Contents

Chapter 5

Canonical paths and matchings

Coupling, at least Markovian coupling, is not a universally applicable method for proving rapid mixing. In this chapter, we define a natural MC on matchings in a graph G and show that its mixing time is bounded by a polynomial in the size of G. Anil Kumar and Ramesh [3] studied a very similar MC to this one, and demonstrated that every Markovian coupling for it takes expected exponential time (in the size of G) to coalesce. In the light of their result, it seems we must take an alternative approach, sometimes called the "canonical paths" method.

5.1 Matchings in a graph

Consider an undirected graph $G = (V, E)$ with vertex set V of size n, and edge set E of size m. Recall that the set of edges $M \subseteq E$ is a *matching* if the edges of M are pairwise vertex disjoint. The vertices that occur as endpoints of edges of M are said to be *covered* by M; the remaining vertices are *uncovered*. For a given graph $G = (V, E)$, we are interested in sampling from the set of matchings of G according to the distribution

$$\pi(M) = \frac{\lambda^{|M|}}{Z} \tag{5.1}$$

where $Z := \sum_M \lambda^{|M|}$, and the sum is over matchings M of all sizes. In statistical physics, the edges in a matching are referred to as "dimers" and the uncovered vertices as "monomers." The probability distribution defined in (5.1) characterises the *monomer-dimer system* specified by G and λ. The

1. Select $e = \{u, v\} \in E$ u.a.r.
2. There are three mutually exclusive (but not exhaustive) possibilities.

   ($\uparrow$) If u and v are not covered by X_0, then $M \leftarrow X_0 \cup \{e\}$.

   ($\downarrow$) If $e \in X_0$, then $M \leftarrow X_0 \setminus \{e\}$.

   ($\leftrightarrow$) If u is uncovered and v is covered by some edge $e' \in X_0$ (or vice versa, with the roles of u and v reversed), then $M' \leftarrow M \cup \{e\} \setminus \{e'\}$.

   If none of the above situations obtain, then $M \leftarrow X_0$.
3. With probability $\min\{1, \pi(M)/\pi(X_0)\}$ set $X_1 \leftarrow M$; otherwise, set $X_1 \leftarrow X_0$. (This form of acceptance probability is known as the *Metropolis filter.*)

Figure 5.1: An MC for sampling weighted matchings

normalising factor Z is the *partition function* of the system. The parameter $\lambda \in \mathbb{R}^+$ can be chosen to either favour smaller ($\lambda < 1$) or larger ($\lambda > 1$) matchings, or to generate them from the uniform distribution ($\lambda = 1$).

Note that computing Z exactly is a hard problem. For if it could be done efficiently, one could compute $Z = Z(\lambda)$ at a sequence of distinct values of λ, and then extract the coefficients of $Z(\lambda)$ by interpolating the computed values. (Observe that $Z(\lambda)$ is a polynomial in λ.) But the highest-order coefficient is just the number of perfect matchings in G. It follows from Theorem 2.2 that evaluating $Z(\lambda)$ at (say) integer points $\lambda \in \mathbb{N}$ is #P-hard. Indeed, with a little more work, one can show that evaluating $Z(\lambda)$ at the particular point $\lambda = 1$ (i.e., counting the number of matchings in G) is #P-complete. Although it is unlikely that Z can be computed efficiently, nothing stops us from having an efficient approximation scheme, in the FPRAS sense of §3.1.

We construct an MC for sampling from distribution (5.1) as shown in Figure 5.1. As usual, denote the state space of the MC by Ω, and its transition matrix by P. Consider two adjacent matchings M and M' with $\pi(M) \leq \pi(M')$. By *adjacent* we just mean that $P(M, M') > 0$, which is equivalent to $P(M', M) > 0$. The transition probabilities between M and

M' may be written

$$P(M, M') = \frac{1}{m}, \quad \text{and}$$
$$P(M', M) = \frac{1}{m}\frac{\pi(M)}{\pi(M')},$$

giving rise to the symmetric form

$$\pi(M)P(M, M') = \pi(M')P(M'M) = \frac{1}{m}\min\left\{\pi(M), \pi(M')\right\}. \tag{5.2}$$

The above equality makes clear that the MC is time-reversible, and that its stationary distribution (appealing Lemma 3.7) is π.

Remarks 5.1. (a) The transition probabilities are easy to compute: since a transition changes the number of edges in the current matching by at most one, the acceptance probability in step 3 is either 1 or $\min\{\lambda, \lambda^{-1}\}$, and it is easy to determine which.

(b) Broder [8] was the first to suggest sampling matching by simulating an appropriate MC. His proposal was to construct an MC whose states are perfect matchings (i.e., covering all vertices) and near-perfect matchings (i.e., leaving exactly two vertices uncovered). The MC on all matchings presented in Figure 5.1 was introduced by Jerrum and Sinclair [36].

(c) Time reversibility is a property of MC's that is frequently useful to us; in particular, as we have seen on several occasions, it permits easy verification of the stationary distribution of the MC. However, we shall not make use of the property in the remainder of the chapter, and all the results will hold in the absence of time reversibility.

5.2 Canonical paths

The key to demonstrating rapid mixing using the "canonical paths" technique lies in setting up a suitable multicommodity flow problem. For any pair $x, y \in \Omega$, we imagine that we have to route $\pi(x)\pi(y)$ units of distinguishable fluid from x to y, using the transitions of the MC as "pipes." To obtain a good upper bound on mixing time we must route the flow evenly, without creating particularly congested pipes. To formalise this, we need a measure for congestion.

For any pair $x, y \in \Omega$, define a canonical path $\gamma_{xy} = (x = z_0, z_1, \dots, z_\ell = y)$ from x to y through pairs (z_i, z_{i+1}) of states adjacent in the MC, and let

$$\Gamma := \{\gamma_{xy} \mid x, y \in \Omega\}$$

be the set of all canonical paths. The *congestion* $\varrho = \varrho(\Gamma)$ of the chain is defined by

$$\varrho(\Gamma) := \max_{t=(u,v)} \Bigg\{ \underbrace{\frac{1}{\pi(u)P(u,v)}}_{(\text{capacity of } t)^{-1}} \underbrace{\sum_{x,y\colon \gamma_{xy} \text{ uses } t} \pi(x)\pi(y)\,|\gamma_{xy}|}_{\text{total flow through } t} \Bigg\}. \tag{5.3}$$

where t runs over all transitions, i.e., all pairs of adjacent states of the chain, and $|\gamma_{xy}|$ denotes the length of the path γ_{xy}.

We want to show that if ϱ is small then so is the mixing time of the MC. Consider some arbitrary "test" function $f : \Omega \to \mathbb{R}$. The variance of f (with respect to π) is

$$\operatorname{Var}_\pi f := \sum_{x\in\Omega} \pi(x)\big[f(x) - \mathrm{E}_\pi f\big]^2 = \sum_{x\in\Omega} \pi(x) f(x)^2 - (\mathrm{E}_\pi f)^2, \tag{5.4}$$

where

$$\mathrm{E}_\pi f := \sum_{x\in\Omega} \pi(x) f(x).$$

It is often convenient to work with an alternative, possibly less familiar expression for variance, namely

$$\operatorname{Var}_\pi f = \frac{1}{2} \sum_{x,y\in\Omega} \pi(x)\pi(y)\big(f(x) - f(y)\big)^2. \tag{5.5}$$

Equivalence of (5.4) and (5.5) follows from the following sequence of identities:

$$\begin{aligned}
&\frac{1}{2} \sum_{x,y\in\Omega} \pi(x)\pi(y)\big(f(x) - f(y)\big)^2 \\
&\qquad = \sum_{x,y\in\Omega} \big[\pi(x)\pi(y)f(x)^2 - \pi(x)\pi(y)f(x)f(y)\big] \\
&\qquad = \sum_{x\in\Omega} \pi(x)f(x)^2 \sum_{y\in\Omega} \pi(y) - \sum_{x\in\Omega} \pi(x)f(x) \sum_{y\in\Omega} \pi(y)f(y) \\
&\qquad = \sum_{x\in\Omega} \pi(x)f(x)^2 - (\mathrm{E}_\pi f)^2 \\
&\qquad = \operatorname{Var}_\pi f.
\end{aligned}$$

The variance $\mathrm{Var}_\pi f$ measures the "global variation" of f over Ω. By contrast, the *Dirichlet form*

$$\mathcal{E}_\pi(f,f) := \frac{1}{2} \sum_{x,y\in\Omega} \pi(x)P(x,y)\big(f(x)-f(y)\big)^2 \tag{5.6}$$

measures the "local variation" of f with respect to the transitions of the MC. The key result relating the congestion ϱ to local and global variation is the following.

Theorem 5.2 (Diaconis and Stroock [20]; Sinclair [57]). *For any function $f : \Omega \to \mathbb{R}$,*

$$\mathcal{E}_\pi(f,f) \geq \frac{1}{\varrho} \mathrm{Var}_\pi f. \tag{5.7}$$

where $\varrho = \varrho(\Gamma)$ is the congestion, defined in (5.3), with respect to any set of canonical paths Γ.

Remarks 5.3. (a) An inequality such as (5.7), which bounds the ratio of the local to the global variation of a function, is often termed a *Poincaré inequality.*

(b) If the congestion ϱ is small, then high global variation of a function entails high local variation. This in turn entails, as we shall see presently, short mixing time.

Proof of Theorem 5.2. We follow Sinclair [57, Thm. 5].

$$\begin{aligned}
2\,\mathrm{Var}_\pi f &= \sum_{x,y\in\Omega} \pi(x)\pi(y)\big(f(x)-f(y)\big)^2 \\
&= \sum_{x,y\in\Omega} \pi(x)\pi(y)\Bigg(\sum_{(u,v)\in\gamma_{xy}} 1\cdot\big(f(u)-f(v)\big)\Bigg)^2 && (5.8)\\
&\leq \sum_{x,y\in\Omega} \pi(x)\pi(y)\,|\gamma_{xy}| \sum_{(u,v)\in\gamma_{xy}} \big(f(u)-f(v)\big)^2 && (5.9)\\
&= \sum_{u,v\in\Omega} \sum_{\substack{x,y:\\(u,v)\in\gamma_{xy}}} \pi(x)\pi(y)\,|\gamma_{xy}|\,\big(f(u)-f(v)\big)^2
\end{aligned}$$

$$
\begin{aligned}
&= \sum_{u,v\in\Omega} \big(f(u)-f(v)\big)^2 \sum_{\substack{x,y:\\(u,v)\in\gamma_{xy}}} \pi(x)\pi(y)\,|\gamma_{xy}| \\
&\le \sum_{u,v\in\Omega} \big(f(u)-f(v)\big)^2\, \pi(u)P(u,v)\,\varrho \qquad (5.10)\\
&= 2\varrho\,\mathcal{E}_\pi(f,f).
\end{aligned}
$$

Equality (5.8) is a "telescoping sum," inequality (5.9) is Cauchy-Schwarz, and inequality (5.10) is from the definition of ϱ. □

For the following analysis, we modify the chain by making it "lazy." In each step, the lazy MC stays where it is with probability $\frac{1}{2}$, and otherwise makes the transition specified in Figure 5.1. Formally, the transition matrix of the lazy MC is $P_{\mathrm{zz}} := \frac{1}{2}(I+P)$, where I is the identity matrix. It is straightforward to show that the lazy MC is ergodic if the original MC is, in which case the stationary distribution of the two is identical. (In fact, irreducibility of the original MC is enough to guarantee ergodicity of the lazy MC.)

Exercise 5.4. Verify these claims about the lazy MC.

Remarks 5.5. (a) This laziness doubles the mixing time, but ensures that the eigenvalues of the transition matrix are all non-negative, and avoids possible parity conditions that would lead to the MC being periodic or nearly so. In an implementation, to simulate $2t$ steps of the lazy MC, one would generate a sample T from the binomial distribution $\mathrm{Bin}(2t,\frac{1}{2})$, and then simulate T steps of the original, non-lazy MC. Thus, in practice, efficiency would not be compromised by laziness.

(b) The introduction of the lazy chain may seem a little unnatural. At the expense of setting up a little machinery, it can be avoided by using a continuous-time MC rather than a discrete-time MC as we have done. Some other parts of our development would also become smoother in the continuous-time setting. We shall return to this point at the end of the chapter.

Before picking up the argument, some extra notation will be useful. If f is any function $f:\Omega\to\mathbb{R}$ then $P_{\mathrm{zz}}f:\Omega\to\mathbb{R}$ denotes the function defined by

$$[P_{\mathrm{zz}}f](x) := \sum_{y\in\Omega} P_{\mathrm{zz}}(x,y)f(y).$$

The function $P_{zz}f$ is the "one-step averaging" of f. Similarly, $P_{zz}^t f$, defined in an analogous way, is the "t-step averaging" of f: it specifies the averages of f over t-step evolutions of the MC, starting at each of the possible states. If the MC is ergodic (as here), then $P_{zz}^t f$ tends to the constant function $\mathrm{E}_\pi f$ as $t \to \infty$. (Observe that $\mathrm{E}_\pi(P_{zz}f) = \mathrm{E}_\pi f$ and hence $\mathrm{E}_\pi(P_{zz}^t f) = \mathrm{E}_\pi f$; in other words, t-step averaging preserves expectations.) Thus we can investigate the mixing time of the MC by seeing how quickly $\mathrm{Var}_\pi(P_{zz}^t f)$ tends to 0 as $t \to \infty$. This is the idea we shall now make rigorous.

Theorem 5.6. *For any function $f : \Omega \to \mathbb{R}$,*

$$\mathrm{Var}_\pi(P_{zz}f) \le \mathrm{Var}_\pi f - \frac{1}{2}\mathcal{E}_\pi(f,f). \tag{5.11}$$

Proof. We follow closely Mihail's [51] derivation. Consider the one-step averaging of f with respect to the lazy chain:

$$\begin{aligned}[P_{zz}f](x) &= \sum_{y\in\Omega} P_{zz}(x,y)f(y)\\ &= \frac{1}{2}f(x) + \frac{1}{2}\sum_{y\in\Omega} P(x,y)f(y)\\ &= \frac{1}{2}\sum_{y\in\Omega} P(x,y)\big(f(x)+f(y)\big). \qquad (5.12)\end{aligned}$$

For convenience, assume[1] $\mathrm{E}_\pi f = 0$, and hence $\mathrm{E}_\pi(P_{zz}f) = 0$. Then the left-hand side of (5.11) is bounded above as follows:

$$\begin{aligned}\mathrm{Var}_\pi(P_{zz}f) &= \sum_{x\in\Omega}\pi(x)\big\{[P_{zz}f](x)\big\}^2\\ &= \frac{1}{4}\sum_{x\in\Omega}\pi(x)\left(\sum_{y\in\Omega}\sqrt{P(x,y)}\big(f(x)+f(y)\big)\cdot\sqrt{P(x,y)}\right)^2 \qquad (5.13)\\ &\le \frac{1}{4}\sum_{x\in\Omega}\pi(x)\sum_{y\in\Omega}P(x,y)\big(f(x)+f(y)\big)^2\sum_{z\in\Omega}P(x,z) \qquad (5.14)\\ &= \frac{1}{4}\sum_{x,y\in\Omega}\pi(x)P(x,y)\big(f(x)+f(y)\big)^2, \qquad (5.15)\end{aligned}$$

[1] Otherwise add or subtract a constant, an operation that leaves unchanged the quantities of interest, namely $\mathrm{Var}_\pi f$, $\mathrm{Var}_\pi(P_{zz}f)$ and $\mathcal{E}_\pi(f,f)$.

where step (5.13) uses (5.12), and inequality (5.14) is Cauchy-Schwarz. To get at the right-hand side of (5.11) we use yet another expression for the variance of f:

$$\begin{aligned}\operatorname{Var}_\pi f &= \frac{1}{2}\sum_{x\in\Omega}\pi(x)f(x)^2 + \frac{1}{2}\sum_{y\in\Omega}\pi(y)f(y)^2 \\ &= \frac{1}{2}\sum_{x,y\in\Omega}\pi(x)f(x)^2P(x,y) + \frac{1}{2}\sum_{x,y\in\Omega}\pi(x)P(x,y)f(y)^2 \\ &= \frac{1}{2}\sum_{x,y\in\Omega}\pi(x)P(x,y)\big(f(x)^2+f(y)^2\big). \qquad (5.16)\end{aligned}$$

Subtracting (5.15) from (5.16) yields

$$\begin{aligned}\operatorname{Var}_\pi f - \operatorname{Var}_\pi(P_{\mathrm{zz}}f) &\geq \frac{1}{4}\sum_{x,y\in\Omega}\pi(x)P(x,y)\big(f(x)-f(y)\big)^2 \\ &= \frac{1}{2}\,\mathcal{E}_\pi(f,f),\end{aligned}$$

as required. □

Combining Theorem 5.2 and Theorem 5.6 gives:

Corollary 5.7. *For any function $f : \Omega \to \mathbb{R}$,*

$$\operatorname{Var}_\pi(P_{\mathrm{zz}}f) \leq \left(1 - \frac{1}{2\varrho}\right)\operatorname{Var}_\pi f,$$

where $\varrho = \varrho(\Gamma)$ is the congestion, defined in (5.3), with respect to any set of canonical paths Γ.

Remark 5.8. The algebraic manipulation in the proof of Theorem 5.6 seems mysterious. The discussion of the continuous-time setting at the end of the chapter will hopefully clarify matters a little.

We can now use Corollary 5.7 to bound the mixing time of the chain, by using a special function f. For a subset $A \subseteq \Omega$ of the state space, we consider its indicator function

$$f(x) := \begin{cases} 1, & \text{if } x \in A; \\ 0, & \text{otherwise.}\end{cases}$$

Then we have $\operatorname{Var}_\pi f \leq 1$ and therefore

$$\operatorname{Var}_\pi(P^t_{\mathrm{zz}}f) \leq \left(1-\frac{1}{2\varrho}\right)^t \leq \exp\left\{\frac{-t}{2\varrho}\right\},$$

where $P^t_{\mathrm{zz}} f$ is the t-step averaging of f. Fix some starting state $x \in \Omega$ and set

$$t = \left\lceil 2\varrho \left(2 \ln \varepsilon^{-1} + \ln \pi(x)^{-1} \right) \right\rceil .$$

This gives

$$\mathrm{Var}_\pi(P^t_{\mathrm{zz}} f) \le \exp\left\{ -2 \ln \varepsilon^{-1} - \ln \pi(x)^{-1} \right\} = \varepsilon^2 \pi(x).$$

On the other hand,

$$\begin{aligned} \mathrm{Var}_\pi(P^t_{\mathrm{zz}}) &\ge \pi(x) \left([P^t_{\mathrm{zz}} f](x) - \mathrm{E}_\pi(P^t_{\mathrm{zz}} f) \right)^2 \\ &= \pi(x) \left([P^t_{\mathrm{zz}} f](x) - \mathrm{E}_\pi\, f \right)^2, \end{aligned}$$

which implies

$$\varepsilon \ge \left| [P^t_{\mathrm{zz}} f](x) - \mathrm{E}_\pi\, f \right| = \left| P^t_{\mathrm{zz}}(x, A) - \pi(A) \right|$$

for all A. This in turn means that the total variation distance $D_{\mathrm{tv}}(P^t_{\mathrm{zz}}(x, \cdot\,), \pi)$ is bounded by ε, and we obtain the following corollary:

Corollary 5.9. *The mixing time of the lazy MC is bounded by*

$$\tau_x(\varepsilon) \le 2\varrho \left(2 \ln \varepsilon^{-1} + \ln \pi(x)^{-1} \right),$$

where $\varrho = \varrho(\Gamma)$ is the congestion, defined in (5.3), with respect to any set of canonical paths Γ.

Remark 5.10. The factor 2 in front of the bound on mixing time is an artifact of using the lazy MC.

5.3 Back to matchings

In the previous section, we saw how a general technique (canonical paths) can be used to bound the Poincaré constant of an MC, and how that constant in turn bounds the mixing time. Let's apply this machinery to the matching chain presented in Figure 5.1. Our ultimate goal is to derive a polynomial upper bound on mixing time:

Proposition 5.11. *The mixing time τ of the MC on matchings of a graph G (refer to Figure 5.1) is bounded by*

$$\tau(\varepsilon) \le nm\bar{\lambda}^2 \big(4 \ln \varepsilon^{-1} + 2n \ln n + n\, |\ln \lambda| \big),$$

where n and m are the number of vertices and edges of G, respectively, and $\bar{\lambda} = \max\{1, \lambda\}$.

Remark 5.12. It is possible, with a little extra work, to improve the upper bound in Proposition 5.11 by a factor of $\bar{\lambda}$: see Exercise 5.17.

The first step is to define the set Γ of canonical paths. Given two matchings I (initial) and F (final), we need to connect I and F by a canonical path γ_{IF} in the adjacency graph of the matching MC. Along this path, we will have to lose or gain at least the edges in the symmetric difference $I \oplus F$; these edges define a graph of maximum degree two, which decomposes into a collection of paths and even-length cycles, each of them alternating between edges in I and edges in F. If we fix some ordering of the vertices in V, we obtain a unique ordering of the connected components of $(V, I \oplus F)$, by smallest vertex. Within each connected component we may identify a unique "start vertex": in the case of a cycle this will be the smallest vertex, and the case of a path the smaller of the two endpoints. We imagine each path to be oriented away from its start vertex, and each cycle to be oriented so that the edge in I adjacent to the start vertex acquires an orientation away from the start vertex. In Figure 5.2 — which focuses on a particular transition $t = (M, M')$ on the canonical path from I to F — the m connected components of $I \oplus F$ are denoted $P_1, \dots, P_m$.

To get from I to F, we now process the components of $(V, I \oplus F)$ in the order $P_1, \dots, P_m$. In each cycle, we first remove the edge in I incident to the start vertex using a $\downarrow$-transition; with a sequence of $\leftrightarrow$-transitions following the cycle's orientation, we then replace I- by F-edges; finally, we perform a $\uparrow$-transitions to add the edge in F incident to the start vertex. In every path, if the start vertex is incident to an F-edge, we use $\leftrightarrow$-transitions along the path and finish by a $\uparrow$-transition in case the path has odd length. If the start vertex is incident to an I-edge, we start with a $\downarrow$-transition, then use $\leftrightarrow$-transitions along the path, and finish with an $\uparrow$-transition in case the path has even length. This concludes the description of the canonical path γ_{IF}. Each transition t on a canonical path γ_{IF} can be though of a contributing to the processing of a certain connected component of $I \oplus F$; we call this the *current* component (or cycle, or path, if we want to be more specific).

Denote by

$$\mathrm{cp}(t) := \big\{(I, F) \mid t \in \gamma_{IF}\big\}$$

the set of pairs $(I, F) \in \Omega$ whose canonical path γ_{IF} uses transition t. To bound the mixing time of the MC, we need to bound from above the

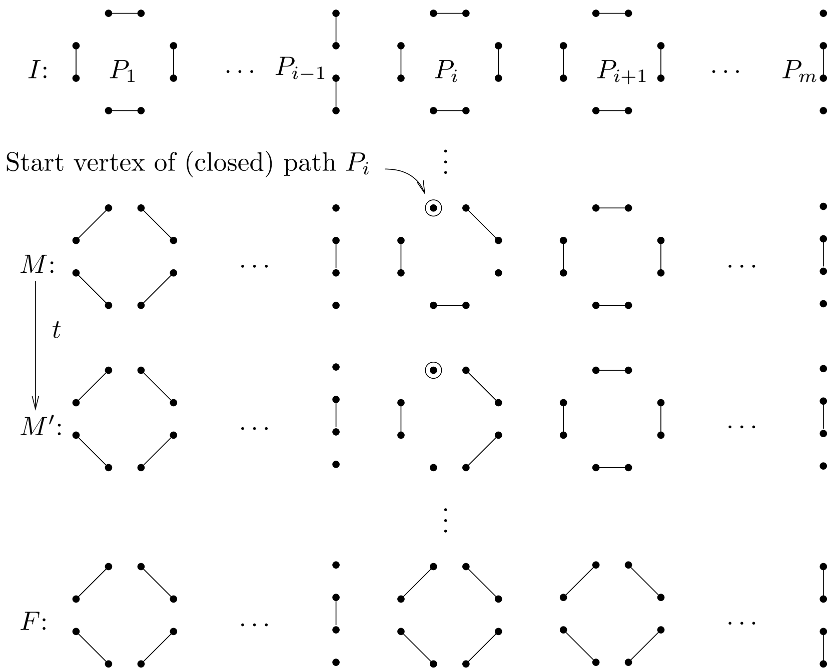

Figure 5.2: A step in a canonical path between matchings

congestion

$$\varrho = \max_{t=(M,M')} \left\{ \frac{1}{\pi(M)P(M,M')} \sum_{(I,F)\in\mathrm{cp}(t)} \pi(I)\pi(F)\,|\gamma_{IF}| \right\} \tag{5.17}$$

(c.f. (5.3)), where the maximum is over all transitions $t = (M, M')$. It is not immediately clear how to do this, as the sum is over a set we don't have a ready handle on. Suppose, however, that were able to construct, for each transition $t = (M, M')$, an injective function $\eta_t : \mathrm{cp}(t) \to \Omega$ such that

$$\pi(I)\pi(F) \lessapprox \pi(M)P(M,M')\,\pi(\eta_t(I,F)), \tag{5.18}$$

for all $(I, F) \in \mathrm{cp}(t)$, where the relational symbol "$\lessapprox$" indicates that the left-hand side is larger than the right-hand side by at most a polynomial factor in the "instance size," i.e., some measure of G and λ. Then it would

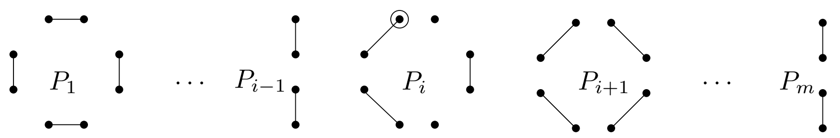

Figure 5.3: The corresponding encoding $\eta_t(X, Y)$.

follow that

$$\begin{aligned}
\varrho &\lessapprox \max_t \left\{ \sum_{(I,F)\in \mathrm{cp}(t)} \pi(\eta_t(I,F))\,|\gamma_{IF}| \right\} && \text{from (5.17) and (5.18)} \\
&\lessapprox \max_t \left\{ \sum_{(I,F)\in \mathrm{cp}(t)} \pi(\eta_t(I,F)) \right\} && \text{since } |\gamma_{IF}| \le n \\
&\le 1 && \text{since } \eta_t \text{ is injective.}
\end{aligned}$$

In other words, the congestion ϱ (and hence the mixing time of the MC) is polynomial in the instance size, as we should like.

We now complete the programme by defining an encoding η_t with the appropriate properties, and making exact the calculation just performed. To this end, fix a transition $t = (M, M')$. If t is a $\leftrightarrow$-transition, $(I, F) \in \mathrm{cp}(t)$, and the current component (with respect to the canonical path γ_{IF}) is a cycle, then we say that t is *troublesome* (with respect to the path γ_{IF}). If t is troublesome, then we denote by $e_{IFt} \in I$ the (unique) edge in I that is adjacent to the start vertex of the cycle being processed by t. For all $(I, F) \in \mathrm{cp}(t)$, define

$$\eta_t(I,F) = \begin{cases} \big(I \oplus F \oplus (M \cup M')\big) \setminus \{e_{IFt}\}, & \text{if } t \text{ is troublesome;} \\ I \oplus F \oplus (M \cup M'), & \text{otherwise.} \end{cases}$$

Roughly speaking, the encoding $C = \eta_t(I, F)$ agrees with I on the components that have been completely processed, and with F on the components that have not been touched yet. Moreover, C agrees with I and F on the edges common to both. (See Figure 5.3.) The crucial properties of η_t are described in the following sequence of claims.

Claim 5.13. *For all transitions t and all pairs $(I, F) \in \mathrm{cp}(t)$, the encoding $C = \eta_t(I, F)$ is a matching; thus η_t is a function with range Ω, as required.*

Proof. Consider the set of edges $A = I \oplus F \oplus (M \cup M')$, and suppose that some vertex, u say, has degree two in A. (Since $A \subseteq I \cup F$, no vertex degree can exceed two.) Then A contains edges $\{u, v_1\}, \{u, v_2\}$ for distinct vertices v_1, v_2, and since $A \subseteq I \cup F$, one of these edges must belong to I and the other to F. Hence both edges belong to $I \oplus F$, which means that neither can belong to $M \cup M'$. Following the form of $M \cup M'$ along the canonical path, however, it is clear that there can be at most one such vertex u; moreover, this happens precisely when t is a troublesome transition and u is the start vertex of the current cycle. Our definition of η_t removes one of the edges adjacent to u in this case, so all vertices in C have degree at most one, i.e., C is indeed a matching. □

Claim 5.14. *For every transition t, the function $\eta_t : \mathrm{cp}(t) \to \Omega$ is injective.*

Proof. Let t be a transition, and $(I, F) \in \mathrm{cp}(t)$. We wish to show that the pair (I, F) can be uniquely reconstructed from a knowledge only of t and $\eta_t(I, F)$. It is immediate from the definition of η_t that the symmetric difference $I \oplus F$ can be recovered from $C = \eta_t(I, F)$ using the relation

$$I \oplus F = \begin{cases} \big(C \oplus (M \cup M')\big) \cup \{e_{IFt}\}, & \text{if } t \text{ is troublesome;} \\ C \oplus (M \cup M'), & \text{otherwise.} \end{cases}$$

Of course, we don't know, a priori, the identity of the edge e_{IFt}. However, once we have formed the set $C \oplus (M \cup M')$ we can see that e_{IFt} is the unique edge that forms a cycle when added to the current path. There is a slightly delicate issue here: how do we know whether we are in the troublesome case or not? In other words, how to we know whether the current component is a cycle or a path? The answer lies in the convention for choosing the start vertex. It can be checked that choosing the lowest vertex as start vertex leads to a path being oriented in the opposite sense to a cycle in this potentially ambiguous situation.

Given $I \oplus F$, we can at once infer the sequence of paths $P_1, P_2, \ldots, P_m$ that have to be processed along the canonical path from I to F, and the transition t tells us which of these, P_i say, is the current one. The partition of $I \oplus F$ into I and F is now straightforward: I agrees with C on paths $P_1, \ldots, P_{i-1}$, and with M on paths $P_{i+1}, \ldots, P_m$. On the current path, P_i, the matching I agrees with C on the already processed part, and with M on the rest. (If t is troublesome, then the edge e_{IFt} also belongs to I.) Finally, the reconstruction of I and F is completed by noting that $I \cap F = M \setminus (I \oplus F)$, which is immediate from the definition of the paths. Hence I and F can be uniquely recovered from $C = \eta_t(I, F)$, so η_t is injective. □

Claim 5.15. *For all transitions $t = (M, M')$ and all pairs $(I, F) \in \mathrm{cp}(t)$,*

$$\pi(I)\pi(F) \le m\bar{\lambda}^2\pi(M)P(M, M')\,\pi(\eta_t(I, F)),$$

where $\bar{\lambda} := \max\{1, \lambda\}$.

Proof. Let $C = \eta_t(I, F)$, and consider the expressions

$$\lambda^{|I|}\lambda^{|F|} \quad \text{and} \quad \lambda^{|M\cup M'|}\lambda^{|C|},$$

which are closely related to the quantities

$$\pi(I)\pi(F) \quad \text{and} \quad \pi(M)P(M, M')\,\pi(\eta_t(I, F))$$

of interest. Each edge $e \in E$ contributes a factor 1, λ or λ^2 to $\lambda^{|I|}\lambda^{|F|}$, according to whether e is in neither, exactly one, or both of I and F. A similar observation can be made about $\lambda^{|M\cup M'|}\lambda^{|C|}$. If $e \notin I$ and $e \notin F$ then $e \notin M \cup M'$ and $e \notin C$, and the contribution to both expressions is 1. If $e \in I$ and $e \in F$ then $e \in M \cup M'$ and $e \in C$ and the contribution to both expressions is λ^2. If $e \in I \oplus F$ then $e \in (M \cup M') \oplus C$ and the contribution to both expressions is λ, with one possible exception: if t is troublesome and $e = e_{IFt}$ then there is a contribution λ to $\lambda^{|I|}\lambda^{|F|}$ and 1 to $\lambda^{|M\cup M'|}\lambda^{|C|}$. Thus,

$$\lambda^{|I|}\lambda^{|F|} \le \bar{\lambda}\,\lambda^{|M\cup M'|}\lambda^{|C|}.$$

Dividing by Z^2, the square of the partition function, it follows that

$$\pi(I)\pi(F) \le \bar{\lambda}^2\pi(M)\pi(C) \quad \text{and} \quad \pi(I)\pi(F) \le \bar{\lambda}^2\pi(M')\pi(C),$$

where we have used the fact that $|M|, |M'| \ge |M \cup M'| - 1$. Then

$$\begin{aligned}\pi(I)\pi(F) &\le \bar{\lambda}^2 \min\left\{\pi(M), \pi(M')\right\}\pi(C)\\ &= m\bar{\lambda}^2\pi(M)P(M, M')\pi(C) \qquad \text{by (5.2)},\end{aligned}$$

yielding the required inequality. □

Now we are ready to evaluate the congestion ϱ.

Proposition 5.16. *With a set of canonical paths Γ defined as in this section, the congestion $\varrho = \varrho(\Gamma)$ of the MC on matchings of a graph G (refer to Figure 5.1) is bounded by $\varrho \le nm\bar{\lambda}^2$, where n and m are the number of vertices and edges of G, respectively, and $\bar{\lambda} = \max\{1, \lambda\}$.*

Proof. We just need to make precise the rough calculation following (5.18).

$$\begin{aligned}
\varrho &= \max_{t=(M,M')} \left\{ \frac{1}{\pi(M)P(M,M')} \sum_{(I,F)\in \mathrm{cp}(t)} \pi(I)\pi(F)\,|\gamma_{IF}| \right\} \\
&\le m\bar{\lambda}^2 \sum_{(I,F)\in \mathrm{cp}(t)} \pi(\eta_t(I,F))\,|\gamma_{IF}| && \text{by Claim 5.15} \\
&\le nm\bar{\lambda}^2 \sum_{(I,F)\in \mathrm{cp}(t)} \pi(\eta_t(I,F)) && \text{since } |\gamma_{IF}| \le n \\
&\le nm\bar{\lambda}^2 && \text{by Claim 5.14.}
\end{aligned}$$

□

The sought-for bound on mixing time follows immediately.

Proof of Proposition 5.11. Combine Corollary 5.9 and Proposition 5.16, noting the crude bound $\ln \pi(x)^{-1} \le n \ln n + \frac{1}{2} n\,|\ln \lambda|$, which holds uniformly over $x \in \Omega$. □

Exercise 5.17. Show how to tighten the upper bound in Proposition 5.11 by a factor $\bar{\lambda}$. Since Claim 5.15 is essentially tight when t is troublesome, it is necessary to improve somehow the inequality

$$\sum_{(I,F)\in \mathrm{cp}(t)} \pi(\eta_t(I,F)) \le 1,$$

by studying carefully the range of η_t. See Jerrum and Sinclair [36], specifically the proof of their Proposition 12.4.

5.4 Extensions and further applications

Let G be a graph with at least one perfect matching (i.e., matching that covers all vertices of G). In the limit, as $\lambda \to \infty$, the partition function $Z(\lambda)$ counts the number of perfect matchings in G. However, the bound on mixing time provided by Proposition 5.11 grows unboundedly with increasing λ, so it is not clear whether the MC we have studied in this chapter provides us with a FPAUS for perfect matchings in G. At first we might hope that it is not necessary to set λ very large; perhaps the distribution (5.1) already places sufficient probability on the totality of perfect matchings at some quite modest λ. (According to Proposition 5.11, we need λ to be bounded by a

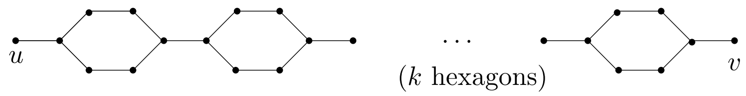

Figure 5.4: A graph with many "near perfect" matchings.

polynomial in n, the number of vertices in G, to achieve a FPAUS/FPRAS for perfect matchings.)

Unfortunately, there are graphs (see Figure 5.4) for which the perfect matchings make an insignificant contribution to distribution (5.1) unless λ is exponentially large in n. This claim follows from the these easily verified properties of the illustrated graph: (i) it has a unique perfect matching, and (ii) it has 2^k matchings that cover all vertices apart from u and v. The question of whether there exists an FPRAS (equivalently, by the observations of Chapter 3, an FPAUS) for perfect matchings in a general graph is still open at the time of writing. However, progress has been made in some special cases, that of bipartite graphs being perhaps the most interesting.

The problem of counting perfect matchings in a *bipartite* graph is of particular significance, since is is equivalent to evaluating the permanent of a $0,1$-matrix. (Refer to problems #BIPARTITEPM and 0,1-PERM of Chapter 2.) Recently, Jerrum, Sinclair and Vigoda [37] presented an FPRAS for the permanent of a $0,1$-matrix (in fact a general matrix with non-negative entries) using MC simulation. Noting that the counterexample of Figure 5.4 is bipartite, it is clear that we need to introduce a more sophisticated MC to achieve this result. In very rough terms, it is necessary to weight matchings according not just to the *number* of uncovered vertices but also their *locations*. In this way it is possible to access perfect matchings from near-perfect ones via a "staircase" of relatively small steps. Full details may be found in [37].

The canonical paths technique has also been applied by Jerrum and Sinclair to the ferromagnetic Ising model [35] and by Morris and Sinclair to "knapsack solutions" [52]. The latter application is particularly interesting for its use of random canonical paths.

5.5 Continuous time

It is possible to gain a better understanding of Theorem 5.6 and Corollary 5.7 by moving to continuous time.

Associated with any discrete-time MC $(X_t : t \in \mathbb{N})$ is a "continuised" MC $(\widetilde{X}_t : t \in \mathbb{R}^+)$. (We use tilde to distinguish continuous-time notions from their discrete-time analogues.) The MC $(\widetilde{X}_t)$ makes jumps at times $(t_1, t_2, t_3, \dots)$ where the time increments $t_{i+1} - t_i$, for $i \in \mathbb{N}$, are independent r.v's that are exponentially distributed with mean 1. (Here we use the convention $t_0 = 0$.) Between the jumps, i.e., in the intervals $[t_i, t_{i+1})$, for $i \in \mathbb{N}$, the value of $\widetilde{X}_t$ is constant. The jumps, when they occur, are governed by the same transition matrix P as the original MC (X_t). Informally, we have replaced deterministic time-1 holds between jumps by random, exponential, mean-1 holds. See Norris [53] for a proper treatment of continuous-time MC's.

The continuous-time MC has an "infinitesimal description" $\Pr(\widetilde{X}_{t+dt} = y \mid \widetilde{X}_t = x) = P(x, y)\,dt$ for all $x \neq y$. As a consequence, the distribution of $\widetilde{X}_t$ has a particularly pleasant form:

$$\widetilde{P}^t(x, y) := \Pr(\widetilde{X}_t = y \mid \widetilde{X}_0 = x) = \exp\{(P - I)t\},$$

where I is the identity matrix.[2] As in the discrete-time case, we aim to bound the rate of convergence of $(\widetilde{X}_t)$ to stationarity by analysing the decay of the variance

$$\operatorname{Var}_\pi(\widetilde{P}^t f) := \sum_{x \in \Omega} \pi(x) \bigl\{[\widetilde{P}^t f](x)\bigr\}^2, \tag{5.19}$$

where the function $\widetilde{P}^t f : \Omega \to \mathbb{R}$ is defined by

$$[\widetilde{P}^t f](x) := \sum_{y \in \Omega} \widetilde{P}^t(x, y) f(y), \tag{5.20}$$

and $f : \Omega \to \mathbb{R}$ is any test function with $\mathrm{E}_\pi f = 0$.

By calculus, starting with (5.19) and (5.20), we may derive (calculation left to the reader):

$$\frac{d}{dt} \operatorname{Var}_\pi(\widetilde{P}^t f) = 2 \sum_{x, y \in \Omega} \pi(x) \bigl(P(x, y) - I(x, y)\bigr)\, [\widetilde{P}^t f](x)\, [\widetilde{P}^t f](y).$$

[2]The exponential function applied to matrices can be understood as a convergent sum $\exp Q := I + Q + Q^2/2! + Q^3/3! + \cdots$.

Hence, setting $t = 0$, we obtain

$$\begin{aligned}\frac{d}{dt}\operatorname{Var}_\pi(\widetilde{P}^t f)\Big|_{t=0} &= 2\sum_{x,y\in\Omega}\pi(x)\big(P(x,y)-I(x,y)\big)f(x)f(y)\\ &= 2\sum_{x,y\in\Omega}\pi(x)P(x,y)f(x)f(y) - 2\operatorname{Var}_\pi f\\ &= -2\,\mathcal{E}_\pi(f,f),\end{aligned}$$

a continuous-time analogue of Theorem 5.6.

Applying Theorem 5.2, we see that $\operatorname{Var}_\pi(\widetilde{P}^t f)$ is bounded by the solution of the differential equation $\dot{v} = -(2/\varrho)v$, and hence

$$\operatorname{Var}_\pi(\widetilde{P}^t f) \le \exp\left\{-\frac{2t}{\varrho}\right\}\operatorname{Var}_\pi f, \tag{5.21}$$

a continuous-time analogue of Corollary 5.7.

Exercise 5.18. Follow through in detail the calculations sketched above.

Remarks 5.19. (a) The rate of decay of variance promised by (5.21) is faster than Corollary 5.7 by a factor 4. A factor 2 is explained by the avoidance of the lazy MC, but the remaining factor 2 is "real." This suggests that the calculation in Theorem 5.6 is not only a little mysterious, but also gives away a constant factor.

(b) Simulating the continuised MC is unproblematic, and can be handled by a device similar to that employed in the case of the lazy MC (c.f. Remarks 5.5). To obtain a sample from the distribution of $\widetilde{X}_t$: (i) generate a sample T from the Poisson distribution with mean t, and then (ii) simulate the discrete-time MC for T steps.

Bibliography

[1] David Aldous. Random walks on finite groups and rapidly mixing Markov chains. In *Seminar on probability, XVII*, pages 243–297. Springer, Berlin, 1983.

[2] David Aldous and James Fill. Reversible Markov chains and random walks on graphs. `http://www.stat.berkeley.edu/~aldous/book.html`.

[3] V. S. Anil Kumar and H. Ramesh. Coupling vs. conductance for the Jerrum-Sinclair chain. *Random Structures Algorithms*, 18(1):1–17, 2001.

[4] Catherine Bandle. *Isoperimetric Inequalities and Applications*. Pitman (Advanced Publishing Program), Boston, Mass., 1980.

[5] Edward A. Bender. The asymptotic number of non-negative integer matrices with given row and column sums. *Discrete Math.*, 10:217–223, 1974.

[6] Piotr Berman and Marek Karpinski. On some tighter inapproximability results (extended abstract). In *Automata, languages and programming (Prague, 1999)*, pages 200–209. Springer, Berlin, 1999.

[7] Béla Bollobás. *Modern Graph Theory*. Springer-Verlag, New York, 1998.

[8] Andrei Z. Broder. How hard is it to marry at random? (on the approximation of the permanent). In *Proceedings of the 18th Annual ACM Symposium on Theory of Computing (STOC)*, pages 50–58. ACM Press, 1986. Erratum in *Proceedings of the 20th Annual ACM Symposium on Theory of Computing*, 1988, p. 551.

[9] R. L. Brooks. On colouring the nodes of a network. *Proc. Cambridge Philos. Soc.*, 37:194–197, 1941.

[10] Russ Bubley and Martin Dyer. Graph orientations with no sink and an approximation for a hard case of #SAT. In *Proceedings of the Eighth Annual ACM-SIAM Symposium on Discrete Algorithms (New Orleans, LA, 1997)*, pages 248–257, New York, 1997. ACM.

[11] Russ Bubley and Martin Dyer. Path coupling: a technique for proving rapid mixing in Markov chains. In *Proceedings of the 38th Symposium on Foundations of Computer Science (FOCS)*, pages 223–231. IEEE Computer Society Press, 1997.

[12] Russ Bubley and Martin Dyer. Path coupling, Dobrushin uniqueness, and approximate counting. Technical Report 97.04, School of Computer Studies, University of Leeds, January 1997.

[13] Russ Bubley and Martin Dyer. Faster random generation of linear extensions. *Discrete Math.*, 201(1-3):81–88, 1999.

[14] Russ Bubley, Martin Dyer, and Catherine Greenhill. Beating the 2Δ bound for approximately counting colourings: a computer-assisted proof of rapid mixing. In *Proceedings of the Ninth Annual ACM-SIAM Symposium on Discrete Algorithms (San Francisco, CA, 1998)*, pages 355–363, New York, 1998. ACM Press.

[15] Russ Bubley, Martin Dyer, and Mark Jerrum. An elementary analysis of a procedure for sampling points in a convex body. *Random Structures Algorithms*, 12(3):213–235, 1998.

[16] Charles J. Colbourn, J. Scott Provan, and Dirk Vertigan. A new approach to solving three combinatorial enumeration problems on planar graphs. *Discrete Appl. Math.*, 60(1-3):119–129, 1995.

[17] Colin Cooper and Alan M. Frieze. Mixing properties of the Swendsen-Wang process on classes of graphs. *Random Structures Algorithms*, 15(3-4):242–261, 1999.

[18] Persi Diaconis. *Group representations in probability and statistics*. Institute of Mathematical Statistics, Hayward, CA, 1988.

[19] Persi Diaconis and Laurent Saloff-Coste. Comparison theorems for reversible Markov chains. *Ann. Appl. Probab.*, 3(3):696–730, 1993.

[20] Persi Diaconis and Daniel Stroock. Geometric bounds for eigenvalues of Markov chains. *Ann. Appl. Probab.*, 1(1):36–61, 1991.

[21] Alexander Dinghas. Über eine Klasse superadditiver Mengenfunktionale von Brunn-Minkowski-Lustenikschem typus. *Math. Zeitschr.*, 68:111–125, 1957.

[22] Martin Dyer, Alan Frieze, and Mark Jerrum. On counting independent sets in sparse graphs. In *Proceedings of the 40th Symposium on Foundations of Computer Science (FOCS)*, pages 210–217. IEEE Computer Society Press, 1999.

[23] Martin Dyer, Alan Frieze, and Ravi Kannan. A random polynomial-time algorithm for approximating the volume of convex bodies. *J. Assoc. Comput. Mach.*, 38(1):1–17, 1991.

[24] Martin Dyer and Catherine Greenhill. A more rapidly mixing Markov chain for graph colorings. *Random Structures Algorithms*, 13(3-4):285–317, 1998.

[25] Martin Dyer and Catherine Greenhill. On Markov chains for independent sets. *J. Algorithms*, 35(1):17–49, 2000.

[26] H. G. Eggleston. *Convexity.* Cambridge University Press, New York, 1958.

[27] Tomás Feder and Milena Mihail. Balanced matroids. In *Proceedings of the 24th Annual ACM Symposium on Theory of Computing (STOC)*, pages 26–38. ACM Press, 1992.

[28] Uriel Feige and Carsten Lund. On the hardness of computing the permanent of random matrices. *Comput. Complexity*, 6(2):101–132, 1996/97.

[29] Joan Feigenbaum and Lance Fortnow. Random-self-reducibility of complete sets. *SIAM J. Comput.*, 22(5):994–1005, 1993.

[30] Michael R. Garey and David S. Johnson. *Computers and Intractability: a Guide to the Theory of NP-Completeness.* W. H. Freeman and Co., San Francisco, Calif., 1979.

[31] Leslie Ann Goldberg. Computation in permutation groups: counting and randomly sampling orbits. In *Surveys in combinatorics, 2001 (Sussex)*, pages 109–143. Cambridge Univ. Press, Cambridge, 2001.

[32] Oded Goldreich. *Introduction to Complexity Theory*. Lecture Notes Series of the Electronic Colloquium on Computational Complexity. `http://www.eccc.uni-trier.de/eccc/`, 1999.

[33] G. R. Grimmett and D. R. Stirzaker. *Probability and Random Processes*. The Clarendon Press Oxford University Press, New York, second edition, 1992.

[34] Mark Jerrum. Computational Pólya theory. In *Surveys in combinatorics, 1995 (Stirling)*, pages 103–118. Cambridge Univ. Press, Cambridge, 1995.

[35] Mark Jerrum and Alistair Sinclair. Polynomial-time approximation algorithms for the Ising model. *SIAM J. Comput.*, 22(5):1087–1116, 1993.

[36] Mark Jerrum and Alistair Sinclair. The Markov chain Monte Carlo method: an approach to approximate counting and integration. In Dorit S. Hochbaum, editor, *Approximation Algorithms for NP-hard Problems*, pages 482–520. PWS, 1996.

[37] Mark Jerrum, Alistair Sinclair, and Eric Vigoda. A polynomial-time approximation algorithm for the permanent of a matrix with non-negative entries. *Electronic Colloquium on Computational Complexity*, TR00-079, 2000.

[38] Mark R. Jerrum, Leslie G. Valiant, and Vijay V. Vazirani. Random generation of combinatorial structures from a uniform distribution. *Theoret. Comput. Sci.*, 43(2-3):169–188, 1986.

[39] Ravi Kannan, László Lovász, and Miklós Simonovits. Random walks and an $O^*(n^5)$ volume algorithm for convex bodies. *Random Structures Algorithms*, 11(1):1–50, 1997.

[40] Richard M. Karp, Michael Luby, and Neal Madras. Monte Carlo approximation algorithms for enumeration problems. *J. Algorithms*, 10(3):429–448, 1989.

[41] P. W. Kasteleyn. Graph theory and crystal physics. In Frank Harary, editor, *Graph Theory and Theoretical Physics*, pages 43–110. Academic Press, 1967.

[42] Torgny Lindvall. *Lectures on the coupling method*. John Wiley & Sons Inc., New York, 1992. A Wiley-Interscience Publication.

[43] Torgny Lindvall and L. C. G. Rogers. Coupling of multidimensional diffusions by reflection. *Ann. Probab.*, 14(3):860–872, 1986.

[44] L. Lovász and M. Simonovits. Random walks in a convex body and an improved volume algorithm. *Random Structures Algorithms*, 4(4):359–412, 1993.

[45] László Lovász and M. D. Plummer. *Matching Theory.* North-Holland, 1986.

[46] Michael Luby and Eric Vigoda. Approximately counting up to four (extended abstract). In *Proceedings of the 29th Annual ACM Symposium on Theory of Computing (STOC)*, pages 682–687. ACM Press, 1997.

[47] Michael Luby and Eric Vigoda. Fast convergence of the Glauber dynamics for sampling independent sets. *Random Structures Algorithms*, 15(3-4):229–241, 1999.

[48] Neal Madras and Dana Randall. Markov chain decomposition for convergence rate analysis. To appear in the *Annals of Applied Probability.*

[49] Meena Mahajan and V. Vinay. A combinatorial algorithm for the determinant. In *Proceedings of the 8th Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 730–738. ACM/SIAM, 1997.

[50] Lisa McShine. Random sampling of labeled tournaments. *Electron. J. Combin.*, 7(1):Research Paper 8, 9 pp. (electronic), 2000.

[51] Milena Mihail. Conductance and convergence of Markov chains: a combinatorial treatment of expanders. In *Proceedings of the 30th Symposium on Foundations of Computer Science (FOCS)*, pages 526–531. IEEE Computer Society Press, 1989.

[52] Ben Morris and Alistair Sinclair. Random walks on truncated cubes and sampling 0-1 knapsack solutions. In *Proceedings of the 40th Symposium on Foundations of Computer Science (FOCS)*, pages 230–240. IEEE Computer Society Press, 1999.

[53] J. R. Norris. *Markov chains.* Cambridge University Press, Cambridge, 1997.

[54] Christos H. Papadimitriou. *Computational Complexity.* Addison-Wesley Publishing Company, Reading, MA, 1994.

[55] Neil Robertson, P. D. Seymour, and Robin Thomas. Permanents, Pfaffian orientations, and even directed circuits. *Ann. of Math. (2)*, 150(3):929–975, 1999.

[56] Janos Simon. On the difference between one and many. In *Automata, Languages and Programming (Fourth Colloq., Univ. Turku, Turku, 1977), Lecture Notes in Computer Science, Vol. 52*, pages 480–491. Springer, Berlin, 1977.

[57] Alistair Sinclair. Improved bounds for mixing rates of Markov chains and multicommodity flow. *Combin. Probab. Comput.*, 1(4):351–370, 1992.

[58] Alistair Sinclair. *Algorithms for Random Generation and Counting: a Markov Chain Approach.* Birkhäuser Boston Inc., Boston, MA, 1993.

[59] P. Tetali and S. Vempala. Random sampling of Euler tours. *Algorithmica*, 30(3):376–385, 2001.

[60] Seinosuke Toda and Mitsunori Ogiwara. Counting classes are at least as hard as the polynomial-time hierarchy. *SIAM J. Comput.*, 21(2):316–328, 1992.

[61] W. T. Tutte. *Graph Theory*, volume 21 of *Encyclopedia of Mathematics: Combinatorics*. Addison-Wesley, 1984.

[62] L. G. Valiant. Completeness classes in algebra. In *Proceedings of the 11th annual ACM Symposium on Theory of Computing (STOC)*, pages 249–261. ACM, 1979.

[63] L. G. Valiant. The complexity of computing the permanent. *Theoret. Comput. Sci.*, 8(2):189–201, 1979.

[64] L. G. Valiant and V. V. Vazirani. NP is as easy as detecting unique solutions. *Theoret. Comput. Sci.*, 47(1):85–93, 1986.

[65] Leslie G. Valiant. The complexity of enumeration and reliability problems. *SIAM J. Comput.*, 8(3):410–421, 1979.

[66] J. H. van Lint and R. M. Wilson. *A course in combinatorics*. Cambridge University Press, Cambridge, 1992.

[67] Eric Vigoda. Improved bounds for sampling colorings. *J. Math. Phys.*, 41(3):1555–1569, 2000.