Ph.D. course on Network Dynamics, Fall 2013

Lectures 1 and 2: Random walks and Markov chains

1 Introduction

Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ be an *undirected graph* of finite size $n := |\mathcal{V}|$. Throughout, we shall assume that $\mathcal{G}$ is *connected*, i..e, every pair of nodes in $\mathcal{V}$ can be joined by a path. Let $A \in \{0,1\}^{\mathcal{V}\times\mathcal{V}}$ be its adjacency matrix, defined by $A_{uv} = 1$ if $\{u,v\} \in \mathcal{E}$ and $A_{uv} = 0$ if $\{u,v\} \notin \mathcal{E}$. A *random walk* on $\mathcal{G}$ is a discrete-time Markov chain $V(t)$ with state space $\mathcal{V}$ and transition probability matrix $P = D^{-1}A$, where $D := \operatorname{diag}(d)$ and $d = A\mathbb{1}$ is the degree vector of $\mathcal{G}$. In simple words, at each time $t = 0, 1, \ldots$, the random walk $V(t)$ jumps from a node $v \in \mathcal{V}$ to one of its neighbors in $\mathcal{G}$ with uniform probability.

In fact, one can consider more general Markov chains than random walks. Let P be a *stochastic matrix* on $\mathcal{V}$, i.e., $P_{uv} \geq 0$ for all $u, v \in \mathcal{V}$, and $P\mathbb{1} = \mathbb{1}$. Then, a discrete-time Markov chain with transition probability matrix P is a stochastic process $V(t)$ on $\mathcal{V}$ such that

$$\mathbb{P}(V(t+1) = v | V(0), V(1), \ldots, V(t) = u) = P_{uv}\,, \qquad \forall u, v \in \mathcal{V}\,, t \geq 0\,.$$

Observe that, if $\mu(t)$ is the probability distribution of $V(t)$, whose entries are given by $\mu_v(t) := \mathbb{P}(V(t) = v)$, and $P(t)$ stands for the t-step transition probability matrix, whose entries are given by $P_{ij}(t) = \mathbb{P}(V(t) = j | V(0) = i)$, then one has

$$\mu(t+1) = P'\mu(t)\,, \qquad P(t+1) = PP(t)\,. \tag{1}$$

In particular, (1) implies that $\mu(t) = (P')^t\mu(0)$ and $P(t) = P^t$, for $t \geq 0$.

A probability distribution π (i.e., a nonnegative vector such that $\mathbb{1}'\pi = 1$) is called *stationary* (or *invariant*) for a stochastic matrix P if

$$P'\pi = \pi \,.$$

The probabilistic interpretation is that, if $V(0)$ has distribution π, then so does $V(1)$, and in fact $V(t)$ for every time $t \geq 0$. Existence of a stationary distribution π for every stochastic matrix P is a standard fact.[1] For example, a one line computation (check it!) shows that for the random walk on an undirected graph $\mathcal{G}$, $\pi = d/\mathbb{1}'d$ is a stationary distribution (in fact the unique one, since we assumed $\mathcal{G}$ to be connected, see below). For general stochastic matrices P there is not such an easy explicit formula for π. Note that a stochastic matrix P has uniform stationary distribution $\pi = n^{-1}\mathbb{1}$ if and only if $P'\mathbb{1} = \mathbb{1}$, i.e., if not only it sums up to 1 on each row, but it does so also on each column. In this case the matrix P is called *doubly stochastic*. This is a very special case! In particular, the stochastic matrix associated to the random walk on a graph $\mathcal{G}$ is doubly stochastic if and only if the graph is regular, i.e., every node has the same degree.

On the other hand, if a stochastic matrix P is *irreducible* (i.e., if the directed graph obtained by putting a link from u to v whenever $P_{uv} > 0$ is strongly connected), then the stationary distribution π is unique (in fact, it also has all strictly positive entries –prove it as an exercise!). To prove uniqueness of the stationary distribution, observe that it is equivalent to $P - I$ having rank $n - 1$, which in turns is equivalent to

$$(I - P)x = 0 \tag{2}$$

having only constant solutions. But this easily follows from considering the maximum entry of x, say x_u, and observing that $x_u = \sum_v P_{uv}x_v$ implies that x_v are also maximal, and the argument can be iterated until covering the whole node set $\mathcal{V}$ (that one does not get stack earlier follows exactly from irreducibility). In fact, the matrix $I - P$ is often called the *Laplacian* matrix, and solutions of equation (2) are called *harmonic* vectors.

In fact, if P is a stochastic irreducible matrix, then not only is its stationary distribution π unique, but one also has that $(P')^t\mu \to \pi$ for all probability distributions μ. I.e., irrespective of the distribution of its initial state $V(0)$,

[1] E.g., since P' maps the simplex of probability distributions (i.e., the convex compact set of those nonnegative vectors x such that $\mathbb{1}'x = 1$) in itself it should have a fixed point. (Brouwer's fixed-point theorem) Alternatively, one can use the Perron-Frobenius theory.

the Markov chain $V(t)$ will be distributed according to π in the limit of large t. We will not prove this for general irreducible stochastic matrices now (one can find such a proof in standard textbooks) but rather concentrate on a special class of stochastic matrices, namely time-reversible matrices, for which we will get explicit bounds on the speed of convergence in terms of eigenvalues of P.

A stochastic matrix P is called time-*reversible* (or simply reversible) if there exists a probability distribution π such that

$$\pi_u P_{uv} = \pi_v P_{vu}\,, \qquad u, v \in \mathcal{V}\,. \tag{3}$$

Observe that the condition above implies in particular that π is invariant. The probabilistic interpretation, which also explains the terminology, is that, if $V(0)$ has distribution π satisfying (3), then the Markov chain $V(t)$ with transition probability matrix P is such that, for all $t \geq 0$, the vector $(V(0), V(1), \ldots, V(t-1), V(t))$ has the same joint probability distribution of the time-reversed vector $(V(t), V(t-1), \ldots, V(1), V(0))$. (You can verify this coincides with (3) for $t = 1$, and then generalize it by induction on t).

In particular, one can easily check that the stochastic matrix $P = D^{-1}A$ associated to the random walk on an undirected graph $\mathcal{G}$ is reversible, using its stationary distribution $\pi = d/(\mathbb{1}'d)$. More in general, one can consider a connected weighted graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, W)$, where W is a weight matrix such that $W_{uv} = W_{vu} > 0$ for all $\{u, v\} \in \mathbb{E}$ and $W_{uv} = 0$ if $\{u, v\} \notin \mathcal{E}$, put $d = W\mathbb{1}$, and $D := \mathrm{diag}\,(d)$. Then, $P = D^{-1}W$ is a reversible stochastic matrix with stationary distribution $\pi = d/(\mathbb{1}'d)$. However, (3) is not satisfied by every stochastic matrix P with stationary distribution π: to get a counterexample, simply try with the stochastic matrix associated to the random walk on a directed graph. In fact, in a sense, the subclass of reversible stochastic matrices is as rich as the subclass of undirected weighted graphs. To every (connected) and possibly directed weighted graph one can associate in a natural way an (irreducible) stochastic matrix, which turns out to be reversible if and only if the weighted graph was undirected.

Random walks and Markov chains can also be considered in continuous time. A continuous-time random walk $V(t)$ on an undirected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ is a Markov random process on $\mathcal{V}$ spends a mean-1 exponentially distributed random time on a node v, then jumping to a neighbor node uniformly chosen among the neighbors of v. More in general, to every transition-rate matrix $Q \in \mathbb{R}^{\mathcal{V} \times \mathcal{V}}$ with nonnegative off-diagonal entries and such that $Q\mathbb{1} = 0$, we can associate a continuous-time Markov chain on $\mathcal{V}$ such that, conditioned $V(t) = u$, the chain will wait a mean-$|Q_{uu}|$ exponentially distributed random time and then jump to another node $v \in \mathcal{V}$ randomly chosen with probability Q_{uv}. The probability distribution $\mu(t)$ and the time-t transition probability matrix $K(t)$ of the Markov chain $V(t)$ will then satisfy the differential equations

$$\frac{\mathrm{d}}{\mathrm{d}t}\mu(t) = Q'\mu(t)\,, \qquad \frac{\mathrm{d}}{\mathrm{d}t}K(t) = QK(t)\,. \tag{4}$$

The above are the continuous-time analogous of (1) and are sometimes referred to as Komogorov's forward and, respectively, backward, equations. In particular, they imply that $\mu(t) = \exp(Q'(t))\mu(0)$ and $K(t) = \exp(tQ)$ for $t \geq 0$. A probability distribution π is stationary of $Q'\pi = 0$, and the chain is reversible if $\pi_u Q_{uv} = \pi_v Q_{uv}$ for all $u \neq v \in \mathcal{V}$.

2 Spectral gap of reversible Markov chains

For an irreducible reversible stochastic matrix P with stationary probability distribution π, define the matrix

$$M := \Pi^{1/2} P \Pi^{-1/2}\,, \qquad \Pi := \mathrm{diag}\,(\pi)\,.$$

It is immediate to check that M is symmetric. Then, M (and hence P since they are similar matrices) has real eigenvalues $1 = \lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n \geq -1$. The following result provides an estimate of the rate of convergence to π for the probability distribution of a Markov chain started from arbitrary initial state.

Theorem 1. Let P be a reversible stochastic matrix with stationary distribution π. Then,

$$|P_{ij}(t) - \pi_j| \leq \lambda^t \sqrt{\frac{\pi_j}{\pi_i}}\,, \qquad i, j \in \mathcal{V}\,, \quad t \geq 0\,,$$

where $\lambda := \max\{\lambda_2, |\lambda_n|\}$ is the largest-in-module eigenvalue of P.

Proof. For $1 \leq k \leq n$, let $a_{(k)}$ be the eigenvector of M corresponding to eigenvalue λ_k. Since M is symmetric, such eigenvectors can be chosen to form an orthonormal basis of $\mathbb{R}^{\mathcal{V}}$, so that M admits the representation

$$M = \sum_{1 \leq k \leq n} \lambda_k a_{(k)} a'_{(k)}\,.$$

It follows that, for all $t \geq 0$,

$$\Pi^{1/2} P \Pi^{-1/2} = M^t = \sum_{1 \leq k \leq n} \lambda_k^t a_{(k)} a'_{(k)} \,. \tag{5}$$

Now, recall that $\lambda_1 = 1$ and check that $a_{(1)}$ can be chosen to have components $\sqrt{\pi_v}$, for all $v \in \mathcal{V}$. For $1 \leq i \leq n$, let $e_{(i)}$ be the vector whose i-th entry equals 1 and all whose other entries equal 0. By equating the (i,j)-th entry of the leftmost and rightmost side of (5), one gets

$$\begin{aligned} \sqrt{\pi_i} P_{ij}(t) \frac{1}{\sqrt{\pi_j}} - \sqrt{\pi_i \pi_j} &= \left| \sum_{k=2}^{n} \lambda_k^t e'_{(i)} a_{(k)} a'_{(k)} e_{(j)} \right| \\ &\leq \lambda^t \left(\sum_{k=2}^{n} (e'_{(i)} a_{(k)})^2 \right)^{1/2} \left(\sum_{k=2}^{n} (a'_{(k)} e_{(j)})^2 \right)^{1/2} \\ &\leq \lambda^t \,, \end{aligned} \tag{6}$$

where the first inequality follows from Cauchy-Schwartz, and the second one from the fact that $\sum_{k=1}^{n} (e'_{(i)} a_{(k)})^2 = ||e_{(i)}||_2^2 = 1$, since $\{a_{(k)} : 1 \leq k \leq n\}$ is an orthonormal basis of $\mathbb{R}^{\mathcal{V}}$. The claim now follows by multiplying both sides of (6) by $\sqrt{\pi_j / \pi_i}$. □

Observe that Theorem 1 guarantees that the probability distribution of a reversible Markov chain converges to π exponentially fast in time, provided that $\lambda_2 < 1$ and $\lambda_n > -1$. Now, $\lambda_2 < 1$ is equivalent to irreducibility of P, i.e., connectedness of the graph, which is our standing assumption. On the other hand, one could find irreducible reversible stochastic matrices P with $\lambda_n = 1$. This occurs if and only if the graph associated to P is bipartite, i.e., its node set $\mathcal{V}$ can be partitioned in two subsets such that links exist only between pairs of nodes belonging to different subset, but there are no links joining nodes from the same subset. Notice that, in this case, it is not Theorem 1 to be weak, but what really matters is that the probability distribution of the Markov chain does not converge to π. Indeed, at even time instants $t = 0, 2, 4, \ldots$, the chain $V(t)$ will always be in the node subset to which its initial state $V(0)$ belongs, while, at odd times $t = 1, 3, 5, \ldots$, it will be in the complementary node subset, and this prevents its distribution to converge. However, in practice one can often 'forget' about λ_n with the following trick. Define the *lazy* version of a stochastic matrix P as $P_{(\text{lazy})} = \frac{1}{2}(P+I)$. This is the transition probability matrix of a Markov chain which, conditioned on $V(t) = u$, waits a mean-2 geometrically distributed random time and then moves to another state v with probability P_{uv}. In practice, laziness has the effect of slowing down the original chain by a factor 2. Observe that the stationary distribution of $P_{(\text{lazy})}$ coincides with one of P, while their eigenvalues satisfy $\lambda_k^{(\text{lazy})} = (1+\lambda_k)/2$, hence they all belong to the interval $[0,1]$. Hence, Theorem 1 convergence of the probability distribution of the lazy version of the Markov chain with transition probability matrix P to the invariant distribution π at rate $((\lambda_2+1)/2)^t$.

The quantity $1-\lambda_2$ is called the *spectral gap* of P, while its inverse

$$\tau_{\text{rel}} := \frac{1}{1-\lambda_2}$$

is called the *relaxation time* of P. The spectral gap of a reversible stochastic matrix admits the following useful *variational characterization.* For two vectors $f, g \in \mathbb{R}^{\mathcal{V}}$, we define the *Dirichlet form*

$$\mathcal{E}(f,g) := \frac{1}{2} \sum_{u,v\in\mathcal{V}} \pi_u P_{uv}(f_u - f_v)(g_u - g_v)\,.$$

Then, the following result holds true

Theorem 2 (Variational characterization of the spectral gap)**.** Let P be a stochastic matrix reversible with respect to the distribution π, and let λ_2 be its second largest eigenvalue. Then,the spectral gap satisfies

$$1-\lambda_2 = \min\left\{\frac{\mathcal{E}(f,f)}{\sum_v \pi_v f_v^2} \,:\, f \neq 0,\, \pi' f = 0\right\}.$$

Proof. □

One of the great advantages of the variational characterization above is that it allows for estimating the spectral gap in terms of the geometry of the graph associated to the Markov chain. In particular, define the *conductance* of a reversible matrix P as

$$\Phi := \min_{\substack{\mathcal{U}\subseteq\mathcal{V}:\\ 0<\pi(\mathcal{U})\leq 1/2}} \frac{\sum_{u\in\mathcal{U}}\sum_{v\in\mathcal{V}\setminus\mathcal{U}} \pi_u P_{uv}}{\pi(\mathcal{U})}\,. \tag{7}$$

In the above, the denominator $\pi(\mathcal{U}) := \sum_{u\in\mathcal{U}} \pi_u$ is the size of the set $\mathcal{U}$, as measured by π, while the numerator $\sum_{u\in\mathcal{U}}\sum_{v\in\mathcal{V}\setminus\mathcal{U}} \pi_u P_{uv}$ can be thought of as measuring the strength of the interconnection between $\mathcal{U}$ and its complement set $\mathcal{V} \setminus \mathcal{U}$. In fact, the quantity Φ is often referred to as the *bottleneck ratio*, the *isoperimetric constant*, or the *Cheeger constant*. When P is the stochastic matrix associated to the random walk on $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, one has that

$$\Phi = \min_{0 < \frac{\mathrm{vol}(\mathcal{U})}{\mathbb{1}' d} \leq \frac{1}{2}} \frac{|\partial(\mathcal{U})|}{\mathrm{vol}(\mathcal{U})} \,,$$

where $\mathrm{vol}(\mathcal{U}) := \sum_{u \in \mathcal{U}} d_u$ stands for the volume of a set $\mathcal{U}$, while $\partial(\mathcal{U})$ denotes its boundary, i.e., the set of links connecting $\mathcal{U}$ to its complementary set $\mathcal{V} \setminus \mathcal{V}$.

Theorem 3 (Cheeger's inequality)**.** The spectral gap of a stochastic matrix P reversible with respect to a stationary distribution π satisfies

$$\frac{1}{2}\Phi^2 \leq 1 - \lambda_2 \leq 2\Phi \,,$$

In continuous time, the situation is somehow simpler since, in case of reversibility, convergence is dictated by the spectral gap of Q, without any need of introducing laziness. In particular, the following result is the continuous-time analogous of Theorem 1.

Theorem 4. The time-t transition probability matrix $K(t)$ of a continuous-time Markov chain with reversible transition rate matrix Q satisfies

$$|K_{ij}(t) - \pi_j| \leq \sqrt{\frac{\pi_j}{\pi_i}} \exp(-\beta t) \,, \qquad \forall i, j \in \mathcal{V},\, t \geq 0 \,,$$

where β is the second smallest eigenvalue of $-Q$.

Proof. □

3 Total variation distance and mixing time

In this section, we shall analyze the speed of convergence to equilibrium for not necessarily reversible stochastic matrices. Call a stochastic matrix P *aperiodic* if the maximum common divisor of the length of all the cycles in the directed graph associated to P equals 1. In particular, this is true if $P_{vv} > 0$ for at least one v (hence, e.g., for lazy chains). For simple random walks on an undirected graph this amounts to say that the graph is not bipartite. Throughout, we shall deal with irreducible aperiodic stochastic matrices P and denote by π their stationary distribution.

We start by introducing a metric for probability distributions. Let the total variation distance between two probability distributions μ and ν be defined as

$$||\nu - \mu||_{TV} = \frac{1}{2} \sum_v |\nu_v - \mu_v| \,.$$

Sometimes the definition is without the factor 1/2, however the convention we are choosing here will prove convenient later. In particular, it makes $||\mu - \nu||_{TV} \leq 1$. A simple but very useful property of the total variation distance is that (see [**?**, Proposition 4.2])

$$||\nu - \mu||_{TV} = \max_{\mathcal{U} \subseteq \mathcal{V}} \{\mu(\mathcal{U}) - \nu(\mathcal{U})\} \ ,$$

so that the total variation distance can be thought of as the maximum difference between the probability put on subset by the two distributions. Now, a key observation is that every stochastic matrix (not necessarily reversible or even irreducible) is *non-expansive* in total variation distance, i.e., $||P'\mu - P'\nu||_{TV} \leq ||\mu - \nu||_{TV}$ for all probability distributions μ and ν. When proving it, it becomes clear that, if for every u, v there exists some w such that $P_{uw}P_{uv} > 0$ (hence, in particular, if all the entries of P are strictly positive), then P is contractive in total variation, i.e., there exists $\alpha \in (0, 1)$ such that $||P'\mu - P'\nu||_{TV} \leq \alpha||\mu - \nu||_{TV}$ for all probability distributions μ and ν. We define the *mixing time* of P as

$$\tau_{\text{mix}} := \inf \left\{ t \geq 1 : ||(P^t)'\mu - (P^t)'\nu||_{TV} \leq e^{-1} \,, \ \forall \mu, \nu \right\} \,. \tag{8}$$

Here, the choice of the constant e^{-1} is rather arbitrary but convenient, replacing e^{-1} with any $\alpha \in (0, 1/2)$ would have worked as well. Observe that, because of convexity of the total variation distance separately in both arguments, one has that

$$\Delta(t) := \sup_{\mu,\nu} ||(P^t)'\mu - (P^t)'\nu||_{TV} = \max_{i,j \in \mathcal{V}} ||(P^t)_{i,\cdot} - (P^t)_{j,\cdot}||_{TV} \,.$$

Another important property is *submultiplicativity* of $\Delta(t)$, i.e., the fact that $\Delta(s + t) \leq \Delta(s)\Delta(t)$, for all $s, t \geq 0$. (see [**?**, Lemma 4.12]) In particular, this implies that

$$||(P^t)'\mu - \pi||_{TV} \leq ||(P^t)'\mu - (P^t)'\pi||_{TV} \leq \exp(-\lfloor t/\tau_{\text{mix}} \rfloor) \,,$$

so that τ_{mix} is the time of exponential decay to 0 of the distance between the probability distribution of an arbitrarily started chain and the stationary distribution.

For reversible matrices, Theorem 1 yields the following:

Corollary 1. Let P be a reversible stochastic matrix with stationary distribution π. Then,

$$\tau_{\text{mix}} \leq \left\lceil \frac{\log(2e/\pi_*)}{\log(\lambda^{-1})} \right\rceil ,$$

where $\pi_* := \min_v \pi_v$, and $\lambda := \max\{\lambda_2, -\lambda_n\}$. Hence, the mixing time of $P_{(\text{lazy})} = (P + I)/2$ satisfies

$$\tau_{\text{mix}}^{(\text{lazy})} \leq \lceil 2\tau_{\text{rel}} \log(2e/\pi_*) \rceil ,$$

where $\tau_{\text{rel}} := 1/(1 - \lambda_2)$ stands for the relaxation time of P.

Proof. Exercise □

Hence, for reversible Markov chains one can use Cheeger's inequality (i.e., Theorem 3) to get an upper bound on the mixing time in terms of the conductance

$$\tau_{\text{mix}}^{(\text{lazy})} \leq \left\lceil \frac{2\log(e^2/\pi_*)}{\Phi^2} \right\rceil . \tag{9}$$

In fact, the mixing time is finite for every, not necessarily reversible, aperiodic and irreducible stochastic matrix P (see, e.g., [**?**, Theorem 4.9]). In Section 5, we will learn another technique to derive upper bounds on the mixing time which works irrespective of any reversibility assumption. Before doing that, we propose the following lower bound on the mixing time in terms of conductance, which is valid without any reversibility assumption.

Theorem 5. Let P be an irreducible stochastic matrix with stationary distribution π. Then,

$$\tau_{\text{mix}} \geq C\frac{1}{\Phi} \tag{10}$$

where $C := (e - 2)/(2e)$.

Proof. □

Note that, besides constants, the righthand side of (9) differs from the one of (10) by the exponent 2 on the conductance Φ, as well as the multiplicative factor $\log(1/\pi_*)$. The former shows up because of the use of Cheeger's inequality for estimating the spectral gap, while the latter is due to the use of Theorem 1.

4 Hitting times

Let $V(t)$ be a Markov chain with transition probability matrix P. The hitting time and return time on a nonenmpty subset $\mathcal{A} \subseteq \mathcal{V}$ are the random variables

$$T_{\mathcal{A}} := \inf\{t \geq 0 : V(t) \in \mathcal{A}\}, \qquad T_{\mathcal{A}}^{+} := \inf\{t \geq 1 : V(t) \in \mathcal{A}\},$$

respectively. Let also

$$\gamma_{va} := \mathbb{P}(V(T_{\mathcal{A}}) = a | V(0) = v), \qquad v \in \mathcal{V},\, a \in \mathcal{A},$$

be the probability of hitting a specific node a before any other node in $\mathcal{A}$.

There are useful identities which allow one to compute hitting probabilities and expected hitting times. On the one hand, one has that

$$\gamma_{va} = \sum_{w} P_{vw}\gamma_{wa}, \quad \forall v \in \mathcal{V} \setminus \mathcal{A}, \qquad \gamma_{aa} = 1, \qquad \gamma_{ba} = 0, \quad \forall b \in \mathcal{A} \setminus \mathcal{A}.$$

The above can be rewritten more compactly as

$$\Upsilon_{\mathcal{V}\setminus\mathcal{A}}(I - P)\gamma = 0, \qquad \Upsilon_{\mathcal{A}}\gamma = I, \tag{11}$$

where $\gamma \in \mathbb{R}^{\mathcal{V}\times\mathcal{A}}$ is the matrix with entries γ_{av}, and $\Upsilon_{\mathcal{V}\setminus\mathcal{A}}$ and $\Upsilon_{\mathcal{A}}$ are the projection matrices on $\mathcal{A}$ and $\mathcal{V} \setminus \mathcal{A}$, respectively. The reader is encourages to compare equation (11) with (2).

On the other hand, if $\tau_v^{\mathcal{A}} := \mathbb{E}[T_{\mathcal{A}} | V(0) = v]$ stands for the expected hitting time on $\mathcal{A}$ when starting from node $v \in \mathcal{V}$, then

$$\tau_v^{\mathcal{A}} = 1 + \sum_{w} P_{vw}\tau_w^{\mathcal{A}}, \quad \forall v \in \mathcal{V} \setminus \mathcal{A}, \qquad \tau_a^{\mathcal{A}} = 0, \quad \forall a \in \mathcal{A},$$

which can be written more compactly as

$$\Upsilon_{\mathcal{V}\setminus\mathcal{A}}(I - P)\tau^{\mathcal{A}} = \mathbb{1}, \qquad \Upsilon_{\mathcal{A}}\tau^{\mathcal{A}} = 0.$$

Example 1. For the simple random walk on the line with node set $\{0, 1, \ldots, n\}$, and $\mathcal{A} := \{0, n\}$, one has $\gamma_{k,n} = k/n = 1 - \gamma_{0,n}$, and $\tau^{\mathcal{A}} = k(n-k)$ for all $k = 0, \ldots, n$.

5 Coupling

An approach to get upper bounds on the mixing time τ_{mix} alternative to the one in Section 3 is based on *coupling* techniques. The key idea of coupling is the following: Consider two Markov chains, $V_1(t)$ and $V_2(t)$, started in possibly different initial states $V_1(0), V_2(0) \in \mathcal{V}$, and moving simultaneously with the same, not necessarily reversible, transition probability matrix P. This means that the joint transition probabilities

$$Q_{(i,u)(j,v)} := \mathbb{P}(V_1(t+1) = j, V_2(t+1) = v | V_1(t) = i, V_2(t) = u)$$

satisfy

$$\sum_v Q_{(i,u)(j,v)} = P_{ij} \,, \qquad \sum_j Q_{(i,u)(j,v)} = P_{uv} \,, \qquad \forall i, j, u, v \in \mathcal{V} \,. \tag{12}$$

A *Markov coupling* for a stochastic matrix P is any stochastic matrix Q on the product space $\mathcal{V}^2$ satisfying (12). One Markov coupling is obviously the trivial one $Q_{(i,u)(j,v)} = P_{ij}P_{uv}$, which corresponds to $V_1(t)$ and $V_2(t)$ moving independently on $\mathcal{V}$ with transition probability matrix P. However, one can often find less trivial, and more useful, couplings.

The following is a basic result connecting coupling with mixing. Define the *coupling time*

$$T_{\text{couple}} := \min \{ t \geq 0 \,, V_1(t) = V_2(t) \} \,,$$

i.e., T_{couple} denotes the random time that $V_1(t)$ and $V_2(t)$ happen to be in the same node. Then, one has the following

Proposition 1. *Let P be a stochastic matrix on $\mathcal{V}$. For every coupling Q of two Markov chains $V_1(t)$ and $V_2(t)$ with transition probability matrix P,*

$$\Delta(t) \leq \max_{i,u} \{ \mathbb{P}(T_{couple} > t | V_1(0) = i, V_2(0) = u) \} \,, \qquad \forall t \geq 0 \,.$$

Proof. □