Markov Chains and Spectral Clustering

Ning Liu and William J. Stewart

Department of Computer Science
North Carolina State University, Raleigh, NC 27695-8206, USA
nliu@ncsu.edu,
billy@ncsu.edu

Abstract. The importance of Markov chains in modeling diverse systems, including biological, physical, social and economic systems, has long been known and is well documented. More recently, Markov chains have proven to be effective when applied to internet search engines such as Google's PageRank model [7], and in data mining applications wherein data trends are sought. It is with this type of Markov chain application that we focus our research efforts. Our starting point is the work of Fiedler who in the early 70's developed a spectral partitioning method to obtain the minimum cut on an undirected graph (symmetric system). The vector that results from the spectral decomposition, called the Fiedler vector, allows the nodes of the graph to be partitioned into two subsets. At the same time that Fiedler proposed his spectral approach, Stewart proposed a method based on the dominant eigenvectors of a Markov chain — a method which was more broadly applicable to nonsymmetric systems. Enlightened by these, somewhat orthogonal, results and combining them together, we show that spectral partitioning can be viewed in the framework of state clustering on Markov chains. Our research results to date are two-fold. First, we prove that the second eigenvector of the signless Laplacian provides a heuristic solution to the NP-complete state clustering problem which is the dual of the minimum cut problem. Second, we propose two clustering techniques for Markov chains based on two different clustering measures.

Keywords: spectral clustering, graph partitioning, Markov chains, eigenvector.

1 Introduction

The aim of clustering is to group objects together on the basis of similarity measurement criteria. Roughly speaking, clustering algorithms can be divided into two classes: hierarchical and partitioning [4]. In this paper, we focus on graph partitioning which provides clustering information for graph nodes [1,6].

A graph $G = (V, E)$, where V is a set of vertices and E is a set of edges, can be partitioned into two disjoint subsets V_1 and V_2, $V_1 \cup V_2 = V$, by deleting those edges that connect V_1 and V_2. A common objective is to partition the graph in such a way that the partition has minimum cost, e.g., the minimum number of

edges cut. In a weighted graph this involves ensuring that the sum of the weights on the removed edges is as small as possible. In graph theoretic language, this is called the minimum cut. Given two sets of vertices V_1 and V_2 which partition a graph, the numerical value assigned to the cut, called the "cut value", is given by

$$\text{cut}(V_1, V_2) = \sum_{i \in V_1, \, j \in V_2} w_{ij}$$

where w_{ij} is the weight on the edge that connects vertex i to vertex j. Spectral graph partitioning is an important and effective heuristic approach for finding good solutions to the minimum cut problem. It was introduced in the early 1970s [3], and popularized in the early 1990s [11].

The starting point for applying spectral partitioning on a graph is to create a matrix representation of the graph, e.g., the Laplacian matrix. Fiedler first derived well known representations of the Laplacian to show its connection to the minimum cut problem [3] i.e., minimizing the total weight of edges cut. Besides minimum cut, a number of different graph partitioning objectives have also been proposed, such as maximum association [12] and normalized cut (balanced minimum cut) [12]. Maximum association and minimum cut can be viewed as dual problems of each other. In this paper we introduce the signless Laplacian matrix [2] to model the maximum association in spectral clustering, and show that the eigenvalues of the signless Laplacian have important physical interpretations.

Markov chains are widely used in modeling diverse systems and in data mining applications, e.g., applied to internet search engines such as Google's PageRank model [7]. Spectral based methods can also be applied to a Markov chain for the purposes of state clustering analysis. The dominant eigenvectors of the transition probability matrix of a Markov chain allow the states of the Markov chain to be arranged into meaningful groups [13]. Most recently, an interesting observation of spectral clustering provided by Meila et al.[8] is that spectral clustering can be depicted in the framework of Markov random walks on graph structure. It would appear that the results from Fiedler and Stewart have been unified, but in fact this is not the case as we shall explain later.

The rest of this paper is organized as follows. Section 2.1 reviews the basic procedure of spectral clustering by the Laplacian matrix. Section 2.2 introduces a new spectral method to cluster graph nodes with the objective of maximum association, the dual problem of minimum cut. Section 2.3 shows how to construct the stochastic matrix of a random walk from a graph so as to perform spectral clustering. In Section 2.4, we introduce two clustering measures on the states of Markov chains, first the normalized cut measure and second the distance measure of states from the steady state, and we propose a novel technique of clustering on graph nodes which is based on incorporating these two clustering techniques. Section 3 describes an application of NCD Markov chains while Section 4 presents our conclusions.

2 Spectral Graph Partitioning and State Clustering

2.1 Minimum Cut via Eigen-Analysis

In this section, we first review the basic procedure of spectral clustering. The starting point is to create the *Laplacian matrix* of a graph. Representations of the Laplacian show connections to the cut problem [3]. The Laplacian matrix L of an undirected graph $G = (V, E)$ with n vertices and m edges is the $n \times n$ matrix whose elements are as follows.

$$L_{ij} = \begin{cases} \sum_k w_{ik}, & \text{if } i = j \\ -w_{ij}, & \text{if } i \neq j, \ i \text{ and } j \text{ are adjacent} \\ 0, & \text{otherwise,} \end{cases}$$

The matrix L can be obtained as $L = D - A$, where A is the adjacency matrix of G and D is a diagonal matrix whose i^{th} diagonal element is $D_{ii} = \sum_k w_{ik}$. It is well known that the Laplacian matrix of a graph has a number of interesting properties [3].

1. L is symmetric positive semi-definite. As such its eigenvalues are all real and non-negative. Furthermore the eigenvectors of L constitute a full set of n real and orthogonal vectors.
2. $Le = 0$, where e is a column vector whose elements are all equal to 1. Thus 0 is the smallest eigenvalue of L and e is its corresponding eigenvector.
3. For any vector x, we have

$$x^T L x = \sum_{\{i,j\} \in E} w_{ij} (x_i - x_j)^2. \tag{1}$$

Given a partition of V into V_1 and V_2 ($V_1 \cup V_2 = V$), a partition vector p is defined as

$$p_i = \begin{cases} +1, & \text{vertex } i \in V_1, \\ -1, & \text{vertex } i \in V_2. \end{cases} \tag{2}$$

Clearly we see that $p^T p = n$. Given a Laplacian matrix L and a partition vector p, we have, from Equation (1),

$$p^T L p = \sum_{\{i,j\} \in E} w_{ij} (p_i - p_j)^2.$$

Observe from Equation (2), that the weight of edges within each set V_1 or V_2 is not counted in this sum, while the weight of each edge connecting a vertex of V_1 to a vertex in V_2 is multiplied by a factor of 4. Given that we have defined the cut value as $\text{cut}(V_1, V_2) = \sum_{i \in V_1, j \in V_2} w_{ij}$, it follows that $p^T L p = 4 \, \text{cut}(V_1, V_2)$ and the Rayleigh quotient is

$$\frac{p^T L p}{p^T p} = \frac{1}{n} \cdot 4 \, \text{cut}(V_1, V_2). \tag{3}$$

A well known result from matrix computation [5,10] is that the maximum and minimum of the Rayleigh quotient can be obtained as the largest and smallest eigenvalues of the Laplacian matrix L:

$$\lambda_{max} = \max_{x \neq 0} \frac{x^T L x}{x^T x} \quad \text{and} \quad \lambda_{min} = \min_{x \neq 0} \frac{x^T L x}{x^T x}, \tag{4}$$

where λ_{max} and λ_{min} are the largest and smallest eigenvalues of L respectively, and the x that achieves the maximum or minimum is the corresponding eigenvector of λ_{max} or λ_{min}. Therefore, the minimum value of the Rayleigh quotient is zero, which is the smallest eigenvalue of L corresponding to the eigenvector e. This partition vector indicates that all the vertices of the graph are in the same set, which means nothing is cut. This is the trivial partition. It is the second smallest eigenvalue (referred to as the Fiedler value) of L that provides the optimal value, and its corresponding eigenvector (referred to as the Fiedler vector) gives the real valued partition vector for our minimum cut problem.

Once the eigenvector corresponding to the second smallest eigenvalue has been computed, we can partition the vertices into two subsets. In the ideal case, the eigenvector takes on only two discrete values and the signs of the values tell us exactly how to partition the graph. However, the eigenvectors assume continuous values and we need to choose a splitting point to partition it into two parts. Here we take value zero. Actually, there are many other different ways of choosing such a splitting point. One can take the median value as the splitting point or one can search for the splitting point such that the resulting partition has the best cut value. We can also group close values together to partition the eigenvector into several parts.

2.2   Node Clustering in Graphs

Sometimes we wish to determine clustering information for a graph G, i.e., to figure out which vertices in G have strong mutual cohesion. The problem of maximizing the total weight within two clusters is the dual of the minimum cut problem. In our work, we propose a new method for modeling this dual problem by introducing an eigen-analysis of the signless Laplacian matrix M. We shall see that the clustering solution from an eigen-analysis of the signless Laplacian coincides with the solution of the minimum cut problem. We now proceed through the details of this dual problem and describe our approach to solving it.

Equation (1) suggests how we might find the maximum total weight within clusters. If we change the minus sign in Equation (1) and construct a matrix M that satisfies

$$p^T M p = \sum_{\{i,j\} \in E} w_{ij}(p_i + p_j)^2, \tag{5}$$

then we see that the edges connecting the two subsets V_1 and V_2 do not contribute to the value of this equation, but instead the edges within each cluster contribute

4 times their weight. The dual problem can then be solved, and the optimal p which maximizes $p^T M p$ provides our sought-after clustering information.

To obtain the matrix M, we need the unoriented incidence matrix I_G of G. I_G is an $n \times m$ matrix, wherein each row is associated with a vertex and each column is associated with an edge, and the column of I_G corresponding to an edge from vertex i to j has zeros everywhere except in positions i and j: the two nonzero elements are given by $\sqrt{w_{ij}}$.

Given any vector x, let k be the element of $I_G{}^T x$ that corresponds to the edge $\{i, j\}$. Then,

$$(I_G{}^T x)_k = \sqrt{w_{ij}}(x_i + x_j).$$

If we now construct the matrix $M = I_G I_G{}^T$, then

$$x^T M x = x^T I_G I_G{}^T x = (x^T I_G)_{1 \times m}(I_G{}^T x)_{m \times 1}, \quad \text{for all} \ \ x.$$

This implies that

$$x^T M x = \sum_{\{i,j\} \in E} w_{ij}(x_i + x_j)^2.$$

Replacing x with the partition vector p yields Equation (5). The matrix M is called the signless Laplacian matrix [2]. It can also be obtained as $M = D + A$, where A is the adjacency matrix of the graph G and D is a diagonal matrix with diagonal entry $D_{ii} = \sum_k w_{ik}$. The Rayleigh quotient $p^T M p / p^T p$ provides a quantitative evaluation of the cohesion of vertices within clusters. We can maximize the Rayleigh quotient $p^T M p / p^T p$ to obtain the node clusters in graphs. Actually, because of the relation $M = 2D - L$, maximizing $p^T M p / p^T p$ is equivalent to minimizing $p^T L p / p^T p$ in the minimum cut problem.

By the Perron-Frobenius theorem[10], since M is real-symmetric, nonnegative and irreducible, all its eigenvalues are real and the eigenvector corresponding to the largest eigenvalue is the only eigenvector of M whose elements are all nonzero and positive. We already know that the eigenvalues of the signless Laplacian matrix M can be depicted as the maximum, minimum and intermediate values of the Rayleigh quotient. Since the eigenvector corresponding to the largest eigenvalue of M is positive, this means that all the vertices are in one cluster; it does not satisfy our objective of getting two clusters of vertices. Therefore the second largest eigenvalue of M is the maximum value that we want and it approximately estimates the value of the cohesion of vertices within the clusters. The associated eigenvector generates the clustering information among these vertices of the graph G. The clustering result indicated here should coincide with the minimum cut result.

This leads us to ask why we still need to model the dual problem if the results from the dual and the original are the same. One significant reason is that the eigenvalues of the signless Laplacian have important physical interpretations. They provide a quantitative evaluation of total weight of edges within clusters. Since the eigenvalues of the Laplacian matrix provide a measure of the total weight of edges connecting clusters, the ratio of these eigenvalues can be used to analyze and evaluate the effects of partitioning and clustering.

Spectral clustering not only provides a bi-partitioning result, i.e. two clusters of graph nodes, but also multiple clusters. Roughly, there are two classes of methods in spectral clustering to obtain multiple clusters: (a) recursive bi-partitioning, and (b) using multiple eigenvectors. In Section 3, we will show an example of using multiple eigenvectors to generate more than two clusters.

2.3   State Clustering on Markov Chains and Graphs

Spectral based methods may also be applied to a directed graph viewed as a Markov chain for clustering purposes. In a Markov chain, the state transition probability diagram can be viewed as a graph with directed edges. It is well known that when a Markov chain is irreducible and aperiodic with a finite state space, its stochastic matrix has a single eigenvalue on the unit circle, the eigenvalue 1, and the left-hand eigenvector corresponding to this unit eigenvalue is the stationary distribution vector. In [13] it is shown that the right-hand eigenvector of a stochastic matrix corresponding to the subdominant eigenvalue, i.e., the eigenvalue with second largest modulus closest to, but strictly less than, 1, provides the information necessary to cluster states into coherent groups. This approach is based on the concept of distance (the number of transition steps) from each state to the stationary distribution.

Therefore, the spectral approach can be applied not only in the symmetric case, e.g., Laplacian and signless Laplacian matrices, but also in the asymmetric case such as Markov chains. The clusters of graph nodes in spectral clustering are based on the minimum cut sense, while the clusters of Markov chains states are in the sense of distance from each state to the steady state. Actually these two clustering measures can be connected in the framework of a random walk on a graph topology. In other words, spectral clustering can be viewed in the framework of state clustering on Markov chains [8].

It has been shown that the adjacency matrix A of a graph can be converted to a transition probability matrix P to generate a random walk on a graph as:

$$P = D^{-1}A.$$

D is a diagonal matrix whose diagonal elements are the vertex degree. Using the Laplacian matrix $L = D - A$ and the signless Laplacian matrix $M = D + A$, we obtain the following formulas showing their relationship with P:

$$I - P = D^{-1}L \quad \text{and} \quad I + P = D^{-1}M$$

Proposition 1. If an eigenpair (λ, v) is a solution of the eigenvalues/vectors problem $Pv = \lambda v$, then the pair (λ, v) is also a solution of the generalized eigenvalues/vectors problems $(1 - \lambda)Dv = Lv$ and $(1 + \lambda)Dv = Mv$.

Proof: Given a pair (λ, v) which satisfies the eigenvalues/vectors problem $Pv = \lambda v$, since $P = D^{-1}A$ and $L = D - A$, we have [8]

$$Pv = \lambda v \quad \Rightarrow \quad D^{-1}Av = \lambda v \quad \Rightarrow \quad D^{-1}(D - L)v = \lambda v$$

$$\Rightarrow \quad Iv - D^{-1}Lv = \lambda v \quad \Rightarrow \quad (1-\lambda)Dv = Lv \tag{6}$$

It is the same for the signless Laplacian matrix $M = D + A$:

$$Pv = \lambda v \quad \Rightarrow \quad D^{-1}Av = \lambda v \quad \Rightarrow \quad D^{-1}(M-D)v = \lambda v$$

$$\Rightarrow \quad D^{-1}Mv - Iv = \lambda v \quad \Rightarrow \quad (1+\lambda)Dv = Mv \tag{7}$$

It is known that the eigenvectors of the generalized eigenvalue problem $(1-\lambda)Dv = Lv$ provide a heuristic solution for the minimum balanced cut on a graph [12]. Therefore, from the proposition above, we can assert that the right eigenvectors of P also provide the balanced cut solution, just like the eigenvectors of the generalized problem on M.

Before we discuss the details of clustering information given by the right-hand eigenvectors of P, let us first look at an interesting property of P. This property is only true for the P of random walk generated on a graph, and not for all transition probability matrices.

Proposition 2. All the eigenvalues of the probability matrix P derived from a random walk on a graph are real.

Proof: $P = D^{-1}A$, though not symmetric does has a symmetric structure. Since

$$D^{1/2}PD^{-1/2} = D^{-1/2}AD^{-1/2},$$

the matrix P and the symmetric matrix $D^{-1/2}AD^{-1/2}$ are similar. Similar matrices share the same eigenvalues. Therefore all eigenvalues of P are real.

Proposition 2 provides and alternative way to calculate the eigenvectors and eigenvalues of P: it is easier to compute the eigenvectors/eigenvalues of a symmetric matrix than an unsymmetric one.

Let λ be an eigenvalue of P generated from a graph, and let x_R and x_L be the corresponding right and left eigenvectors of P respectively. Then

$$Px_R = \lambda x_R \quad \text{and} \quad P^T x_L = \lambda x_L.$$

First, let us focus on x_R. For the eigenvalue problem of P:

$$Px_R = \lambda x_R \quad \Rightarrow \quad D^{-1}Ax_R = \lambda x_R.$$

If we premultiply $D^{1/2}$ on both sides, we obtain

$$D^{-1/2}AD^{-1/2}(D^{1/2}x_R) = \lambda(D^{1/2}x_R).$$

Therefore, the value λ and the vector $D^{1/2}x_R$ are respectively an eigenvalue and eigenvector of the matrix $D^{-1/2}AD^{-1/2}$. Second, let us now consider x_L. Since A is symmetric:

$$P^T x_L = \lambda x_L \quad \Rightarrow \quad AD^{-1}x_L = \lambda x_L.$$

If we now premultiply $D^{-1/2}$ on both sides we obtain

$$D^{-1/2}AD^{-1/2}(D^{-1/2}x_L) = \lambda(D^{-1/2}x_L).$$

Hence the value λ and the vector $D^{-1/2}x_L$ are respectively an eigenvalue and eigenvector of the matrix $D^{-1/2}AD^{-1/2}$. Now we see that x_R and x_L can be obtained from calculating the eigenvectors of the matrix $D^{-1/2}AD^{-1/2}$. Let v be the eigenvector of $D^{-1/2}AD^{-1/2}$ corresponding to the eigenvalue λ. Then

$$x_R = D^{-1/2}v \quad \text{and} \quad x_L = D^{1/2}v.$$

Let us now return to the clustering information provided by the probability matrix P of a random walk on a graph. Spectral clustering is in the measure of minimum balanced cut, while Markov chains clustering is in the measure of distance of states from the steady state. Now a question rises: do these two kinds of clustering results coincide in the right eigenvectors of P? The answer is not simply "yes" or "no", but instead depends on the eigenvalues of P.

Clustering on Markov chains uses the right eigenvector of P corresponding to the eigenvalue with modulus closest to 1. If the eigenvalue closest to the unit circle is positive, then it will be shifted to the eigenvalue closest to 0 in the generalized eigenvalue problem of L in Equation (6) which is used in spectral clustering, and the corresponding eigenvector gives the minimum balanced cut on the graph. In this case the two clustering results coincide. However if the eigenvalue of P closest to the unit circle is negative, then it will be shifted to the eigenvalue closest to 2 in Equation (6) and so the corresponding eigenvector will not give the clustering information based on the minimum balanced cut of a graph. Now the two clustering results are distinct.

2.4   Two Clustering Techniques for Markov Chains

After generating a random walk on a graph via $P = D^{-1}A$, we can perform a clustering analysis on the Markov chain of this graph. If the Markov chain is ergodic, then spectral decomposition of the transition probability matrix P provides two clustering measures of graph nodes.

The first measure is based on the minimum balanced cut. If we sort the eigenvalues of the probability matrix P in descending order, the eigenvector associated with the second largest eigenvalue (should be positive)— which is shifted to the second smallest one in the generalized eigenvalue problem (6) and the second largest one in (7) — gives clusters of states based on the minimum balanced cut from the spectral clustering point of view. There is much material available concerning the use of the eigenvectors of P having positive eigenvalues to carry out spectral clustering [9].

The second measure is based on the "distance" of each state to the steady state. If we sort the modulus of eigenvalues in descending order, the eigenvector associated with the second largest one (not necessarily positive) gives clusters of states based on this distance measure from a Markov chain point of view. If the second largest one is positive, then the two clustering measures coincide; otherwise they do not. Previous research has paid much more attention to the

first measure than to the second, due primarily to the fact that clustering using the first measure is closely associated with the structure of real world networks. However the second measure is also related to the graph structure and we should not simply ignore the information it provides.

Let us focus on the case when the eigenvalue having the second largest modulus is negative. How should we identify the clustering information from the corresponding eigenvector? An element with small amplitude in the eigenvector indicates the corresponding state is close to the steady state and perhaps we can say that it belongs to a group of states which are more closely linked to each other. On the other hand, an element with large amplitude indicates the corresponding state is far from the steady state and possibly belongs to a group in which states are not closely linked to each other. That means the clustering information should be identified through close amplitude values of elements in the eigenvector. The eigenvector associated with a negative eigenvalue whose modulus is very close to 1 indicates the role or position of states (graph nodes) in each cluster and also the cluster structure.

Fig. 1. Joint complete and star graph

A concrete example may help in understanding the clustering techniques for Markov chains. Figure 1 is an unweighted graph with 11 vertices. After forming the transition probability matrix P of the associated random walk on this graph, we find that second largest eigenvalue of P is $\lambda_1 = 0.8852$ while the eigenvalue with second largest modulus is $\lambda_2 = -0.9336$. The corresponding eigenvectors are v_1 and v_2.

$$(\mathbf{v_1}, \mathbf{v_2}) = \begin{pmatrix} -0.1423 & 0.0105 \\ -0.1423 & 0.0105 \\ -0.1423 & 0.0105 \\ -0.1423 & 0.0105 \\ -0.0770 & -0.0707 \\ 0.2581 & -0.3085 \\ 0.2581 & -0.3085 \\ 0.2581 & -0.3085 \\ 0.2581 & -0.3085 \\ 0.2581 & -0.3085 \\ 0.2285 & 0.2880 \end{pmatrix}.$$

Based on the first clustering measure from v_1, we obtain cluster 1 {vertices 1,...,5} and cluster 2 {vertices 6,...,11}. Moving to the second clustering measure, we make two observations:

(a) Eigenvector v_2 shows that the first five elements associated with cluster 1 have relatively small modulus while others associated with cluster 2 have relatively large modulus. This means that cluster 1 has a structure in which states are closely linked to each other, while cluster 2 has the opposite structure. This observation can be confirmed by the figure. The left part of figure is a clique and right part is a star.

(b) If we sort the values of elements in v_2, then the value of vertex 5 is closer to cluster 2 than other vertices in cluster 1 while the value of vertex 11 is closer to cluster 1 than other vertices in cluster 2. This is reasonable, because vertex 5 and 11 have the possibility of transitioning to the other cluster in a single step; they are the connecting vertices between clusters. This can also be confirmed by the observation that vertex 5 in v_2 has a modulus which is relatively large compared with other vertices in cluster 1; vertex 11 has relatively smaller modulus in cluster 2.

Therefore, the eigenvector associated with a negative eigenvalue whose modulus is very close to 1 indicates firstly, the cluster structure and secondly, the role or position of states (graph nodes) in each cluster. When we perform a clustering analysis on P, it is appropriate to combine both clustering techniques, using not only the eigenvectors with positive eigenvalues but also those with negative eigenvalues to obtain more comprehensive information concerning each cluster of states.

There are two applications of Markov chains associated with the two types of clustering techniques. One is associated with nearly completely decomposable (NCD) Markov chains, and the other with nearly periodic Markov chains. In NCD Markov chains, there are several positive eigenvalue very close to 1. In this case, the eigenvectors corresponding to this kind of eigenvalue will provide the first clustering measure information. In nearly periodic Markov chains, there are multiple eigenvalues close to the unit circle but not close to 1. In this case, the corresponding eigenvectors provide the second clustering measure information.

To summarize, spectral decomposition of P not only provides the clusters of nodes satisfying the minimum balanced cut, the same as that obtained from the Laplacian and the signless Laplacian matrix, but also it provides additional clustering information concerning each cluster's structure and the role or position of graph nodes in each cluster.

3   NCD Markov Chains

The right-hand eigenvectors of the transition probability matrix P provide clustering information concerning the states of a Markov chain — especially the eigenvector corresponding to the second largest eigenvalue of P. This idea can be applied to a Markov chain with NCD structure to find nearly uncoupled

blocks. It is frequently the case that the state space of a Markov chain can be partitioned into disjoint subsets, with strong interaction between states within subsets but with weak interaction between subsets themselves. Such problems are called nearly completely decomposable (NCD) Markov chains.

For example, given the following transition probability matrix P (Courtois [13]) with NCD structure as indicated,

$$P = \begin{pmatrix} 0.85 & 0.0 & 0.149 & 0.0009 & 0.0 & 0.00005 & 0.0 & 0.00005 \\ 0.1 & 0.65 & 0.249 & 0.0 & 0.0009 & 0.00005 & 0.0 & 0.00005 \\ 0.1 & 0.8 & 0.0996 & 0.0003 & 0.0 & 0.0 & 0.0001 & 0.0 \\ 0.0 & 0.0004 & 0.0 & 0.7 & 0.2995 & 0.0 & 0.0001 & 0.0 \\ 0.0005 & 0.0 & 0.0004 & 0.399 & 0.6 & 0.0001 & 0.0 & 0.0 \\ 0.0 & 0.00005 & 0.0 & 0.0 & 0.00005 & 0.6 & 0.2499 & 0.15 \\ 0.00003 & 0.0 & 0.00003 & 0.00004 & 0.0 & 0.1 & 0.8 & 0.0999 \\ 0.0 & 0.00005 & 0.0 & 0.0 & 0.00005 & 0.1999 & 0.25 & 0.55 \end{pmatrix},$$

we compute the eigenvalues to be

$$\textbf{Eigenvalues} = \begin{pmatrix} 1.0\ 0.9998\ 0.9985\ 0.7500\ 0.5501\ 0.4000\ 0.3007\ -0.1495 \end{pmatrix}.$$

The eigenvalues $\lambda_2 = 0.9998$ and $\lambda_3 = 0.9985$ are very close to the unit eigenvalue 1, and very close to each other. From the discussion in Section 2.4, the second and third right-hand eigenvectors are approximately of equal importance, and hence the third cannot be ignored. Both of them provide clustering information. The corresponding right-hand eigenvectors of λ_2 and λ_3 are

$$\textbf{v}_2, \textbf{v}_3 = \begin{pmatrix} -0.3536 & -0.4876 \\ -0.3536 & -0.4878 \\ -0.3536 & -0.4883 \\ -0.3536 & 0.3783 \\ -0.3536 & 0.3777 \\ 0.3536 & 0.0073 \\ 0.3536 & 0.0073 \\ 0.3536 & 0.0073 \end{pmatrix},$$

These two eigenvectors can be used to obtain the NCD blocks for the Markov chain. Based on v_2, the first five states form one cluster, the remaining states form the other cluster. After checking v_3, it is obvious that the first five states can yet again be partitioned into two clusters corresponding to values that are close to one another This leads to three blocks of states for this NCD Markov chain, namely $\{1,2,3\}$, $\{4,5\}$ and $\{6,7,8\}$.

4   Conclusion

In this paper, we introduced the concept of modeling the dual problem of graph partitioning, i.e., graph clustering using a spectral method applied to the signless

Laplacian matrix. This spectral algorithm generates clusters values (eigenvalues) that can be compared with cut values. Furthermore, we proposed two clustering techniques for Markov chains. These two clustering measures obtained from the probability matrix allow us to obtain more comprehensive information of clusters on graph nodes.

References

1. Brand, M., Huang, K.: A unifying theorem for spectral embedding and clustering. In: 9th International Conference on Artificial Intelligence and Statistics (2003)
2. Cvetkovic, D., Rowlinson, P., Simic, S.K.: Signless Laplacians of finite graphs. Linear Algebra and its Applications 423, 155–171 (2007)
3. Fiedler, M.: Algebraic connectivity of graphs. Czecheslovak Mathematical Journal 23, 298–305 (1973)
4. Filippone, M., Camastra, F., Masulli, F., Rovetta, S.: A survey of kernel and spectral methods for clustering. Pattern Recognition 41(1), 176–190 (2008)
5. Golub, G.H., Van Loan, C.F.: Matrix computations, 3rd edn. Johns Hopkins University Press (1996)
6. Higham, D.J., Kibble, M.: A unified view of spectral clustering. University of Strathclyde Mathematics Research Report (2004)
7. Hilgers, P.V., Langville, A.N.: The five greatest applications of Markov Chains (2006)
8. Meila, M., Shi, J.: Learning segmentation by random walks. In: NIPS, pp. 873–879 (2000)
9. Meila, M., Shi, J.: A random walks view of spectral segmentation (2001)
10. Meyer, C.D.: Matrix analysis and applied linear algebra. SIAM (2000)
11. Pothen, A., Simon, H., Liou, K.-P.: Partitioning sparse matrices with eigenvectors of graphs. SIAM Journal on Matrix Analysis and Applications 11(3), 430–452 (1990)
12. Shi, J., Malik, J.: Normalized Cuts and Image Segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence 22, 888–905 (2000)
13. Stewart, W.J.: Introduction to the Numerical Solution of Markov Chains. Princeton (1994)