(x)g(x)$. Now,

$$h(x)g'(x) = \left(\sqrt{x} - \sqrt{y}\right)\left(\frac{1-a}{2\sqrt{(1-a)x + ay}} - \frac{b}{2\sqrt{bx + (1-b)y}}\right)$$

and,

$$h'(x)g(x) = \frac{1}{2\sqrt{x}}\left(\sqrt{(1-a)x+ay} - \sqrt{bx+(1-b)y}\right).$$

Multiplying both sides of the equation $h(x)g'(x) = h'(x)g(x)$ by

$$2\sqrt{x}\sqrt{(1-a)x+ay}\sqrt{bx+(1-b)y},$$

we get the equation

$$\sqrt{x}\left(\sqrt{x}-\sqrt{y}\right)\left((1-a)\sqrt{bx+(1-b)y} - b\sqrt{(1-a)x+ay}\right)$$
$$= ((1-a)x+ay)\sqrt{bx+(1-b)y} - (bx+(1-b)y)\sqrt{(1-a)x+ay}.$$

In order to solve this equation for x, we rearrange it and cancel some terms to get:

$$((1-b)y + b\sqrt{x}\sqrt{y})\sqrt{(1-a)x+ay} = (ay + (1-a)\sqrt{x}\sqrt{y})\sqrt{bx+(1-b)y}.$$

Canceling a factor of $\sqrt{y}$ on both sides yields

$$((1-b)\sqrt{y} + b\sqrt{x})\sqrt{(1-a)x+ay} = (a\sqrt{y} + (1-a)\sqrt{x})\sqrt{bx+(1-b)y}.$$

Squaring both sides yields

$$((1-b)^2y + 2b(1-b)\sqrt{x}\sqrt{y} + b^2x)((1-a)x+ay)$$
$$= (a^2y + 2a(1-a)\sqrt{x}\sqrt{y} + (1-a)^2x)(bx+(1-b)y).$$

Rearranging this we get

$$((1-b)^2y+b^2x)((1-a)x+ay)-(a^2y+(1-a)^2x)(bx+(1-b)y)$$
$$=(2a(1-a)(bx+(1-b)y)-2b(1-b)((1-a)x+ay))\sqrt{x}\sqrt{y}.$$

Rearranging again we get

$$b(1-a)(b-(1-a))x^2+(ab^2+(1-a)(1-b)^2-ba^2-(1-b)(1-a)^2)xy$$
$$+a(1-b)((1-b)-a)y^2$$
$$=(2b(1-a)(a-(1-b))x+2a(1-b)(1-a-b)y)\sqrt{x}\sqrt{y}.$$

Factoring, we get

$$-b(1-a)(1-a-b)x^2+(ab(b-a)+(1-a)(1-b)(a-b))xy+a(1-b)((1-a-b)y^2$$
$$=2(1-a-b)(a(1-b)y-b(1-a)x)\sqrt{x}\sqrt{y}.$$

Factoring again we get

$$(1-a-b)(a(1-b)y^2+(a-b)xy-b(1-a)x^2)$$
$$=2(1-a-b)(a(1-b)y-b(1-a)x)\sqrt{x}\sqrt{y}.$$

Canceling the $1-a-b$ term (noting that (III.8) holds when $1-a-b=0$) and moving everything to one side we get

$$a(1-b)y^2+(a-b)xy-b(1-a)x^2-2a(1-b)y\sqrt{y}\sqrt{x}+2b(1-a)x\sqrt{x}\sqrt{y}=0.$$

The left hand side factors as

$$(\sqrt{x}-\sqrt{y})^2(b(1-a)x-a(1-b)y).$$

Since (III.8) holds if $\sqrt{x}=\sqrt{y}$, the critical point we are interested in is $x=\frac{a(1-b)}{b(1-a)}y$. At this critical point we have

$$\begin{aligned} g\left(\frac{a(1-b)}{b(1-a)}y\right) &= \sqrt{\frac{a(1-b)}{b}y+ay}-\sqrt{\frac{a(1-b)}{1-a}y+(1-b)y} \\ &= \left(\sqrt{\frac{a}{b}}-\sqrt{\frac{1-b}{1-a}}\right)\sqrt{y} \\ &= \left(\sqrt{a(1-a)}-\sqrt{b(1-b)}\right)\frac{\sqrt{y}}{\sqrt{b(1-a)}}, \end{aligned}$$

and

$$\begin{aligned} h\left(\frac{a(1-b)}{b(1-a)}y\right) &= \sqrt{\frac{a(1-b)}{b(1-a)}y}-\sqrt{y} \\ &= \left(\sqrt{a(1-b)}-\sqrt{b(1-a)}\right)\frac{\sqrt{y}}{\sqrt{b(1-a)}}. \end{aligned}$$

So,

$$f\left(\frac{a(1-b)}{b(1-a)}y\right)=\frac{\sqrt{a(1-a)}-\sqrt{b(1-b)}}{\sqrt{a(1-b)}-\sqrt{b(1-a)}}=\sqrt{(1-a)(1-b)}-\sqrt{ab}$$

Thus the maximum of $(f(x))^2$ is θ, and the lemma is proved. □

III.4 An Inequality

We prove an inequality that is similar in flavor to Lemma III.1. The proof takes more work than that of Lemma III.1, and is seemingly stronger, yet when it is used for the reconstruction bound, one is prompted to make subsequent rough bounds that yield an end result that is inferior. We briefly describe the motivation for the inequality in terms of the reconstruction problem, prove the inequality, and then show why it fails to be useful. Recall that X_{Π}^{0} and X_{Π}^{1} are (random) configurations on the cutset Π given a 0 and a 1 at the root respectively. Rather than bound the total variation distance by the Hellinger distance, we bound it by the L^2 distance:

$$\begin{aligned} d_{TV}(X_{\Pi}^{0}, X_{\Pi}^{1}) &= \frac{1}{2}\sum_{\sigma} \left|\mathbb{P}\left(X_{\Pi}^{0}=\sigma\right) - \mathbb{P}\left(X_{\Pi}^{1}=\sigma\right)\right| \\ &\le \frac{\sqrt{2}}{2}\sqrt{\sum_{\sigma} \frac{\left(\mathbb{P}\left(X_{\Pi}^{0}=\sigma\right) - \mathbb{P}\left(X_{\Pi}^{1}=\sigma\right)\right)^2}{\mathbb{P}\left(X_{\Pi}^{0}=\sigma\right) + \mathbb{P}\left(X_{\Pi}^{1}=\sigma\right)}} \end{aligned}$$

If we then mimic the proof of Theorem 1, we would need to employ the use of the following inequality:

$$\sum_{\alpha_1,\ldots,\alpha_n} \frac{\left(\prod_{i=1}^{k} x_i^{\alpha_i} - \prod_{i=1}^{k} y_i^{\alpha_i}\right)^2}{\prod_{i=1}^{k} x_i^{\alpha_i} + \prod_{i=1}^{k} y_i^{\alpha_i}} \le \sum_{i=1}^{k}\sum_{\alpha_i} \frac{(x_i^{\alpha_i} - y_i^{\alpha_i})^2}{x_i^{\alpha_i} + y_i^{\alpha_i}} \tag{III.9}$$

for some k, where for each i, $\sum_{\alpha_i} x_i^{\alpha_i} = 1 = \sum_{\alpha_i} y_i^{\alpha_i}$, the sum being over some finite set A_i with $\alpha_i \in A_i$. This is Corollary III.3 below. We first prove a related inequality which will lead to III.9 for the case $k=2$:

Lemma III.3. *Let $\psi(x,y) = \frac{xy}{x+y}$ for $x+y \neq 0$ and 0 otherwise. Let m and n be natural numbers. Let $w_i, x_i, y_j, z_j \geq 0$ for each $i = 1,\ldots,m$ and $j = 1,\ldots,n$. If $\sum_{i=1}^{m} w_i \leq 1$,*

$\sum_{i=1}^{m} x_i \leq 1$, $\sum_{j=1}^{n} y_j \leq 1$, *and* $\sum_{j=1}^{n} z_j \leq 1$, *then*

$$\sum_{i=1}^{m} \psi(w_i, x_i) + \sum_{j=1}^{n} \psi(y_j, z_j) - \sum_{i=1}^{m}\sum_{j=1}^{n} \psi(w_i y_j, x_i z_j) \leq \frac{1}{2}\left(\min_{i \mid w_i x_i \neq 0} \frac{w_i^2 + x_i^2}{(w_i + x_i)^2} + \min_{j \mid y_j z_j \neq 0} \frac{y_j^2 + z_j^2}{(y_j + z_j)^2}\right)$$

We note for convenience that in the case that $w_i, x_i, y_j, z_j > 0$ for each *i*,*j*, this is the same as:

$$\sum_{i=1}^{m} \frac{w_i x_i}{w_i + x_i} + \sum_{j=1}^{n} \frac{y_j z_j}{y_j + z_j} - \sum_{i=1}^{m}\sum_{j=1}^{n} \frac{w_i y_j x_i z_j}{w_i y_j + x_i z_j} \leq \frac{1}{2}\left(\min_{i} \frac{w_i^2 + x_i^2}{(w_i + x_i)^2} + \min_{j} \frac{y_j^2 + z_j^2}{(y_j + z_j)^2}\right)$$

Proof. We use the method of Lagrange multipliers. Let $\mathbf{w} = (w_i)_{i=1}^{m}$, $\mathbf{x} = (x_i)_{i=1}^{m}$, $\mathbf{y} = (y_j)_{j=1}^{n}$, $\mathbf{z} = (z_j)_{j=1}^{n}$ (where we now consider these to be variables and not fixed), and let

$$f(\mathbf{w}, \mathbf{x}, \mathbf{y}, \mathbf{z}) := \sum_{i=1}^{m} \psi(w_i, x_i) + \sum_{j=1}^{n} \psi(y_j, z_j) - \sum_{i=1}^{m}\sum_{j=1}^{n} \psi(w_i y_j, x_i z_j).$$

(We consider f to be likewise defined for all possible dimensions of $\mathbf{w}$, $\mathbf{x}$, $\mathbf{y}$, and $\mathbf{z}$ – and only for those $\mathbf{w}$, $\mathbf{x}$, $\mathbf{y}$, and $\mathbf{z}$ that satisfy the conditions of the lemma.) Let $g_w(\mathbf{w}) = w_1 + \cdots + w_m$, and let g_x, g_y, and g_z be similarly defined. We want to maximize $f(\mathbf{w}, \mathbf{x}, \mathbf{y}, \mathbf{z})$ over the constraints $g_w(\mathbf{w}) = W$, $g_x(\mathbf{x}) = X$, $g_y(\mathbf{y}) = Y$, and $g_z(\mathbf{z}) = Z$, where $0 < W, X, Y,$ and $Z \leq 1$ are all constants. Let S be the set of $(\mathbf{w}, \mathbf{x}, \mathbf{y}, \mathbf{z})$ that satisfy the constraints with $0 < w_i, x_i, y_j, z_j$ for all i and j. If f achieves its maximum at some point $(\overline{\mathbf{w}}, \overline{\mathbf{x}}, \overline{\mathbf{y}}, \overline{\mathbf{z}}) \in S$, then there exists λ_w, λ_x, λ_y,

λ_z such that $(\overline{\mathbf{w}}, \overline{\mathbf{x}}, \overline{\mathbf{y}}, \overline{\mathbf{z}})$ is a solution of the equation

$$\begin{aligned}\nabla f(\mathbf{w},\mathbf{x},\mathbf{y},\mathbf{z}) =& \lambda_w \nabla g_w(\mathbf{w},\mathbf{x},\mathbf{y},\mathbf{z}) + \lambda_x \nabla g_x(\mathbf{w},\mathbf{x},\mathbf{y},\mathbf{z}) \\ &+ \lambda_y \nabla g_y(\mathbf{w},\mathbf{x},\mathbf{y},\mathbf{z}) + \lambda_z \nabla g_z(\mathbf{w},\mathbf{x},\mathbf{y},\mathbf{z}).\end{aligned}$$

Since ∇g_w, ∇g_x, ∇g_y, and ∇g_z are each vectors of 0's and 1's (with a 1 in the coordinates corresponding to the appropriate symbol: w, x, y, or z), this breaks into the system of equations:

$$f_{w_i} = \frac{x_i^2}{(w_i + x_i)^2} - \sum_{j=1}^{n} \frac{y_j x_i^2 z_j^2}{(w_i y_j + x_i z_j)^2} = \lambda_w$$

$$f_{x_i} = \frac{w_i^2}{(w_i + x_i)^2} - \sum_{j=1}^{n} \frac{z_j w_i^2 y_j^2}{(w_i y_j + x_i z_j)^2} = \lambda_x$$

$$f_{y_j} = \frac{z_j^2}{(y_j + z_j)^2} - \sum_{i=1}^{m} \frac{w_i x_i^2 z_j^2}{(w_i y_j + x_i z_j)^2} = \lambda_y$$

$$f_{z_j} = \frac{y_j^2}{(y_j + z_j)^2} - \sum_{i=1}^{m} \frac{x_i w_i^2 y_j^2}{(w_i y_j + x_i z_j)^2} = \lambda_z$$

and thus $(\overline{\mathbf{w}}, \overline{\mathbf{x}}, \overline{\mathbf{y}}, \overline{\mathbf{z}})$ solves the system:

$$w_i f_{w_i} + x_i f_{x_i} = \frac{w_i x_i}{w_i + x_i} - \sum_{j=1}^{n} \frac{w_i y_j x_i z_j}{w_i y_j + x_i z_j} = w_i \lambda_w + x_i \lambda_x$$

$$y_j f_{y_j} + z_j f_{z_j} = \frac{y_j z_j}{y_j + z_j} - \sum_{i=1}^{m} \frac{w_i y_j x_i z_j}{w_i y_j + x_i z_j} = y_j \lambda_y + z_j \lambda_z$$

Summing the first equation over i and the second equation over j, we have that

$(\overline{\mathbf{w}}, \overline{\mathbf{x}}, \overline{\mathbf{y}}, \overline{\mathbf{z}})$ is a solution of:

$$\sum_{i=1}^{m} \frac{w_i x_i}{w_i + x_i} - \sum_{i=1}^{m}\sum_{j=1}^{n} \frac{w_i y_j x_i z_j}{w_i y_j + x_i z_j} = W\lambda_w + X\lambda_x$$

$$\sum_{j=1}^{n} \frac{y_j z_j}{y_j + z_j} - \sum_{i=1}^{m}\sum_{j=1}^{n} \frac{w_i y_j x_i z_j}{w_i y_j + x_i z_j} = Y\lambda_y + Z\lambda_z$$

Combining these two, we have that $(\overline{\mathbf{w}}, \overline{\mathbf{x}}, \overline{\mathbf{y}}, \overline{\mathbf{z}})$ is a solution of:

$$\sum_{i=1}^{m}\sum_{j=1}^{n} \frac{w_i y_j x_i z_j}{w_i y_j + x_i z_j} = \frac{1}{2}\left(\sum_{i=1}^{m} \frac{w_i x_i}{w_i + x_i} + \sum_{j=1}^{n} \frac{y_j z_j}{y_j + z_j} - W\lambda_w - X\lambda_x - Y\lambda_y - Z\lambda_z\right)$$

and thus,

$$\begin{aligned} f(\overline{\mathbf{w}}, \overline{\mathbf{x}}, \overline{\mathbf{y}}, \overline{\mathbf{z}}) &= \frac{1}{2}\left(\sum_{i=1}^{m} \frac{\overline{w}_i \overline{x}_i}{\overline{w}_i + \overline{x}_i} + \sum_{j=1}^{n} \frac{\overline{y}_j \overline{z}_j}{\overline{y}_j + \overline{z}_j} + W\lambda_w + X\lambda_x + Y\lambda_y + Z\lambda_z\right) \\ &\leq \frac{1}{2}\left(\sum_{i=1}^{m} \frac{\overline{w}_i \overline{x}_i}{\overline{w}_i + \overline{x}_i} + \sum_{j=1}^{n} \frac{\overline{y}_j \overline{z}_j}{\overline{y}_j + \overline{z}_j} + \lambda_w + \lambda_x + \lambda_y + \lambda_z\right). \end{aligned}$$

We now seek bounds for $\lambda_w + \lambda_x$ and $\lambda_y + \lambda_z$.

$$\lambda_w + \lambda_x = \frac{\overline{w}_i^2 + \overline{x}_i^2}{(\overline{w}_i + \overline{x}_i)^2} - \sum_{j=1}^{n} \frac{\overline{y}_j \overline{z}_j \left(\overline{w}_i^2 \overline{y}_j + \overline{x}_i^2 \overline{z}_j\right)}{\left(\overline{w}_i \overline{y}_j + \overline{x}_i \overline{z}_j\right)^2}$$

$$\lambda_y + \lambda_z = \frac{\overline{y}_j^2 + \overline{z}_j^2}{\left(\overline{y}_j + \overline{z}_j\right)^2} - \sum_{i=1}^{m} \frac{\overline{w}_i \overline{x}_i \left(\overline{y}_j^2 \overline{w}_i + \overline{z}_j^2 \overline{x}_i\right)}{\left(\overline{w}_i \overline{y}_j + \overline{x}_i \overline{z}_j\right)^2}$$

for all i, j. The function $g(s,t) = \frac{as^2+bt^2}{(as+bt)^2}$ can be shown to be bounded below by

$\frac{1}{a+b}$ and so we have:

$$\lambda_w + \lambda_x \leq \frac{\overline{w}_i^2 + \overline{x}_i^2}{(\overline{w}_i + \overline{x}_i)^2} - \sum_{j=1}^{n} \frac{\overline{y}_j \overline{z}_j}{\overline{y}_j + \overline{z}_j}$$

$$\lambda_y + \lambda_z \leq \frac{\overline{y}_j^2 + \overline{z}_j^2}{(\overline{y}_j + \overline{z}_j)^2} - \sum_{i=1}^{m} \frac{\overline{w}_i \overline{x}_i}{\overline{w}_i + \overline{x}_i}$$

which yields:

$$f(\overline{\mathbf{w}}, \overline{\mathbf{x}}, \overline{\mathbf{y}}, \overline{\mathbf{z}}) \leq \frac{1}{2}\left(\frac{\overline{w}_i^2 + \overline{x}_i^2}{(\overline{w}_i + \overline{x}_i)^2} + \frac{\overline{y}_j^2 + \overline{z}_j^2}{(\overline{y}_j + \overline{z}_j)^2} \right)$$

for all i, j. So the lemma is proved in the case that the maximum is achieved in the set S for some W, X, Y, and Z. If the maximum is not achieved in the set S for any W, X, Y, and Z, then some of the w_i, x_i, y_j, z_j are zero. But when these are zero, then those indices do not show up in the sums for the function f. By simply ignoring these indices we are back in the previous case, but for smaller m and/or n. □

Theorem III.1. *With the above definitions, we have that* $f(\mathbf{w}, \mathbf{x}, \mathbf{y}, \mathbf{z}) \leq \frac{1}{2}$ *whenever* $(\mathbf{w}, \mathbf{x}, \mathbf{y}, \mathbf{z})$ *is in the domain of* f*.*

Proof. When the i-th coordinate of w and x agree, then we have that $\frac{w_i^2 + x_i^2}{(w_i + x_i)^2} = \frac{1}{2}$. Similarly for $\mathbf{y}$ and $\mathbf{z}$. So if $\mathbf{w}$ and $\mathbf{x}$ agree somewhere and $\mathbf{y}$ and $\mathbf{z}$ agree somewhere, then $f(\mathbf{w}, \mathbf{x}, \mathbf{y}, \mathbf{z}) \leq \frac{1}{2}$. Otherwise, we prove the bound by introducing an additional coordinate to each of the vectors $\mathbf{w}$, $\mathbf{x}$, $\mathbf{y}$,and $\mathbf{z}$, being the same for each and very small, borrowing from some other coordinate to preserve the sum condition. Specifically, given $(\mathbf{w}, \mathbf{x}, \mathbf{y}, \mathbf{z})$ in the domain of f with $\dim(\mathbf{w}) = \dim(\mathbf{x}) = m$ and $\dim(\mathbf{y}) = \dim(\mathbf{z}) = n$ assume that $w_k \neq 0 \neq x_k$ for some coordinate k and $y_\ell \neq 0 \neq z_\ell$ for some coordinate ℓ (if there is no such k for a given $(\mathbf{w}, \mathbf{x}, \mathbf{y}, \mathbf{z})$ then $f(\mathbf{w}, \mathbf{x}, \mathbf{y}, \mathbf{z}) = \sum_{j=1}^{n} \frac{y_j z_j}{y_j + z_j} \leq \frac{1}{2}$ by multi-dimensional

Jensen's inequality (similarly for ℓ)). Form the sequences $\widetilde{\mathbf{w}}(\varepsilon) \in \mathbb{R}^{m+1}$, $\widetilde{\mathbf{x}}(\varepsilon) \in \mathbb{R}^{m+1}$, $\widetilde{\mathbf{y}}(\varepsilon) \in \mathbb{R}^{n+1}$,$\widetilde{\mathbf{z}}(\varepsilon) \in \mathbb{R}^{n+1}$ defined by

$$\widetilde{w_i}(\varepsilon) := \begin{cases} w_i & \text{if } i \neq k \text{ and } i \leq m \\ w_k - \varepsilon & \text{if } i = k \\ \varepsilon & \text{if } i = m+1 \end{cases}$$

$$\widetilde{x_i}(\varepsilon) := \begin{cases} x_i & \text{if } i \neq k \text{ and } i \leq m \\ x_k - \varepsilon & \text{if } i = k \\ \varepsilon & \text{if } i = m+1 \end{cases}$$

$$\widetilde{y_i}(\varepsilon) := \begin{cases} y_i & \text{if } i \neq k \text{ and } i \leq n \\ y_k - \varepsilon & \text{if } i = k \\ \varepsilon & \text{if } i = n+1 \end{cases}$$

$$\widetilde{z_i}(\varepsilon) := \begin{cases} z_i & \text{if } i \neq k \text{ and } i \leq n \\ z_k - \varepsilon & \text{if } i = k \\ \varepsilon & \text{if } i = n+1 \end{cases}$$

then $(\widetilde{\mathbf{w}}(\varepsilon), \widetilde{\mathbf{x}}(\varepsilon), \widetilde{\mathbf{y}}(\varepsilon), \widetilde{\mathbf{z}}(\varepsilon))$ satisfies the conditions of the previous Lemma III.3 for small enough ε and so $f(\widetilde{\mathbf{w}}(\varepsilon), \widetilde{\mathbf{x}}(\varepsilon), \widetilde{\mathbf{y}}(\varepsilon), \widetilde{\mathbf{z}}(\varepsilon)) \leq \frac{1}{2}$ for small ε. Since $f(\widetilde{\mathbf{w}}(\varepsilon), \widetilde{\mathbf{x}}(\varepsilon), \widetilde{\mathbf{y}}(\varepsilon), \widetilde{\mathbf{z}}(\varepsilon))$ is a continuous function of ε, we have that $f(\mathbf{w}, \mathbf{x}, \mathbf{y}, \mathbf{z}) = f(\widetilde{\mathbf{w}}(0), \widetilde{\mathbf{x}}(0), \widetilde{\mathbf{y}}(0), \widetilde{\mathbf{z}}(0)) \leq \frac{1}{2}$ □

Corollary III.2. *With the above definitions, we have*

$$\sum_{i=1}^{m} \sum_{j=1}^{n} \frac{(w_i y_j - x_i z_j)^2}{w_i y_j + x_i z_j} \leq \sum_{i=1}^{m} \frac{(w_i - x_i)^2}{w_i + x_i} + \sum_{j=1}^{n} \frac{(y_j - z_j)^2}{y_j + z_j}$$

Proof.

$$\sum_{i=1}^{m}\sum_{j=1}^{n}\frac{(w_iy_j - x_iz_j)^2}{w_iy_j + x_iz_j} - \sum_{i=1}^{m}\frac{(w_i - x_i)^2}{w_i + x_i} - \sum_{j=1}^{n}\frac{(y_j - z_j)^2}{y_j + z_j}$$

$$= \sum_{i=1}^{m}\sum_{j=1}^{n}\frac{(w_iy_j + x_iz_j)^2 - 4w_ix_iy_jz_j}{w_iy_j + x_iz_j}$$

$$- \sum_{i=1}^{m}\frac{(w_i + x_i)^2 - 4w_ix_i}{w_i + x_i} - \sum_{j=1}^{n}\frac{(y_j + z_j)^2 - 4y_jz_j}{y_j + z_j}$$

$$= 2 - 4\sum_{i=1}^{m}\sum_{j=1}^{n}\frac{w_ix_iy_jz_j}{w_iy_j + x_iz_j} - 2 + 4\sum_{i=1}^{m}\frac{w_ix_i}{w_i + x_i} - 2 + 4\sum_{j=1}^{n}\frac{y_jz_j}{y_j + z_j}$$

$$= -2 + 4\left(\sum_{i=1}^{m}\frac{w_ix_i}{w_i + x_i} + \sum_{j=1}^{n}\frac{y_jz_j}{y_j + z_j} - \sum_{i=1}^{m}\sum_{j=1}^{n}\frac{w_ix_iy_jz_j}{w_iy_j + x_iz_j}\right) \leq 0$$

□

We now extend this result to more than just two pairs of sequences:

Corollary III.3. *For each i from 1 to k, let $(x_i^{\alpha_i})_{\alpha_i \in A_i}$ and $(y_i^{\alpha_i})_{\alpha_i \in A_i}$ be sequences with $\sum_{\alpha_i \in A_i} x_i^{\alpha_i} = 1 = \sum_{\alpha_i \in A_i} y_i^{\alpha_i}$, then*

$$\sum_{(\alpha_1,\ldots,\alpha_n)\in A_1\times\cdots\times A_n}\frac{\left(\prod_{i=1}^{k} x_i^{\alpha_i} - \prod_{i=1}^{k} y_i^{\alpha_i}\right)^2}{\prod_{i=1}^{k} x_i^{\alpha_i} + \prod_{i=1}^{k} y_i^{\alpha_i}} \leq \sum_{i=1}^{k}\sum_{\alpha_i\in A_i}\frac{(x_i^{\alpha_i} - y_i^{\alpha_i})^2}{x_i^{\alpha_i} + y_i^{\alpha_i}}$$

Proof. This follows by induction and Corrolary III.2 since

$$\sum_{(\alpha_2,\ldots,\alpha_n)\in A_2\times\cdots\times A_n}\prod_{i=2}^{k} x_i^{\alpha_i} = \prod_{i=2}^{k}\sum_{\alpha_i\in A_i} x_i^{\alpha_i} = 1$$

□

We can now apply this to our situation with probabilities of configurations on a tree:

$$\sum_{\sigma\in\Omega_v} \frac{\left(\mathbb{P}\left(X_v^0=\sigma\right)-\mathbb{P}\left(X_v^1=\sigma\right)\right)^2}{\mathbb{P}\left(X_v^0=\sigma\right)+\mathbb{P}\left(X_v^1=\sigma\right)}$$

$$=\sum_{\sigma} \frac{\left(\prod_{w\in C(v)}\mathbb{P}(X_v^0(w)=\sigma_w)-\prod_{w\in C(v)}\mathbb{P}(X_v^1(w)=\sigma_w)\right)^2}{\prod_{w\in C(v)}\mathbb{P}(X_v^0(w)=\sigma_w)+\prod_{w\in C(v)}\mathbb{P}(X_v^1(w)=\sigma_w)}.$$

Applying Corollary III.3, we get that this is less than

$$\sum_{w\in C(v)}\sum_{\sigma\in\Omega_w} \frac{\left(\mathbb{P}\left(X_v^0(w)=\sigma\right)-\mathbb{P}\left(X_v^1(w)=\sigma\right)\right)^2}{\mathbb{P}\left(X_v^0(w)=\sigma\right)+\mathbb{P}\left(X_v^1(w)=\sigma\right)}$$

$$=\sum_{w\in C(v)}\sum_{\sigma\in\Omega_w} \frac{(1-a-b)^2\left(\mathbb{P}\left(X_w^0(w)=\sigma\right)-\mathbb{P}\left(X_w^1(w)=\sigma\right)\right)^2}{(1-a+b)\mathbb{P}\left(X_w^0(w)=\sigma\right)+(1+a-b)\mathbb{P}\left(X_w^1(w)=\sigma\right)}. \quad \text{(III.10)}$$

Making use of the inequality $\frac{x+y}{(1-a+b)x+(1+a-b)y} \leq \frac{1}{1-|a-b|}$ for $0<x,y<1$, we get that (III.10) is less than

$$\frac{(1-a-b)^2}{1-|a-b|}\sum_{w\in C(v)}\sum_{\sigma\in\Omega_w} \frac{\left(\mathbb{P}\left(X_w^0(w)=\sigma\right)-\mathbb{P}\left(X_w^1(w)=\sigma\right)\right)^2}{\mathbb{P}\left(X_w^0(w)=\sigma\right)+\mathbb{P}\left(X_w^1(w)=\sigma\right)}. \quad \text{(III.11)}$$

Defining

$$\phi(v)=\sum_{\sigma\in\Omega_v} \frac{\left(\mathbb{P}\left(X_v^0=\sigma\right)-\mathbb{P}\left(X_v^1=\sigma\right)\right)^2}{\mathbb{P}\left(X_v^0=\sigma\right)+\mathbb{P}\left(X_v^1=\sigma\right)}$$

and $\theta=\frac{(1-a-b)^2}{1-|a-b|}$, we have proved that

$$\phi(v)\leq\theta\sum_{w\in C(v)}\phi(w)$$

For every $v\in T(\Pi)$ with $v\notin\Pi$. This inequality can be used to get a bound on for the reconstruction problem that is inferior to that which was proved earlier in this paper, but for which the constant θ is the same for $a=0$, $a=b$, and $a=b$.

The problem is that the bounds in (III.11) are not good. If we compare the results that we get for the reconstruction problem by using Corollary III.3 versus Lemma III.1, we see that their efficiency comes down to how well the functions $\frac{(x-y)^2}{x+y}$ and $\left(\sqrt{x}-\sqrt{y}\right)^2$ handle the respective equivalent of (III.11). That is, both methods lead to an inequality of the form

$$\phi_f(v) \leq \theta \sum_{w \in C(v)} \phi_f(w)$$

Where θ is a constant and the function ϕ_f is of the form

$$\phi(v) = \sum_{\sigma \in \Omega_v} f\left(\mathbb{P}\left(X_v^0 = \sigma\right), \mathbb{P}\left(X_v^1 = \sigma\right)\right).$$

In our bounds for the reconstruction problem, the function f was given by either $f(x,y) = \frac{(x-y)^2}{x+y}$ or $f(x,y) = \left(\sqrt{x}-\sqrt{y}\right)^2$. These two functions are related by

$$\left(\sqrt{x}-\sqrt{y}\right)^2 \leq \frac{(x-y)^2}{x+y}$$

For all $0 \leq x, y \leq 1$ (with appropriate limits at $(0,0)$). If we can establish an inequality of the form

$$\phi_f(v) \leq \theta_{fg} \sum_{w \in C(v)} \phi_g(w),$$

We might be able to get a better constant θ_{fg} if we use f and g that satisfy

$$\left(\sqrt{x}-\sqrt{y}\right)^2 \leq f(x,y) < g(x,y) \leq \frac{(x-y)^2}{x+y}$$

Since in this case, relaxing conditions of the functions f and g might yield a smaller $\theta_f g$ than in our previous bounds where we took $f = g$.

REFERENCES

[1] A. D. Barbour, Lars Holst, and Svante Janson, *Poisson approximation*, Oxford Studies in Probability, vol. 2, Oxford University Press, New York, 1992.

[2] Béla Bollobás, *Random graphs*, second ed., Cambridge Studies in Advanced Mathematics, vol. 73, Cambridge University Press, Cambridge, 2001.

[3] Christian Borgs, Jennifer Chayes, Elchanan Mossel, and Sebastien Roch, *The kesten-stigum reconstruction bound is tight for roughly symmetric binary channels*, 47th Annual IEEE Symposium on Foundations of Computer Science (2006), 518–530.

[4] Rick Durrett, *Random graph dynamics*, Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press, Cambridge, 2007.

[5] Paul Erdős and Alfréd Rényi, *On the evolution of random graphs*, Bulletin of the International Statistical Institute **38** (1961), 343–347.

[6] William Evans, Claire Kenyon, Yuval Peres, and Leonard J. Schulman, *Broadcasting on trees and the Ising model*, The Annals of Applied Probability **10** (2000), no. 2, 410–433.

[7] Nikolaos Fountoulakis and Bruce Reed, *The evolution of the mixing rate of a simple random walk on the giant component of a random graph*, Random Structures Algorithms **33** (2008), no. 1, 68–86.

[8] W. Keith Hastings, *Monte carlo sampling methods using markov chains and their applications*, Biometrika **57** (1970), no. 1, 97–109.

[9] Benjamini Itai, Gady Kozma, and Nicholas Wormald, *The mixing time of the giant component of a random graph*, (2006), Preprint.

[10] Mark Jerrum and Alistair Sinclair, *Approximating the permanent*, SIAM Journal on Computing **18** (1989), no. 6, 1149–1178.

[11] Harry Kesten and Bernt P. Stigum, *Additional limit theorems for indecomposable multidimensional Galton-Watson processes*, Annals of Mathematical Statistics **37** (1966), 1463–1481.

[12] David A. Levin, Yuval Peres, and Elizabeth L. Wilmer, *Markov chains and mixing times*, American Mathematical Society, Providence, RI, 2009.

[13] James B. Martin, *Reconstruction thresholds on regular trees*, Discrete random walks, Association of Discrete Mathematics and Theoretical Computer Science, Nancy, France, 2003, pp. 191–204.

[14] Nicholas Metropolis, Arianna W. Rosenbluth, Marshall N. Rosenbluth, Augusta H. Teller, and Edward Teller, *Equation of state calculations by fast computing machines*, The Journal of Chemical Physics **21** (1953), no. 6, 1087–1092.

[15] Elchanan Mossel, *Reconstruction on trees: beating the second eigenvalue*, The Annals of Applied Probability **11** (2001), no. 1, 285–300.

[16] Reza Rejaie, Daniel Stutzbach, Nick Duffield, Subhabrata Sen, and Walter Willinger, *On unbiased sampling for unstructured peer-to-peer networks*, (2006), Preprint.