Chapter 7

Inapproximability

Not all counting problems are efficiently approximable. We open with a simple example.

Fact 7.1. *Unless* RP = NP *there can be no FPRAS for the number of Hamilton cycles in a graph* G.

Informally: assuming, as seems likely, that there exist predicates in NP that admit no polynomial-time randomised algorithm, then no FPRAS for Hamilton cycles can exist. Still informally: the reason is that an FPRAS for Hamilton cycles would, in particular, need to distinguish the zero from non-zero case with reasonable probability.

To apply a rigorous interpretation to Fact 7.1, we need to divert briefly into randomised complexity classes, in particular RP and BPP. A predicate $\varphi : \Sigma^* \to \{0,1\}$ is in the class RP if there is a polynomial-time witness-checking predicate[1] $\chi : \Sigma^* \times \Sigma^* \to \{0,1\}$ and a polynomial p such that:

(i) if $\neg\varphi(x)$ then $\neg\chi(x,w)$ for all putative witnesses $w \in \Sigma^{p(|x|)}$;

(ii) if $\varphi(x)$ then $\Pr[\chi(x,w)] \geq \frac{1}{2}$, where w is assumed to be chosen u.a.r. from the set $\Sigma^{p(|x|)}$.

The predicate φ is in the class BPP if there exist χ and p, as above, satisfying:

(i′) if $\neg\varphi(x)$ then $\Pr[\chi(x,w)] \leq \frac{1}{4}$;

(ii′) if $\varphi(x)$ then $\Pr[\chi(x,w)] \geq \frac{3}{4}$,

[1] Refer to Chapter 2 for the general setting.

where, again, w is assumed to be chosen u.a.r. from the set $\Sigma^{p(|x|)}$. Thus RP (respectively, BPP) is the set of predicates that can be decided in randomised polynomial time with one-sided (respectively, two-sided) error.

Remarks 7.2. (a) There is no significance in the exact thresholds $\frac{1}{2}$, $\frac{1}{4}$ and $\frac{3}{4}$ appearing in the above definitions. By designing appropriate simulations, one can show that $\frac{1}{2}$ can be replaced by any constant strictly between 0 and 1, and $\frac{1}{4}$ and $\frac{3}{4}$ by any constants c_1, c_2 with $0 < c_1 < c_2 < 1$.

(b) It is immediate from the definition of RP that RP $\subseteq$ NP. No similar inclusion is known for BPP.

Now, comparing the definition of BPP with that of FPRAS, we see that the existence of an FPRAS for the number of Hamilton cycles in a graph G would immediately imply that the decision problem — is G Hamiltonian? — is in BPP. Since the decision problem is NP-complete, it would follow that NP $\subseteq$ BPP. The apparently stronger conclusion RP $=$ NP follows from the complexity-theoretic fact:

Fact 7.3. *If* NP $\subseteq$ BPP *then* NP $\subseteq$ RP *(and hence* RP $=$ NP*).*

See, e.g., Papadimitriou's textbook [67, Problem 11.5.18].

Remark 7.4. The converse to Fact 7.1 is also true: if RP $=$ NP then there is an FPRAS for the number of Hamilton cycles in a graph. Whereas Fact 7.1 is trivial, its converse is not, relying as it does on the bisection method of Valiant and Vazirani [77]. See Chapter 10 of Goldreich's lecture notes [38].

Of course, Hamiltonicity is not a distinguished NP-complete problem. More generally we have:

Fact 7.5. *(Informal statement.) If the decision version of a counting problem is* NP*-complete, then the counting problem itself does not admit an FPRAS unless* RP $=$ NP*.*

Exercise 7.6. Provide a formal statement of Fact 7.5 using the notion of witness-checking predicates.

Fact 7.5 instantly yields a large number of counting problems that, for a rather trivial reason, do not admit an FPRAS (under a reasonable complexity-theoretic assumption). We now turn to an example that does not admit an FPRAS for some non-trivial (though only slightly non-trivial) reason.

Let us consider the independent sets counting problem:

Name. #IS.

Instance. A graph G.

Output. The number of independent sets[2] of all sizes in G.

The decision version of #IS is trivial, since every graph has the empty set of vertices as an independent set. Nevertheless, we shall see that #IS is hard to approximate under some reasonable complexity-theoretic assumption. We shall make use of the optimisation version of #IS:

Name. MaxIS.

Instance. A graph G.

Output. The size of a maximum independent set in G.

MaxIS is a classical NP-complete[3] problem: see, e.g., Garey and Johnson [36, GT20].

Proposition 7.7. *There is no FPRAS for* #IS *unless* RP = NP.

Proof. We use a reduction from MaxIS. Let $G = (V, E)$ be an instance of MaxIS. We want to construct a graph $G' = (V', E')$, being an instance of #IS, in such a way that *typical* independent sets in G' reveal *maximum* independent sets in G.

The construction replaces vertices by blocks of r vertices and edges by complete bipartite graphs between blocks; formally,

$$V' = V \times \{0, \dots, r-1\},$$

and

$$E' = \big\{\{(u,i),(v,j)\} : \{u,v\} \in E \text{ and } i,j \in \{0 \dots r-1\}\big\}.$$

(See Figure 7.1.)

Each independent set I' in G' projects to an independent set

$$I = \big\{v \in V : \text{there exists } i \in \{0 \dots r-1\} \text{ such that } (v,i) \in I'\big\}$$

in G. (Since each edge of G corresponds to a complete bipartite subgraph in G', the set I is indeed independent in G.) Suppose $|I| = k$; then there are $(2^r - 1)^k$ independent sets I' in G' that project to the specific independent set I in G. We consider the two complementary situations:

---

[2] An independent set in graph G is a subset $U \subseteq V(G)$ of the vertex set of G such that no edge of G has both endpoints in U.

[3] To make formal sense of this claim, one would need to make MaxIS into a decision problem. This could be done, in the usual way, by specifying a bound $k \in \mathbb{N}$ as part of the problem instance and asking whether G has an independent set of size at least k.

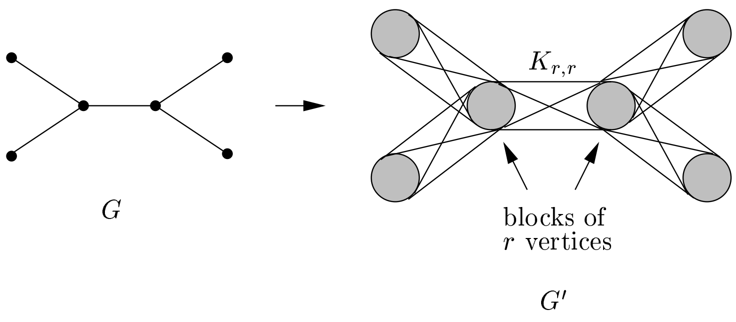

Figure 7.1: The construction.

(a) An independent set of size k exists in G. Then there are at least $(2^r - 1)^k$ independent sets in G'.

(b) The maximum independent set in G has size less than k. Then there are at most $2^n(2^r - 1)^{k-1}$ independent sets in G', where $n = |V|$.

Setting $r = n + 2$, we have

$$(2^r - 1)^k = (2^{n+2} - 1)(2^r - 1)^{k-1} \geq 2 \times 2^n(2^r - 1)^{k-1};$$

in other words, the minimum possible number of independent sets in case (a) exceeds the maximum possible number in case (b) by a factor 2. An FPRAS for #IS would be able to distinguish cases (a) and (b) with high probability, providing us with a polynomial-time randomised algorithm (with two-sided error) for MAXIS. As we have seen, this would imply RP = NP. $\square$

Remark 7.8. Note that the reduction proves something much stronger than the non-existence of an FPRAS for #IS. It shows (under the assumption RP $\neq$ NP) that there in no polynomial time randomised algorithm that approximates the number of independent sets even to within any fixed exponential factor. To see this, simply set $r = cn$ with $c > 1$. The statement can be strengthened even further: see Dyer, Frieze and Jerrum [27].

7.1 Independent sets in a low degree graph

Proposition 7.7 is evidence that the number of independent sets in a graph is hard to approximate in general, so we need to restrict the class of problem instances to make progress. One interesting way to do this is to place a bound Δ on the maximum degree of the instance G. Then we can investigate how the computational difficulty of of #IS varies as Δ does. On the positive side we have the following result.

Theorem 7.9 (Luby and Vigoda). *There is an FPRAS for* #IS *when* $\Delta = 4$.

Proof (sketch). As usual, it is enough to be able to sample independent sets almost uniformly at random in polynomial time.

Independent sets are sampled using an MC based on edge updates. View an independent set I in graph $G = (V, E)$ as a function $I : V \to \{0, 1\}$, where $I(v) = 1$ has the interpretation that v is in the independent set. The state space of the MC is the set of all independent sets in G. Transition probabilities are specified by the following trial, where $X_0 : V \to \{0, 1\}$ is the initial independent set.

1. Choose an edge $\{u, w\} \in E$, u.a.r.
2. Begin to construct a new independent set I as follows: with equal probability ($\frac{1}{3}$ in each case) set (a) $I(u) := 0$ and $I(w) := 0$; (b) $I(u) := 0$ and $I(w) := 1$; or (c) $I(u) := 1$ and $I(w) := 0$. (Note that these three cases correspond to the three possible restrictions of an independent set in G to the edge $\{u, w\}$.)
3. For all $v \in V \setminus \{u, w\}$ set $I(v) := X_0(v)$.
4. If I is an independent set then $X_1 := I$, otherwise $X_1 := X_0$.

Informally, we are using edge-updates with Metropolis acceptance probabilities.

This MC can be shown to be rapidly mixing using the path-coupling method. Two independent sets are considered to be adjacent if they differ at exactly one vertex. If adjacent independent sets are considered to be at distance 1, the derived path-metric is just Hamming distance. Suppose X_0 and Y_0 are adjacent; on the basis of a case analysis of moderate complexity it is possible to conclude that the expected Hamming distance between X_1 and Y_1 is at most 1. (For a regular graph with no small cycles there are four "good edges" $\{u, w\}$ whose selection may cause the distance to decrease, and twelve "bad edges" which may cause the distance to increase. In the worst case, these two effects are exactly in balance.) It follows that the mixing time of the MC scales quadratically with n. $\square$

Exercise 7.10. Complete the proof of Theorem 7.9. To keep technical complexity to a minimum, assume the graph G is triangle-free, i.e., contains no cycles of length 3. In case you need to refer to it, a complete analysis (in a more general setting where vertices in the independent set are given weight or "fugacity" λ) is given by Luby and Vigoda [58]. Theorem 7.9 corresponds to the case $\lambda = 1$ of their result. Dyer and Greenhill [30] also obtain a generalisation of Theorem 7.9, using a slightly different MC. Their proof has the advantage of dispensing with triangle-freeness.

According to Theorem 7.9, approximately counting independent sets in a graph G is tractable provided the maximum degree $\varDelta$ is small enough. We know that $\varDelta = 4$ is small enough, so what about $\varDelta = 5, 6, \ldots$? The reduction described in Proposition 7.7 constructs graphs of arbitrarily large degree, so it apparently leaves open the possibility that there is an FPRAS for #IS for any fixed degree bound $\varDelta$. However, if we look afresh at the construction of Theorem 7.9 in the light of inapproximability results for the optimisation problem MaxIS, we discover that there is a definite upper bound on $\varDelta$. This idea is due to Luby and Vigoda [58].

Proposition 7.11. *There is no FPRAS for* #IS *when* $\varDelta = 1188$*, unless* RP $=$ NP.

Proof. We know that MaxIS is NP-hard when restricted to graphs of maximum degree 4. A result of Berman and Karpinski [6, Thm 1(iv)] tells us more: for any $\varepsilon > 0$, it is NP-hard to determine the size of a maximum independent set in a graph G to within ratio of $\frac{73}{74} + \varepsilon$, even when G is restricted to have maximum degree 4. (By "determining the size… within ratio ϱ" we mean computing a number $\hat{k}$ such that $\varrho k \le \hat{k} \le k$, where k is the size of a maximum independent set in G.) In other words, the problem MaxIS is polynomial-time (Turing) reducible to the approximate version of MaxIS, in which we ask for a result within ratio $\frac{73}{74} + \varepsilon$. This result, like many other inapproximability results for optimisation problems, rests on the powerful theory of probabilistically checkable proofs (PCP).

So let G be a graph of maximum degree 4. Using our construction from the proof of Theorem 7.7 with $r = 297$, we obtain a graph G' of maximum degree 1188. We shall see that even a rough approximation to the *number* of independent sets in G' will provide a close (within ratio $\frac{73}{74} + \varepsilon$) approximation to the *size* of the largest independent set in G. Thus the existence of an FPRAS for #IS in graphs of maximum degree 1188 would imply the existence of a polynomial-time randomised algorithm (with two-sided error) for MaxIS. As before, this would in turn imply RP $=$ NP.

We define J' to be the collection of all independent sets in G'. Let k be the size of a maximum independent set in G. We have

$$(2^r - 1)^k \le |J'| \le 2^n (2^r - 1)^k,$$

or, taking the natural logarithm,

$$k \ln(2^r - 1) \le \ln |J'| \le n \ln 2 + k \ln(2^r - 1).$$

Consider the following estimate for k:

$$\hat{k} = \frac{\ln |J'| - n \ln 2}{\ln(2^r - 1)};$$

it is clear that

$$k - \frac{n \ln 2}{\ln(2^r - 1)} \le \hat{k} \le k.$$

Recall that Brooks's theorem [8, 10] asserts that any graph of maximum degree $\Delta \ge 3$ that does not contain $K_{\Delta+1}$ as a connected component is Δ-colourable. Assuming, as we may, that G is connected, it follows that G is 4-colourable. Since any (and hence in particular the largest) of the four colour classes is an independent set, $k \ge n/4$. Thus

$$k \left(1 - \frac{4 \ln 2}{\ln(2^r - 1)}\right) \le \hat{k} \le k.$$

Note that, when $r = 297$,

$$\frac{4 \ln 2}{\ln(2^r - 1)} < \frac{1}{74}.$$

If we had an FPRAS for #IS restricted to graphs of maximum degree 1188 then we would be able to approximate $|J'|$ (with high probability) within arbitrarily small constant relative error, and $\ln |J'|$ (and hence $\hat{k}$) within arbitrarily small constant additive error. But this in turn would provide an approximation to the size of the largest independent set in G (with high probability) within ratio $\frac{73}{74} + \varepsilon$. $\square$

One might suspect that the degree bound $\Delta = 1188$ in Proposition 7.11 is quite a bit larger than necessary, and this is indeed the case. Indeed, simply by tightening the analysis of the construction used in the proof of Proposition 7.11, one can reduce the degree Δ in its statement by 10–20%.

Exercise 7.12. Using the same reduction, but improved estimates, show that Proposition 7.11 holds for some Δ less than 1100. (I think $\Delta = 964$ is achievable.)

Using a technically more involved reduction, Dyer, Frieze and Jerrum have shown that $\Delta = 1188$ may be replaced by $\Delta = 25$. That still leaves a large gap between what is known to be tractable ($\Delta = 4$) and intractable ($\Delta = 25$); no doubt the upper bound could be reduced slightly at the expense of additional technical complexity, but a substantial gap would still remain.

To explore further the boundary between tractable and intractable requires us, at present, to accept more circumstantial evidence. Consider any MC on independent sets of a graph on n vertices. Let $b(n) \le n$ be any function of n and suppose the Hamming distance between successive states X_t and X_{t-1} of the MC is uniformly bounded by $b(n)$. We will say that the MC is $b(n)$-*cautious*. (Recall that we are viewing independent sets as functions $V \to \{0,1\}$.) Thus a $b(n)$-cautious MC is not permitted to change the status of more than $b(n)$ vertices in G at any step. Ideally, for ease of implementation, we would wish to have $b(n)$ a constant (as in the proposals of Luby and Vigoda [58], and Dyer and Greenhill [30]). However, we are able show that no $b(n)$-cautious chain on independent sets can mix rapidly unless $b(n) = \Omega(n)$, even when $\Delta = 6$. Thus any chain that *does* mix rapidly on graphs of maximum degree 6 must change the status of a sizeable proportion of the vertices at each step.

Theorem 7.13 (Dyer, Frieze and Jerrum). *There exists an infinite family of regular bipartite graphs of degree* 6*, together with constants* $\delta, \gamma > 0$*, such that the following is true: any* δn*-cautious MC on independent sets of these graphs has exponential mixing time, in the sense that* $\tau\left(\frac{1}{4}\right) = \Omega(\exp(\gamma n))$.

Dyer, Frieze and Jerrum's proof of Theorem 7.13 provides an explicit value for δ, namely $\delta = 0.35$. We present a simplified version of the proof here that does not attempt to estimate δ. The idea underlying the proof is very simple: if the state space of an MC has a tight "constriction" then its mixing time will be long. This intuition may be formalised as follows.

Claim 7.14. *Consider an MC with state space* Ω*, transition matrix* P*, and stationary distribution* π*. Let* $A \subset \Omega$ *be a set of states such that* $\pi(A) \le \frac{1}{2}$*, and* $M \subset \Omega$ *be a set of states that forms a "barrier" in the sense that* $P(i,j) = 0$ *whenever* $i \in A \setminus M$ *and* $j \in \overline{A} \setminus M$*. Then the mixing time* τ *of the MC satisfies* $\tau\left(\frac{1}{4}\right) \ge \pi(A)/4\pi(M)$.

We defer the proof of the claim to the end of the chapter.

Proof of Theorem 7.13. Our counterexample to rapid mixing (or, more precisely, family of counterexamples indexed by n) is a random regular bipartite graph G of degree $\Delta = 6$, with n vertices on the left and n on the right. Denote the left and right vertex sets by V_1 and V_2 respectively. The random graph model is simple. A *pairing* is one of the $n!$ possible bijections between left and right vertices viewed as a regular bipartite graph of degree 1. Select Δ pairings, independently and u.a.r., and form the union: the result is a bipartite graph G of maximum degree Δ. Since the pairings may not be disjoint, the graph G may not be regular; we return to this point later.

Let $J(\alpha, \beta)$ be the collection of all independent sets in G having αn vertices on the left and βn on the right. For a given set of αn vertices $U_1 \subseteq V_1$ and βn vertices $U_2 \subseteq V_2$, what is the probability that a random pairing will avoid joining some element in U_1 to some element in U_2? Well, the "image" of U_1 under the pairing is a random αn-subset of V_2, so the answer is the same as the probability that a random αn-subset of V_2 is disjoint from U_2; but the latter probability is just

$$\binom{(1-\beta)n}{\alpha n} \bigg/ \binom{n}{\alpha n}.$$

Thus the expected size of $J(\alpha, \beta)$ for a random G chosen according to our model is just

$$\mathbb{E}\,|J(\alpha,\beta)| = \binom{n}{\alpha n}\binom{n}{\beta n}\left[\binom{(1-\beta)n}{\alpha n} \bigg/ \binom{n}{\alpha n}\right]^{\Delta}.$$

(By linearity of expectation, the required quantity is simply the number of possible candidates (U_1, U_2), times the probability that all Δ pairings avoid connecting U_1 and U_2.) By Stirling's approximation we have

$$\mathbb{E}\,|J(\alpha,\beta)| = \exp\big(\varphi(\alpha,\beta)\, n(1+o(1))\big)$$

where

$$\begin{aligned} \varphi(\alpha,\beta) &= -\alpha\ln\alpha - \beta\ln\beta - \Delta(1-\alpha-\beta)\ln(1-\alpha-\beta) \\ &\quad + (\Delta-1)\big((1-\alpha)\ln(1-\alpha) + (1-\beta)\ln(1-\beta)\big). \end{aligned} \tag{7.1}$$

We treat φ as a function of real arguments α and β, even though a combinatorial interpretation is possible only when αn and βn are integers. Then φ is defined on the triangle

$$\mathcal{T} = \big\{(\alpha,\beta) : \alpha, \beta \geq 0 \text{ and } \alpha + \beta \leq 1\big\},$$

and is clearly symmetrical in α, β. (The function φ is defined by equation (7.1) on the interior of $\mathcal{T}$, and can be extended to the boundary by taking limits.)

Now set $\Delta = 6$. By calculus, $\varphi(\alpha, \alpha)$ has a unique maximum in the range $[0, \frac{1}{2})$; numerically $\varphi(\alpha, \alpha)$ is uniformly less than 0.704 in this range. Consider the region $\mathcal{D} = \{(\alpha, \beta) \in \mathcal{T} : |\alpha - \beta| \leq \delta\}$, where δ is a small positive constant. (This is the δ in the statement of the theorem.) For sufficiently small $\delta > 0$,

$$\varphi(\alpha, \beta) \leq 0.705, \quad \text{for all } (\alpha, \beta) \in \mathcal{D}.$$

For, if not, there would be an infinite sequence (α_i, β_i) of points in $\mathcal{T}$, all satisfying $\varphi(\alpha, \beta) > 0.705$, which approach the diagonal $\alpha = \beta$ arbitrarily closely. By compactness, there would be a subsequence of (α_i, β_i) converging to some point on the diagonal, contradicting continuity of φ. So, by Markov's inequality, with very high probability,[4]

$$\left| \bigcup_{(\alpha,\beta)\in\mathcal{D}} J(\alpha, \beta) \right| \leq e^{0.706n}, \tag{7.2}$$

where the union is over α, β which are multiples of $1/n$.

Denote by $\mathcal{L}$ and $\mathcal{R}$ the two connected regions of $\mathcal{T} \setminus \mathcal{D}$. We need a lower bound on the number of independent sets in these regions which exceeds the upper bound (7.2). With this in mind, define

$$\theta(\alpha) = -\alpha \ln \alpha - (1 - \alpha) \ln(1 - \alpha) + (\ln 2)(1 - \Delta\alpha).$$

for $\alpha < \Delta^{-1}$. Then, for *any* graph G in the space of random graphs, the total number of independent sets I with $|I \cap V_1| = \alpha n$ is (crudely) at least

$$|J(\alpha, *)| \geq \binom{n}{\alpha n} 2^{(1-\Delta\alpha)n} = \exp\big(\theta(\alpha)\, n(1 - o(1))\big).$$

(Choose αn vertices from V_1; then choose any subset of vertices from the at least $(1 - \Delta\alpha)n$ unblocked vertices in V_2.) Set $\Delta = 6$ as before and $\alpha^* = 0.015$. Then, by numerical computation, $\theta(\alpha^*)$ is greater than 0.708. In other words,

$$\left| \bigcup_{(\alpha,\beta)\in\mathcal{L}} J(\alpha, \beta) \right| \geq e^{0.708n}, \tag{7.3}$$

[4] "With very high probability" may be taken to mean "with probability differing from 1 by an amount decaying exponentially fast with n."

for all sufficiently large n, with a similar bound for $\mathcal{R}$. Comparing (7.2) and (7.3), we see that, with very high probability, the number of approximately balanced independent sets is smaller, by an exponential factor, than the number with a sizeable imbalance in either direction. Specifically, the former is smaller than the latter by a factor $e^{\gamma n}$, where $\gamma = 0.002$.

The $(n+n)$-vertex graph whose existence is guaranteed by Theorem 7.13 (ignoring for a moment the regularity requirement) is any graph from the space of random graphs under consideration that exhibits the exponential gap just described. (A randomly chosen graph will do with high probability.) The remainder of our argument concerns such a graph.

Now consider a δn-cautious MC. Let $A = \bigcup_{\alpha \geq \beta} J(\alpha, \beta)$ denote the set of leftward leaning independent sets, and assume, without loss of generality, that A is no larger than its complement $\overline{A} = \Omega \setminus A$. Denote by M the set of approximately balanced independent sets $M = \bigcup_{(\alpha,\beta)\in\mathcal{D}} J(\alpha, \beta)$.

Since the MC is δn-cautious, it cannot make a transition from A to $\overline{A}$ directly, but only by using intermediate states in M. Now, we know from inequalities (7.2) and (7.3) that

$$|A| \geq e^{\gamma n}\,|M|. \tag{7.4}$$

If we are prepared to weaken the theorem slightly by dropping the condition that the graphs be regular, we can immediately complete the proof by appealing to Claim 7.14.

We may address the regularity issue by reference to a standard result about the union-of-pairings model for random bipartite graphs. Provided Δ is taken as constant, Bender [5] has shown that Δ-regular graphs occur in our random graph model with probability bounded away from 0. Since we prove that random graphs of maximum degree 6, with very high probability, have the property we seek, it follows that random Δ-regular graphs (in the induced probability space), with very high probability, have the property too. $\square$

It only remains to present the missing proof.

Proof of Claim 7.14. Denote by π_t the t-step distribution of the MC. First note that

$$\begin{aligned}\|\pi_{t+1} - \pi_t\|_{\mathrm{TV}} = \|\pi_t P - \pi_{t-1} P\|_{\mathrm{TV}} &= \frac{1}{2} \max_{\|z\|_\infty \leq 1} (\pi_t - \pi_{t-1}) P z \\ &\leq \frac{1}{2} \max_{\|w\|_\infty \leq 1} (\pi_t - \pi_{t-1}) w \\ &= \|\pi_t - \pi_{t-1}\|_{\mathrm{TV}},\end{aligned}$$

since $\|Pz\|_\infty \le \|z\|_\infty$. Hence, by induction, $\|\pi_{t+1} - \pi_t\|_{\mathrm{TV}} \le \|\pi_1 - \pi_0\|_{\mathrm{TV}}$ and, further, using the triangle inequality, $\|\pi_t - \pi_0\|_{\mathrm{TV}} \le t\,\|\pi_1 - \pi_0\|_{\mathrm{TV}}$. Now, for $\emptyset \subset S \subset \Omega$, define

$$\Phi(S) = \frac{1}{\pi(S)} \sum_{i\in S} \sum_{j\in \overline{S}} \pi(i)P(i,j).$$

The quantity $\Phi = \min\{\Phi(S) : S \subset \Omega \text{ and } 0 < \pi(S) \le \frac{1}{2}\}$ is sometimes called the "conductance" of the MC. (Conductance is normally considered in the context of time-reversible Markov chains. However, both the definition and the line of argument employed here apply to non-time-reversible chains.) Now

$$\begin{aligned}\sum_{i\in A}\sum_{j\in\overline{A}} \pi(i)P(i,j) &\le \sum_{i\in A}\sum_{j\in\overline{A}\cap M} \pi(i)P(i,j) + \sum_{i\in A\cap M}\sum_{j\in\overline{A}} \pi(i)P(i,j)\\ &\le \pi(\overline{A}\cap M) + \pi(A\cap M)\\ &= \pi(M).\end{aligned}$$

In short, $\Phi(A)\,\pi(A) \le \pi(M)$. So setting

$$\pi_0(i) = \begin{cases} \pi(i)/\pi(A), & \text{if } i \in A;\\ 0, & \text{otherwise,}\end{cases}$$

we have

$$\|\pi_1 - \pi_0\|_{\mathrm{TV}} = \frac{1}{2}\sum_{j\in\Omega}\left|\sum_{i\in\Omega}\pi_0(i)P(i,j) - \pi_0(j)\right| \tag{7.5}$$

$$= \sum_{j\in\overline{A}}\sum_{i\in A}\pi_0(i)P(i,j) \tag{7.6}$$

$$= \Phi(A).$$

(To see equality (7.6), observe that the terms in (7.5) with $j \in A$ make a contribution to the sum that is equal to that made by the terms with $j \in \overline{A}$. Now simply restrict the sum to terms with $j \in \overline{A}$.) But $\|\pi_0 - \pi\|_{\mathrm{TV}} \ge \frac{1}{2}$, since $\pi(A) \le \frac{1}{2}$, and hence

$$\|\pi_t - \pi\|_{\mathrm{TV}} \ge \|\pi_0 - \pi\|_{\mathrm{TV}} - \|\pi_t - \pi_0\|_{\mathrm{TV}} \ge \frac{1}{2} - t\,\Phi(A).$$

Thus we cannot achieve $\|\pi_t - \pi\|_{\mathrm{TV}} \le \frac{1}{4}$ until

$$t \ge \frac{1}{4\,\Phi(A)} \ge \frac{\pi(A)}{4\pi(M)}.$$

By an averaging argument there must exist some initial state $x_0 \in A$ for which $\tau_{x_0}\left(\frac{1}{4}\right) \geq \pi(A)/4\pi(M)$. $\square$