

# Clustering behaviour in Markov chains with eigenvalues close to one

Jane Breen<sup>a,\*</sup>, Emanuele Crisostomi<sup>b</sup>, Mahsa Faizrahnemoon<sup>c</sup>, Steve Kirkland<sup>a</sup>, Robert Shorten<sup>d</sup>

<sup>a</sup>*Department of Mathematics, University of Manitoba, Winnipeg MB, R3T 2N2, Canada.*

<sup>b</sup>*Department of Energy, Systems, Territory and Constructions Engineering, University of Pisa, 56121, Pisa, Italy.*

<sup>c</sup>*Department of Mathematics, Simon Fraser University, Burnaby, BC V5A 1S6, Canada.*

<sup>d</sup>*School of Electrical & Electronic Engineering, University College Dublin, Belfield, Dublin 4, Ireland.*

---

## Abstract

Finite, discrete, time-homogeneous Markov chains are frequently used as a simple mathematical model of real-world dynamical systems. In many such applications, an analysis of clustering behaviour in the states of the system is desirable, and it is well-known that the eigendecomposition of the transition matrix  $A$  of the Markov chain can provide such insight. Clustering methods based on the sign pattern in the second eigenvector of  $A$  are frequently used when  $A$  has dominant eigenvalues that are real. In this paper, we present a method to include an analysis for complex eigenvalues of  $A$  which are close to 1. Since a real spectrum is not guaranteed in most applications, this is a valuable result in the area of spectral clustering in Markov chains.

*Keywords:* Markov chain, clustering, complex eigenvalue

*2010 MSC:* 15B51, 60J10, 15A18

---

## 1. Introduction

A finite, discrete, time-homogeneous Markov chain refers to a mathematical model of a system which occupies, at any given time, one of a finite number of states  $\{s_1, \dots, s_n\}$  and transitions between states in discrete time-steps, according to prescribed transition probabilities. In particular, for any pair of states  $s_i$  and  $s_j$ , there is a given probability  $a_{ij}$  that the system moves to state  $s_j$  in one time-step, given that it is currently in state  $s_i$ . A Markov chain is memoryless, meaning that the movement of the system in the next time-step depends only on the current state the system occupies. A Markov chain can be

---

\*Corresponding author

*Email addresses:* `breenj3@myumanitoba.ca` (Jane Breen), `emanuele.crisostomi@gmail.com` (Emanuele Crisostomi), `mfaizrah@sfu.ca` (Mahsa Faizrahnemoon), `Stephen.Kirkland@umanitoba.ca` (Steve Kirkland), `robert.shorten@ucd.ie` (Robert Shorten)

represented by a row-stochastic matrix  $A = [a_{ij}]$  – that is, a matrix in which each row sums to 1 – referred to as the *probability transition matrix* of the chain.

The applications of Markov chains are diverse, including their use as a model of traffic in road networks (see [2]), and of web traffic on the world wide web (see [11]), and in molecular conformational dynamics in drug design (see [5]). Complex dynamical systems such as this last example can be difficult to analyse, but a Markov chain approach to analysing the same system is both simple and powerful; see [8] for an introduction to modelling complex dynamical systems using a Markov chain. The advantages of modelling with Markov chains are broad; given only a transition matrix, many features of the real-world system can be determined with simple techniques in linear algebra. The long-term behaviour of the system is encapsulated in the stationary distribution vector, which is simply a left eigenvector of  $A$  corresponding to the eigenvalue 1. The short-term behaviour can be examined using the mean first passage time from the  $i^{\text{th}}$  state to the  $j^{\text{th}}$  state,  $m_{i,j}$  – i.e. the expected time it will take to reach state  $s_j$  for the first time if the system begins in state  $s_i$ . An interesting parameter of a Markov chain which has been the recent subject of extensive study is *Kemeny’s constant*, denoted  $\mathcal{K}(A)$ , which is interpreted in terms of the “expected time to mixing” in the system (see [10]), and gives an overall measure of how ‘well-connected’ the system is, due to its interpretation in terms of the expected length of a trip from a ‘randomly’ selected initial state to a ‘randomly’ selected destination state.

Of great interest in many systems that can be modelled using Markov chains is the manifestation of clustering behaviour and how it may be predicted. Clustering behaviour is usually characterised by the existence of collections of states of the Markov chain for which the system, if starting in a state in a cluster, is unlikely to leave that collection of states in the short term. That is, the expected number of time-steps until the chain is in a state outside of that cluster is relatively large. In the extreme, this results in a *nearly uncoupled* system; that is, the stochastic matrix in question can be considered as a perturbed block diagonal matrix, where the diagonal blocks represent the clusters or *almost invariant aggregates* of the chain. It is not difficult to show that for such a matrix, the second-largest (or *subdominant*) eigenvalue(s) must be close to 1 due to the continuity of eigenvalues. However, much of the work regarding the determination of this clustered behaviour or *near uncoupling* is concerned with the converse question: given that a stochastic irreducible matrix  $A$  has eigenvalues  $\lambda$  close to 1, what can be said about the clustering behaviour of the Markov chain represented by  $A$ ? This is the question that we consider in this article.

In a stochastic model of molecular dynamics, clusters are referred to as *metastable states*, representing different conformations of the molecule. Identification of these metastable conformations is extremely important in drug design (see [4]) and biomolecular research (see [13]). An algorithm for this purpose, based on *Perron cluster analysis* (that is, analysing the cluster of eigenvalues around the Perron eigenvalue or spectral radius 1), is developed in [3], and improved upon in [5]. However, the Markov chains considered in these models are always reversible, and hence the eigenvalues of the stochastic matrices are always

real. An attempt to generalise without relying on the reversibility assumption is given in [12], which uses the *singular value decomposition* instead of the eigendecomposition of the matrix. Further assumptions were found to be necessary for this algorithm in [15].

In [2], a detailed description of a Markov chain model for an urban traffic network is presented. Clustering in such a network corresponds to ‘communities’ in the network, and this is measured using Kemeny’s constant and the second largest (real) eigenvalue of the transition matrix, which is found to be the basis of an expression of a bound on the probabilities of moving from one subset of states to another, and also a bound on the mean first passage times between those subsets. Since we cannot depend on the assumption of real eigenvalues in this model of road traffic – or in any general Markov chain model – we wish to generalise the result in this paper to the case that the stochastic matrix representing the chain has complex eigenvalues.

We now give an expository discussion of the main theoretical result in [2, Appendix 6.1], regarding evidence of clustering in a Markov chain derived from the existence of a real eigenvalue near 1. The method of proof of this result inspires this paper.

**Theorem 1.1.** *Let  $A$  be an irreducible stochastic matrix and suppose that  $\lambda \in \mathbb{R}$  is an eigenvalue of  $A$ . Let  $v = [v_1^\top \mid -v_2^\top \mid \mathbf{0}^\top]^\top$  be a corresponding  $\lambda$  eigenvector (with  $v_1 > 0$  and  $v_2 > 0$ ) and let us partition the matrix  $A$  conformally as*

$$\left[ \begin{array}{c|c|c} A_{11} & A_{12} & A_{13} \\ \hline A_{21} & A_{22} & A_{23} \\ \hline A_{31} & A_{32} & A_{33} \end{array} \right]$$

and label the subsets of the partition as  $S_1, S_2$ , and  $S_3$  respectively. Then:

- (a)  $\rho(A_{11}), \rho(A_{22}) > \lambda$ .
- (b) There are subsets  $\tilde{S}_1 \subseteq S_1$ ,  $\tilde{S}_2 \subseteq S_2$ , and positive vectors  $\tilde{w}_1^\top, \tilde{w}_2^\top$  with supports on  $\tilde{S}_1, \tilde{S}_2$  respectively such that  $\tilde{w}_1^\top \mathbb{1} = \tilde{w}_2^\top \mathbb{1} = 1$  and

$$\sum_{i \in \tilde{S}_1} \tilde{w}_1(i) \sum_{j \notin \tilde{S}_1} a_{ij} = 1 - \rho(A_{11}) \leq 1 - \lambda, \quad (1.1)$$

and

$$\sum_{i \in \tilde{S}_2} \tilde{w}_2(i) \sum_{j \notin \tilde{S}_2} a_{ij} = 1 - \rho(A_{22}) \leq 1 - \lambda. \quad (1.2)$$

- (c) For any  $j \in \tilde{S}_2$ ,

$$\sum_{i \in \tilde{S}_1} \tilde{w}_1(i) m_{ij} \geq \frac{1}{1 - \rho(A_{11})} \geq \frac{1}{1 - \lambda} \quad (1.3)$$

and for any  $j \in \tilde{S}_1$ ,

$$\sum_{i \in \tilde{S}_2} \tilde{w}_2(i) m_{ij} \geq \frac{1}{1 - \rho(A_{22})} \geq \frac{1}{1 - \lambda}, \quad (1.4)$$

where  $m_{ij}$  are entries of the mean first passage matrix.

We first remark that the bounds in this theorem hold for any real eigenvalue  $\lambda$  of  $A$ . However, the result is intended to provide evidence of clustering in a system when  $\lambda$  is close to 1. The justification for this primarily relies on parts (b) and (c) of the above; (b) essentially implies that transitions out of a certain subset of the index set of  $A_{11}$  are rare (that is, the corresponding transition probabilities are small) if  $\lambda$  is close to 1, while (c) indicates that the expected times for the system to move from one subset of states to the other are large.

The drawback of this result is that lower bounds for  $\rho(A_{11})$  and  $\rho(A_{22})$  are determined only by a real eigenvalue and its corresponding eigenvector. Our question is this – can any clustering behaviour be determined from a complex eigenvalue and corresponding eigenvector? That is, given  $\lambda \in \mathbb{C}$  an eigenvalue of  $A$  where  $\lambda = \alpha + i\beta$ , some  $\alpha, \beta$ , can we:

- (i) define a conformal partition of a corresponding eigenvector for  $\lambda$  and the matrix  $A$ ;
- (ii) determine lower bounds for the spectral radii of  $A_{11}$  and  $A_{22}$  (the principal submatrices determined by the index set of this partition); and
- (iii) conclude equivalent statements about the clustering properties of  $A$  in Theorem 1.1, parts (b) and (c).

**Remark 1.2.** A brief examination of the proof in [2] of Theorem 1.1 will determine that indeed, (b) and (c) are proven independent of the fact that  $\lambda$  is real; moreover, given lower bounds for  $\rho(A_{11}), \rho(A_{22})$ , these may be substituted for  $\lambda$  in (1.1), (1.2), (1.3) and (1.4). Therefore our goal in this work is to define a partition and determine appropriate lower bounds on the spectral radii of corresponding principal submatrices which will be close to 1 under certain circumstances. By Theorem 1.1 (b) and (c), this will allow us to conclude that there is evidence of clustering in the Markov chain.

Another interesting line of inquiry for future work would be to quantify a measure of clustering which incorporates both mean first passage times and transition probabilities in the manner described above, and to study the behaviour of eigenvalues of transition matrices which display clustering according to this measure. Note that such a measure exists which quantifies clustering solely in terms of transition probabilities (see [9]), but that in practice, considering mean first passage times in addition to transition probabilities gives a more robust interpretation of clustering behaviour. Other characterisations are also feasible, and the suggestion we make is just one possibility.

The remainder of this article is structured as follows: in Section 2, we establish a method for determining a conformal partition using the real part of the complex eigenvector corresponding to the eigenvalue  $\lambda = \alpha + i\beta$  and the corresponding lower bounds. This is our main result, presented in a discursive format and formalised in the statement of Theorem 2.1 at the end of the section; in particular, we do not supply a formal proof. In Section 3, we describe a parallel method for determining a conformal partition using the imaginary part of the eigenvector corresponding to  $\lambda$ , which can produce entirely new evidence of clustering behaviour in the Markov chain corresponding to the transition matrix

in question. Since the mathematics is identical to that in Section 2, this is presented in less detail, and again formalised in Theorem 3.1 without a formal proof. In Section 4 we give some remarks about the implementation of this method as an algorithm, and present some examples to illustrate the method. The results of Sections 2 and 3 can also be found in [6].

Finally, we remark that this work is inspired by the empirical observations in [7] which anticipated that clustering behaviour could be detected from complex eigenvectors. A rationale is given in Section 2.2 of [7] as to why this may be plausible, although we will emphasise later in Remark 3.2 that the mere existence of complex eigenvalues of large modulus or close to 1 is not sufficient to conclude the presence of clustering behaviour. Furthermore, in a simulated model of a bus network, the authors of [7] demonstrate that clusters in the Markov chain can be determined by visual inspection of the second eigenvector. We will analyse this example further in Section 4 using the theory developed in Sections 2 and 3.

## 2. A conformal partition with respect to the real part of an eigenvector

Suppose that  $A$  is an irreducible stochastic matrix of order  $n$  with an eigenvalue  $\lambda = \alpha + i\beta$ . It will be important later to consider  $\alpha$  to be close to 1, and  $\beta$  close to 0, but for now we assume only that  $\alpha, \beta > 0$ . Let  $x + iy$  be a corresponding eigenvector for  $A$ , where  $x, y \in \mathbb{R}^n$ . It follows from equating real and complex coefficients in the standard eigenequation that

$$Ax = \alpha x - \beta y \tag{2.1}$$

and

$$Ay = \beta x + \alpha y. \tag{2.2}$$

Now consider the following: we partition the system (i.e. the matrix  $A$  and the vectors  $x$  and  $y$ ) according to where  $x$  is positive, negative, and zero. That is, we have

$$\left[ \begin{array}{c|c|c} A_{11} & A_{12} & A_{13} \\ \hline A_{21} & A_{22} & A_{23} \\ \hline A_{31} & A_{32} & A_{33} \end{array} \right] \left[ \begin{array}{c} x_1 \\ x_2 \\ 0 \end{array} \right] = \left[ \begin{array}{c} \alpha x_1 - \beta y_1 \\ \alpha x_2 - \beta y_2 \\ -\beta y_3 \end{array} \right]$$

where  $x_1 > 0$  and  $x_2 < 0$ , entrywise. Let  $S_1, S_2, S_3$  denote the index sets of the partition. Note that  $S_3$  may be empty.

This gives:

$$A_{11}x_1 + A_{12}x_2 = \alpha x_1 - \beta y_1,$$

and since  $A_{12}x_2$  is entrywise nonpositive,

$$A_{11}x_1 \geq \alpha x_1 - \beta y_1.$$

Now consider the diagonal matrix  $X_1 := \text{diag}(x_1)$ , and observe that  $A_{11}x_1 = A_{11}X_1\mathbb{1}$ . Furthermore, from the above,

$$X_1^{-1}A_{11}X_1\mathbb{1} \geq X_1^{-1}(\alpha x_1 - \beta y_1).$$

By a well-known result (see [14, Cor.1 to Thm.1]) we know that the spectral radius of a nonnegative matrix lies between its minimum and maximum row sums; hence

$$\rho(X_1^{-1}A_{11}X_1) \geq \min(X_1^{-1}A_{11}X_1\mathbb{1}) \geq \min_j \left( \frac{\alpha x_1(j) - \beta y_1(j)}{x_1(j)} \right)$$

(where  $x_i(j)$  denotes the  $j^{\text{th}}$  entry of  $x_i$ , for example). Since  $A_{11}$  and  $X_1^{-1}A_{11}X_1$  are similar matrices, their spectral radii are equal, and it follows that

$$\rho(A_{11}) \geq \alpha - \beta \max_j \left( \frac{y_1(j)}{x_1(j)} \right). \quad (2.3)$$

Similarly, we may show that

$$\rho(A_{22}) \geq \alpha - \beta \max_j \left( \frac{y_2(j)}{x_2(j)} \right). \quad (2.4)$$

We pause here to note two things. The first is that if  $\alpha$  is close to 1, and  $\beta$  close to 0, then these lower bounds are each close to 1, indicating that the vertices indexed by  $S_1$  and  $S_2$  display some clustering behaviour in the Markov chain represented by  $A$ , in the manner described in Theorem 1.1, parts (b) and (c).

The second thing to note is that it is vital to our further discussion on these clusters that we assume that

$$\alpha x_1 - \beta y_1 > 0 \quad \text{and} \quad \alpha x_2 - \beta y_2 < 0, \quad (2.5)$$

in order that the lower bounds in (2.3) and (2.4) are positive. We will consider (2.5) as an additional hypothesis that must be satisfied in order to state the result, because although (2.3) and (2.4) still hold in the case that (2.5) does not, they are worthless to us, since the spectral radius will always be nonnegative. Therefore this result (and subsequent results) are only relevant in the case that the hypotheses in (2.5) hold.

### 2.1. Expanding the index sets of the partition

The bounds derived in this section correspond to Theorem 2.1, 1(ii) and 2(ii).

It is possible to optimise the bounds in (2.3) and (2.4) by adding a little flexibility in how the partition is determined. We allow the option of expanding the first cluster indexed by  $S_1$  to include indices corresponding to positive entries of  $y_3$ . Similarly, we expand the second index set  $S_2$  by including entries corresponding to negative entries of  $y_3$ .

Formally, we define the new partition index sets as

$$\tilde{S}_1 = S_1 \cup \{j \mid y_3(j) > 0\}, \quad \text{and} \quad \tilde{S}_2 = S_2 \cup \{j \mid y_3(j) < 0\}.$$

Equivalently, we may consider the vector  $x + ty$  for some  $t > 0$  and partition, as before, according to where  $x + ty$  is positive, negative, and zero, producing the respective partition sets  $\tilde{S}_1, \tilde{S}_2$ , and  $\tilde{S}_3 = \{j \mid x(j) = y(j) = 0\}$ . In particular, we must choose  $t > 0$  sufficiently small to achieve the above; we require

$$t < \frac{-x(j)}{y(j)}, \quad \text{for all } j \text{ such that } x(j)y(j) < 0.$$

This ensures that  $x(j) + ty(j) > 0$  for all  $j \in \tilde{S}_1$ , and that  $x(j) + ty(j) < 0$  for all  $j \in \tilde{S}_2$ .

Considering  $A(x+ty)$ , and partitioning with respect to where  $x+ty$  is positive, negative, and zero, using (2.1) and (2.2) we obtain:

$$\left[ \begin{array}{c|c|c} \tilde{A}_{11} & \tilde{A}_{12} & \tilde{A}_{13} \\ \hline \tilde{A}_{21} & \tilde{A}_{22} & \tilde{A}_{23} \\ \hline \tilde{A}_{31} & \tilde{A}_{32} & \tilde{A}_{33} \end{array} \right] \left[ \begin{array}{c} \tilde{x}_1 + \tilde{t}\tilde{y}_1 \\ \tilde{x}_2 + \tilde{t}\tilde{y}_2 \\ 0 \end{array} \right] = \left[ \begin{array}{c} \alpha(\tilde{x}_1 + \tilde{t}\tilde{y}_1) + \beta(\tilde{t}\tilde{x}_1 - \tilde{y}_1) \\ \alpha(\tilde{x}_2 + \tilde{t}\tilde{y}_2) + \beta(\tilde{t}\tilde{x}_2 - \tilde{y}_2) \\ (\alpha t - \beta)\tilde{y}_3 \end{array} \right], \quad (2.6)$$

where  $\tilde{A}_{ij}$  denote the submatrices corresponding to the new partition with index sets  $\tilde{S}_1, \tilde{S}_2, \tilde{S}_3$ .

Note that this ‘repartition’ is not substantially different; we simply allow the option of including some extra states in each cluster by including indices corresponding to positive entries of  $y_3$  to  $S_1$ , and indices corresponding to negative entries of  $y_3$  to  $S_2$ . Moreover,  $S_1 \subseteq \tilde{S}_1$  and  $S_2 \subseteq \tilde{S}_2$ . Interpreting in terms of prospective clustering behaviour, we are simply allowing the possible addition of more states into our existing index set to determine more information about the clustering behaviour of our chain.

Proceeding as before, we have, from (2.6):

$$\tilde{A}_{11}(\tilde{x}_1 + \tilde{t}\tilde{y}_1) \geq \alpha(\tilde{x}_1 + \tilde{t}\tilde{y}_1) + \beta(\tilde{t}\tilde{x}_1 - \tilde{y}_1) \quad (2.7)$$

$$\Rightarrow \rho(\tilde{A}_{11}) \geq \alpha + \beta \min_j \left( \frac{\tilde{t}\tilde{x}_1(j) - \tilde{y}_1(j)}{\tilde{x}_1(j) + \tilde{t}\tilde{y}_1(j)} \right), \quad (2.8)$$

and similarly

$$\rho(\tilde{A}_{22}) \geq \alpha + \beta \min_j \left( \frac{\tilde{t}\tilde{x}_2(j) - \tilde{y}_2(j)}{\tilde{x}_2(j) + \tilde{t}\tilde{y}_2(j)} \right). \quad (2.9)$$

These lower bounds are increasing functions of  $t$  and so they are optimised by taking the limit as  $t$  approaches

$$\min \left\{ \frac{-x(j)}{y(j)} \mid x(j)y(j) < 0 \right\}. \quad (2.10)$$

If the set above is empty, then  $t$  is unbounded. This occurs if and only if wherever  $x$  is positive,  $y$  is positive, and where  $x$  is negative,  $y$  is negative. It follows that our lower bounds would then be

$$\rho(\tilde{A}_{11}) \geq \alpha + \beta \min_j \left( \frac{\tilde{x}_1(j)}{\tilde{y}_1(j)} \right),$$

and

$$\rho(\tilde{A}_{22}) \geq \alpha + \beta \min_j \left( \frac{\tilde{x}_2(j)}{\tilde{y}_2(j)} \right),$$

by taking the limit as  $t \rightarrow \infty$ .

As in the first case, we need to ensure that the right-hand side of (2.8) is positive in order for our conclusions to be worthwhile. It is not difficult to show that if  $S_1 = \tilde{S}_1$ , the first hypothesis in (2.5) is sufficient to ensure positivity, but if  $S_1$  is a proper subset of  $\tilde{S}_1$ , the additional case that  $\tilde{x}_1(j) = 0$  and  $\tilde{y}_1(j) > 0$  is considered. In this case we obtain the additional restriction that

$$t > \frac{\beta}{\alpha}.$$

Since  $t$  is also bounded above by (2.10), it is necessary that  $\frac{\beta}{\alpha}$  is also strictly less than this. In other words, it is necessary that

$$\alpha x_1 + \beta y_1 > 0 \quad \text{and} \quad \alpha x_2 + \beta y_2 < 0. \quad (2.11)$$

It remains to consider the possibility that, in our lower bounds (2.8) and (2.9), we divide by zero when setting  $t$  equal to the minimum entry of the set in (2.10). Of course, by choosing  $t$  in this way, we do set at least one entry of either the vector  $\tilde{x}_1 + t\tilde{y}_1$  or the vector  $\tilde{x}_2 + t\tilde{y}_2$  equal to zero, but since we choose the minimum entry of

$$\frac{t\tilde{x}_i - \tilde{y}_i}{\tilde{x}_i + t\tilde{y}_i}$$

for  $i = 1, 2$ , this only presents issues if *every* entry of either  $\tilde{x}_1 + t\tilde{y}_1$  or  $\tilde{x}_2 + t\tilde{y}_2$  is equal to zero. This means that linear dependence occurs between the vectors  $\tilde{x}_i$  and  $\tilde{y}_i$  for  $i = 1$  or  $i = 2$ . This presents yet another condition: that the vectors  $x_i$  and  $y_i$  must be linearly independent (since linear independence of  $x_i$  and  $y_i$  implies linear independence of  $\tilde{x}_i$  and  $\tilde{y}_i$ ).

Finally, we note that there is a possibility that no extra states are included in one or both of the clusters; that is,  $\tilde{S}_1 = S_1$ , or  $\tilde{S}_2 = S_2$ . In this case, it is easily shown that the new lower bounds (2.8) and (2.9) are an improvement on the lower bounds (2.3) and (2.4), respectively.

## 2.2. An alternate expansion of the index sets of the partition

The bounds derived in this section correspond to Theorem 2.1, 1(iii) and 2(iii).

We now consider an alternate partition derived from the index sets  $S_1$  and  $S_2$ , where we allow the inclusion of indices corresponding to negative entries of  $y_3$  to  $S_1$ , and positive entries of  $y_3$  to  $S_2$ . That is, we define new index sets

$$\bar{S}_1 = S_1 \cup \{j \mid y_3(j) < 0\} \quad \text{and} \quad \bar{S}_2 = S_2 \cup \{j \mid y_3(j) > 0\}.$$

This is equivalent to considering the vector  $x + ty$ , where  $t$  is negative and sufficiently small, and partitioning according to where  $x + ty$  is positive, negative and zero, denoting these new index sets  $\bar{S}_1, \bar{S}_2, \bar{S}_3$ . This may provide a different partition than before, but the computation of the lower bounds is exactly the same. Since we observed that the expressions for the lower bounds in (2.8) and (2.9) were increasing in  $t$ , and  $t$  is negative, we choose  $t \rightarrow 0$  to optimise these lower bounds for the spectral radii.

We now summarize the results of this section in the following theorem:

**Theorem 2.1.** *Let  $A$  be an  $n \times n$  irreducible and stochastic matrix, let  $\lambda = \alpha + i\beta$  be an eigenvalue of  $A$ , with  $\alpha, \beta > 0$ , and let  $x + iy$  be a right eigenvector of  $A$  corresponding to  $\lambda$ . For  $i = 1, 2, 3$ , let  $S_i, \tilde{S}_i$ , and  $\bar{S}_i$  be the index sets described in Section 2, let  $x_i, y_i, \tilde{x}_i, \tilde{y}_i, \bar{x}_i, \bar{y}_i$  be the subvectors of  $x$  and  $y$  corresponding to the index sets  $S_i, \tilde{S}_i$ , and  $\bar{S}_i$ , and let  $A_{ii}, \tilde{A}_{ii}$  and  $\bar{A}_{ii}$  be the principal submatrices of  $A$  corresponding to the index sets  $S_i, \tilde{S}_i$ , and  $\bar{S}_i$ . Then:*

1. If  $\alpha x_1 - \beta y_1 > 0$ ,

$$(i) \quad \rho(A_{11}) \geq \alpha - \beta \cdot \max_j \left\{ \frac{y_1(j)}{x_1(j)} \right\}.$$

(ii) If  $y_1 > 0$  and  $y_2 < 0$ , then

$$\rho(\tilde{A}_{11}) \geq \alpha + \beta \cdot \min_j \left\{ \frac{\tilde{x}_1(j)}{\tilde{y}_1(j)} \right\}.$$

Otherwise, if  $x_1$  and  $y_1$  are linearly independent,  $\alpha x_1 + \beta y_1 > 0$ , and  $\alpha x_2 + \beta y_2 < 0$ , then

$$\rho(\tilde{A}_{11}) \geq \alpha + \beta \cdot \min_j \left\{ \frac{t\tilde{x}_1(j) - \tilde{y}_1(j)}{\tilde{x}_1(j) + t\tilde{y}_1(j)} \right\},$$

where  $t > 0$  and is bounded above by

$$\min \left\{ \frac{-x(j)}{y(j)} \mid x(j)y(j) < 0 \right\}.$$

$$(iii) \quad \rho(\bar{A}_{11}) \geq \alpha - \beta \cdot \max_j \left\{ \frac{\bar{y}_1(j)}{\bar{x}_1(j)} \right\}.$$

2. If  $\alpha x_2 - \beta y_2 < 0$ ,

$$(i) \quad \rho(A_{22}) \geq \alpha - \beta \cdot \max_j \left\{ \frac{y_2(j)}{x_2(j)} \right\}.$$

(ii) If  $y_1 > 0$  and  $y_2 < 0$ , then

$$\rho(\tilde{A}_{22}) \geq \alpha + \beta \cdot \min_j \left\{ \frac{\tilde{x}_2(j)}{\tilde{y}_2(j)} \right\}.$$

Otherwise, if  $x_2$  and  $y_2$  are linearly independent,  $\alpha x_1 + \beta y_1 > 0$ , and  $\alpha x_2 + \beta y_2 < 0$ , then

$$\rho(\tilde{A}_{22}) \geq \alpha + \beta \cdot \min_j \left\{ \frac{t\tilde{x}_2(j) - \tilde{y}_2(j)}{\tilde{x}_2(j) + t\tilde{y}_2(j)} \right\},$$

where  $t > 0$  and is bounded above by

$$\min \left\{ \frac{-x(j)}{y(j)} \mid x(j)y(j) < 0 \right\}.$$

$$(iii) \quad \rho(\bar{A}_{22}) \geq \alpha - \beta \cdot \max_j \left\{ \frac{\bar{y}_2(j)}{\bar{x}_2(j)} \right\}.$$

### 3. A conformal partition with respect to the imaginary part of an eigenvector

We could also begin by partitioning the system with respect to where  $y$  is positive, negative, and zero. This potentially gives an entirely different partition, and we use the same approach to find lower bounds for the spectral radii of the submatrices of  $A$  corresponding to the index sets of this partition. For the purpose of avoiding congested notation, we will re-use  $y_1, y_2, x_1, x_2$ , and  $x_3$  in this section to denote the components of the vectors  $y$  and  $x$  once the system has been partitioned, where this time it is understood that  $y_1 > 0$  and  $y_2 < 0$ . Similarly, the submatrices  $A_{ij}$  now represent something different than in Sections 1 and 2.

In particular, we consider:

$$\left[ \begin{array}{c|c|c} A_{11} & A_{12} & A_{13} \\ \hline A_{21} & A_{22} & A_{23} \\ \hline A_{31} & A_{32} & A_{33} \end{array} \right] \left[ \begin{array}{c} y_1 \\ y_2 \\ 0 \end{array} \right] = \left[ \begin{array}{c} \beta x_1 + \alpha y_1 \\ \beta x_2 + \alpha y_2 \\ \beta x_3 \end{array} \right]$$

where  $y_1 > 0$  and  $y_2 < 0$ , entrywise. Let  $T_1, T_2, T_3$  denote the index sets of the partition. This equation produces, in the same way as before, the following inequalities:

$$\rho(A_{11}) \geq \alpha + \beta \min_j \left( \frac{x_1(j)}{y_1(j)} \right) \quad (3.1)$$

$$\rho(A_{22}) \geq \alpha + \beta \min_j \left( \frac{x_2(j)}{y_2(j)} \right). \quad (3.2)$$

We are once again in a situation where, if  $\alpha \approx 1$  and  $\beta \approx 0$ , these lower bounds are close to 1, indicating clustering behaviour in the Markov chain represented by the transition matrix  $A$ .

Note that we require new hypotheses in order to ensure these lower bounds are positive. In particular, we need

$$\alpha y_1 + \beta x_1 > 0 \quad \text{and} \quad \alpha y_2 + \beta x_2 < 0. \quad (3.3)$$

Pursuing the same analysis with repartitioning results in the following theorem:

**Theorem 3.1.** Let  $A$  be an  $n \times n$  irreducible and stochastic matrix, let  $\lambda = \alpha + i\beta$  be an eigenvalue of  $A$ , with  $\alpha, \beta > 0$ , and let  $x + iy$  be a right eigenvector of  $A$  corresponding to  $\lambda$ . For  $i = 1, 2, 3$ , let  $T_i$  denote the index sets obtained by partitioning according to where  $y$  is positive, negative and zero, and let  $x_i$  and  $y_i$  denote the subvectors of  $x$  and  $y$  corresponding to the index set  $T_i$ . Also, let

$$\tilde{T}_1 = T_1 \cup \{j \mid x_3(j) > 0\}, \quad \tilde{T}_2 = T_2 \cup \{j \mid x_3(j) < 0\}$$

and

$$\bar{T}_1 = T_1 \cup \{j \mid x_3(j) < 0\}, \quad \bar{T}_2 = T_2 \cup \{j \mid x_3(j) > 0\}.$$

and let  $\tilde{x}_i, \tilde{y}_i$ , and  $\bar{x}_i, \bar{y}_i$  be the subvectors of  $x$  and  $y$  corresponding to the index sets  $\tilde{T}_i$  and  $\bar{T}_i$ , respectively. Finally, let  $A_{ii}, \tilde{A}_{ii}$  and  $\bar{A}_{ii}$  be the principal submatrices of  $A$  corresponding to the index sets  $T_i, \tilde{T}_i$ , and  $\bar{T}_i$ . Then:

1. If  $\alpha y_1 + \beta x_1 > 0$ ,

- (i)  $\rho(A_{11}) \geq \alpha + \beta \cdot \min_j \left\{ \frac{x_1(j)}{y_1(j)} \right\}$ .
- (ii)  $\rho(\tilde{A}_{11}) \geq \alpha + \beta \cdot \min_j \left\{ \frac{\tilde{x}_1(j)}{\tilde{y}_1(j)} \right\}$ .
- (iii) If  $x_1 < 0$  and  $x_2 > 0$ , then

$$\rho(\bar{A}_{11}) \geq \alpha - \beta \cdot \max_j \left\{ \frac{\bar{y}_1(j)}{\bar{x}_1(j)} \right\}.$$

Otherwise, if  $x_1$  and  $y_1$  are linearly independent,  $\alpha y_1 - \beta x_1 > 0$ , and  $\alpha y_2 - \beta x_2 < 0$ , then

$$\rho(\bar{A}_{11}) \geq \alpha + \beta \cdot \min_j \left\{ \frac{\bar{x}_1(j) - s\bar{y}_1(j)}{s\bar{x}_1(j) + \bar{y}_1(j)} \right\},$$

where  $s < 0$  and is bounded below by

$$\min \left\{ \frac{-y(j)}{x(j)} \mid x(j)y(j) > 0 \right\}.$$

2. If  $\alpha y_2 + \beta x_2 < 0$ ,

- (i)  $\rho(A_{22}) \geq \alpha + \beta \cdot \min_j \left\{ \frac{x_2(j)}{y_2(j)} \right\}$ .
- (ii)  $\rho(\tilde{A}_{22}) \geq \alpha + \beta \cdot \min_j \left\{ \frac{\tilde{x}_2(j)}{\tilde{y}_2(j)} \right\}$ .
- (iii) If  $x_1 < 0$  and  $x_2 > 0$ , then

$$\rho(\bar{A}_{22}) \geq \alpha - \beta \cdot \max_j \left\{ \frac{\bar{y}_2(j)}{\bar{x}_2(j)} \right\}.$$

Otherwise, if  $x_2$  and  $y_2$  are linearly independent,  $\alpha y_1 - \beta x_1 > 0$ , and  $\alpha y_2 - \beta x_2 < 0$ , then

$$\rho(\bar{A}_{22}) \geq \alpha + \beta \cdot \min_j \left\{ \frac{\bar{x}_2(j) - s\bar{y}_2(j)}{s\bar{x}_2(j) + \bar{y}_2(j)} \right\},$$

where  $s < 0$  and is bounded below by

$$\min \left\{ \frac{-y(j)}{x(j)} \mid x(j)y(j) > 0 \right\}.$$

**Remark 3.2.** We note here the importance of the hypotheses in (2.5) and (3.3), and stress that it is not sufficient to simply determine an eigenvalue of the transition matrix which is sufficiently close to 1 and conclude that the associated Markov chain must exhibit clustering behaviour.

It is easily determined that the  $n \times n$  transition matrix

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

has eigenvalues  $\{e^{\frac{2\pi j i}{n}} \mid j = 0, 1, \dots, n-1\}$ , and hence for  $n$  large enough,  $A$  will have an eigenvalue  $\lambda = e^{\frac{2\pi i}{n}}$  sufficiently close to 1, with large real part and small imaginary part. However, the Markov chain represented by  $A$  has no clustering behaviour, since the chain transitions cyclically through the states in a deterministic fashion.

The eigenvector  $x + iy$  corresponding to  $\lambda = e^{\frac{2\pi i}{n}} = \cos(\frac{2\pi}{n}) + i \sin(\frac{2\pi}{n})$  is

$$\begin{bmatrix} \cos(0) \\ \cos(\frac{2\pi}{n}) \\ \cos(\frac{4\pi}{n}) \\ \vdots \\ \cos(\frac{2(n-1)\pi}{n}) \end{bmatrix} + i \begin{bmatrix} \sin(0) \\ \sin(\frac{2\pi}{n}) \\ \sin(\frac{4\pi}{n}) \\ \vdots \\ \sin(\frac{2(n-1)\pi}{n}) \end{bmatrix}.$$

Partitioning the system with respect to the sign pattern of  $x$  as in Section 2, we have

$$x_1 = \begin{bmatrix} \cos(0) \\ \cos(\frac{2\pi}{n}) \\ \vdots \\ \cos(\frac{2k\pi}{n}) \\ \cos(-\frac{2k\pi}{n}) \\ \vdots \\ \cos(-\frac{2\pi}{n}) \end{bmatrix}, \quad y_1 = \begin{bmatrix} \sin(0) \\ \sin(\frac{2\pi}{n}) \\ \vdots \\ \sin(\frac{2k\pi}{n}) \\ \sin(-\frac{2k\pi}{n}) \\ \vdots \\ \sin(-\frac{2\pi}{n}) \end{bmatrix},$$

where

$$k = \begin{cases} \lfloor \frac{n}{4} \rfloor, & n \not\equiv 0 \pmod{4}; \\ \frac{n}{4} - 1, & n \equiv 0 \pmod{4}. \end{cases}$$

Let  $\alpha + i\beta = \lambda = \cos(\frac{2\pi}{n}) + i \sin(\frac{2\pi}{n})$ . Considering the first of the hypotheses in (2.5), we have that each entry of  $\alpha x_1 + \beta y_1$  is of the form

$$\cos(\frac{2\pi}{n}) \cos(\frac{2\pi j}{n}) - \sin(\frac{2\pi}{n}) \sin(\frac{2\pi j}{n}) = \cos(\frac{2\pi(j+1)}{n})$$

for  $j \in \{-k, \dots, -1, 0, 1, \dots, k\}$ . Hence there is an entry of the vector  $\alpha x_1 - \beta y_1$  which is negative (in particular, the  $k + 1$  entry), and so the hypothesis does not hold. Similarly,  $\alpha x_2 - \beta y_2$  can be found to have a positive entry; furthermore due to the structure of the matrix and its eigenvectors, it is similarly determined that the hypotheses do not hold for the other case of partitioning with respect to  $y$ . Hence we conclude nothing about clustering behaviour of the Markov chain with transition matrix  $A$ , as expected.

#### 4. Simulations and examples

In this section we produce some numerical examples, determined somewhat randomly, in order to test this method of detecting clusters. To produce such examples for which we can test this process, we require irreducible stochastic matrices which have complex eigenvalues appropriately close to 1, and which have a certain degree of clustered structure built in. To design an appropriate matrix with a prescribed eigenvalue, we make use of the following technique, which is described below in a general matrix theory setting but is more well-known in spectral graph theory in the context of an *equitable partition* of the adjacency matrix of a graph (see [1, Section 2.3]).

Consider a square block matrix

$$X = \left[ \begin{array}{c|c|c|c} X_{11} & X_{12} & \cdots & X_{1m} \\ \hline X_{21} & X_{22} & \cdots & X_{2m} \\ \hline \vdots & \vdots & \ddots & \vdots \\ \hline X_{m1} & X_{m2} & \cdots & X_{mm} \end{array} \right]$$

and suppose that  $X$  has been partitioned into these blocks in such a way that each block  $X_{ij}$  has constant row sums  $q_{ij}$ . That is,  $X_{ij} \mathbb{1}_j = q_{ij} \mathbb{1}_{k_i}$  (where  $k_i$  is the number of rows in  $X_{ij}$ ).

Now consider that  $\lambda$  is an eigenvalue of  $X$  with right eigenvector  $[c_1 \mathbb{1}_{k_1} \mid c_2 \mathbb{1}_{k_2} \mid \cdots \mid c_m \mathbb{1}_{k_m}]^\top$  if and only if

$$\begin{bmatrix} q_{11} & q_{12} & \cdots & q_{1m} \\ q_{21} & q_{22} & \cdots & q_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ q_{m1} & q_{m2} & \cdots & q_{mm} \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_m \end{bmatrix} = \lambda \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_m \end{bmatrix}.$$

That is, the eigenvalues of this quotient matrix  $Q = [q_{ij}]$  are contained in the spectrum of  $X$ . We make use of this technique here by determining some  $3 \times 3$  matrix with appropriate eigenvalues to act as the quotient matrix, and then constructing a larger matrix where the first two diagonal blocks represent clusters of the chain.

Consider the  $3 \times 3$  matrix

$$B = \begin{bmatrix} \frac{5}{6} & 0 & \frac{1}{6} \\ \frac{3}{4} & \frac{1}{6} & \frac{1}{12} \\ \frac{2}{3} - 12t^2 & \frac{1}{3} + 12t^2 & 0 \end{bmatrix}$$

which, for any  $t \in [0, \frac{\sqrt{2}}{6}]$ , is nonnegative, stochastic, and irreducible, with eigenvalues  $1, \pm it$ . Taking a convex combination of this matrix and the identity will produce a  $3 \times 3$  matrix

$$C := (1 - s)I + sB, \quad 0 \leq s \leq 1$$

which has eigenvalues  $1, (1 - s) \pm i(st)$ .

Now let  $T_{11}$ ,  $T_{22}$  and  $T_{33}$  be irreducible stochastic matrices of orders  $n_1, n_2$ , and  $n_3$  respectively. Further, let  $T_{ij}$  be an  $n_i \times n_j$  rectangular row-stochastic matrix, for  $i, j \in \{1, 2, 3\}, i \neq j$ . Then

$$A = \left[ \begin{array}{c|c|c} c_{11}T_{11} & c_{12}T_{12} & c_{13}T_{13} \\ \hline c_{21}T_{21} & c_{22}T_{22} & c_{23}T_{23} \\ \hline c_{31}T_{31} & c_{32}T_{32} & c_{33}T_{33} \end{array} \right]$$

is an irreducible stochastic matrix of order  $n_1 + n_2 + n_3$  which has in its spectrum the complex eigenvalues  $(1 - s) \pm i(st)$ .

For  $s$  chosen appropriately, these eigenvalues are close to 1. Furthermore, the first two diagonal blocks of  $A$  represent potential clusters in that the probability of transitioning to another state indexed in the same block is high ( $\geq 1 - s$ ) relative to the probability of transitioning to a state indexed in another block. Note that although these transition matrices are constructed to have certain properties, we emphasise that this construction may involve many elements of randomness in order to effectively test the method presented in this article.

- $t$  is chosen uniformly at random from the interval  $[0, \frac{\sqrt{2}}{6}]$ .
- $s$  is chosen uniformly at random from the interval  $[0, 0.2]$ , so that the complex eigenvalue  $\lambda$  under consideration has real part  $\alpha \in [0.8, 1]$  and imaginary part  $\beta \in [0, \frac{\sqrt{2}}{30}]$ .
- The order  $n$  of  $A$  is fixed.  $n_1$  is an integer chosen at random from  $\{3, 4, \dots, \lfloor \frac{n}{2} \rfloor\}$ .  $n_2$  is an integer chosen at random from  $\{3, 4, \dots, n - n_1 - 3\}$ , and  $n_3 := n - n_1 - n_2$ . These integers are chosen in this way to avoid trivial or degenerate cases.
- The matrices  $T_{ij}$  are chosen as uniform stochastic random matrices respecting a zero pattern determined randomly with density  $p := 0.7$ .

We present a single example in detail that has been constructed in this way in order to illustrate both the construction and the analysis of the eigenvectors which produces the partitions and associated potential clusters.

**Example 4.1.**  $A$  is a  $100 \times 100$  matrix, with  $n_1 = 32$ ,  $n_2 = 53$ , and  $n_3 = 15$ .  $A$  has an eigenvalue  $\lambda = 0.9067 + 0.0106i$ ; i.e. the values of  $s$  and  $t$  chosen are 0.0933 and 0.1138, respectively.

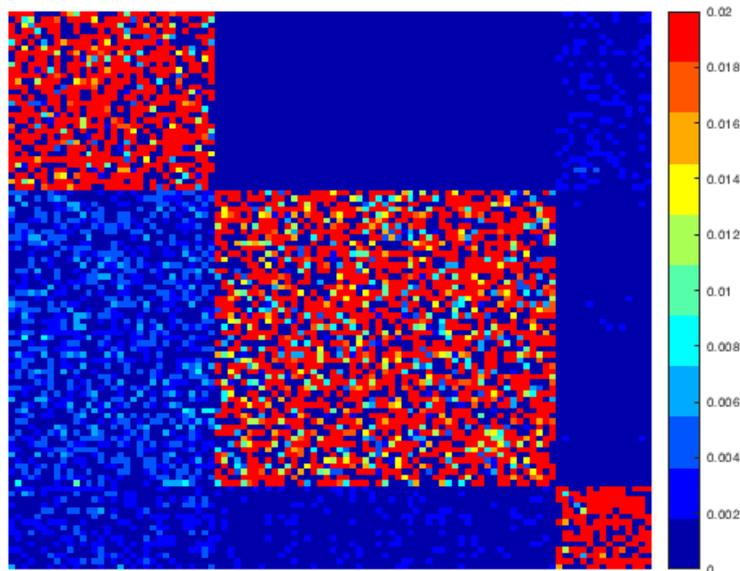


Figure 4.1: A heat map of the transition matrix  $A$  in Example 4.1. Evidence of clustering behaviour is indicated by the largely red blocks on the diagonal, indicating high probability of staying within one group of states, and low probability (blue) of leaving.

A heat map for  $A$  is given in Fig. 4.1, which illustrates the clustering behaviour of the Markov chain with transition matrix  $A$ . The index sets of the constructed clusters are  $\{1, 2, \dots, 32\}$ ,  $\{33, 34, \dots, 85\}$  and  $\{86, 87, \dots, 100\}$ .

The results of the analysis according to Theorem 2.1 are summarized in the following table:

Thm 2.1	index set	spectral radius	lower bound
1(i)	$\{1, 2, \dots, 32\}$	0.9844	0.9052
1(ii)	$\{1, 2, \dots, 32\}$	0.9844 <sup>1</sup>	0.9844 <sup>1</sup>
1(iii)	$\{1, 2, \dots, 32\}$	0.9844	0.9052
2(i)	$\{33, 34, \dots, 100\}$	0.9348	0.8932
2(ii)	$\{33, 34, \dots, 100\}$	0.9348	0.9150
2(iii)	$\{33, 34, \dots, 100\}$	0.9348	0.8932

We see that the partitions determined according to Theorem 2.1 recover the first cluster exactly, and present the second two clusters together with evidence of clustering behaviour on the union of those two sets of states. Note that although the ‘repartitioning’ process of Theorem 2.1 does not actually change the partitions, the calculation of the lower bounds in both cases is an improvement on the initial lower bound.

The results of the analysis according to Theorem 3.1 are summarized in the following table:

Thm 3.1	index set	spect. rad.	lower bound
1(i)	hypothesis not satisfied: $\alpha y_1 + \beta x_1 \not\approx 0$	-	-
1(ii)	hypothesis not satisfied: $\alpha y_1 + \beta x_1 \not\approx 0$	-	-
1(iii)	hypothesis not satisfied: $\alpha y_1 + \beta x_1 \not\approx 0$	-	-
2(i)	$\{33, \dots, 85, 88, 89, 90, 99\}$	0.9224	0.9150
2(ii)	$\{33, \dots, 85, 88, 89, 90, 92, 99\}$	0.9224	0.9150
2(iii)	hypothesis not satisfied: $\alpha x_2 + \beta y_2 \not\approx 0$	-	-

In this case, certain hypotheses do not hold, and hence we obtain less information about the clustering behaviour of this matrix from the partitions obtained with respect to the sign pattern of the imaginary part of the eigenvector. However, we note that while in the previous analysis, the second and third constructed clusters were presented as a single cluster, we see some differentiation here in that the partition obtained from Theorem 3.1 2(i) produces the index set  $\{33, \dots, 85, 88, 89, 90, 99\}$  with equally strong evidence of clustering on this subset of the state space, which is essentially composed of the second constructed cluster, plus four states from the third constructed cluster.

We produce 10,000 such matrices with in-built clustered structure, then run this same analysis of the eigenvectors corresponding to eigenvalues close to 1 using Theorems 2.1 and 3.1. The goal is to determine how frequently we can recover the clusters that we constructed, and to determine lower bounds on the spectral radii of the corresponding submatrices. The magnitude of these lower bounds is an indicator of the ‘tightness’ of the detected cluster, in the manner described in Theorem 1.1, parts (b) and (c).

Our results are as follows:

---

<sup>1</sup>The spectral radius and the lower bound in this case differ in the  $10^{-15}$  position.

- The first cluster is recovered exactly in 99.6% of cases.
- The second cluster is recovered exactly in 9.12% of cases.
- The third cluster is recovered exactly in 0.4% of cases.
- The first and second clusters are both recovered exactly in 8.94% of cases. The first and third are recovered exactly in 0.27% of cases. The second and third are never both recovered exactly in the same simulation; neither are all three ever recovered exactly using this analysis.
- In *every* case that the first cluster is recovered, the second and third clusters are both partially recovered in that they are each produced in some larger subset of states on which the Markov chain displays clustering behaviour. In 96.74% of these cases where the second cluster is not recovered, the second cluster is produced by this algorithm along with some states from the third cluster as an index set of states on which the Markov chain displays clustering behaviour, in a similar manner as discussed in Example 4.1. This occurs with the third cluster in only 6 cases out of a total of 9039.

We note that in the extreme majority of these examples, the first constructed cluster is recovered exactly, and the second and third are often grouped together as a single cluster. We expect that this is an artefact of the particular  $3 \times 3$  matrix used to produce these examples. In particular, the spectral radii of the submatrices corresponding to the first, second and third clusters are  $1 - \frac{s}{6}$ ,  $1 - \frac{5s}{6}$ , and  $1 - s$ , respectively. Hence it is not unexpected that the first cluster is recovered more frequently, as by this measure alone it represents a ‘stronger’ or ‘tighter’ cluster than the others.

As a final remark, we observe that once this tightest cluster is determined, we could use a ‘divide-and-conquer’ approach and consider a principal submatrix with the rows and columns corresponding to this cluster removed. Renormalizing the rows of this matrix will produce a stochastic transition matrix which represents a parallel Markov chain acting only on the states of the second and third cluster. It is possible that we could then perform a similar analysis on this new matrix, if it had eigenvalues appropriately close to one, and there is a chance that we could recover the second and third clusters separately.

To illustrate, we refer to Example 4.1, whose first cluster on the states  $\{1, 2, \dots, 32\}$  was recovered exactly. We consider the matrix  $\hat{A}$  constructed as above by appropriately renormalizing the principal submatrix obtained from  $A$  by deleting the first 32 rows and columns. This matrix has an eigenvalue equal to 0.9437. By considering the sign pattern of the associated eigenvector and applying Theorem 1.1, we conclude evidence of clustering behaviour on the states indexed by  $\{33, 34, \dots, 85\}$  and  $\{86, 87, \dots, 100\}$ ; that is, we recover both constructed clusters exactly. An investigation into possible divide-and-conquer approaches will be the subject of a forthcoming article.

**Example 4.2.** In this example, we apply Theorems 2.1 and 3.1 to the transition matrix for a simulated bus network discussed in [7] in which states represent bus stops in a network and transition probabilities are simulated probabilities of passengers moving from one stop to another through the bus network. In [7], the authors observed evidence of clustering behaviour based on the mean first passage matrix (shown in Fig. 4.2a) and observed that the eigenvector corresponding to the eigenvalue of second largest modulus (which was, in this case, complex) could be used to determine these clusters. A plot of the entries of this vector is given in Fig. 4.2b in which it can clearly be seen that the entries are clustered, with clusters indexed by  $\{1, 2, \dots, 6\}$ ,  $\{7, 8, \dots, 11\}$ , and  $\{12, 13, \dots, 17\}$ , which are the subsets of states for which the mean first passage times seem to predict clustering.

The following tables summarize the findings of Theorems 2.1 and 3.1 with regard to this eigenvalue  $\lambda = 0.9998 + 0.0001i$  and its corresponding eigenvector.

Thm 2.1	index set	spectral radius	lower bound
1(i)	$\{12, 13, \dots, 17\}$	0.9998	0.9997
1(ii)	$\{12, 13, \dots, 17\}$	0.9998 <sup>1</sup>	0.9998 <sup>1</sup>
1(iii)	$\{12, 13, \dots, 17\}$	0.9998	0.9997
2(i)	$\{1, 2, \dots, 11\}$	0.9999	0.9996
2(ii)	$\{1, 2, \dots, 11\}$	0.9999	0.9996
2(iii)	$\{1, 2, \dots, 11\}$	0.9999	0.9996

Thm 3.1	index set	spect. rad.	lower bound
1(i)	$\{7, 8, \dots, 16\}$	0.9999	0.9997
1(ii)	$\{7, 8, \dots, 17\}$	0.9999	0.9997
1(iii)	hypothesis not satisfied: $\alpha x_1 + \beta y_1 \neq 0$	-	-
2(i)	$\{1, 2, \dots, 6\}$	0.9999 <sup>2</sup>	0.9999 <sup>2</sup>
2(ii)	$\{1, 2, \dots, 6\}$	0.9999 <sup>3</sup>	0.9999 <sup>3</sup>
2(iii)	hypothesis not satisfied: $\alpha x_1 + \beta y_1 \neq 0$	-	-

With some examination of this information, one can recover all three clusters.

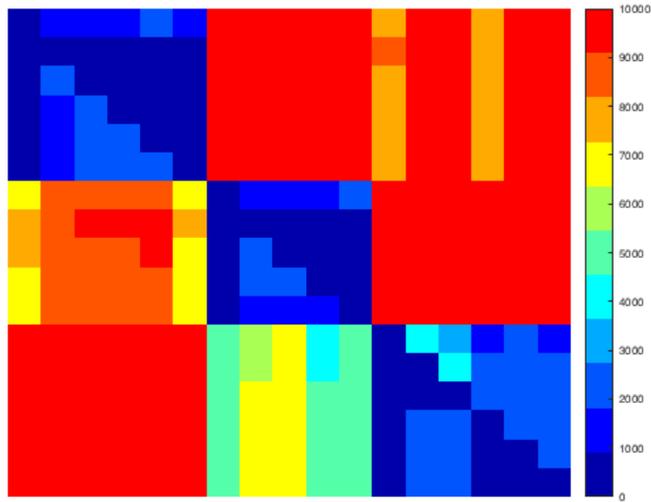
*Acknowledgements:*

This research was supported in part by the University of Manitoba Graduate Fellowship (Jane Breen), by Science Foundation Ireland under grant 11/PI/1177 (Mahsa Faizrah-nemoun and Robert Shorten), and by a Discovery Grant from the Natural Sciences and

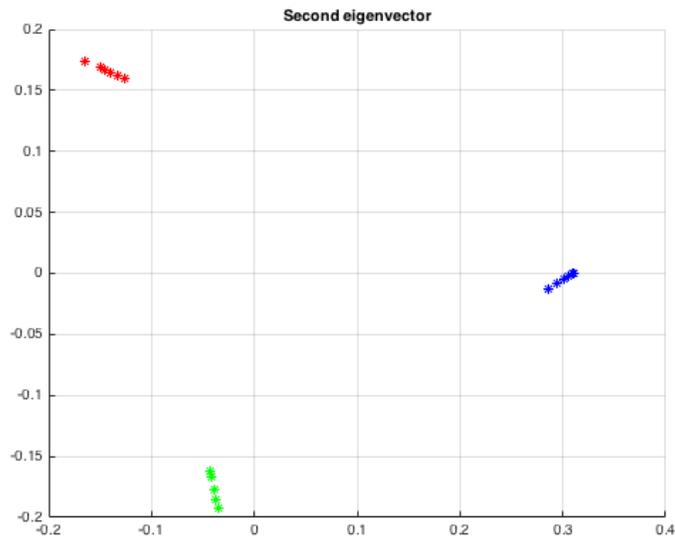
<sup>1</sup>The spectral radius and the lower bound in this case differ in the  $10^{-6}$  position.

<sup>2</sup>The spectral radius and the lower bound in this case differ in the  $10^{-5}$  position.

<sup>3</sup>The spectral radius and the lower bound in this case differ in the  $10^{-5}$  position.



(a) Heat map of the mean first passage matrix for the transition matrix of the simulated bus network. Evidence of clustering behaviour is given by the relatively low (blue) values for mfp times between states in the same cluster, and high values for mfp times between clusters.



(b) Entries of the eigenvector corresponding to the eigenvalue of second-largest modulus

Figure 4.2: Evidence of clustering in the transition matrix for the bus network in [7], discussed in Example 4.2.

Engineering Research Council of Canada under grant number RGPIN/6123-2014 (Steve Kirkland).

The authors wish to express their thanks to an anonymous reviewer, whose comments and suggestions improved this article.

## References

- [1] Andries E. Brouwer and Willem H. Haemers. *Spectra of Graphs*. Springer New York, 2012.
- [2] Emanuele Crisostomi, Stephen 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.
- [3] Peter Deuffhard, Wilhelm Huisinga, Alexander Fischer, and Christof Schütte. Identification of almost invariant aggregates in reversible nearly uncoupled Markov chains. *Linear Algebra and its Applications*, 315(1-3):39–59, 2000.
- [4] Peter Deuffhard and Christof Schütte. Molecular conformation dynamics and computational drug design. *Applied Mathematics Entering the 21st Century. Proceedings ICIAM*, pages 91–119, 2004.
- [5] Peter Deuffhard and Marcus Weber. Robust Perron cluster analysis in conformation dynamics. *Linear Algebra and its Applications*, 398:161–184, 2005.
- [6] Mahsa Faizrahnemoon. *Real-data modelling of transportation networks*. PhD thesis, National University of Ireland, Maynooth, 2016.
- [7] Mahsa Faizrahnemoon, Arieh Schlote, Lorenzo Maggi, Emanuele Crisostomi, and Robert Shorten. A big-data model for multi-modal public transportation with application to macroscopic control and optimisation. *International Journal of Control*, 88(11):2354–2368, 2015.
- [8] Gary Froyland. Extracting dynamical behavior via Markov models. *Nonlinear Dynamics and Statistics*, pages 281–312, 2001.
- [9] D. J. Hartfiel and Carl D. Meyer. On the structure of stochastic matrices with a subdominant eigenvalue near 1. *Linear Algebra Appl.*, 272:193–203, 1998.
- [10] Jeffrey Hunter. Mixing times with applications to perturbed Markov chains. *Linear Algebra and its Applications*, 417(1):108–123, 2006.
- [11] Mark Levene and George Loizou. Kemeny’s constant and the random surfer. *The American Mathematical Monthly*, 109(8):741–745, 2002.

- [12] Volker Mehrmann, Daniel B. Szyld, and Elena Virnik. An SVD approach to identifying metastable states of Markov chains. *Electronic Transactions on Numerical Analysis*, 29:46–69, 2008.
- [13] Christof Schütte and Wilhelm Huisinga. Biomolecular conformations can be identified as metastable sets of molecular dynamics. *Handbook of Numerical Analysis*, 10:699–744, 2003.
- [14] Eugene Seneta. *Non-negative matrices and Markov chains*. Springer-Verlag, New York, 1981.
- [15] Ryan M. Tifenbach. On an SVD-based algorithm for identifying meta-stable states of Markov chains. *Electronic Transactions on Numerical Analysis*, 38:17–33, 2011.

### **Appendix: Flowcharts for Theorem 2.1 and Theorem 3.1**

On the next two pages, we provide flowcharts for each of Theorem 2.1 and Theorem 3.1 that clearly display the dependencies of the lower bounds on the technical hypotheses.

