Electronic Transactions on Numerical Analysis.
Volume 40, pp. 120-147, 2013.

ISSN 1068-9613.

ETNA
Kent State University
http://etna.math.kent.edu

A COMBINATORIAL APPROACH TO NEARLY UNCOUPLED MARKOV CHAINS I: REVERSIBLE MARKOV CHAINS*

RYAN M. TIFENBACH†

Abstract. A Markov chain is a sequence of random variables $x_0, x_1, \ldots$ that take values in a state space $\mathcal{S}$. A set $\mathcal{E} \subseteq \mathcal{S}$ is referred to as an almost invariant aggregate if transitions from x_t to x_{t+1} where $x_t \in \mathcal{E}$ and $x_{t+1} \notin \mathcal{E}$ are exceedingly rare. A Markov chain is referred to as nearly uncoupled if there are two or more disjoint almost invariant aggregates contained in its state space. Nearly uncoupled Markov chains are characterised by long periods of relatively constant behaviour punctuated by sudden, extreme changes. We present an algorithm for producing almost invariant aggregates of a nearly uncoupled reversible Markov chain. This algorithm utilises the stochastic complement to iteratively reduce the order of the given state space.

Key words. nearly uncoupled Markov chain, reversible Markov chain, stochastic complement, stochastic matrix

AMS subject classifications. 15A18, 15A51, 60J10, 60J20, 65F15

1. Preliminaries.

This article contains material previously presented in [14]. A discrete-time, time-homogeneous *Markov chain* on a finite state space $\mathcal{S}$ is a sequence of random variables

$$X = (x_0, x_1, x_2, \ldots)$$

that take values in $\mathcal{S}$ and satisfy the *Markov property.* The Markov property is the statement that for all $t \geq 0$ and $i, j \in \mathcal{S}$, the probability

$$\mathbb{P}\left[x_{t+1} = j \,|\, x_t = i\right]$$

is independent of t and of the values taken on by $x_0, \ldots, x_{t-1}$ (if $t \geq 1$). That is, for all $i, j \in \mathcal{S}$, the probability of $x_{t+1} = j$ given that $x_t = i$ is uniquely determined by i and j.

For the remainder of this work, we will simply use the label Markov chain with the implicit understanding that we only consider the discrete-time, time-homogeneous case and a finite state space.

Let X be a Markov chain on $\mathcal{S}$. For each $i, j \in \mathcal{S}$, the value

$$a_{ij} = \mathbb{P}\left[x_{t+1} = j \,|\, x_t = i\right]$$

is referred to as the ijth transition probability of X. The square matrix A on the index set $\mathcal{S}$ whose entries are the transition probabilities is referred to as the transition matrix of X. Such a matrix is *stochastic:* it is square, entrywise nonnegative, and the sum of the entries in each row is 1. These facts are summarised by the equations

$$A \geq 0 \text{ and } A\mathbb{1} = \mathbb{1}.$$

The symbol $\mathbb{1}$ is used for the column vector with every entry equal to 1. We refer to the index set of a stochastic matrix as its state space.

Let A be a stochastic matrix on the state space $\mathcal{S}$ and $\mathcal{C}_1, \mathcal{C}_2 \subseteq \mathcal{S}$ be nonempty. The *submatrix*

$$A(\mathcal{C}_1, \mathcal{C}_2) = [a_{ij}]_{i \in \mathcal{C}_1, j \in \mathcal{C}_2}$$

*Received August 6, 2012. Accepted March 6, 2013. Published online on June 29, 2013. Recommended by D. Szyld.

†Department of Mathematics and Statistics, Memorial University of Newfoundland, St. John's, NL, Canada (ryan.tifenbach@mun.ca).

ETNA
Kent State University
http://etna.math.kent.edu

is obtained by deleting each of the ith rows where $i \notin \mathcal{C}_1$ and each of the jth columns where $j \notin \mathcal{C}_2$. When $\mathcal{C}_1 = \mathcal{C}_2 = \mathcal{C}$, we refer to the matrix $A(\mathcal{C}, \mathcal{C}) = A(\mathcal{C})$ as the *principal submatrix* of A on $\mathcal{C}$.

The subscript t is referred to as the time parameter of X. If $x_t = i$ and $x_{t+1} = j$, we say that the Markov chain visits the states i and j at times t and $t+1$, respectively, and that the Markov chain transitions from i to j at time $t+1$.

We refer to a diagonal matrix where each diagonal entry is positive as a positive diagonal matrix. We use the notation π_i to refer to the ith diagonal entry of a diagonal matrix Π.

Let X be a Markov chain with transition matrix A. If there is a positive diagonal matrix Π such that ΠA is symmetric, we refer to X as a *reversible Markov chain* and A as a *reversible stochastic matrix*. When ΠA is symmetric, we have $\pi_i a_{ij} = \pi_j a_{ji}$ for all i and j. Let A be an irreducible reversible stochastic matrix, let Π be a positive diagonal matrix such that ΠA is symmetric, and let $\pi = \Pi \mathbb{1}$. Then,

$$\pi^T A = \mathbb{1}^T \Pi A = \mathbb{1}^T A^T \Pi = \mathbb{1}^T \Pi = \pi^T,$$

and we see that π is a left-eigenvector of A. It is straightforward to show that when A is reversible and π is a left-eigenvector associated with 1 (in particular, when π is a stationary distribution) that ΠA is symmetric, where Π is the unique diagonal with $\Pi \mathbb{1} = \pi$.

Reversible Markov chains are characterised by symmetry in time. If the Markov chain X is reversible, then

$$\lim_{t \to \infty} \mathbb{P}\left[x_{t-1} = j \,|\, x_t = i\right] = \mathbb{P}\left[x_{t+1} = j \,|\, x_t = i\right].$$

That is, when X is reversible, then as $t \to \infty$, for any i and j, if the Markov chain visits state i at time t, the probability that the previous state visited is j converges to the probability that the next state visited is j. If we imagine "rewinding" a recording of such a Markov chain, the result would be indistinguishable from the original.

EXAMPLE 1.1. The canonical example of a reversible Markov chain is the random walk on a graph. Let G be a weighted undirected graph with vertex set $\mathcal{S}$ and weight function w. We impose the restriction that G does not have any isolated vertices. We label the edges of G via their endpoints: the edge ij is an edge joining the vertices i and j. As the graph G is undirected, the edge ji is identical to the edge ij. The weight function w assigns a positive number to every edge in G: for all pairs of vertices $i, j \in \mathcal{S}$ (not necessarily distinct), we have $w_{ij} = w_{ji} > 0$ if the edge ij is present in G and $w_{ij} = 0$ if the edge ij is not present in G. The random walk on G is a Markov chain on the state space equal to $\mathcal{S}$, where transitions occur along the edges of G. The transition probabilities are determined by the weight function:

$$\mathbb{P}\left[x_{t+1} = j \,|\, x_t = i\right] = \frac{w_{ij}}{\sum_{k \in \mathcal{S}} w_{ik}}.$$

As long as G contains no isolated vertices, the above is a uniquely determined (by G and w) Markov chain on $\mathcal{S}$. Let A be the transition matrix of the random walk X

$$a_{ij} = \frac{w_{ij}}{\sum_{k \in \mathcal{S}} w_{ik}}.$$

For each $i \in \mathcal{S}$, let

$$\pi_i = \sum_{k \in \mathcal{S}} w_{ik},$$

ETNA
Kent State University
http://etna.math.kent.edu

and let Π be the diagonal matrix with the ith diagonal entry equal to π_i. Then the ijth entry of ΠA is equal to w_{ij} —this matrix is symmetric by definition.

The treatment in this work makes extensive use of the assumption of reversibility. In a future piece, we will present a similar analysis of the nonreversible case.

Let $\mathcal{S}$ be a finite collection. A probability distribution vector on $\mathcal{S}$ is a column vector with nonnegative entries indexed by $\mathcal{S}$, the sum of which is 1. A probability distribution vector v models some random variable x on $\mathcal{S}$: the value v_i is the probability that $x = i$.

Let X be a Markov chain with transition matrix A. A stationary vector of X is a probability distribution vector π such that $\pi^T A = \pi^T$.

Two matrices are *permutation similar* (denoted by $\cong$) if one can be obtained from the other by simultaneously reordering the rows and columns. That is, $A \cong B$ if there is a bijection f such that $a_{f(i),f(j)} = b_{ij}$ for all i and j.

A Markov chain X is irreducible if for any $i, j \in \mathcal{S}$ (not necessarily distinct) it is possible for the Markov chain to visit i and then, at some later time, visit j. A square matrix A is reducible if its indices can be reordered so that

$$A \cong \begin{bmatrix} B_1 & 0 \\ C & B_2 \end{bmatrix},$$

where each B_i is square and has order greater than or equal to 1. The matrix A is irreducible if it is not reducible. A Markov chain is irreducible if and only if its associated transition matrix is irreducible.

A square matrix B is *substochastic* if it is square, entrywise nonnegative, and the sum of the entries in each row is less than or equal to 1. We refer to B as *properly substochastic* if it has no principal submatrices that are stochastic. An application of the well-known Perron-Frobenius theorem implies that a substochastic matrix B is properly substochastic if and only if every eigenvalue λ of B has $|\lambda| < 1$. Moreover, if A is an irreducible stochastic matrix and B is a principal submatrix of A of strictly smaller order, then B is properly substochastic. A substochastic matrix B is reversible in precisely the same manner as a stochastic matrix: B is reversible if there is a positive diagonal matrix Π such that ΠB is symmetric. Principal submatrices of reversible stochastic matrices are reversible substochastic matrices.

In the following two sections we define the stochastic complement and the concept of a nearly uncoupled Markov chain. In Section 4 we present a collection of algorithms which use the stochastic complement to analyse the nearly uncoupled structure of a given Markov chain and in Section 5 we present a number of examples of such an analysis. The appendix contains the calculation of a specific lower bound concerning reversible stochastic complements of which we make extensive use—the reader may wish to examine the appendix before reading Section 4.

2. The stochastic complement. We introduce the concept of a stochastic complement. Expositions of the properties of the stochastic complement appear in [10, 14].

DEFINITION 2.1. *Let X be an irreducible Markov chain on the state space $\mathcal{S}$. Let $\mathcal{C}$ be a proper nonempty subset of $\mathcal{S}$ and express the transition matrix of X as*

$$A \cong \begin{bmatrix} B_1 & B_{12} \\ B_{21} & B_2 \end{bmatrix},$$

where B_1 is the principal submatrix of A corresponding to $\mathcal{S} \setminus \mathcal{C}$ and B_2 is the principal submatrix corresponding to $\mathcal{C}$. We refer to the matrix

$$A \setminus \mathcal{C} = B_1 + B_{12} (I - B_2)^{-1} B_{21}$$

ETNA
Kent State University
http://etna.math.kent.edu

as the stochastic complement which removes $\mathcal{C}$ or as the stochastic complement on $\mathcal{S} \setminus \mathcal{C}$. When X and A are reducible, the stochastic complement $A \setminus \mathcal{C}$ is defined as above, as long as the principal submatrix B_2 is properly substochastic.

When we are removing a single state via a stochastic complement, we will use the notation $A \setminus i$ rather than $A \setminus \{i\}$. We will preserve indices/state labels in stochastic complements throughout this work. That is, if the state space of the Markov chain associated with A is $\mathcal{S}$, then the entries of $A \setminus \mathcal{C}$ are indexed by $\mathcal{S} \setminus \mathcal{C}$. For example, consider the stochastic matrix

$$A = \begin{bmatrix} 0.1 & 0.2 & 0.7 \\ 0.3 & 0.4 & 0.3 \\ 0.2 & 0.8 & 0.0 \end{bmatrix}$$

with the regular indices of $\mathcal{S} = \{1, 2, 3\}$. Then,

$$A \setminus 2 = \begin{bmatrix} 0.1 & 0.7 \\ 0.2 & 0.0 \end{bmatrix} + \begin{bmatrix} 0.2 \\ 0.8 \end{bmatrix} \left(I - \begin{bmatrix} 0.4 \end{bmatrix}\right)^{-1} \begin{bmatrix} 0.3 & 0.3 \end{bmatrix} = \begin{bmatrix} 0.2 & 0.8 \\ 0.6 & 0.6 \end{bmatrix}.$$

We will view this matrix as having the indices $\mathcal{S} \setminus 2 = \{1, 3\}$. The $(1, 3)$-entry of $A \setminus 2$ is 0.8 and $A \setminus 2$ has no $(1, 2)$-entry, for instance.

The following proposition summarises some results in [10, 14].

PROPOSITION 2.2. *Let X be a Markov chain on the state space $\mathcal{S}$ with transition matrix A. Let $\mathcal{C} \subseteq \mathcal{S}$, and let $\hat{A} = A \setminus \mathcal{C}$ be a stochastic complement. Then, the matrix $\hat{A}$ is a stochastic matrix. Moreover, for each $i, j \notin \mathcal{C}$, the value $\hat{a}_{ij}$ is the probability that after visiting i, the state j is the very first state not in $\mathcal{C}$ that the Markov chain transitions into. That is, given $x_{t'} = i$,*

$$\hat{a}_{ij} = \mathbb{P}\left[x_T = j\right],$$

where

$$T = \inf\left\{t \geq t' + 1 \,|\, x_t \notin \mathcal{C}\right\}.$$

Thus, we can view the stochastic complement which removes $\mathcal{C}$ as a method of editing out the collection $\mathcal{C}$ from the Markov chain X. A stochastic complement $A \setminus \mathcal{C}$ models the Markov chain obtained by "fast-forwarding" through the states in $\mathcal{C}$, ignoring any time spent in that collection.

In [10], it is noted that the stochastic complement is a specific instance of the more general Schur complement. The following proposition can be deduced from known properties of the Schur complement; see [16].

PROPOSITION 2.3. *Let X be an irreducible Markov chain on the state space $\mathcal{S}$ with transition matrix A. Let $\mathcal{C}_1$ and $\mathcal{C}_2$ be nonempty disjoint subsets of $\mathcal{S}$ such that $\mathcal{C} = \mathcal{C}_1 \cup \mathcal{C}_2 \neq \mathcal{S}$. Then, the stochastic complement on $\mathcal{S} \setminus \mathcal{C}$ can be formed via two stochastic complements by first removing $\mathcal{C}_1$ and then removing $\mathcal{C}_2$. That is,*

$$A \setminus (\mathcal{C}_1 \cup \mathcal{C}_2) = (A \setminus \mathcal{C}_1) \setminus \mathcal{C}_2.$$

We will make extensive use of this proposition. In particular, we will often form stochastic complements by removing states one at a time. This will allow us to calculate a stochastic complement without having to calculate a matrix inverse. Let A be a stochastic matrix on the state space $\mathcal{S}$, and suppose that $a_{ii} < 1$. Express

$$A \cong \begin{bmatrix} B & v \\ w^T & a_{ii} \end{bmatrix},$$

ETNA
Kent State University
http://etna.math.kent.edu

where the second position corresponds to state i and the first to the remainder of the state space, $\mathcal{S} \setminus i$. (We further suppose that the ordering of $\mathcal{S} \setminus i$ has not been altered in this expression.) Then,

$$A \setminus i = B + \frac{1}{1 - a_{ii}} v w^T.$$

PROPOSITION 2.4. *Let X be a reversible Markov chain with transition matrix A and state space $\mathcal{S}$. Let Π be a positive diagonal matrix such that ΠA is symmetric. Let $\hat{A} = A \setminus \mathcal{C}$ be a stochastic complement, and let $\hat{\Pi} = \Pi(\mathcal{S} \setminus \mathcal{C})$ be the principal submatrix of Π on the collection $\mathcal{S} \setminus \mathcal{C}$. Then, the matrix $\hat{A}$ is reversible via the fact that $\hat{\Pi}\hat{A}$ is symmetric.*

Proof. Express

$$A \cong \begin{bmatrix} B_1 & B_{12} \\ B_{21} & B_2 \end{bmatrix} \text{ and } \Pi \cong \begin{bmatrix} \Pi_1 & 0 \\ 0 & \Pi_2 \end{bmatrix},$$

where the first diagonal block corresponds to $\mathcal{S}\setminus\mathcal{C}$ and the second corresponds to $\mathcal{C}$. Since ΠA is symmetric, we have

$$\Pi_1 B_{12} = B_{21}^T \Pi_2 \text{ and } \Pi_2 B_{21} = B_{12}^T \Pi_1.$$

As well, since $\Pi_1 B_1$ and $\Pi_2 B_2$ are symmetric, it holds that

$$\Pi_1 B_1 = B_1^T \Pi_1 \text{ and } \Pi_2(I - B_2) = (I - B_2^T)\Pi_2.$$

The second equality further implies that

$$(I - B_2^T)^{-1}\Pi_2 = \Pi_2(I - B_2)^{-1}.$$

So, we simply calculate the transpose of $\Pi_1(A \setminus \mathcal{C})$:

$$\begin{aligned}(\Pi_1(A \setminus \mathcal{C}))^T &= \left(B_1 + B_{12}(I - B_2)^{-1}B_{21}\right)^T \Pi_1 = B_1^T\Pi_1 + B_{21}^T(I - B_2^T)^{-1}B_{12}^T\Pi_1 \\ &= \Pi_1 B_1 + B_{21}^T(I - B_2^T)^{-1}\Pi_2 B_{21} = \Pi_1 B_1 + B_{21}^T\Pi_2(I - B_2)^{-1}B_{21} \\ &= \Pi_1 B_1 + \Pi_1 B_{12}(I - B_2)^{-1}B_{21} = \Pi_1(A \setminus \mathcal{C}). \qquad \square\end{aligned}$$

3. Nearly uncoupled Markov chains.

DEFINITION 3.1. *Let X be a Markov chain on the state space $\mathcal{S}$. Let $\mathcal{E}$ be a nonempty proper subset of $\mathcal{S}$, and let ϵ be a small (near zero) positive constant. We refer to $\mathcal{E}$ as an almost invariant aggregate with respect to ϵ if, given $x_t \in \mathcal{E}$, the probability that $x_{t+1} \notin \mathcal{E}$ is less than or equal to ϵ.*

The definition of an almost invariant aggregate given above is somewhat nebulous—the probability that $x_t \in \mathcal{E}$ depends both on the parameter t and on the initial distribution of the Markov chain. Below, we make this more specific by showing two different manners in which a collection $\mathcal{E}$ may be considered to be almost invariant.

We use the symbol $\mathbb{1}$ to refer to the column vector with every entry equal to 1. Let B be a substochastic matrix. We define the *error vector of* B to be the nonnegative vector

$$\gamma_B = \mathbb{1} - B\mathbb{1} = (I - B)\mathbb{1}.$$

The error vector is a measure of how "close" B is to being stochastic. If $\gamma_B = 0$, then B is stochastic; if every entry of γ_B is near 0, then every row sum of B is near 1 and thus B is "nearly" stochastic.

ETNA
Kent State University
http://etna.math.kent.edu

Let X be an irreducible Markov chain with transition matrix A, and let π be the unique stationary distribution. Let $\mathcal{E}$ be a nonempty proper subset of the associated state space, let $B = A(\mathcal{E})$ be the associated principal submatrix, and let $\gamma = \gamma_B$. For each $i \in \mathcal{E}$, the ith entry of γ is the probability of transitioning from i to a state not in $\mathcal{E}$:

$$\gamma(i) = \mathbb{P}\left[x_{t+1} \notin \mathcal{E} \,|\, x_t = i\right].$$

If $\gamma \leq \epsilon\mathbb{1}$ (if every entry of γ is less than or equal to ϵ), then the probability of transitioning from a state in $\mathcal{E}$ to a state not in $\mathcal{E}$ is less than or equal to ϵ. Thus, regardless of any other conditions, if this condition holds, then transitions that exit $\mathcal{E}$ have a probability less than or equal to ϵ of occurring. Such an occurrence clearly implies that $\mathcal{E}$ is an almost invariant aggregate.

However, there is a weaker condition which still implies that transitions exiting $\mathcal{E}$ are a rare occurrence. Let π be the unique stationary distribution of A, and let $\hat{\pi} = \pi(\mathcal{E})$ be the subvector corresponding to $\mathcal{E}$. If the Markov chain has an initial distribution equal to π, then the value

$$\frac{\hat{\pi}^T(I - B)\mathbb{1}}{\hat{\pi}^T\mathbb{1}} = \frac{\hat{\pi}^T\gamma_B}{\hat{\pi}^T\mathbb{1}}$$

is the probability that $x_{t+1} \notin \mathcal{E}$ given $x_t \in \mathcal{E}$. Moreover, for any initial distribution, it holds that

$$\frac{\hat{\pi}^T(I - B)\mathbb{1}}{\hat{\pi}^T\mathbb{1}} = \lim_{t\to\infty} \mathbb{P}\left[x_{t+1} \notin \mathcal{E} \,|\, x_t \in \mathcal{E}\right].$$

That is, $\frac{\hat{\pi}^T\gamma_B}{\hat{\pi}^T\mathbb{1}}$ is the expected long-term probability of transitioning from a state in $\mathcal{E}$ to a state not in $\mathcal{E}$. Thus, if

$$\frac{\hat{\pi}^T\gamma_B}{\hat{\pi}^T\mathbb{1}} \leq \epsilon,$$

then transitions from $\mathcal{E}$ to $\mathcal{S} \setminus \mathcal{E}$ become rare as $t \to \infty$. In [3, 4], this value is referred to as the π-coupling measure of a collection $\mathcal{E}$, and is denoted by $w_\pi(\mathcal{E})$:

$$w_\pi(\mathcal{E}) = \frac{\hat{\pi}^T\gamma_B}{\hat{\pi}^T\mathbb{1}}.$$

It is easily shown that the first criterion above implies the second. That is, if the principal submatrix B represents transitions within $\mathcal{E}$ and $\gamma_B \leq \epsilon\mathbb{1}$, then $w_\pi(\mathcal{E}) \leq \epsilon$.

We are interested in the following problem: given a reversible stochastic matrix A and a small positive value ϵ, can we produce a partition of the state space $\mathcal{S}$ into almost invariant aggregates. We present an algorithm which utilises the stochastic complement to produce a candidate partition.

This problem is investigated in-depth in [3, 4, 13]. There, the approach utilises the spectral decomposition of the stochastic matrix in question. When a stochastic matrix is nearly uncoupled, it has multiple eigenvalues very close to the number 1. The eigenvalues of a stochastic matrix that are near to 1 are referred to as a *Perron cluster*. In [3, 4], the eigenvectors associated with the Perron cluster are used to produce a candidate partition through a process referred to as *Perron cluster cluster analysis.*

In [6], a similar approach that utilises the singular value decomposition rather than the spectral decomposition is proposed. We present an additional discussion concerning this approach in [15].

ETNA
Kent State University
http://etna.math.kent.edu

The imprecision of Definition 3.1 is, in a sense, imposed by the applications of the concept. The analysis of almost invariant aggregates appears in pharmaceutical design; see [4, 13], for example. A biomolecule often has multiple molecular configurations it may assume; these configurations may have a significant effect on the chemical properties of the molecule. In practise, the transitions from configuration to configuration have been modelled well as Markov chains. Moreover, these Markov chains, in general, exhibit a nearly uncoupled structure. Identification of the almost invariant aggregates of these configurations is useful in the pharmaceutical drug design.

In this context, almost invariant aggregates do not occur as artifacts of the model or its assumptions, but rather arise naturally in the processes themselves. The states spaces of configurations of biomolecules seem to possess partitions into aggregates such that when a molecule assumes a configuration in one aggregate, it will continue to be observed to do so for a relatively long period of time.

A natural explanation why this occurs has to do with energy requirements of certain molecular states. Transitions from one such aggregate to another are those which require an influx of energy to occur. For example, the molecule may need to be heated or charged in order for a certain transformation of its configuration to take place. On the other hand, transitions within an aggregate are those which require a relatively low amount of energy to occur or, possibly, involve a release of energy.

These concrete observations, however, do not impose any specific requirements on a definition of the phenomena. The π-coupling measure defined above seems to be robust enough that it can be used to evaluate potential uncouplings in a variety of applications. The first criterion that $\mathcal{E}$ is an almost invariant aggregate if $\gamma_{\mathcal{E}} \leq \epsilon \mathbb{1}$ for some small ϵ is of more theoretical interest and we use it primarily in our discussions of the algorithms themselves rather than in the analysis of their performance.

3.1. Clustered graphs. Nearly uncoupled Markov chains are very much related to a graph theoretic concept known as clustering. Roughly speaking, a *cluster* in a graph is a collection of vertices X such that

1. edges that join two vertices in X are relatively common, and
2. edges that join a vertex in X to a vertex not in X are relatively uncommon.

This concept is typically only of interest when the graph in question contains a large number of vertices. Graphs where a relatively high number of vertices are contained in clusters are referred to as *highly clustered.*

Somewhat intuitively, if a graph is highly clustered, the associated random walk is nearly uncoupled and *vice versa*—methods for constructing almost invariant aggregates of a random walk on a graph often correspond to methods for constructing clusters in that graph.

For example, the PCCA and SVD-based methods [3, 4, 6] are derived from the method of partitioning a graph's vertices using a *Fiedler* vector. The Fiedler vector of a graph is an eigenvector associated with the second smallest eigenvalue of the corresponding *Laplacian matrix*; see [2, 5]. Partitioning the vertex set of a graph into two sets X and Y so that the members of X have positive entries in the Fiedler vector and the members of Y have negative entries generally has the property that edges joining two vertices both in one of X or Y are much more common than edges with one endpoint in each of X and Y.

In [11], a method of producing clusters that utilises the convergence of the random walk is introduced—the stochastic consensus clustering algorithm (SCCA) looks for collections of states that, roughly speaking, tend to be visited in quick succession.

Clusters in large graphs or networks are of great interest in data-mining [9] and analysis of social networks [1, 12].

ETNA
Kent State University
http://etna.math.kent.edu

4. Error-reducing algorithms. Let A be a reversible stochastic matrix on the state space $\mathcal{S}$, and let $\mathcal{E} \subseteq \mathcal{S}$. We refer to the number

$$\nu_A(\mathcal{E}) = \mathbb{1}^T \gamma_{A(\mathcal{E})} = \mathbb{1}^T (I - A(\mathcal{E}))\mathbb{1} = \sum_{i \in \mathcal{E}} \sum_{j \notin \mathcal{E}} a_{ij}$$

as the *total error* of $\mathcal{E}$. If we express the matrix A as

$$A \cong \begin{bmatrix} A(\mathcal{E}) & A_{12} \\ A_{21} & A(\mathcal{S} \setminus \mathcal{E}) \end{bmatrix},$$

then the total error of $\mathcal{E}$ is simply the sum of the entries in the off-diagonal block A_{12}, that is, $\nu_A(\mathcal{E}) = \mathbb{1}^T A_{12} \mathbb{1}$. If the total error of $\mathcal{E}$ is small, then $\mathcal{E}$ is, evidently, an almost invariant aggregate. However, the converse does not seem to reasonably hold. If $\mathcal{E}$ is an almost invariant aggregate and $|\mathcal{E}|$ is relatively large, then we may have $\nu_A(\mathcal{E})$ correspondingly large.

Let $\tilde{A} = A \setminus \mathcal{C}$ be a stochastic complement derived from A. We refer to $\tilde{A}$ as *error-reducing* if for every almost invariant aggregate $\mathcal{E} \subseteq \mathcal{S}$, we have $\mathcal{E} \nsubseteq \mathcal{C}$ and

$$\nu_{\tilde{A}}(\mathcal{E} \setminus \mathcal{C}) \leq \nu_A(\mathcal{E}).$$

A stochastic complement $A \setminus \mathcal{C}$ is error-reducing as long as $\mathcal{C}$ does not contain an entire almost invariant aggregate and the removal of $\mathcal{C}$ does not increase the total error of any almost invariant aggregates.

Let B be a reversible substochastic matrix on the state space $\mathcal{S}$. Let f be a bijection from $\{1, 2, \ldots, n\}$ to the state space $\mathcal{S}$ (where $n = |\mathcal{S}|$). We refer to f as a *lower-weighted reordering* of $\mathcal{S}$ if for all $i < j$, it holds that $b_{f(j)f(i)} \leq b_{f(i)f(j)}$.

LEMMA 4.1. *Let A be a reversible stochastic matrix on the state space $\mathcal{S}$ and let Π be a positive diagonal such that ΠA is symmetric. Let f be an ordering of $\mathcal{S}$ such that*

$$\pi_{f(1)} \geq \pi_{f(2)} \geq \cdots \geq \pi_{f(n)},$$

where $n = |\mathcal{S}|$. Then, f is a lower-weighted reordering of $\mathcal{E}$.

Proof. Let A, Π, and f be as above. Since ΠA is symmetric, we have $\pi_i a_{ij} = \pi_j a_{ji}$ for all i and j. Suppose that $i < j$.

If $a_{f(i)f(j)} = 0$, then the facts $\pi_{f(i)} a_{f(i)f(j)} = \pi_j a_{f(j)f(i)}$, $\pi_{f(i)} > 0$ and $\pi_{f(j)} > 0$ imply that $a_{f(j)f(i)} = 0$. So in this case, it holds that $a_{f(i)f(j)} \geq a_{f(j)f(i)}$.

If $a_{f(i)f(j)} > 0$, then $\pi_{f(i)} a_{f(i)f(j)} = \pi_j a_{f(j)f(i)}$ implies that

$$\frac{a_{f(j)f(i)}}{a_{f(i)f(j)}} = \frac{\pi_{f(i)}}{\pi_{f(j)}} \geq 1.$$

So in either case we obtain the inequality $a_{f(i)f(j)} \geq a_{f(j)f(i)}$. ∎

LEMMA 4.2. *Let A be a reversible stochastic matrix on the state space $\mathcal{S}$. Let $\mathcal{E} \subseteq \mathcal{S}$, and let $i \notin \mathcal{E}$. Then,*

$$\nu_{A \setminus i}(\mathcal{E}) \leq \nu_A(\mathcal{E}),$$

as long as $A \setminus i$ exists.

Proof. Express

$$A \cong \begin{bmatrix} B & v & * \\ w^T & a_{ii} & * \\ * & * & * \end{bmatrix},$$

ETNA
Kent State University
http://etna.math.kent.edu

where the first position corresponds to $\mathcal{E}$, the second to i, and the third to the remainder of the state space; we have only labelled those entries that will appear in our calculations. We calculate

$$A \setminus i \cong \begin{bmatrix} B + \frac{1}{1-a_{ii}} vw^T & * \\ * & * \end{bmatrix}.$$

So,

$$\nu_{A\setminus i}(\mathcal{E}) = \mathbb{1}^T \left(I - B - \frac{1}{1-a_{ii}} vw^T \right) \mathbb{1} \leq \mathbb{1}^T (I - B)\, \mathbb{1} = \nu_A(\mathcal{E}). \qquad \square$$

LEMMA 4.3. *Let A be a reversible stochastic matrix on the state space $\mathcal{S}$. Let f be a lower-weighted reordering of A, let $\mathcal{E} \subseteq \mathcal{S}$ contain at least two states, and let*

$$i = \max \{k \leq n \,|\, f(k) \in \mathcal{E}\},$$

where $n = |\mathcal{S}|$. Then,

$$\nu_{A\setminus f(i)}(\mathcal{E} \setminus i) \leq \nu_A(\mathcal{E}).$$

Proof. Let $i_1 < \cdots < i_m$ be the indices mapped to $\mathcal{E}$ by f:

$$\mathcal{E} = \{f(i_1), \ldots, f(i_m)\}.$$

By assumption, $m \geq 2$ and

$$\max \{k \leq n \,|\, f(k) \in \mathcal{E}\} = i_m.$$

Express $\tilde{A}$ so that the first m states are $f(i_1), \ldots, f(i_m)$ (in that order). We partition this expression of $\tilde{A}$ into blocks

$$\tilde{A} \cong \begin{bmatrix} B & u & C \\ v^T & a & w^T \\ * & * & * \end{bmatrix},$$

where the first position corresponds to

$$\tilde{\mathcal{E}} = \{f(i_1), \ldots, f(i_{m-1})\}$$

the second corresponds to $f(i_m)$ (u, v, and w are column vectors), and the third to the remainder of the indices. We note that since f is a lower-weighted reordering, we have $u \leq v$ (entrywise).

We calculate

$$\tilde{A} = A \setminus f(i_m) \cong \begin{bmatrix} B + \frac{1}{1-a} uv^T & C + \frac{1}{1-a} uw^T \\ * & * \end{bmatrix}.$$

Now we have

$$\nu_A(\mathcal{E}) = \mathbb{1}^T C \mathbb{1} + w^T \mathbb{1}.$$

ETNA
Kent State University
http://etna.math.kent.edu

Then, the total error of $\tilde{\mathcal{E}}$ (with respect to $\tilde{A}$) satisfies

$$\begin{aligned}\nu_{\tilde{A}}(\tilde{\mathcal{E}}) &= \mathbb{1}^T C \mathbb{1} + \frac{1}{1-a}\mathbb{1}^T u w^T \mathbb{1} \leq \mathbb{1}^T C \mathbb{1} + \frac{1}{1-a}\mathbb{1}^T v w^T \mathbb{1} \\ &\leq \mathbb{1}^T C \mathbb{1} + \frac{1}{1-a}(1-a) w^T \mathbb{1} = \mathbb{1}^T C \mathbb{1} + w^T \mathbb{1} = \nu_A(\mathcal{E}).\end{aligned}$$

The first inequality is due to the fact that $u \leq v$ (entrywise) and the second is due to the stochastic property: since $v^T \mathbb{1} + a + w^T \mathbb{1} = 1$, we have $v^T \mathbb{1} = 1 - a - w^T \mathbb{1} \leq 1 - a$. ∎

We will utilise Algorithm 1 in our stochastic complement based algorithms. The inputs of the Choose-algorithm are a reversible stochastic matrix A and an index i; the output is either an index $j \neq i$ such that

1. $0 < a_{ji} \leq a_{ij}$ and
2. $a_{ik} \leq a_{ij}$, for all $k \neq i$,

or the value 0 if no such j exists. Let A be reversible with stationary distribution π, and let Choose$(A, i) = j \neq 0$, then the fact that $\pi_i a_{ij} = \pi_j a_{ji}$ implies that $\pi_i \leq \pi_j$.

We now present our first stochastic complement based algorithm, the Perron-ordered algorithm. The inputs of Algorithm 2 are a reversible stochastic matrix A that has been reordered via its stationary distribution and a small positive value ϵ. That is, the algorithm implicitly assumes that if $\tilde{A} = A \setminus \mathcal{C}$ is a stochastic complement formed from A and if $i < j$ are not contained in $\mathcal{C}$, then $\tilde{a}_{ji} \leq \tilde{a}_{ij}$.

We note that reordering A via some output of Algorithm 3 is not sufficient to ensure the above property. The state space of A needs to be ordered via its stationary distribution—if a stationary distribution is not known or calculable, Algorithm 4 should be used instead.

The output of the Perron-ordered algorithm, Algorithm 2, is a partition of the state space $\mathcal{S}$ into candidate aggregates. We claim that when the value ϵ is well-chosen, the output aggregates are very good candidates for almost invariant aggregates.

Let A be a reversible stochastic matrix on the state space $\mathcal{S}$, let $n = |\mathcal{S}|$, and let $0 < \epsilon < 1$. For each $s \geq 1$,

1. let i_s be the state removed during the sth iteration of the while loop,
2. let

$$\mathcal{C}_s = \{i_1, i_2, \ldots, i_s\}$$

be the collection of states removed after s iterations,

3. and let $j_s \notin \mathcal{C}_s$ be the state identified by the Choose algorithm, i.e.,

$$j_s = \text{Choose}(A \setminus \mathcal{C}_{s-1}, i_s),$$

One can show, inductively, that after s iterations

1. the current stored list of aggregates,

$$\{\mathcal{E}_i \,|\, i \notin \mathcal{C}_s\},$$

is a partition of $\mathcal{S}$,

2. the stored value of m_i is the number of states contained in $\mathcal{E}_i$, and
3. for any stationary distribution π of A and any $i \notin \mathcal{C}_s$,

$$\pi_i = \max_{k \in \mathcal{E}_i}\{\pi_k\}.$$

ETNA
Kent State University
http://etna.math.kent.edu

Algorithm 1 Choose(A, i).

Select an index $j \neq i$ such that a_{ij} is maximal among the off-diagonal entries in the ith row of A; if this maximal value is attained more than once, choose such a j that has minimal a_{ji}.
if $a_{ij} < a_{ji}$ or $a_{ij} = 0$ **then**
  return 0
else
  return j
end if

Algorithm 2 The Perron-ordered algorithm.

$B := A$
for all $i \in \mathcal{S}$ **do**
  $\mathcal{E}_i := \{i\}$
  $m_i := 1$
end for
$K := \{i \,|\, b_{ii} < 1 - \epsilon\}$
while K is nonempty **do**
  Let $i \in K$ be the member of K closest to $|\mathcal{S}|$ that has Choose$(B, i) \neq 0$:

$$i := \max \{k \in K \,|\, \text{Choose}(B, k) \neq 0\} .$$

  if No such $i \in K$ exists **then**
    Exit the **while** loop.
  else
    $j :=$ Choose(B, i)
    $\mathcal{E}_j := \mathcal{E}_j \cup \mathcal{E}_i$
    $m_j := m_j + m_i$
    Delete $\mathcal{E}_i$ from storage.
    $B := B \setminus i$
    $K := K \setminus i$
    for all $k \in K$ **do**
      if $b_{kk} \geq \frac{(1-\epsilon)^2}{1+(m_k-2)\epsilon}$ **then**
        $K := K \setminus k$
      end if
    end for
  end if
end while
return $\{\mathcal{E}_k\}$

Now, let $s \geq 1$ and suppose that

1. there is a partition of the state space $\mathcal{S}$ into almost invariant aggregates,
2. the stochastic complement $A \setminus \mathcal{C}_s$ is error-reducing, and
3. for $t = 1, \ldots, s$, each pair of states i_t and j_t is contained in a single member of the partition into almost invariant aggregates.

Suppose that the algorithm does not terminate after the sth iteration and consider the pair of states $i = i_{s+1}$ and $j = j_{s+1}$. Let $\mathcal{E}$ be the almost invariant aggregate containing i.

First, suppose that every member of $\mathcal{E}$ aside from i is contained in $\mathcal{C}_s$. Since every pair (i_t, j_t) is contained in a single almost invariant aggregate, every member of the stored partition $\{\mathcal{E}_i \,|\, i \notin \mathcal{C}_s\}$ contains states from exactly one almost invariant aggregate, hence it

ETNA
Kent State University
http://etna.math.kent.edu

Algorithm 3 Reorder(B).

```
Let n be the order of B.
f := (1, 2, ..., n)
r := 1
s := 2
while s ≤ n do
  if b_{f(r)f(t)} ≤ b_{f(t)f(r)} for t = s, ..., n then
    r := r + 1
    if r = s then
      s := s + 1
    end if
  else
    Let t be such that s ≤ t ≤ n and b_{f(r)f(t)} > b_{f(t)f(r)}.
    (f(r), f(r+1), ..., f(t)) := (f(t), f(r), f(r+1), ..., f(t-1))
    s := s + 1
  end if
end while
return f
```

holds that $\mathcal{E} = \mathcal{E}_i$ and $m_i = |\mathcal{E}|$. However, Proposition A.10 then implies that

$$b_{ii} \geq \frac{(1-\epsilon)^2}{1+(m_i-2)\epsilon},$$

where $B = A \setminus \mathcal{C}_s$. (This is the result of an extensive calculation found in the appendix.) This is a contradiction because when the state i is removed, we have

$$b_{ii} < \frac{(1-\epsilon)^2}{1+(m_i-2)\epsilon}.$$

So, we suspect that the state i is contained in some almost invariant aggregate $\mathcal{E}$ such that $\mathcal{E} \setminus \mathcal{C}_{s+1}$ is nonempty. The stochastic complement $B = A \setminus \mathcal{C}_s$ is error-reducing and the entry b_{ij} is larger than any other b_{ik}—this suggests that j is contained in this same almost invariant aggregate. Then, when we form the stochastic complement $A \setminus \mathcal{C}_{s+1}$, the fact that π_i is minimal (among all states satisfying the above) together with Lemmas 4.1, 4.2, and 4.3, suggests that this new stochastic complement is error-reducing as well.

Thus, inductively, every correct association made (every union $\mathcal{E}_j \cup \mathcal{E}_i$ where i and j are members of the same almost invariant aggregate) seems to make further correct associations more likely. The algorithm terminates when it cannot locate any further states i that seem to be "safe" to remove.

Next, we present the lower-weighted algorithm (Algorithm 4), intended for use on reversible matrices A for which the stationary distribution π is unknown. We recommend using this process rather than simply calculating a stationary distribution for the following reason. When the stochastic matrix A is nearly uncoupled, there is a potentially large number of eigenvalues near to the eigenvalue 1. Thus, the eigenvector problem $x^T = x^T A$ is badly conditioned—computed solutions to this problem are very sensitive to perturbation and roundoff error making them unreliable as actual eigenvectors. The lower-weighted algorithm bypasses this problem by proceeding without any knowledge of the eigenvectors of A.

The lower-weighted algorithm utilises the reorder algorithm, Algorithm 3, to construct lower-weighted reorderings of the state space; the input of Algorithm 3 is a reversible stochas-

ETNA
Kent State University
http://etna.math.kent.edu

Algorithm 4 The lower-weighted algorithm.

$B := A$
for all $i \in \mathcal{S}$ **do**
  $\mathcal{E}_i := \{i\}$
  $m_i := 1$
end for
$f :=$ Reorder(B)
$K := \{i \,|\, b_{ii} < 1 - \epsilon\}$
while K is nonempty **do**
  if Choose$(B, i) = 0$ for all $i \in K$ **then**
    Exit the **while** loop.
  else
    Let k be the largest index such that $f(k) \in K$ and Choose$(B, f(k)) \neq 0$
    $i := f(k)$
    $j :=$ Choose(B, i)
    $\mathcal{E}_j = \mathcal{E}_j \cup \mathcal{E}_i$
    $m_j := m_j + m_i$
    Delete $\mathcal{E}_i$ from storage.
    $B := B \setminus i$
    $f :=$ Reorder(B)
  end if
end while
return $\{\mathcal{E}_k\}$

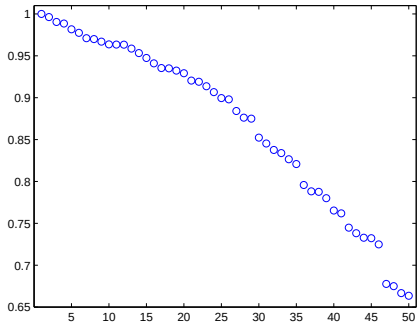


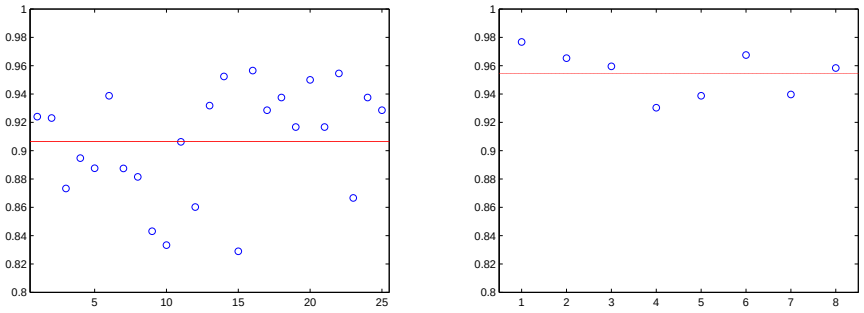
FIG. 5.1. *The* 50 *eigenvalues of the collaboration network nearest to* 1.

tic matrix and the output is a lower-weighted reordering f. The reorder algorithm is a modified version of a depth-first search. Within Algorithm 3, the function f is expressed as its range ordered by its domain, i.e., the function $f = (2, 3, 1)$ has $f(1) = 2$, $f(2) = 3$, and $f(3) = 1$.

The lower-weighted algorithm proceeds in much the same manner as the Perron-ordered algorithm; it simply recalculates a lower-weighted reordering at every step in order to ensure that the complements formed are error-reducing. (The reordering needs to be recalculated after every stochastic complement—if the matrix A is lower-weighted, it is not necessarily true that $A \setminus i$ is lower-weighted.)

5. Examples. First, we apply our Perron-ordered algorithm to a collaboration network, previously presented in [12]. The network represents published collaborations between a

ETNA
Kent State University
http://etna.math.kent.edu



(a) $\epsilon = 0.05$ (b) $\epsilon = 0.01$

FIG. 5.2. *Algorithm 2 applied to a collaboration network.*

group of researchers. Each node in the network is a given researcher; the edges and their weights are determined by papers published. If a paper has k authors, it contributes a weight of $1/\binom{k}{2}$ to every edge joining pairs of these k authors. (Each published work contributes a total weight of 1 to the edges of the network.) The full data set includes 1589 authors. However, the full network is not connected. We will analyse only the largest connected component (as in [12]). The remainder of this section refers only to this largest connected component, which consists of 379 researchers.

This collaboration network is interesting because it highly clustered. This is reflected in its spectrum—the transition matrix of the random walk on this network has a large number of eigenvalues near 1 and no clear gap separating the eigenvalues near 1 from the remainder. We present the 50 eigenvalues of the transition matrix nearest to 1 in Figure 5.1.

We apply our Perron-ordered algorithm to the transition matrix of the random walk on this network in order to detect "clusters"—groups of nodes that are well-connected to each other and poorly connected to the remainder of the nodes. We apply it twice, once each with the inputs $\epsilon = 0.05$ and $\epsilon = 0.01$. As the random walk is derived from a known network, the stationary distribution of the random walk is known—it is a scalar multiple of the vector of the degrees of the nodes. Thus, we evaluate our output using the π-coupling measure,

$$w_\pi(\mathcal{E}) = \frac{\pi(\mathcal{E})^T(I - A(\mathcal{E}))\mathbb{1}}{\pi(\mathcal{E})^T\mathbb{1}} = \frac{\pi(\mathcal{E})^T\gamma_{A(\mathcal{E})}}{\pi(\mathcal{E})^T\mathbb{1}} .$$

(As well, knowledge of the stationary distribution is required to apply the Perron-ordered algorithm.) We find that applying Algorithm 2 with inputs $\epsilon = 0.05$ and $\epsilon = 0.01$ results in partitions of the state space into 25 and 8 aggregates, respectively. The π-coupling measures of these aggregates together with the mean π-coupling measures (drawn as a red line) are displayed in Figure 5.2.

We present the network in Figure 5.3. The partition obtained with $\epsilon = 0.01$ is represented by the colouring of the vertices. The layout of the vertices has not been influenced by the partition. We graphed this network with the software package PAJEK*—the software attempts (as much as possible) to arrange the vertices so that the edges are drawn with short lengths and so that few edges cross. The partition produced by the Perron-ordered algorithm coincides well with this layout.

This collaboration network is analysed in [12]. There, it is used to illustrate a concept known as modularity. The modularity of a vertex is a measure of how "central" that vertex is within the network. It was found that ten vertices have relatively high modularity.

*http://vlado.fmf.uni-lj.si/pub/networks/pajek/

ETNA
Kent State University
http://etna.math.kent.edu

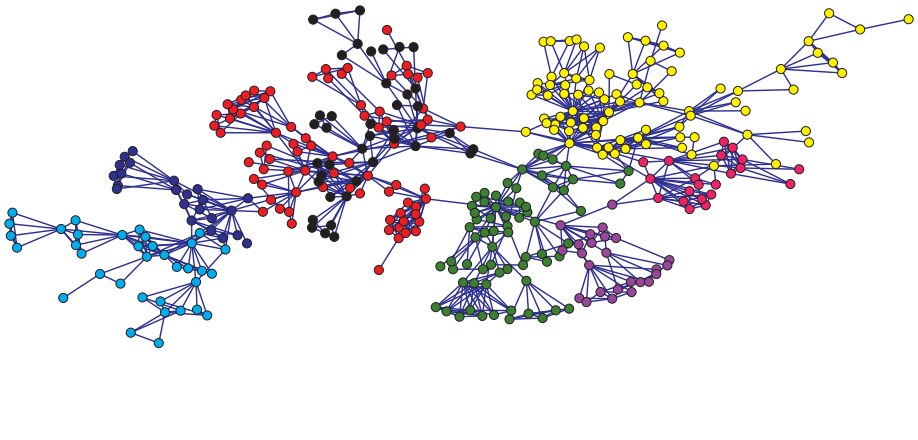


FIG. 5.3. *The largest connected component of the collaboration network with a partition into almost invariant aggregates.*

Next, we briefly summarise some of our results concerning a simulation of the molecular transitions n-pentane. We provide this example in order to compare the performance of our algorithms with that of competing methods. The stochastic matrix in question was obtained from a Markov chain Monte Carlo simulation of the molecular states of the n-pentane molecule; see [3, 13, 14] for details.

The matrix in question is a 255×255 reversible stochastic matrix with known stationary distribution. It is previously analysed in [3, 6].

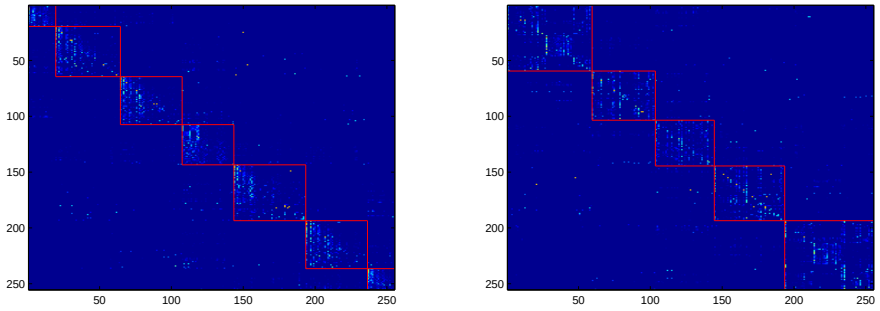
The Perron cluster analysis technique decides in advance how many aggregates will be formed. The number is chosen by selecting a cluster of eigenvalues of the matrix near to 1 (the Perron cluster)—the number of such eigenvalues is the number of aggregates the algorithm will produce. In [3], it is noted that the eight eigenvalues of the n-pentane transition matrix nearest to 1 are (approximately) 1, 0.986, 0.984, 0.982, 0.975, 0.941, 0.938, and 0.599. In addition to the eigenvalue 1, there are four eigenvalues approximately equal to 0.98 and a further two approximately equal to 0.94 with the remainder being not close to 1. This suggests either 5 or 7 aggregates. The PCCA algorithm is applied twice to produce partitions into 5 and 7 aggregates.

In [6], a similar approach that uses singular vectors rather than eigenvectors is applied to the same matrix. The SVD-based approach is applied only once to produce a partition into 7 aggregates.

We find that applications of the lower-weighted and Perron-ordered algorithms with inputs of $\epsilon = 0.01$ and 0.005 produce partitions into 7 and 5 aggregates, respectively. We report the *minimum* π-coupling measures of the output partitions in Table 5.1. That is, every member of each partition produced has a π-coupling measure greater than or equal to the given value. The minimum π-coupling measures of the outputs of our stochastic complement-based algorithms in each case exceed those of the competing Perron cluster and SVD-based methods, at least slightly. The transition matrix is displayed in Figure 5.4—the results obtained here compare very favourably with those in [3, 6].

6. Conclusion. The algorithms presented here are an efficient manner to construct almost invariant aggregates of a given stochastic matrix. Each of the iterative steps applied

ETNA
Kent State University
http://etna.math.kent.edu



(a) *The lower-weighted algorithm with* $\epsilon = 0.01$. (b) *The Perron-ordered algorithm with* $\epsilon = 0.005$.

FIG. 5.4. *Decouplings of the* n*-pentane transition matrix obtained via the stochastic complement based algorithms.*

TABLE 5.1
Stochastic complement based and other algorithms applied to the n*-pentane transition matrix.*

Algorithm	ϵ	Aggregates	Min. π-coupling measure
Lower-weighted	0.01	7	0.921
	0.005	5	0.979
Perron-ordered	0.01	7	0.920
	0.005	5	0.979
PCCA	n/a	7	0.918
	n/a	5	0.976
SVD-based	n/a	7	0.876

to a matrix of order n require of the order of n^2 floating point operations. Thus, the algorithms operate in polynomial time of order n^3 with respect to the order of the input matrices. The storage requirements are minimal—the amount of additional storage necessary (assuming the input matrix is already in storage) is simply the amount required to store a second copy of the input matrix. This is in contrast to the eigenvector and singular vector based approaches [3, 4, 6, 13], which are actually convergent algorithms rather than deterministic ones.

This efficiency allows the stochastic complement based algorithms to be run many times at little cost with varying values for the input ϵ. In addition to allowing the selection of an optimal output, such a process may be of utility in analysing the decoupled properties of the Markov chain more thoroughly. These multiple output partitions can be combined to form hierarchical decompositions of the state space. For example, the 25 aggregates formed by applying the Perron-ordered algorithm to the collaboration network with $\epsilon = 0.05$ form a refinement of the aggregates obtained with $\epsilon = 0.01$.

An advantage of this approach is its independence of spectral methods. When a stochastic matrix A is nearly uncoupled, it possesses multiple eigenvalues near to 1 (its Perron cluster). If this Perron cluster is sufficiently large, the calculation of the eigenvectors associated with these eigenvalues is somewhat unreliable—these calculations are very sensitive to floating point round-off errors, for example. Moreover, spectral methods tend to rely upon heuristic or convergent methods, especially in their calculation of eigenvectors and singular vectors. The algorithms presented here terminate after a finite and well-bounded number of steps, producing an output partition uniquely determined by the input matrix.

ETNA
Kent State University
http://etna.math.kent.edu

Appendix A. A lower bound concerning stochastic complements of reversible substochastic matrices.

We construct a lower bound on a specific term relating to stochastic complements of reversible substochastic matrices. This lower bound is used in Algorithms 2 and 4 to determine whether or not a given state is "safe" to be removed via a stochastic complement.

A.1. Definitions and problem statement.

DEFINITION A.1. *Let B be a properly substochastic matrix, and let $\mathcal{C}$ be the associated state space. If the order of B is 1, that is, if $\mathcal{C} = \{i\}$ and $B = [b_{ii}]$, we define $\alpha_B(i) = b_{ii}$. If $\mathcal{C}$ contains two or more states, then for each $i \in \mathcal{C}$, we express*

$$B \cong \begin{bmatrix} b_{ii} & v^T \\ w & C \end{bmatrix},$$

and define

$$\alpha_B(i) = b_{ii} + v^T(I - C)^{-1}w.$$

We note that given B and $\mathcal{C}$ as above,

$$[\alpha_B(i)] = B \setminus \{j \in \mathcal{C} \,|\, j \neq i\},$$

if we define $B \setminus \emptyset = B$.

Let X be a Markov chain on the state space $\mathcal{S}$. For $\mathcal{C} \subseteq \mathcal{S}$, we define

$$E_{\mathcal{C}} = \inf_{t \geq 1} \{t : x_t \notin \mathcal{C}\}.$$

If $x_0 \in \mathcal{C}$, we refer to $t = E_{\mathcal{C}}$ as the *first exit time* out of $\mathcal{C}$, and we say that the Markov chain exits $\mathcal{C}$ at time t. As well, for each $i \in \mathcal{S}$,

$$T_i = \inf_{t \geq 1} \{t : x_t = i\}$$

is the *first passage time* into i.

Let A be a reversible stochastic matrix on the state space $\mathcal{S}$, let $\mathcal{C} \subseteq \mathcal{S}$, $B = A(\mathcal{C})$, and suppose further that B is irreducible. Via Proposition 2.2, it can be shown that for any $i \in C$, the number $\alpha_B(i)$ is the probability of transitioning from i to i (in one or more steps) without first exiting $\mathcal{C}$. That is,

$$\alpha_B(i) = \mathbb{P}\left[T_i < E_{\mathcal{C}} \,|\, x_0 = i\right].$$

Let B be an irreducible reversible substochastic matrix and suppose that Π and Π' are positive diagonal matrices such that ΠB and $\Pi B'$ are symmetric. For any $i \neq j$ with $b_{ij} \neq 0$, we have

$$\pi_i b_{ij} = \pi_j b_{ji} \text{ and } \pi'_i b_{ij} = \pi'_j b_{ji},$$

further implying that

$$\frac{b_{ji}}{b_{ij}} = \frac{\pi_i}{\pi_j} = \frac{\pi'_i}{\pi'_j}.$$

So when $b_{ij} \neq 0$, the ratios $\pi_i/\pi_j = \pi'_i/\pi'_j$ are uniquely determined by B. Since B is irreducible, this implies that every ratio $\pi_i/\pi_j = \pi'_i/\pi'_j$ is uniquely determined by B.

ETNA
Kent State University
http://etna.math.kent.edu

The above reasoning implies that for any such matrix B, the matrices Π that symmetrise B via a left-multiplication are uniquely determined up to a multiplication by a positive scalar. In light of this, we will define a canonical positive diagonal for each such B. Let B be an irreducible reversible substochastic matrix. We define $\Pi = \Pi_B$ to be the unique positive diagonal such that ΠB is symmetric and such that the largest entry in Π is 1. That is, $\Pi = \Pi_B$ satisfies

$$\pi_i b_{ij} = \pi_j b_{ji} \text{ and } \max_{i \in \mathcal{C}} \{\pi_i\} = 1,$$

where $\mathcal{C}$ is the state space of B.

DEFINITION A.2. *Let $n \geq 1$ be a positive integer and $\epsilon < 1$ be a positive real number. We define $\mathcal{B}(n, \epsilon) = \{B\}$ to be the collection of $n \times n$ substochastic matrices B such that*

1. *B is irreducible and reversible and*
2. *$\gamma_B = (I - B)\mathbb{1} \leq \epsilon \mathbb{1}$.*

We note that for all $B \in \mathcal{B}(n, \epsilon)$, it holds that $B\mathbb{1} \geq (1 - \epsilon)\mathbb{1}$.

DEFINITION A.3. *Let $n \geq 1$, and let $\epsilon < 1$ be a positive real number. Let $B \in \mathcal{B}(n, \epsilon)$, and let $\Pi = \Pi_B$. We define $\alpha(B)$ to be the minimum value of $\alpha_B(i)$ subject to $\pi_i = 1$,*

$$\alpha(B) = \min_{i \ni \pi_i = 1} \{\alpha_B(i)\}.$$

We note that if $B \in \mathcal{B}(n, \epsilon)$ and $\Pi = \Pi_B$, then for every index i of B, either $\pi_i < 1$ or $\alpha_B(i) \geq \alpha(B)$.

The problem we solve is the following. Given a positive integer $n \geq 2$ and a positive real number $\epsilon < 1$, we calculate the number

$$\alpha(\mathcal{B}) = \inf_{B \in \mathcal{B}} \{\alpha(B)\}$$

and characterise those reversible substochastic matrices $B \in \mathcal{B}$ that have $\alpha(B) = \alpha(\mathcal{B})$.

A.2. Preliminaries.

LEMMA A.4. *Let $B \in \mathcal{B}(n, \epsilon)$ where $n \geq 2$. Then we can express*

$$B \cong \begin{bmatrix} a & v^T \\ w & C \end{bmatrix},$$

where, in addition to the fact that B is irreducible and substochastic, it holds that

1. *$a + v^T \mathbb{1} \geq 1 - \epsilon$,*
2. *$C\mathbb{1} + w \geq (1 - \epsilon)\mathbb{1}$,*
3. *$\alpha(B) = a + v^T (I - C)^{-1} w$, and*
4. *there is a positive diagonal matrix Q, such that $Q \leq I$, $QC = C^T Q$, and $Qw = v$.*

Proof. Let $\Pi = \Pi_B$. Since

$$\alpha(B) = \min_{i \ni \pi_i = 1} \{\alpha_B(i)\},$$

there is an index i such that $\pi_i = 1$ and $\alpha(B) = \alpha_B(i)$. Express

$$B \cong \begin{bmatrix} b_{ii} & v^T \\ w & C \end{bmatrix},$$

ETNA
Kent State University
http://etna.math.kent.edu

where the first row and column corresponds to such a state i and the principal submatrix C corresponds to the remainder of the state space. The first two claims are direct consequences of the fact that

$$\gamma_B = (I - B)\mathbb{1} \leq \epsilon\mathbb{1}.$$

The third claim is simply a restatement of the fact that $\alpha_B(i) = \alpha(B)$.

Finally, since $\pi_i = 1$ and $\pi_j \leq 1$ for all $j \in \mathcal{C}$, we have

$$\Pi_B \cong \begin{bmatrix} 1 & 0 \\ 0 & Q \end{bmatrix},$$

where $Q \leq I$ via the same correspondence as B. The fourth claim is a consequence of the fact that $\Pi_B B = B^T \Pi_B$. □

The following lemma combines a few well-known results from the theory of M-matrices; see [8, Section 2.5].

LEMMA A.5. *Let X, Y, and Z be nonnegative square matrices of order $m \geq 1$ such that*

1. *Z is irreducible,*
2. *X and Y are positive diagonal matrices,*
3. *$X \leq Y$ with a strict inequality in at least one diagonal entry, and*
4. *$Z\mathbb{1} \leq X\mathbb{1}$ with a strict inequality in at least one position.*

Then, the matrices $(X - Z)^{-1}$ and $(Y - Z)^{-1}$ are defined and satisfy (entrywise)

$$0 < (Y - Z)^{-1} < (X - Z)^{-1}.$$

A real matrix X is *positive definite* if it is symmetric and every eigenvalue of X is positive. We note that real positive definite matrices are nonsingular.

LEMMA A.6. *Let X be a real positive definite matrix, and let v be a nonzero real vector. Then,*

$$\left(v^T X v\right)\left(v^T X^{-1} v\right) \geq \left(v^T v\right)^2 = \|v\|^4,$$

with equality if and only if v is an eigenvector of X.

Proof. We make use of some well-known facts from linear algebra.

First, the Cauchy-Schwarz inequality (as it applies to real spaces of column vectors) is the following proposition: let v and w be nonzero real column vectors, then

$$v^T w \leq \|v\|\,\|w\| = (v^T v)^{1/2}(w^T w)^{1/2},$$

with equality if and only if $v = \beta w$ for some nonzero real number β.

Second, we make use of the following propositions taken from [7, Chapter 7]: let X be a real positive definite matrix, then

1. there is a unique real positive definite matrix labelled $X^{1/2}$ and referred to as the square root of X such that
$$(X^{1/2})^2 = X,$$
2. the matrix X^{-1} is itself real and positive definite, and
3. the square root of X^{-1} is the inverse of $X^{1/2}$,
$$(X^{-1})^{1/2} = (X^{1/2})^{-1},$$
and we label this matrix $X^{-1/2}$.

ETNA
Kent State University
http://etna.math.kent.edu

(We have modified the results in [7] slightly as we are only interested in the real case).

Now, let X be a real positive definite matrix and v be a nonzero real vector. Then,

$$v^T v = v^T X^{1/2} X^{-1/2} v = (X^{1/2}v)^T (X^{-1/2}v).$$

So, via the Cauchy-Schwarz inequality,

$$v^T v \leq \left\| X^{1/2}v \right\| \left\| X^{-1/2}v \right\| = (v^T X v)^{1/2} (v^T X^{-1} v)^{1/2}.$$

Squaring every term in this expression yields the expression in the statement above. Furthermore, we note that equality holds if and only if

$$X^{1/2}v = \beta X^{-1/2}v,$$

for some real number β. When this occurs, a left-multiplication by $X^{1/2}$ leads to $Xv = \beta v$. ∎

Let $B \in \mathcal{B}(n,\epsilon)$, and let $\Pi = \Pi_B$. We note that $\Pi_B = I$ if and only if B is symmetric. If B is symmetric, then

$$\alpha(B) = \min_{i \ni \pi_i = 1} \{\alpha_B(i)\} = \min\{\alpha_B(i)\},$$

and we have $\alpha_B(i) \geq \alpha(B)$ for all i. As well, if B is symmetric, the expression of B found in Lemma A.4 is

$$B \cong \begin{bmatrix} a & v^T \\ v & C \end{bmatrix},$$

where C is symmetric and $\alpha(B) = a + v^T(I - C)^{-1}v$.

LEMMA A.7. *Let $B \in \mathcal{B}(n,\epsilon)$. If B is not symmetric, then there is a symmetric substochastic matrix $\hat{B} \in \mathcal{B}(n,\epsilon)$ such that $\alpha(\hat{B}) < \alpha(B)$.*

Proof. Suppose that $B \in \mathcal{B}(n,\epsilon)$ is not symmetric. Express

$$B \cong \begin{bmatrix} a & v^T \\ w & C \end{bmatrix} \text{ and } \Pi = \Pi_B \cong \begin{bmatrix} 1 & 0 \\ 0 & Q \end{bmatrix}$$

as in Lemma A.4. Thus, $Q \leq I$, $Qw = v$, and QC is symmetric. The assumption that B is not symmetric implies that $Q \neq I$. We note that since $B \in \mathcal{B}(n,\epsilon)$, we have

1. B is irreducible,
2. $1 - \epsilon \leq a + v^T \mathbb{1} \leq 1$, and
3. $(1-\epsilon)\mathbb{1} \leq C\mathbb{1} + w \leq \mathbb{1}$.

Let

$$\hat{B} = \begin{bmatrix} a & v^T \\ v & \hat{C} \end{bmatrix},$$

where

$$\hat{C} = QC + (1-\epsilon)(I - Q).$$

We claim that $\hat{B}$ is a symmetric member of $\mathcal{B}(n,\epsilon)$ and $\alpha(\hat{B}) < \alpha(B)$. Since $0 \leq Q \leq I$, $\hat{B}$ is nonnegative. For every $i \neq j$, we have $\hat{b}_{ij} = \pi_i b_{ij}$, so the fact that B is irreducible implies that $\hat{B}$ is irreducible. As well, the fact that QC is symmetric implies that $\hat{B}$ is symmetric. So we next need to show that $\hat{B}$ is substochastic and $\gamma_{\hat{B}} \leq \epsilon\mathbb{1}$.

ETNA
Kent State University
http://etna.math.kent.edu

By assumption, we have $1-\epsilon \leq a + v^T\mathbb{1} \leq 1$. Next, the inequality $C\mathbb{1} + w \geq (1-\epsilon)\mathbb{1}$ implies that

$$\begin{aligned}\hat{C}\mathbb{1} + v &= (QC + (1-\epsilon)(I-Q))\,\mathbb{1} + Qw = Q(C\mathbb{1}+w) + (1-\epsilon)(I-Q)\mathbb{1} \\ &\geq Q\left((1-\epsilon)\mathbb{1}\right) + (1-\epsilon)(I-Q)\mathbb{1} = (1-\epsilon)\mathbb{1}.\end{aligned}$$

As well, the inequalities $C\mathbb{1} + w \leq \mathbb{1}$ and $0 \leq Q \leq I$ imply that

$$\begin{aligned}\hat{C}\mathbb{1} + v &= Q(C\mathbb{1}+w) + (1-\epsilon)(I-Q)\mathbb{1} \leq Q\mathbb{1} + (1-\epsilon)(I-Q)\mathbb{1} \\ &= (1-\epsilon)\mathbb{1} + \epsilon Q\mathbb{1} \leq (1-\epsilon)\mathbb{1} + \epsilon\mathbb{1} = \mathbb{1}.\end{aligned}$$

Hence, we obtain that $(1-\epsilon)\mathbb{1} \leq \hat{B}\mathbb{1} \leq \mathbb{1}$. Thus, $\hat{B}$ is a symmetric member of $\mathcal{B}(n,\epsilon)$. We now show that $\alpha(\hat{B}) < \alpha(B)$.

Since $\hat{B}$ is symmetric, we have $\Pi_{\hat{B}} = I$, which implies $\alpha_{\hat{B}}(1) \geq \alpha(\hat{B})$. We note that $Qw = v$, thus we obtain $w = Q^{-1}v$. We calculate

$$\alpha(B) = a + v^T\left(I-C\right)^{-1}w = a + v^T\left(I-C\right)^{-1}Q^{-1}v = a + v^T\left(Q-QC\right)^{-1}v$$

and

$$\begin{aligned}\alpha_{\hat{B}}(1) &= a + v^T\left(I-\hat{C}\right)^{-1}v = a + v^T\left(I-(QC+(1-\epsilon)(I-Q))\right)^{-1}v \\ &= a + v^T\left(Q+\epsilon(I-Q)-QC\right)^{-1}v.\end{aligned}$$

Permute the indices (if necessary) so that

$$C \cong \begin{bmatrix} C_1 & & 0 \\ & \ddots & \\ 0 & & C_l \end{bmatrix},\ Q \cong \begin{bmatrix} Q_1 & & 0 \\ & \ddots & \\ 0 & & Q_l \end{bmatrix},\ \text{and}\ v \cong \begin{bmatrix} v_1 \\ \vdots \\ v_l \end{bmatrix},$$

where each C_k is irreducible. We expand our formulae above for $\alpha(B)$ and $\alpha_{\hat{B}}(1)$:

$$\alpha(B) = a + \sum_{k=1}^{l} v_k^T(Q_k - Q_kC_k)^{-1}v_k$$

and

$$\alpha_{\hat{B}}(1) = a + \sum_{k=1}^{l} v_k^T\left(Q_k + \epsilon(I-Q_k) - Q_kC_k\right)^{-1}v_k.$$

If $Q_k = I$, the kth terms from the two sums are equal. If $Q_k \neq I$, we apply Lemma A.5 with $X = Q_k$, $Y = Q_k + \epsilon(I-Q_k)$, and $Z = Q_kC_k$ to see that entrywise

$$0 < \left(Q_k + \epsilon(I-Q_k) - Q_kC_k\right)^{-1} < (Q_k - Q_kC_k)^{-1}.$$

Since B is irreducible, every v_k has at least one positive term. Thus, if $Q_k \neq I$, it holds that

$$v_k^T\left(Q_k + \epsilon(I-Q_k) - Q_kC_k\right)^{-1}v_k < v_k^T(Q_k - Q_kC_k)^{-1}v_k.$$

Since $Q \neq I$, there is at least one $Q_k \neq I$, and so we arrive at

$$\alpha(\hat{B}) \leq \alpha_{\hat{B}}(1) < \alpha(B). \qquad \square$$

LEMMA A.8. *Let $B \in \mathcal{B}(n,\epsilon)$ be symmetric. Suppose that there is an index i such that $\alpha_B(i) = \alpha(B)$ and $\gamma_B(i) < \epsilon$. Then, there is also a symmetric substochastic matrix $\hat{B} \in \mathcal{B}(n,\epsilon)$ such that*

ETNA
Kent State University
http://etna.math.kent.edu

1. $\alpha(B) > \alpha(\hat{B})$,
2. $\gamma_{\hat{B}}(i) = \epsilon$, *and*
3. *for all* $j \neq i$, $\gamma_{\hat{B}}(j) = \gamma_B(j)$.

Proof. Let the state i be such that $\alpha_B(i) = \alpha(B)$ and $\gamma_B(i) < \epsilon$. Without loss of generality, we assume that $i = 1$. By assumption, B is symmetric. Via Lemma A.4, we express

$$B = \begin{bmatrix} a & v^T \\ v & C \end{bmatrix},$$

where C is symmetric and $\alpha(B) = \alpha_B(1) = a + v^T(I - C)^{-1}v$. Now, we have

$$\gamma_B(1) = 1 - a - v^T\mathbb{1},$$

hence, $\gamma_B(1) < \epsilon$ implies that $a + v^T\mathbb{1} > 1 - \epsilon$.

First, suppose that $v^T\mathbb{1} \leq 1 - \epsilon$, then we have $a > 1 - \epsilon - v^T\mathbb{1} \geq 0$. Let

$$\hat{B} = \begin{bmatrix} 1 - \epsilon - v^T\mathbb{1} & v^T \\ v & C \end{bmatrix}.$$

We have

$$\alpha(\hat{B}) \leq \alpha_{\hat{B}}(1) = 1 - \epsilon - v^T\mathbb{1} + v^T(I - C)^{-1}v < a + v^T(I - C)^{-1}v = \alpha(B).$$

So, we next assume that $v^T\mathbb{1} > 1 - \epsilon$. Let R be the diagonal matrix with $r_i = v(i)$, thus $R\mathbb{1} = v$. For real numbers z with $0 \leq z < 1$, let $v(z) = (1-z)v$, and let $C(z) = C + zR$. As long as $z < 1$, the matrix $C(z)$ is properly substochastic, so $(1 - C(z))^{-1}$ is nonnegative. We will first show that the function

$$f(z) = v(z)^T(I - C(z))^{-1}v(z) = (1 - z)^2 v^T(I - C(z))^{-1}v$$

is strictly decreasing in z over the interval $z \in [0, 1)$. We note that

$$\frac{d}{dz}C(z) = R.$$

We will use the fact that when $z \mapsto Y$ is an entrywise differentiable function $\mathbb{R} \mapsto \mathbb{R}^{n\times n}$, the derivative of Y^{-1} is given by

$$\frac{d}{dz}Y^{-1} = -Y^{-1}\left(\frac{d}{dz}Y\right)Y^{-1}.$$

We calculate

$$\begin{aligned}
\frac{df}{dz} &= (1 - z)^2 v^T \left(\frac{d}{dz}(I - C(z))^{-1}\right) v + \left(\frac{d}{dz}(1 - z)^2\right) v^T(I - C(z))^{-1}v \\
&= (1 - z)^2 v^T(I - C(z))^{-1}\left(-\frac{d}{dz}(I - C(z))\right)(I - C(z))^{-1}v \\
&\quad + \left(\frac{d}{dz}(1 - z)^2\right) v^T(I - C(z))^{-1}v \\
&= (1 - z)^2 v^T(I - C(z))^{-1}R(I - C(z))^{-1}v - 2(1 - z)v^T(I - C(z))^{-1}v \\
&= v(z)^T(I - C(z))^{-1}R(I - C(z))^{-1}v(z) - 2v(z)^T(I - C(z))^{-1}v \\
&= v(z)^T(I - C(z))^{-1}R(I - C(z))^{-1}v(z) - 2v(z)^T(I - C(z))^{-1}R\mathbb{1} \\
&= v(z)^T(I - C(z))^{-1}R\left((I - C(z))^{-1}v(z)) - 2\mathbb{1}\right).
\end{aligned}$$

ETNA
Kent State University
http://etna.math.kent.edu

An application of Lemma A.5, together with the fact that B is irreducible shows that the vector $v(z)^T(I-C(z))^{-1}R$ is entrywise nonnegative with at least one positive entry (as long as $0 \leq z < 1$). We will show that the vector $(I-C(z))^{-1}v(z) - 2\mathbb{1}$ has every entry negative. We note that

$$C(z)\mathbb{1} + v(z) = C\mathbb{1} + zR\mathbb{1} + (1-z)v = C\mathbb{1} + zv + (1-z)v = C\mathbb{1} + v \leq \mathbb{1};$$

thus, $v(z) \leq \mathbb{1} - C(z)\mathbb{1} = (I - C(z))\mathbb{1}$. This implies that

$$(I - C(z))^{-1}v(z) \leq (I - C(z))^{-1}(I - C(z))\mathbb{1} = \mathbb{1} < 2\mathbb{1},$$

and so

$$(I - C(z))^{-1}v(z) - 2\mathbb{1} < 0.$$

Hence, we have shown that $f(z) < f(0)$ as long as $0 < z < 1$. Let z_0 be such that

$$v(z_0)^T\mathbb{1} = (1 - z_0)v^T\mathbb{1} = 1 - \epsilon.$$

Since $1 - \epsilon < v^T\mathbb{1} \leq 1$, we have $0 < z_0 \leq \epsilon < 1$. Let

$$\hat{B} = \begin{bmatrix} 0 & v(z_0)^T \\ v(z_0) & C(z_0) \end{bmatrix}.$$

Since $v(z_0)$ is a positive scalar multiple of v and $C(z_0)$ is equal to the sum of C and a nonnegative diagonal matrix, $\hat{B}$ is an irreducible nonnegative matrix. The sum of the entries in the first row of $\hat{B}$ is $1 - \epsilon$ and the sum of the entries in any other row is equal to the sum of the entries in the corresponding row of B. Thus, we obtain

$$(1 - \epsilon)\mathbb{1} \leq \hat{B}\mathbb{1} \leq \mathbb{1}.$$

Finally, $C(z_0)$ is symmetric since C is symmetric. Thus, $\hat{B}$ is a symmetric member of $\mathcal{B}(n, \epsilon)$ and $Q_{\hat{B}} = I$. Then, we note that

$$\alpha(\hat{B}) \leq \alpha_{\hat{B}}(1) = f(z_0)$$

and

$$f(z_0) < a + f(0) = a + bv^T(I - C)^{-1}v = \alpha(B). \qquad \square$$

Let $\mathcal{B} = \mathcal{B}(n, \epsilon)$. In calculating the value

$$\alpha(\mathcal{B}) = \inf_{B \in \mathcal{B}} \{\alpha(B)\},$$

it is sufficient to find a lower bound for $\alpha(B)$ where B is a symmetric member of $\mathcal{B}$ (via Lemma A.7), and $\alpha(B) = \alpha_B(i)$ where $\gamma_B(i) = \epsilon$ (via Lemma A.8).

A.3. A lower bound concerning stochastic complements of reversible substochastic matrices. We now calculate the value of

$$\alpha(\mathcal{B}) = \inf_{B \in \mathcal{B}} \{\alpha(B)\},$$

where $\mathcal{B} = \mathcal{B}(n, \epsilon)$. For $n = 1$, the problem is trivial. In this case, $\mathcal{B} = \{[b] : 1 - \epsilon \leq b \leq 1\}$. For $B = [b] \in \mathcal{B}$, we have $\alpha(B) = b$, so in this case,

$$\alpha(\mathcal{B}) = \inf_{B \in \mathcal{B}} \{\alpha(B)\} = 1 - \epsilon.$$

ETNA
Kent State University
http://etna.math.kent.edu

PROPOSITION A.9. *Let n be a positive integer greater than or equal to 2 and ϵ be a positive real number strictly less than 1, and let $\mathcal{B} = \mathcal{B}(n, \epsilon)$. Then,*

$$\alpha(\mathcal{B}) = \frac{(1-\epsilon)^2}{1+(n-2)\epsilon} \ .$$

Moreover, a matrix $B \in \mathcal{B}$ satisfies $\alpha(B) = \alpha(\mathcal{B})$ if and only if

$$B \cong \begin{bmatrix} 0 & \frac{1-\epsilon}{n-1}\mathbb{1}^T \\ \frac{1-\epsilon}{n-1}\mathbb{1} & C \end{bmatrix},$$

where C is an $(n-1)\times(n-1)$ symmetric nonnegative matrix such that

$$C\mathbb{1} = (1-\epsilon)\mathbb{1} - \frac{1-\epsilon}{n-1}\mathbb{1} = \frac{(1-\epsilon)(n-2)}{n-1}\mathbb{1}.$$

Proof. By Lemmas A.4, A.7, and A.8, we simply have to calculate a lower bound for $a + v^T(I-C)^{-1}v$, where

1. the matrix C is symmetric, nonnegative, and has order $n-1$,
2. the vector v is nonnegative, has order $n-1$, and satisfies $v^T\mathbb{1} \le 1-\epsilon$,
3. the matrix
$$B = \begin{bmatrix} a & v^T \\ v & C \end{bmatrix}$$
is substochastic and irreducible,
4. $a + v^T\mathbb{1} = 1-\epsilon$, and
5. $C\mathbb{1} + v \ge (1-\epsilon)\mathbb{1}$.

Let C, v, and a satisfy the above conditions, and let $m = n-1 \ge 1$ be the order of C and v. Let

$$r = C\mathbb{1} + v - (1-\epsilon)\mathbb{1}.$$

We note that $r \ge 0$. Let R be the diagonal matrix of order m with ith diagonal entry equal to r_i. As in the proof of Lemma A.7, express

$$C \cong \begin{bmatrix} C_1 & & 0 \\ & \ddots & \\ 0 & & C_l \end{bmatrix}, \ R \cong \begin{bmatrix} R_1 & & 0 \\ & \ddots & \\ 0 & & R_l \end{bmatrix}, \text{ and } v \cong \begin{bmatrix} v_1 \\ \vdots \\ v_l \end{bmatrix},$$

where each C_k is irreducible. As B is irreducible, each v_k has at least one positive entry. An application of Lemma A.5 with $Y = I + R_k$, $X = I$, and $Z = C_k$ shows that if $R_k \ne 0$, then the matrix $(I + R_k - C_k)^{-1}$ exists and entrywise it holds that

$$0 < (I + R_k - C_k)^{-1} < (I - C_k)^{-1}.$$

Thus, the matrix $(I+R-C)^{-1}$ is entrywise nonnegative. Let

$$\alpha' = a + v^T(I+R-C)^{-1}v.$$

Then,

$$\alpha' = a + v^T(I+R-C)^{-1}v = a + \sum_{i=1}^{k} v_i^T(I+R_i-C_i)^{-1}v_i$$

$$\le a + \sum_{i=1}^{k} v_i^T(I-C_i)^{-1}v_i = \alpha(B),$$

ETNA
Kent State University
http://etna.math.kent.edu

with equality if and only if $R = 0$. We note that $R = 0$ holds if and only if $B\mathbb{1} = (1-\epsilon)\mathbb{1}$.

Now, let $C' = C - R$, so that

$$\alpha' = a + v^T (I - C')^{-1} v.$$

Although the matrix C' may have negative entries, the matrix $(I - C')^{-1} = (I + R - C)^{-1}$ is entrywise nonnegative (as noted above). Since $R\mathbb{1} = C\mathbb{1} + v - (1-\epsilon)\mathbb{1}$, we have the relation $C'\mathbb{1} + v = (1-\epsilon)\mathbb{1}$ implying that

$$v = (1-\epsilon)\mathbb{1} - C'\mathbb{1} = (I - C')\mathbb{1} - \epsilon\mathbb{1}.$$

Thus, we obtain

$$\begin{aligned} v^T (I - C')^{-1} v &= \left(\mathbb{1}^T (I - C') - \epsilon \mathbb{1}^T\right) (I - C')^{-1} \left((I - C')\mathbb{1} - \epsilon\mathbb{1}\right) \\ &= \mathbb{1}^T (I - C')\mathbb{1} - 2\epsilon \mathbb{1}^T \mathbb{1} + \epsilon^2 \mathbb{1}^T (I - C')^{-1} \mathbb{1}, \end{aligned}$$

as well as

$$a = 1 - \epsilon - v^T \mathbb{1} = 1 - \epsilon - \left(\mathbb{1}^T (I - C') - \epsilon \mathbb{1}^T\right) \mathbb{1} = 1 - \epsilon - \mathbb{1}^T (I - C')\mathbb{1} + \epsilon \mathbb{1}^T \mathbb{1}.$$

Hence,

$$\begin{aligned} \alpha' &= a + v^T (I - C')^{-1} v \\ &= 1 - \epsilon - \mathbb{1}^T (I - C')\mathbb{1} + \epsilon \mathbb{1}^T \mathbb{1} + \mathbb{1}^T (I - C')\mathbb{1} - 2\epsilon \mathbb{1}^T \mathbb{1} + \epsilon^2 \mathbb{1}^T (I - C')^{-1} \mathbb{1} \\ &= 1 - \epsilon - \epsilon \mathbb{1}^T \mathbb{1} + \epsilon^2 \mathbb{1}^T (I - C')^{-1} \mathbb{1} = 1 - (m+1)\epsilon + \epsilon^2 \mathbb{1}^T (I - C')^{-1} \mathbb{1}. \end{aligned}$$

(The vector $\mathbb{1}$ in the above expression has order m and so $\mathbb{1}^T \mathbb{1} = m$). Thus, in order to calculate a lower bound for α' we simply need to calculate a lower bound for $\mathbb{1}^T (I - C')^{-1} \mathbb{1}$.

Now, C and C' are symmetric and $C - C' = R$, where R is a positive semidefinite matrix (R is a nonnegative diagonal matrix). The largest positive eigenvalue of C' is less than or equal to the largest positive eigenvalue of C (see [7, Corollary 7.7.4], for example). The matrix C is properly substochastic as it is a principal submatrix of an irreducible substochastic matrix. The largest positive eigenvalue of C is thus strictly less than 1. Altogether, C' is a symmetric real matrix whose eigenvalues are strictly less than 1, further implying that $I - C'$ is a positive definite real matrix.

By Lemma A.6, we have

$$\left(\mathbb{1}^T (I - C')^{-1} \mathbb{1}\right) \left(\mathbb{1}^T (I - C')\mathbb{1}\right) \geq \left(\mathbb{1}^T \mathbb{1}\right)^2,$$

with equality if and only if $\mathbb{1}$ is an eigenvector of C'. Note that $C'\mathbb{1} + v = (1-\epsilon)\mathbb{1}$ implies that $\mathbb{1}$ is an eigenvector of C' if and only if v is a scalar multiple of $\mathbb{1}$. Hence, we find

$$\mathbb{1}^T (I - C')^{-1} \mathbb{1} \geq \frac{(\mathbb{1}^T \mathbb{1})^2}{\mathbb{1}^T (I - C')\mathbb{1}} = \frac{m^2}{\mathbb{1}^T (I - C')\mathbb{1}} \quad ,$$

with equality if and only if v is a scalar multiple of $\mathbb{1}$. As well, we have the identities $\mathbb{1}^T \mathbb{1} = m$ and

$$\begin{aligned} \mathbb{1}^T (I - C')\mathbb{1} &= \mathbb{1}^T \mathbb{1} - \mathbb{1}^T C'\mathbb{1} = \mathbb{1}^T \mathbb{1} - \mathbb{1}^T \left((1-\epsilon)\mathbb{1} - v\right) \\ &= \epsilon \mathbb{1}^T \mathbb{1} + v^T \mathbb{1} = m\epsilon + v^T \mathbb{1} \leq m\epsilon + (1-\epsilon) = 1 + (m-1)\epsilon. \end{aligned}$$

ETNA
Kent State University
http://etna.math.kent.edu

(Recall that $v^T\mathbb{1} \leq 1-\epsilon$.) Thus,

$$\mathbb{1}^T(I-C')^{-1}\mathbb{1} \geq \frac{m^2}{1+(m-1)\epsilon},$$

with equality if and only if v is a scalar multiple of $\mathbb{1}$ and $v^T\mathbb{1} = 1-\epsilon$. These two conditions uniquely identify v: when they both hold, we have

$$v = \frac{1-\epsilon}{m}\mathbb{1}.$$

So in total we arrive at

$$\begin{aligned}\alpha(B) \geq \alpha' &= 1-(m+1)\epsilon+\epsilon^2\mathbb{1}^T(I-C')^{-1}\mathbb{1} \geq 1-(m+1)\epsilon+\epsilon^2\frac{m^2}{1+(m-1)\epsilon} \\ &= \frac{(1-(m+1)\epsilon)(1+(m-1)\epsilon)+m^2\epsilon^2}{1+(m-1)\epsilon} = \frac{(1-\epsilon)^2}{1+(m-1)\epsilon},\end{aligned}$$

with equality if and only if the matrix

$$B = \begin{bmatrix} a & v^T \\ v & C \end{bmatrix}$$

satisfies

1. $a + v^T\mathbb{1} = 1-\epsilon$,
2. $v = \frac{1-\epsilon}{m}\mathbb{1}$, and
3. $C\mathbb{1} + v = (1-\epsilon)\mathbb{1}$.

These three conditions together imply that $a = 0$ and $C\mathbb{1} = \frac{(1-\epsilon)(m-1)}{m}\mathbb{1}$. Substituting $m = n-1$, we obtain the formulae in the statement of the proposition. □

Let $n \geq 1$, $\epsilon < 1$, and let $\mathcal{B} = \mathcal{B}(n,\epsilon)$. We note that the above formula for $\alpha(\mathcal{B})$ agrees with that in the case $n = 1$. As noted, when $n = 1$,

$$\alpha(\mathcal{B}) = 1-\epsilon = \frac{(1-\epsilon)^2}{1-\epsilon} = \frac{(1-\epsilon)^2}{1+(n-2)\epsilon}.$$

For $n = 1$ or 2, the matrices $B \in \mathcal{B}$ that have

$$\alpha(B) = \alpha(\mathcal{B}) = \frac{(1-\epsilon)^2}{1+(n-2)\epsilon} = (1-\epsilon)^2$$

are unique; they are

$$B = [1-\epsilon] \text{ or } B = \begin{bmatrix} 0 & 1-\epsilon \\ 1-\epsilon & 0 \end{bmatrix},$$

respectively.

However, this minimum for $\alpha(B)$ is not uniquely attained for $n \geq 3$. For example, the matrices

$$B_1 = \begin{bmatrix} 0 & \frac{1-\epsilon}{2} & \frac{1-\epsilon}{2} \\ \frac{1-\epsilon}{2} & \frac{1-\epsilon}{2} & 0 \\ \frac{1-\epsilon}{2} & 0 & \frac{1-\epsilon}{2} \end{bmatrix}$$

ETNA
Kent State University
http://etna.math.kent.edu

and

$$B_2 = \begin{bmatrix} 0 & \frac{1-\epsilon}{2} & \frac{1-\epsilon}{2} \\ \frac{1-\epsilon}{2} & 0 & \frac{1-\epsilon}{2} \\ \frac{1-\epsilon}{2} & \frac{1-\epsilon}{2} & 0 \end{bmatrix}$$

satisfy

$$\alpha(B_1) = \alpha(B_2) = \alpha(\mathcal{B}) = \frac{(1-\epsilon)^2}{1+\epsilon} \ .$$

PROPOSITION A.10. *Let A be a reversible stochastic matrix on the state space $\mathcal{S}$. Let $\mathcal{E} \subseteq \mathcal{S}$ be such that $B = A(\mathcal{E})$ satisfies $B\mathbb{1} \geq (1-\epsilon)\mathbb{1}$ where $0 < \epsilon < 1$. Suppose further that no proper subsets of $\mathcal{E}$ satisfy this condition. Let Π be a positive diagonal matrix with ΠA being symmetric, and let $\mathcal{C} \subseteq \mathcal{S}$ be such that $\mathcal{E} \setminus \mathcal{C} = \{i\}$ where*

$$\pi_i = \max_{j \in \mathcal{E}} \{\pi_j\}.$$

Let $m = |\mathcal{E}|$ and $\tilde{A} = A \setminus \mathcal{C}$. Then,

$$\tilde{a}_{ii} \geq \frac{(1-\epsilon)^2}{1+(m-2)\epsilon} \ .$$

Proof. If $m = 1$, the claim is trivial as we have $a_{ii} \geq 1-\epsilon$ implying that

$$\tilde{a}_{ii} \geq a_{ii} \geq 1-\epsilon = \frac{(1-\epsilon)^2}{1+(m-2)\epsilon} \ .$$

As well, if $\mathcal{E} \cup \mathcal{C} = \mathcal{S}$, then $\tilde{A} = [1]$, and the claim holds.

Let $\mathcal{C}_1 = \mathcal{C} \setminus \mathcal{E}$, and let $A' = A \setminus \mathcal{C}$. Then, $B' = A'(\mathcal{E})$ satisfies $B' \geq B$ and thus $B' \in \mathcal{B}(m, \epsilon)$. Then, since $\tilde{A} = A' \setminus \mathcal{C}_2$ and $\mathcal{E} = \{i\} \cup \mathcal{C}_2$, we have

$$\tilde{a}_{ii} = \alpha_{B'}(i) \geq \frac{(1-\epsilon)^2}{1+(m-2)\epsilon} \ ,$$

via Proposition A.9. □

REFERENCES

[1] V. BLONDEL, C. DE KERCHOVE, E. HUENS, AND P. VAN DOOREN, *Social leaders in graphs*, in Positive Systems, Proceedings of the Second Multidisciplinary International Symposium on Positive Systems: Theory and Applications (POSTA 06), C. Commault and N. Marchand, eds., Lecture Notes in Control and Information Sciences, 341, Springer, Berlin, 2006, pp. 231–237.

[2] F. R. CHUNG, *Spectral Graph Theory*, AMS, Providence, 1997.

[3] P. DEUFLHARD, W. HUISINGA, A. FISCHER, AND C. SCHÜTTE, *Identification of almost invariant aggregates in reversible nearly uncoupled Markov chains*, Linear Algebra Appl., 315 (2000), pp. 39–59.

[4] P. DEUFLHARD AND M. WEBER, *Robust Perron cluster analysis in conformation dynamics*, Linear Algebra Appl., 398 (2005), pp. 161–184.

[5] M. FIEDLER, *A property of eigenvectors of nonnegative symmetric matrices and its application to graph theory*, Czechoslovak Math. J., 25 (1975), pp. 619–633.

[6] D. FRITZSCHE, V. MEHRMANN, D. SZYLD, AND E. VIRNIK, *An SVD approach to identifying metastable states of Markov chains*, Electron. Trans. Numer. Anal., 29 (2008), pp. 46–69. http://etna.math.kent.edu/vol.29.2007-2008/pp46-69.dir

[7] R. A. HORN AND C. R. JOHNSON, *Matrix Analysis*, Cambridge University Press, Cambridge, 1985.

[8] ———, *Topics in Matrix Analysis*, Cambridge University Press, Cambridge, 1991.

ETNA
Kent State University
http://etna.math.kent.edu

[9] G. LINDEN, B. SMITH, AND J. YORK, *Amazon.com recommendations: item-to-item collaborative filtering*, IEEE Internet Comp., 7 (2003), pp. 76–80.
[10] C. MEYER, *Stochastic complementation, uncoupling Markov chains, and the theory of nearly reducible systems*, SIAM Rev., 31 (1989), pp. 240–272.
[11] C. MEYER AND C. WESSELL, *Stochastic data clustering*, SIAM J. Matrix Anal. Appl, 33 (2012), pp. 1214–1236.
[12] M. NEWMAN, *Finding community structure in networks using the eigenvectors of matrices*, Phys. Rev. E, 74 (2006), 036104, 19 pp.
[13] C. SCHÜTTE, A. FISCHER, W. HUISINGA, AND P. DEUFLHARD, *A direct approach to conformational dynamics based on hybrid Monte Carlo*, J. Comp. Phys., 151 (1999), pp. 146–168.
[14] R. M. TIFENBACH, *A combinatorial approach to nearly uncoupled Markov chains*, PhD Thesis, Hamilton Institute, National University of Ireland, Maynooth, 2011.
[15] ———, *On an SVD-based algorithm for identifying meta-stable states of Markov chains*, Electron. Trans. Numer. Anal., 38 (2011), pp. 17–33.
http://etna.math.kent.edu/vol.38.2011/pp17-33.dir
[16] F. ZHANG (ed.), *The Schur Complement and its Applications*, Springer, New York, 2005.