Fastest Expected Time to Mixing for a Markov Chain on a Directed Graph

Steve Kirkland*
Hamilton Institute
National University of Ireland Maynooth
Ireland
stephen.kirkland@nuim.ie

Abstract

For an irreducible stochastic matrix T, the Kemeny constant $K(T)$ measures the expected time to mixing of the Markov chain corresponding to T. Given a strongly connected directed graph D, we consider the set Σ_D of stochastic matrices whose directed graph is subordinate to D, and compute the minimum value of K, taken over the set Σ_D. The matrices attaining that minimum are also characterised, thus yielding a description of the transition matrices in Σ_D that minimise the expected time to mixing. We prove that $K(T)$ is bounded from above as T ranges over the irreducible members of Σ_D if and only if D is an intercyclic directed graph, and in the case that D is intercyclic, we find the maximum value of K on the set Σ_D. Throughout, our results are established using a mix of analytic and combinatorial techniques.

Keywords: Stochastic matrix; Directed graph; Kemeny constant.

AMS Classification Numbers: 15A51; 60J10; 15A42.

*This material is based upon works supported in part by the Science Foundation Ireland under Grant No. SFI/07/SK/I1216b.

1 Introduction and preliminaries

Let T be an irreducible stochastic matrix of order n, and denote the stationary distribution of T by π^T. Fix an index i between 1 and n, and recall that the *Kemeny constant* for T is given by $K(T) = \sum_{j=1}^n m_{ij}\pi_j$, where for each $i, j = 1, \ldots, n, m_{ij}$ denotes the mean first passage time from state i to state j (here we take the convention that $m_{ii} = 0$). It turns out that, remarkably, $K(T)$ is independent of the choice of i ([9]). Despite its probabilistic formulation, the Kemeny constant can be computed from the eigenvalues of T as follows: denoting the eigenvalues of T by $1 \equiv \lambda_1, \lambda_2, \ldots, \lambda_n$, we have (see [12])

$$K(T) = \sum_{j=2}^{n} \frac{1}{1-\lambda_j}. \tag{1}$$

Indeed, that expression is used (in [8], for example) to show that $K(T) \geq \frac{n-1}{2}$, with equality holding if T happens to be the adjacency matrix of a directed n-cycle. Observe that the right hand side of (1) is well-defined for any stochastic matrix T having 1 as a simple (i.e. algebraically and geometrically simple) eigenvalue. Consequently, we slightly extend the definition of the Kemeny constant to the class of stochastic matrices having 1 as a simple eigenvalue, and take $K(T)$ to be given by (1) for all such matrices.

The Kemeny constant admits several interpretations for the Markov chain associated with an irreducible stochastic matrix. In [8], it is shown that $K(T)+1$ coincides with the expected time to mixing for the chain. Here is the idea: let Y be a random variable with probability distribution given by π^T; sample Y, say $Y = j$ (with probability π_j), and start the chain $\{X_m\}$ at state $X_0 = i$; define the *time to mixing*, M, to be the minimum $k \geq 1$ such that $X_k = j$. It then follows that the expected value for M coincides with $K(T) + 1$. In a somewhat different direction, using the fact that $K(T) = \sum_{i=1}^n \pi_i \sum_{j=1}^n m_{ij}\pi_j$, the Kemeny constant is interpreted in [12] as the mean first passage time from an unknown initial state to an unknown destination state.

It is also noteworthy that the Kemeny constant provides a measure of the conditioning of the stationary distribution under perturbation of the underlying transition matrix. Specifically, if T and $T + E$ are two irreducible stochastic matrices of order n with stationary distributions π^T and $\tilde{\pi}^T$ respectively, then as shown in [8], we have

$$||\pi^T - \tilde{\pi}^T||_1 \leq K(T)||E||_\infty. \tag{2}$$

For any stochastic matrix T of order n, the *directed graph associated with* $T, \mathcal{D}(T)$, is the directed graph on vertices labeled $1, \ldots, n$, such that for each $i, j = 1, \ldots, n, i \to j$ is an arc in $\mathcal{D}(T)$ if and only if $t_{ij} > 0$. Note that $\mathcal{D}(T)$ carries qualitative information about the Markov chain associated with T, since the arcs of $\mathcal{D}(T)$ correspond to the transitions that are possible in a single step of the Markov chain. (We refer the reader to [2] for background on the interplay between square matrices and their directed graphs.) In this paper, we consider the effect of the combinatorial structure of $\mathcal{D}(T)$ on the value of $K(T)$. Specifically, for a strongly connected directed graph D on n vertices, we define the set Σ_D as follows:

$$\Sigma_D = \{T | T \text{ is stochastic and } n \times n \text{ and for each } i, j = 1, \ldots, n,$$
$$i \to j \text{ is an arc in } \mathcal{D}(T) \text{ only if } i \to j \text{ is an arc in } D\}.$$

Observe that Σ_D is a compact, convex set of matrices, whose irreducible members are dense in Σ_D. One of our main results, Theorem 2.6, provides a formula for $\min\{K(T) | T \in \Sigma_D\}$, while Theorem 2.13 characterises the matrices yielding that minimum value, thus identifying those transition matrices in Σ_D that minimise the expected time to mixing. The following example illustrates the scenario that we address in this paper.

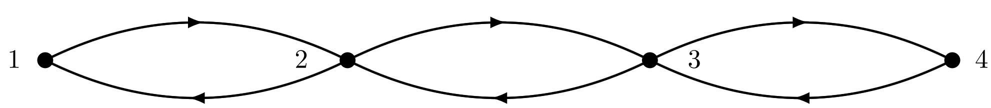

Figure 1: Directed graph D for Example 1.1

Example 1.1 Consider the directed graph D shown in Figure 1. A typical irreducible matrix $T \in \Sigma_D$ has the form $T = \begin{bmatrix} 0 & 1 & 0 & 0 \\ x & 0 & 1-x & 0 \\ 0 & 1-y & 0 & y \\ 0 & 0 & 1 & 0 \end{bmatrix}$, where $x, y \in (0, 1)$.

It is straightforward to determine that the eigenvalues of such a T are given by $1, -1, \sqrt{xy}$, and $-\sqrt{xy}$. Consequently we find that $K(T) = \frac{1}{2} + \frac{2}{1-xy}$. In particular we have $K(T) > \frac{5}{2}$ for any irreducible $T \in \Sigma_D$; note also that $K(T)$ is unbounded from above as T ranges over the irreducible members of Σ_D.

One of the key techniques employed in this paper involves the *group inverse* of $I - T$, which we now briefly outline. For a stochastic matrix T having 1 as a simple eigenvalue, the singular matrix $I - T$ is known to have a group inverse, $(I-T)^{\#}$, that can be characterised as the unique matrix such that $(I-T)(I-T)^{\#} = (I-T)^{\#}(I-T), (I-T)(I-T)^{\#}(I-T) = (I-T)$, and $(I-T)^{\#}(I-T)(I-T)^{\#} = (I-T)^{\#}$. If the eigenvalues of T are given by $1, \lambda_2, \ldots, \lambda_n$, then the eigenvalues of $(I-T)^{\#}$ are given by $0, \frac{1}{1-\lambda_2}, \ldots, \frac{1}{1-\lambda_n}$. In particular, we find that $K(T) = trace((I-T)^{\#})$. If T is a stochastic matrix with 1 as a simple eigenvalue, it follows from Lemma 3.3 of [11] that there is a neighbourhood of T such that $(I-\tilde{T})^{\#}$ is a well-defined continuous function for any stochastic matrix $\tilde{T}$ in that neighbourhood. In particular we see that K is continuous in some neighbourhood of T. We refer the interested reader to [3] for further information on generalised inverses.

Throughout the paper, we make use of standard facts on stochastic matrices. The reader may refer to [14] for the necessary background. Background material on directed graphs may be found in [5].

We close this section with a remark on the title of this paper. There is an existing body of work on the so-called fastest mixing Markov chain on a graph (see [1]). Results in that area focus on reversible Markov chains having a specified undirected graph, and on the *subdominant eigenvalue* of the corresponding transition matrix – i.e. the eigenvalue of second largest modulus after 1. The object is then to identify the reversible Markov chain respecting that graph which minimizes the modulus of the subdominant eigenvalue of the corresponding transition matrix. Our results in Section 2 bear a philosophical resemblance to those on fastest mixing Markov chains, for we consider an underlying combinatorial structure (a directed graph), and a measure of how quickly a Markov chain mixes ($K(T) + 1$ in our case); we then identify those transition matrices that simultaneously respect the combinatorial structure and minimise our measure of mixing time.

2 The minimum Kemeny constant on a directed graph

Throughout this section, we take D to be a strongly connected directed graph on n vertices, and we let k denote the length of a longest cycle in D. Let $\mu(D) = inf\{K(T)|T \in \Sigma_D$ and T has 1 as a simple eigenvalue$\}$. We begin with a useful result that leads to an upper bound on $\mu(D)$.

Lemma 2.1 *Suppose that $T \in \Sigma_D$ has the form*

$$T = \left[\begin{array}{c|c} C & 0 \\ \hline X & N \end{array}\right], \tag{3}$$

where N is nilpotent, and C is the adjacency matrix of a directed cycle of length ℓ. Then $K(T) = \frac{2n-\ell-1}{2}$.

Proof. It is straightforward to determine that

$$(I-T)^{\#} = \left[\begin{array}{c|c} (I-C)^{\#} & 0 \\ \hline (I-N)^{-1}X(I-C)^{\#} - \frac{1}{\ell}(I-N)^{-1}\mathbf{1}\mathbf{1}^T & (I-N)^{-1} \end{array}\right],$$

where we use $\mathbf{1}$ to denote an all ones vector of the appropriate order. Hence $K(T) = K(C) + n - \ell$. From Theorem 3 of [10], we find that each diagonal entry of $(I-C)^{\#}$ is equal to $\frac{\ell-1}{2\ell}$, so that $K(C) = \frac{\ell-1}{2}$; the conclusion now follows. $\square$

Corollary 2.2 *We have $\mu(D) \leq \frac{2n-k-1}{2}$.*

Proof. Note that there is a spanning subgraph of D such that each vertex has outdegree 1, and which contains exactly one directed cycle, of length k. Indeed, such a subgraph $\tilde{D}$ can be constructed as follows. Begin by identifying a k-cycle in D, let V_0 denote the subset consisting of the vertices on that cycle, and let A_0 denote the collection of arcs on that cycle. Then, for each $l \geq 0$ such that $|\cup_{p=0}^{l} V_p| < n$, let V_{l+1} denote the set of all vertices in D from which there is an out-arc to some vertex in V_l; for each $i \in V_{l+1}$, select a single vertex $j_i \in V_l$ such that $i \to j_i$ is an arc in D, and let A_{l+1} denote a collection of arcs $i \to j_i, i \in V_{l+1}$. For some smallest index m we have $|\cup_{p=0}^{m} V_p| = n$, and now we let $\tilde{D}$ be the (spanning) subgraph of D whose arc set is $\cup_{p=0}^{m} A_p$.

Let A be the adjacency matrix of such a subgraph $\tilde{D}$, and note then that $A \in \Sigma_D$. Observe that A can be written in the form (3), with C as the adjacency matrix of the directed k-cycle. From Lemma 2.1, we have $K(A) = \frac{2n-k-1}{2}$, from which the conclusion follows. $\square$

Our next result shows that if T is a stochastic matrix such that $K(T)$ is not too large, then the non-Perron eigenvalues of T are bounded away from 1.

Lemma 2.3 *Suppose that A is a stochastic matrix of order n having 1 as a simple eigenvalue, and let $\lambda \neq 1$ be an eigenvalue of A. If $K(A) \leq n$, then $|1-\lambda| \geq \frac{1-\cos\left(\frac{2\pi}{n}\right)}{n}$.*

Proof. Suppose first that $\lambda \in \mathbb{R}$. In that case, we have $\frac{1}{|1-\lambda|} = \frac{1}{1-\lambda} \leq K(A) \leq n$, so that $|1-\lambda| \geq \frac{1}{n}$, and the desired inequality follows.

Next we suppose that λ is complex, say with $\lambda = x + iy$. We then have $n \geq K(A) \geq \frac{1}{1-\lambda} + \frac{1}{1-\bar{\lambda}} = \frac{2(1-x)}{(1-x)^2+y^2}$. From Theorem 2 of [6] we have $|y| \leq (1-x)\frac{\sin\left(\frac{2\pi}{n}\right)}{1-\cos\left(\frac{2\pi}{n}\right)}$, so that $y^2 \leq (1-x)^2\frac{\sin^2\left(\frac{2\pi}{n}\right)}{(1-\cos\left(\frac{2\pi}{n}\right))^2}$. It now follows that $\frac{2(1-x)}{(1-x)^2+y^2} \geq \frac{1-\cos\left(\frac{2\pi}{n}\right)}{1-x}$. Hence we find that $|1-\lambda| \geq 1-x \geq \frac{1-\cos\left(\frac{2\pi}{n}\right)}{n}$. $\square$

Next, we show that while $\mu(D)$ is defined as an infimum, it is in fact a minimum.

Lemma 2.4 *There is a matrix $S \in \Sigma_D$ such that S has 1 as a simple eigenvalue, and $K(S) = \mu(D)$.*

Proof. From the definition of $\mu(D)$, we find that there is a sequence of matrices $T_m \in \Sigma_D$, each with 1 as a simple eigenvalue, such that $K(T_m) \to \mu(D)$ as $m \to \infty$. As Σ_D is compact, there is a subsequence T_{m_j} of T_m such that T_{m_j} converges in Σ_D as $j \to \infty$. Denote $\lim_{j\to\infty} T_{m_j}$ by S.

Since $K(T_{m_j}) \leq n$ for all sufficiently large j, we find from Lemma 2.3 that for all such j, and any eigenvalue $\lambda \neq 1$ of T_{m_j}, $|1-\lambda| \geq \frac{1-\cos\left(\frac{2\pi}{n}\right)}{n}$. It now follows that the matrix S has 1 as a simple eigenvalue. Thus, the function K is continuous in a neighbourhood of S, and so we find that $K(S) = \lim_{j\to\infty} K(T_{m_j}) = \mu(D)$. $\square$

Our next technical result shows that there is a matrix with special structure that minimises K.

Lemma 2.5 *There is a $(0,1)$ matrix $A \in \Sigma_D$ such that 1 is a simple eigenvalue of A and $K(A) = \mu(D)$.*

Proof. Appealing to Lemma 2.4, let S be a matrix in Σ_D having 1 as a simple eigenvalue, such that $K(S) = \mu(D)$. If S is a $(0,1)$ matrix, there is nothing to show, so suppose that some row of S contains at least two positive entries. For concreteness, we take $s_{ip}, s_{iq} > 0$, for indices $i, p, q \in \{1, \ldots, n\}$ with $p \neq q$. We claim that there is another matrix in $\Sigma_D, \hat{S}$ say, such that $K(\hat{S}) = \mu(D)$, and in addition such that $\hat{S}$ has fewer nonzero entries than S does. The conclusion will then follow via an iterative argument.

Let $Q = I - S$, and for each $t \in [-s_{ip}, s_{iq}]$, let $E_t = te_i(e_p - e_q)^T$. Observe that $S + E_t \in \Sigma_D$ for all such t. Let π^T denote the stationary distribution for S. From Lemma 3.3 of [11], we find that for each $t \in [-s_{ip}, s_{iq}]$ such that $S + E_t$ has 1 as a simple eigenvalue, we have

$$(Q - E_t)^{\#} = Q^{\#}(I - E_tQ^{\#})^{-1} - \mathbf{1}\pi^T(I - E_tQ^{\#})^{-1}Q^{\#}(I - E_tQ^{\#})^{-1},$$

provided that $I - E_tQ^{\#}$ is invertible.

From the Sherman-Morrison formula (see [7] for example) we find that for any t such that $1 - t(e_p - e_q)^TQ^{\#}e_i \neq 0$, we have $(I - E_tQ^{\#})^{-1} = I + \frac{t}{1-t(e_p-e_q)^TQ^{\#}e_i}e_i(e_p - e_q)^TQ^{\#}$. Observe in particular that $(I - E_tQ^{\#})^{-1}\mathbf{1} = \mathbf{1}$ for any such t.

Next, we consider $K(S+E_t)$, and note that for all t such that $|t|$ is sufficiently small, we have

$$\begin{aligned} K(S + E_t) = trace((Q - E_t)^{\#}) = \\ trace(Q^{\#}) + \frac{t}{1 - t(e_p - e_q)^TQ^{\#}e_i} trace(Q^{\#}e_i(e_p - e_q)^TQ^{\#}) \\ -trace(\mathbf{1}\pi^T(I - E_tQ^{\#})^{-1}Q^{\#}(I - E_tQ^{\#})^{-1}). \end{aligned}$$

Recalling that for any square rank one matrix ab^T, we have $trace(ab^T) = b^Ta$, we find that $trace(Q^{\#}e_i(e_p - e_q)^TQ^{\#}) = (e_p - e_q)^TQ^{\#}Q^{\#}e_i$. Also, $trace(\mathbf{1}\pi^T(I - E_tQ^{\#})^{-1}Q^{\#}(I - E_tQ^{\#})^{-1}) = \pi^T(I - E_tQ^{\#})^{-1}Q^{\#}(I - E_tQ^{\#})^{-1})\mathbf{1} = \pi^T(I - E_tQ^{\#})^{-1}Q^{\#}\mathbf{1} = 0$. Consequently, we have

$$\begin{aligned} K(S + E_t) = trace(Q^{\#}) + \frac{t}{1 - t(e_p - e_q)^TQ^{\#}e_i} Q^{\#}e_iQ^{\#}(e_p - e_q)^T = \\ K(S) + \frac{t}{1 - t(e_p - e_q)^TQ^{\#}e_i}(e_p - e_q)^TQ^{\#}Q^{\#}e_i. \end{aligned}$$

From the fact that S minimises K over Σ_D, we deduce that $(e_p - e_q)^TQ^{\#}Q^{\#}e_i$ must be zero, otherwise we could select a small (positive or negative) t so that $K(S+E_t) < \mu(D)$, a contradiction. Consequently we find that for all $t \in [-s_{ip}, s_{iq}]$ such that $1-t(e_p-e_q)^T Q^{\#} e_i \neq 0, K(S+E_t) = trace((Q-E_t)^{\#}) = trace(Q^{\#}) = \mu(D)$. Now select $t_0 = \begin{cases} s_{iq}, & \text{if } (e_p-e_q)^T Q^{\#} e_i \leq 0 \\ -s_{ip}, & \text{if } (e_p-e_q)^T Q^{\#} e_i > 0 \end{cases}$, so that $1-t_0(e_p-e_q)^T Q^{\#} e_i \geq 1$. Then $S+E_{t_0} \in \Sigma_D$, has 1 as a simple eigenvalue, has one more zero entry than S does, and satisfies $K(S+E_{t_0}) = \mu(D)$. The conclusion now follows. $\square$

Next, we present one of the main results of this paper.

Theorem 2.6 *Let D be a strongly connected directed graph on n vertices; denote the length of a longest cycle in D by k. Then*

$$\mu(D) = \frac{2n-k-1}{2}. \tag{4}$$

Proof. By Lemma 2.5, there is a $(0,1)$ matrix $A \in \Sigma_D$ having 1 as a simple eigenvalue, and such that $K(A) = \mu(D)$. Since A is $(0,1)$ with 1 as a simple eigenvalue, it follows that A can be written in the form (3), where C is the adjacency matrix of a directed cycle, say of length ℓ, and where N is nilpotent. By Lemma 2.1, we have $\mu(D) = K(A) = \frac{2n-\ell-1}{2} \geq \frac{2n-k-1}{2}$. Applying Corollary 2.2, we also have $\mu(D) \leq \frac{2n-k-1}{2}$, whence $\ell = k$; formula (4) now follows. $\square$

Corollary 2.7 *Let $T \in \Sigma_D$ be irreducible with stationary distribution π^T, and denote the corresponding mean first passage times by $m_{ij}, i, j = 1, \ldots, n$. For each index $i = 1, \ldots, n$, there is an index $j \neq i$ such that $m_{ij} \geq \frac{2n-k-1}{2(1-\pi_i)}$.*

Proof. From Theorem 2.6, we have $K(T) \geq \frac{2n-k-1}{2}$. Since $K(T) = \sum_{l \neq i} m_{il}\pi_l$, it follows that $\frac{K(T)}{1-\pi_i}$ is a weighted average of the quantities $m_{il}, l = 1, \ldots, n, l \neq i$. The conclusion now follows. $\square$

Remark 2.8 Observe that if T is the adjacency matrix of a directed n-cycle, then Corollary 2.7 asserts that for each $i = 1, \ldots, n$ there is a $j \neq i$ such that $m_{ij} \geq \frac{n}{2}$. If n happens to be even, then we can always find a $j \neq i$ so that in fact $m_{ij} = \frac{n}{2}$.

Example 2.9 Suppose that T is an irreducible tridiagonal stochastic matrix. From the structure of T, we find that length of a longest cycle in $\mathcal{D}(T)$ is 2.

Hence, $K(T) \geq \frac{2n-3}{2}$ by Theorem 2.6. Thus we have a generalisation of the observations made in Example 1.1.

Our next sequence of results is aimed at characterising the matrices $T \in \Sigma_D$ such that $K(T) = \mu(D)$. We begin with a continuity result for minimisers of K.

Lemma 2.10 *Let T_j be a sequence of matrices in Σ_D such that $K(T_j) = \mu(D)$ for all $j \in \mathbb{N}$. If the sequence T_j converges to S, then $K(S) = \mu(D)$.*

in *Proof.* Since $K(T_j) = \mu(D)$ for each j, we find from Theorem 2.6 and Lemma 2.3 that for any index j and eigenvalue $\lambda \neq 1$ of T_j, we have $|1-\lambda| \geq \frac{1-\cos\left(\frac{2\pi}{n}\right)}{n}$. Since the non-Perron eigenvalues of T_j are bounded away from 1, uniformly in j, it follows that 1 is a simple eigenvalue of S. Hence K is continuous in a neighbourhood of S, from which we conclude that $K(S) = \mu(D)$. □

Corollary 2.11 *Suppose that $T \in \Sigma_D$ and that $K(T) = \mu(D)$. Suppose also that there are indices i, p, q with $p \neq q$ such that $t_{ip}, t_{iq} > 0$. Letting $S = T + (-t_{ip})e_i(e_p - e_q)^T$, we have that $K(S) = \mu(D)$.*

Proof. Here we adopt the approach of Lemma 2.5. Let $Q = I - T$, and for each $s \in [-t_{ip}, t_{iq}]$, let $A_s = T + se_i(e_p - e_q)^T$. Then for each such s, we have $K(A_s) = \mu(D) + \frac{s(e_p-e_q)^T Q^\# Q^\# e_i}{1-s(e_p-e_q)^T Q^\# e_i}$, provided that $1 - s(e_p - e_q)^T Q^\# e_i \neq 0$. As in Lemma 2.5, we deduce that $(e_p - e_q)^T Q^\# Q^\# e_i = 0$, so that $K(A_s) = \mu(D)$ for each $s \in [-t_{ip}, t_{iq}]$ such that $1 - s(e_p - e_q)^T Q^\# e_i \neq 0$.

Next, select a sequence $s_m \in [-t_{ip}, t_{iq}]$ such that $1 - s_m(e_p - e_q)^T Q^\# e_i \neq 0$ for all $m \in \mathbb{N}$, and such that $s_m \to -t_{ip}$ as $m \to \infty$. Then $A_{s_m} \to S$ as $m \to \infty$, and $K(A_{s_m}) = \mu(D)$ for all $m \in \mathbb{N}$. The conclusion now follows from Lemma 2.10. □

The following proposition establishes the combinatorial structure of matrices that minimise K over Σ_D.

Proposition 2.12 *Suppose that $A \in \Sigma_D$ and that $K(A) = \mu(D)$. Then every cycle in $\mathcal{D}(A)$ has length k, and any pair of cycles in $\mathcal{D}(A)$ must intersect.*

Proof. We proceed by induction on the number of arcs in $\mathcal{D}(A)$, and note that if $\mathcal{D}(A)$ has just two arcs, then the result is immediate.

Suppose that the conclusion holds for directed graphs with $m \geq 2$ arcs, and that $\mathcal{D}(A)$ has $m+1$ arcs. If each vertex of $\mathcal{D}(A)$ has outdegree one, then since 1 is necessarily a simple eigenvalue of A, it follows that $\mathcal{D}(A)$ has just one cycle. Since $K(A) = \mu(D)$, it follows that this cycle must have length k, as desired.

Suppose that some vertex of $\mathcal{D}(A)$ has outdegree at least two. Without loss of generality, we assume that $1 \to i$ and $1 \to j$ in $\mathcal{D}(A)$. From Corollary 2.11, it follows that we can find $A_1, A_2 \in \Sigma_D$ such that $K(A_1) = K(A_2) = \mu(D)$, and such that $\mathcal{D}(A_1) = \mathcal{D}(A) \setminus \{1 \to i\}$ and $\mathcal{D}(A_2) = \mathcal{D}(A) \setminus \{1 \to j\}$. Note that every cycle in $\mathcal{D}(A)$ not using the arc $1 \to i$ is a cycle in $\mathcal{D}(A_1)$, and so by the induction hypothesis, every such cycle has length k. On the other hand, any cycle in $\mathcal{D}(A)$ that uses the arc $1 \to i$ cannot use the arc $1 \to j$, and so is a cycle in $\mathcal{D}(A_2)$; again by the induction hypothesis, such a cycle must have length k. Hence, every cycle in $\mathcal{D}(A)$ has length k.

Now select two cycles C_1 and C_2 in $\mathcal{D}(A)$. If neither includes the arc $1 \to i$, then both are in $\mathcal{D}(A_1)$ and hence they must intersect by the induction hypothesis. Evidently if both C_1 and C_2 include the arc $1 \to i$ then they intersect. So, without loss of generality we may assume that C_1 includes the arc $1 \to i$, while C_2 does not. If C_2 includes the arc $1 \to j$, then it certainly intersects C_1, while if C_2 does not include the arc $1 \to j$, then both C_1 and C_2 are in $\mathcal{D}(A_2)$. Again the induction hypothesis applies, and we find that C_1 and C_2 intersect. That completes the induction step, and the conclusion follows. $\square$

Recall that an irreducible stochastic matrix T is *periodic with period* m if the greatest common divisor of the cycle lengths in $\mathcal{D}(T)$ is equal to m. In that case, the vertices of $\mathcal{D}(T)$ can be partitioned into subsets $S_1, \ldots, S_m$ such that $i \to j$ is an arc in $\mathcal{D}(T)$ only if there is an index $\ell = 1, \ldots, m$ such that $i \in S_\ell$ and $j \in S_{\ell+1}$ (with the convention that $S_{m+1} \equiv S_1$). These subsets $S_1, \ldots, S_m$ are known as the *cyclically transferring classes* for T.

We are now in a position to characterise the matrices that minimise K over Σ_D.

Theorem 2.13 *Suppose that $A \in \Sigma_D$. We have $K(A) = \mu(D)$ if and only if A can be written in the form*

$$A = \left[\begin{array}{c|c} A_0 & 0 \\ \hline X & N \end{array}\right], \tag{5}$$

where N is nilpotent (or empty in the case that $k = n$) and where A_0 is irreducible, and k-cyclic with one of its cyclically transferring classes of cardinality one.

Proof. Suppose that $K(A) = \mu(D)$; then A has 1 as a simple eigenvalue, and it follows that we may write A as

$$A = \left[\begin{array}{c|c} A_0 & 0 \\ \hline X & N \end{array}\right],$$

where A_0 is irreducible and the spectral radius of N is strictly less than 1. By Proposition 2.12, all cycles of $\mathcal{D}(A)$ have length k, and any two cycles intersect. Hence, all cycles of $\mathcal{D}(A_0)$ have length k, and any two cycles intersect; applying Theorem 6.2 of [4], we thus find that A_0 must be k-cyclic with one of its cyclically transferring classes having cardinality one.

It is straightforward to see that $K(A) = trace((I-A_0)^{\#})+trace((I-N)^{-1})$. Suppose for concreteness that A_0 is $m \times m$; from the structure of A_0, we find that its eigenvalues are $e^{\frac{2\pi ij}{k}}, j = 0, \ldots, k-1$, and 0 with algebraic multiplicity $m-k$. Hence $trace((I-A_0)^{\#}) = \frac{k-1}{2} + m - k$. Note also that $trace((I-N)^{-1}) \geq n - m$, with equality holding only if N is nilpotent. Thus we have $\frac{2n-k-1}{2} = \frac{k-1}{2} + m - k + trace((I-N)^{-1}) \geq \frac{k-1}{2} + m - k + n - m = \frac{2n-k-1}{2}$. We thus conclude that N must be nilpotent, as desired.

The converse is readily established. □

Remark 2.14 Suppose that T is an irreducible stochastic matrix of order n. From Theorem 2.6, we recover the known result that $K(T) \geq \frac{n-1}{2}$, while from Theorem 2.13, we find that $K(T) = \frac{n-1}{2}$ if and only if T is the adjacency matrix of a directed n-cycle.

Remark 2.15 Let D be a strongly connected directed graph. It is interesting to note that any matrix in Σ_D that minimises the Kemeny constant necessarily has a subdominant eigenvalue of modulus 1. Thus we see that by using $K(T)+1$ as a measure for the time to mixing, we obtain very different results than by using the modulus of the subdominant eigenvalue as a measure of the time to mixing.

3 An upper bound for intercyclic directed graphs

In light of the lower bound on K established in Theorem 2.6, it is natural to wonder about the structure of the directed graphs D such that $K(T)$ is bounded from above as T ranges over Σ_D. In this section, we address that question. We begin with a useful observation.

Remark 3.1 It is shown in Lemma 6.1 of [4] that if D contains a pair of vertex-disjoint cycles, then $K(T)$ is not bounded from above as T ranges over the irreducible matrices in Σ_D.

Recall that a directed graph is *intercyclic* if it has the property that any pair of its cycles intersect. A complete characterisation of this class of directed graphs is given in [13]. Observe that by Remark 3.1, if a directed graph D has the property that $K(T)$ is bounded from above as T ranges over the irreducible members of Σ_D, then necessarily D must be intercyclic.

The following technical result will be useful in the sequel.

Lemma 3.2 *Let D be an intercyclic directed graph. Then for any matrix $T \in \Sigma_D$, the number 1 is a simple eigenvalue of T.*

Proof. Fix $T \in \Sigma_D$; for each cycle C in $\mathcal{D}(T)$, let $w(C)$ denote the product of the entries in T corresponding to the arcs of C. Since any pair of cycles in $\mathcal{D}(T)$ intersect, it follows that the characteristic polynomial of T can be written as $\det(\lambda I - T) = \lambda^n - \sum_{C\in\mathcal{D}(T)} w(C)\lambda^{n-|C|}$, where the sum is taken over all cycles $C \in \mathcal{D}(T)$, and where $|C|$ denotes the number of vertices on the cycle C. It now follows from Descartes' rule of signs that $\det(\lambda I - T)$ has precisely one positive root, which is necessarily equal to 1. Hence 1 is a simple eigenvalue of T. $\square$

Lemma 3.2 leads to a continuity result for K.

Corollary 3.3 *Let D be an intercyclic directed graph. Then K is a continuous function on Σ_D; in particular, there is a matrix $A \in \Sigma_D$ such that $K(A) = \max\{K(T)|T \in \Sigma_D\}$.*

Proof. Let T be a stochastic matrix with 1 as a simple eigenvalue. Then $(I-T)^{\#}$ is continuous in a neighbourhood of T, and hence so is $K(T) = trace((I-T)^{\#})$. The other conclusion follows readily. $\square$

Here is the main result of this section.

Theorem 3.4 *Suppose that D is an intercyclic directed graph, and let g denote the length of a shortest cycle in D. Then* $\max\{K(T)|T \in \Sigma_D\} = \frac{2n-g-1}{2}$.

Proof. By Lemma 3.3, K attains its maximum value on Σ_D. Arguing as in Lemma 2.5, it is readily shown that in fact there is a $(0,1)$ matrix A in Σ_D for which $K(A)$ is maximum. As A has 1 as a simple eigenvalue and is $(0,1)$, it follows that A can be written in the form (3), where C is the adjacency matrix of a directed cycle of length ℓ, say. It then follows that $K(A) = \frac{2n-\ell-1}{2} \le \frac{2n-g-1}{2}$. On the other hand, we can readily produce a matrix T in Σ_D such that $\mathcal{D}(T)$ contains a single cycle of length g, and for which $K(T) = \frac{2n-g-1}{2}$. Consequently, it must be the case that $\max\{K(T)|T \in \Sigma_D\} = \frac{2n-g-1}{2}$, as desired. $\square$

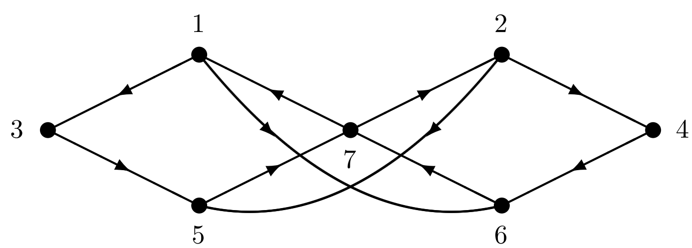

Figure 2: Directed graph D for Example 3.5

Example 3.5 We close the paper with an example that illustrates the results of this section. Consider the directed graph D shown in Figure 2. It is straightforward to see that D is intercyclic (since vertex 7 is on every cycle), and that the shortest and longest cycle lengths are 3 and 4, respectively.

Suppose that we have a matrix $T \in \Sigma_D$. Then there are parameters $x, y, a \in [0,1]$ such that

$$T = \begin{bmatrix} 0 & 0 & 1-x & 0 & 0 & x & 0 \\ 0 & 0 & 0 & 1-y & y & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ a & 1-a & 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

Let U be the submatrix of $I - T$ consisting of its first six columns, and let V be the 6×7 matrix $V = \left[\begin{array}{c|c} I & -\mathbf{1} \end{array} \right]$. We find from Theorem 7.8.2 of [3] that $(I-T)^{\#} = U(VU)^{-2}V$, from which it follows that $K(T) = trace(U(VU)^{-2}V) = trace((VU)^{-1})$. A direct computation now shows that $K(T) = 3 + \frac{6}{4-ax-(1-a)y}$.

Consequently, we find that $K(T) \geq \frac{9}{2} = \frac{14-4-1}{2}$, with equality holding if and only if $ax + (1-a)y = 0$, while $K(T) \leq 5 = \frac{14-3-1}{2}$, with equality holding if and only if $ax + (1-a)y = 1$.

Acknowledgement: The author is grateful to Robert Shorten and Selim Solmaz for a conversation that prompted the investigation in this paper.

References

[1] S. Boyd, P. Diaconis, L. Xiao, Fastest mixing Markov chain on a graph, SIAM Rev. 46 (2004) 667–689.

[2] R. Brualdi, H. Ryser, Combinatorial Matrix Theory, Cambridge University Press, Cambridge, 1991.

[3] S. Campbell, C. Meyer, Generalized Inverses of Linear Transformations, Dover Publications, New York, 1991.

[4] M. Catral, S. Kirkland, M. Neumann, N. Sze, The Kemeny constant for finite homogeneous ergodic Markov chains, J. Sci. Comput., to appear.

[5] G. Chartrand, L. Lesniak, Graphs and Digraphs, fourth edition, Chapman and Hall, Boca Raton, 2005.

[6] N. Dmitriev, E. Dynkin, On the characteristic numbers of a stochastic matrix, C. R. (Doklady) Acad. Sci. URSS (N.S.) 49 (1945) 159162.

[7] R. Horn, C. Johnson, Matrix Analysis, Cambridge University Press, Cambridge, 1985.

[8] J. Hunter, Mixing times with applications to perturbed Markov chains, Lin. Alg. Appl. 417 (2006) 108–123.

[9] J. Kemeny, J. Snell, Finite Markov Chains, Van Nostrand, Princeton, 1960.

[10] S. Kirkland, The group inverse associated with an irreducible periodic nonnegative matrix, SIAM J. Matrix Anal. Appl. 16 (1995) 1127–1134.

[11] S. Kirkland, A combinatorial approach to the conditioning of a single entry in the stationary distribution for a Markov chain, Electronic J. Linear Algebra 11 (2004) 168–179.

[12] M. Levene, and G. Loizou, Kemeny's constant and the random surfer, Amer. Math. Monthly 109 (2002) 741–745.

[13] W. McCuaig, Intercyclic digraphs, in: N. Robertson, P. Seymour (Eds.), Graph Structure Theory, Contemp. Math., 147, Amer. Math. Soc., Providence, 1993, pp. 203–245.

[14] E. Seneta, Non–Negative Matrices and Markov Chains, Springer, New York, 1981.