Markovian Coupling vs. Conductance for the Jerrum-Sinclair Chain*

V.S. Anil Kumar & H. Ramesh

Abstract

We show that no Markovian Coupling argument can prove rapid mixing of the Jerrum-Sinclair Markov chain for sampling almost uniformly from the set of perfect and near perfect matchings of a given graph. In particular, we show that there exists a bipartite graph G such that any Markovian coupling argument on the Jerrum-Sinclair Markov chain for G must necessarily take time exponential in the number of vertices in G. This holds even when the coupling argument is Time-Variant, i.e., the transition probabilities used by the coupling process depend upon the history of the process. In contrast, the above Markov chain on G has been shown to mix in polynomial time using conductance arguments.

1 Introduction

The method of using a rapidly mixing Markov chain to construct an almost uniform sampling scheme for an exponentially large population was pioneered by Broder [2] in his seminal paper on the approximation of the permanent. Since then this technique has been applied to many problems in approximate counting and sampling [16].

Techniques for showing Rapid Mixing. The main challenge in using the above technique is to bound the mixing time of the Markov chain in question. Broder used a technique called *Coupling* to show that the mixing time was polynomial, provided the ratio of the number of near perfect matchings to perfect matchings is polynomial. Coupling was known earlier as a method to prove that a given chain achieves the right stationary distribution [17, 21]. Aldous [1] was the first to use coupling to show rapid mixing. We will describe this technique in detail shortly. However, as pointed out by Mihail [20], there was an error in Broder's coupling argument. Jerrum and Sinclair [14] later showed that Broder's chain (actually a slight variant of it) mixes

*Department of Computer Science and Automation, Indian Institute of Science, Bangalore, 560012. {vsa,ramesh}@csa.iisc.ernet.in.

rapidly using a completely different technique, which involved showing that the underlying graph had large *Conductance*.

The Coupling Method. The coupling argument to show that a given Markov chain $\mathcal{X}$ mixes rapidly starting from initial distribution π' proceeds as follows. Let $\mathcal{S}$ denote the state space of $\mathcal{X}$ and π be its stationary distribution. A new process, called the joint or coupling process, is defined on the state space $\mathcal{S} \times \mathcal{S}$. The initial distribution for the joint process is $\pi' \times \pi$. The moves of the joint process need only satisfy the following constraints: at every instant t, the marginal distribution with respect to the first component be identical to the distribution of $\mathcal{X}$ at time t starting from π', and the marginal distribution with respect to the second component be identical to the distribution of $\mathcal{X}$ at time t starting from π. The Coupling Lemma [12, 1] states that the time required for the probability that the two components are not identical to fall below some parameter ϵ (called the coupling time), is an upper bound on the *mixing time*, i.e., the time taken by $\mathcal{X}$ to get ϵ-close to its stationary distribution. Therefore, to show rapid mixing of $\mathcal{X}$, it suffices to obtain a joint process which satisfies the above marginal distribution constraints and has a small coupling time.

A sufficient but not necessary condition for satisfying the above constraints is the following: the marginal transition probabilities of the joint process with respect to each of the two components be identical to the corresponding transition probabilities in $\mathcal{X}$. Coupling processes in which this strong condition is imposed are called *Markovian*, as opposed to *Non-Markovian* couplings, in which the constraints of the above paragraph can be imposed used weaker restrictions. Note that even for Markovian couplings, the transition probabilities used by the joint process could be *Time-Variant*, i.e., functions of the history of the process. Non-Markovian couplings are more powerful because of their ability to "look" into the future.

The coupling technique is described in detail in [12, 17]; the former deals only with Markovian couplings. Griffeath[10] showed that there always exists a maximal coupling, i.e., a Non-Markovian coupling whose coupling time equals the mixing time of the Markov chain. Thus, in

principle, it is possible to show rapid mixing (if the chain in question indeed mixes rapidly) by setting up an appropriate Non-Markovian coupling and bounding its coupling time. However, the latter task is typically hard.

Coupling v/s Conductance. Our original aim was to study the following question: is the Markovian coupling method as powerful as the conductance method for showing rapid mixing?

There are many situations where coupling has been used to show rapid mixing, e.g., estimating the volume of convex bodies [5], counting the number of linear extensions of a partial order [4], counting k-colourings [3, 11], counting independent sets [19]. There are others in which conductance has been used, e.g., estimating the volume of convex bodies [9, 18], counting matchings in dense graphs [14], estimating the partition function in the Ising Model [15].

To the best of our knowledge, most instances of coupling in the theoretical computer science literature seem to actually be Markovian couplings. The one exception we know of is the paper by Czumaj et al [7], which uses a Non-Markovian coupling. One reason for the popularity of Markovian couplings is the relative ease of setting up and analysing Markovian couplings over Non-Markovian[1] ones.

Mihail [20] and Sinclair [22] point out that Markovian coupling seems ill-suited to prove rapid mixing for chains which lack symmetry. Recent work by Burdzy and Kendall [6] gave the first formal result on the the weakness of Markovian coupling. They showed the existence of chains for which no *efficient* Markovian coupling exists, i.e., there is no Markovian coupling which couples at the maximum possible exponential rate, given by the spectral gap. However, this does not seem to preclude the existence of Markovian couplings which couple at "approximately" this maximum rate.

Our Result. We show that no Markovian coupling argument can prove rapid mixing for the Jerrum-Sinclair chain. In particular, we show that there exists a graph G such that any Markovian coupling argument on the Jerrum-Sinclair chain [14] for sampling perfect and near perfect matchings in G from a distribution which is ϵ-close to uniform, even if Time-Variant, must take at least $O(c^n \ln \frac{1}{\epsilon})$ time, for some constant $c > 1$. The ratio of near perfect matchings to perfect matchings in G will be polynomial, so that the above chain can be shown to mix in polynomial time for G using conductance arguments.

The bipartite graph G above will have n vertices on either side along with some additional properties, which will help prove lower bounds on the coupling time. Consider any

coupling process for the Jerrum-Sinclair chain on graph G. Such a process can be viewed as a directed graph $\mathcal{M}$, where the vertices are pairs from the set of all perfect matchings and near perfect matchings in G. We will assume that vertices in $\mathcal{M}$ in which both components of the pair are identical have transitions to a special sink state with probability 1. The transition probabilities at other vertices in process $\mathcal{M}$ could be arbitrary and functions of history, as long as they satisfy the appropriate marginal constraints. We show that any specification of these transition probabilities must have coupling time exponential in n.

Note that we will be able to show the above lower bound on the coupling time only if $\mathcal{M}$ starts at certain vertices in $\mathcal{M}$; in fact, the coupling time starting at a vertex will be exponential in its "distance" to the sink. We will also show that any coupling strategy must start with most of its probability mass on vertices which are indeed far away from the sink. As further evidence of the weakness of Markovian coupling, we show that the hitting time from *any* non-sink vertex in $\mathcal{M}$ to the sink is at least exponential in n. This is in contrast to the coupling time, which is exponential starting only at certain vertices.

Our proof has two main aspects.

Bounding Transition Probabilities. The key observation we make is that the process $\mathcal{M}$ behaves essentially like a linear chain (a chain in which state i has transitions only to state $i - 1$ and state $i + 1$) having a single sink (i.e., state 0), with probabilities of transitions towards the sink being only a constant fraction of those away from the sink. Showing this involves bounding the probabilities of *distance increasing* and *distance decreasing* transitions for any coupling strategy. The *distance* at each vertex in $\mathcal{M}$ is defined as the symmetric difference between the two matchings or near perfect matchings associated with that vertex. We show lower bounds on the probabilities of distance increasing transitions and upper bounds on the probabilities of distance decreasing from every non-sink vertex. Certain properties of graph G will be used critically in this process. An important point to be noted is that these bounds need to hold for every coupling strategy, and therefore we are constrained to using only the marginal constraints and not the exact nature of the coupling strategy in deriving them.

Bounding Coupling Time and Hitting Times. The second aspect involves transforming the chain $\mathcal{M}$ to an almost linear chain $\mathcal{K}$ and analyzing the coupling and hitting times of this chain. The linearity of $\mathcal{K}$ makes the bounding of coupling and hitting times easy. One irritant is that $\mathcal{K}$ is not strictly linear, e.g., there may be transitions from a vertex in $\mathcal{K}$ to vertices up to 8 steps towards and away from the sink. In addition, the transitions probabilities to vertices towards the sink could be higher than those to vertices away from the sink for some suffix of $\mathcal{K}$.

---

[1]The coupling used by Broder [2] was a Time-Variant Markovian coupling according to our definition, though the paper uses the term Non-Markovian

Actually, the chains $\mathcal{K}$ obtained for bounding the coupling time and the hitting times will be different. Our technique for obtaining $\mathcal{K}$ to bound the hitting times is outlined below and might be of independent interest as well.

We perform what we call *hitting time decreasing* transformations to obtain chain $\mathcal{K}$ (these transformations will be different from those used above for bounding coupling time). These transformations involve altering the transitions at many vertices simultaneously, each alteration being in favour of vertices with lower hitting time. For example, if b has smaller hitting time than c, then the probability of transition from a to b is increased and that from a to c is decreased by the corresponding amount. We will show that such alterations only lead to reduction in the hitting time of *every* vertex. Thus hitting times in $\mathcal{K}$ are all lower than the corresponding hitting times in $\mathcal{M}$.

Time-Variant Markovian Couplings. For Time-Variant Markovian coupling strategies, the above proof for lower bounding the coupling time will still hold, as this proof relies only upon bounds on probability transitions, not the actual values. And these bounds must be respected by any Time-Invariant strategy as well.

Our proof for lower bounding hitting times breaks down for Time-Variant Markovian couplings. However, it is well known that there must exist a Time-Invariant Markovian coupling strategy with hitting times only infinitesimally larger than that of any Time-Variant Markovian strategy. The exponential lower bound on the hitting times for any Time-Variant Markovian strategy will also follow.

Road-map. Section 2 gives some preliminary definitions and a description of the Jerrum-Sinclair chain. A formal description of the Markovian coupling method appears in Section 2.1. Section 3 shows the existence of a graph G with certain properties for which we claim that any coupling strategy on Jerrum and Sinclair's chain will require time exponential in n. In Section 4, we will set up the coupling chain, and bound the probabilities of distance modifying transitions in this chain. Section 5 will describe the bounding of coupling times in this chain. Sections 4 and 5 will hold for Time-Variant Markovian couplings as well. Section 6 will describe the key theorem used in lower bounding of hitting times for Markovian couplings. Further details of this part will appear in the full version of the paper.

2 Preliminaries

Let G denote any bipartite graph with $n + n$ vertices and let V_1, V_2 denote the left and the right sets, respectively, of vertices in G. Let m denote the number of edges in G.

Notations for Markov Chains. For a Markov chain X, let $p_X(a, b)$ denote the probability of the transition from vertex

a to vertex b. Let $h_X(a, b)$ denote the hitting time from a to b in X.

2.1 The Markovian Coupling Method.

Suppose we wish to show that X converges to its stationary distribution π starting at distribution π_1. Let π denote the stationary distribution of X.

The Joint process $\mathcal{M}$. We take another copy Y of X, and consider a process $\mathcal{M}$ whose state space is the Cartesian product of the state spaces of X and Y. This process is defined as follows. The initial distribution for the states of $\mathcal{M}$ is just $\pi_1 \times \pi$. Moves in $\mathcal{M}$ from one joint state to another are made according to probabilities given by a family of functions for each joint state, as described below.

Transitions in $\mathcal{M}$. For states in $\mathcal{M}$ in which both components are identical, there is a transition with probability 1 to an extra *sink* vertex. Transitions for all other states are specified as follows. For each such state $v = (x, y)$ in $\mathcal{M}$, the coupling strategy specifies a family $F_v = \{f_v^{W_1}, f_v^{W_2}, \cdots\}$ of probability vectors, where $\{W_1, W_2, \cdots\}$ is the set of all possible *histories*, i.e., all finite length sequences of states that could be visited by process $\mathcal{M}$ since its inception and until it reaches v. The next move of the process $\mathcal{M}$ from state v is made according to the vector $f_v^{W_i}$, where W_i is the current history. In other words, the move from state v to state w at any instant when the history is W_i is made with probability $f_v^{W_i}(w)$.

Properties of F_v. Any Markovian coupling argument involves specifying the above families F_v with each $f_v^{W_i} \in F_v$ constrained as below. Let $v = (x, y)$ denote the state of $\mathcal{M}$ under consideration, where $x \neq y$. Recall that if $x = y$ then v transits to a sink with probability 1.

1. For each $x' \in X$, consider the set of states $T(x')$ in $\mathcal{M}$ whose first entry is x'; then $\sum_{w \in T(x')} f_v^{W_i}(w)$ equals $p_X(x, x')$, the probability of moving from x to x' in chain X.

2. For each $y' \in Y$, consider the set of states $T(y')$ in $\mathcal{M}$ whose second entry is y'; then $\sum_{w \in T(y')} f_v^{W_i}(w)$ equals $p_Y(y, y')$, the probability of moving from y to y' in chain Y.

Time-Variant and Time-Invariant Markovian Couplings. For each vertex v, if f_v^W is identical for all histories, the coupling is said to be *Time-Invariant*. Otherwise, it is said to be *Time-Variant*.

The Coupling Lemma. Define the *coupling time*, t_ϵ, to be the time at which the probability of being at non-sink states in $\mathcal{M}$ falls below ϵ starting from the distribution $\pi_1 \times \pi$, for

any given $\epsilon < 1$. The *Coupling Lemma* [12, 1, 8] states that t_ϵ is an upper bound on the time required for the chain X to become ϵ-close to π, starting from the initial distribution π_1. We will lower bound t_ϵ by $\Omega(c^n \ln \frac{1}{\epsilon})$, for some constant $c > 1$, for Time-Variant Markovian coupling strategies.

2.2 The Jerrum-Sinclair Chain

The states in this chain are the perfect and near perfect matchings in a bipartite graph $G = (V_1, V_2, E)$. At each state M, this chain does nothing with probability 1/2 and chooses an edge e in E uniformly at random with probability 1/2. The following action is then performed in the latter case.

1. If $e \in M$ and M is a perfect matching, e is deleted.

2. Suppose M is a near perfect matching. Let $e = (u, v)$. If u, v are both unmatched in M, then e is added. If exactly one of them is unmatched, then e is added and e' is deleted, where e' is the edge in M incident on whichever of u, v is matched; we call this a *swap* move.

3. If none of the above conditions holds, then no transition is made.

It was shown by Jerrum and Sinclair[14] that the above chain mixes fast if the ratio of the number of near perfect matchings to the number of perfect matchings is a polynomial in n.

3 Existence of Graph G

We will require the graph G to have the following properties.

Lemma 1 *There exists a $n + n$ bipartite graph G with the following properties.*

1. *The ratio of the number of near-perfect matchings to perfect matchings is polynomial in n and there are $\Omega(\frac{n!}{c^n})$ perfect matchings, for some constant $c > 1$.*

2. *Each vertex has degree at least αn, for some constant $\alpha < 1/2$.*

3. *For every pair of vertices, the intersection of their neighbourhoods has size at most $\alpha n/2$.*

4 The Coupling Chain $\mathcal{M}$ for Graph G.

All properties mentioned in this section will hold for Time-Variant Markovian couplings.

The Coupling Chain. Consider a coupling chain $\mathcal{M}$ on the set of tuples (x, y) of perfect and near perfect matchings of G with a transition from (x, y) to (x', y') if $x \to x'$ and $y \to y'$ are transitions in the original chain. The probabilities for the transitions are determined by the specific coupling strategy used.

Partitioning the States. The states of the above coupling chain are partitioned into layers $L(i)$, $i = 0 \dots 2n$, where $L(i)$ contains all (x, y) such that $|x \oplus y| = i$. Each set $L(i)$ is further partitioned into two sets $Top(i)$ and $Bot(i)$, where $Bot(i) = \{(x, y) | \exists$ vertex $v \in G$ which is unmatched in exactly one of $x, y\}$ and $Top(i) = \{(x, y) |$ either both x and y are perfect matchings or both are near perfect matchings with the same unmatched vertices $\}$.

A move in $\mathcal{M}$ from $L(i)$ to $L(j)$, is called *leftwards* or *distance reducing* if $j < i$, and *rightwards* or *distance increasing* if $j > i$. We assume that all transitions from vertices in $L(0)$ are to a new *sink vertex* with probability 1.

Initial Distribution. Note that for any perfect or near perfect matching M, the number of perfect and near perfect matchings with symmetric difference at most $n/4$ with respect to M is bounded by $n\binom{n}{n/4}(\frac{n}{4})!$, which is a negligible fraction of $\frac{n!}{c^n}$, the number of perfect matchings in G (see Lemma 1). Thus almost all of the probability mass of the initial distribution $\pi_1 \times \pi$ will lie on vertices in $L(i)$, $i \geq n/4$. Here π_1 is the initial distribution for the Jerrum-Sinclair chain on graph G and π is its stationary distribution. For simplicity, we assume that the chain $\mathcal{M}$ begins at some vertex in $L(i)$, $i \geq n/4$.

4.1 Bounds of Transition Probabilities in $\mathcal{M}$

We bound the transition probabilities from each vertex. Ideally, we would like that each vertex has small leftwards probability and large rightwards probability. The leftwards probability from every vertex is indeed small. However, note that small leftwards probability does not automatically imply large rightwards probability as there could be transitions within the same layer. Therefore, it is not sufficient to just show that the leftwards probability is small. As we shall show in Lemma 3 and Lemma 4, vertices in $Bot()$ have small leftwards probability and large rightwards probability for any coupling strategy, as desired. Unfortunately, for vertices in $Top()$, there are coupling strategies which ensure low or zero rightwards probability. However, as we show in Lemma 5 and Lemma 6, the transition probabilities from vertices in $Bot()$ to vertices in $Top()$ is small in any coupling strategy; further, from vertices in $Top()$, there are transitions only to vertices in $Bot()$. This effectively ensures that the lack of large rightwards probabilities at vertices in $Top()$ does not translate to large leftwards probabilities on the whole.

There is one additional caveat. Lemma 3 and 4 show that leftwards probabilities of vertices in $Bot(i)$ are small and rightwards probabilities are large, only as long as the symmetric difference i is some small enough constant fraction of n. In later sections, we will therefore have to handle layers with symmetric difference larger than some number i_0 (which will be a constant fraction of n) differently from other layers.

Lemma 2 *No transition in $\mathcal{M}$ can change the distance by more than 4.*

Proof: Each transition involves at most 4 edges. ∎

Lemma 3 *For any coupling strategy, the sum of the transition probabilities from $(x, y) \in Bot(i)$ to vertices in $L(j)$, $j < i$, is at most $\frac{2i+1}{2m}$.*

Proof: Without loss of generality, assume x is a near perfect matching. A move is distance reducing only if one of the following holds.

1. Suppose the move by x is a swap move. Then the edge swapped out of x must not be in $x \cap y$. Note that if the edge which is swapped out of x is in $x \cap y$ then the move by x will increase the distance by 2 and then the move by y can at most decrease the distance back by 2. The number of choices of the edge swapped in so that the edge swapped out is not in $x \cap y$ is at most $2|x - y| \leq 2i$, and therefore the probability of such an edge being chosen is at most $\frac{2i}{2m}$ in any coupling strategy.

2. Suppose the move by x takes it to a perfect matching. In this case, the distance may or may not decrease depending upon what move y makes. But the total probability of x moving to a perfect matching is at most $\frac{1}{2m}$ in any coupling strategy.

The total probability of distance reduction is thus $\frac{2i+1}{2m}$. ∎

Lemma 4 *For any coupling strategy, the sum of the transition probabilities from $(x, y) \in Bot(i)$ to tuples in $Bot(i+1) \cup Bot(i+2) \cup Bot(i+3) \cup Bot(i+4)$ is at least $\frac{\alpha n/2 - i - 2}{2m}$.*

Proof: Since $(x, y) \in Bot(i)$, one of the following conditions holds:

1. x is a near perfect matching and y is a perfect matching.

2. x is a perfect matching and y is a near perfect matching.

3. x and y are near perfect matchings with at most one common unmatched vertex.

We consider the first and the third cases in turn. The second case is symmetric to the first.

x **is near perfect,** y **is perfect** Let $a \in V_1$ and $b \in V_2$ be the unmatched vertices in x. We consider just one situation in which the distance increases and the resulting tuple (x', y') is also in $Bot(j)$, $j > i$. This situation is when x picks edge $e = (a, u)$, where $(u, u') \in x \cap y$ and moves to $x' = x + e - (u, u')$. The distance between x and y increases by 2 in the process, i.e., $|x' \oplus y| = |x \oplus y| + 2$. y can move to y' either by doing nothing or by deleting some edge, which still results in a net increase of distance by at least 1. Further u' and b are the unmatched vertices in x' and since $(u', b) \notin y$, at least one of them is matched in y', implying that $(x', y') \in Bot(j)$, $j > i$.

The number of edges $e = (a, u)$, where $(u, u') \in x \cap y$ is at least $\alpha n - |x - y|$. Therefore, probability of choosing such edges is at least $\frac{\alpha n - |x - y|}{2m} \geq \frac{\alpha n - |x \oplus y|}{2m} \geq \frac{\alpha n - i}{2m}$, for any coupling strategy.

x, y **are both near perfect.** Second, consider the case when x and y are both near perfect matchings. Suppose x has vertices $a \in V_1$ and $b \in V_2$ unmatched and y has vertices $c \in V_1$ and $d \in V_2$ unmatched. Without loss of generality, assume that b and d are distinct while c could be the same as a.

We will look at just one class of moves for x, which will increase distance by 2. These moves will occur with probability at least $\frac{\alpha n/2 - |x \oplus y|}{2m}$. Then, we will show that at most two choices of moves for y can reduce the distance if x makes a move in the above class; further, these choices will reduce the distance by 2 making the net change in distance 0 and each will occur with probability at most $\frac{1}{2m}$. In addition, we will show that if x makes a move in the above class and y does not make any of the above two possible moves then the resulting tuple (x', y') will be in $Bot(j)$, $j > i$. Thus the net probability of a move to $Bot(j)$, $j > i$, is at least $\frac{\alpha n/2 - |x \oplus y|}{2m} - \frac{2}{2m} = \frac{\alpha n/2 - i - 2}{2m}$, for any coupling strategy. Note that in calculating the probability that x makes a move in the above class and y does not make any of the above two possible moves, we have not assumed anything about the probability distribution of the joint move.

The above class of moves for x involves choosing an edge $e = (b, u)$, where u is not adjacent to d and (u, u') is a matching edge in both x and y, for some $u' \in V_2$. By Lemma 1, there are at least $\alpha n/2$ vertices in V_1 which are adjacent to b but not to d. Thus, there are at least $\alpha n/2 - |x - y| \geq \alpha n/2 - |x \oplus y|$ edges which are in $x \cap y$ and whose endpoint in V_1 is adjacent to b but not to d. Therefore, the above choice of moves occurs with probability at least $\frac{\alpha n/2 - |x \oplus y|}{2m}$ and results in an increase in distance by 2. The only moves for y which could decrease the distance back by 2 are when it chooses the unique edge $(c, c') \in x$, if any, or the unique edge $(d, d') \in x$. If neither of these hap-

pens, then y either does nothing or swaps in an edge which is not in x; the distance can not decrease back by 2 in either case. The probability of y choosing (c, c') or (d, d') is $\frac{2}{2m}$, for any coupling strategy. Finally, $u' \in V_2$ is unmatched in x' and must be matched in y', as $(u, u') \in y$ and (d, u) is not an edge in G; therefore (x', y') is in $Bot(j)$, $j > i$. ∎

Lemma 5 *For any coupling strategy, the sum of the transition probabilities from $(x, y) \in Bot(i)$ to tuples in $Top(i), Top(i+1)$ is at most $\frac{i+3}{2m}$.*

Proof: We enumerate all possibilities of moving to $Top(j)$, $j = i, i+1$, from (x, y) and verify the claim. Without loss of generality, assume x is a near perfect matching. There are two cases. The vertex moved to is either a pair of perfect matchings or a pair of near perfect matchings with identical unmatched vertices. The former possibility happens with probability at most $\frac{1}{2m}$ for any coupling strategy, because exactly one edge can make x a perfect matching. We consider the latter possibility next. There are two cases now.

First, consider the case when x is a near perfect matching with vertices a and b unmatched and y is a perfect matching. When x moves to another near perfect matching x', one of a, b remains unmatched. Then y must delete either the matching edge incident on a or that incident on b to achieve the same pair of unmatched vertices as in x'; this happens with probability at most $\frac{2}{2m}$, for any coupling strategy.

Second, consider the case when x is a near perfect matching with vertices a and b unmatched and y is a near perfect matching with vertices c and d unmatched, $b \neq d$. We show that if $a \neq c$ then there are exactly two swap moves for x which will lead to the unmatched vertices in x', y' being identical. The probability of these moves is at most $\frac{2}{2m}$. We also show that if $a = c$, then there are at most $|x \oplus y|$ moves for x which will lead to the unmatched vertices in x', y' being identical and (x', y') being in $Top(i) \cup Top(i+1)$ (actually, (x', y') will always be in $Top(i)$ for these moves). The probability of these moves is at most $\frac{|x \oplus y|}{2m} = \frac{i}{2m}$.

Suppose $a \neq c$. Then x must choose either the edge (a, c'), if any, such that $(c, c') \in x$, or the edge (d', b), if any, such that $(d', d) \in x$. The first leads to b, c being unmatched and the second to a, d being unmatched. These are the only pairs that can be unmatched in x' if the unmatched vertices in x', y' are identical.

Next, suppose $a = c$. Since x' and y' have to have the same unmatched vertices, a must either be matched in both of x', y' or be unmatched in both of them. In fact, a must be unmatched in x' and y', otherwise b would be unmatched in x' and d in y' and since $b \neq d$, (x', y') will not be in $Top(j)$ for any j. Let e denote the other common unmatched vertex in x', y'. Let (e', e) be the unique edge in x incident on e. We claim that (e', e) must be in $x - y$, for if $(e', e) \in x \cap y$

then x must choose (e', b) and y must choose (e', d) so that the unmatched vertices become identical; but this causes the distance to increase by 2, and (x', y') would end up in $Top(i+2)$ instead of $Top(i) \cup Top(i+1)$. Thus, there are only $|x - y| \leq |x \oplus y|$ moves for x. ∎

Lemma 6 *For any coupling strategy, the sum of the transition probabilities from $(x, y) \in Top(i)$ to tuples in $Top(j)$ is 0 for all $j \neq i$. Further, the sum of the transition probabilities to tuples in $Bot(j)$ is 0 for all $j < i - 2$.*

Proof: First, suppose x and y are both perfect matchings. There are two possibilities: either only one of x, y deletes an edge or both x and y delete an edge each. If only one of x, y deletes an edge, the resulting tuple is in $Bot(i-1)$ or $Bot(i+1)$. If both x and y delete the same edge, the distance remains unchanged and (x', y') remains in $Top(i)$. And if x and y delete different edges, the distance changes by at most two and the resulting unmatched vertices in x', y' are not identical. Therefore, (x', y') is in one of $Bot(i - 2), \ldots, Bot(i + 2)$ in this case.

Second, suppose both x and y are near perfect matchings with unmatched vertices $a \in V_1$ and $b \in V_2$. The possible moves are:

1. Both x and y add (a, b): the distance remains unchanged and $(x', y') \in Top(i)$.

2. x adds edge (a, b) and y does not add (a, b): x becomes a perfect matching while y remains a near perfect matching. y could, of course, make a swap move. But no matter what swap move it makes, the distance must always increase either by 1 or by 3. Therefore $(x', y') \in Bot(i + 1) \cup \cdots \cup Bot(i + 3)$. The case when y adds edge (a, b) and x doesn't is symmetrical.

3. x swaps in an edge $(a, c), c \neq b$, and y doesn't move: the distance changes by at most 2 and the unmatched vertices in x', y' are not identical. The case when y swaps in such an edge and x does not move is symmetrical.

4. Both x and y swap in the same edge $(a, c), c \neq b$: Let (c_1, c) be the matching edge in x and (c_2, c) be the matching edge in y. If $c_1 = c_2$, the distance is unchanged and (x', y') remains in $Top(i)$. And if $c_1 \neq c_2$, the distance reduces by two but (x', y') is in $Bot(i-2)$ because x' has c_1, b unmatched while y' has c_2, b unmatched.

5. x and y pick distinct edges $(a, c), (a, d), c \neq d$, respectively: Let (c', c) be the matching edge in x and (d', d) be the matching edge in y. If $c' = d'$, the distance does not change and (x', y') is in $Top(i)$. And if $c' \neq d'$, then the unmatched vertices in x', y' are not identical and distance cannot decrease, i.e., $(x', y') \in Bot(i) \cup \cdots \cup Bot(i + 4)$.

6

6. x picks $(a, c), c \neq b$ and y picks $(d, b), d \neq a$: $(x', y') \in Bot(j)$ for some j because a is unmatched in y' but not in x'. The distance cannot reduce in the process since $(a, c), (b, d)$ are not present in $x \cup y$. The case when y picks $(a, c), c \neq b$ and x picks $(d, b), d \neq a$ is symmetric.

■

5 Bounding Coupling Time

We now lower bound the coupling time t_ϵ starting in some state in $L(i)$, $i \geq n/4$. By Lemma 6, we can as well assume that the starting vertex is in $Bot(i)$ and not $Top(i)$.

Definitions. Let p denote the maximum over all vertices $v \in Bot(i)$, $1 \leq i \leq i_0$, of the sum of the transition probabilities from v to vertices in $Bot(j)$, $j < i$, and to vertices in $Top(j)$, $j \leq i + 1$. Here $i_0 = \frac{\alpha n}{6 + 14.3^7}$ (recall α from Lemma 1). Let q denote the minimum over all vertices $v \in Bot(i)$, $1 \leq i \leq i_0$, of the sum of the transition probabilities from v to vertices in $Bot(j)$, $j > i$. By Lemmas 3 and 5, $p \leq \frac{7\alpha n}{4m(3 + 7.3^7)}$, and by Lemma 4, $q \geq \frac{7\alpha n}{4m(3 + 7.3^7)} 3^7$. Therefore, $\frac{p}{q} \leq \frac{1}{3^7}$. These constants are all coarse calculations. The fact that the above ratio is small will be critical in showing the exponential lower bound on coupling time. Note that if we had considered $i > i_0$ as well in defining p, q, this ratio need not be small. All vertices in $L(i)$, $i > i_0$, will be called *tail* vertices.

Chain $\mathcal{K}$. We now obtain a new chain $\mathcal{K}$ from $\mathcal{M}$. All tail vertices will be retained as such in $\mathcal{K}$. Vertices in $Bot(i)$, $1 \leq i \leq i_0$, and the sink vertex will also be retained. The transition probabilities amongst these vertices will change however. There will not be any transitions in $\mathcal{K}$ between two vertices in the same $Bot(i)$, $i \leq i_0$. This is done is two steps.

Removing Vertices in $Top(i)$, $i \leq i_0$. In Step 1, we construct chain $\mathcal{M}'$ from $\mathcal{M}$ as follows. Consider each vertex $v \in Bot(i)$ in turn, $i \leq i_0$. We delete all outgoing transitions from v that lead to vertices in $Top()$ and redirect this probability to vertices of $Bot()$ as below. For each vertex w in $Bot(j)$, j possibly equal to i, we take all possible walks which start at v, end at w, and visit only vertices in $Top()$ on the way. We add the sum of the probabilities of all these walks to the original transition probability from v to w.

Note that by Lemmas 2 and 6, any vertex v in $Bot(i)$ in $\mathcal{M}'$ has transitions only to vertices w in $Bot(j)$, $i - 6 \leq j \leq i + 8$.

Lemma 7 *Consider any vertex $v \in Bot(i)$, $i \leq i_0$. The sum of the transition probabilities in $\mathcal{M}'$ from v to vertices w in $Bot(j)$, $j < i$, is at most p. In addition, the sum of* the transition probabilities from v to vertices w in $Bot(j)$, $j > i$, is at least q.

Proof: Let $p_1 = \sum_{w \in Bot(j), j < i} p_\mathcal{M}(v, w)$ and $p_2 = \sum_{w \in Top(j), j < i + 2} p_\mathcal{M}(v, w)$. Note that $p_1 + p_2 \leq p$, by definition. Consider all walks starting at v and ending at vertices in $Bot(j)$, $j < i$, and visiting only vertices in $Top()$ on the way. By Lemma 6, only vertices in $Top(i - 4) \cup \cdots \cup Top(i + 1)$ can be visited by these walks. The sum of the probabilities of all these walks is at most p_2. The first part of the lemma now follows because the additional probabilities given to transitions from v to vertices w in $Bot(j)$, $j < i$, is at most p_2.

The second part of the lemma holds trivially, as transition probabilities amongst vertices in $Bot()$ are never decreased.
■

Removing Transitions within a Layer. In Step 2, we construct chain $\mathcal{K}$ from $\mathcal{M}'$ as follows. Consider each vertex $v \in Bot(i)$ in turn, $i \leq i_0$. We delete all outgoing transitions from v that lead to other vertices in $Bot(i)$ and redirect this probability to vertices of $Bot(j)$, $j \neq i$, as below. For each vertex w in $Bot(j)$, $j \neq i$, we take all possible walks which start at v, end at w, and visit only vertices in $Bot(i)$ on the way. We add the sum of the probabilities of all these walks to the original transition probability from v to w.

Clearly, any vertex $v \in Bot(i)$ in $\mathcal{K}$ has transitions only to vertices w in $Bot(j)$, $i - 6 \leq j \leq i + 8$. Further, $p_\mathcal{K}(v, w) = 0$, for all $v, w \in Bot(i), i \leq i_0$.

Lemma 8 *For each vertex $v \in Bot(i)$ in $\mathcal{K}$, $i \leq i_0$, $\frac{\sum_{w \in Bot(j), j < i} p_\mathcal{K}(v, w)}{\sum_{w \in Bot(j), j > i} p_\mathcal{K}(v, w)} \leq \frac{p}{q} \leq \frac{1}{3^7}$, and $\sum_{w \in Bot(j), j < i} p_\mathcal{K}(v, w) + \sum_{w \in Bot(j), j > i} p_\mathcal{K}(v, w) = 1$*

Proof: Consider all walks starting at v and staying completely within $Bot(i)$. We partition these walks into classes $W(v')$, where v' is a vertex in $Bot(i)$. Those walks whose last vertex is v' are in $W(v')$. Consider any walk $U \in W(v')$. Let $p_{\mathcal{M}'}(U)$ denote the probability of this walk in $\mathcal{M}'$. Then U contributes $\sum_{w \in Bot(j), j < i} p_{\mathcal{M}'}(U) p_{\mathcal{M}'}(v', w)$ additional probability to the transitions from v to the left (i.e., to $Bot(j)$, $j < i$), and $\sum_{w \in Bot(j), j > i} p_{\mathcal{M}'}(U) p_{\mathcal{M}'}(v', w)$ from v to the right. This is because any walk starting at v and ending at vertex $w \notin Bot(i)$ but otherwise staying in $Bot(i)$ can be written as a concatenation of some walk U and a transition (v', w) above. By Lemma 7, $\sum_{w \in Bot(j), j < i} p_{\mathcal{M}'}(v', w) \leq p$ and $\sum_{w \in Bot(j), j > i} p_{\mathcal{M}'}(v', w) \geq q$. Thus the ratio of these two additional probabilities contributed by walk U is at most $\frac{p}{q} \leq \frac{1}{3^7}$. Since each walk contributes in the same ratio the lemma follows. ■

Lemma 9 *The coupling time in $\mathcal{M}$ starting at a vertex $v \in Bot(i)$ is at least the coupling time in $\mathcal{K}$ starting at the same vertex.*

Proof: Any walk W in $\mathcal{M}$ from v to the sink has a corresponding walk W' in $\mathcal{K}$ obtained by short-circuiting portions of W within $Top(i)$, $i \leq i_0$, and portions of W of length more than 1 which stay within some $Bot(i)$, $i \leq i_0$. A number of such walks W could map to the same walk W' in $\mathcal{K}$. It follows from the transformations made above that the sum of the probabilities of all these walks in $\mathcal{M}$ equals the probability of W' in $\mathcal{K}$. The lemma follows. ■

Bounding Coupling Time in $\mathcal{K}$. Recall that our starting vertex is in $Bot(i)$, $i \geq n/4 > i_0$. Since left transitions in $\mathcal{K}$ span at most 6 layers, we can as well assume that the starting vertex is in $Bot(j)$, $i_0 - 6 \leq j \leq i_0$. Define the *escape probability* $p_{esc}(v)$, $v \in Bot(i), i \leq i_0$, to be the sum of the probabilities of all *escape walks* for v, i.e., walks that start at vertex v, end at the sink, and stay to the left of $L(i_0 - 6)$.

Lemma 10 *For all $v \in Bot(i)$, $i_0 - 6 \leq i \leq i_0$, $p_{esc}(v)$ is at most $1/3^{i-1}$.*

Proof: We estimate the total sum of the probabilities of escape walks starting at vertex $v \in Bot(i)$.

Recall that in the chain $\mathcal{K}$, each left step involves jumping at most 6 vertices and each right step involves jumping at most 8 vertices. Recall $\mathcal{K}$ has no transitions between vertices in the same $Bot(j)$, for $j \leq i_0$.

There are at most 2^s LR-sequences of length s, where an LR-sequence is a sequence of lefts and rights, having at least $\frac{s+i}{7}$ lefts. Clearly, each escape walk for v of length s corresponds to some such LR-sequence. This is because $\#l + \#r = s$ and $\#l \geq \frac{i + \#r}{6}$ for each escape walk, where $\#l$ is the number of left moves and $\#r$ the number of right moves in the escape walk. However, many escape walks could correspond to the same LR-sequence, as there could be several left-going or right-going edges from each vertex. Further, note that not all LR-sequences above need have associated escape walks.

Using Lemma 8, we show below that the total probability of all escape walks of length s which correspond to a particular LR-sequence is at most $(\frac{1}{3^7+1})^{\frac{s+i}{7}}$. Then $p_{esc}(v) \leq \sum_{s=0}^{\infty} \sum_{\text{LR-sequences of length } s} (\frac{1}{3^7+1})^{\frac{s+i}{7}} \leq \sum_{s=0}^{\infty} 2^s (\frac{1}{3^7+1})^{\frac{s+i}{7}} \leq \sum_{s=0}^{\infty} \frac{1}{3^i (3/2)^s} \leq \frac{1}{3^{i-1}}$.

It remains to show that the total probability of all escape walks of length s which correspond to a particular LR-sequence with $\#l$ left moves is at most $(\frac{1}{3^7+1})^{\#l}$. Consider the tree T comprising all possible walks which start at v and confirm to this LR-sequence (i.e., if the first element in this sequence is an L then the children of the root v correspond to all vertices reachable from v by making a single

left move; further, if the second element is an R, then the children of any child w of the root correspond to all vertices reachable from w by making a single right move, and so on). These walks include all escape walks for v which correspond to this particular LR-sequence as well. We bound the probabilities of all walks in this tree by $(\frac{1}{3^7+1})^{\#l}$. This is done by induction on the height of the tree as follows.

The induction hypothesis states that for every subtree S of the above tree, the sum of the probabilities of walks in this subtree is at most $(\frac{1}{3^7+1})^{\#l_S}$, where $\#l_S$ is the number of layers in S corresponding to left moves. Note that $\#l_T = \#l$.

For the base case, consider subtrees S having just one layer of edges (i.e., corresponding to walks of length 1). If $\#l_S = 0$ then the claim is true because all these walks of length 1 are right-going transitions from a single vertex and must have total probability at most 1. And if $\#l_S = 1$ then the claim is true because all these walks of length 1 are left-going transitions from a single vertex and must have total probability at most $(\frac{1}{3^7+1})$, by Lemma 8. The inductive step is now easy to show by combining this argument for the base case with the induction hypothesis. ■

Theorem 1 *t_ϵ starting at any vertex $v \in Bot(i), i_0 - 6 \leq i \leq i_0$, is $\Omega(3^{i_0} \ln \frac{1}{\epsilon}) = \Omega(c^n \ln \frac{1}{\epsilon})$, for some constant $c > 1$.*

Proof: We show below that the sum of the probabilities of all walks starting at vertex $v \in A = \{i_0 - 6, \ldots, i_0\}$, picked from any probability distribution on A, and hitting vertices in A at least t times is at least $(1 - p_{max})^t$, where $p_{max} = \max_{w \in A}(p_{esc}(w))$.

If $t \leq \frac{1}{p_{max}} \ln \frac{1}{\epsilon}$, the sum of the probabilities of all walks that visit vertices in A at least t times is more than ϵ. Therefore the sum of the probabilities of walks of length less than t that hit the sink is less than $1 - \epsilon$. Since t_ϵ is defined to be the time needed for the probability of hitting the sink to be at least $1 - \epsilon$, it follows that $t_\epsilon \geq \frac{1}{p_{max}} \ln \frac{1}{\epsilon}$. The theorem follows from Lemma 10.

It needs to be shown that the sum of the probabilities of walks that visit vertices in A at least t times is at least $(1 - p_{max})^t$ for any finite t. By definition, $p_{esc}(v)$ is the sum of the probabilities of escape walks starting at v; these do not visit vertices in A. Hence, the sum of the probabilities of walks starting at v and visiting A at least once is $1 - p_{esc}(v) \geq 1 - p_{max}$ by definition of p_{max}. If the starting vertex v is picked from A using any distribution on A, the sum of the probabilities of the above walks is still at least $1 - p_{max}$. Assume inductively that the sum of the probabilities of escape walks that visit vertices in A at least $t - 1$ times is $(1 - p_{max})^{t-1}$, $t \geq 2$. If we consider the $(t-1)st$ occurrence of a vertex of A among paths that visit A at least $t-1$ times, a probability distribution is induced on A. The process of visiting A at least t times can be viewed

as visiting A $t-1$ times and then taking paths which visit A at least once. From the base case the total probability of visiting A at least once starting at any distribution on A is at least $(1 - p_{max})$. Therefore the sum of the probabilities of walks that visit vertices in A at least t times is at least $(1 - p_{max})^t$, as required. ∎

6 Bounding Hitting Times

Without loss of generality, we will consider only Time-Invariant Markovian couplings.

Sub-stochastic Matrices. We will need the following fact about *sub-stochastic matrices*, i.e., those matrices in which every row sum is at most 1 and at least one row sum is strictly less than 1. Consider any sub-stochastic matrix M of dimension $n \times n$. Associated with M is a Markov chain $\mathcal{L}$ with $n+1$ vertices; we call L the graph *associated* with M. The transition probabilities amongst the last n of these vertices are as given by M; the first vertex is a sink and each of the last n vertices move to the sink with probability equal to the deficit (relative to 1) in their respective row sums in M.

Lemma 11 *For any sub-stochastic matrix M, if there is a path from every vertex to the sink in the associated graph then, $M^k \to 0$ as $k \to \infty$.*

A Key Theorem. Consider a Markov Chain X which has a sink vertex s and a path from each vertex to the sink in the associated graph. Suppose, we perform the following operation repeatedly on X: pick up vertices u, v, w, where $u, v \neq s$, reduce the transition probability from u to v by some amount and increase the transition probability from u to w by the same amount. Let Y denote the new chain obtained in the process. The following theorem holds.

Theorem 2 *Suppose there is a path in Y from each vertex to s. The the following two facts hold.*

1. *Suppose that for each triple u, v, w on which the above change is made, $h_X(v, s) \le h_X(w, s)$. Then, for all vertices $a \in X$, $h_Y(a, s) \ge h_X(a, s)$.*

2. *Suppose that for each triple u, v, w on which the above change is made, $h_X(v, s) \ge h_X(w, s)$. Then, for all vertices $a \in X$, $h_Y(a, s) \le h_X(a, s)$.*

Proof: Consider the vector $\hat{f}$ whose ath component is $h_X(a, s)$, $a \in X$, $a \neq s$, (the vector is the list of hitting times to s for each vertex $a \in X$, $a \neq s$). Let $\hat{g}$ be the corresponding vector for Y. Let F be the probability transition matrix for X, restricted to non-sink vertices, i.e., $F(a, b) = p_X(a, b)$, $a, b \in X$, $a, b \neq s$. Let G be the corresponding matrix for Y. F is sub-stochastic because there is

a path in X from each vertex to the sink. G is sub-stochastic as well for the same reason.

Since $h_X(a, s) = 1 + \sum_b$ in the neighbourhood of a $(p_X(a, b) h_X(b, s))$, $\hat{f}$ satisfies $\hat{f} = F\hat{f} + \hat{b}$, where $\hat{b}$ is the all 1s vector. Similarly, $\hat{g} = G\hat{g} + \hat{b}$. We consider each of the two situation in the theorem in turn.

First, consider the first situation. We will show below that $(F - G)\hat{f} \le 0$. It follows that $\hat{f} = F\hat{f} + \hat{b} \le G\hat{f} + \hat{b} \le G^2\hat{f} + G\hat{b} + \hat{b} \le \ldots \le G^k\hat{f} + (G^{k-1} + G^{k-2} + \cdots + I)\hat{b}$, for all $k \ge 1$. By Lemma 11, G^k tends to 0 as k tends to ∞, and therefore $(G^{k-1} + G^{k-2} + \cdots + I)$ tends to $(I - G)^{-1}$. Therefore, $\hat{f} \le (I - G)^{-1}\hat{b} = \hat{g}$, as required.

Second, consider the second situation. We will show below that $(F - G)\hat{f} \ge 0$. It follows that $\hat{f} = F\hat{f} + \hat{b} \ge G\hat{f} + \hat{b} \ge G^2\hat{f} + G\hat{b} + \hat{b} \ge \ldots \ge G^k\hat{f} + (G^{k-1} + G^{k-2} + \cdots + I)\hat{b}$, for all $k \ge 1$. As before, $\hat{f} \ge (I - G)^{-1}\hat{b} = \hat{g}$, as required.

It remains to be shown that $(F - G)\hat{f} \le 0$ in the first situation and $(F - G)\hat{f} \ge 0$ in the second. We show these two facts below.

Note that the ath entry $(F - G)\hat{f}$ is just $\sum_{d \in X, d \neq s}(p_X(a, d) - p_Y(a, d))h_X(d, s)$; denote this sum by $\hat{s}(a)$. Consider the operations performed on X above to obtain Y. Consider the Markov chains $X = Z_1, Z_2, \ldots, Z_k = Y$ obtained in the process, where Z_i is the chain after the ith operation. Consider the sum $\hat{s}_i(a) = \sum_{d \in X, d \neq s}(p_{Z_i}(a, d) - p_{Z_{i+1}}(a, d))h_X(d, s)$ for all vertices $a \in X$, $a \neq s$. Then $\hat{s}(a) = \sum_i \hat{s}_i(a)$, for all a. It then suffices to show that each $\hat{s}_i(a)$ is non-positive in the first situation and non-negative in the second.

Let u, v, w be the vertices involved in the operation used to obtain Z_{i+1} from Z_i. For all vertices $e \neq u$, $\hat{s}_i(e)$ is easily seen to be 0. Further, $\hat{s}_i(u) = \sum_{d \in X, d \neq s}(p_{Z_i}(u, d) - p_{Z_{i+1}}(u, d))h_X(d, s) = (p_{Z_i}(u, v) - p_{Z_{i+1}}(u, v))h_X(v, s) + (p_{Z_i}(u, w) - p_{Z_{i+1}}(u, w))h_X(w, s)$. Note that $(p_{Z_i}(u, v) - p_{Z_{i+1}}(u, v)) = -(p_{Z_i}(u, w) - p_{Z_{i+1}}(u, w))$. Since $h_X(v, s) \le h_X(w, s)$ in the first situation, $\hat{s}_i(u)$ is non-positive, and since $h_X(v, s) \ge h_X(w, s)$ in the second situation, $\hat{s}_i(u)$ is non-negative, as required. ∎

Definitions. Transformations of the above kind which satisfy the first condition in Theorem 2 are called *Hitting Time Increasing* transformations. Similarly, those which satisfy the second condition are called *Hitting Time Decreasing* transformations. We use the above theorem repeatedly to simplify the chain $\mathcal{M}$.

The proof of Lemma 11 is straightforward and is omitted here. The rest of the proof uses hitting time decreasing transformations to obtain an almost-linear chain $\mathcal{K}$ (which is different from the chain obtained in Section 5) and then bounds hitting times in $\mathcal{K}$.

6.1 Modifying $\mathcal{M}$ to $\mathcal{K}$.

We obtain $\mathcal{K}$ from $\mathcal{M}$ by performing hitting time decreasing transformations. These are done in 3 steps. First, we replace each set $Bot(i)$ and $Top(i)$ by single representative vertices, i.e., those which have the least hitting time in their respective sets. Second, we remove all the representatives of vertices in $Top(i)$. Finally, we truncate the chain by eliminating representatives of vertices in layers $L(i)$, $i \geq \frac{n}{c}$, for some suitable constant c. This constant is related to the fact that left going probabilities are low only when the symmetric difference is not too large (see Section 4).

Step 1. Let $b(i)$ denote the vertex in $Bot(i)$ with the smallest hitting time to s (s is the sink vertex introduced at the beginning of Section 4) in $\mathcal{M}$. Let $t(i)$ denote the vertex in $Top(i)$ with the smallest hitting time to s in $\mathcal{M}$. The $b(i)$s and the $t(i)$s are together called *representative* vertices. We obtain a new chain $\mathcal{M}'$ using a hitting time decreasing transformation as follows.

For each vertex $a \in \mathcal{M}$, we modify its outgoing transition probabilities as follows. For each set $Bot(j)$, $j > 0$, we set $p_{\mathcal{M}'}(a, b(j)) = \sum_{w \in Bot(j)} p_{\mathcal{M}}(a, w)$ and $p_{\mathcal{M}'}(a, w) = 0$ for all $w \in Bot(j) - \{b(j)\}$. Similarly, for each set $Top(j)$, $j > 0$, we set $p_{\mathcal{M}'}(a, t(j)) = \sum_{w \in Top(j)} p_{\mathcal{M}}(a, w)$ and $p_{\mathcal{M}'}(a, w) = 0$ for all $w \in Top(j) - \{t(j)\}$.

The following lemma relates the hitting times in $\mathcal{M}'$ to those in $\mathcal{M}$.

Lemma 12 *For each $w \in \mathcal{M}'$, $h_{\mathcal{M}'}(w, s) \leq h_{\mathcal{M}}(w, s)$. Each vertex in $\mathcal{M}'$ has a path to s.*

Note that only the representative vertices have incoming edges in the new chain $\mathcal{M}'$. All other vertices could have outgoing edges but no incoming edges. Therefore, the hitting time to s of any of these vertices is at least the minimum hitting time to s over all representatives. Therefore, it suffices to bound the minimum hitting time to s for just representative vertices. We will ignore non-representative vertices in M'.

Lemma 13 *There is no transition in $\mathcal{M}'$ from any vertex $t(i)$ to another vertex $t(j)$, $j \neq i$.*

Step 2. Next, we obtain a new chain $\mathcal{M}''$ from $\mathcal{M}'$ by removing the vertices $t(i)$, $i \geq 0$, and modifying the transition probabilities as follows. This is also a hitting time decreasing transformation. By Lemma 13, a vertex $t(i) \in \mathcal{M}'$, $i > 0$ has transitions only to itself and to the $b()$s; let $st(i)$ be that vertex $b(j)$ which has the least hitting time to s amongst all $b()$s to which $t(i)$ has an edge in $\mathcal{M}'$. For $t(0)$, which has an edge only to s, define $st(0)$ to be s.

The vertices of $\mathcal{M}''$ comprise of the vertices $b(j) \in \mathcal{M}'$ and the sink vertex s. For each pair of vertices $w, w' \in \mathcal{M}''$, we set $p_{\mathcal{M}''}(w, w') = p_{\mathcal{M}'}(w, w') + \sum_{i | st(i) = w'} p_{\mathcal{M}'}(w, t(i))$.

Lemma 14 *For each $w \in \mathcal{M}''$, $h_{\mathcal{M}''}(w, s) \leq h_{\mathcal{M}'}(w, s)$. Each vertex in $\mathcal{M}''$ has a path to s.*

The vertices in $\mathcal{M}''$ are $s, b(1), b(2), \cdots, b(2n)$. We will rename these $c(0), c(1), c(2), \cdots c(2n)$.

Step 3. We will transform $\mathcal{M}''$ to a new Markov chain $\mathcal{K}$ using a hitting time decreasing transformation. Let $c(n')$ be one of the vertices in $c((\alpha n)/174 + 1), \ldots, c((\alpha n)/174 + 6)$, whichever has the smallest hitting time to $c(0)$ in $\mathcal{M}''$ (break ties arbitrarily). The transitions in $\mathcal{K}$ are obtained from the transitions in $\mathcal{M}''$ as follows. For each vertex a in $\mathcal{M}''$, $p_{\mathcal{K}}(a, c(i)) = 0$ if $i \geq (\alpha n)/174 + 1$ and $i \neq n'$, and $p_{\mathcal{K}}(a, c(n')) = \sum_{i > (\alpha n)/174} p_{\mathcal{M}''}(a, c(i))$. All other transitions in $\mathcal{M}''$ are retained as such in $\mathcal{K}$.

Lemma 15 *For all $i > 0$, $h_{\mathcal{K}}(c(i), c(0)) \leq h_{\mathcal{M}''}(c(i), c(0))$. Each vertex in $\mathcal{K}$ has a path to s.*

Note that no $c(i)$ such that $i > (\alpha n)/174$, $i \neq n'$, has any incoming transition. Thus, to lower bound the hitting times in $\mathcal{K}$, it suffices to consider only the vertices $c(0), \ldots, c(\frac{\alpha n}{174}), c(n')$. We will rename these vertices $d(0), \ldots, d(n'')$ and ignore the remaining vertices. Note that $d(n'')$ has transitions only to the left or to itself. Also note that $\mathcal{K}$ has the following properties.

Lemma 16 *1. Each vertex $d(i) \in \mathcal{K}$, $0 \leq i \leq n''$, has transitions only to $d(j)$, where $i - 6 \leq j \leq i + 8$.*

2. $\sum_{j < i} p_{\mathcal{K}}(d(i), d(j)) \leq \frac{3i+4}{2m}$, for all $j < n''$, $i \leq n''$.

3. $\sum_{j > i} p_{\mathcal{K}}(d(i), d(j)) \geq (\alpha n/2 - i - 2)/2m$, for all $i < n''$, $j \leq n''$.

It now remains to to lower bound the hitting times in $\mathcal{K}$. This is straightforward. Details of this and proofs of the above lemmas will appear in the full version.

7 Conclusions

We have shown that Markovian coupling is not powerful enough to prove rapid mixing of the Jerrum-Sinclair chain on certain graphs. Curiously, our argument breaks down for Broder's original chain [2], due to the one-sided nature of its moves (the edge swapped in is always incident on the unmatched vertex in V_1, the left side of G). It would be interesting to see whether our result can be extended to Broder's chain as well.

8 Acknowledgements

References

[1] D. Aldous, Random walks on finite groups and rapidly mixing Markov chains, *Séminaire de Probabilités XVII 1981/82*, Springer Lecture Notes in Mathematics 986, 1986, pp. 243-297.

[2] A.Z. Broder, How hard is it to marry at random? (On the approximation of the permanent). *Proceedings of the 18th ACM Symposium on Theory of Computation*, 1986, pp.50-58.

[3] R. Bubley and M. Dyer, Path coupling: a technique for proving rapid mixing in Markov chains, *Proceedings of the 38th Annual IEEE Symposium on Foundations of Computer Science*, 1997, pp.223-231.

[4] R. Bubley and M. Dyer, Faster random generation of linear extensions, *Proceedings of the 9th ACM-SIAM Symposium on Discrete Algorithms*, 1998, pp.350-354.

[5] R. Bubley, M. Dyer and M. Jerrum, An elementary of a procedure for sampling points in a convex body. *Random Structures and Algorithms*, 12, 1998, pp. 213-235.

[6] K. Burdzy and W. Kendall, Efficient Markovian couplings: examples and counter examples. Research Report 331, University of Warwick, Dept. of Statistics, 1998.

[7] A. Czumaj, P. Kanarek, M. Kutylowski, K. Lorys, Delayed path coupling and generating random permutations via distributed stochastic processes. *Proceedings of the 10th ACM-SIAM Symposium on Discrete Algorithms*, 1999, pp. 271-280..

[8] P. Diaconis, *Group Representations in Probability and Statistics*, Institute of Mathematical Statistics, 1988.

[9] M. Dyer, A. Frieze and R. Kannan, A random polynomial time algorithm for approximating the volume of convex bodies, *Journal of the ACM*, 38, 1991, pp. 1-17.

[10] D. Griffeath, A maximal coupling for Markov chains. *Z. Wahrscheinlichkeitstheorie verw.*, Gebiete 31, 1975, pp. 95-106.

[11] M. Jerrum, A very simple algorithm for estimating the number of k-colourings of of a low degree graph, *Random Structures and Algorithms*, 7, 1995, pp. 157-165.

[12] M. Jerrum, Mathematical foundations of the Markov chain monte carlo method, *preprint*.

[13] M. Jerrum, An analysis of a Monte Carlo algorithm for estimating the permanent, *Technical Report ECS-LFCS-91-164*, Dept. of Computer Science, Univ. of Edinburg, 1991.

[14] M. Jerrum and A. Sinclair, Approximating the permanent, *SIAM Jl. on Computing* 18, 1989, pp. 1149-1178.

[15] M. Jerrum and A. Sinclair, Polynomial time approximation algorithms for the Ising model, *SIAM Jl. on Computing*, 22, 1993, pp. 1087-1116.

[16] R. Kannan, Markov chains and polynomial time algorithms, *Proceedings of the 35th Annual IEEE Symposium on Foundations of Computer Science*, 1994, pp.656-671.

[17] T. Lindvall, *Lectures on the Coupling Method*, Wiley & Sons, New York, 1992.

[18] L. Lovasz and M. Simonovits, Random walks in a convex body and an improved volume algorithm, *Random Structures and Algorithms*, 4, 1993, pp. 359-412.

[19] M. Luby and E. Vigoda, Approximately counting up to four, *Proceedings of the 29th ACM Symposium on Theory of Computation*, 1997, pp.682-687.

[20] M. Mihail, On coupling and the approximation of the permanent, *Information Processing Letters* 30, 1989, pp. 91-95.

[21] J.S. Rosenthal, Convergence rates for Markov chains, *SIAM Review* 37, 1995, pp. 387-405.

[22] A. Sinclair, *Randomised algorithms for counting and generating combinatorial structures*, Birkhäuser, 1993.