But from (3.7), using again the fact about the exponential function, we have

$$e^{-\varepsilon/2}\varrho_1\varrho_2\dots\varrho_m \le \mu_1\mu_2\dots\mu_m \le e^{\varepsilon/2}\varrho_1\varrho_2\dots\varrho_m\,,$$

which combined with the previous inequality implies

$$e^{-\varepsilon}\varrho_1\varrho_2\dots\varrho_m \le \overline{Z}_1\overline{Z}_2\dots\overline{Z}_m \le e^{\varepsilon}\varrho_1\varrho_2\dots\varrho_m$$

with probability at least $\frac{3}{4}$. Since $\overline{Z}_1\overline{Z}_2\dots\overline{Z}_m = N^{-1}$ and $\varrho_1\varrho_2\dots\varrho_m = |\mathcal{M}(G)|^{-1}$, our estimator N for $|\mathcal{M}(G)|$ satisfies requirement (3.1). Thus the algorithm that computes computes N as above is an FPRAS for $|\mathcal{M}(G)|$.

The run-time of the algorithm is dominated by the number of samples required, which is $sm \le 75\varepsilon^{-2}m^2$, multiplied by the time-per-sample, which is $T(n,m,\varepsilon)$; the claimed time-bound is immediate. □

Exercise 3.5. Prove a result analogous to Proposition 3.4 with (proper vertex) q-colourings of a graph replacing matchings. Assume that the number of colours q is strictly greater than the maximum degree Δ of G. There is no need to repeat all the calculation, which is in fact identical. The key thing is to obtain an inequality akin to (3.5), but for colourings in place of matchings.

In light of the connection between approximate counting and almost uniform sampling, methods for sampling from complex combinatorially defined sets gain additional significance. The most powerful technique known to us is Markov chain simulation.

3.3 Markov chains

We deal exclusively in this section with discrete-time Markov chains on a finite state space Ω. Many of the definitions and claims extend to countable state spaces with only minor complication. In Chapter 6 we shall need to employ Markov chains with continuous state spaces, but the corresponding definitions and basic facts will be left until they are required. See Grimmett and Stirzaker's textbook [33] for a more comprehensive treatment.

A sequence $(X_t \in \Omega)_{t=0}^{\infty}$ of random variables (r.v's) is a *Markov chain* (MC), with state space Ω, if

$$\Pr[X_{t+1} = y \mid X_t = x_t, X_{t-1} = x_{t-1}, \dots, X_0 = x_0] = \Pr[X_{t+1} = y \mid X_t = x_t], \tag{3.10}$$

for all $t \in \mathbb{N}$ and all $x_t, x_{t-1}, \dots, x_0 \in \Omega$. Equation (3.10) encapsulates the *Markovian property* whereby the history of the MC prior to time t is forgotten. We deal only with *(time-) homogeneous* MCs, i.e., ones for which the right-hand side of (3.10) is independent of t. In this case, we may write

$$P(x,y) := \Pr[X_{t+1} = y \mid X_t = x],$$

where P is the *transition matrix* of the MC. The transition matrix P describes single-step transition probabilities; the t-step transition probabilities P^t are given inductively by

$$P^t(x,y) := \begin{cases} I(x,y), & \text{if } t = 0; \\ \sum_{y' \in \Omega} P^{t-1}(x,y')P(y',y), & \text{if } t > 0, \end{cases}$$

where I denotes the identity matrix $I(x,y) := \delta_{xy}$. Thus $P^t(x,y) = \Pr[X_t = y \mid X_0 = x]$.

A *stationary distribution* of an MC with transition matrix P is a probability distribution $\pi : \Omega \to [0,1]$ satisfying

$$\pi(y) = \sum_{x \in \Omega} \pi(x)P(x,y).$$

Thus if X_0 is distributed as π then so is X_1 (and hence so is X_t for all $t \in \mathbb{N}$). A finite MC always has at least one stationary distribution. An MC is *irreducible* if, for all $x, y \in \Omega$, there exists a $t \in \mathbb{N}$ such that $P^t(x,y) > 0$; it is *aperiodic* if $\gcd\{t : P^t(x,x) > 0\} = 1$ for all $x \in \Omega$.[5] A (finite-state) MC is *ergodic* if it is both irreducible and aperiodic.

Theorem 3.6. *An ergodic MC has a unique stationary distribution π; moreover the MC tends to π in the sense that $P^t(x,y) \to \pi(y)$, as $t \to \infty$, for all $x \in \Omega$.*

Informally, an ergodic MC eventually "forgets" its starting state. Computation of the stationary distribution is facilitated by the following little lemma:

Lemma 3.7. *Suppose P is the transition matrix of an MC. If the function $\pi' : \Omega \to [0,1]$ satisfies*

$$\pi'(x)P(x,y) = \pi'(y)P(y,x), \quad \text{for all } x, y \in \Omega, \tag{3.11}$$

and

$$\sum_{x \in \Omega} \pi'(x) = 1,$$

then π' is a stationary distribution of the MC. If the MC is ergodic, then clearly $\pi' = \pi$ is the unique stationary distribution.

Proof. We just need to check that π' is invariant. Suppose X_0 is distributed as π'. Then

$$\Pr[X_1 = y] = \sum_{x \in \Omega} \pi'(x)P(x,y) = \sum_{x \in \Omega} \pi'(y)P(y,x) = \pi'(y).$$

□

[5] In the case of an irreducible MC, it is sufficient to verify the condition $\gcd\{t : P^t(x,x) > 0\} = 1$ for just one state $x \in \Omega$.

Remark 3.8. Condition (3.11) is known as *detailed balance*. An MC for which it holds is said to be *time reversible*. Clearly, Lemma 3.7 cannot be applied to non-time-reversible MCs. This is not a problem in practice, since all the MCs we consider are time reversible. In fact, it is difficult in general to determine the stationary distribution of large non-time-reversible MCs, unless there is some special circumstance, for example symmetry, that can be taken into consideration. Furthermore, all the usual methods for constructing MCs with specified stationary distributions produce time-reversible MCs.

Example 3.9. Here is a natural (time homogeneous) MC whose state space is the set $\mathcal{M}(G)$ of all matchings (of all sizes) in a specified graph $G = (V, E)$. The transition matrix of the MC is defined implicitly, by an experimental trial. Suppose the initial state is $X_0 = M \in \mathcal{M}(G)$. The next state X_1 is the result of the following trial:

1. With probability $\frac{1}{2}$ set $X_1 \leftarrow M$ and halt.
2. Otherwise, select $e \in E(G)$ and set $M' \leftarrow M \oplus \{e\}$.[6]
3. If $M' \in \mathcal{M}(G)$ then $X_1 \leftarrow M'$ else $X_1 \leftarrow M$.

Since the MC is time homogeneous, it is enough to describe the first transition; subsequent transitions follow an identical trial. Step 1 may seem a little unnatural, but we shall often include such a looping transition to avoid a certain technical complication. Certainly its presence ensures that the MC is aperiodic. The MC is also irreducible, since it is possible to reach the empty matching from any state by removing edges (and reach any state from the empty matching by adding edges). Thus the MC is ergodic and has a unique stationary distribution.

Exercise 3.10. Demonstrate, using Lemma 3.7, that the stationary distribution of the MC of Example 3.9 is uniform over $\mathcal{M}(G)$.

Exercise 3.10 and Proposition 3.4, taken together, immediately suggest an approach to estimating the number of matchings in a graph. Simulate the MC on $\mathcal{M}(G)$ for T steps, starting at some fixed state X_0, say $X_0 = \emptyset$, and return the final state X_T. If T is sufficiently large, this procedure will satisfy the requirements of an almost uniform sampler for matchings in G. Then the method of Proposition 3.4 may be used to obtain a randomised approximation scheme for the number of matchings $|\mathcal{M}(G)|$. Whether this approach is feasible depends crucially on the rate of convergence of the MC to stationarity. We shall prove in Chapter 5 that a modification[7] of the MC described in Example 3.9 does in fact come "close" to stationarity in a polynomial number of steps (in the size of the graph G), hence yielding an FPRAS for the number of matchings in a graph.

[6] The symbol $\oplus$ denotes symmetric difference.

[7] In fact, by comparing the original and modified MCs [19], one can show that the MC as presented in Example 3.9 also converges in polynomially many steps.

Chapter 4

Coupling and colourings

The outline of our programme is now clear: in order to count (approximately) it is enough to be able to sample (almost) uniformly; in order to sample we may simulate an appropriately defined MC. For this approach to be feasible, however, it is important that the MC in question is "rapidly mixing," i.e., that it converges to near-equilibrium in time polynomial (hopefully of small degree) in the size of the problem instance. Since the state space is usually of exponential size as a function of the problem size — think of the number of matchings in a graph as a function of the size of the graph — this is a distinctly non-trivial requirement. We shall presently formalise the rate of convergence to equilibrium in terms of the "mixing time" of the MC. The classical theory of MCs has little to say concerning non-asymptotic bounds on mixing time, and most of the techniques we use have been specially developed for the task in hand. However, there is one classical device, namely coupling, that can be applied in certain situations. As it is the most elementary approach to bounding mixing times, we study it first.

4.1 Colourings of a low-degree graph

Anil Kumar and Ramesh [3] present persuasive evidence that the coupling argument is not applicable to the MC on matchings that was defined at the end of the previous chapter. We therefore use a somewhat simpler example, namely colourings of a low-degree graph. Let $G = (V, E)$ be an undirected graph, and Q a set of q colours. A (proper) q-colouring of G is a an assignment $\sigma : V \to Q$ of colours to the vertices of G such that $\sigma(u) \neq \sigma(v)$ for all edges $\{u, v\} \in E$. In general, even deciding existence of a q-colouring in G is computationally intractable, so we need to impose some condition on G and q.

Denote by Δ the maximum degree[1] of any vertex in G. Brooks' theorem asserts that a q-colouring exists when $q \geq \Delta$, provided $\Delta \geq 3$ and G does not

[1] The *degree* of a vertex is the number of edges incident at that vertex.

M. Jerrum *Counting, Sampling and Integrating: Algorithms and Complexity*

1. Select a vertex $v \in V$, u.a.r.
2. Select a colour $c \in Q \setminus X_0(\Gamma(v))$, u.a.r.
3. $X_1(v) \leftarrow c$ and $X_1(u) \leftarrow X_0(u)$ for all $u \neq v$.

Figure 4.1: Trial defining an MC on q-colourings.

contain $K_{\Delta+1}$ as a connected component [7, 9].[2] The proof of Brooks' theorem is effective, and yields a polynomial-time algorithm for constructing a q-colouring. It is also best possible in the (slightly restricted) sense that there are pairs, for example $q = 3$, $\Delta = 4$, which just fail the condition of the theorem, and for which the problem of deciding q-colourability is NP-complete, even when restricted to graphs of maximum degree Δ. So if we are aiming at an efficient sampling procedure for q-colourings we should certainly assume $q \geq \Delta$. Moreover, to approximate the number of q-colourings using the reduction of Exercise 3.5 we need to assume further that $q > \Delta$. Before we complete the work of this section, we shall need to strengthen this condition still further.

So let $G = (V, E)$ be a graph of maximum degree Δ and let Ω denote the set of all q-colourings of G, for some $q > \Delta$. Denote by $\Gamma(v) = \{u : \{u, v\} \in E(G)\}$ the set of vertices in G that are adjacent to v. Consider the (time-homogeneous) MC (X_t) on Ω whose transitions are defined by the experimental trial presented in Figure 4.1. Here we are considering a colouring as a function $V \to Q$, so $X_0(u)$ denotes the colour of vertex u in the initial state, and $X_0(\Gamma(v)) = \{X_0(u) : u \in \Gamma(v)\}$ denotes the set of all colours applied to neighbours of v. Note that the assumption $q > \Delta$ makes it easy to construct a valid initial state X_0.

Exercises 4.1.

1. Prove that the above MC is irreducible (and hence ergodic) under the (stronger) assumption $q \geq \Delta + 2$. Further prove, using Lemma 3.7, that its (unique) stationary distribution is uniform over Ω.

2. [Alan Sokal.] Exhibit a sequence of connected graphs of increasing size, with $\Delta = 4$, such that the above MC fails to be irreducible when $q = 5$. (Hint: as a starting point, construct a "frozen" 5-colouring of the infinite square lattice, i.e., the graph with vertex set $\mathbb{Z} \times \mathbb{Z}$ and edge set $\big\{(i, j), (i', j') : |i - i'| + |j - j'| = 1\big\}$. The adjective "frozen" applied to a state is intended to indicate that the only transition available from the state is a loop (with probability 1) to the same state.)

3. Design an MC on q-colourings of an arbitrary graph G of maximum degree Δ that is ergodic, provided only that $q \geq \Delta + 1$. The MC should be easily implementable, otherwise there is no challenge! (Hint: use transitions based on edge updates rather than vertex updates.)

[2] K_r denotes the complete graph on r vertices.

We shall show that (X_t) is rapidly mixing, provided $q \geq 2\Delta + 1$, which we assume from now on. (The reader may be assured that this is the very last time we shall strengthen the lower bound on the number of colours!) This result will provide us with a simple and efficient sampling procedure for q-colourings in low-degree graphs.

Suppose (X_t) is any ergodic MC on countable state space Ω, with transition matrix P and initial state $X_0 = x \in \Omega$. For $t \in \mathbb{N}$, the distribution of X_t (the t step distribution) is naturally denoted $P^t(x, \cdot)$. Let π denote the the stationary distribution of the MC, i.e., the limit of $P^t(x, \cdot)$ as $t \to \infty$. Recall the definition of total variation distance from (3.2). We measure the rate of convergence to stationarity of (X_t) by its *mixing time* (from initial state x):

$$\tau_x(\varepsilon) := \min\{t : \|P^t(x, \cdot) - \pi\|_{\mathrm{TV}} \leq \varepsilon\}. \tag{4.1}$$

Lemma 4.2. *The total variation distance $\|P^t(x, \cdot) - \pi\|_{\mathrm{TV}}$ of the t-step distribution from stationarity is a non-increasing function of t.*

Exercise 4.3. Prove Lemma 4.2. (A proof is given at the end of the chapter.)

In the light of Lemma 4.2, the following definition of mixing time is equivalent to (4.1):

$$\tau_x(\varepsilon) := \min\{t : \|P^s(x, \cdot) - \pi\|_{\mathrm{TV}} \leq \varepsilon, \text{ for all } s \geq t\}.$$

In other words, once the total variation distance becomes smaller than ε it stays smaller than ε.

Often we would like to make a statement about mixing time that is independent of the initial state, in which case we take a worst-case view and write

$$\tau(\varepsilon) = \max_{x \in \Omega} \tau_x(\varepsilon);$$

we shall refer to $\tau(\varepsilon)$ simply as the *mixing time*.

Remark 4.4. Sometimes the further simplification of setting ε to some constant, say $\varepsilon = \frac{1}{4}$, is made. The justification for this runs as follows. If τ is the first time t at which $\|P^t(x, \cdot) - \pi\|_{\mathrm{TV}} \leq \frac{1}{4}$, then it can be shown [2, Chap. 2, Lemma 20] that $\|P^{k\tau}(x, \cdot) - \pi\|_{\mathrm{TV}} \leq 2^{-k}$ for every $k \in \mathbb{N}$.

Our aim in the next section is to show that the mixing time $\tau(\varepsilon)$ of the MC on colourings is bounded by a polynomial in n and $\log \varepsilon^{-1}$.

Proposition 4.5. *Suppose G is a graph on n vertices of maximum degree Δ. Assuming $q \geq 2\Delta + 1$, the mixing time $\tau(\varepsilon)$ of the MC of Figure 4.1 is bounded above by*

$$\tau(\varepsilon) \leq \frac{q - \Delta}{q - 2\Delta} \, n \ln\left(\frac{n}{\varepsilon}\right).$$

Taking the instance size n into account is a prominent feature of applications of MCs in computer science, especially as compared with classical Markov chain theory. Observe that Proposition 4.5, combined with Proposition 3.4, implies the existence of an FPRAS for q-colourings in graphs of low enough degree.

Corollary 4.6. *Suppose G is a connected graph of maximum degree Δ, and $q \geq 2\Delta + 1$. Then there is an FPRAS for counting q-colourings in G. Denote by n the number of vertices in G and by m the number of edges. Then the running time of this FPRAS as a function of n, m and the error tolerance ε (regarding Δ and q as fixed) is bounded by $cnm^2\varepsilon^{-2}\max\{\ln(m/\varepsilon), 1\}$ for some constant c.*

4.2 Bounding mixing time using coupling

Coupling as a proof technique was discovered by Doeblin in the 1930s. However, its more recent popularity as a tool for bounding mixing time owes much to Aldous. Actually, we shall be using only a restricted form of coupling, namely Markovian coupling.

We start with a ground (time homogeneous) MC (Z_t) with state space Ω and transition matrix P. A *(Markovian) coupling* for (Z_t) is an MC (X_t, Y_t) on $\Omega \times \Omega$, with transition probabilities defined by:

$$\begin{aligned} \Pr[X_1 = x' \mid X_0 = x, Y_0 = y] &= P(x, x'), \\ \Pr[Y_1 = y' \mid X_0 = x, Y_0 = y] &= P(y, y'). \end{aligned} \tag{4.2}$$

Equivalently, with $\widehat{P} : \Omega^2 \to \Omega^2$ denoting the transition matrix of the coupling,

$$\begin{aligned} \sum_{y' \in \Omega} \widehat{P}((x, y), (x', y')) &= P(x, x'), \\ \sum_{x' \in \Omega} \widehat{P}((x, y), (x', y')) &= P(y, y'). \end{aligned}$$

Thus, the sequence of r.v.'s (X_t) viewed in isolation forms an MC with transition matrix P, as does the sequence (Y_t).

The easy way to achieve (4.2) would be to assume independence of (X_t) and (Y_t), i.e., that

$$\widehat{P}((x, y), (x', y')) = P(x, x')P(y, y').$$

But this is not necessary, and for our application not desirable. Instead, we are after some correlation that will tend to bring (X_t) and (Y_t) together (whatever their initial states) so that $X_t = Y_t$ for some quite small t. Note that once $X_t = Y_t$, we can arrange quite easily for X_s to be equal to Y_s, for all $s \geq t$, while continuing to satisfy (4.2): just choose a transition from X_s and let Y_s copy it.

The following simple lemma, which is the basis of the coupling method, was perhaps first made explicit by Aldous [1, Lemma 3.6]; see also Diaconis [18, Chap. 4, Lemma 5].

Lemma 4.7 (Coupling Lemma). *Let (X_t, Y_t) be any coupling, satisfying (4.2), based on a ground MC (Z_t) on Ω. Suppose $t : [0,1] \to \mathbb{N}$ is a function satisfying the condition: for all $x, y \in \Omega$, and all $\varepsilon > 0$*

$$\Pr[X_{t(\varepsilon)} \neq Y_{t(\varepsilon)} \mid X_0 = x, Y_0 = y] \leq \varepsilon.$$

Then the mixing time $\tau(\varepsilon)$ of (Z_t) is bounded above by $t(\varepsilon)$.

Proof. Denote by P the transition matrix of (Z_t). Let $A \subseteq \Omega$ be arbitrary. Let $X_0 = x \in \Omega$ be fixed, and Y_0 be chosen according to the stationary distribution π of (Z_t). For any $\varepsilon \in (0,1)$ and corresponding $t = t(\varepsilon)$,

$$\begin{aligned} P^t(x, A) &= \Pr[X_t \in A] \\ &\geq \Pr[X_t = Y_t \wedge Y_t \in A] \\ &= 1 - \Pr[X_t \neq Y_t \vee Y_t \notin A] \\ &\geq 1 - (\Pr[X_t \neq Y_t] + \Pr[Y_t \notin A]) \\ &\geq \Pr(Y_t \in A) - \varepsilon \\ &= \pi(A) - \varepsilon. \end{aligned}$$

Hence, by the second part of definition (3.2), $\|P^t(x, \cdot) - \pi\|_{\mathrm{TV}} \leq \varepsilon$. □

Remark 4.8. Actually we established the stronger conclusion

$$\|P^t(x, \cdot) - P^t(y, \cdot)\|_{\mathrm{TV}} \leq \varepsilon, \quad \text{for all pairs } x, y \in \Omega.$$

This slightly different notion of l_1-convergence corresponds to a slightly different notion of mixing time. This new mixing time has certain advantages, notably submultiplicativity: see Aldous and Fill [2] for more detail.

Let's now see how these ideas may be applied to the q-colouring MC of Figure 4.1. We need to define a coupling on Ω^2 such that the projections onto the first and second coordinates are faithful copies of the original MC in the sense of (4.2). Moreover, we wish the coupling to *coalesce*, i.e., reach a state where $X_t = Y_t$, as soon as possible. Figure 4.2 presents what seems at first sight to be a reasonable proposal. Note that if you hide the random variable Y_1 then the companion random variable X_1 is distributed exactly as if we had used the trial presented in Figure 4.1. (By symmetry, a similar statement could be made about Y_1.) Thus the coupling condition (4.2) is satisfied.

We have argued that the coupling in Figure 4.2 is correct, but how efficient is it? Intuitively, provided we can arrange for $\Pr[c_x = c_y]$ in step 2 to be large, we ought to reach a state with $X_t = Y_t$ (i.e., coalescence) in not too many steps. The Coupling Lemma will then provide a good upper bound on mixing time. In order to understand what is involved in maximising $\Pr[c_x = c_y]$, the following exercise may be useful.

1. Select a vertex $v \in V$ u.a.r.
2. Select a pair of colours (c_x, c_y) from some joint distribution on
$$\big(Q \setminus X_0(\Gamma(v))\big) \times \big(Q \setminus Y_0(\Gamma(v))\big)$$
that has the "correct" marginal distributions; specifically, the distribution of c_x (respectively c_y) should be uniform over $Q \setminus X_0(\Gamma(v))$ (respectively $Q \setminus Y_0(\Gamma(v))$. This joint distribution will be chosen so as to maximise $\Pr[c_x = c_y]$.
3. Set $X_1(v) \leftarrow c_x$ and $Y_1(v) \leftarrow c_y$.

Figure 4.2: A coupling for the MC on colourings

Exercise 4.9. Suppose that $Q = \{0, 1, \ldots, 6\}$, $X_0(\Gamma(v)) = \{3, 6\}$ and $Y_0(\Gamma(v)) = \{4, 5, 6\}$. Thus the sets of legal colours for v in X_1 and Y_1 are $c_x \in \{0, 1, 2, 4, 5\}$ and $c_y \in \{0, 1, 2, 3\}$, respectively. Construct a joint distribution for (c_x, c_y) such that c_x is uniform on $\{0, 1, 2, 4, 5\}$, c_y is uniform on $\{0, 1, 2, 3\}$, and $\Pr[c_x = c_y] = \frac{3}{5}$. Show that your construction is optimal.

The best that can be done in general is as follows.

Lemma 4.10. *Let U be a finite set, A, B be subsets of U, and Z_a, Z_b be random variables, taking values in U. Then there is a joint distribution for Z_a and Z_b such that Z_a (respectively Z_b) is uniform and supported on A (respectively B) and, furthermore,*
$$\Pr[Z_a = Z_b] = \frac{|A \cap B|}{\max\{|A|, |B|\}}$$

Exercise 4.11. Prove Lemma 4.10 and show that the result is best possible. (Assuming your construction in Exercise 4.9 is reasonably systematic, it should be possible to adapt it to the general situation.)

Remark 4.12. The term "coupling" does not have a precise agreed meaning, but its general sense is the following. A pair or perhaps a larger collection of r.v's is given. A coupling is a joint distribution of the several variables that has the correct marginals — i.e., each r.v., when observed independently of the others, has the correct probability distribution — but, taken together, the variables are seen to be correlated. Usually the correlation aims to "bring the r.v's closer together" in some sense. Lemma 4.10 is a special example of an optimal coupling of two r.v's that Lindvall calls the *γ-coupling* [42, §I.5]. The coupling of MCs, as captured in condition (4.2), is another example of the concept.

We are now well prepared for the main result of the chapter.

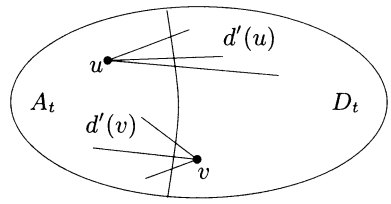

Figure 4.3: Two ways to count the edges spanning the cut (A_t, D_t).

Proof of Proposition 4.5. We analyse the coupling of Figure 4.2 using the joint distribution for the colour-pair (c_x, c_y) that is implicit in Lemma 4.10. Thus, letting

$$\xi := |Q \setminus X_0(\Gamma(v))| \qquad (= \#\text{ legal colours for } v \text{ in } X_1),$$
$$\eta := |Q \setminus Y_0(\Gamma(v))| \qquad (= \#\text{ legal colours for } v \text{ in } Y_1),$$

and

$$\zeta := \big|\big(Q \setminus X_0(\Gamma(v))\big) \cap \big(Q \setminus Y_0(\Gamma(v))\big)\big| \qquad (= \#\text{ common legal colours}),$$

the probability that the same colour is chosen for both X_1 and Y_1 in step 2 is just

$$\Pr[c_x = c_y] = \frac{\zeta}{\max\{\xi, \eta\}}. \tag{4.3}$$

Consider the situation that obtains after the coupling has been run for t steps. Let $A_t \subseteq V$ be the set of vertices on which the colourings X_t and Y_t agree, and $D_t = V \setminus A_t$ be the set on which they disagree. Let $d'(v)$ denote the number of edges incident at vertex v that have one endpoint in A_t and one in D_t. Clearly,

$$\sum_{v \in A_t} d'(v) = \sum_{u \in D_t} d'(u) = m',$$

where m' is the number of edges of G that span A_t and D_t. (The situation is visualised in Figure 4.3.) We want to prove that the disagreement set D_t tends to get smaller and smaller.

In one transition, the size of the disagreement set D_t changes by at most one. We therefore need to consider just three cases: increasing/decreasing by one or remaining constant. In fact, we just need to compute the probability of the first two, since the third can be got by complementation.

Consider first the probability that $|D_{t+1}| = |D_t| + 1$. For this event to occur, the vertex v selected in step 1 must lie in A_t, and the new colours c_x and c_y selected in step 2 must be different. Observing that the quantities ξ, η and ζ satisfy the linear inequalities

$$\begin{aligned} \xi - \zeta &\le d'(v), \\ \eta - \zeta &\le d'(v), \quad \text{and} \\ \xi, \eta &\ge q - \Delta, \end{aligned} \tag{4.4}$$

we deduce, from (4.3), that

$$\begin{aligned}\Pr[c_x = c_y] &\geq \frac{\max\{\xi,\eta\} - d'(v)}{\max\{\xi,\eta\}} \\ &\geq 1 - \frac{d'(v)}{q-\Delta},\end{aligned}$$

conditional on v being selected in step (1). Thus

$$\begin{aligned}\Pr\left[|D_{t+1}| = |D_t| + 1\right] &= \frac{1}{n}\sum_{v\in A_t} \Pr\left[c_x \neq c_y \;\middle|\; v \text{ selected}\right] \\ &\leq \frac{1}{n}\sum_{v\in A_t}\frac{d'(v)}{q-\Delta} = \frac{m'}{(q-\Delta)n}. \qquad (4.5)\end{aligned}$$

Now consider the probability that $|D_{t+1}| = |D_t| - 1$. For this event to occur, the vertex v selected in step 1 must lie in D_t, and the new colours c_x and c_y selected in step 2 must be the same. The analogues of inequalities (4.4) in this case are

$$\begin{aligned}\xi - \zeta &\leq \Delta - d'(v), \\ \eta - \zeta &\leq \Delta - d'(v), \quad \text{and} \\ \xi, \eta &\geq q - \Delta.\end{aligned}$$

Proceeding as in the previous case,

$$\begin{aligned}\Pr[c_x = c_y] &\geq \frac{\max\{\xi,\eta\} - \Delta + d'(v)}{\max\{\xi,\eta\}} \\ &= 1 - \frac{\Delta - d'(v)}{\max\{\xi,\eta\}} \\ &\geq \frac{q - 2\Delta + d'(v)}{q-\Delta},\end{aligned}$$

conditional on v being selected in step (1). Hence

$$\begin{aligned}\Pr\left[|D_{t+1}| = |D_t| - 1\right] &\geq \frac{1}{n}\sum_{v\in D_t}\frac{q - 2\Delta + d'(v)}{q-\Delta} \\ &\geq \frac{q-2\Delta}{(q-\Delta)n}|D_t| + \frac{m'}{(q-\Delta)n} \qquad (4.6)\end{aligned}$$

Define

$$a = \frac{q-2\Delta}{(q-\Delta)n} \quad \text{and} \quad b = b(m') = \frac{m'}{(q-\Delta)n},$$

so that $\Pr\left[|D_{t+1}| = |D_t|+1\right] \leq b$ and $\Pr\left[|D_{t+1}| = |D_t|-1\right] \geq a|D_t|+b$. Provided $a > 0$, i.e., $q > 2\Delta$, the size of the set D_t tends to decrease with t, and hence,

intuitively at least, the event $D_t = \emptyset$ should occur with high probability for some $t \leq T$ with T not too large. Since $D_t = \emptyset$ is precisely the event that coalescence has occurred, it only remains to confirm this intuition, and quantify the rate at which D_t converges to the empty set. From equations (4.5) and (4.6),

$$\begin{aligned} \mathrm{E}\left[\,|D_{t+1}| \;\middle|\; D_t\right] &\leq b(|D_t|+1) + (a|D_t|+b)(|D_t|-1) \\ &\quad + (1-a|D_t|-2b)|D_t| \\ &= (1-a)|D_t|. \end{aligned}$$

Thus $\mathrm{E}\,|D_t| \leq (1-a)^t|D_0| \leq (1-a)^t n$, and, because $|D_t|$ is a non-negative integer random variable, $\Pr[\,|D_t| \neq 0] \leq n(1-a)^t \leq ne^{-at}$. Note that $\Pr[D_t \neq \emptyset] \leq \varepsilon$, provided $t \geq a^{-1}\ln(n\varepsilon^{-1})$, establishing the result. □

Remark 4.13. With a little care, the argument can be pushed to $q = 2\Delta$, though the bound on mixing time worsens by a factor of about n^2. (The r.v. D_t behaves in the boundary case rather like an unbiased random walk, and therefore its expected time to reach the origin $D_t = 0$ is longer; refer, e.g., to Dyer and Greenhill [24], in particular their Theorem 2.1.)

The (direct) coupling technique described here has been used in a number of other applications, such as approximately counting independent sets in a low-degree graph (Luby and Vigoda [46])[3] and estimating the volume of a convex body (Bubley, Dyer and Jerrum [15]).[4] In practice, the versatility of the approach is limited by our ability to design couplings that work well in situations of algorithmic interest. The next section reports on a new technique that promises to extend the effective range of the coupling argument by providing us with a powerful design tool.

4.3 Path coupling

The coupling technique described and illustrated in the previous section is conceptually very simple and appealing. However, in applying the method to concrete situations we face a technical difficulty, which began to surface even in §4.2: how do we encourage (X_t) and (Y_t) to coalesce, while satisfying the demanding constraints (4.2)? Path coupling is an engineering solution to this problem, invented by Bubley and Dyer [11, 12]. Their idea is to define the coupling only on pairs of "adjacent" states, for which the task of satisfying (4.2) is relatively easy, and then to extend the coupling to arbitrary pairs of states by composition of adjacent couplings along a path. The approach is not entirely distinct from classical coupling, and the Coupling Lemma still plays a vital role.

[3] Though the subsequent journal article [47] uses the more sophisticated path coupling method, which will be described presently.

[4] The latter application draws inspiration from Lindvall and Rodgers's [43] idea of coupling diffusions by reflection.

1. Select $p \in [n-1]$ according to the distribution f, and $r \in \{0,1\}$ u.a.r.
2. If $r = 1$ and $X_0 \circ (p, p+1) \in \Omega$, then $X_1 := X_0 \circ (p, p+1)$; otherwise, $X_1 := X_0$.

Figure 4.4: Trial defining an MC on linear extensions of a partial order $\prec$.

We illustrate path coupling in the context of a MC on linear extensions of a partial order. We are given a partially ordered set $(V, \prec)$, where $V = [n] = \{0, 1, \ldots, n-1\}$. Denote by $\operatorname{Sym} V$ the symmetric group on V. We are interested in sampling, u.a.r., a member of the set

$$\Omega = \left\{ g \in \operatorname{Sym} V : g(i) \prec g(j) \Rightarrow i \leq j, \text{ for all } i, j \in V \right\}$$

of linear extensions of $\prec$. In forming a mental picture of the the set Ω, the following characterisation may be helpful: $g \in \Omega$ iff the linear order

$$g(0) \sqsubset g(1) \sqsubset \cdots \sqsubset g(n-1) \tag{4.7}$$

extends, or is consistent with, the partial order $\prec$.

As usual, we propose to sample from Ω by constructing an ergodic MC on state space Ω, whose stationary distribution is uniform. The transitions from one linear extension to another are obtained by pre-composing the current linear extension with a random transition $(p, p+1)$. Instead of selecting $p \in [n-1]$ uniformly, we select p from a probability distribution f on $[n-1]$ that gives greater weight to values near the centre of the range. It is possible that this refinement actually reduces the mixing time; in any case, it leads to a simplification of the proof. Formally, the transition probabilities of the MC are defined by the experimental trial presented in Figure 4.4. Note that composition "$\circ$" is to be read right to left, so that (assuming $r = 1$): $X_1(p) = X_0(p+1)$, $X_1(p+1) = X_0(p)$ and $X_1(i) = X_0(i)$, for all $i \notin \{p, p+1\}$.

Provided the probability distribution f is supported on the whole interval $[n-1]$, this MC is irreducible and aperiodic. It is easy to verify, for example using Lemma 3.7, that the stationary distribution of the MC is uniform. As in §3.3, the explicit loop probability of $\frac{1}{2}$ is introduced mainly for convenience in the proof. However, some such mechanism for destroying periodicity is necessary in any case if we wish to treat the empty partial order consistently.

Our analysis of the mixing time of the MC using path coupling will closely follow that of Bubley and Dyer [13]. To apply path coupling, we need first to decide on an adjacency structure for the state space Ω. In this instance we decree that two states g and g' (linear extensions of $\prec$) are adjacent iff $g' = g \circ (i, j)$ for some transposition (i, j) with $0 \leq i < j \leq n-1$; in this case, the distance $d(g, g')$ from g to g' is defined to be $j - i$. Note that the notions of adjacency and distance are symmetric with respect to interchanging g and g', so we can regard this imposed

1. Select $p \in [n-1]$ according to the distribution f, and $r_x \in \{0,1\}$ u.a.r. If $j - i = 1$ and $p = i$, set $r_y := 1 - r_x$; otherwise, set $r_y := r_x$.
2. If $r_x = 1$ and $X_0 \circ (p, p+1) \in \Omega$ then set $X_1 := X_0 \circ (p, p+1)$; otherwise, set $X_1 := X_0$.
3. If $r_y = 1$ and $Y_0 \circ (p, p+1) \in \Omega$ then set $Y_1 := Y_0 \circ (p, p+1)$; otherwise, set $Y_1 := Y_0$.

Figure 4.5: A possible coupling for the MC on linear extensions.

adjacency structure as a weighted, undirected graph on vertex set Ω; let us refer to this structure as the adjacency graph. It is easily verified that the shortest path in the adjacency graph between two adjacent states is the direct one using a single edge. Thus d may be extended to a metric on Ω by defining $d(g,h)$, for arbitrary states g and h, to be the length of a shortest path from g to h in the adjacency graph.

Next we define the coupling. We need to do this just for adjacent states, as the extension of the coupling via shortest paths to arbitrary pairs of states will be automatic. Suppose that $(X_0, Y_0) \in \Omega^2$ is a pair of states related by $Y_0 = X_0 \circ (i,j)$ for some transposition (i,j) with $0 \le i < j \le n-1$. then the transition to (X_1, Y_1) in the coupling is defined by the experimental trial presented in Figure 4.5. We need to show:

Lemma 4.14. *For adjacent states X_0 and Y_0,*

$$\mathrm{E}\left[d(X_1, Y_1) \mid X_0, Y_0\right] \le \varrho\, d(X_0, Y_0), \tag{4.8}$$

where $\varrho < 1$ is a constant depending on f. For a suitable choice for f, one has $\varrho = 1 - \alpha$, where $\alpha = 6/(n^3 - n)$.

Informally, Lemma 4.14 says that distance between pairs of states in the coupled process tends to decrease: exactly the situation we encountered earlier in the context of the MC on q-colourings. Before proceeding with the proof of Lemma 4.14, let us pause to consider why it is sufficient to establish (4.8) just for adjacent states.

Lemma 4.15. *Suppose a coupling (X_t, Y_t) has been defined, as above, on adjacent pairs of states, and suppose that the coupling satisfies the contraction condition (4.8) on adjacent pairs. Then the coupling can be extended to all pairs of states in such a way that (4.8) holds unconditionally.*

Proof. Suppose $X_0 = x_0 \in \Omega$ and $Y_0 = y_0 \in \Omega$ are now arbitrary. Denote by $P(\cdot,\cdot)$ the transition probabilities of the MC on linear extensions. Let $x_0 = z^{(0)}, z^{(1)}, \ldots, z^{(\ell)} = y_0$ be a shortest path from x_0 to y_0 in the adjacency graph.

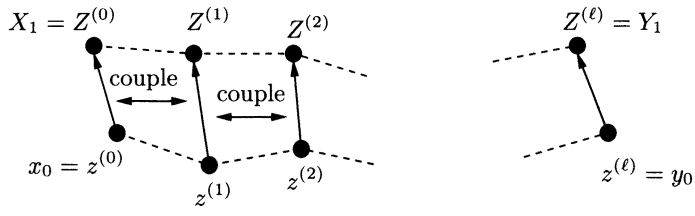

Figure 4.6: Extending a coupling along a shortest path

(Assume a deterministic choice rule for resolving ties.) First select $Z^{(0)} \in \Omega$ according to the probability distribution $P(z^{(0)}, \cdot\,)$. Now select $Z^{(1)}$ according to the probability distribution induced by the transition $(z^{(0)}, z^{(1)}) \mapsto (Z^{(0)}, Z^{(1)})$ in the coupled process, conditioned on the choice of $Z^{(0)}$; then select $Z^{(2)}$ according to the probability distribution induced by the transition $(z^{(1)}, z^{(2)}) \mapsto (Z^{(1)}, Z^{(2)})$, conditioned on the choice of $Z^{(1)}$; and so on, ending with $Z^{(\ell)}$. (The procedure is visualised in Figure 4.6.)

Let $X_1 = Z^{(0)}$ and $Y_1 = Z^{(\ell)}$. It is routine to verify, by induction on path length ℓ, that Y_1 has been selected according to the (correct) distribution $P(y_0, \cdot\,)$. Moreover, by linearity of expectation and (4.8),

$$\begin{aligned} \mathrm{E}\left[d(X_1, Y_1) \mid X_0 = x_0, Y_0 = y_0\right] &\le \sum_{i=0}^{\ell-1} \mathrm{E}\, d(Z^{(i)}, Z^{(i+1)}) \\ &\le \varrho \sum_{i=0}^{\ell-1} d(z^{(i)}, z^{(i+1)}) \\ &= \varrho\, d(x_0, y_0). \end{aligned}$$

□

So we see that it is enough to establish the contraction property (4.8) for adjacent pairs of states.

Proof of Lemma 4.14. If $p \notin \{i-1, i, j-1, j\}$ then the tests made in steps (2) and (3) either both succeed or both fail. Thus $Y_1 = X_1 \circ (i, j)$ and $d(X_1, Y_1) = j - i = d(X_0, Y_0)$. Summarising:

$$d(X_1, Y_1) = d(X_0, Y_0), \quad \text{if } p \notin \{i-1, i, j-1, j\}. \tag{4.9}$$

Next suppose $p = i-1$ or $p = j$. These cases are symmetrical, so we consider only the former. With probability at least $\frac{1}{2}$, the tests made in steps (2) and (3)

both fail, since $\Pr[r_x = r_y = 0] = \frac{1}{2}$. If this happens, clearly, $d(X_1, Y_1) = j - i = d(X_0, Y_0)$. Otherwise, with probability at most $\frac{1}{2}$, one or other test succeeds. If they both succeed, then

$$\begin{aligned} Y_1 &= Y_0 \circ (i-1, i) \\ &= X_0 \circ (i, j) \circ (i-1, i) \\ &= X_1 \circ (i-1, i) \circ (i, j) \circ (i-1, i) \\ &= X_1 \circ (i-1, j), \end{aligned}$$

and $d(X_1, Y_1) = j - i + 1 = d(X_0, Y_0) + 1$; if only one (say the one in step 2) succeeds, then $Y_1 = Y_0 = X_0 \circ (i, j) = X_1 \circ (i-1, i) \circ (i, j)$, and $d(X_1, Y_1) \leq j - i + 1 = d(X_0, Y_0) + 1$. Summarising:

$$\mathrm{E}\left[d(X_1, Y_1) \mid X_0, Y_0, p = i - 1 \vee p = j\right] \leq d(X_0, Y_0) + \frac{1}{2}. \tag{4.10}$$

Finally suppose $p = i$ or $p = j-1$. Again, by symmetry, we need only consider the former. There are two subcases, depending on the value of $j - i$. The easier subcase is $j - i = 1$. If $r_x = 1$ then $r_y = 0$ and

$$X_1 = X_0 \circ (i, i+1) = Y_0 \circ (i, i+1) \circ (i, i+1) = Y_0 = Y_1,$$

with a similar conclusion when $r_x = 0$. Thus $d(X_1, Y_1) = 0 = d(X_0, Y_0) - 1$. The slightly harder subcase is the complementary $j - i \geq 2$. The crucial observation is that $X_0 \circ (i, i+1), Y_0 \circ (i, i+1) \in \Omega$ and hence the tests in steps (2) and (3) either both succeed or both fail, depending only on the value of $r_x = r_y$. To see this, observe that

$$X_0(i) \not\succ X_0(i+1) = Y_0(i+1) \not\succ Y_0(j) = X_0(i),$$

from which we may read off the fact that $X_0(i)$ and $X_0(i+1)$ are incomparable in $\prec$. The same argument applies equally to $Y_0(i)$ and $Y_0(i+1)$. If $r_x = 0$ there is no change in state; otherwise, if $r_x = 1$,

$$\begin{aligned} X_1 &= X_0 \circ (i, i+1) \\ &= Y_0 \circ (i, j) \circ (i, i+1) \\ &= Y_1 \circ (i, i+1) \circ (i, j) \circ (i, i+1) \\ &= Y_1 \circ (i+1, j), \end{aligned}$$

and $d(X_1, Y_1) = j - i - 1 = d(X_0, Y_0) - 1$. Summarising both the $j - i = 1$ and $j - i \geq 2$ subcases:

$$\mathrm{E}\left[d(X_1, Y_1) \mid X_0, Y_0, p = i \vee p = j - 1\right] \leq e(X_0, Y_0), \tag{4.11}$$

where

$$e(X_0, Y_0) = \begin{cases} 0, & \text{if } d(X_0, Y_0) = 1; \\ d(X_0, Y_0) - \frac{1}{2}, & \text{otherwise.} \end{cases}$$

Note that, in the case $j-i=1$, inequality (4.11) covers just one value of p, namely $p=i=j-1$, instead of two; however, this effect is exactly counterbalanced by an expected reduction in distance of 1 instead of just $\frac{1}{2}$. Combining (4.9)–(4.11) we obtain

$$\mathrm{E}\left[d(X_1,Y_1) \mid X_0,Y_0\right] \le d(X_0,Y_0) - \frac{-f(i-1)+f(i)+f(j-1)-f(j)}{2}.$$

Specialising the probability distribution $f(\cdot)$ to be $f(i) := \alpha(i+1)(n-i-1)$ — where $\alpha := 6/(n^3-n)$ is the appropriate normalising constant — we have, by direct calculation, $-f(i-1)+f(i)+f(j-1)-f(j) = 2\alpha(j-i)$. Since $d(X_0,Y_0) = j-i$, we obtain (4.8) with $\varrho = 1-\alpha$. □

From Lemmas 4.14 and 4.15 it is now a short step to:

Proposition 4.16 (Bubley and Dyer). *The mixing time of the MC on linear extensions (refer to Figure 4.4) is bounded by*

$$\tau(\varepsilon) \le (n^3-n)(2\ln n + \ln \varepsilon^{-1})/6.$$

Proof. By iteration, $\mathrm{E}\left[d(X_t,Y_t) \mid X_0,Y_0\right] \le \varrho^t d(X_0,Y_0)$. For any pair of linear extensions g and h, there is a path in the adjacency graph using only *adjacent* transpositions (i.e., length one edges) that swaps each incomparable pair at most once. Thus $d(X_0,Y_0) \le \binom{n}{2} \le n^2$, and

$$\Pr[X_t \ne Y_t] \le \mathrm{E}\, d(X_t,Y_t) \le (1-\alpha)^t n^2.$$

The latter quantity is less than ε, provided $t \ge (n^3-n)(2\ln n + \ln\varepsilon^{-1})/6$. The result follows directly from Lemma 4.7. □

David Wilson has recently derived a similar $O(n^3 \log n)$ bound on mixing time when f is uniform, i.e, when the transposition $(p,p+1)$ is selected u.a.r.

Exercises 4.17. 1. Use Proposition 4.16 to construct an FPRAS for linear extensions of a partial order.

2. Reprove Proposition 4.5 using path coupling. Note the significant simplification over the direct coupling proof.

New applications of path coupling are regularly being discovered. Bubley, Dyer and Greenhill [14] have presented an FPRAS for q-colourings of a low degree graph that extends the range of applicability of the one described earlier. They were able, for example, to approximate in polynomial time the number of 5-colourings of a graph of maximum degree 3, thus "beating the 2Δ bound" that appeared to exist following the result described in §4.1. Vigoda [66], in a path-coupling tour de force, was able to beat the 2Δ bound uniformly over all sufficiently large Δ; specifically, he proved rapid mixing whenever $q > \frac{11}{6}\Delta$. It is fair to say

that neither of these improvements would have been possible without the aid of path coupling.

Dyer and Greenhill have also considered independent sets in a low degree graph [25], and obtained a result similar to, but apparently incomparable with, that of Luby and Vigoda [47]. Bubley and Dyer (again) applied path coupling to establish rapid mixing of a natural Markov chain on sink-free orientations of an arbitrary graph [10]. McShine [50] presents a particularly elegant application of path coupling to sampling tournaments. One further example must suffice: Cooper and Frieze [17] have applied path coupling to analyse the "Swendsen-Wang process," which is commonly used to sample configurations of the "random cluster" or ferromagnetic Potts model in statistical physics.

Finally, for those who skipped Exercise 4.3, here is the missing proof.

Proof of Lemma 4.2. The claim is established by the following sequence of (in-) equalities:

$$\begin{aligned}
2\,\|P^{t+1}(x,\,\cdot\,)-\pi\|_{\mathrm{TV}} &= \sum_{y\in\Omega}\bigl|P^{t+1}(x,y)-\pi(y)\bigr| \\
&= \sum_{y\in\Omega}\left|\sum_{z\in\Omega}P^t(x,z)\,P(z,y)-\sum_{z\in\Omega}\pi(z)\,P(z,y)\right| \\
&\le \sum_{y\in\Omega}\sum_{z\in\Omega}\bigl|P^t(x,z)-\pi(z)\bigr|\,P(z,y) \qquad (4.12)\\
&= \sum_{z\in\Omega}\bigl|P^t(x,z)-\pi(z)\bigr|\sum_{y\in\Omega}P(z,y) \\
&= 2\,\|P^t(x,\,\cdot\,)-\pi\|_{\mathrm{TV}},
\end{aligned}$$

where (4.12) is the triangle inequality. □

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 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,

M. Jerrum *Counting, Sampling and Integrating: Algorithms and Complexity*

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

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

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

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 the vertices of G) 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 MCs 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 $\gamma_{xy} = (x = z_0, z_1, \ldots, z_\ell = y)$ 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

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

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

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

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

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

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

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

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; Sinclair). *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] whose proof in turn is inspired by Diaconis and Stroock [20].

$$
\begin{aligned}
2\operatorname{Var}_\pi f &= \sum_{x,y\in\Omega} \pi(x)\pi(y)\big(f(x)-f(y)\big)^2 \\
&= \sum_{x,y\in\Omega} \pi(x)\pi(y)\Bigg(\sum_{(u,v)\in\gamma_{xy}} 1\cdot\big(f(u)-f(v)\big)\Bigg)^2 && (5.8)\\
&\le \sum_{x,y\in\Omega} \pi(x)\pi(y)\,|\gamma_{xy}| \sum_{(u,v)\in\gamma_{xy}} \big(f(u)-f(v)\big)^2 && (5.9)\\
&= \sum_{u,v\in\Omega} \sum_{\substack{x,y:\\(u,v)\in\gamma_{xy}}} \pi(x)\pi(y)\,|\gamma_{xy}|\,\big(f(u)-f(v)\big)^2 \\
&= \sum_{u,v\in\Omega} \big(f(u)-f(v)\big)^2 \sum_{\substack{x,y:\\(u,v)\in\gamma_{xy}}} \pi(x)\pi(y)\,|\gamma_{xy}| \\
&\le \sum_{u,v\in\Omega} \big(f(u)-f(v)\big)^2\,\pi(u)P(u,v)\,\varrho && (5.10)\\
&= 2\varrho\,\mathcal{E}_\pi(f,f).
\end{aligned}
$$

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 $P_{zz} := \frac{1}{2}(I+P)$, 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 $P_{zz}f : \Omega \to \mathbb{R}$ denotes the function defined by

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

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:

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

[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 step (5.14) relies on the fact that the square of the expectation of a r.v. is no greater than the expectation of its square. To get at the right-hand side of (5.11) we use yet another expression for the variance of f:

$$\begin{aligned}\operatorname{Var}_\pi f &= \frac{1}{2}\sum_{x\in\Omega}\pi(x)f(x)^2 + \frac{1}{2}\sum_{y\in\Omega}\pi(y)f(y)^2 \\ &= \frac{1}{2}\sum_{x,y\in\Omega}\pi(x)f(x)^2P(x,y) + \frac{1}{2}\sum_{x,y\in\Omega}\pi(x)P(x,y)f(y)^2 \\ &= \frac{1}{2}\sum_{x,y\in\Omega}\pi(x)P(x,y)\big(f(x)^2 + f(y)^2\big). \qquad (5.15)\end{aligned}$$

Subtracting (5.14) from (5.15) yields

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

as required. □

Combining Theorem 5.2 and Theorem 5.6 gives:

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

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

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

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

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

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

This gives

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

On the other hand,

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

which implies

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

for all A. This in turn means that the total variation distance $\|P_{zz}^t(x, \cdot) - \pi\|_{\mathrm{TV}}$ is bounded by ε, and we obtain the following corollary:

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

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

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*

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

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.

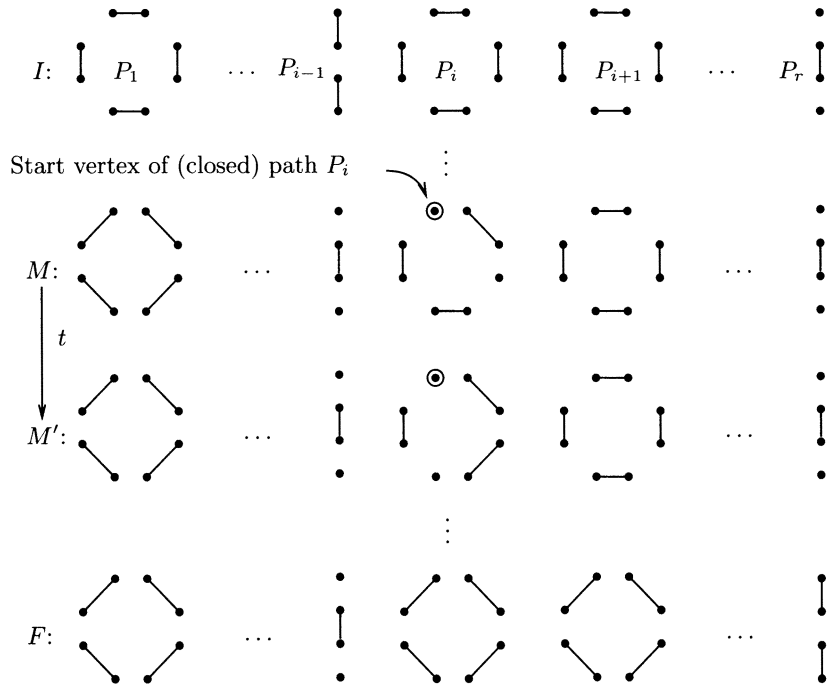

Figure 5.2: A step in a canonical path between matchings

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 r connected components of $I \oplus F$ are denoted $P_1, \ldots, P_r$.

To get from I to F, we now process the components of $(V, I \oplus F)$ in the order $P_1, \ldots, P_r$. 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 thought of as 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

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

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 congestion

$$\varrho = \max_{t=(M,M')} \left\{ \frac{1}{\pi(M)P(M,M')} \sum_{(I,F)\in \mathrm{cp}(t)} \pi(I)\pi(F)\,|\gamma_{IF}| \right\} \tag{5.16}$$

(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

$$\pi(I)\pi(F) \lessapprox \pi(M)P(M,M')\,\pi(\eta_t(I,F)), \tag{5.17}$$

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 follow that

$$\begin{aligned}
\varrho &\lessapprox \max_t \left\{ \sum_{(I,F)\in \mathrm{cp}(t)} \pi(\eta_t(I,F))\,|\gamma_{IF}| \right\} && \text{from (5.16) and (5.17)} \\
&\lessapprox \max_t \left\{ \sum_{(I,F)\in \mathrm{cp}(t)} \pi(\eta_t(I,F)) \right\} && \text{since } |\gamma_{IF}| \le n \\
&\le 1 && \text{since } \eta_t \text{ is injective.}
\end{aligned}$$

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

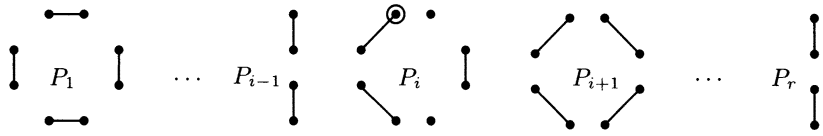

Figure 5.3: The corresponding encoding $\eta_t(X, Y)$.

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 $P_1, P_2, \ldots, P_r$ 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 $P_{i+1}, \ldots, P_r$. 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 $(I, F) \in \mathrm{cp}(t)$,*

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

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,

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

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

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

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

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

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 $\varrho \leq nm\bar{\lambda}^2$, 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 (5.17).

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

□

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 $\ln \pi(x)^{-1} \leq n \ln n + \frac{1}{2} n\,|\ln \lambda|$, 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

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

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

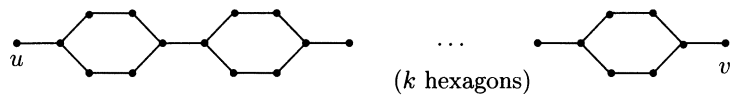

Figure 5.4: A graph with many "near perfect" matchings.

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 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.