Minimising the largest mean first passage time of a Markov chain: the influence of directed graphs☆

Jane Breen*, Steve Kirkland

Department of Mathematics, University of Manitoba, Winnipeg MB, R3T 2N2, Canada

Abstract

For a Markov chain described by an irreducible stochastic matrix A of order n, the mean first passage time $m_{i,j}$ measures the expected time for the Markov chain to reach state j for the first time given that the system begins in state i, thus quantifying the short-term behaviour of the chain. In this article, a lower bound for the maximum mean first passage time is found in terms of the stationary distribution vector of A, and some matrices for which equality is attained are produced. The main objective of this article is to characterise the directed graphs for which any stochastic matrix A respecting this directed graph attains equality in this lower bound, producing a class of Markov chains with optimal short-term behaviour.

Keywords: Markov chain, mean first passage time, stochastic matrix, directed graph, Hessenberg matrix
2010 MSC: 15B51, 05C50, 60J10

1. Introduction

A stochastic matrix A is an entrywise nonnegative matrix whose rows sum to 1, i.e. $A\mathbb{1} = \mathbb{1}$, where $\mathbb{1}$ represents the vector of all ones. Stochastic matrices are central to the study of Markov chains, which are a type of probabilistic

☆Funding: This work was supported in part by NSERC and by the University of Manitoba Graduate Fellowship.

*Corresponding author

Email addresses: `breenj3@myumanitoba.ca` (Jane Breen), `Stephen.Kirkland@umanitoba.ca` (Steve Kirkland)

model describing dynamical systems which move between some finite number of states in discrete time-steps, where these transitions between states depend only on the present state occupied by the system. The connection with stochastic matrices is the following: given a Markov chain describing a system with a finite state space indexed by the integers $\{1, 2, \dots, n\}$, construct a matrix such that the $(i, j)^{th}$ entry is the probability of the system transitioning from state i to state j in a single time step. This is a stochastic matrix, referred to as the *probability transition matrix*, or simply *transition matrix* of the chain. This transition matrix A wholly represents the Markov chain, in that given an initial vector u_0 describing the probabilities that the process is in one of the various states at time 0, the probability distribution across all states after k time-steps is the vector $u_0^\top A^k$, for $k \geq 1$.

Representing a Markov chain by a stochastic matrix in this way enables us to analyse the long-term behaviour of the modelled system using basic techniques from linear algebra. If the transition matrix A is *irreducible* – i.e. for any pair of indices i, j, there exists some $m \in \mathbb{N}$ such that the $(i, j)^{th}$ entry of A^m is positive – then by the Perron-Frobenius theorem, A must have a strictly positive left eigenvector $w = [w_1\, w_2\, \cdots\, w_n]^\top$ corresponding to the eigenvalue 1. This eigenvector, when normalised so that the entries sum to 1 (thus producing a probability distribution) is referred to as the *stationary distribution vector* of the chain. It is an important quantity for the following reason: in the case that the transition matrix A is also *primitive* – i.e. there exists some $m \in \mathbb{N}$ for which every entry of A^m is positive – the iterates of the chain converge to $w^\top$ independent of the initial distribution. This is powerful in analysing the underlying system, since we can then say that the probability the Markov chain is in the i^{th} state in the long term is the i^{th} entry w_i of this eigenvector $w^\top$. Thus the long-term behaviour of the modelled system is summarised by a fundamental feature of the corresponding stochastic matrix.

The short-term behaviour of a system modelled by a Markov chain is considered as follows. Define $F_{i,j}$ to be a random variable representing the first passage time from state i to state j; i.e. the number of time steps elapsed (≥ 1) before the system reaches state j for the first time, given that it began in state i. The expected value of this random variable, then, is a key quantity of interest. It is referred to as the *mean first passage time from i to j*, denoted $m_{i,j}$. In the special case that $i = j$, $m_{i,i}$ is referred to as the *mean first return time* to state i. This facilitates the construction of the *matrix of mean first passage times* $M = [m_{i,j}]$ which is the unique solution (see [15, Section 6.1]) to the equation

$$M = A(M - M_{diag}) + J, \tag{1.1}$$

where M_{diag} is the diagonal matrix $diag([m_{1,1} \cdots m_{n,n}])$ and J is the matrix containing all ones. It is the off-diagonal entries of M which are the main subject of this article. We find a lower bound (in terms of the stationary distribution vector) for the maximum mean first passage time between any two distinct states in the system, thus giving some insight into how "well-connected" the Markov chain is.

Remark 1.1. Another measure of "connectivity" in a Markov chain with transition matrix A is the quantity $\kappa_i(A)$ given for some index $1 \leq i \leq n$ by

$$\kappa_i(A) = \sum_{\substack{j=1 \\ j \neq i}}^{n} m_{i,j} w_j,$$

which may be interpreted as the expected length of time to reach a randomly-chosen destination state from initial state i. Remarkably, this was shown to be independent of the choice of i in [11], and so $\kappa_i(A)$ is referred to as *Kemeny's constant* (for any i), and denoted instead by $\mathcal{K}(A)$. An alternative formula for Kemeny's constant was given in [17] as

$$\mathcal{K}(A) = \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j \neq i}}^{n} w_i m_{i,j} w_j,$$

admitting the interpretation of Kemeny's constant as the expected number of time-steps it takes the Markov chain system to move from a randomly-selected initial state to a randomly-selected destination state (where selection occurs with probabilities given by the stationary distribution). Thus a low value of Kemeny's constant observed for a transition matrix A implies that the Markov chain represented by A is, in a sense, "well-connected". A lower bound for $\mathcal{K}(A)$ (for A a stochastic matrix of order n) is given in [10] by $\mathcal{K}(A) \geq \frac{n-1}{2}$, while more recently, a lower bound in terms of the stationary distribution vector of A was determined in [14]. We will pursue a similar result here, but for the more widely-ranging mean first passage times, giving an alternate understanding of how "well-connected" the system is in the short term.

We begin by determining a lower bound on the maximum mean first passage time into state j in terms of the j^{th} entry of the stationary distribution vector.

Proposition 1.2. *Let $A = [a_{ij}]$ be an $n \times n$ stochastic irreducible transition matrix, with stationary vector w and mean first passage matrix M. Then for every $1 \leq j \leq n$,*

$$\max_{\substack{1 \leq i \leq n \\ i \neq j}} m_{i,j} \geq \frac{1}{w_j} - 1.$$

Equality is attained in this lower bound if and only if

1. *$a_{jj} = 0$; and*
2. *a_{jk} is nonzero only if $m_{k,j} = \frac{1}{w_j} - 1$.*

Proof. Without loss of generality, suppose $j = n$. Partition off the last row and column of A:

$$A = \left[\begin{array}{c|c} T & (I-T)\mathbb{1} \\ \hline r^\top & 1 - r^\top \mathbb{1} \end{array}\right].$$

From the eigen-equation for $w^\top$, it follows readily that

$$w_n = \frac{1}{1 + r^\top (I-T)^{-1}\mathbb{1}}. \tag{1.2}$$

Now suppose the mean first passage matrix M is partitioned conformally, so that

$$M = \left[\begin{array}{c|c} \hat{M} & y \\ \hline s^\top & m_{n,n} \end{array}\right].$$

By examining the matrix equation (1.1) using these block-partitioned matrices, it is easily seen that $y = (I-T)^{-1}\mathbb{1}$, i.e.

$$(I-T)^{-1}\mathbb{1} = \begin{bmatrix} m_{1,n} \\ m_{2,n} \\ \vdots \\ m_{n-1,n} \end{bmatrix}.$$

Hence

$$
\begin{aligned}
r^\top (I-T)^{-1}\mathbb{1} &= \sum_{i=1}^{n-1} r_i m_{i,n} \\
&\le \sum_{i=1}^{n-1} r_i \left(\max_{1\le i\le n-1} m_{i,n} \right) \\
&= r^\top \mathbb{1} \left(\max_{1\le i\le n-1} m_{i,n} \right) \\
&\le \max_{1\le i\le n-1} m_{i,n},
\end{aligned} \tag{1.3}
$$

since $r^\top \mathbb{1} \le 1$, and so we may conclude from (1.2) that

$$\max_{1\le i\le n-1} m_{i,n} \ge \frac{1}{w_n} - 1. \tag{1.4}$$

To investigate when equality holds in (1.4), simply examine the string of inequalities in (1.3), and observe that equality holds in the first one if and only if $r_i > 0 \Rightarrow m_{i,n}$ is maximum. Equality holds in the second inequality in (1.3) if and only if $r^\top \mathbb{1} = 1$; i.e. $a_{nn} = 0$.

□

Remark 1.3. In the proof of Proposition 1.2 an expression is obtained for $m_{i,n}$ as

$$m_{i,n} = e_i^\top (I-T)^{-1}\mathbb{1}, \quad \text{for } i = 1, \ldots, n-1.$$

The same techniques of partitioning may be applied after an appropriate permutation of the rows and columns of A and M to produce the following expressions for $m_{i,j}$:

$$m_{i,j} = \begin{cases} e_i^\top (I - A_{(j)})^{-1}\mathbb{1}, & \text{if } i < j \\ e_{i-1}^\top (I - A_{(j)})^{-1}\mathbb{1}, & \text{if } j < i, \end{cases}$$

where $A_{(j)}$ denotes the principal submatrix of order $n-1$ obtained by deleting the j^{th} row and column from A. We also note that by examining (1.2) along with (1.1) in block-partitioned form, it may be determined that $m_{n,n} = \frac{1}{w_n}$, and in general,

$$m_{i,i} = \frac{1}{w_i} \quad \text{for all } i = 1, \ldots, n.$$

The interested reader can find alternative derivations of these expressions using so-called absorbing chain techniques in [18, Theorem 4.5, pp.128-130].

Proposition 1.2 furnishes a lower bound on the overall maximum mean first passage time.

Corollary 1.4. *Let A be an $n \times n$ stochastic irreducible matrix, with stationary vector w and mean first passage matrix $M = [m_{i,j}]$. Then*

$$\max_{i \neq j} m_{i,j} \geq \frac{1}{\min_k w_k} - 1. \tag{1.5}$$

Observe that we can think of this result in terms of an optimisation problem – if the stationary distribution has been specified, then that places a lower bound on the maximum off-diagonal entry in the mean first passage matrix. Thus, if A yields equality in (1.5) then A has optimal performance (in terms of mean first passage times) subject to having a specified long-term behaviour (i.e. stationary distribution).

Such a result is of great value in many applications of Markov chains, particularly in those where mean first passage times may be used to determine some key feature of the modelled system. For example, Markov chains can be used as a model for traffic on an urban road network (see [6]) by letting the state space be a set of road segments and transition probabilities be determined by "turning probabilities" observed in reality. In this context, the mean first passage times represent average travel times between locations, a key aspect in the efficiency of the network. By producing a lower bound on the maximum expected travel time between two locations, we provide an indication of how well-connected the network might be.

Example 1.5. Let w be any probability vector, ordered so that w_1 is the smallest entry. Then the transition matrix

$$A = \frac{1}{1 - w_1}(\mathbb{1} w^\top - w_1 I)$$

is readily seen to yield equality in (1.5).

Example 1.6. Consider a stochastic companion matrix

$$A = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ a_1 & a_2 & a_3 & \cdots & a_n \end{bmatrix}$$

where the a_j's are nonnegative and sum to 1. It is straightforward to show that the stationary vector $w^\top$ of A is given by

$$w_j \quad = \quad \frac{\sum_{k=1}^{j} a_k}{\sum_{k=1}^{n}(n+1-k)a_k}, \qquad j = 1, \ldots, n.$$

Observe that the w_j's are nondecreasing, that $a_1 = \frac{w_1}{w_n}$, and that $a_j = \frac{w_j - w_{j-1}}{w_n}$, for $j = 2, \ldots, n$. Further, it can be shown that the mean first passage matrix M is given by

$$M = \begin{bmatrix} \frac{1}{w_1} & 1 & 2 & \cdots & n-2 & n-1 \\ \frac{1}{w_1} - 1 & \frac{1}{w_2} & 1 & 2 & \cdots & n-2 \\ \frac{1}{w_1} - 2 & \frac{1}{w_2} - 1 & \frac{1}{w_3} & 1 & \cdots & n-3 \\ \vdots & \vdots & \vdots & \ddots & & \vdots \\ \frac{1}{w_1} - (n-1) & \frac{1}{w_2} - (n-2) & \frac{1}{w_3} - (n-3) & \cdots & \frac{1}{w_{n-1}} - 1 & \frac{1}{w_n} \end{bmatrix}.$$

In particular, w_1 is the minimum entry in $w^\top$, and since $w_1 \leq \frac{1}{n}$, we also have $\frac{1}{w_1} - 1 \geq n - 1$; hence the maximum off-diagonal entry in M is $\frac{1}{w_1} - 1$, so that equality holds in (1.5).

The next example we consider is a very particular family of stochastic matrices, first discussed in [14] in the context of the characterisation of stochastic matrices A for which equality is attained in the following lower bound on Kemeny's constant:

$$\mathcal{K}(A) \geq \sum_{j=1}^{n} (j-1) w_j, \tag{1.6}$$

where the stationary vector $w^\top = \begin{bmatrix} w_1 & \cdots & w_n \end{bmatrix}$ is ordered so that the entries are in nondecreasing order. As we have already discussed, Kemeny's constant provides a measure of efficiency in a Markov chain, and so if this family of matrices can also be shown to attain equality in the lower bound on the mean first passage times given in (1.5), then these transition matrices represent Markov chains that can be considered to have optimal performance in two ways, subject to their having a specified stationary distribution.

Example 1.7. Let A be an irreducible stochastic matrix of order n with stationary distribution vector w, and suppose that the entries of w are in nondecreasing order. Then equality holds in (1.6) for A if and only if A is permutation equivalent to a matrix in the following family, for a fixed index k, $1 \leq k \leq n-1$:

$$\left[\begin{array}{cccc|ccccc|c}
0 & 0 & \cdots & 0 & 1 & 0 & \cdots & 0 & 0 & 0 \\
\frac{w_1}{w_2} & 0 & \cdots & 0 & \frac{w_2-w_1}{w_2} & 0 & \cdots & 0 & 0 & 0 \\
 & \ddots & & \vdots & \vdots & \vdots & & \vdots & \vdots & \\
0 & \cdots & \frac{w_{k-1}}{w_k} & 0 & \frac{w_k-w_{k-1}}{w_k} & 0 & \cdots & 0 & 0 & 0 \\
\hline
0 & 0 & \cdots & 0 & 0 & 1 & \cdots & 0 & 0 & 0 \\
0 & 0 & \cdots & 0 & 0 & 0 & 1 & & 0 & \\
\vdots & \vdots & & \vdots & \vdots & \vdots & & \ddots & \vdots & \vdots \\
0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 1 & 0 \\
0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 & 1 \\
\hline
0 & \cdots & 0 & \frac{w_k}{w_n} & \frac{w_{k+1}-w_k}{w_n} & \frac{w_{k+2}-w_{k+1}}{w_n} & \cdots & \frac{w_{n-2}-w_{n-3}}{w_n} & \frac{w_{n-1}-w_{n-2}}{w_n} & \frac{w_n-w_{n-1}}{w_n}
\end{array}\right].$$

To clarify the transition matrix, observe that the state space is partitioned as $\{1,\ldots,k\},\{k+1,\ldots,n-1\},\{n\}$. Note that there are two degenerate cases, when $k=1$ and $k=n-1$; each produces a stochastic matrix with a companion matrix pattern.

To prove that these matrices also yield equality in (1.5), we present the mean first passage matrix here, and leave it to the reader to confirm that it is the unique solution to the equation (1.1).

Fix an index k. Then $m_{i,j}$ is given by:

$$m_{i,j} = \begin{cases}
\frac{1}{w_j}(1-\sum_{r=1}^{j-1} w_r) + \frac{1}{w_i}(\sum_{r=1}^{i-1} w_r) & 1 \leq i < j \leq k; \\
\frac{1}{w_j} - \frac{1}{w_i}(\sum_{r=1}^{i-1} w_r) & 1 \leq j < i \leq k; \\
\frac{1}{w_j}(1-\sum_{r=1}^{j-1} w_r) - (i-k) & k+1 \leq i \leq n,\ 1 \leq j \leq k; \\
(j-k) + \frac{1}{w_i}(\sum_{r=1}^{i-1} w_r) & 1 \leq i \leq k,\ k+1 \leq j \leq n; \\
j-i & k+1 \leq i < j \leq n; \\
\frac{1}{w_j} + (j-i) & k+1 \leq j < i \leq n; \\
\frac{1}{w_i} & i=j.
\end{cases}$$

It may be determined that the largest off-diagonal entry of M occurs in the $(k+1,1)$ position, and is equal to $\frac{1}{w_1}-1$. We remark that results from Section 2 will render this a consequence of the fact that these matrices fall inside a larger family of stochastic matrices for which equality holds in (1.5).

These last examples clearly have very distinctive structure, indicating that the combinatorial influence of the *directed graph* of the transition matrix on the range of possible mean first passage times may be significant, and is worth investigating. That is the main objective of this paper.

In the following sections, we freely make use of basic concepts and definitions from graph theory, and refer the reader to [3] for the necessary background.

2. Directed graphs for which equality is always attained

Definition 2.1. Given an $n \times n$ matrix A, the *directed graph* of A, denoted $\mathcal{D}(A)$, consists of n vertices labelled 1 to n, with an arc (i, j) from vertex i to vertex j if and only if $a_{ij} \neq 0$, for $1 \leq i, j \leq n$.

The directed graph of a transition matrix for a Markov chain is a convenient way to visualise the movement of the system. Vertices represent states of the chain, and arcs represent possible transitions between them. If the arcs are considered to be weighted with the corresponding transition probabilities, this directed graph represents the Markov chain in its entirety. This representation is even more natural in such applications as the vehicular traffic model, where directed graph reflects the structure of the urban road network. Of course, the directed graph plays a greater role in analysing the transition matrix of a Markov chain than simple visual representation. To give an example, a matrix A is irreducible if and only if $\mathcal{D}(A)$ is *strongly connected*; that is, for any pair of distinct vertices i, j, there exists a directed path from i to j in $\mathcal{D}(A)$.

Since the network is predetermined in many applications, the following type of question is often posed: given a directed graph D that our transition matrix A must 'respect' (i.e. $\mathcal{D}(A)$ is some subgraph of D), what are the properties of A? In other words, what is the influence of the directed graph on the properties of the Markov chain? We formalise this with the following definition: given a directed graph D, let S_D denote the set of all stochastic matrices A such that $a_{ij} > 0$ only if (i, j) is an arc in D. This object has been used to determine the combinatorial influence of a given directed graph on the stationary vector (see [2, 13]) and Kemeny's constant (see [5, 12]) for all $A \in S_D$, and we now hope to do the same for mean first passage times. Given the structure noticed in the examples achieving equality in the lower bound (1.5) on the maximum mean first passage time, this seems like an appropriate thing to consider. We pose the following question: can we characterise the directed graphs D for which equality holds in the lower bound (1.5) for *every* irreducible $A \in S_D$? An answer to this question will determine the types of networks for which any Markov chain is "optimal" in the sense of minimising the maximum mean first passage time between distinct states.

As previously discussed in Remark 1.1, mean first passage times and Kemeny's constant are related concepts, although the mean first passage times encompass a larger quantity of information, while $\mathcal{K}(A)$ is a single-valued parameter of a Markov chain. Regardless, in similar questions asked of the range of values of $\mathcal{K}(A)$ for all $A \in S_D$ where D is a given directed graph (see [5, 12]) the cycle structure of D played an important role. Indeed, we will see a similar theme arising in the characterisation of digraphs D for which (1.5) holds for all $A \in S_D$. This is not unexpected, as the structure of D is already known to influence the mean first passage times of transition matrices in the family S_D. Recall the following result for mean first passage times: for a strongly connected directed graph D and some irreducible $A \in S_D$ with mean first passage times denoted by $m_{i,j}$, for any triple of indices i, j, k we have

$$m_{i,j} \leq m_{i,k} + m_{k,j}, \tag{2.1}$$

with equality if and only if k is distinct from i and j and every path in D from vertex i to vertex j passes through vertex k. This inequality was proven in [9] and a separate proof is given in [15, Theorem 6.2.1] which allows for the characterisation of equality, demonstrating the influence the network can have on mean first passage times between states of any Markov chain on that network; i.e. regardless of the transition probabilities. We will make use of this inequality several times in the sequel.

Remark 2.2. To demonstrate the usefulness of the 'triangle inequality' (2.1), consider the mean first passage matrix $M = [m_{i,j}]$ derived for the transition matrix given in Example 1.6. This could easily be derived using the observations that $m_{j,j} = \frac{1}{w_j}$ for $1 \leq j \leq n$; that $m_{i,j} = j - i$, for $2 \leq i < j \leq n$; and finally, that $m_{i,j} = \frac{1}{w_j} - (i - j)$, for $1 \leq j < i \leq n$, with this last observation following from the equality case of (2.1), as for $j < i$,

$$m_{j,i} + m_{i,j} = m_{j,j}$$

since a first passage from j to j must pass through i, and $m_{j,i} = i - j$.

Our first result Proposition 2.5 is a necessary condition on the cycle structure of directed graphs D for which equality holds for all $A \in S_D$. To prove it, we require the following definitions and a technical lemma whose proof further reinforces the idea that there is a strong relationship between the directed graph and mean first passage times of associated Markov chains.

Definition 2.3. A directed graph D is said to be *minimally strong* if D is strongly connected and the removal of any arc in D results in a digraph which is not strongly connected.

A matrix A is said to be *nearly reducible* if its directed graph $\mathcal{D}(A)$ is minimally strong, or equivalently, if A is an irreducible matrix such that setting any nonzero entry of A to 0 results in a reducible matrix.

Lemma 2.4. *Let D be a strongly connected directed graph on $n \geq 2$ vertices, labelled $v_1, v_2, \ldots, v_n$, and suppose that for some index j there exists an index $k \neq j$ such that for all irreducible $A \in S_D$ with mean first passage matrix M,*

$$m_{k,j} = \max_{\substack{1 \leq l \leq n \\ l \neq j}} m_{l,j}. \tag{2.2}$$

Then:

(a) For every cycle in D which contains v_k but not v_j, if there is a vertex v_i on the cycle with more than one arc issuing from it, then there is an arc from v_i to v_k in D.

(b) If there is a cycle in D containing both v_k and v_j, then there must be an arc from v_j to v_k in D.

Proof. Let D be a strongly connected digraph of order n such that the hypothesis (2.2) holds for all irreducible $A \in S_D$. Without loss of generality, suppose $j = n$, and let $D \setminus \{v_n\}$ denote the digraph obtained from D by the removal of v_n and all incident arcs. We will assume that there is a cycle in $D \setminus \{v_n\}$ containing v_k which permits a vertex v_i with outdegree greater than one (recalling that the outdegree of a vertex refers to the number of arcs issuing from it), such that (v_i, v_k) is not an arc in D. Then we will construct an irreducible matrix $\tilde{A} \in S_D$ such that

$$e_k^\top (I - \tilde{A}_{(n)})^{-1} \mathbb{1} \ < \ \max_{1 \leq l \leq n-1} e_l^\top (I - \tilde{A}_{(n)})^{-1} \mathbb{1},$$

contradicting (2.2).

To this end, first consider a directed cycle of length l with the following weighted adjacency matrix C. Here the weight of one arc in the cycle (and corresponding matrix entry) is equal to $t < 1$, while the remaining arcs on the cycle have weight 1:

$$C = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ t & 0 & 0 & \cdots & 0 \end{bmatrix}. \tag{2.3}$$

We have

$$\begin{aligned} (I-C)^{-1}\mathbb{1} &= \begin{bmatrix} 1 & -1 & 0 & \cdots & 0 \\ 0 & 1 & -1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 & -1 \\ -t & 0 & \cdots & 0 & 1 \end{bmatrix}^{-1} \mathbb{1} \\ &= \frac{1}{1-t}\begin{bmatrix} 1 & 1 & 1 & \cdots & 1 \\ 1 & t & 1 & \cdots & 1 \\ 1 & t & t & \cdots & 1 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & t & t & \cdots & t \end{bmatrix}\mathbb{1} \\ &= \frac{1}{1-t}\begin{bmatrix} l \\ t+(l-1) \\ 2t+(l-2) \\ \vdots \\ (l-1)t+1 \end{bmatrix}. \end{aligned}$$

With $t < 1$ this results in a uniquely maximum entry in the first position.

Now construct a matrix T of order $n-1$ in the pattern of $D \setminus \{v_n\}$ so that for some permutation matrix P,

$$PTP^{\top} = \left[\begin{array}{c|c} C & O \\ \hline X & N \end{array}\right] \tag{2.4}$$

where C is as in (2.3), with some row of C other than the first corresponding to v_k, and X and N chosen appropriately so that the rows sum to 1 and N

is nilpotent. To see that such a choice of X and N is possible, let V denote the subset of vertices in D which are not contained in the cycle represented by C. We construct a digraph D_1 of D by choosing a maximal subdigraph for which the directed graph induced by V does not contain a cycle. Then any matrix in S_{D_1} has the appropriate form (2.4) when we remove the row and column corresponding to v_n; that is, the matrix N is nilpotent. This is to ensure that $(I-T)^{-1}$ exists.

We have

$$\begin{aligned} \max_{1\le i\le n-1} e_i^\top (I-T)^{-1}\mathbb{1} &\ge \max_{1\le m\le l} e_m^\top (I-C)^{-1}\mathbb{1} \\ &> e_k^\top (I-C)^{-1}\mathbb{1} \\ &= e_k^\top (I-T)^{-1}\mathbb{1}. \end{aligned} \tag{2.5}$$

We will now show the existence of an irreducible stochastic matrix $\tilde{A} \in S_D$ such that $(I-\tilde{A}_{(n)})^{-1}$ is 'as close as we like' to $(I-T)^{-1}$ – that is, for a chosen matrix norm $\|\cdot\|$ and given $\varepsilon > 0$, we can find a matrix $\tilde{A}$ for which $\|(I-\tilde{A}_{(n)})^{-1} - (I-T)^{-1}\| < \varepsilon$.

Without loss of generality (and for ease of notation) suppose that the permutation matrix P above is the identity matrix. Then it is v_l which has outdegree greater than 1.

Let A be an $n \times n$ matrix such that $\mathcal{D}(A) \subseteq D$, and $A_{(n)} = T$. In particular, let

$$A = \left[\begin{array}{c|c|c} C & O & 0 \\ \hline X & N & y \\ \hline * & * & * \end{array}\right]$$

where $y = \mathbb{1} - X\mathbb{1} - N\mathbb{1}$ (for appropriately-sized vectors $\mathbb{1}$) and the last row may be chosen arbitrarily, so long as the row sums to 1, and $a_{ni} \neq 0$ only if (v_n, v_i) is an arc in D. Constructed in this way, A is reducible and strictly substochastic – the l^{th} row sums to $t < 1$. Furthermore,

$$e_k^\top (I-A_{(n)})^{-1}\mathbb{1} < \max_{1\le i\le l} e_i^\top (I-A_{(n)})^{-1}\mathbb{1}, \quad \text{from (2.5)}.$$

We first construct a new stochastic matrix $\hat{A}$ from A which retains the property that $e_k^\top (I-\hat{A}_{(n)})^{-1}\mathbb{1}$ is not maximum. To do this, we need only focus on the l^{th} row of A, corresponding to the vertex v_l with outdegree greater than 1.

- *Case 1:* (v_l, v_n) is an arc in D.
  Then $\hat{A}$ may be constructed by setting $\hat{A} = A + (1-t)e_l e_n^\top$. In this case, $\hat{A}_{(n)} = A_{(n)}$.

- *Case 2:* (v_l, v_n) is not an arc in D.
  It may be assumed that there is an arc from v_l to v_{l+p}, for some $1 \leq p < n-l$ (as otherwise, since D is strongly connected, there is some other vertex v_m on the cycle from which such an arc exists). Then $\hat{A}$ is constructed by setting

$$\hat{A} = A + (1-t)e_l e_{l+p}^\top.$$

Then $\hat{A}_{(n)} = A_{(n)} + (1-t)e_l e_{l+p}^\top$ (where the vectors are resized), and by the Sherman-Morrison-Woodbury formula (see [8, Section 0.7.4]):

$$(I - \hat{A}_{(n)})^{-1} = (I - A_{(n)})^{-1} + \frac{(1-t)(I - A_{(n)})^{-1} e_l e_{l+p}^\top (I - A_{(n)})^{-1}}{1 - (1-t)e_{l+p}^\top (I - A_{(n)})^{-1} e_l}.$$

The row sums of $(I - \hat{A}_{(n)})^{-1}$ are obtained by multiplying on the right by $\mathbb{1}$:

$$(I - \hat{A}_{(n)})^{-1}\mathbb{1} = (I - A_{(n)})^{-1}\mathbb{1} + \frac{(1-t)(I - A_{(n)})^{-1} e_l e_{l+p}^\top (I - A_{(n)})^{-1}\mathbb{1}}{1 - (1-t)e_{l+p}^\top (I - A_{(n)})^{-1} e_l},$$

and considering only the first l row sums, we have on the right-hand side:

$$\frac{1}{1-t}\begin{bmatrix} l \\ t + (l-1) \\ 2t + (l-2) \\ \vdots \\ (l-1)t + 1 \end{bmatrix} + \gamma_t \begin{bmatrix} 1 \\ 1 \\ 1 \\ \vdots \\ t \end{bmatrix},$$

where

$$\gamma_t = \frac{(1-t)e_{l+p}^\top (I - A_{(n)})^{-1}\mathbb{1}}{1 - (1-t)e_{l+p}^\top (I - A_{(n)})^{-1} e_l},$$

i.e. a scalar. Hence the maximum entry in the first l positions of $(I - \hat{A}_{(n)})^{-1}\mathbb{1}$ still occurs in the first position, not the one corresponding to vertex v_k.

In either case $\hat{A}$ is a matrix in S_D for which

$$e_k^\top (I - \hat{A}_{(n)})^{-1}\mathbb{1} < \max_{1\le i\le n-1} e_i^\top (I - \hat{A}_{(n)})^{-1}\mathbb{1}. \tag{2.6}$$

Next consider that $\hat{A}$ may be a reducible member of S_D. However, irreducible matrices in S_D are dense in S_D, and so given $\delta > 0$, there exists $\tilde{A} \in S_D$ irreducible such that $\|\hat{A}_{(n)} - \tilde{A}_{(n)}\| < \delta$.

Finally, recall that $f : B \mapsto B^{-1}$ is a continuous function on the set of real invertible matrices, and so for all $\varepsilon > 0$ there exists $\delta > 0$ such that $\|B_1 - B_2\| < \delta$ implies $\|B_1^{-1} - B_2^{-1}\| < \varepsilon$. Given $\varepsilon > 0$, we can find an irreducible $\tilde{A} \in S_D$ so that $\|\hat{A}_{(n)} - \tilde{A}_{(n)}\| < \delta$. Then $\|(I - \hat{A}_{(n)}) - (I - \tilde{A}_{(n)})\| < \delta$, which implies

$$\|(I - \hat{A}_{(n)})^{-1} - (I - \tilde{A}_{(n)})^{-1}\| < \varepsilon.$$

Hence we conclude using (2.6) that there exists an irreducible matrix $\tilde{A} \in S_D$ with mean first passage times $M = [m_{i,j}]$ for which

$$m_{k,n} = e_k^\top (I - \tilde{A}_{(n)})^{-1}\mathbb{1} \;<\; \max_{1\le i\le n-1} e_i^\top (I - \tilde{A}_{(n)})^{-1}\mathbb{1} = \max_{1\le i\le n-1} m_{i,n},$$

contradicting our hypothesis (2.2). We conclude that if (2.2) holds for all $A \in S_D$ for some digraph D, then D has the property that for any cycle containing v_k in $D \setminus \{v_j\}$, if a vertex v_i on the cycle has outdegree greater than 1 then (v_i, v_k) must be an arc in D.

To prove (b), suppose that (2.2) holds for all irreducible $A \in S_D$, and that there is a cycle in D containing v_k and v_j, but that (v_j, v_k) is not an arc in D. Let v_m be the vertex on the cycle to which there is an arc from v_j. Then we can immediately construct a reducible $\hat{A} \in S_D$ such that for some permutation matrix P,

$$P\hat{A}P^\top = \left[\begin{array}{c|ccccc} N & & & X & & \\ \hline & 0 & 1 & 0 & \cdots & 0 \\ & 0 & 0 & 1 & \cdots & 0 \\ O & \vdots & \vdots & \vdots & \ddots & \vdots \\ & 0 & 0 & 0 & \cdots & 1 \\ & 1 & 0 & 0 & \cdots & 0 \end{array}\right]$$

where v_j corresponds to the last row, v_m to the first row in the second diagonal block, and v_k to some other index in the second block of the partition. Without loss of generality, assume $P = I$, so that $j = n$. Clearly, if l is the length of the cycle in question (and hence the size of the second block in the partition) then $e_m^\top (I - \hat{A}_{(n)})^{-1}\mathbb{1} = l - 1$, while $e_k^\top (I - \hat{A}_{(n)})^{-1}\mathbb{1} < l - 1$. Using a continuity argument as before, we may conclude the existence of an irreducible $\tilde{A} \in S_D$ for which $m_{k,n} < \max_i m_{i,n}$, contradicting the hypothesis. Hence (v_j, v_k) must be an arc in D.

□

Proposition 2.5. *Let D be a minimally strong directed graph on n vertices, with the property that for any irreducible $A \in S_D$ with stationary vector w and mean first passage matrix M,*

$$\max_{\substack{1 \le i,j \le n \\ i \ne j}} m_{i,j} = \frac{1}{\min_k w_k} - 1.$$

Then D is a directed cycle.

Proof. Let D be minimally strong, and let $A \in S_D$. Then A is nearly reducible, and by [3, Theorem 3.3.4] (and without loss of generality on the ordering of the rows and columns) A has the following form:

$$\left[\begin{array}{ccccc|c} 0 & 1 & 0 & \cdots & 0 & \\ 0 & 0 & 1 & \cdots & 0 & \\ \vdots & \vdots & \vdots & \ddots & \vdots & e_p e_k^\top \\ 0 & 0 & 0 & \cdots & 1 & \\ 0 & 0 & 0 & \cdots & 0 & \\ \hline & & a e_j e_1^\top & & & T \end{array}\right] \tag{2.7}$$

where the first diagonal block has order p ($1 \le p \le n-1$), and T is a substochastic, nearly reducible matrix of order $n - p$. In particular, if $\hat{D} = \mathcal{D}(T)$, then $T = \hat{T} - a e_j e_l^\top$, for some fixed $\hat{T} \in S_{\hat{D}}$, $a \in (0, 1)$, and index l such that $\hat{t}_{jl} > a$. Moreover, $1 \le j, k \le n - p$, and $(p + j, p + k)$ is not an arc in D (as this would contradict the assumption that D is minimally strong). Furthermore, note that if $p = n - 1$ then we are done, so henceforth take p to be at most $n - 2$. For ease of notation, let $v_1, v_2, \ldots, v_p$ denote the vertices corresponding to the first p rows and columns of A, and $\hat{v}_1, \hat{v}_2, \ldots, \hat{v}_{n-p}$ the vertices of $\hat{D}$.

The stationary vector w of this matrix A can be computed as

$$w^\top = \frac{1}{p + e_k^\top (I-T)^{-1}\mathbb{1}} \left[\mathbb{1}_p^\top \mid \quad e_k^\top (I-T)^{-1} \right], \tag{2.8}$$

or using stochastic complementation (see [15, Section 5.1]) as

$$w^\top = \frac{1}{pa\tilde{w}_j + 1} \left[a\tilde{w}_j \mathbb{1}_p^\top \mid \quad \tilde{w}^\top \right]$$

where $\tilde{w}$ is the stationary vector of the stochastic complement $\tilde{T} = T + ae_j e_l^\top = \hat{T} + ae_j(e_k - e_l)^\top$. Note that for sufficiently small values of a, the minimum entry in w occurs in the first p positions and the lower bound for the maximum mean first passage times is $\frac{1}{w_p} - 1$.

We claim that equality holds in the lower bound (1.5) for all matrices $A \in S_D$ only if for all $\hat{T} \in S_{\hat{D}}$,

$$\hat{m}_{k,j} = \max_{1 \le i \le n-p} \hat{m}_{i,j} \tag{2.9}$$

where $\hat{M}$ is the mean first passage matrix for $\hat{T}$.

Proof of claim: Fix $\hat{T} \in S_{\hat{D}}$ and let $a \in (0,1)$ be chosen sufficiently small so that when A is formed as in (2.7), the minimum entry of the stationary vector w of A occurs in the first p positions. Let M be the mean first passage matrix for A, and suppose that equality holds in (1.5). Now consider the mean first passage times into state p, recalling that $m_{i,p} = e_i^\top (I - A_{(p)})^{-1}\mathbb{1}$:

$$(I - A_{(p)})^{-1} = \left[\begin{array}{ccccc|c} 1 & -1 & 0 & \cdots & 0 & \\ 0 & 1 & -1 & \cdots & 0 & \\ \vdots & \vdots & \ddots & \ddots & \vdots & O \\ 0 & 0 & \cdots & 1 & -1 & \\ 0 & 0 & \cdots & 0 & 1 & \\ \hline & & -ae_je_1^\top & & & I - T \end{array}\right]^{-1}$$

$$= \left[\begin{array}{cccc|c} 1 & 1 & \cdots & 1 & \\ 0 & 1 & \cdots & 1 & O \\ \vdots & \vdots & \ddots & \vdots & \\ 0 & 0 & \cdots & 1 & \\ \hline & J & & & (I - T)^{-1} \end{array}\right]$$

where J denotes the all-ones matrix of order $(n-p) \times (p-1)$. Then

$$(I - A_{(p)})^{-1}\mathbb{1} = \left[\begin{array}{c} p-1 \\ p-2 \\ \vdots \\ 1 \\ \hline (p-1)\mathbb{1} + (I-T)^{-1}\mathbb{1} \end{array}\right].$$

So

$$\max_{\substack{1 \le i \le n \\ i \ne p}} m_{i,p} = (p-1) + \max_{1 \le i \le n-p} e_i^\top (I-T)^{-1}\mathbb{1}$$

while the lower bound on $\max m_{i,p}$ is

$$\frac{1}{w_p} - 1 = p + e_k^\top (I-T)^{-1}\mathbb{1} - 1.$$

Hence for equality to hold in the lower bound (1.5) on the mean first passage times of A as assumed, it is required that

$$e_k^\top (I-T)^{-1}\mathbb{1} = \max_{1 \le i \le n-p} e_i^\top (I-T)^{-1}\mathbb{1}. \tag{2.10}$$

Observe, however, that $e_i^\top(I-T)^{-1}\mathbb{1} = m_{p+i,1}$, for $i = 1, \ldots, n-p$. Furthermore, since every path from a vertex in $\hat{D}$ to v_1 must pass through $\hat{v}_j$, it follows from the equality case of (2.1) that

$$m_{p+i,1} = m_{p+i,p+j} + m_{p+j,1} \quad \text{for all } i = 1, \ldots, n-p.$$

Hence (2.10) is equivalent to the condition that $m_{p+k,p+j} = \max_i m_{p+i,p+j}$. However, note that these mean first passage times are calculated as the sums of the lower rows of $(I - A_{(p+j)})^{-1}$ and that the principal submatrix $A_{(p+j)}$ has a lower left block of zeros, and its lower diagonal block is $T_{(j)}$. Recall that $T = \hat{T} - ae_je_l^\top$, and hence $T_{(j)} = \hat{T}_{(j)}$. Hence

$$\begin{aligned} m_{p+i,p+j} &= e_i^\top(I - \hat{T}_{(j)})^{-1}\mathbb{1} \\ &= \hat{m}_{i,j}, \end{aligned}$$

where $\hat{M}$ is the mean first passage matrix of $\hat{T}$. Hence (2.10) is equivalent to the condition that

$$\hat{m}_{k,j} = \max_{1 \le i \le n-p} \hat{m}_{i,j},$$

as claimed.

Suppose that $n - p \ge 2$. Using Lemma 2.4 applied to $\hat{D}$, it may be asserted that if (2.9) must hold for all $\hat{T} \in S_{\hat{D}}$, then if $\mathcal{C} \subseteq \hat{D}$ is a cycle through $\hat{v}_k$ which does not contain $\hat{v}_j$, then there is exactly one vertex (say $\hat{v}_m$) with out-degree greater than one, and $(\hat{v}_m, \hat{v}_k)$ is also an arc in $\hat{D}$. Since $\hat{v}_j$ is not in $\mathcal{C}$, there is a path in $\hat{D}$ from $\hat{v}_m$ to $\hat{v}_j$ which does not use the arc $(\hat{v}_m, \hat{v}_k)$. The existence of such a path determines the construction of a path from $\hat{v}_m$ to $\hat{v}_k$ in D, via $\hat{v}_j$ and $v_1, v_2, \ldots, v_p$, which does not use the arc $(\hat{v}_m, \hat{v}_k)$. Hence if this arc is deleted from D, the directed graph remains strongly connected, contradicting the assumption that D is minimally strong.

In the case that there is no such cycle $\mathcal{C}$ – that is, every cycle through $\hat{v}_k$ in $\hat{D}$ also contains $\hat{v}_j$ – then from Lemma 2.4, there must be an arc from $\hat{v}_j$ to $\hat{v}_k$. This contradicts the hypothesis that D is minimally strong.

Hence we must have $n - p = 1$, and the lower diagonal block in (2.7) is trivial. Therefore the only minimally strong directed graph D for which equality can hold for all $A \in S_D$ is the directed cycle on n vertices.

□

Since the question posed at the beginning of this section insists that *every* $A \in S_D$ must satisfy equality in the lower bound on mean first passage times, this result gives a great restriction on the directed graph D. Since D is strongly connected, it must contain the directed n-cycle as a subdigraph, and no other minimally strong digraphs may appear as a subdigraph of D. Thus each such D satisfying our requirements is built on this underlying cycle. Recall that such a cycle passing through every vertex of the digraph exactly once is referred to as a *Hamilton cycle*.

Proposition 2.6. *Let D be a strongly connected directed graph on n vertices such that for every irreducible $A \in S_D$ with stationary distribution vector w and mean first passage matrix M, equality holds in the lower bound* (1.5). *Then D has a unique Hamilton cycle.*

Proof. The existence of a Hamilton cycle as a subdigraph of D is a corollary to Proposition 2.5. To show uniqueness of the Hamilton cycle, suppose that D has two distinct Hamilton cycles as subdigraphs. The adjacency matrices of these subgraphs are both elements of S_D, which we denote A and A' and assume without loss of generality that the vertices of D are ordered in such a way that

$$A = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ 1 & 0 & 0 & \cdots & 0 \end{bmatrix}$$

and that A' has its last row different from the last row of A.

The stationary vector of A and of A' is clearly equal to $w^\top = \frac{1}{n}\mathbb{1}^\top$, and it is easily seen that w is also the stationary vector of any convex combination of A and A'; that is, any matrix in S_D of the form

$$A_t := (1-t)A + tA', \quad \text{for } 0 \le t \le 1.$$

We now consider the mean first passage times for A_t: in particular the mean first passage times into state n, which are given by $(I - A_{t_{(n)}})^{-1}\mathbb{1}$. First, however, note that

$$(I - A_{(n)})^{-1} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ 0 & 1 & \cdots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}, \tag{2.11}$$

and that $m_{j,n} = n - j$.

Now

$$\begin{aligned}
(I - A_{t_{(n)}})^{-1} &= (I - (1-t)A_{(n)} - tA'_{(n)})^{-1} \\
&= (I - A_{(n)} - t(A'_{(n)} - A_{(n)}))^{-1} \\
&= \left[(I - A_{(n)})(I - t(I - A_{(n)})^{-1}(A'_{(n)} - A_{(n)}))\right]^{-1} \\
&= (I - t(I - A_{(n)})^{-1}(A'_{(n)} - A_{(n)}))^{-1}(I - A_{(n)})^{-1} \\
&= [I + t(I - A_{(n)})^{-1}(A'_{(n)} - A_{(n)})](I - A_{(n)})^{-1} + O(t^2)
\end{aligned}$$

for small values of t. The maximum mean first passage time in column n then is given by the maximum entry of

$$\begin{aligned}
(I - A_{t_{(n)}})^{-1}\mathbb{1} &= (I - A_{(n)})^{-1}\mathbb{1} + t(I - A_{(n)})^{-1}(A'_{(n)} - A_{(n)})(I - A_{(n)})^{-1}\mathbb{1} + O(t^2) \\
&= \begin{bmatrix} n-1 \\ n-2 \\ \vdots \\ 1 \end{bmatrix} + t \begin{bmatrix} 1 & 1 & \cdots & 1 \\ 0 & 1 & \cdots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix} (A'_{(n)} - A_{(n)}) \begin{bmatrix} n-1 \\ n-2 \\ \vdots \\ 1 \end{bmatrix} + O(t^2).
\end{aligned}$$

Recalling (2.11), it may be determined that the maximum entry occurs in the first row; that is,

$$m_{1,n} = (n-1) + t\mathbb{1}^\top A'_{(n)} \begin{bmatrix} n-1 \\ n-2 \\ \vdots \\ 1 \end{bmatrix} - t \begin{bmatrix} 0 & 1 & \cdots & 1 \end{bmatrix} \begin{bmatrix} n-1 \\ n-2 \\ \vdots \\ 1 \end{bmatrix} + O(t^2).$$

Since the column sums of A' are all equal to one, $\mathbb{1}^\top A'_{(n)} = \mathbb{1}^\top - e_j^\top$, where j is the index of the nonzero entry in the n^{th} row of A' (equivalently, the index of the vertex to which an arc issues from vertex n in the Hamilton cycle represented by A'). Hence

$$\begin{aligned}
m_{1,n} &= n - 1 + t(\mathbb{1}^\top - e_j^\top) \begin{bmatrix} n-1 \\ n-2 \\ \vdots \\ 1 \end{bmatrix} - t\,\frac{n(n-1)}{2} + O(t^2) \\
&= n - 1 + t\left(\frac{n(n-1)}{2} - (n-j) - \frac{(n-1)(n-2)}{2}\right) + O(t^2) \\
&= n - 1 + t(j-1) + O(t^2),
\end{aligned}$$

and since by assumption j cannot equal 1, we have that for sufficiently small values of t, $m_{1,n} > n-1$.

Thus the existence of two Hamilton cycles in D produces a matrix $A_t \in S_D$ for which the maximum mean first passage time is strictly greater than the lower bound, and we have a contradiction. □

2.1. Hessenberg cycles

In this section we discuss a particular class of directed graphs, first recalling the definition of a *lower Hessenberg matrix* to be an $n \times n$ matrix A such that $a_{ij} \neq 0$ only if $i+1 \leq j$. An analogous definition holds for *upper Hessenberg matrices*, where matrix entries are nonzero only on the first subdiagonal and above.

Definition 2.7. Let D be a strongly connected directed graph on n vertices, with adjacency matrix $\mathcal{A}$. D is said to be a *Hessenberg graph* if there exists a permutation matrix P such that $P\mathcal{A}P^\top$ is a (lower) Hessenberg matrix. D is called a *Hessenberg cycle* if D is Hessenberg and $e_n^\top(P\mathcal{A}P^\top) = e_1$; that is, after some relabelling of the vertices, there is exactly one arc issuing from vertex n to vertex 1.

Without loss of generality, we will assume when dealing with a Hessenberg digraph D that the vertices have been ordered in such a way that the adjacency matrix $\mathcal{A}$ is Hessenberg, so as to remove the need for the permutation matrix P. In addition, if D is a Hessenberg cycle, we assume without loss of generality that $e_n^\top \mathcal{A} = e_1^\top$, for the same reason. If necessary, we may refer to D as a Hessenberg cycle *with respect to the index* j if it is the j^{th} vertex which corresponds to the last row when the permutation matrix is applied. We note also that the rows and columns of an upper Hessenberg matrix may be simultaneously permuted to produce a lower Hessenberg matrix; for this reason we have limited the definition above and the discussions hereafter to lower Hessenberg matrices.

An example of a Hessenberg cycle on six vertices may be seen in Figure 1, and since this example displays all possible arcs, any Hessenberg cycle on six vertices is a subdigraph of this one. Note in particular that a Hessenberg cycle on n vertices contains a unique Hamilton cycle of the form

$$1 \to 2 \to \cdots \to (n-1) \to n \to 1.$$

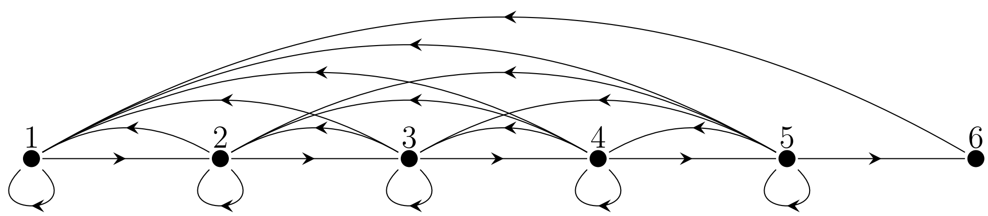

Figure 1: A Hessenberg cycle on six vertices, displaying all admissible arcs.

Remark 2.8. Notice that the directed graph of the matrix in Example 1.7 is a Hessenberg cycle with respect to the first vertex. The unique Hamilton cycle in this digraph is

$$(k{+}1) \to (k{+}2) \to \cdots \to (n{-}1) \to n \to k \to (k{-}1) \to \cdots \to 2 \to 1 \to (k{+}1).$$

This may be used to re-order the rows and columns of the transition matrix of Example 1.7 to obtain lower Hessenberg form, with last row equal to $e_1^\top$.

Proposition 2.9. *Let D be a Hessenberg cycle of order n. Then for all irreducible $A \in S_D$ with stationary vector w,*

$$w_n \leq w_k \qquad \text{for all } k = 1, \ldots, n.$$

Proof. The proof uses induction on n. Suppose D is a Hessenberg cycle on 2 vertices. Then every irreducible $A \in S_D$ has the form

$$\begin{bmatrix} t & 1-t \\ 1 & 0 \end{bmatrix},$$

for some $t \in [0, 1)$. This has stationary vector

$$w = \begin{bmatrix} \frac{1}{2-t} & \frac{1-t}{2-t} \end{bmatrix}.$$

In this case $w_2 \leq w_1$ for any choice of t; i.e. for all $A \in S_D$.

Now assume that for $m < n$, every stochastic $m \times m$ Hessenberg matrix A with $e_m^\top A = e_1$ with stationary vector w satisfies $w_m \leq w_k$ for all $k = 1, \ldots m$. Consider a Hessenberg cycle D on n vertices. For any $A \in S_D$ with stationary vector w, we have the eigen-equation

$$\begin{bmatrix} w_1 & w_2 & \cdots & w_n \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} & 0 & \cdots & 0 \\ a_{21} & a_{22} & a_{23} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{n-1,1} & a_{n-1,2} & a_{n-1,3} & \cdots & a_{n-1,n} \\ 1 & 0 & 0 & \cdots & 0 \end{bmatrix} = \begin{bmatrix} w_1 & w_2 & \cdots & w_n \end{bmatrix}$$

giving

$$a_{11}w_1 + a_{21}w_2 + \ldots + w_n = w_1.$$

Hence $w_n \leq w_1$.

Now perform stochastic complementation (see [15, Section 5.1]) on the first row and column of A, in which case $S_1 = [1]$ and S_2 is a stochastic Hessenberg matrix of order $n-1$, with its $(n-1)^{th}$ row equal to $e_1^\top$. Recall that the stationary vector for A can be written in terms of the stationary vectors of the stochastic complements; in particular, $w^\top = [w_1 \mid \gamma\tilde{w}]$, where $\tilde{w}$ is the stationary vector for S_2. By the induction hypothesis, we know that

$$\tilde{w}_{n-1} \leq \tilde{w}_k \quad \text{for all } k = 1, \ldots, n-1.$$

But this implies that

$$\gamma\tilde{w}_{n-1} \leq \gamma\tilde{w}_k \quad \text{for all } k = 1, \ldots, n-1,$$

and hence

$$w_n \leq w_j, \quad \text{for all } j = 2, \ldots, n$$

and therefore by induction, the hypothesis holds for all Hessenberg cycles D. □

Proposition 2.10. *Let D be a Hessenberg cycle of order n. Then for every irreducible $A \in S_D$ with stationary vector w and mean first passage matrix $M = [m_{i,j}]$,*

$$\max_{\substack{1 \leq i,j \leq n \\ i \neq j}} m_{i,j} = m_{1,n} = \frac{1}{w_n} - 1.$$

Proof. Let D be a Hessenberg cycle and let $A \in S_D$ with mean first passage matrix M and stationary vector w. First, observe from the equality case of (2.1) that

$$m_{n,n} = m_{n,1} + m_{1,n},$$

as every path from n to n must go through vertex 1. Furthermore, since the only arc issuing from vertex n is to vertex 1, it follows that $m_{n,1} = 1$. Given that the mean first return time $m_{n,n} = \frac{1}{w_n}$, we obtain

$$m_{1,n} = \frac{1}{w_n} - 1.$$

Next, note that for a Hessenberg cycle D (as shown for example in Figure 1) it is clear that for $i < j$, any path from vertex i to vertex j must pass through every vertex indexed by integers between i and j. Therefore by (2.1),

$$m_{i,j} = m_{i,k} + m_{k,j}, \quad \text{for any } k,\ i < k < j. \tag{2.12}$$

We now show that $m_{1,n} \geq m_{i,j}$, for all $i, j = 1, \ldots, n$, $i \neq j$. First, suppose that $i < j$. Then it follows from (2.12) that

$$m_{i,n} = m_{i,j} + m_{j,n}$$

and hence $m_{i,n} > m_{i,j}$ for $i < j < n$. It also follows from (2.12) that

$$m_{1,n} = m_{1,i} + m_{i,n}$$

and hence $m_{1,n} > m_{i,n}$. Therefore if $i < j$, we have shown that

$$m_{1,n} > m_{i,j}.$$

Now suppose that $i > j$. By (2.1),

$$m_{i,j} \leq m_{i,n} + m_{n,j}.$$

Since we have observed that $m_{n,1} = 1$, and $m_{n,j} = m_{n,1} + m_{1,j}$ by (2.12), we have

$$m_{i,j} \leq m_{i,n} + m_{1,j} + 1.$$

Now, any mean first passage time is at least 1. So $m_{j,i} \geq 1$, and hence from the above

$$m_{i,j} \leq m_{1,j} + m_{j,i} + m_{i,n}.$$

But if $i > j$, then every path in D from vertex 1 to vertex i passes through vertex j. Hence $m_{1,j} + m_{j,i} = m_{1,i}$. We also observe that every path from vertex 1 to vertex n passes through vertex i, and so $m_{1,i} + m_{i,n} = m_{1,n}$. Therefore

$$m_{i,j} \leq m_{1,n}.$$

$\square$

Proposition 2.11. *Let D be a strongly connected directed graph such that the following hold:*

(a) D has a unique Hamilton cycle;

(b) there exists an index l such that for all irreducible $A \in S_D$ with stationary vector w, $w_l = \min_k w_k$.

Then D is a Hessenberg cycle (with respect to the index l).

Proof. Let D be such a directed graph, satisfying conditions (a) and (b). Without loss of generality, we can assume

1. The unique Hamilton cycle is $1 \to 2 \to \cdots \to n \to 1$;
2. The index l for which $w_l = \min_k w_k$ for all $A \in S_D$ with stationary vector w is $l := n$.

Now suppose that D is not a Hessenberg cycle. Then there are two cases to consider: either there is an arc (j, k) in D with $j + 1 < k$, or there is more than one arc issuing from vertex n.

Case 1: Suppose there is an arc (j, k) in D with $j + 1 < k$. Then there exists a family of matrices $A_t \in S_D$ such that $\mathcal{D}(A_t)$ consists only of the directed cycle as above, and the arc (j, k). Then for any $t \in (0, 1)$, A_t is of the form:

$$A_t = \begin{bmatrix} 0 & 1 & 0 & \cdots & & \cdots & 0 & 0 \\ 0 & 0 & 1 & \cdots & & \cdots & 0 & 0 \\ \vdots & \vdots & & \ddots & & & & \\ & & & & 1-t & \cdots & t & \\ \vdots & \vdots & & & & \ddots & & \\ & & & & & & & 1 \\ 1 & 0 & 0 & \cdots & & \cdots & 0 & 0 \end{bmatrix}.$$

It is easily seen by examination that $w^\top A_t = w^\top$ gives

$$w_n = w_j, \quad \text{and} \quad (1 - t)w_j = w_{j+1}.$$

That is, $w_n = \frac{1}{1-t} w_{j+1}$. Hence $w_n > w_{j+1}$ for any choice of t, $0 < t < 1$, contradicting (b).

Case 2: Suppose that D contains an arc (n,k), some $k \neq n$. Then there exists a family of matrices $A_t \in S_D$ of the form

$$A_t = \begin{bmatrix} 0 & 1 & 0 & \cdots & \cdots & 0 \\ 0 & 0 & 1 & \cdots & \cdots & 0 \\ \vdots & \vdots & & \ddots & & \vdots \\ \vdots & \vdots & & & \ddots & \vdots \\ 0 & 0 & \cdots & & \cdots & 1 \\ 1-t & 0 & \cdots & t & \cdots & 0 \end{bmatrix}.$$

Then the eigenequation for the stationary vector w of A_t produces

$$w_1 = (1-t)w_n,$$

and hence $w_1 < w_n$, contradicting (b) as before. $\square$

This final corollary summarises the results from this section and provides a characterisation of all directed graphs D for which equality is attained in the lower bound (1.5) on $\max_{i \neq j} m_{i,j}$ for all $A \in S_D$, with the extra stipulation that the stationary vectors of matrices in S_D must always have the minimum entry occurring in the same position.

Corollary 2.12. *D is a strongly connected directed graph on n vertices such that for every irreducible $A \in S_D$ with stationary vector w and mean first passage matrix M:*

(a) $w_n = \min_k w_k$*;*

(b) $\displaystyle\max_{\substack{1 \leq i,j \leq n \\ i \neq j}} m_{i,j} = \frac{1}{w_n} - 1,$

if and only if D is a Hessenberg cycle.

Remark 2.13. Let D be the Hessenberg cycle on n vertices with all possible arcs, and suppose we are given some probability vector $w = [w_1\, w_2\, \cdots\, w_n]^\top$ with $w_1 \geq w_2 \geq \cdots \geq w_n$. Then it is not difficult to show that there exists a parametrised family of matrices in S_D with w as their stationary distribution vector. This is determined simply by examining the eigenequation $w^\top A = w^\top$

for an arbitrary matrix A with the Hessenberg cycle pattern. In particular, the entries on the superdiagonal may be expressed as

$$\begin{aligned} a_{j,j+1} &= \frac{w_{j+1}}{w_j}\sum_{k=1}^{j} a_{j+1,k} + \frac{w_{j+2}}{w_j}\sum_{k=1}^{j} a_{j+2,k} + \cdots + \frac{w_n}{w_j} \\ &= \sum_{i=1}^{n-j}\sum_{k=1}^{j} \frac{w_{j+1}}{w_j} a_{j+1,k}. \end{aligned}$$

Considering the diagonal entries of A to be constrained so that $a_{j,j} = 1 - \sum_{k=1}^{j-1} a_{j,k} - a_{j,j+1}$, we can consider every entry below the main diagonal to be free parameters, constrained only by the fact that the matrix must be nonnegative and stochastic. Certainly this family is nonempty, as we may set all of these parameters to zero and still obtain a matrix $A \in S_D$ with $w^\top A = w^\top$.

2.1.1. $M/G/1$ queues

Stochastic upper Hessenberg matrices arise as transition matrices for a Markov chain model of a certain type of queueing system, where customers arrive to the system, spend a certain amount of time waiting for service, and leave after they have been served. The model is as follows: let $\mathcal{S} = \{s_0, s_1, s_2, \ldots\}$ be our state space, where s_i denotes the state that there are i customers in the system, including one who is being served. Transitions between states are governed by arrivals and departures to and from the system; i.e. new customers joining the queue or customers leaving the queue once they have been served. Suppose that the length of a time-step is chosen so that at most one customer is served in a single time-step. Then $s_i \to s_j$ is a permitted transition if and only if $j > i$ (new arrivals), $j = i$ (one arrival and one departure), or $j = i - 1$ (no arrivals, one departure). Hence the transition matrix A is upper Hessenberg.

This Markov chain model of a queue is a description of the *embedded Markov chain* for an $M/G/1$ queue (see [1, Chapter 5]). This is a stochastic process of arrivals and departures where arrivals are assumed to be Markovian (governed by a Poisson process), service times have a general distribution, and there is one server. These assumptions allow the above Markov chain construction to describe the behaviour of the queueing system. Such models can be applied to familiar, mundane queues such as patients at a doctor's office, or vehicles awaiting service at a mechanic. They can also be used to

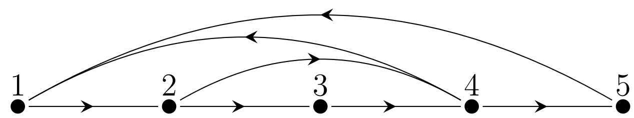

Figure 2: A non-Hessenberg digraph D on 5 vertices for which equality is attained in our lower bound for all $A \in S_D$.

examine communication systems, where 'customers' are voice or data traffic awaiting transmission. These applications have existed since the very first publication on queueing theory (see [7]) and are adapted in recent times to be of use with modern technology, such as video transmission (see [16]) and web server performance (see [4]).

The result we have proven in this section applies to $M/G/1/K$ queues – that is, an $M/G/1$ queue with finite capacity, resulting in a finite state space and transition matrix. This means that if the system has reached capacity, any arriving customers must leave without joining the queue. The Hessenberg cycle which we have examined in this section represents a queueing system with the particular feature that once the system is empty, with zero customers waiting in the queue, it then fills to full capacity, in a 'bulk arrival'. Our result says that for queues of this type, the expected time to reach the state that there are no customers in the queue from the state that the queue is full is optimal, relative to the given stationary vector.

2.2. Non-Hessenberg digraphs which achieve equality

Our initial objective was to characterise all directed graphs D for which equality held in the lower bound (1.5) for all irreducible $A \in S_D$. The results of the previous section determine the family of so-called Hessenberg cycles as the characterisation of all such digraphs for which the additional condition holds: that the minimum entry of the stationary vector of a transition matrix in S_D always occurs in the same position.

Consider now the example D in Figure 2, which is not Hessenberg due to the arc $(2, 4)$ (and there is no permutation of the vertices resulting in a

Hessenberg cycle). Any matrix A in S_D is of the form

$$A = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & a & 1-a & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 1-b & 0 & 0 & 0 & b \\ 1 & 0 & 0 & 0 & 0 \end{bmatrix}$$

for some $0 < a, b < 1$. The stationary vector of A is computed as

$$w^\top = \tfrac{1}{3+a+b} \begin{bmatrix} 1 & 1 & a & 1 & b \end{bmatrix}$$

and the mean first passage matrix is

$$M = \begin{bmatrix} 3+a+b & 1 & \frac{3+b}{a} - 1 - b & 2+a & \frac{3+a}{b} \\ 2+a+b & 3+a+b & \frac{3+b}{a} - 2 - b & 1+a & \frac{3+a}{b} - 1 \\ 2+b & 3+b & \frac{3+b}{a} + 1 & 1 & \frac{3+a}{b} - 1 - a \\ 1+b & 2+b & \frac{3+b}{a} & 3+a+b & \frac{3+a}{b} - 1 - 2a \\ 1 & 2 & \frac{3+b}{a} - b & 3+a & \frac{3+a}{b} + 1 \end{bmatrix}.$$

The minimum entry of the stationary vector is either w_3 or w_5, depending on whether $a > b$ or $a < b$. Further, the maximum entry in the mean first passage matrix is either

$$m_{4,3} = \frac{3+b}{a} \quad \text{or} \quad m_{1,5} = \frac{3+a}{b},$$

which are $\frac{1}{w_3} - 1$ and $\frac{1}{w_5} - 1$ respectively. Moreover, when $a < b$, w_3 is minimum and $\max_{i \neq j} m_{i,j} = m_{4,3}$, and when $a > b$, w_5 is minimum and $\max_{i \neq j} m_{i,j} = m_{1,5}$, and hence equality holds in the lower bound (1.5) for all $A \in S_D$.

This example of order five can be generalised to a directed graph of order n with the same properties as D, shown in Figure 3. The stationary vector of a matrix in this family is

$$w^\top = \tfrac{1}{3+a+b} \begin{bmatrix} 1 & 1 & a & 1 & \cdots & 1 & b \end{bmatrix}$$

and the mean first passage matrix is

$$\begin{bmatrix} (n-2)+a+b & 1 & \frac{(n-2)+b}{a} - (n-4) - b & 2+a & \cdots & (n-3)+a & \frac{(n-2)+a}{b} \\ (n-3)+a+b & (n-2)+a+b & \frac{(n-2)+b}{a} - (n-3) - b & 1+a & \cdots & (n-4)+a & \frac{(n-2)+a}{b} - 1 \\ (n-3)+b & (n-3)+a+b & \frac{(n-2)+b}{a} + 1 & 1 & \cdots & (n-4) & \frac{(n-2)+a}{b} - 1 - a \\ (n-4)+b & (n-3)+b & \frac{(n-2)+b}{a} & (n-2)+a+b & \cdots & (n-5) & \frac{(n-2)+a}{b} - 2 - a \\ \vdots & \vdots & \vdots & \vdots & & \vdots & \vdots \\ 2+b & 3+b & \frac{(n-2)+b}{a} - (n-6) & 4+a+b & \cdots & 1 & \frac{(n-2)+a}{b} - (n-4) - a \\ 1+b & 2+b & \frac{(n-2)+b}{a} - (n-5) & 3+a+b & \cdots & (n-2)+a+b & \frac{(n-2)+a}{b} - (n-3) - a \\ 1 & 2 & \frac{(n-2)+b}{a} - (n-5) - b & 3+a & \cdots & (n-2)+a & \frac{(n-2)+a}{b} + 1 \end{bmatrix}.$$

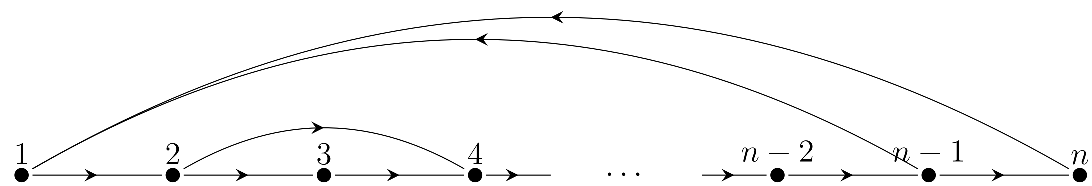

Figure 3: A non-Hessenberg digraph D on n vertices for which equality is attained in our lower bound for all $A \in S_D$.

Again, the minimum entry of the stationary vector depends on whether $a > b$ or $a < b$, and in either case, the maximum off-diagonal entry of the mean first passage matrix is $\frac{1}{\min_k w_k} - 1$.

2.2.1. Some observations

From the existence of the above class of examples, we can observe that the digraph characterisation problem becomes significantly more difficult when we relax the constraint that the minimum entry of the stationary vectors of these matrices occurs in a common position over the whole class. In particular, equality is attained in this class of examples due to the following features of the directed graphs in this class: letting $\mathcal{J}$ denote the index set of indices j for which it is possible w_j is minimum for some $A \in S_D$, we have

- for all $j \in \mathcal{J}$, the maximum entry in the j^{th} column of the mean first passage matrix (bar $m_{j,j}$) is $m_{j+1,j}$;
- for all $j \in \mathcal{J}$, $m_{j+1,j} = \frac{1}{w_j} - 1$;
- the set of candidates for maximum off-diagonal mean first passage time is $\{m_{j+1,j} \mid j \in \mathcal{J}\}$, and $m_{j+1,j}$ is the overall maximum precisely when $w_j = \min_k w_k$.

For these reasons, it would seem that to answer the more general question, results are sought on directed graphs for which equality holds in the column lower bound of Proposition 1.2 for all transition matrices respecting the digraph. Further, showing that these mean first passage times are maximum in the whole mean first passage matrix becomes difficult without the particular structure of a Hessenberg digraph and hence the tools used in Proposition 2.10, particularly the 'triangle inequality' for mean first passage times in (2.1).

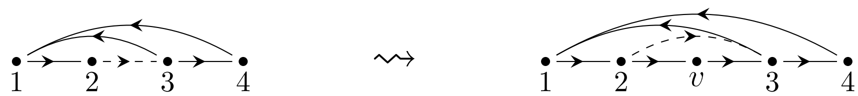

Figure 4: An illustration of a process for constructing the previous order 5 example from an order 4 Hessenberg cycle

A reasonable starting point would be to determine a method for constructing new families of directed graphs from the Hessenberg cycles, thereby using some of the tools we have already produced. Notice in Figure 4 that our order 5 example above is constructed from an order 4 Hessenberg cycle by introducing a new vertex v between vertices 2 and 3 on the Hamilton cycle, but retaining the original arc from 2 to 3. This produces a non-Hessenberg directed graph which can then be examined for the property that mean first passage times are limited by the stationary vector, using the information we have about the mean first passage times of transition matrices respecting the original Hessenberg cycle. Note that a similar construction exists to create the order n non-Hessenberg digraph in Figure 3 from a Hessenberg cycle of order $n-1$. A developed construction method which is proven to produce families of new non-Hessenberg directed graphs with equality attainment in (1.5) would shed some light on the general characterisation question.

Acknowledgement: The authors would like to express their thanks to an anonymous reviewer, whose constructive comments helped to improve this paper.

References

[1] U Narayan Bhat. *An Introduction to Queueing Theory: Modeling and Analysis in Applications*. Birkhäuser, 2015.

[2] Jane Breen and Steve Kirkland. Stationary vectors of stochastic matrices subject to combinatorial constraints. *Electronic Journal of Linear Algebra*, 28(1):4, 2015.

[3] Richard Brualdi and Herbert Ryser. *Combinatorial Matrix Theory*, volume 39. Cambridge University Press, 1991.

[4] Jianhua Cao, Mikael Andersson, Christian Nyberg, and Maria Kihl. Web server performance modeling using an M/G/1/K* PS queue. In

10th International Conference on Telecommunications, 2003, volume 2, pages 1501–1506. IEEE, 2003.

[5] Minerva Catral, Steve Kirkland, Michael Neumann, and Nung-Sing Sze. The Kemeny constant for finite homogeneous ergodic Markov chains. *Journal of Scientific Computing*, 45(1-3):151–166, 2010.

[6] Emanuele Crisostomi, Steve Kirkland, and Robert Shorten. A Google-like model of road network dynamics and its application to regulation and control. *International Journal of Control*, 84(3):633–651, 2011.

[7] Agner Krarup Erlang. The theory of probabilities and telephone conversations. *Nyt Tidsskrift for Matematik B*, 20(33-39):16, 1909.

[8] Roger Horn and Charles Johnson. *Matrix Analysis*. Cambridge University Press, 2nd edition, 2013.

[9] Jeffrey Hunter. Stationary distributions and mean first passage times of perturbed Markov chains. *Linear Algebra and its Applications*, 410:217–243, 2005.

[10] Jeffrey Hunter. Mixing times with applications to perturbed Markov chains. *Linear Algebra and its Applications*, 417(1):108–123, 2006.

[11] John Kemeny and James Laurie Snell. *Finite Markov Chains*, volume 356. van Nostrand Princeton, NJ, 1960.

[12] Steve Kirkland. Fastest expected time to mixing for a Markov chain on a directed graph. *Linear Algebra and its Applications*, 433(11):1988–1996, 2010.

[13] Steve Kirkland. Load balancing for Markov chains with a specified directed graph. *Linear and Multilinear Algebra*, 62(11):1491–1508, 2014.

[14] Steve Kirkland. On the Kemeny constant and stationary distribution vector for a Markov chain. *Electronic Journal of Linear Algebra*, 27(1):23, 2014.

[15] Steve Kirkland and Michael Neumann. *Group Inverses of M-Matrices and their Applications*. CRC Press, 2013.

[16] Nikolaos Laoutaris and Ioannis Stavrakakis. Adaptive playout strategies for packet video receivers with finite buffer capacity. In *IEEE International Conference on Communications, 2001*, volume 3, pages 969–973. IEEE, 2001.

[17] Mark Levene and George Loizou. Kemeny's constant and the random surfer. *The American Mathematical Monthly*, 109(8):741–745, 2002.

[18] Eugene Seneta. *Non-negative Matrices and Markov Chains.* Springer-Verlag, New York, 1981.