Proceedings Paper:
Dyer, ME, Jerrum, MR and Muller, H (2016) On the switch Markov chain for perfect matchings. In: Proceedings of the Twenty-Seventh Annual ACM-SIAM Symposium on Discrete Algorithms. Twenty-Seventh Annual ACM-SIAM Symposium on Discrete Algorithms, 10-12 Jan 2016, Arlington, Virginia, USA. SIAM , pp. 1972-1983. ISBN 978-1-611974-12-6

https://doi.org/10.1137/1.9781611974331.ch138

On the switch Markov chain for perfect matchings*

Martin Dyer
School of Computing
University of Leeds
Leeds, LS2 9JT, UK.
m.e.dyer@leeds.ac.uk

Mark Jerrum
School of Mathematical Sciences
Queen Mary University of London
London, E1 4NS, UK.
m.jerrum@qmul.ac.uk

Haiko Müller
School of Computing
University of Leeds
Leeds, LS2 9JT, UK.
h.muller@leeds.ac.uk

Abstract

We study a simple Markov chain, the switch chain, on the set of all perfect matchings in a bipartite graph. This Markov chain was proposed by Diaconis, Graham and Holmes as a possible approach to a sampling problem arising in Statistics. They considered several classes of graphs, and conjectured that the switch chain would mix rapidly for graphs in these classes. Here we settle their conjecture almost completely. We ask: for which graph classes is the Markov chain ergodic and for which is it rapidly mixing? We provide a precise answer to the ergodicity question and close bounds on the mixing question. We show for the first time that the mixing time of the switch chain is polynomial in the class of *monotone* graphs. This class was identified by Diaconis, Graham and Holmes as being of particular interest in the statistical setting.

1 Introduction

Counting perfect matchings in a bipartite graph, or computing the permanent of 0-1 matrix, has been one of the most central problems in Computer Science. This, and the computationally equivalent problem of sampling matchings uniformly at random, also has practical applications, in Statistics and other areas. In [8], Diaconis, Graham and Holmes considered the applications of the 0-1 permanent to Statistics, in particular where the 0-1 matrix has recognisable structure, which they called *truncated* or *interval-restricted*.

The truncated 0-1 matrices are those for which the columns can be permuted to give the *consecutive 1's* property on rows. That is, no two 1's in any row are separated by one or more 0's. Diaconis, Graham and Holmes [8] considered "one-sided" truncation, where the consecutive 1's appear at the left of each row, and "two-sided" truncation, where the consecutive 1's can appear anywhere in a row. For two-sided truncation, they considered two special cases. In the first, both the rows and columns can be permuted so that they have the consecutive 1's property. In the second, the rows and columns can be permuted so that the consecutive 1's have a "staircase" presentation, the *monotone* case, which is of particular interest in statistical applications [10].

Diaconis, Graham and Holmes [8] proposed a Markov chain for sampling perfect matchings in a bipartite graph, which we call the *switch chain*. They

*The work described here was partially supported by EPSRC grants "Computational Counting" (refs. EP/I011935/1 and EP/I012087/1) and "Randomized algorithms for computer networks" (ref. EP/M004953/1)

showed ergodicity of the chain for the truncated matrices considered in [8], and conjectured that it would converge rapidly. Computing the 0-1 permanent is a well-solved problem from a theoretical viewpoint. It is #P-complete to compute exactly [30], but there is a polynomial time approximation algorithm [16]. However, the switch chain gives a simpler and more practical algorithm than that of [16], making it worthy of consideration. Thus Diaconis, Graham and Holmes's conjecture was subsequently studied in the PhD theses of Matthews [22] and Blumberg [4], and we will discuss their results.

A 0-1 matrix is equivalently the biadjacency matrix of a bipartite graph, and we will study the graphs corresponding to the matrices considered by Diaconis, Graham and Holmes [8]. We identify the largest hereditary graph class in which the switch chain is ergodic: chordal bipartite graphs. We show that the graphs considered in [8] form an ascending sequence within this class. We examine the mixing time behaviour of the switch chain for graphs from these classes, extending work of [8], [4] and [22].

In particular, we show for the first time that the mixing time of the switch chain is polynomial on monotone graphs. This is proved by a novel application of a simple combinatorial lemma, the solution to the so-called *mountain climbing problem* [13, 18, 24, 28, 31]. Though this lemma is well known, there appears to be no worst-case analysis of this problem in the literature. We provide such an analysis in the full paper [9].

After this paper was written, we learned that Bhatnagar, Randall, Vazirani and Vigoda [3] had used a similar approach for a different problem. They analysed the Jerrum-Sinclair chain [15] for generating random bichromatic matchings in graphs that have edges partitioned into two colour classes.

For further information on Markov chains, see [1, 14, 20]. For the graph-theoretic background, see [6, 12, 27, 31].

1.1 Notation and definitions

Let $\mathbb{N} = \{1, 2, \dots\}$ denote the natural numbers. If $n \in \mathbb{N}$, let $[n] = \{1, 2, \dots, n\}$ and, if $n_1, n_2 \in \mathbb{N}$, let $[n_1, n_2] = \{n_1, n_1 + 1, \dots, n_2\}$. We will also use the notation $[n]' = \{1', 2', \dots, n'\}$ and $[n_1, n_2]' = \{n_1', (n_1+1)', \dots, n_2'\}$. Here the prime serves only to distinguish i from i'. Ordering and arithmetic for $[n]'$ elements follows that for $[n]$. Thus, for example, $1' < 2'$ and $1' + 2' = 3'$.

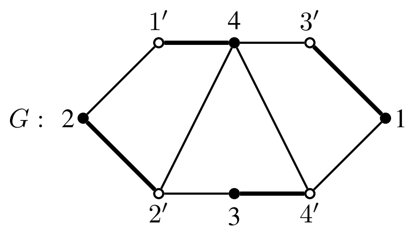
A graph $G = (V, E)$ is *bipartite* if its vertex set $V = [m] \cup [n]'$ and there is no (undirected) edge $(v, w) \in E$ such that $v, w \in [m]$ or $v, w \in [n]'$. Thus V comprises two independent sets $[m]$ and $[n]'$. Bipartite graphs $G_1 = ([m] \cup [n]', E_1)$ and $G_2 = ([m] \cup [n]', E_2)$ are *isomorphic* if there are permutations σ of $[m]$ and τ of $[n]'$ such that $(j, k') \in E_1$ if and only if $(\sigma_j, \tau_{k'}) \in E_2$. If $G = ([m] \cup [n]', E)$, we consider $[m]$ and $[n]'$ to have the usual linear ordering, and we will abuse notation by denoting these ordered sets simply by $[m]$ and $[n]'$. Then $A(G)$ denotes the $m \times n$ *biadjacency matrix* of G, with rows indexed by $[m]$ and columns by $[n]'$, such that $A(i, j') = 1$ if $(i, j') \in E$, and $A(i, j') = 0$ otherwise. The neighbourhood in G of a vertex $v \in [m] \cup [n]'$ will be denoted by $\mathcal{N}(v)$. To avoid trivialities, we will assume that G has no isolated vertices, unless explicitly stated otherwise.

A *matching* in a bipartite graph $G = ([m] \cup [n]', E)$ is a set of independent edges, that is, no two edges in the set share an endpoint. A *perfect* matching is a set of edges such that every vertex of G lies in exactly one edge. For a bipartite graph $G = ([m] \cup [n]', E)$ this requires $m = n$, and n independent edges in E. In particular, G can have no isolated vertices. We will call a bipartite graph with $m = n$ *balanced*. Equivalently, a perfect matching may be viewed as n independent 1's in the $n \times n$ 0-1 matrix $A(G)$. Thus a perfect matching M has edge set $\{(i, \pi_i') : i \in [n]\}$, where π is a permutation of $[n]$. Equivalently, M has edge set $\{(\sigma_j, j') : j \in [n]\}$, where σ is a permutation of $[n]$. Note that $\sigma = \pi^{-1}$ as elements of the symmetric group S_n. We may identify the matching M with the permutations π and σ. An example is shown in Fig. 1 below.

The total number of perfect matchings in a bipartite graph G is called the *permanent* of the matrix $A(G)$. We will denote this by per(A) when $A = A(G)$.

1.2 Computing the permanent

Valiant [30] showed that computing the permanent *exactly* is #P-complete for a general 0-1 matrix. No



$$A(G):\quad \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \end{array} \begin{array}{c} \begin{array}{cccc} 1' & 2' & 3' & 4' \end{array} \\ \left[\begin{array}{cccc} 0 & 0 & \mathbf{1} & 1 \\ 1 & \mathbf{1} & 0 & 0 \\ 0 & 1 & 0 & \mathbf{1} \\ \mathbf{1} & 1 & 1 & 1 \end{array}\right] \end{array}$$

Fig. 1: Bipartite graph with perfect matching $\sigma = (3241),\ \pi = \sigma^{-1} = (4213)$.

algorithm running in sub-exponential time is known for the exact evaluation of the permanent of 0-1 matrices.

Jerrum, Sinclair and Vigoda [16] showed that the 0-1 permanent has a *fully polynomial randomised approximation scheme* (FPRAS), using an algorithm for randomly sampling perfect matchings. This improved a Markov chain algorithm of Jerrum and Sinclair [15], which was not guaranteed to have polynomial time convergence for all bipartite graphs. The algorithm of [16] is simple, but involves polynomially many repetitions of a polynomial-length sequence of related Markov chains. The best bound known for the running time of this algorithm is $O(n^7 \log^4 n)$, due to Bezáková, Štefankovič, Vazirani and Vigoda [2].

Jerrum, Valiant and Vazirani [17] showed that sampling almost uniformly at random and approximate counting have equivalent computational complexity for many combinatorial problems, including the permanent. Technically, the problem must be *self-reducible*.

From the viewpoint of computational complexity, these results entirely settle the question of sampling and counting perfect matchings in bipartite graphs. However, simpler methods have been proposed for special cases of this problem, and here we consider one such proposal.

1.3 The switch chain

Diaconis, Graham and Holmes [8] proposed the following Markov chain for sampling perfect matchings from a balanced bipartite graph $G = ([n] \cup [n]', E)$ almost uniformly at random, which we will call the *switch chain*. A transition of the chain will be called a *switch*. Diaconis, Graham and Holmes [8] called this a “transposition”. The switch chain generalises the transposition chain for generating random permutations.

Switch chain

Let the perfect matching M_t at time t be the permutation π of $[n]$.

(1) Set $t \leftarrow 0$, and let M_0 be any perfect matching of G.

(2) Choose $i, j \in [n]$, uniformly at random, so $(i, \pi'_i),\ (j, \pi'_j) \in M_t$.

(3) If $i \neq j$ and $(i, \pi'_j), (j, \pi'_i)$ are both in E, set $M_{t+1} \leftarrow M_t \setminus \{(i, \pi'_i), (j, \pi'_j)\} \cup \{(i, \pi'_j), (j, \pi'_i)\}$.

(4) Otherwise, set $M_{t+1} \leftarrow M_t$.

(5) Set $t \leftarrow t + 1$. If $t < t_{\max}$, repeat from step (2). Otherwise, stop.

Note that the switch chain is invariant under isomorphisms of G, so properties of the chain can be investigated from the viewpoint of graph theory. An example of a switch is shown in Fig. 2, with the edges $(4, 1'),\ (2, 2')$ being switched for $(4, 2'),\ (2, 1')$.

2 Graph classes

2.1 Chordal bipartite graphs

The first question we might ask about the switch chain is: for which class of graphs is it ergodic? We wish to have a graph-theoretic answer to this question, so that we can recognise membership of graphs in the class. Therefore, we restrict attention to *hereditary* graph classes, that is, those for which all (vertex) induced subgraphs of every graph in the class are also in the class. Hereditary classes are central in modern graph theory, and are most usually characterised by describing a minimal set of “excluded subgraphs”, induced subgraphs which cannot be present.

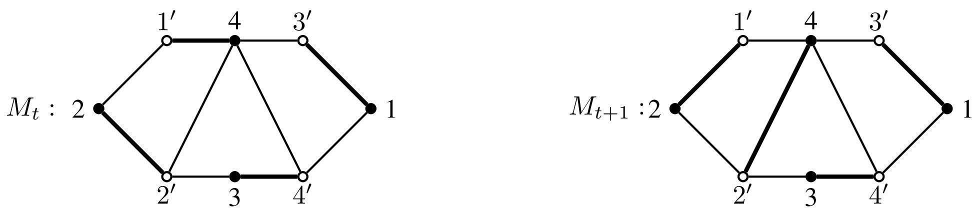


Fig. 2: A step of the switch chain

For example, *perfect graphs* are those which exclude all odd-length cycles (*odd holes*) of length at least 5, or their complements (*odd antiholes*) [7]. Thus, in particular, the class of perfect graphs contains all bipartite graphs, which exclude all odd holes and antiholes. All the graphs we consider here are bipartite, and hence perfect.

In our application, there is a further technical reason for preferring to work with hereditary graph classes. We then have self-reducibility for most problems in #P, including the permanent. This property implies the equivalence between almost uniform sampling and approximate counting referred to in Section 1.2. See [17] for details.

The switch chain is ergodic on a graph $G = (V, E)$ if the state space of the chain, the set of perfect matchings, is connected by switches. We extend this to include graphs with no perfect matching, where the state space is empty. Then we will say that a graph G is *hereditarily ergodic* if, for every $U \subseteq V$, the induced subgraph $G[U]$ is ergodic. A class of graphs will be called hereditarily ergodic if every graph in the class is hereditarily ergodic.

Diaconis, Graham and Holmes [8] observed that the switch chain is not ergodic for all bipartite graphs. They gave the example shown in Fig. 3: This graph has two perfect matchings, but the switch chain cannot move between them. This is because the graph is a chordless 6-cycle. In fact, this property characterises the class of graphs for which the switch chain is not ergodic, as we now show.

We say a graph G is *chordal bipartite* if it has no chordless cycle of length other than four. The class of chordal bipartite graphs is clearly hereditary. Note that the definition of chordal bipartite graphs is an excluded subgraph characterisation. To show that the switch chain is ergodic for this class, we require the following "excluded submatrix" characterisation.

A Γ (Gamma) in a 0-1 matrix is an induced submatrix of the form $\Gamma : \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}$. A matrix is called Γ-free if it has no such induced submatrix. Lubiw [21] gave the following characterisation.

Theorem 1 (Lubiw). *A bipartite graph is chordal bipartite if and only if it is isomorphic to a graph $G = ([m] \cup [n]', E)$ such that $A(G)$ is Γ-free. This property can be used to recognise chordal bipartite graphs in $O(|E| \log |E|)$ time.* $\square$

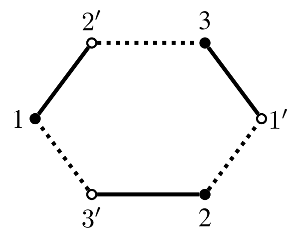
This was improved to $O(|E|)$ time by Uehara [29]. Then, for the switch chain, we have

Lemma 2. *Chordal bipartite graphs are the largest hereditary class of bipartite graphs in which the switch chain is ergodic. In this class, the diameter of the chain is at most n, for $G = ([n] \cup [n]', E)$.*

Proof. Clearly any graph with an induced cycle of length greater than 4 cannot be in the class, so we need only show ergodicity for chordal bipartite graphs. The chain is aperiodic, since there is a loop probability at least $1/n$ at each step, from choosing $i = j$ in step (2). Thus we must show that the chain is irreducible. From Theorem 1, we may suppose that $A(G)$ is a Γ-free presentation.

Let $\mathcal{G} = (\Omega, \mathcal{E})$ be the transition graph of the switch chain, with Ω the set of perfect matchings in G, and $\mathcal{E}$ the set of transitions. We must show that $\mathcal{G}$ is connected, and has diameter at most n. Let π and σ be any two perfect matchings in G, and let $\text{dist}(\pi, \sigma) = |\{i : \pi'_i \neq \sigma'_i\}|$. Note that $\text{dist}(\pi, \sigma) \leq n$, and $\text{dist}(\pi, \sigma) = 0$ implies $\pi = \sigma$.

Let k be the smallest index such that $\pi'_k \neq \sigma'_k$ and, without loss of generality, suppose $\pi'_k > \sigma'_k$. Then there exists $\ell > k$ such that $\pi'_\ell = \sigma'_k$, and hence $\pi'_\ell \neq \sigma'_\ell$. Then we have (k, σ'_k), (k, π'_k), $(\ell, \sigma'_k) \in$



$$\begin{array}{c|ccc} & 1' & 2' & 3' \\ \hline 1 & 0 & 1 & 1 \\ 2 & 1 & 0 & 1 \\ 3 & 1 & 1 & 0 \end{array}$$

Fig. 3: A non-ergodic graph

E, $\ell > k$ and $\pi'_k > \sigma'_k$. The Γ-free property of $A(G)$ then implies $(\ell, \pi'_k) \in E$. Thus we have (k, π'_k), $(\ell, \pi'_\ell) \in \pi$ and (k, π'_ℓ), $(\ell, \pi'_k) \in E$. Therefore $\tau \in \Omega$ and $(\pi, \tau) \in \mathcal{E}$, where $\tau = \pi \setminus \{(k, \pi'_k), (\ell, \pi'_\ell)\} \cup \{(k, \pi'_\ell), (\ell, \pi'_k)\}$.

Note that $\tau'_i = \pi'_i$ for $i \neq k, \ell$. However, $\pi'_k \neq \sigma'_k$, but $\tau'_k = \pi'_\ell = \sigma'_k$. Also $\pi'_\ell \neq \sigma'_\ell$, but $\tau'_\ell = \pi'_k = \sigma'_\ell$ if $\pi'_k = \sigma'_\ell$. Thus $\text{dist}(\pi, \sigma) - 2 \leq \text{dist}(\tau, \sigma) \leq \text{dist}(\pi, \sigma) - 1$. Hence there is a path of at most n edges in $\mathcal{G}$ between any pair of matchings π, σ. Therefore the diameter of $\mathcal{G}$ is at most n. □

Computing the permanent exactly is #P-complete for the class of chordal bipartite graphs [23], though this result does not extend even to chordal bipartite graphs of bounded degree. The complexity of exact computation of the permanent is unknown for all the subclasses of chordal bipartite graphs considered below, with the exception of *chain graphs*, which we discuss in Section 2.5.

2.2 Convex graphs

The largest class of graphs considered by Diaconis, Graham and Holmes [8] were those with "two-sided restrictions". These are bipartite graphs G for which $A(G)$ has the *consecutive 1's* property. These have been called *convex* graphs in the graph theory literature. They were introduced by Glover [11], who gave a simple greedy algorithm for finding a maximum matching in such a graph. The consecutive 1's property can be recognised in $O(|E|)$ time by, for example, the well-known algorithm of Booth and Lueker [5].

A bipartite graph is *convex* if it is isomorphic to a graph $G = ([m] \cup [n]', E)$ such that $\mathcal{N}(i)$ is an interval $[\alpha'_i, \beta'_i] \subseteq [n]'$ for all $i \in [m]$. Note that this property remains true under an arbitrary permutation of $[m]$. Then, it is easy to show that

Lemma 3. *Convex graphs are a proper hereditary subclass of chordal bipartite graphs.* □

2.3 Biconvex graphs

Diaconis, Graham and Holmes [8] considered the following subclass of convex graphs. A graph $G = ([m] \cup [n]', E)$ is *biconvex* if it is convex and $\mathcal{N}(j')$ is an interval $[\alpha_{j'}, \beta_{j'}] \subseteq [n]$ for all $j' \in [n]'$. Thus $A(G)$ has the consecutive 1's property for both rows and columns.

Lemma 4. *Biconvex graphs are a proper hereditary subclass of convex graphs.* □

Thus we know that the switch chain converges eventually on biconvex graphs, but how quickly is this guaranteed to occur? Unfortunately, the convergence may be exponentially slow. Both Matthews [22] and Blumberg [4] gave the following examples $\mathcal{G}_k = \big([n] \cup [n]', \mathcal{E}_k\big)$, where $n = 2k - 1$:

$$(i, j') \in \mathcal{E}_k \iff \begin{cases} 1 \leq i < k, k' \leq j' \leq (k+i)'; \\ i = k, 1' \leq j' \leq n'; \\ k < i \leq n, (i-k)' \leq j' \leq k'. \end{cases}$$

We omit the proof of slow mixing here. A sketch is given in the full paper [9], but see [4] or [22] for details.

2.4 Monotone graphs

Diaconis, Graham and Holmes [8] considered a structured subclass of biconvex graphs, which they called *monotone*, and showed that the switch chain is ergodic on monotone graphs. However, note that Lemma 2 gives a stronger result for the larger class of chordal bipartite graphs. Diaconis, Graham and Holmes [8] conjectured that the switch chain mixes rapidly in the class MONOTONE. An example is shown in Fig. 4.

A bipartite graph $G = ([m] \cup [n]', E)$ will be called *monotone* if it is isomorphic to a convex graph such that $\alpha'_i \le \alpha'_j$ and $\beta'_i \le \beta'_j$ for all $i, j \in [m]$ with $i < j$. Thus $A(G)$ has a "staircase" presentation, and we assume that G is labelled accordingly. First, we can show that, if G is row-monotone, it is also column-monotone.

Lemma 5. *A monotone graph is biconvex, and $\alpha_{i'} \le \alpha_{j'}$, $\beta_{i'} \le \beta_{j'}$ if $i', j' \in [n]'$ and $i' < j'$.* □

Next we give a "forbidden submatrix" characterisation of monotone graphs.

Lemma 6. *A bipartite graph is monotone if and only if it is isomorphic to a graph G such that $A(G)$ has none of the following as an induced 2×2 submatrix:*

$$\ulcorner \text{ (Gamma)} : \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}, \quad \lrcorner \text{ (backwards L)} : \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix},$$

$$/ \text{ (slash)} : \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}.$$

□

We note that Le [19] has independently given a somewhat tighter result. Using Lemma 7 below, [19, Cor. 1] implies that that only the submatrices $\ulcorner$ and $\lrcorner$ need be forbidden to obtain the class of monotone graphs.

A *bipartite permutation graph* is a *permutation graph* which is also bipartite. A graph $G = (V, E)$ is a permutation graph if there are permutations π, σ of V so that $(\pi_i, \pi_j) \in E$ if and only if $\pi_i < \pi_j$ and $\sigma_i > \sigma_j$. This can be given a *intersection* representation, where π, σ are points on two parallels and, for all $v \in V$, $v \in \pi$ is connected by a line to $v \in \sigma$. Then $(v, w) \in E$ if and only if corresponding lines (v, v) and (w, w) cross. Spinrad, Brandstädt and Stewart [26] studied this class of graphs, and gave $O(|E|)$ time algorithms for recognising membership in the class, and for constructing the intersection representation. Our reason for introducing this class of graphs is that the bipartite permutation graphs are precisely the monotone graphs.

Lemma 7. *A graph is monotone if and only if it is a bipartite permutation graph.*

Proof. The condition of Lemma 6 is equivalent to the following. If $(i, k'), (j, \ell') \in E$ with $i < j$ and $k' > \ell'$, then $(i, \ell'), (j, k') \in E$. The conclusion now follows from the characterisation of bipartite permutation graphs given in [26], in particular Definition 3 and Theorem 1. □

2.5 Chain graphs

Diaconis, Graham and Holmes called the simplest class of graphs they considered "one-sided restriction" graphs. These are usually called *chain graphs* in the graph theory literature [32], and form a proper subclass of monotone graphs. An example is shown in Fig. 5.

A chain graph is isomorphic to a graph $G = ([m] \cup [n]', E)$ where $\mathcal{N}(i) = [a_i]'$ for all $i \in [m]$, with $a_1 \le a_2 \le \cdots \le a_m$. Hence chain graphs are a proper hereditary subclass of monotone graphs, given by taking $\alpha'_i = 1'$, $\beta'_i = a'_i$, for all $i \in [n]$. It is then easy to show that $\mathcal{N}(j') = [b_j, m]$ for all $j' \in [n]'$, with $b_1 \ge b_2 \ge \cdots \ge b_n$. Diaconis, Graham and Holmes [8] observed that there is a "classical" explicit formula for the permanent of a chain graph G. Thus the permanent can can be evaluated exactly in FP for chain graphs, in fact in $O(n)$ time. This is easily proved, but we omit the details here.

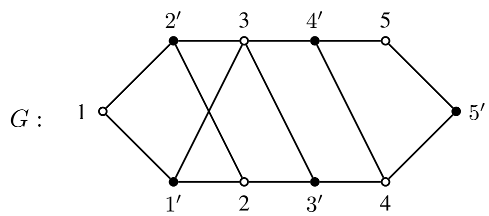
Regarding the switch chain, Matthews [22] showed, using a coupling argument, that the mixing time for chain graphs is bounded by $O(n^3 \log n)$.

3 Analysis of the switch chain

In Section 2 we have shown that the hereditary graph classes considered by Diaconis, Graham and Holmes [8] form an ascending sequence:

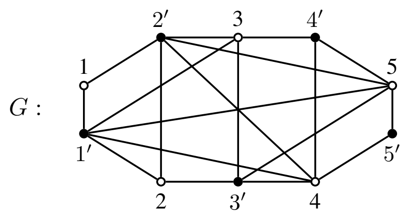
$$\text{CHAIN} \subset \text{MONOTONE} \subset \text{BICONVEX} \subset \text{CONVEX} \subset \text{CHORDAL BIPARTITE}.$$

We know from Lemma 2 that the switch chain is ergodic for bipartite graphs in all these classes. We have observed that the switch chain may have exponential mixing time in the class BICONVEX, and the switch chain has mixing time $O(n^3 \log n)$ in the class CHAIN [22]. Therefore we need only determine whether or not the class MONOTONE has polynomial mixing time. The remainder of this section will be devoted to showing that this class does indeed have



$$A(G): \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array} \begin{array}{c} \begin{array}{ccccc} 1' & 2' & 3' & 4' & 5' \end{array} \\ \left[\begin{array}{ccccc} 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 \end{array}\right] \end{array}$$

Fig. 4: A monotone graph



$$A(G): \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array} \begin{array}{c} \begin{array}{ccccc} 1' & 2' & 3' & 4' & 5' \end{array} \\ \left[\begin{array}{ccccc} 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \end{array}\right] \end{array}$$

Fig. 5: A chain graph

polynomial mixing time.

For an ergodic Markov chain on state space Ω, transition probabilities $\mathcal{P} : \Omega^2 \to [0,1]$ and stationary distribution π, the variation distance at time t, starting in state $x \in \Omega$, is $\Delta_x(t) = \max_{A\subset\Omega} |\mathcal{P}^t(x,A) - \pi(A)|$. The *mixing time*, $T_{\text{mix}}(\varepsilon) = \max_{x\in\Omega} \min\{t : \Delta_x(t) \le \varepsilon\}$, is the first time that the variation distance falls below ε, maximised over the starting state x. It is usual to measure this as a function of problem size n, suppressing the dependence on ε, which may be set to some conventional value, usually $\varepsilon = 1/e$.

3.1 Canonical paths and flows

Suppose the problem size is n. The method requires constructing paths of transitions of the chain $X = Z_1 \to Z_2 \to \cdots \to Z_\ell = Y$, between each pair of states X and Y in the state space Ω, so that the length of each path is at most polynomial in n. This is often easy to achieve, but to obtain a good upper bound on mixing time it is essential that the paths are "spread out" over the state space, and do not overload any particular transition. The degree of success in achieving this end is measured by the congestion of the set of paths. Denote the (canonical) path from X to Y by γ_{XY}. Then the congestion ϱ of the chosen paths is

$$\varrho = \max_{(Z,Z^\dagger)} \frac{\sum_{X,Y:\gamma_{XY}\ni(Z,Z^\dagger)} \pi(X)\pi(Y)\,|\gamma_{XY}|}{\pi(Z)\mathcal{P}(Z,Z^\dagger)}, \tag{1}$$

where π is the stationary distribution of the chain, $|\gamma_{XY}|$ is the length of the path γ_{XY}, the maximisation is over all transitions $Z \to Z^\dagger$, i.e., all pairs $Z^\dagger \neq Z$ with $\mathcal{P}(Z,Z^\dagger) > 0$, and the sum is over all paths that use the transition $Z \to Z^\dagger$. A precise relationship between congestion and mixing time is given in [14, Cor. 5.9], based on Sinclair [25].

Lemma 8. *Let $T_{\text{mix}}(\varepsilon)$ be the mixing time of a symmetric Markov chain $\mathcal{M}$. If ϱ is the congestion with respect to any set of canonical paths, then $T_{\text{mix}}(\varepsilon) \le 2\varrho(\ln|\Omega| + 2\ln\varepsilon^{-1})$.* □

Thus a bound on mixing time follows from a bound on congestion. When π is uniform on Ω, this can be done by ensuring that the number of canonical paths through any transition is bounded by a "small" multiple of $|\Omega|$. Thus a strategy for obtaining a bound on congestion is the following. Fix a transition $Z \to Z^\dagger$. For every canonical path γ_{XY} from X to Y that uses transition $(Z, Z^\dagger)$, specify an *encoding* $W \in \Omega$, such that, given W and g additional bits of information, we can identify X and Y uniquely. Then, if $|\gamma_{XY}| \le \ell$, for all X, Y, and $\mathcal{P}(Z,Z^\dagger) \ge \mu$, for all $Z, Z^\dagger$, we have $\rho \le 2^g\ell/\mu$.

3.2 Construction of canonical paths

Our goal is to construct canonical paths between arbitrary pairs X, Y of perfect matchings in G. In general, $(V, X \cup Y)$ is a subgraph of G, composed of alternating cycles $C_1 \cup \cdots \cup C_s = X \oplus Y$, and isolated edges $X \cap Y$. These cycles are ordered deterministically in some way, for example, according to the smallest unprimed vertex in each cycle. Then we switch each of the cycles in order, using the procedure we described below. The isolated edges are left untouched. Thus, it is sufficient to construct the canonical path for a single alternating cycle.

In fact, we may specialise the canonical path construction even further. Since MONOTONE is a hereditary class, if H is any alternating cycle in G, it is a Hamilton cycle in a smaller monotone graph $G[V(H)]$. Thus we assume that $G[V(H)] = G$ in the remainder of this section, and let H be the (Hamilton) cycle with vertices $(u_1, v_1', u_2, v_2', \ldots, u_n, v_n')$. So our initial and final matchings are $X = \{(u_1, v_1'), (u_2, v_2'), \ldots, (u_n, v_n')\}$ and $Y = \{(u_2, v_1'), (u_3, v_2'), \ldots, (u_1, v_n')\}$ where we choose $u_1 = n$ as the initial vertex of the cycle.

With each pair $(u, v') \in [n] \times [n]'$, we associate a point $p = (v, n - u + 1)$ in $\mathbb{R}^2$. In particular, the points $\{p_i = (v_i, n - u_i + 1) : i \in [n]\}$ represent the edges in X, and $\{q_i = (v_i, n - u_{i+1} + 1) : i \in [n]\}$ represent those in Y, interpreting u_{n+1} as u_1. This mapping assigns Cartesian coordinates to the entries of $A(G)$ such that the x coordinate increases with increasing column number, and the y-coordinate decreases with increasing row number. Denote the x- and y-coordinates of point $p \in \mathbb{R}^2$ by $x(p)$ and $y(p)$, so $p = (x(p), y(p))$.

Let $P = \{p_1, p_2, \ldots, p_n\} \cup \{q_1, q_2, \ldots, q_n\} \subset [n]^2$. The alternating (Hamilton) cycle $X \cup Y$ corresponds to the cyclic sequence $(p_1, q_1, p_2, q_2, \ldots, p_n, q_n)$. Join adjacent points in this sequence by line segments, omitting (q_n, p_1), to yield a continuous path Π from p_1 to q_n. This path consists of alternating horizontal and vertical segments. By the choice $u_1 = n$ for the initial vertex, we have $y(p_1) = y(q_n) = 1$, i.e., that the path begins and ends at the lowest point, in the final row of the matrix. The path reaches the highest point at $x(q_k) = x(p_{k+1}) = n$, in the first row of the matrix.

The following lemma, inspired by the "mountain climbing problem" (see, e.g. [28]) is proved in the full version of this paper [9].

Lemma 9. *Suppose* Π *is as above. There are continuous functions* $\alpha, \beta : [0,1] \to \Pi$ *satisfying* $\alpha(0) = p_1$*,* $\alpha(1) = q_k$*,* $\beta(0) = q_n$*,* $\beta(1) = p_{k+1}$*, and* $y(\alpha(t)) = y(\beta(t))$ *for all* $t \in [0,1]$*. Moreover the set of events* $T = \big\{t \in [0,1] : \alpha(t) \in P$ *or* $\beta(t) \in P\big\}$ *has cardinality at most* n^2*.* □

The trajectories of $\alpha(t)$ and $\beta(t)$ do not generally move uniformly along Π It may be necessary for either or both of $\alpha(t)$ and $\beta(t)$ to retreat along Π in order to make progress later.

We regard the points points P, and the path Π, as being contained in the *board* $[n]^2$, on which we move n *tokens*. Movements of the tokens correspond to switches on the path from X to Y. The tokens are initially on $p_1, p_2, \ldots, p_n$, representing the matching X. We move these n tokens to $q_1, q_2, \ldots, q_n$, representing Y, in a manner consistent with switches in the graph. At each step, we relocate two tokens. These two tokens are the endpoints of the diagonal of some axis-aligned rectangle, say R. We switch these tokens to the endpoints of the opposite diagonal of R. For this to be a valid switch in the graph, the new locations must correspond to 1's in the matrix $A(G)$. Ensuring that this at every step requires the tokens to be moved in a particular order.

As t increases from from 0 to 1, the *foci* $\alpha(t)$ and $\beta(t)$ move along Π. We move tokens in the neighbourhood of the foci according to certain rules. (See Fig. 6 for an example.) If we remove $\alpha(t)$ and $\beta(t)$ from Π, we separate it into three connected pieces. Denote the points in P lying in the middle piece by $P_U = P_U(t)$ and the remaining points by $P_L = P_L(t)$. Note that $P_L(t) \cup P_U(t) = P$, except at events t, when $\alpha(t) \in P$ or $\beta(t) \in P$. We ensure that tokens in P_U are in their original locations (i.e., at points of the form p_i), while those in P_L are in their final locations (i.e., at points of the form q_i). When $t = 1$, $P_u = P$, and all tokens are at their final location.

The arrangement of tokens on the board at time t (viewed as a subset of $[n]^2$) will be called the *con-*

figuration, and denoted $\sigma = \sigma(t)$. Since σ should correspond to a perfect matching in G, we insist that it contains one point from every row, $x \in [n]$, and column, $y \in [n]$, of the board. The basic underlying strategy is to keep the tokens on the points P as far as possible. If this can done effectively, we can construct an encoding of the current state (as described in Section 3.1), by forming a perfect matching W from the points of P that are not in the current configuration σ.

As the foci move, there are time periods (open t-intervals) when both $\alpha(t)$ and $\beta(t)$ are on (open) vertical segments. During these periods $y(\alpha(t))$ and $y(\beta(t))$ are either both increasing or both decreasing. Call these *v-periods*. They are separated by *h-periods*, during which one of $\alpha(t)$ or $\beta(t)$ is stationary and the other moves horizontally. During v-periods, $\sigma(t)$ is constant and well defined. We will not examine configurations during h-periods, so the definition there is unimportant. During a v-period, $\alpha(t), \beta(t) \notin P$, so $P_L(t) \cup P_U(t) = P$, and all points of P are assigned to $P_L(t)$ or $P_U(t)$. So assume that $\alpha(t)$ and $\beta(t)$ are both on (open) vertical segments.

For convenience we use a local labelling around $\alpha(t)$ and $\beta(t)$. Let a_1 and a_2 be the lower and upper ends of the line segment containing $\alpha(t)$. Continue the labelling $\ldots, a_0, a_1, a_2, a_3, \ldots$ along Π as far as neeeded. This is a local labelling of some subsequence of $p_1, q_1, \ldots, p_n, q_n$. Similarly, we label the points around $\beta(t)$. So b_1 and b_2 are the lower and upper ends of the line segment containing $\beta(t)$.

If $\sigma \subset [n]^2$ is a configuration of tokens, a *hole-pair* is a pair H of adjacent points $H = \{p_i, q_i\}$ or $H = \{q_i, p_{i+1}\}$ of P such that $\sigma \cap H = \emptyset$. As the foci move, we maintain the following Invariant I:

- I1 $\{a_1, a_2\}$ is a hole-pair.
- I2 If $x(a_1) < x(b_1)$ then $\{b_2, b_3\}$ is a hole-pair; otherwise $\{b_0, b_1\}$ is a hole-pair.
- I3 The are no hole-pairs beyond these two.

A number of consequences follow from I1–I3: (i) $\sigma(t)$ is completely determined by $\alpha(t)$ and $\beta(t)$, (ii) $|\sigma \cap P| = n - 1$, (iii) $\sigma \cap P_L \subseteq \{q_1, q_2, \ldots, q_n\}$, and (iv) $\sigma \cap P_U \subseteq \{p_1, p_2, \ldots, p_n\}$. (Working around Π from a hole-pair, successive points in P must be alternately in and out of σ, demonstrating (i). The other conclusions follow from this argument.) Invariant I may fail after a token-switch, but when this happens it will be reinstated at the following switch. There may be a single token (and exceptionally three) lying outside P (i.e., $|\sigma \setminus P| = 1$ or $|\sigma \setminus P| = 3$). Such tokens are called *dislocations*, and denoted by d (or d' or d''). See Fig. 6 for an example.

Initially $\sigma = \{p_1, \ldots, p_n\}$ so Invariant I is not satisfied. A similar remark applies to the final configuration $\sigma = \{q_1, \ldots, q_n\}$. We will see how to finesse this issue later. Let us assume that Invariant I is satisfied, and that t is in a v-period, so that $\alpha(t)$ and $\beta(t)$ move upwards on vertical line segments (a_1, a_2) and (b_1, b_2). The situation when $\alpha(t)$ and $\beta(t)$ move downwards can be handled by symmetry. Depending on the ordering of $y(a_2)$ and $y(b_2)$ one of two events occurs first: either $y(\alpha(t)) = y(a_2)$ (an α-event) or $y(\beta(t)) = y(b_2)$ (a β-event). We consider the situation just before and just after the event, and what action must be taken to maintain the invariant.

The proof now proceeds by case analysis. There are eight cases I–IV and I*–IV*, and these are exhaustive. First we split on whether $x(a_1) < x(b_1)$ (Cases I and II) or $x(a_1) > x(b_1)$ (Cases III and IV). Then we split on whether $y(a_2) < y(b_2)$ (Cases I and III) or $y(a_2) > y(b_2)$ (Cases II and IV). Finally we split on whether $\alpha(t)$ and $\beta(t)$ have the same horizontal relationship after the event (unstarred cases) or opposite (starred cases). For space reasons we will consider only Case I. For the remaining cases see the full paper [9].

Case I. This case is $x(a_1) < x(b_1)$, $y(a_2) < y(b_2)$ and $x(a_3) < x(b_1)$, see Fig. 6. In this figure, the dotted-and-dashed line is $y = y(\alpha(t)) \; (= y(\beta(t)))$, the current y-coordinate of the foci. Points with tokens are grey and points without tokens are white. The configuration in the left diagram is changed to that in the right. The current location of the dislocation is d. Note that d is always the intersection of the vertical line through a_1 and a_2, and either the horizontal line through b_0 and b_1, or through b_2 and b_3, since the board contains a token on every horizontal and vertical line.

We switch the tokens at a_3 and d. This means moving the tokens to the endpoints of the opposite diagonal of an axis-aligned rectangle with diagonal (a_3, d). In this case, the tokens move to a_2 and a new

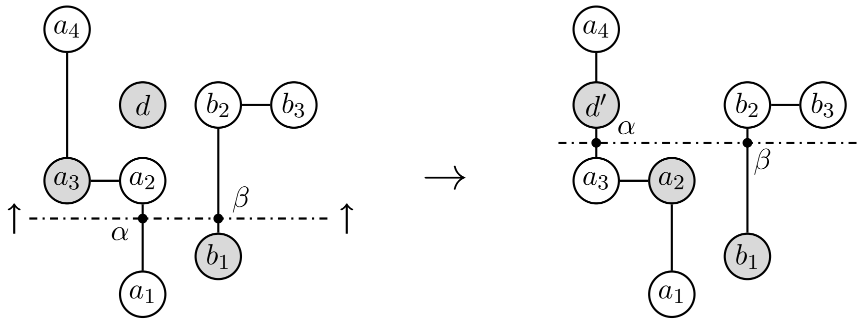


Fig. 6: Case I: $x(a_1) < x(b_1)$, $y(a_2) < y(b_2)$ and $x(a_3) < x(b_1)$

dislocation d'. We must check that d' corresponds to a 1 in $A(G)$. Here use the fact that G is monotone. We see that d' is above a_3 and left of b_2. (Note that the position of a_4 is purely diagrammatic, and does not imply this.) Since a_3 and b_2 correspond to 1's in the matrix, so does d'. It is now easy to check that I1–I3 have been preserved.

We noted that the initial and final configurations do not satisfy the invariant. So, to start the procedure, we create one horizontal and one vertical hole-pair. We do this by adding three "virtual points" p'_1, p_{n+1} and q_{n+1}. Suppose the coordinates of p_1 are $(h, 1)$. Delete p_1 and add $p'_1 = (h, 0)$, $p_{n+1} = (n+1, 1)$ and $q_{n+1} = (n+1, 0)$. In G, this corresponds to adding new vertices $n+1$ and $(n+1)'$ and new edges $(n+1, h')$, $(n, (n+1)')$ and $(n+1, (n+1)')$, together with any others needed to preserve monotonicity. We add a token to q_{n+1}, leaving the existing tokens in place. The token at p_1 now becomes dislocation d. Place α and β just below d and p_{n+1}. The invariant is satisfied, and we can now start the canonical path construction as described earlier. A similar construction can be used to finish the path.

3.3 Encoding and congestion

In Lemma 9, reverse the role of α and β, so that $\alpha(0) = q_n$, $\alpha(1) = p_{k+1}$, $\beta(0) = p_1$ and $\beta(1) = q_k$. Place the n tokens initially on the points $\{q_1, \ldots, q_n\}$, and denote the configuration at time t by $\sigma'(t)$. Since the trajectories of $\alpha(t)$ and $\beta(t)$ are oblivious of the tokens, $P_L = P_L(t)$ and $P_U = P_U(t)$ are unchanged. According to the invariant, the configuration σ' satisfies $\sigma' \cap P_U \subseteq \{q_1, \ldots, q_n\}$, $\sigma' \cap P_L \subseteq \{p_1, \ldots, p_n\}$ and $|\sigma' \cap P| = n - 1$. At any legal time t, then $\big|(\sigma(t) \cup \sigma'(t)) \cap P\big| = 2n - 2$. Consider a canonical path $X = Z_0 \to \cdots \to Z_\ell = Y$ constructed as in Section 3.2. Some of cases involve two switches. Then we call the configuration between the two switches (and the corresponding perfect matching Z_i) *transitory*. Non-transitory configurations are of the form $\sigma(t)$ for some t, and these configurations satisfy Invariant I. If Z_i is not transitory, consider a time t at which configuration $\sigma(t)$ corresponds to Z_i. Then $\sigma'(t)$ is a near complement to $\sigma(t)$ with respect to P, and its corresponding perfect matching Z'_i is a near complement to Z_i with respect to $X \cup Y$. Specifically, since $\big|(\sigma(t) \cup \sigma'(t)) \cap P\big| = 2n - 2$, we have $\big|(Z_i \cup Z'_i) \cap (X \cup Y)\big| = 2n - 2$.

If $(Z, Z^\dagger)$ is a transition of the switch chain, we wish to provide each canonical path through $(Z, Z^\dagger)$ with a unique encoding. In fact, our encoding will be an element of $\Omega \times [4n^2]$. Suppose the transition is (Z_i, Z_{i+1}) on a canonical path from X to Y. We can suppose that $C = X \cup Y$ is a single cycle. At least one of Z_i and Z_{i+1} is not transitory, say, Z_i. Our encoding will be the Z'_i above, with some additional data. Since $C' = Z_i \cup Z'_i$, we provide the identity of the edges in $C' \setminus C$: there are at most $2n^2$ possibilities. Now we must add two edges so that the result is a cycle, but there are only two ways this can be done. Finally, we need to signal Z'_i corresponds to Z_i and not Z_{i+1}: a further two possibilities. This gives us our encoding within the set $\Omega \times [8n^2]$.

We have all the quantities needed for the calculation of the congestion ϱ. From the definition of the switch chain, $\mathcal{P}(Z, Z^\dagger) = 2n^{-2}$. From Lemma 9, the maximum length of a canonical path is n^2. Substituting these values into (1) yields $\varrho \le |\Omega|^{-1}(n^2/2)(8n^2)\,|\Omega|\,n^2 = 4n^6$. Then, from

Lemma 8, noting that the state space Ω has cardinality at most $n!$, we obtain the bound on mixing time.

Theorem 10. *The switch Markov chain has mixing time* $\tau(\varepsilon) < 8n^6(n \ln n + 2 \ln \varepsilon^{-1}) = O(n^7 \log n)$ *for any graph* $G = ([n] \cup [n]', E)$ *in the class* MONOTONE.

We note that the algorithm of [2] has running time $O(n^7 \log^4 n)$, but with no bound given on the implied constant. It may be possible to improve our analysis, but it is highly unlikely that we could match the $O(n^2 \log n)$ bound conjectured by Diaconis, Graham and Holmes [8].

References

[1] D. ALDOUS AND J. A. FILL, *Reversible Markov chains and random walks on graphs*, 2002. `http://www.stat.berkeley.edu/~aldous/RWG/book.html`. Unfinished monograph, recompiled 2014.

[2] I. BEZÁKOVÁ, D. STEFANKOVIC, V. V. VAZIRANI, AND E. VIGODA, *Accelerating simulated annealing for the permanent and combinatorial counting problems*, SIAM Journal on Computing, 37 (2008), pp. 1429–1454.

[3] N. BHATNAGAR, D. RANDALL, V. VAZIRANI, AND E. VIGODA, *Random bichromatic matchings*, in LATIN 2006: Theoretical Informatics, vol. 3887 of Lecture Notes in Computer Science, Springer, 2006, pp. 190–201.

[4] O. BLUMBERG, *Permutations with interval restrictions*, PhD thesis, Stanford University, 2012.

[5] K. S. BOOTH AND G. S. LUEKER, *Testing for the consecutive ones property, interval graphs, and graph planarity using PQ-tree algorithms*, Journal of Computer and System Sciences, 13 (1976), pp. 335–379.

[6] A. BRANDSTÄDT, V. B. LE, AND J. P. SPINRAD, *Graph Classes: A Survey*, vol. 3 of Monographs on Discrete Mathematics and Applications, SIAM, Philadelphia, 1999.

[7] M. CHUDNOVSKY, N. ROBERTSON, P. SEYMOUR, AND R. THOMAS, *The strong perfect graph theorem*, Annals of Mathematics, 164 (2006), pp. 51–229.

[8] P. DIACONIS, R. GRAHAM, AND S. P. HOLMES, *Statistical problems involving permutations with restricted positions*, in State of the art in probability and statistics, vol. 36 of Lecture Notes–Monograph Series, Institute of Mathematical Statistics, 2001, pp. 195–222.

[9] M. DYER, M. JERRUM, AND H. MÜLLER, *On the switch Markov chain for perfect matchings*, 2015. `http://arxiv.org/abs/1501.07725`

[10] B. EFRON AND V. PETROSIAN, *Nonparametric methods for doubly truncated data*, Journal of the American Statistical Association, 94 (1999), pp. 824–834.

[11] F. GLOVER, *Maximum matching in a convex bipartite graph*, Naval Research Logistics Quarterly, 14 (1967), pp. 313–316.

[12] M. C. GOLUMBIC, *Algorithmic graph theory and perfect graphs*, vol. 57 of Annals of Discrete Mathematics, North-Holland, Amsterdam, 2004.

[13] J. E. GOODMAN, J. PACH, AND C. K. YAP, *Mountain climbing, ladder moving, and the ring width of a polygon*, Amer. Math. Monthly, 96 (1989), pp. 494–510.

[14] M. JERRUM, *Counting, sampling and integrating: algorithms and complexity*, ETH Zürich Lectures in Mathematics, Birkhäuser, Basel, 2003.

[15] M. JERRUM AND A. SINCLAIR, *Approximating the permanent*, SIAM Journal on Computing, 18 (1989), pp. 1149–1178.

[16] M. JERRUM, A. SINCLAIR, AND E. VIGODA, *A polynomial-time approximation algorithm for the permanent of a matrix with non-negative entries*, Journal of the ACM, 51 (2004), pp. 671–697.

[17] M. JERRUM, L. VALIANT, AND V. VAZIRANI, *Random generation of combinatorial structures from a uniform distribution*, Theoretical Computer Science, 43 (1986), pp. 169–188.

[18] T. KELETI, *The mountain climbers' problem*, Proceedings of the American Mathematical Society, 117 (1993), pp. 89–97.

[19] V. B. LE, *On opposition graphs, coalition graphs, and bipartite permutation graphs*, Discrete Applied Mathematics, 168 (2014), pp. 26–33.

[20] D. A. LEVIN, Y. PERES, AND E. L. WILMER, *Markov chains and mixing times*, American Mathematical Society, 2006.

[21] A. LUBIW, *Doubly lexical orderings of matrices*, SIAM Journal on Computing, 16 (1987), pp. 854–879.

[22] J. MATTHEWS, *Markov chains for sampling matchings*, PhD thesis, Edinburgh University, 2008.

[23] Y. OKAMOTO, R. UEHARA, AND T. UNO, *Counting the number of matchings in chordal and chordal bipartite graph classes*, in Graph-Theoretic Concepts in Computer Science, vol. 5911 of Lecture Notes in Computer Science, Springer, 2010, pp. 296–307.

[24] I. PAK, *Lectures on discrete and polyhedral geometry*, 2010. `http://www.math.ucla.edu/~pak/geompol8.pdf`

[25] A. SINCLAIR, Improved bounds for mixing rates of Markov chains and multicommodity flow, Combinatorics Probability and Computing, 1 (1992), pp. 351–370.

[26] J. SPINRAD, A. BRANDSTÄDT, AND L. STEWART, *Bipartite permutation graphs*, Discrete Applied Mathematics, 18 (1987), pp. 279 – 292.

[27] J. P. SPINRAD, *Efficient Graph Representations*, Fields Institute monographs, The Fields Institute for Research in Mathematical Sciences, American Mathematical Soc., 2003.

[28] A. TUCKER, *The parallel climbers puzzle*, Math Horizons (1995), pp. 22–24.

[29] R. UEHARA, *Linear time algorithms on chordal bipartite and strongly chordal graphs*, in Automata, Languages and Programming, 29th International Colloquium, 2002, pp. 993–1004.

[30] L. G. VALIANT, *The complexity of computing the permanent*, Theoretical Computer Science, 8 (1979), pp. 189–201.

[31] D. B. WEST, *Introduction to Graph Theory*, Prentice Hall, 2nd ed., 2001.

[32] M. YANNAKAKIS, *Computing the minimum fill-in is NP-complete*, SIAM J. Alg. Disc. Meth., 2 (1981), pp. 77–79.