Chapter 3
Markov Chains

3.1 Introduction

Probabilistic methods have long been applied to solve discrete mathematics problems (see, for example, Erdős [39]–[40], and Alon and Spencer [3] for a recent and comprehensive treatment on probabilistic methods). Similarly, connections between Markov chains and graph theory have long been made (see Harary [59]). Our contribution here is to apply properties of Markov chains to the Hamiltonian cycle problem and to take advantage of the still emerging theory of perturbed Markov chains in this context. In Section 3.2, we give a brief introduction to Markov chains and various perturbations that we employ to obtain our results in this and subsequent chapters. More specifically, in Section 3.3, we present results on how fundamental matrices of Markov chains can be used to solve the Hamiltonian cycle problem, using their top-left matrix elements. In Section 3.4, we show that Hamiltonian cycles can be seen as variance minimisers of first hitting times and demonstrate a Hamiltonian gap that differentiates between the Hamiltonian and non-Hamiltonian graphs. Sections 3.3 and 3.4 make greater use of probabilistic methods than the remainder of this book. A reader unfamiliar with the latter could proceed to Chapter 4 and beyond with only a minimal loss of continuity.

3.2 Markov Chains and Perturbations

A *stochastic process* is a collection of random variables X_t, $t \geq 0$, which take values from a set $\mathcal{S}$ called a *state space*. There are two types of stochastic processes: *continuous-time* where $t \in [0, \infty)$, and *discrete-time* where $t \in \{0, 1, 2, \ldots\}$. In discrete-time case, each time point t is also called a *stage*. A stochastic process is *Markov* if the knowledge of the past does not influence the future, other than through the present. Mathematically, a stochastic pro-

V. S. Borkar et al., *Hamiltonian Cycle Problem and Markov Chains,*
International Series in Operations Research & Management Science 171,
DOI 10.1007/978-1-4614-3232-6_3, © Springer Science+Business Media, LLC 2012

cess is Markov if, given $t_0 \leq t_1 \leq \cdots \leq t_n$, the following equality is satisfied for every set $\{X_{t_n} \leq x_n\}$

$$P\{X_{t_n} \leq x_n | X_{t_{n-1}}, \ldots, X_{t_1}\} = P\{X_{t_n} \leq x_n | X_{t_{n-1}}\}, \text{ for all } n \in \mathbb{N}.$$

Such a stochastic process is a *Markov chain* if $\mathcal{S}$ is discrete, and a *Markov process* otherwise. Significant applications of Markov chains and processes include those in geostatistics, stock market fluctuations, population processes, modelling games of chance such as Monopoly and, most recently, internet search engines including Google. Here, we are only concerned with finite discrete-time Markov chains, where $\mathcal{S} = \{1, 2, \ldots, N\}$ with $N = |\mathcal{S}| < \infty$ and $t \in \{0, 1, \ldots\}$. We recommend Kemeny and Snell [70] for an excellent introduction to finite Markov chains.

In other words, a discrete-time stochastic process X_t is a Markov chain if, given any $s_1, \ldots, s_n \in \mathcal{S}$ and $t_0 \leq t_1 \leq \cdots \leq t_n$, the following equality holds

$$P\{X_{t_n} = s_n | X_{t_{n-1}} = s_{n-1}, \ldots, X_{t_1} = s_1\} = P\{X_{t_n} = s_n | X_{t_{n-1}} = s_{n-1}\},$$

for all $n \in \mathbb{N}$. The *one-step transition probability*

$$P_{ij}^{t,t+1} = P\{X_{t+1} = j | X_t = i\} \tag{3.1}$$

is the probability that the system moves from state i at time t to state j at time $t+1$. If this probability does not depend on t, that is, if for $t \neq t'$,

$$P_{ij}^{t,t+1} = P_{ij}^{t',t'+1},$$

then the Markov chain X_t is *time-homogeneous*, with *stationary transition probabilities* P_{ij}. Unless otherwise indicated, all Markov chains we deal with in this book are time-homogeneous. The $N \times N$ *one-step probability transition matrix* $\mathbf{P}$ has entries $p_{ij} = P_{ij}$, where $p_{ij} \geq 0$ and $\sum_{j \in \mathcal{S}} p_{ij} = 1$. Similarly, for $i, j \in \mathcal{S}$ and for $t = 0, 1, 2, \ldots$, we define the *n-step transition probability*

$$P_{ij}^{(n)} = P\{X_{t+n} = j | X_t = i\}. \tag{3.2}$$

This conditional probability does not depend on t as we assume that all Markov chains considered are time-homogeneous. The $N \times N$ *n-step probability transition matrix* $\mathbf{P}^{(n)}$ has entries $p_{ij}^{(n)} = P_{ij}^{(n)}$. For $m, n \geq 0$, the matrix form of the Chapman-Kolmogorov equation is given by

$$\mathbf{P}^{(n+m)} = \mathbf{P}^{(m)} \mathbf{P}^{(n)}. \tag{3.3}$$

Following easily from the definitions above, (3.3) implies that the n-step probability transition matrix $\mathbf{P}^{(n)}$ can be obtained by multiplying the one-step transition matrix $\mathbf{P}$ by itself n times.

Every probability transition matrix is a *stochastic matrix*, that is, a matrix where the elements of each row are nonnegative and sum to 1. Consequently, every stochastic matrix is a probability transition matrix of some Markov chain. A *doubly stochastic matrix* is a stochastic matrix in which not only every row but also every column has a sum of unity. A *doubly stochastic deterministic matrix* is a *doubly stochastic matrix* every element of which is either 1 or 0.

Types of Markov Chains Consider a discrete state space $\mathcal{S}$, possibly countably infinite. If it is not possible to leave some state $i \in \mathcal{S}$, that is, $p_{ii} = 1$, then state i is *absorbing*. After starting at state i, if every return to i occurs in multiples of n_i steps, then i has a *period* n_i. Formally, we define $n_i = \gcd\{t : P\{X_t = i | X_0 = i\} > 0\}$, where gcd is the greatest common divisor. A state i is *aperiodic* if $n_i = 1$, and *periodic* with period n_i if $n_i > 1$.

If the probability of never returning to state i after starting at state i is positive, then the state i is *transient*. Formally, $P\{\mathcal{T}_i = \infty\} > 0$, where $\mathcal{T}_i = \inf\{t \geq 1 : X_t = i | X_0 = i\}$ is a random variable representing the first time of returning to i, and is also known as the *first return time* of state i. A state i that is not transient is said to be *recurrent*. For a recurrent state i, if the expectation of the returning time $\mathcal{T}_i$ is finite, that is, $\mathbb{E}[\mathcal{T}_i] < \infty$, then the state i is *positive recurrent*. Otherwise, it is *null recurrent*. A Markov chain is *absorbing* if it has at least one absorbing state.

A Markov chain is *irreducible* if there is a path with positive probability to go from any state to any other state. In an irreducible chain, if one state is recurrent (respectively, positive recurrent or null recurrent) then all states are likewise recurrent (respectively, positive recurrent or null recurrent). Finally, a Markov chain is *ergodic* (also known as *regular*) if it is irreducible and every state is positive recurrent. For finite $\mathcal{S}$, every recurrent state is positive recurrent. In particular, all states of a finite state irreducible Markov chain are positive recurrent.

Distribution Vectors A *distribution vector* $\boldsymbol{\nu}$ is a row vector with nonnegative entries ν_i, the probability of the system being in state i. An *initial distribution vector* $\boldsymbol{\nu}^{(0)}$ is a row vector with entries $\nu_i^{(0)}$, the probability of starting at state i. The distribution vector $\boldsymbol{\nu}^{(n)}$ with entries $\nu_i^{(n)}$, the probability of being at state i after n steps, is given by

$$\boldsymbol{\nu}^{(n)} = \boldsymbol{\nu}^{(0)} \mathbf{P}^n. \tag{3.4}$$

Each of the vectors $\boldsymbol{\nu}, \boldsymbol{\nu}^{(0)}$, and $\boldsymbol{\nu}^{(n)}$ sums to 1.

Stationary Distribution Matrix The *stationary distribution matrix* $\mathbf{P}^*$ is defined as follows

$$\mathbf{P}^* = \lim_{T\to\infty} \frac{1}{T+1} \sum_{t=0}^{T} \mathbf{P}^t. \tag{3.5}$$

It is well-known (see Doob [29]) that the limit in (3.5) exists and $\mathbf{P}^*$ is also known as the *Cesaro-limit matrix*, as it is the long-run average of the powers of $\mathbf{P}$. Every stationary distribution matrix $\mathbf{P}^*$ satisfies the following identity

$$\mathbf{P}^*\mathbf{P} = \mathbf{P}\mathbf{P}^* = \mathbf{P}^*\mathbf{P}^* = \mathbf{P}^*. \tag{3.6}$$

For *ergodic* (or *regular*) Markov chains, the following three properties hold.

(i) If the chain is aperiodic, the stationary distribution matrix $\mathbf{P}^*$ is equivalent to

$$\mathbf{P}^* = \lim_{t\to\infty} \mathbf{P}^t. \tag{3.7}$$

(ii) Every row of $\mathbf{P}^*$ is identical.

(iii) Define $\boldsymbol{e} = (1, \ldots, 1)^{\mathrm{T}} \in \mathbb{R}^N$ and let every row of $\mathbf{P}^*$ be the row-vector $\boldsymbol{q}$. Then, all entries of $\boldsymbol{q}$ are strictly positive and $\boldsymbol{q}$ is the unique solution to the system of linear equations

$$\begin{aligned} \boldsymbol{q}\mathbf{P} &= \boldsymbol{q} \\ \boldsymbol{q}\boldsymbol{e} &= 1. \end{aligned} \tag{3.8}$$

In this case, the vector $\boldsymbol{q}$ is often called the *stationary distribution vector* of the Markov chain.

Let $\mathbf{J} = \boldsymbol{e}\boldsymbol{e}^{\mathrm{T}}$, an $N \times N$ matrix of which every entry is unity. If $\mathbf{P}$ is doubly stochastic and induces an irreducible Markov chain, then $\boldsymbol{q} = 1/N\boldsymbol{e}^{\mathrm{T}}$ satisfies the system of equations (3.8), and consequently,

$$\mathbf{P}^* = 1/N\mathbf{J}. \tag{3.9}$$

In the context of this book, the Markov chains that we are interested in always correspond to an arbitrary but fixed graph G of order N. Such a graph defines a family of *induced Markov chains* that correspond to a family of $N \times N$ probability transition matrices $\mathbf{P}$ such that the entries of the ith row form a probability mass function on the edges of G emanating from vertex i of the graph, for $i = 1, 2, \ldots, N$.

Example 3.1 *We consider a 5-vertex graph depicted in Figure 3.1, and two Markov chains, each with the state space* $\mathcal{S} = \{1, 2, 3, 4, 5\}$.

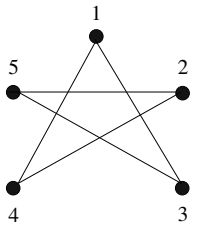

Fig. 3.1: A 5-vertex graph

In the first one, depicted in Figure 3.2, the system can travel from one state to exactly one other state. For example, from state 1, the system can go only to state 3. Therefore, the probability p_{13} *of going from state 1 to state 3 is 1, and* $p_{1i} = 0$ *for all* $i \neq 3, i \in \mathcal{S}$*. Of course, the corresponding graph displayed in Figure 3.2 is a spanning subgraph of the original graph from Figure 3.1.*

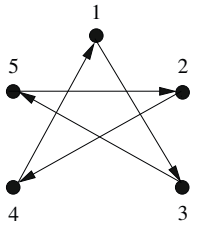

Fig. 3.2: An example of a finite discrete-time Markov chain with five states

The associated probability transition matrix $\mathbf{P}_1$ *is given by*

$$\mathbf{P}_1 = \begin{bmatrix} \cdot & \cdot & 1 & \cdot & \cdot \\ \cdot & \cdot & \cdot & 1 & \cdot \\ \cdot & \cdot & \cdot & \cdot & 1 \\ 1 & \cdot & \cdot & \cdot & \cdot \\ \cdot & 1 & \cdot & \cdot & \cdot \end{bmatrix}. \tag{3.10}$$

In this Markov chain, it is possible to travel from every state to any other state, hence it is irreducible. However, since every state has a period 5, the Markov chain is not aperiodic. Let $\boldsymbol{v} = (1/4, 1/4, 0, 1/4, 1/4)$ *be the initial distribution vector, then the distribution vector after one step is*

$$\boldsymbol{v}^{(1)} = \boldsymbol{v}\mathbf{P}_1 = [\,1/4\ 1/4\ 1/4\ 1/4\ 0\,].$$

After one step, the probabilities of the system being at states 1, 2, 3, 4 and 5 are $1/4, 1/4, 1/4, 1/4$ *and* 0*, respectively. The stationary distribution matrix* $\mathbf{P}_1^*$ *associated with this Markov chain is* $1/5\mathbf{J}$*.*

In the second Markov chain, depicted in Figure 3.3, from state 1, the system can travel to states 3 and 4, each with probability 1/2, and from state 5, the system can travel to state 2 with probability 1/5 and to state 3 with probability 4/5. From any other state, the system can only travel to exactly one other state.

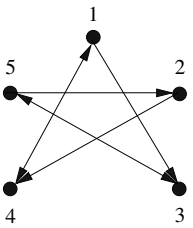

Fig. 3.3: Another example of a finite discrete-time Markov chain with five states

The associated probability transition matrix $\mathbf{P}_2$ *is given by*

$$\mathbf{P}_2 = \begin{bmatrix} \cdot & \cdot & \frac{1}{2} & \frac{1}{2} & \cdot \\ \cdot & \cdot & \cdot & 1 & \cdot \\ \cdot & \cdot & \cdot & \cdot & 1 \\ 1 & \cdot & \cdot & \cdot & \cdot \\ \cdot & \frac{1}{5} & \frac{4}{5} & \cdot & \cdot \end{bmatrix}. \tag{3.11}$$

This Markov chain is irreducible and aperiodic. With the initial distribution vector $\boldsymbol{v} = (1/4, 1/4, 0, 1/4, 1/4)$*, the distribution vector after one step is*

$$\boldsymbol{v}^{(1)} = \boldsymbol{v}\mathbf{P}_2 = [\,1/4\ 1/20\ 13/40\ 3/8\ 0\,].$$

The stationary distribution matrix $\mathbf{P}_2^*$ *associated with this Markov chain is* $\mathbf{P}_2^* = \boldsymbol{e}(2/15, 1/15, 1/3, 2/15, 1/3)$.

For every Markov chain, the matrix $\mathbf{I} - \mathbf{P} + \mathbf{P}^*$ is always invertible, and its inverse is called the *fundamental matrix*. Let $\mathbf{G}(\mathbf{P})$ be the fundamental matrix of a Markov chain specified by the probability transition matrix $\mathbf{P}$, then

$$\mathbf{G}(\mathbf{P}) = (\mathbf{I} - \mathbf{P} + \mathbf{P}^*)^{-1}. \tag{3.12}$$

Perturbations of Markov Chains We introduce two perturbations that have been applied to Markov chains in our line of research:

1. Symmetric linear perturbation ([34; 38; 37]): For a perturbation parameter $\varepsilon \in [0, 1)$ and an $N \times N$ probability transition matrix $\mathbf{P}$, the perturbed matrix $\mathbf{P}^\varepsilon$ is defined as

$$\mathbf{P}^\varepsilon = (1 - \varepsilon)\mathbf{P} + \varepsilon/N\mathbf{J}. \tag{3.13}$$

This symmetric linear perturbation ensures that the Markov chain specified by $\mathbf{P}^\varepsilon$ is always ergodic, while preserving double stochasticity whenever $\mathbf{P}$ is doubly stochastic.

2. Asymmetric linear perturbation ([49; 51; 35]): For a perturbation parameter $\varepsilon \in [0, 1)$ and an $N \times N$ probability transition matrix $\mathbf{P}$, the perturbed matrix $\mathbf{P}^\varepsilon$ is defined as

$$\mathbf{P}^\varepsilon = (1 - \varepsilon)\mathbf{P} + \varepsilon \begin{bmatrix} 1\,0 \cdots 0 \\ 0\,0 \cdots 0 \\ \vdots\,\vdots \quad \vdots \\ 0\,0 \cdots 0 \end{bmatrix} \mathbf{P} + \varepsilon \begin{bmatrix} 0\,0 \cdots 0 \\ 1\,0 \cdots 0 \\ \vdots\,\vdots \ddots \vdots \\ 1\,0 \cdots 0 \end{bmatrix}. \tag{3.14}$$

This asymmetric linear perturbation not only eliminates multiple ergodic classes but also differentiates vertex 1—referred to as the *home vertex* from Chapter 4 onwards—from other vertices. Additionally, it maintains roughly the level of sparsity of the original probability transition matrix $\mathbf{P}$.

Example 3.2 *We revisit a Markov chain introduced in Example 3.1, with the probability transition matrix* $\mathbf{P}_1$ *specified in (3.11)*

$$\mathbf{P}_1 = \begin{bmatrix} \cdot & \cdot & 1 & \cdot & \cdot \\ \cdot & \cdot & \cdot & 1 & \cdot \\ \cdot & \cdot & \cdot & \cdot & 1 \\ 1 & \cdot & \cdot & \cdot & \cdot \\ \cdot & 1 & \cdot & \cdot & \cdot \end{bmatrix}.$$

For $\varepsilon \in (0, 1)$*, applying the symmetric linear perturbation defined in (3.13), we obtain*

$$\mathbf{P}^\varepsilon = \begin{bmatrix} \rho & \rho & 1 - 4/5\varepsilon & \rho & \rho \\ \rho & \rho & \rho & 1 - 4/5\varepsilon & \rho \\ \rho & \rho & \rho & \rho & 1 - 4/5\varepsilon \\ 1 - 4/5\varepsilon & \rho & \rho & \rho & \rho \\ \rho & 1 - 4/5\varepsilon & \rho & \rho & \rho \end{bmatrix},$$

where $\rho = 1/5\varepsilon$. Applying the asymmetric linear perturbation defined in (3.14) gives us

$$\mathbf{P}^{\varepsilon} = \begin{bmatrix} \cdot & \cdot & 1 & \cdot & \cdot \\ \varepsilon & \cdot & \cdot & 1-\varepsilon & \cdot \\ \varepsilon & \cdot & \cdot & \cdot & 1-\varepsilon \\ 1 & \cdot & \cdot & \cdot & \cdot \\ \varepsilon & 1-\varepsilon & \cdot & \cdot & \cdot \end{bmatrix}.$$

3.3 Hitting Times and the Fundamental Matrix

In this section, we derive important relationships between entries of the fundamental matrix $\mathbf{G}(\mathbf{P})$ (associated with the Markov chain defined by the transition matrix $\mathbf{P}$) and the moments of the first hitting times, with respect to a given initial distribution. Indeed, we focus on τ_1, the random variable denoting the first return time of the home state/vertex 1, given that Markov chain starts at that vertex, and we denote by $Var[\tau_1]$ its variance. We show that whenever $\mathbf{P}$ is doubly stochastic the first diagonal entry of $\mathbf{G}(\mathbf{P})$ has the linear form $a + bVar[\tau_1]$, where a and b depend only on N. Of course, if we take vertex k as the home vertex, then the kth diagonal entry of $\mathbf{G}(\mathbf{P})$ can be shown to have an analogous linear form $a + bVar[\tau_k]$.

The above is significant because, subsequently, we shall show that the Hamiltonian cycle problem is equivalent to an optimisation problem, the objective of which is to minimise the top-left element of the fundamental matrix of the Markov chain permissible on a given graph, over the space of Markov chains on the graph that have doubly stochastic transition matrices. Therefore, the HCP is also equivalent to the problem of minimising the variance of τ_1 over the space of doubly stochastic probability transition matrices $\mathbf{P}$ that can be associated with the given graph.

We prove this equivalence by deriving the formulae for the entries of the first column of the fundamental matrix. Consider $\boldsymbol{w} = (\mathbf{I} - \mathbf{P} + \mathbf{P}^*)^{-1}\boldsymbol{r}$, the first column of $\mathbf{G}(\mathbf{P})$, where $\boldsymbol{r} = (1, 0, \ldots, 0)^{\mathrm{T}}$ is an N-dimensional column vector. Then, $(\mathbf{I} - \mathbf{P} + \mathbf{P}^*)\boldsymbol{w} = \boldsymbol{r}$. Recall that for irreducible Markov chains, every row of $\mathbf{P}^*$ is identical to $\boldsymbol{q}$, the stationary distribution vector of $\mathbf{P}$. Therefore, $\mathbf{P}^*\boldsymbol{w} = (\sum_i q_i w_i)\,\boldsymbol{e}$ and hence

$$\boldsymbol{w} - \mathbf{P}\boldsymbol{w} + (\sum_i q_i w_i)\boldsymbol{e} = \boldsymbol{r}. \tag{3.15}$$

We denote by $\mathbb{E}_i[\tau_1]$ the expectation of the first return time to vertex 1, starting from vertex i, for $i \in \mathcal{S}$. Before stating our results, we briefly introduce some concepts and recall some well-known theorems.

A *σ-field* $\mathcal{F}$ over a set Ω is a family of subsets of Ω such that it satisfies the three conditions:

(i) $\Omega \in \mathcal{F}$,
(ii) if $Y \in \mathcal{F}$ then $Y^c \in \mathcal{F}$, where Y^c is the complement of Y,
(iii) a countable union of sets in $\mathcal{F}$ is also in $\mathcal{F}$.

A real random variable X is a measurable function on $(\Omega, \mathcal{F})$, that is, for all intervals $A \subset \mathbb{R}$, the set $\{\omega \in \Omega : X(\omega) \in A\} \in \mathcal{F}$. Let $\mathcal{F}_1 \subset \mathcal{F}_2 \subset \mathcal{F}_3 \subset \cdots$ be an increasing family of σ-fields. Then, a stochastic process $\{X_n\}$ taking values in $\mathbb{R}$ is said to be a *martingale* with respect to $\{\mathcal{F}_n\}$ if

(i) X_n is $\mathcal{F}_n$-measurable for all n, and
(ii) $\mathbb{E}[|X_n|] < \infty$ and $\mathbb{E}[X_{n+1}|\mathcal{F}_n] = X_n$ for all n.

The family $\{\mathcal{F}_n\}$ is often understood from the context, $\mathcal{F}_n = \sigma(X_1, \ldots, X_n)$ being a common choice. A random variable τ taking values in $\{0, 1, 2, \ldots, \infty\}$ is a *stopping time* with respect to $\{\mathcal{F}_n\}$ if $\{\tau \leq n\} \in \mathcal{F}_n$ for all n. Equivalently, $\{\tau = n\} \in \mathcal{F}_n$ for all n. Intuitively, at each n, based on the "observed history" at each n, one knows whether τ has occurred or not.

Let $X_0 = i$ and $p(X_{m-1}, j)$ be the (X_{m-1}, j)th element of the matrix $\mathbf{P}$. We define

$$M_n = \sum_{m=1}^{n} \{w_{X_m} - \sum_j p(X_{m-1}, j) w_j\}, \text{ for } n \geq 1,$$

where w_{X_m} is the X_mth entry of the vector $\boldsymbol{w}$. Then, it is well-known (see Borkar [16, Chapter 3]) that the sequence $\{M_n\}$ is a martingale with respect to the family of σ-fields $\mathcal{F}_n = \sigma(X_i, i \leq n)$, that is, the σ-field generated by the sets of the type $\{\omega : X_i \in A\}$ for $i \leq n$ and intervals $A \subset \mathbb{R}$. The following result, by Doob, can be found in Borkar [16, Chapter 3].

Theorem 3.1. ***Optional Sampling theorem.*** *Let X_i be a martingale with respect to $\{\mathcal{F}_i\}$ and η a bounded stopping time with respect to $\{\mathcal{F}_i\}$. Then*

$$\mathbb{E}[X_\eta] = \mathbb{E}[X_1].$$

Now we are ready to state results concerning the elements of $\boldsymbol{w}$, the first column of the fundamental matrix $\mathbf{G}(\mathbf{P})$.

Theorem 3.2. *The entries w_i of the vector $\boldsymbol{w}$ are given by*

$$w_1 = \frac{1}{2} \frac{\mathbb{E}_1[\tau_1(\tau_1 + 1)]}{\mathbb{E}_1[\tau_1]^2} \tag{3.16}$$

$$w_j = \frac{1}{2} \frac{\mathbb{E}_1[\tau_1(\tau_1 + 1)]}{\mathbb{E}_1[\tau_1]^2} - \frac{\mathbb{E}_j[\tau_1]}{\mathbb{E}_1[\tau_1]}, \quad \textit{for } j \neq 1. \tag{3.17}$$

Proof. By the Optional Sampling theorem, for $i \in \mathcal{S}$ and $n \geq 1$,

$$\mathbb{E}_i[M_{\tau_1 \wedge n}] = \mathbb{E}_i[\sum_{m=1}^{\tau_1 \wedge n} (w_{X_m} - \sum_j p(X_{m-1}, j) w_j)] = 0,$$

where $\tau_1 \wedge n = \min\{\tau_1, n\}$. Since $0 < \mathbb{E}_i[\tau_1] < \infty$, we take the limit as n tends to ∞ and apply the dominated convergence theorem to obtain

$$\mathbb{E}_i[\sum_{m=1}^{\tau_1} (w_{X_m} - \sum_j p(X_{m-1}, j) w_j)] = 0. \tag{3.18}$$

Recall that $X_0 = i$, and $X_{\tau_1} = 1$. Then, for each $i = 1, 2, \ldots, n$, the left-hand side of (3.18) becomes

$$\begin{aligned}
&\mathbb{E}_i[\sum_{m=0}^{\tau_1 - 1} (w_{X_{m+1}} - \sum_j p(X_m, j) w_j)] \\
&= \mathbb{E}_i[\sum_{m=0}^{\tau_1 - 1} (w_{X_m} - \sum_j p(X_m, j) w_j) + w_{X_{\tau_1}} - w_{X_0}] \\
&= w_1 + \mathbb{E}_i[\sum_{m=0}^{\tau_1 - 1} (w_{X_m} - \sum_j p(X_m, j) w_j)] - w_i \\
&= w_1 - \mathbb{E}_i[\sum_{m=0}^{\tau_1 - 1} (\sum_j p(X_m, j) w_j - w_{X_m})] - w_i.
\end{aligned}$$

Since the right-hand side of (3.18) is 0,

$$w_1 - \mathbb{E}_i[\sum_{m=0}^{\tau_1 - 1} (\sum_j p(X_m, j) w_j - w_{X_m})] - w_i = 0. \tag{3.19}$$

Left-multiplying both sides of (3.15) by the limiting matrix $\mathbf{P}^*$ yields

$$\mathbf{P}^* \{\boldsymbol{w} - \mathbf{P}\boldsymbol{w} + (\sum_i q_i w_i)\boldsymbol{e}\} = \mathbf{P}^* \boldsymbol{r}. \tag{3.20}$$

As $\mathbf{P}^* \mathbf{P} = \mathbf{P}^*$, the left-hand side of (3.20) becomes

$$\begin{aligned}
&\mathbf{P}^* \boldsymbol{w} - \mathbf{P}^* \mathbf{P} \boldsymbol{w} + \mathbf{P}^* (\sum_i q_i w_i) \boldsymbol{e} \\
&= \mathbf{P}^* \boldsymbol{w} - \mathbf{P}^* \boldsymbol{w} + \mathbf{P}^* (\sum_i q_i w_i) \boldsymbol{e} \\
&= \mathbf{P}^* (\sum_i q_i w_i) \boldsymbol{e}.
\end{aligned}$$

Therefore, by (3.20), $\mathbf{P}^*(\sum_i q_i w_i)\boldsymbol{e} = \mathbf{P}^*\boldsymbol{r}$, and, consequently, $\sum_i q_i w_i = q_1$. Using this fact, the X_mth row of (3.15) is

$$w_{X_m} - \sum_j p(X_m, j) w_j + q_1 = I\{X_m = 1\},$$

where $I\{X_m = 1\}$ is 1 if $X_m = 1$ and 0 otherwise, as $r_{X_m} = 1$ if $X_m = 1$ and 0 otherwise. Therefore,

$$w_{X_m} - \sum_j p(X_m, j) w_j = I\{X_m = 1\} - q_1,$$

and (3.19) becomes

$$w_1 - \mathbb{E}_i[\sum_{m=0}^{\tau_1 - 1} (w_{X_m} - I\{X_m = 1\} + q_1 - w_{X_m})] - w_i = 0$$

$$w_1 + \mathbb{E}_i[\sum_{m=0}^{\tau_1 - 1} (I\{X_m = 1\} - q_1)] - w_i = 0.$$

Rearranging the last equation yields

$$w_i = w_1 + \mathbb{E}_i[\sum_{m=0}^{\tau_1 - 1} (I\{X_m = 1\} - q_1)]. \tag{3.21}$$

For all $i \neq 1, i \in \mathcal{S}$ and $m < \tau_1$, $X_m \neq 1$ as τ_1 is the first hitting time of vertex 1. Therefore, for $i \neq 1, i \in \mathcal{S}$, (3.21) is the same as

$$\begin{aligned} w_i &= w_1 + \mathbb{E}_i[\sum_{m=0}^{\tau_1 - 1} (-q_1)] \\ &= w_1 + \mathbb{E}_i[-q_1 \tau_1] \\ &= w_1 - q_1 \mathbb{E}_i[\tau_1]. \end{aligned} \tag{3.22}$$

For $i = 1$, $I\{X_0 = 1\} = 1$, and (3.21) reduces to $w_1 = w_1$ since

$$\begin{aligned} w_1 &= w_1 + \mathbb{E}_1[\sum_{m=0}^{\tau_1 - 1} (-q_1) + 1] \\ &= w_1 + \mathbb{E}_i[-q_1 \tau_1] + 1 \\ &= w_1 - q_1 \mathbb{E}_1[\tau_1] + 1 \\ &= w_1, \end{aligned} \tag{3.23}$$

the second last equality comes from the fact that $q_1 \mathbb{E}_1[\tau_1] = 1$, by Borkar [16, Theorem 5.3.2]. As $\sum_i q_i w_i = q_1$, we multiply both sides of (3.22) by q_i, and

sum over all $i \neq 1$ to obtain

$$\begin{aligned}
\sum_{i\neq 1} w_i q_i &= w_1 \sum_{i\neq 1} q_i - q_1 \sum_{i\neq 1} q_i \mathbb{E}_i[\tau_1] \\
\sum_{i} w_i q_i - w_1 q_1 &= w_1 \sum_{i\neq 1} q_i - q_1 \sum_{i\neq 1} q_i \mathbb{E}_i[\tau_1] \\
q_1 &= w_1 \sum_{i} q_i - q_1 \sum_{i\neq 1} q_i \mathbb{E}_i[\tau_1] \\
&= w_1 - q_1 \sum_{i\neq 1} q_i \mathbb{E}_i[\tau_1], \qquad (3.24)
\end{aligned}$$

with the last equality obtained by the property of the stationary distribution vector $\boldsymbol{q}$ (see (3.8)). As $q_1 \mathbb{E}_1[\tau_1] = 1$, $q_1 = \mathbb{E}_1[\tau]^{-1}$, and (3.24) becomes

$$\begin{aligned}
q_1 &= w_1 - q_1 \sum_{i} q_i \mathbb{E}_i[\tau_1] + q_1 q_1 \mathbb{E}_1[\tau_1] \\
&= w_1 - q_1 \sum_{i} q_i \mathbb{E}_i[\tau_1] + q_1.
\end{aligned}$$

Consequently,

$$w_1 = q_1 \sum_{i} q_i \mathbb{E}_i[\tau_1]. \qquad (3.25)$$

By [16, Theorem 5.3.4], equation (3.25) is equivalent to

$$\begin{aligned}
w_1 &= q_1 \frac{\mathbb{E}_1[\sum_{m=0}^{\tau_1 - 1} \mathbb{E}_{X_m}[\tau_1]]}{\mathbb{E}_1[\tau_1]} \\
&= \mathbb{E}[\tau_1]^{-1} \frac{\mathbb{E}_1[\sum_{m=0}^{\tau_1 - 1} \mathbb{E}_{X_m}[\tau_1]]}{\mathbb{E}_1[\tau_1]} \\
&= \frac{\mathbb{E}_1[\sum_{m=0}^{\tau_1 - 1} \mathbb{E}_{X_m}[\tau_1]]}{\mathbb{E}_1[\tau_1]^2} \\
&= \frac{\mathbb{E}_1[\sum_{m=0}^{\tau_1 - 1} (\tau_1 - m)]}{\mathbb{E}_1[\tau_1]^2} \\
&= \frac{1}{2} \frac{\mathbb{E}_1[\tau_1(\tau_1 + 1)]}{\mathbb{E}_1[\tau_1]^2},
\end{aligned}$$

with the last equation obtained by using the occupational measure identity in Pitman [85, p.74]. Hence, we complete the proof for (3.16).

From (3.16) and (3.22), for $i \neq 1$, we obtain

$$w_i = w_1 - q_1 \mathbb{E}_i[\tau_1]$$

$$= \frac{1}{2}\frac{\mathbb{E}_1[\tau_1(\tau_1+1)]}{\mathbb{E}_1[\tau_1]^2} - \mathbb{E}_1[\tau_1]^{-1}\mathbb{E}_i[\tau_1]$$
$$= \frac{1}{2}\frac{\mathbb{E}_1[\tau_1(\tau_1+1)]}{\mathbb{E}_1[\tau_1]^2} - \frac{\mathbb{E}_i[\tau_1]}{\mathbb{E}_1[\tau_1]},$$

which completes the proof for (3.17). □

The following theorem gives analytic expressions for w_1, the top-left element of the fundamental matrix $\mathbf{G}(\mathbf{P})$ for doubly stochastic matrices $\mathbf{P}$. In effect, it shows that this element can be regarded as an objective function

$$w_1 = w_1(\mathbf{P}) = a + bVar[\tau_1|\mathbf{P}].$$

In the next section, the associated optimisation problem will be discussed in much more detail.

Theorem 3.3. *For a given doubly stochastic* $\mathbf{P}$*, we have*

$$w_1 = \frac{1}{N^2}\sum_i \mathbb{E}_i[\tau_1] = \frac{(N+1)}{2N} + \frac{1}{2N^2}\mathbb{E}_1[(\tau_1 - N)^2]. \tag{3.26}$$

Moreover, if $\mathbf{P}$ *is associated with a Hamiltonian cycle and we apply the symmetric linear perturbation defined in (3.13), then* w_1 *simplifies to*

$$w_1^\varepsilon = \frac{1}{2}\frac{(N+1)}{N} + O(\varepsilon). \tag{3.27}$$

Proof. Recall that for a doubly stochastic probability transition matrix $\mathbf{P}$, the stationary distribution vector $\boldsymbol{q}$ of $\mathbf{P}$ is $1/N\boldsymbol{e}^{\mathrm{T}}$, so $q_i = 1/N$ for all i. Since $q_1 = \mathbb{E}_1[\tau_1]^{-1}$, we have $\mathbb{E}_1[\tau_1] = N$. By (3.25),

$$w_1 = q_1\sum_i q_i\mathbb{E}_i[\tau_1] = 1/N^2\sum_i \mathbb{E}_i[\tau_1].$$

In addition, by (3.16),

$$\begin{aligned} w_1 &= \frac{\mathbb{E}_1[\tau_1(\tau_1+1)]}{2(\mathbb{E}_1[\tau_1])^2} \\ &= \frac{\mathbb{E}_1[\tau_1^2] + \mathbb{E}_1[\tau_1]}{2(\mathbb{E}_1[\tau_1])^2} \\ &= \frac{\mathbb{E}_1[(\tau_1 - \mathbb{E}_1[\tau_1])^2] + \mathbb{E}_1[\tau_1]^2 + \mathbb{E}_1[\tau_1]}{2(\mathbb{E}_1[\tau_1])^2} \\ &= \frac{\mathbb{E}_1[\tau_1]^2 + \mathbb{E}_1[\tau_1]}{2(\mathbb{E}_1[\tau_1])^2} + \frac{\mathbb{E}_1[(\tau_1 - \mathbb{E}_1[\tau_1])^2]}{2(\mathbb{E}_1[\tau_1])^2} \\ &= \frac{N+1}{2N} + \frac{\mathbb{E}_1[(\tau_1 - N)^2]}{2N^2}. \end{aligned}$$

For the second part of Theorem 3.3, we need to show that $\mathbb{E}[(\tau_1 - N)^2] = O(\varepsilon)$. The probability that the system travels using at least one weak transition, that is, an edge with probability ε, in N steps is at most $N(N-1)\varepsilon$. Therefore, $P\{\tau_1 \neq N\} \leq N(N-1)\varepsilon$. Also, for the inequality $\tau_1 > kN$ to hold for $k \geq 1$, the system must travel using at least one weak transition in each block of N consecutive steps. Consequently, $P\{\tau_1 > kN\} \leq [N(N-1)\varepsilon]^k$. This leads to

$$\begin{aligned}
\mathbb{E}[(\tau_1 - N)^2] &= \sum_{j\geq 1} j^2 P\{|\tau_1 - N| = j\} \\
&= \sum_{j\geq 1} j^2 P\{\tau_1 = N - j\} + \sum_{j\geq 1} j^2 P\{\tau_1 = N + j\} \\
&\leq N^2 \sum_{j\geq 1} P\{\tau_1 = N - j\} + \sum_{j\geq 1} j^2 P\{\tau_1 = N + j\} \\
&\leq N^2 P\{\tau_1 \neq N\} + \sum_{j\geq 1} j^2 P\{\tau_1 = N + j\} \\
&\leq N^3(N-1)\varepsilon + \sum_{j\geq 1} j^2 P\{\tau_1 = j + N\} \\
&\leq N^3(N-1)\varepsilon + \sum_{k\geq 1} [(k+1)N]^2 P\{\tau_1 > kN\} \\
&\leq N^3(N-1)\varepsilon + \sum_{k\geq 1} [(k+1)N]^2 [N(N-1)\varepsilon]^k \\
&\leq O(\varepsilon).
\end{aligned}$$

This completes the proof. □

3.4 Hamiltonian Cycles as Hitting Time Variance Minimisers

Consider a given graph G with N vertices and recall that every unperturbed probability transition matrix $\mathbf{P}$ induced by G has the property that $p_{ij} = 0$ whenever edge (i, j) is not present in the graph. All unperturbed probability transition matrices considered in this section are assumed to be induced by G. It is clear that if $\mathbf{P}$ is the probability transition matrix induced by a Hamiltonian cycle, then starting from the home vertex 1, the corresponding Markov chain will return to it after exactly N transitions, implying that $Var[\tau_1|\mathbf{P}] = 0$. Furthermore, since in such a case the symmetric linear perturbation defined in (3.13) does not alter the ergodic structure of this chain, it is reasonable to expect that $Var[\tau_1|\mathbf{P}^\varepsilon]$ tends to 0 as ε approaches 0. Hence, in view of Theorem 3.3, it is also reasonable to conjecture that probability transition matrices $\mathbf{P}^\varepsilon_{\text{H}}$ induced by Hamiltonian cycles achieve the minimum

in the optimisation problem

$$\min Var[\tau_1|\mathbf{P}^\varepsilon],$$

over the space of doubly stochastic probability transition matrices $\mathbf{P}^\varepsilon$ induced by the given graph G, provided that $\varepsilon > 0$ is sufficiently small. Equivalently, we conjecture that probability transition matrices $\mathbf{P}^\varepsilon_{\text{H}}$ induced by Hamiltonian cycles achieve the minimum in the optimisation problem

$$\min w_1^\varepsilon(\mathbf{P}^\varepsilon),$$

over the same space of doubly stochastic probability transition matrices, provided that $\varepsilon > 0$ is sufficiently small. This section is devoted to proving these conjectures and to establishing the existence of the *Hamiltonicity gap* property, which demonstrates that the optimal values of these two objective functions can be used to distinguish between Hamiltonian and non-Hamiltonian graphs, without actually requiring that a Hamiltonian cycle be found.

We denote by $\mathbf{P}^\varepsilon$ the probability transition matrix obtained from $\mathbf{P}$ after we apply the symmetric linear perturbation defined in (3.13), by $\mathcal{D}_d$ the finite set of perturbed doubly stochastic deterministic matrices, and by $\mathcal{D}$ the convex set of perturbed doubly stochastic matrices obtained by taking the closed convex hull of $\mathcal{D}_d$. We also write $\mathcal{D}_d$ as the disjoint union $\mathcal{D}_{\text{H}} \cup \mathcal{D}_s$, where $\mathcal{D}_{\text{H}}$ represents the set of Hamiltonian cycles and $\mathcal{D}_s$ the set of disjoint unions of short cycles that cover the graph. All sets $\mathcal{D}_d, \mathcal{D}, \mathcal{D}_{\text{H}}$ and $\mathcal{D}_s$ depend on the perturbation parameter ε, which we suppress. Likewise, we suppress the dependence of aforementioned objective functions on $\mathbf{P}^\varepsilon$, except where ambiguity might arise.

Lemma 3.1. *For $\mathbf{P}^\varepsilon \in \mathcal{D}_s$, the top-left entry w_1^ε of the fundamental matrix $\mathbf{G}(\mathbf{P}^\varepsilon)$ tends to infinity as ε approaches 0.*

Proof. Let $i \in \mathcal{S}$ lie in a short cycle of $\mathbf{P}$ not containing vertex 1. Then, a chain starting at i must make an $\varepsilon-$transition before ever hitting 1. Thus, if ζ denotes the first time it makes an $\varepsilon-$transition, then

$$\begin{aligned}\mathbb{E}_i[\tau_1] &\geq \mathbb{E}_i[\zeta] \\ &= \sum_{m\geq 1} m(N-1)\varepsilon\{1-(N-1)\varepsilon\}^{m-1} \\ &= 1/\{(N-1)\varepsilon\}.\end{aligned}$$

The claim follows from (3.26). □

We say that $\mathbf{P} \in \mathcal{D}$ is a perturbation of Hamiltonian cycle if there exists a $\hat{\mathbf{P}} \in \mathcal{D}_{\text{H}}$ such that $||\mathbf{P} - \hat{\mathbf{P}}|| = C\varepsilon_0$ for prescribed $C, \varepsilon_0 > 0$. Let $\mathcal{D}_p$ denote the set of such matrices $\mathbf{P}$.

Theorem 3.4. *For sufficiently small $\varepsilon > 0$, all minima of w_1^ε on $\mathcal{D}$ are attained on $\mathcal{D}_p$.*

The proof uses the following lemma. We use the notation $\theta(1)$ to denote any function $g(\varepsilon)$ satisfying $\liminf_{\varepsilon \downarrow 0} g(\varepsilon) > 0$ and $\limsup_{\varepsilon \downarrow 0} g(\varepsilon) < \infty$.

Lemma 3.2. *For $\mathbf{P} \in \mathcal{D} \setminus \mathcal{D}_p$,*

$$w_1^\varepsilon \geq \frac{1}{2}\frac{(N+1)}{N} + \theta(1).$$

Proof. By (3.26)

$$w_1^\varepsilon = \frac{1}{2}\frac{N+1}{N} + \frac{1}{2N^2}\mathbb{E}_1[(\tau_1 - N)^2].$$

Thus, it suffices to prove that for $\mathbf{P} \in \mathcal{D} \setminus \mathcal{D}_p$,

$$\mathbb{E}_1[(\tau_1 - N)^2] = \theta(1),$$

which it indeed will be if the weighted digraph of $\mathbf{P}$ contains a short (that is, non-Hamiltonian) cycle containing 1 which has a probability $\theta(1)$. Now, $\mathbf{P}$ is a finite convex combination of elements of $\mathcal{D}_e$, with either

(i) weight $\theta(1)$ for at least one $\hat{\mathbf{P}} \in \mathcal{D}_s$, or
(ii) weights $\theta(1)$ for at least two distinct $\mathbf{P}', \mathbf{P}'' \in \mathcal{D}_{\text{H}}$.

It is easy to see that if (i) and (ii) were false, $\mathbf{P}$ would be in $\mathcal{D}_p$. In case of (i), it is clear that $\mathbf{P}$ has a short cycle containing 1 which has a probability $\theta(1)$. In case of (ii), let $\mathbf{P}$ put the $\theta(1)$ weights on $\mathbf{P}_1, \mathbf{P}_2 \in \mathcal{D}_{\text{H}}$, corresponding to two Hamiltonian cycles $1 \to x_1 \to x_2 \to \cdots \to x_{N-1} \to 1$ and $1 \to y_1 \to y_2 \to \cdots \to y_{N-1} \to 1$ respectively. Let $i = \min\{j > 1 : x_j \neq y_j\}$. Then, $x_i = y_k$ for some $k > i$. Hence, $1 \to \cdots \to x_i \to y_{k+1} \to \cdots \to y_{s-1} \to 1$ defines a short cycle that contains 1 and has a probability $\theta(1)$. This completes the proof. □

Proof of Theorem 3.4. Since w^ε is the unique solution to the well-posed linear system given by (3.15), it depends smoothly on the coefficients thereof by Cramer's rule, hence on ε. From Theorem 3.3, it then follows that

$$w_1^\varepsilon = (N+1)/(2N) + \mathcal{O}(\varepsilon)$$

for $\mathbf{P} \in \mathcal{D}_p$. The claim now follows from Lemma 3.2. □

Directional derivatives Before showing that the minima of w_1^ε on $\mathcal{D}$ are attained at doubly stochastic matrices induced by Hamiltonian cycles, we need to present some results on directional derivatives. We now derive an expression for the directional derivative of our objective function w_1^ε.

Let $\mathbf{P}_0$ and $\mathbf{P}_1$ denote two doubly stochastic matrices in $\mathcal{D}$ and for $0 \leq \lambda \leq 1$, define $\mathbf{P}_\lambda = \lambda\mathbf{P}_1 + (1-\lambda)\mathbf{P}_0$. Correspondingly, define $\kappa_\lambda(i) = \mathbb{E}_i[\tau_1]$ for $i \in \mathcal{S}$, where the dependence of the distribution of τ_1 on the parameter λ is implicit. Also, let $\tilde{\mathbf{P}}_0, \tilde{\mathbf{P}}_1$ and $\tilde{\mathbf{P}}_\lambda$ denote sub-matrices derived, respectively, from $\mathbf{P}_0, \mathbf{P}_1$, and $\mathbf{P}_\lambda$ by deletion of their first row and column. Similarly, for vectors, we use tilde to denote truncations resulting from the omission of the first entry, for instance, $\tilde{\boldsymbol{\kappa}}_\lambda = (\kappa_\lambda(2), \ldots, \kappa_\lambda(N))^{\mathrm{T}}$, $\tilde{\mathbf{e}} = (1, \ldots, 1)^{\mathrm{T}} \in \mathbb{R}^{N-1}$.

By Borkar [15, Lemma 1.3], we know that $\tilde{\boldsymbol{\kappa}}_\lambda$ is the unique solution to

$$\tilde{\boldsymbol{\kappa}}_\lambda = \tilde{\mathbf{e}} + \tilde{\mathbf{P}}_\lambda \tilde{\boldsymbol{\kappa}}_\lambda, \tag{3.28}$$

that is,

$$\tilde{\boldsymbol{\kappa}}_\lambda = (\mathbf{I} - \tilde{\mathbf{P}}_\lambda)^{-1}\tilde{\mathbf{e}}. \tag{3.29}$$

We denote by $\{X_n\}$ the Markov chain governed by $\mathbf{P}_\lambda$, and by p_{ij} the elements of $\mathbf{P}_\lambda$.

Lemma 3.3. *Let $\nu_\lambda(i) = \mathbb{E}[\sum_{m=1}^{\tau_1} I\{X_m = i\}]$ when the initial distribution is the uniform distribution. Then, $\tilde{\boldsymbol{\nu}}_\lambda^{\mathrm{T}} = (\nu_\lambda(2), \ldots, \nu_\lambda(N))$ is the unique solution to*

$$\tilde{\boldsymbol{\nu}}_\lambda^{\mathrm{T}} = 1/N\tilde{\mathbf{e}}^{\mathrm{T}} + \tilde{\boldsymbol{\nu}}_\lambda^{\mathrm{T}}\tilde{\mathbf{P}}_\lambda = 1/N\tilde{\mathbf{e}}^{\mathrm{T}}(\mathbf{I} - \tilde{\mathbf{P}}_\lambda)^{-1}. \tag{3.30}$$

Proof. We define $\zeta_1 = \min\{n \geq 0 : X_n = 1\}$ ($= \infty$ if this set is empty), and $\nu_\lambda^j(i) = \mathbb{E}_j[\sum_{m=0}^{\zeta_1} I\{X_m = i\}]$. Then, for $j \neq 1$, $\zeta_1 = \tau_1$, and

$$\nu_\lambda^j(i) = \mathbb{E}_j[\sum_{m=0}^{\zeta_1} I\{X_m = i\}] = \mathbb{E}_j[\sum_{m=0}^{\tau_1} I\{X_m = i\}].$$

Consider $i, j \neq 1$. We note that for $n \geq 1$, $M_n = \sum_{m=1}^{n}(I\{X_m = i\} - p_{X_{m-1},i})$ is a martingale. By the Optional Sampling theorem (stated in Theorem 3.1) and for $T \geq 1$,

$$\mathbb{E}_j[\sum_{m=1}^{\tau_1 \wedge T}(I\{X_m = i\} - p_{X_{m-1},i})] = 0 \tag{3.31}$$

Letting $T \uparrow \infty$ and using the easily established fact that $\mathbb{E}[\tau_1^2] < \infty$, which implies uniform integrability of the sum above as T varies, we have

$$\mathbb{E}_j[\sum_{m=1}^{\tau_1}(I\{X_m = i\} - p_{X_{m-1},i})] = 0.$$

Thus,

$$\begin{aligned}
\nu_\lambda^j(i) &= \mathbb{E}_j[\sum_{m=0}^{\tau_1} I\{X_m = i\}] \\
&= \delta_{ij} + \mathbb{E}_j[\sum_{m=1}^{\tau_1} I\{X_m = i\}] \\
&= \delta_{ij} + \mathbb{E}_j[\sum_{m=1}^{\tau_1} p_{X_{m-1},i}] \\
&= \delta_{ij} + \mathbb{E}_j[\sum_{m=0}^{\tau_1-1} p_{X_m,i}] \\
&= \delta_{ij} + \mathbb{E}_j[\sum_{m=0}^{\tau_1-1} \sum_{k\neq 1} p_{ki} I\{X_m = k\}]
\end{aligned}$$

since $X_m \neq 1$ for $m < \tau_1$,

$$\begin{aligned}
&= \delta_{ij} + \mathbb{E}_j[\sum_{k\neq 1} \sum_{m=0}^{\tau_1-1} p_{ki} I\{X_m = k\}] \\
&= \delta_{ij} + \sum_{k\neq 1} \mathbb{E}_j[\sum_{m=0}^{\tau_1} I\{X_m = k\}] p_{ki},
\end{aligned}$$

since $X_{\tau_1} = 1 \neq k$. Thus,

$$\nu_\lambda^j(i) = \delta_{ij} + \sum_{k\neq 1} \nu_\lambda^j(k) p_{ki}. \tag{3.32}$$

Since $\nu_\lambda^1(i) = 0$ for $i \neq 1$, we also have

$$\nu_\lambda^1(i) = \delta_{i1} + \sum_{k\neq 1} \nu_\lambda^1(k) p_{ki} = 0.$$

Multiplying both sides of (3.32) by $1/N$ and summing over j, we obtain

$$\sum_{j=1}^{N} 1/N \nu_\lambda^j(i) = 1/N + 1/N \sum_j \sum_{k\neq 1} \nu_\lambda^j(k) p_{ki}.$$

This proves the claim. □

Let $J(\lambda)$ denote our objective as a function of λ, that is, w_1^ε, evaluated along the line segment $\{\mathbf{P}_\lambda : 0 \leq \lambda \leq 1\}$. From (3.26), we have

$$J(\lambda) = 1/N^2 \sum_{i=1}^{N} \kappa_\lambda(i). \tag{3.33}$$

Differentiating with respect to λ on both sides yields

$$J'(\lambda) = 1/N^2 \sum_{i=1}^{N} \kappa'_\lambda(i) = 1/N^2 \sum_{i=2}^{N} \kappa'_\lambda(i), \tag{3.34}$$

because $\kappa_\lambda(1) = \mathbb{E}_1[\tau_1] = N$ for all $\lambda \in [0,1]$ and, consequently, $\kappa'_\lambda(1) = 0$ for all $\lambda \in [0,1]$. From (3.28) and the definition of $\tilde{\mathbf{P}}_\lambda$, we have

$$\tilde{\boldsymbol{\kappa}}'_\lambda = (\tilde{\mathbf{P}}_1 - \tilde{\mathbf{P}}_0)\tilde{\boldsymbol{\kappa}}_\lambda + \tilde{\mathbf{P}}_\lambda \tilde{\boldsymbol{\kappa}}'_\lambda.$$

Therefore,

$$\tilde{\boldsymbol{\kappa}}'_\lambda = (\mathbf{I} - \tilde{\mathbf{P}}_\lambda)^{-1}(\tilde{\mathbf{P}}_1 - \tilde{\mathbf{P}}_0)\tilde{\boldsymbol{\kappa}}_\lambda,$$

and, together with (3.29) and (3.30), this leads to

$$\begin{aligned} J'(\lambda) &= \frac{1}{N^2}\tilde{\mathbf{e}}^{\mathrm{T}}(\mathbf{I} - \tilde{\mathbf{P}}_\lambda)^{-1}(\tilde{\mathbf{P}}_1 - \tilde{\mathbf{P}}_0)\tilde{\boldsymbol{\kappa}}_\lambda \\ &= \frac{1}{N}\tilde{\boldsymbol{\nu}}_\lambda^{\mathrm{T}}(\tilde{\mathbf{P}}_1 - \tilde{\mathbf{P}}_0)\tilde{\boldsymbol{\kappa}}_\lambda \\ &= \frac{1}{N^2}\tilde{\mathbf{e}}^{\mathrm{T}}(\mathbf{I} - \tilde{\mathbf{P}}_\lambda)^{-1}(\tilde{\mathbf{P}}_1 - \tilde{\mathbf{P}}_0)(\mathbf{I} - \tilde{\mathbf{P}}_\lambda)^{-1}\tilde{\mathbf{e}}. \end{aligned} \tag{3.35}$$

Though the following, purely technical, lemma is a straightforward application of the Cauchy-Schwartz inequality, we include its proof for the sake of completeness.

Lemma 3.4. *If $x_m = m$ for $1 \le m \le N$, and $\{y_k\}$ is a permutation of $\{x_j\}$, then $\sum_i x_i y_i$ is maximised when $y_i = x_i$ for all i, and minimised when $y_i = N + 1 - x_i$ for all i.*

Proof. The maximisation claim is immediate from the Cauchy-Schwartz inequality. For the permutation $\{z_i = N + 1 - y_i\}$ of $\{x_j\}$, we have

$$\sum_i x_i z_i \le \sum_i x_i^2,$$

with equality if and only if $x_i = z_i$ for all i. Hence,

$$\sum_i x_i y_i = \sum_i x_i(N + 1 - z_i) \ge \sum_i x_i(N + 1 - x_i),$$

with equality if and only if $y_i = N + 1 - x_i$ for all i. Thus, the minimisation claim also follows. □

We now consider $J'(0)$ in the situations where the doubly stochastic matrix $\mathbf{P}_0 \in \mathcal{D}_{\text{H}}$ is induced by a deterministic transition matrix tracing out a Hamiltonian cycle. We first show that $J'(0) > 0$ on a straight line path from $\mathbf{P}_0$ towards any doubly stochastic $\mathbf{P}_1$ induced by the graph. This shows that deterministic transition matrices inducing Hamiltonian cycles correspond to local minima. Suppose then that $\mathbf{P}_0$ corresponds to a Hamiltonian cycle H_0. Without loss of generality, we assume H_0 is the Hamiltonian cycle $1 \to 2 \to \cdots \to N \to 1$. To start with, we consider $\mathbf{P}_1 \in \mathcal{D}_{\text{H}} \cup \mathcal{D}_s$, other than $\mathbf{P}_0$. That is, $\mathbf{P}_1$ is induced by any deterministic matrix that traces out in the graph either a union of disjoint cycles or a Hamiltonian cycle other than H_0. For each $i \in \mathcal{S}$ and $i \neq 1$, we denote by $m(i)$ the number of steps required to reach vertex 1 from i on H_0 if $\varepsilon = 0$. Then,

$$\begin{aligned} \kappa_0(i) &= \mathbb{E}_i[\tau_1] = m(i) + O(\varepsilon) \\ &= (N - i + 1) + O(\varepsilon). \end{aligned} \tag{3.36}$$

To verify that the suppressed $\varepsilon-$dependence above is, indeed, only of order $O(\varepsilon)$, note that by (3.13) we can write $\tilde{\mathbf{P}}_0 = \tilde{\mathbf{P}}_0(\varepsilon) = \tilde{\mathbf{P}}_0(0) - \varepsilon\tilde{\mathbf{K}}_0$, for some fixed matrix $\tilde{\mathbf{K}}_0$. Now, it follows from (3.29) that (3.36) is simply the ith equation in the system

$$\tilde{\boldsymbol{\kappa}}_0 = (\mathbf{I} - \tilde{\mathbf{P}}_0(\varepsilon))^{-1}\tilde{\mathbf{1}} = (\mathbf{I} - \tilde{\mathbf{P}}_0(0) + \varepsilon\tilde{\mathbf{K}}_0)^{-1}\tilde{\mathbf{1}}. \tag{3.37}$$

It is easy to check that $\mathbf{I} - \tilde{\mathbf{P}}_0(0)$ is invertible and hence that the above equation corresponds to a regularly (rather than a singularly) perturbed system. By well-known results from perturbation theory of linear operators (see Langenhop [73] and Avrachenkov *et al.* [6]), it follows that

$$\tilde{\boldsymbol{\kappa}}_0 = (\mathbf{I} - \tilde{\mathbf{P}}_0(0))^{-1}\tilde{\mathbf{1}} + O(\varepsilon).$$

Also,

$$\mathbb{E}_j[\sum_{\ell=0}^{\tau_1} I\{X_\ell = i\}] = \begin{cases} 1 + O(\varepsilon) & \text{for } j = 1, 2, \ldots, i-1, \\ O(\varepsilon) & \text{for } j = i, \ldots, N. \end{cases} \tag{3.38}$$

Equation (3.38) can be proved by arguments analogous to those used for proving (3.36). Hence, for $i = 2, \ldots, N$,

$$\begin{aligned} \nu_0(i) &= 1/N \sum_{1 \leq j < i} \mathbb{E}_j[\sum_{\ell=1}^{\tau_1} I\{X_\ell = i\}] + O(\varepsilon) \\ &= (i-1)/N + O(\varepsilon). \end{aligned}$$

Thus, by (3.28),

$$\begin{aligned}\tilde{\boldsymbol{\nu}}_0^{\mathrm{T}}\tilde{\mathbf{P}}_0\tilde{\boldsymbol{\kappa}}_0 &= \sum_{i=2}^{N-1}\nu_0(i)\kappa_0(i+1)\\ &= 1/N\{1(N-2)+2(N-3)+\cdots(N-2)1\}+O(\varepsilon)\\ &= 1/N\sum_{r=1}^{N-1} r\{(N-1)-r\}+O(\varepsilon)\\ &= (N-1)^2/2-1/N\sum_{r=1}^{N-1} r^2+O(\varepsilon).\end{aligned}$$

Now, suppose that $\mathbf{P}_1$ is induced by either a Hamiltonian cycle distinct from H_0 or a deterministic matrix that traces out a union of disjoint cycles in the graph. Hence, for every ith row there is a unique j_ith column such that $[\mathbf{P}_1]_{i,j_i}=1$ and $j_i\neq j_k$, for $i\neq k$. Thus,

$$\begin{aligned}\tilde{\boldsymbol{\nu}}_0^{\mathrm{T}}\tilde{\mathbf{P}}_1\tilde{\boldsymbol{\kappa}}_0 &= \sum_{i=2}^{N}\nu_0(i)\kappa_0(j_i)\\ &= 1/N\sum_{i=2}^{N}(i-1)(N-j_i+1)+O(\varepsilon)\\ &= 1/N\sum_{r=1}^{N-1} r\{(N-1)-(j_{r+1}-2)\}+O(\varepsilon)\\ &= (N-1)/N\sum_{r=1}^{N-1} r-1/N\sum_{r=1}^{N-1} ry_r+O(\varepsilon)\\ &= (N-1)^2/2-1/N\sum_{r=1}^{N-1} ry_r+O(\varepsilon),\end{aligned}$$

where $r=i-1$, $y_r=(j_{r+1}-2)$ and $y_r\in\{0,1,2,\ldots,(N-2)\}$ with $y_r\neq y_k$ whenever $r\neq k$. If y_r were allowed to take values only in the set $\{1,2,\ldots,N-1\}$, then by Lemma 3.4 we would have that

$$\sum_{r=1}^{N-1} r^2>\sum_{r=1}^{N-1} ry_r, \tag{3.39}$$

whenever $(y_1,\ldots,y_{N-1})\neq(1,\ldots,N-1)$. However, the inclusion of zero as one of the possible values for y_r can only lower the right-hand side of (3.39). Hence, we have proved that, whenever $\tilde{\mathbf{P}}_1\in\mathcal{D}_{\mathrm{H}}\cup\mathcal{D}_s$ and $\tilde{\mathbf{P}}_1\neq\tilde{\mathbf{P}}_0$,

$$\tilde{\boldsymbol{\nu}}_0^{\mathrm{T}}\tilde{\mathbf{P}}_1\tilde{\boldsymbol{\kappa}}_0-\tilde{\boldsymbol{\nu}}_0^{\mathrm{T}}\tilde{\mathbf{P}}_0\tilde{\boldsymbol{\kappa}}_0>0.$$

Now consider an arbitrary doubly stochastic $\mathbf{P}_1$ other than $\mathbf{P}_0$. By Birkhoff-von Neumann theorem (Bapat and Raghavan [9]),

$$\mathbf{P}_1 = \sum_{i=1}^{M} \gamma_i \mathbf{P}_i^\dagger, \tag{3.40}$$

where $\gamma_i \geq 0$ for all $i, \sum_i \gamma_i = 1$, $\mathbf{P}_i^\dagger \in \mathcal{D}_{\text{H}} \cup \mathcal{D}_s$ correspond to permutation matrices and $M \geq 1$ is the number of permutation matrices induced by the graph. For at least one value of i in the summation (3.40), $\mathbf{P}_i^\dagger \neq \mathbf{P}_0$ and $\gamma_i > 0$. Then, by the preceding strict inequalities and the second equality of (3.35) we have that

$$\begin{aligned} J'(0) &= 1/N(\tilde{\boldsymbol{\nu}}_0^{\text{T}} \tilde{\mathbf{P}}_1 \tilde{\boldsymbol{\kappa}}_0 - \tilde{\boldsymbol{\nu}}_0^{\text{T}} \tilde{\mathbf{P}}_0 \tilde{\boldsymbol{\kappa}}_0) \\ &= 1/N \sum_i \gamma_i (\tilde{\boldsymbol{\nu}}_0^{\text{T}} \tilde{\mathbf{P}}_i^\dagger \tilde{\boldsymbol{\kappa}}_0 - \tilde{\boldsymbol{\nu}}_0^{\text{T}} \tilde{\mathbf{P}}_0 \tilde{\boldsymbol{\kappa}}_0) > 0. \end{aligned}$$

The following main result now follows rather easily.

Theorem 3.5. *If $\mathbf{P}_0$ is induced by a Hamiltonian cycle, then,*

(i) $\mathbf{P}_0$ is a strict local minimum for the cost functional w_1^ε, and

(ii) $\mathbf{P}_0$ is also a global minimum for the cost functional w_1^ε.

Proof. Part (i) was proved above for $\mathbf{P}_0$ corresponding to the Hamiltonian cycle H_0: it is sufficient to observe that for a strict *local* minimum, the quantity

$$\boldsymbol{\nu}_0 \tilde{\mathbf{P}}_1 \boldsymbol{\kappa}_0 - \boldsymbol{\nu}_0 \tilde{\mathbf{P}}_0 \boldsymbol{\kappa}_0$$

remains strictly bounded away from zero as ε approaches 0 for all extremal $\mathbf{P}_1 \neq \mathbf{P}_0$. The effect of considering another Hamiltonian cycle would be only to permute the order of the terms in various summations, without changing the conclusions.

To obtain Part (ii), first note that the above allows us to choose an $\eta_0 > 0$ such that $\mathbf{P}_0$ is the strict local minimum of w_1^ε in the η_0-neighborhood of $\mathbf{P}_0$. As in the proof of Theorem 3.4, choose $\varepsilon > 0$ small enough so that the global minimum of w_1^ε is attained on the η_0-neighborhood of $\mathbf{P}_0$. *Small enough* here is quantified by an upper bound that depends only on N and η_0 (see Borkar *et al.* [17]). The claim follows. □

Recall from Theorem 3.3 that for a doubly stochastic matrix $\mathbf{P} \in \mathcal{D}$, the functional consisting of the top-left element of the fundamental matrix induced by $\mathbf{P}$ is given by

$$w_1^\varepsilon = \frac{N+1}{2N} + \frac{1}{2N^2} \mathbb{E}_1[(\tau_1 - N)^2]. \tag{3.41}$$

We suppressed the dependence on $\mathbf{P}$ on the right-hand side of (3.41), but the expectation term is a function of $\mathbf{P}$, since $\mathbf{P}$ determines the distribution of τ_1. It should now be clear that a consequence of Theorems 3.3 and 3.5 is that whenever the underlying graph G is Hamiltonian, the minimum of the above functional over $\mathbf{P} \in \mathcal{D}$ is given by

$$w_1^{\varepsilon}(\mathbf{P}_{\text{H}}) = \min_{\mathbf{P}\in\mathcal{D}} w_1^{\varepsilon}(\mathbf{P}) = (N+1)/(2N) + O(\varepsilon), \tag{3.42}$$

where $\mathbf{P}_{\text{H}} \in \mathcal{D}_{\text{H}}$ is a probability transition matrix defining any Hamiltonian cycle in the graph.

Hamiltonicity Gap: A lower bound for the non-Hamiltonian case

In this section, we prove that for positive and sufficiently small ε, there exists $\Delta(N) > 0$ such that whenever the graph G is non-Hamiltonian

$$\{\min_{\mathbf{P}\in\mathcal{D}^{\varepsilon}} w_1^{\varepsilon}(\mathbf{P})\} - w_1^{\varepsilon}(\mathbf{P}_{\text{H}}) \geq \Delta(N) - O(\varepsilon).$$

We name the quantity $\Delta(N)$ the *Hamiltonicity gap of order N* because it distinguishes all non-Hamiltonian graphs with N vertices from all Hamiltonian graphs with the same number of vertices.

Before presenting the proof, we note that such a result is reasonable when we consider the possible variability of τ_1—as captured by its variance $\mathbb{E}_1[(\tau_1 - N)^2]$—for both Hamiltonian and non-Hamiltonian graphs. In the former case, it is clear that this variance can be made nearly zero by following a Hamiltonian cycle because the latter would yield a variance actually equal to zero were it not for the (small) perturbation ε. However, if the graph is non-Hamiltonian, perhaps we cannot avail ourselves of such a variance annihilating transition matrix. This intuitive reasoning is made rigorous in the remainder of this section.

The key step in what follows is the derivation of an upper bound on $P\{\tau_1 = N|\mathbf{P}\}$, the probability that the system returns to vertex 1 in N steps, under an arbitrary doubly stochastic matrix $\mathbf{P}$ in a non-Hamiltonian graph.

Lemma 3.5. *Suppose that G is a non-Hamiltonian graph, and let $\mathbf{P}$ be an arbitrary doubly stochastic transition matrix feasible on G.*

(i) If $\varepsilon = 0$, then $P\{\tau_1 = N|\mathbf{P}\} \leq 1/4$.

(ii) If $\varepsilon > 0$ and small, then $P\{\tau_1 = N|\mathbf{P}\} \leq 1/4 + O(\varepsilon)$.

Proof. First, consider the case $\varepsilon = 0$. Let $\mathbf{P}$ be an arbitrary doubly stochastic matrix and let $\{X_t\}_0^{\infty}$ be the Markov chain induced by $\mathbf{P}$ and the starting

state 1. Let $\gamma_1 = (X_0, X_1, \ldots, X_N)$ be a path of N steps through the graph and let $\chi_1 = \{\gamma_1 | X_0 = X_N = 1, X_k \neq 1, k = 1, \ldots, N-1\}$. That is, the event that the first return to 1 occurs after N steps is $\{\tau_1 = N\}$, which is simply the event that γ_1 traces a path within χ_1 and hence

$$P\{\tau_1 = N|\mathbf{P}\} = \sum_{\gamma_1 \in \chi_1} p_{\gamma_1},$$

where p_{γ_1} denotes the probability (under $\mathbf{P}$) of observing the path γ_1. However, because the graph is assumed to be non-Hamiltonian, all the paths in χ_1 that receive a positive probability have the structure

$$\gamma_1 = \gamma_1' \cup \bar{\gamma}_1,$$

where γ_1' consists of a non-self-intersecting "reduced path" from 1 to itself of length $m \leq N-2$ adjoined at some vertex (or vertices) other than 1 by one or more loops of total length $N-m$, that together constitute $\bar{\gamma}_1$. One can think of $\gamma_1' \cup \bar{\gamma}_1$ as the first and second parts of a figure comprising of a basic loop with one or more side-lobes attached to it, each of which is either a loop or a connected union of loops. The simplest instance of this is a figure of eight, with two loops of length m and $N-m$ respectively, attached at a vertex other than 1.

Let p_{γ_1} denote the probability of the original path and p'_{γ_1} that of the reduced path. Let $q = p_{\gamma_1}/p'_{\gamma_1} \leq 1$, which is the contribution to p coming from the loops comprising $\bar{\gamma}_1$. More generally, define $\gamma_0 = \gamma_1', \gamma_1 = \gamma_1' \cup \bar{\gamma}_1, \gamma_2 = \gamma_1' \cup \bar{\gamma}_1 \cup \bar{\gamma}_1, \gamma_3 = \gamma_1' \cup \bar{\gamma}_1 \cup \bar{\gamma}_1 \cup \bar{\gamma}_1, \ldots$. For $n \geq 2$, the paths γ_n from 1 to itself that begin with the same reduced path γ_1' but may repeat exactly the path $\bar{\gamma}_1$ for $n \geq 2$ times, all contribute to the event $\{\tau_1 \neq N\}$, as does $\gamma_0 = \gamma_1'$.

The paths γ_n, for $n \geq 2$, have probabilities $p_{\gamma_1} q^{n-1}$. The total probability that these paths and $\gamma_0 = \gamma_1'$ (but excluding the original γ_1) contribute to $\{\tau_1 \neq N\}$ is

$$\begin{aligned} p_{\gamma_1}/q + \sum_{n \geq 2} p_{\gamma_1} q^{n-1} &= p_{\gamma_1}\{1/q + q/(1-q)\} \\ &= p_{\gamma_1}\{-1 + 1/(q(1-q))\} \\ &\geq 3p_{\gamma_1}. \end{aligned}$$

It follows that

$$P\{\tau_1 \neq N|\mathbf{P}\} \geq \sum_{\gamma_1 \in \chi_1} 3p_{\gamma_1} = 3P\{\tau_1 = N|\mathbf{P}\}.$$

Hence,

$$\begin{aligned} 1 &= P\{\tau_1 < \infty | \mathbf{P}\} \\ &= P\{\tau_1 = N | \mathbf{P}\} + P\{\tau_1 \neq N | \mathbf{P}\} \\ &\geq 4P\{\tau_1 = N | \mathbf{P}\}, \end{aligned}$$

implying $P\{\tau_1 = N|\mathbf{P}\} \leq 1/4$, or, $P\{\tau_1 \neq N|\mathbf{P}\} \geq 3/4$.

Returning to the case when $\varepsilon > 0$ and sufficiently small, we note that in the Markov chain induced by $\mathbf{P}$ there are now two types of transitions: *strong* transitions that correspond to $\mathbf{P}$ assigning a positive probability to edges that are actually in the graph and *weak* transitions that are strictly the result of our perturbation. The latter are of order ε. Thus, the only impact that the perturbation makes on the argument presented above is to introduce an adjustment of order ε. This completes the proof. □

Theorem 3.6. *Consider a non-Hamiltonian graph G of order N, and define $\Delta(N) = 3/(8N^2)$.*

(i) For any $\mathbf{P}$, $\mathbb{E}_1[(\tau_1 - N)^2] \geq 3/4 - O(\varepsilon)$.

(ii) The following lower bound holds:

$$\{\min_{\mathbf{P} \in \mathcal{D}^\varepsilon} w_1^\varepsilon(\mathbf{P})\} - w_1^\varepsilon(\mathbf{P}_{\text{H}}) \geq \Delta(N) - O(\varepsilon).$$

Proof. Let $\mathbf{P}$ be an arbitrary doubly stochastic matrix and $\mathbb{E}_1[(\tau_1 - N)^2]$ be the corresponding variance of the first return time to vertex 1, starting from 1. Clearly,

$$\begin{aligned} \mathbb{E}_1[(\tau_1 - N)^2] &= \sum_{k \geq 1} (k-N)^2 P\{\tau_1 = k | \mathbf{P}\} \\ &\geq \sum_{k \geq 1,\ k \neq N} P\{\tau_1 = k | \mathbf{P}\} \\ &= P\{\tau_1 \neq N | \mathbf{P}\}. \end{aligned}$$

Hence by Part (ii) of Lemma 3.5, we have obtained Part (i), namely

$$\mathbb{E}_1[(\tau_1 - N)^2] \geq 3/4 - O(\varepsilon). \tag{3.43}$$

It now follows from (3.41) that

$$w_1^\varepsilon \geq \frac{N+1}{2N} + \frac{1}{2N^2}(3/4 - O(\varepsilon)) = \frac{N+1}{2N} + \Delta(N) - O(\varepsilon). \tag{3.44}$$

Part (ii) now follows immediately from (3.44) and (3.42). □

In summary, in this chapter, we considered spaces of probability transition matrices of Markov chains induced by a given graph G and certain associated random variables such as first return times to given vertices. We showed that, in a prescribed sense, the variances of the latter are minimised precisely at those Markov chains that are induced by Hamiltonian cycles of the graph whenever it possesses such cycles. Furthermore, we showed that the perturbed variance functional differentiates between all Hamiltonian and non-Hamiltonian graphs of a given order, by means of the size of the gap between minimal values of that functional over the space of doubly stochastic probability transition matrices induced by G. This suggests that stochastic—perhaps even statistical—methods could be brought to bear on this essentially deterministic, combinatorial, problem.