

NORMALIZED CUTS ARE APPROXIMATELY INVERSE EXIT TIMES*

MATAN GAVISH[†] AND BOAZ NADLER[‡]

Abstract. Normalized cut is a widely used measure of separation between clusters in a graph. In this paper we provide a novel probabilistic perspective on this measure. We show that for a partition of a graph into two weakly connected sets, $V = A \uplus B$, the multiway normalized cut is approximately $MNcut \approx 1/\tau_{A \rightarrow B} + 1/\tau_{B \rightarrow A}$, where $\tau_{A \rightarrow B}$ is the unidirectional characteristic exit time of a random walk from subset A to subset B . Using matrix perturbation theory, we show that this approximation is exact to first order in the connection strength between the two subsets A and B , and we derive an explicit bound for the approximation error. Our result implies that for a Markov chain composed of a reversible subset A that is weakly connected to an absorbing subset B , the easy-to-compute normalized cut measure is a reliable proxy for the chain’s spectral gap.

Key words. normalized cut, graph partitioning, Markov chain, characteristic exit time, generalized eigenvalues, perturbation bounds, matrix perturbation theory

AMS subject classifications. 62H30, 15A18, 15A42, 60J20, 60J10, 65H17

DOI. 10.1137/110826928

1. Introduction. Clustering of data is a fundamental problem in many fields. When the data is represented as a symmetric weighted similarity graph, clustering becomes a graph partitioning problem. In this paper we focus on the popular *normalized cut* criteria for graph partitioning. The normalized cut (Ncut) criteria, and its variants based on conductance, are widely used measures of separation between clusters in a graph in a variety of applications [7, 17]. Since optimizing the Ncut measure is in general an NP-hard problem, several researchers suggested various spectral-based approximation methods; see, for example, [3, 15]. These spectral methods, in turn, have interesting connections to kernel K-means clustering and to nonnegative matrix factorizations; see [4, 5].

In this paper we focus on the normalized cut measure itself. Similar to previous works [9, 14], we consider a random walk on the weighted graph, with the probability of jumping from one node to another being proportional to the edge weight between them. Our contribution is a novel Markov chain-based perspective on this measure. Our main result is that for a partition of a graph V into two disjoint sets, $V = A \uplus B$, the multiway normalized cut measure is approximately

$$(1.1) \quad MNcut \left(A \uplus B \right) \approx \frac{1}{\tau_{A \rightarrow B}} + \frac{1}{\tau_{B \rightarrow A}},$$

where $\tau_{A \rightarrow B}$ is the characteristic exit time of a random walk from cluster A to cluster B , as defined explicitly below. Furthermore, we provide an explicit bound for the approximation error in (1.1), which is second order in the connection strengths between the two subsets A and B . Our proof is based on matrix perturbation theory for

*Received by the editors March 8, 2011; accepted for publication (in revised form) by B. Hendrickson March 11, 2013; published electronically June 20, 2013.

<http://www.siam.org/journals/simax/34-2/82692.html>

[†]Department of Statistics, Stanford University, Stanford, CA 94305 (gavish@stanford.edu). This author was supported by a William R. and Sara Hart Kimball Stanford Graduate Fellowship.

[‡]Department of Computer Science and Applied Mathematics, Weizmann Institute of Science, Rehovot 76100, Israel (boaz.nadler@weizmann.ac.il). This author was partly supported by a grant from the ISF.

generalized eigenvalue problems [18]. Interestingly, (1.1) can be viewed as a discrete analogue of a well-known result for continuous diffusion processes with two metastable states [8]; see section 4.

From a mathematical point of view, when the connection strengths between the two clusters are small, the two clusters become nearly uncoupled and the resulting Markov chain *nearly decomposable*. The properties of such singularly perturbed Markov chains have been a subject of several works; see the surveys [16, 1] and the many references therein. Some works considered the behavior of the stationary distribution, others considered mean first passage times, typically focusing on singular Laurent series expansions, and yet others considered the important problem of dimensionality reduction and aggregation of states for such nearly reducible Markov chains; see, for example, [6, 11]. Our focus, in contrast, is on the relation of such Markov chains to the normalized cut measure—an easy-to-compute quantity used in a variety of applied scenarios. To this end, our analysis involves the study of a singularly perturbed Markov chain with transient states, as we consider one of the clusters as absorbing with respect to a random walk that starts at the other cluster. Furthermore, rather than series expansions, we derive explicit error bounds for the approximation in (1.1).

Our result has several implications and potential applications. First, in the context of clustering, (1.1) provides a precise probabilistic statement regarding the relation between the algebraic, easy-to-compute Ncut measure of a given partition and the eigenvalues corresponding to local random walks on each of these partitions. Our result is different from previously derived global spectral bounds that depend on the eigenvalues of the random walk on the entire graph. In particular, our analysis implies that for weakly connected clusters, the normalized cut measure essentially considers only the time it takes for a random walk to *exit* a given cluster, without taking into account its internal structure. As discussed in more detail in [12], this leads to a fundamental limitation of the Ncut measure to detect structures at different scales in a large graph and motivates the use of other measures that also take into account the internal characteristics of potential clusters [13].

Another potential application of our result is in the study of stochastic dynamical systems and in particular their discretization into reversible finite Markov chains [3]. In this field, a fundamental problem is the identification of *almost invariant sets*, e.g., subsets of phase space where a stochastic system stays for a long time before exiting. Our observation above, in the context of clustering, is also relevant here. As complex dynamical systems may exhibit invariant sets at very different spatial and time scales, our analysis suggests that the Ncut measure or its various spectral relaxations may not be appropriate in such cases.

Yet another application of our result is to the numerical evaluation of the spectral gap of nearly decomposable Markov matrices. In the study of stochastic dynamical systems, the equivalent quantity of interest are the mean exit times from *given* or a priori known metastable sets, and their dependence on various parameters, such as temperature, bond-strengths, etc. Upon discretization of such systems, this amounts to the exit times of nearly uncoupled Markov chains. In this context, one may apply (1.1) in the reverse direction: rather than computing the second largest eigenvalue of a potentially very large matrix, and from it the characteristic exit time from this metastable state, one may approximate it by the easy-to-compute normalized cut measure, with explicit guarantees as to the accuracy of this approximation.

The paper is organized as follows. The normalized cut and relevant notation are introduced in section 2. Our main result is stated in section 3, discussed in section 4,

and demonstrated using a simple numerical simulation in section 5. The proof appears in section 6, with some auxiliary results deferred to the appendix.

2. Preliminaries. Let (V, W) be a connected weighted graph with node set $V = \{1, \dots, n\}$ and an $n \times n$ affinity (or edge weight) matrix $W = (W_{i,j})_{1 \leq i, j \leq n}$. The matrix W is assumed to be symmetric with nonnegative entries and such that $W_{ii} > 0$ for all $i = 1, \dots, n$.

2.1. The normalized cut of a graph partition. Consider a partition of the node set into two disjoint “clusters” $V = A \uplus B$ of sizes k and $n - k$ nodes, respectively. Without loss of generality, we denote the nodes by $A = \{1, \dots, k\}$ and $B = \{k + 1, \dots, n\}$. As discussed in the introduction, the normalized cut [17] is a widely used, easy-to-calculate measure of separation for a given graph partition.

DEFINITION 2.1. For a partition $V = A \uplus B$, define

$$\begin{aligned} \text{Cut}(A, B) &= \sum_{i \in A, j \in B} W_{i,j}, \\ \text{Assoc}(A, V) &= \sum_{i \in A, j \in V} W_{i,j}. \end{aligned}$$

Then the normalized cut of A with respect to B is

$$(2.1) \quad \text{Ncut}(A, B) = \frac{\text{Cut}(A, B)}{\text{Assoc}(A, V)},$$

whereas the multiway normalized cut of the partition $V = A \uplus B$ is defined as [10, 19]

$$(2.2) \quad \text{MNcut}(A \uplus B) = \text{Ncut}(A, B) + \text{Ncut}(B, A) = \frac{\text{Cut}(A, B)}{\text{Assoc}(A, V)} + \frac{\text{Cut}(B, A)}{\text{Assoc}(B, V)}.$$

The goal of this paper is to provide a precise probabilistic interpretation of this measure, which holds when the clusters A and B are “well separated” in the sense defined below.

2.2. Notation. For any $i \in V$, denote by

$$(2.3) \quad W^{i,A} = \sum_{j \in A} W_{i,j}, \quad W^{i,B} = \sum_{j \in B} W_{i,j}$$

the total affinity between node i and the clusters A and B , respectively. Similarly, denote

$$W^{A,A} = \sum_{i \in A} W^{i,A} = \sum_{i \in A, j \in A} W_{i,j}, \quad W^{A,B} = \sum_{i \in A} W^{i,B} = \sum_{i \in A, j \in B} W_{i,j}.$$

In this notation, the normalized cut measure (2.1) is

$$(2.4) \quad \text{Ncut}(A, B) = \frac{W^{A,B}}{W^{A,A} + W^{A,B}}.$$

With an eye towards a matrix perturbation approach, we define the connectivity parameter as

$$(2.5) \quad \varepsilon = \max_{i \in A} \frac{W^{i,B}}{W^{i,A}}.$$

In this paper we focus on situations where all nodes $i \in A$ are weakly connected to B , namely, for all $i \in A$, $W^{i,B} \ll W^{i,A}$. Equivalently, $\varepsilon \ll 1$, which allows ε to be used as a small parameter in a matrix perturbation analysis.

2.3. The unidirectional exit time $\tau_{A \rightarrow B}$. Following [9, 14], we define a Markov random walk on our graph (V, W) that jumps from node to node at unit time intervals. We define the probability of jumping from node i to node j as $p_{i,j} = W_{i,j}/D_{i,i}$, where $D_{i,i} = \sum_j W_{i,j}$. For our purposes, it will also be useful to consider random walks on subsets of V , such as a random walk constrained to the cluster A , as well as random walks where some of the nodes are *absorbing*.

In particular, for the unidirectional exit time from cluster A to cluster B , we consider a simplification of the weighted graph (V, W) , in which the cluster B is collapsed into a new, single absorbing state b , with no allowed transitions $b \rightarrow A$. Let $V^{A \cup b} = A \uplus b$ be the node set of this simplified graph whose corresponding affinity matrix is

$$W^{A \cup b} = \left(\begin{array}{ccc|c} W_{1,1} & \dots & W_{1,k} & W^{1,B} \\ \vdots & \ddots & \vdots & \vdots \\ W_{k,1} & \dots & W_{k,k} & W^{k,B} \\ \hline & & 0 & 1 \end{array} \right).$$

Consider a random walk on this simplified graph, $(V^{A \cup b}, W^{A \cup b})$, whose corresponding Markov matrix is $(D^{A \cup b})^{-1}W^{A \cup b}$, where $D^{A \cup b}$ is the row-normalization diagonal matrix, $D_{i,i}^{A \cup b} = \sum_j W_{i,j}^{A \cup b}$. Denote the eigenvalues of this matrix by

$$1 = \lambda_1^{A \cup b} \geq \lambda_2^{A \cup b} \geq \dots \geq \lambda_{k+1}^{A \cup b}.$$

Assume that the set A is connected, so that $1 > \lambda_2^{A \cup b}$, and moreover that $\lambda_2^{A \cup b}$ is the second largest eigenvalue in absolute value. The *characteristic exit time* from A to B is defined as the relaxation time of this chain,

$$(2.6) \quad \tau_{A \rightarrow B} = \frac{1}{1 - \lambda_2^{A \cup b}}.$$

3. Main result. Our main result is a relation between the exit time $\tau_{A \rightarrow B}$ and the normalized cut measure (2.1). To this end, we also consider a random walk constrained to stay in the cluster A . Under the assumption that A is connected, this chain is reversible. We denote by λ_2^A its second largest eigenvalue and by

$$(3.1) \quad \Gamma^A = 1 - \lambda_2^A$$

its spectral gap. Again we assume that the second largest eigenvalue λ_2^A is also the second largest in absolute value.

We are finally ready to state our main result.

THEOREM 3.1. *Let G be a symmetric graph with affinity matrix W . Consider a binary partition of the graph $V = A \uplus B$. Assume that the subset A is connected and that for the random walk constrained to remain in A , the second largest eigenvalue λ_2^A is also second largest in absolute value. Denote by $\varepsilon = \max_{i \in A} \frac{W^{i,B}}{W^{i,A}}$ the connectivity parameter.*

1. For any $\varepsilon \geq 0$,

$$(3.2) \quad \frac{1}{\tau_{A \rightarrow B}} \leq Ncut(A, B).$$

2. Assume that

$$(3.3) \quad 0 < \varepsilon < \frac{\Gamma^A}{8}.$$

Then

$$(3.4) \quad Ncut(A, B) = \frac{1}{\tau_{A \rightarrow B}} + \varepsilon^2 \mathcal{R}(\varepsilon),$$

where

$$0 \leq \mathcal{R}(\varepsilon) \leq \frac{16}{\Gamma^A}.$$

COROLLARY 3.1. Denote $\varepsilon = \max\{\max_{i \in A} \frac{W^{i,B}}{W^{i,A}}, \max_{i \in B} \frac{W^{i,A}}{W^{i,B}}\}$ and assume that

$$0 < \varepsilon < \frac{1}{8} \min\{\Gamma^A, \Gamma^B\}.$$

Then

$$(3.5) \quad MNcut(A, B) = \frac{1}{\tau_{A \rightarrow B}} + \frac{1}{\tau_{B \rightarrow A}} + \varepsilon^2 \mathcal{R}(\varepsilon),$$

where

$$0 \leq \mathcal{R}(\varepsilon) \leq 16 \left(\frac{1}{\Gamma^A} + \frac{1}{\Gamma^B} \right).$$

4. Discussion. For a binary partition, (3.5) gives a novel probabilistic interpretation to the normalized cut measure in terms of characteristic *exit times* from each of the two subsets. For a partition into s disjoint subsets, $V = \bigsqcup_{j=1}^s A_j$, (3.4) generalizes to a relation between the multiway normalized cut measure and the exit times from each of the s clusters,

$$(4.1) \quad MNcut\left(\bigsqcup A_j\right) = \sum_{j=1}^s \frac{1}{\tau_{A_j \rightarrow A_j^c}} + \varepsilon^2 \mathcal{R}(\varepsilon),$$

where $A_j^c = V \setminus A_j$ and a remainder term $\mathcal{R}(\varepsilon) \geq 0$ with a bound analogous to the one in the binary case.

4.1. Limitations. The key assumption limiting the applicability of our result is (3.3), which requires that $\varepsilon = \max_i W^{i,B}/W^{i,A} < (1 - \lambda_2^A)/8$. Hence, our result holds only when A and B are weakly connected, compared to the internal spectral gap of a random walk constrained to remain in A .

In addition, while a small value of ε implies a small normalized cut, since $Ncut(A, B) \leq \varepsilon$, the converse is in general false: it is possible to construct examples where the normalized cut is small yet ε is large. Indeed ε is a “worse-case” rather than an “on-average” measure of separation between clusters. Informally, a small value of ε means that the boundary between the clusters A and B is pronounced, in the sense that each node in A has a much stronger overall connection to nodes in A than to nodes in B .

4.2. Sharpness of the upper bound. A closer inspection of the proof shows that in fact it gives a stronger bound on the residual $\mathcal{R}(\varepsilon)$ from (3.4).

To describe this sharper bound, consider a Markov chain X_t on $A \cup b$ that starts from the stationary distribution π inside A ,

$$\pi_i = \begin{cases} W^{i,A}/W^{A,A}, & i \in A, \\ 0, & i = b. \end{cases}$$

Then the one-step transition probability to the absorbing state b is

$$(4.2) \quad \mathbb{P}_{A \rightarrow B} = \sum_{i \in A} \mathbb{P}(X_1 = b | X_0 = i) \pi_i = \sum_{i \in A} \left[\frac{W^{i,B}}{W^{i,A} + W^{i,B}} \cdot \frac{W^{i,A}}{W^{A,A}} \right].$$

To factor out the small connectivity magnitude ε , we write $\mathbb{P}_{A \rightarrow B} = \varepsilon \mathbb{Q}_{A \rightarrow B}$. By definition, this quantity satisfies that $\mathbb{Q}_{A \rightarrow B} \leq 1$ and measures the one-step absorption probability in units of ε when the Markov chain on A is in its stationary distribution. Then an improved bound on the residual is

$$|\mathcal{R}(\varepsilon)| \leq \frac{16}{\Gamma^A} \min\{2\mathbb{Q}_{A \rightarrow B}, 1\}.$$

Finally, let us remark that the coefficient 16 above is by no means sharp. For example, in the numerical experiment described below, the actual residual observed was approximately $\frac{0.2}{\Gamma^A} \varepsilon^2$.

4.3. Comparison to previous spectral bounds. Several authors derived the following lower bound on the multiway normalized cut of a partition into s clusters $V = \uplus_{j=1}^s A_j$:

$$(4.3) \quad MNcut \left(\uplus_{j=1}^s A_j \right) \geq \sum_{j=1}^s (1 - \lambda_j(V)),$$

where $\lambda_j(V)$ are the eigenvalues of the *global* random walk matrix on the whole graph G [10, 2].

In contrast, our result (4.1) is different, as it provides (under suitable cluster separability conditions) both a lower and a precise upper bound on $MNcut(V)$ in terms of *local* properties of the graph: the between cluster connectivity, internal relaxation times within each cluster, and the characteristic exit times from them. These quantities depend on eigenvalues of various normalized submatrices of G , which are in general different from the global eigenvalues $\lambda_j(V)$.

4.4. Insensitivity to internal structure. One implication of our result is that when A and B are weakly connected, even though $Ncut(A, B)$ contains a normalization of the cut by the volume of a cluster, it nonetheless considers mainly the exit time from A , without taking into account its internal characteristics. For example, a set A may have a very small normalized cut, $Ncut(A, A^c) \ll 1$, and yet contain a disconnected or weakly connected subgraph, but this structure is not captured by the normalized cut measure. As discussed in [12], this leads to some limitations of the normalized cut measure in discovering clusters in graphs.

4.5. Approximation of exit times by *Ncut*. In the study of weakly coupled stochastic dynamical systems, a quantity of interest is the characteristic exit time from a given set of nodes. At low temperatures corresponding to extremely small values of ε , numerical methods for calculating $\lambda_2^{A \cup b}$ and hence $\tau_{A \rightarrow B}$ may be unstable as $\lambda_2^{A \cup b}$ becomes very close to 1. In contrast, the *Ncut* measure can still be calculated efficiently. Our result thus provides a priori guarantees for the accuracy of this approximation to $\lambda_2^{A \cup b}$. Similarly, the *Ncut* approximation can be computed for extremely large graphs, for which eigenvalue computations may be quite challenging or even no longer feasible.

4.6. Analogy to continuum diffusion. Finally, we mention that our result can be viewed as a discrete analogue of known results regarding the second eigenvalue for a continuum diffusion process with two metastable states [8]. Consider a particle performing diffusion in a multidimensional potential with two wells separated by a high potential barrier. In this case, asymptotically as temperature tends to zero (equivalently in our case $\varepsilon \rightarrow 0$), the second (nontrivial) eigenvalue μ_2 of the Fokker–Planck diffusion operator with two metastable states A, B corresponding to the two potential wells is approximately given by

$$\mu_2 \approx \frac{1}{\tau_{A \rightarrow B}} + \frac{1}{\tau_{B \rightarrow A}}.$$

By analogy, this suggests that when there are two well-defined clusters in a discrete finite graph, the second eigenvalue of the whole graph is close to (1.1). A detailed study of this issue, which is related to isoperimetric and Cheeger inequalities on graphs, is an interesting problem but is beyond the scope of this paper.

5. A numerical example. One way to evaluate the tightness of the upper bound (3.4) is to compare the residual $Ncut(A, B) - 1/\tau_{A \rightarrow B}$ with our upper bound of $\frac{16}{\Gamma_A} \varepsilon^2$ in a numerical simulation. To perform this comparison even for very small values of ε where eigenvalue calculation of $\lambda_2^{A \cup b}$ may be inaccurate, we construct a simple class of graphs where $\lambda_2^{A \cup b}$ can be calculated analytically. This allows for a comparison of the true residual against our upper bound, for values of ε close to machine precision.

Assume that the cluster A is a d -regular graph on k nodes, drawn from the uniform distribution, so that the spectral gap Γ^A is reasonably large. When all vertices of A have the same exit probability α to the absorbing state b , it is easy to see that Theorem 3.1 holds with no approximation error, namely, $Ncut(A, b) = 1 - \lambda_2^{A \cup b} = 1/\tau_{A \rightarrow b}$.

A simple example with nonconstant exit probabilities is when these are equal to α on one half of A , and to $r\alpha$ on the other half of A , for some $r > 0$. Consider then two d -regular graphs $(V^{(i)}, W^{(i)})$, $i = 1, 2$, sampled from the uniform distribution on d -regular graphs. Denote $V^{(1)} = \{1, \dots, k\}$ and $V^{(2)} = \{k + 1, \dots, 2k\}$, and let $b = 2k + 1$ be the absorbing state. Define an affinity matrix W on $V = \{1, \dots, 2k + 1\}$ by

$$W_{i,j} = \begin{cases} 1 & W_{i,j}^{(1)} = 1 \text{ or } W_{i,j}^{(2)} = 1 \text{ or } |i - j| = k, \\ \alpha & 1 \leq i \leq k, j = 2k + 1, \\ r\alpha & k + 1 \leq i \leq 2k, j = 2k + 1 \end{cases}$$

so that each node in $V^{(1)}$ is connected to a corresponding node in $V^{(2)}$, to its neighbors in $V^{(1)}$, and to the absorbing state. Note that $\varepsilon = r\alpha/(d + 1)$ and that $\lambda_2^{A \cup b}(\alpha)$ is

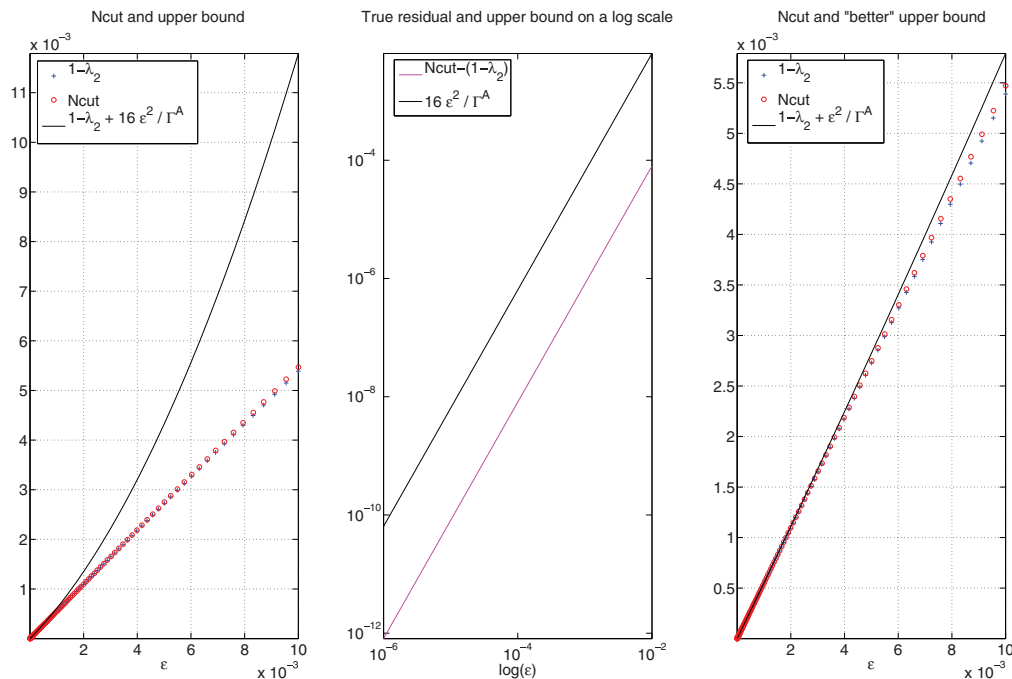


FIG. 5.1. Numerical comparison of the normalized cut and the inverse exit time approximation.

the second eigenvalue of the 3×3 Markov matrix

$$\begin{pmatrix} \frac{d}{d+1+\alpha} & \frac{1}{d+1+\alpha} & \frac{\alpha}{d+1+\alpha} \\ \frac{1}{d+1+r\alpha} & \frac{d}{d+1+r\alpha} & \frac{r\alpha}{d+1+r\alpha} \\ 0 & 0 & 1 \end{pmatrix},$$

allowing us to calculate it analytically.

Figure 5.1 summarizes the result of this experiment for $k = 50$, $d = 7$, $r = 10$, and $\varepsilon \in [10^{-6}, 10^{-2}]$. We compared the exact value of the residual $Ncut(A, b) - \frac{1}{\tau_{A \rightarrow b}} = Ncut(A, b) - (1 - \lambda_2^{A \cup b})$ with the upper bound $\frac{16}{\Gamma^A} \varepsilon^2$. In the case under study, the upper bound is informative yet pessimistic (left panel), as the approximation $Ncut \approx 1 - \lambda_2^{A \cup b}$ is quite accurate. The actual residual $Ncut(A, b) - (1 - \lambda_2^{A \cup b})$ is in this case exactly quadratic in ε (middle panel, log scale), and our quadratic bound is pessimistic by a multiplicative factor of roughly 80. Replacing the constant 16 in our main result, which is presumably not tight due to limitations of our proof technique, with the constant 1, we also compared the residual $Ncut(A, b) - (1 - \lambda_2^{A \cup b})$ with the quantity ε^2 / Γ^A (right panel). The reader is invited to recreate this experiment, replacing the values k , d , r , and α with arbitrary values. The MATLAB script and the original 7-regular graph used here are available on the author's homepage, <http://www.stanford.edu/~gavish/documents/ncut.simulation.m>.

6. Proof of the main result. To simplify the proof, we first introduce the following notation. We write \mathbf{d}_{in} for the vector $(W^{i,A})_{i \in A}$ and \mathbf{d}_{out} for the vector $\varepsilon^{-1}(W^{i,B})_{i \in A}$. Hence $\mathbf{d}_{in}(i)$ is the in-degree of a node $i \in A$, whereas $\mathbf{d}_{out}(i)$ is its out-degree scaled by the connectivity parameter ε , so that $\mathbf{d}_{out}(i) = O(1)$. Similarly,

we define $d_{in} = \sum_{i \in A} \mathbf{d}_{in}(i) = W^{A,A}$ and $d_{out} = \sum_{i \in A} \mathbf{d}_{out}(i) = \varepsilon^{-1}W^{A,B}$. Our main object of interest, $\lambda_2^{A \cup b}$, will be denoted simply by λ_2 , whereas the *Ncut* measure is now given by

$$Ncut(A, B) = \frac{W^{A,B}}{W^{A,A} + W^{A,B}} = \frac{\varepsilon d_{out}}{d_{in} + \varepsilon d_{out}}.$$

For future use, we denote $D_{in} = \text{diag}(\mathbf{d}_{in}(1), \dots, \mathbf{d}_{in}(k))$, $D_{out} = \text{diag}(\mathbf{d}_{out}(1), \dots, \mathbf{d}_{out}(k))$, and $D(\varepsilon) = D_{in} + \varepsilon D_{out}$, so that

$$D^{A \cup b} = \begin{pmatrix} D_{in} + \varepsilon D_{out} & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} D(\varepsilon) & 0 \\ 0 & 1 \end{pmatrix}.$$

Finally, let W^A be the upper $k \times k$ minor of W , namely, $W_{i,j}^A = W_{i,j}$, for $1 \leq i, j \leq k$.

6.1. Passing to a generalized eigenvalue problem. Recall that $\lambda_2^{A \cup b}$ is the second largest root of the characteristic polynomial $\det((D^{A \cup b})^{-1}W^{A \cup b} - \lambda I) = 0$ or, equivalently, of $\det(W^{A \cup b} - \lambda D^{A \cup b}) = 0$. Expanding the latter determinant by the last row of $W^{A \cup b}$ gives

$$\det(W^{A \cup b} - \lambda D^{A \cup b}) = \pm(1 - \lambda) \cdot \det(W^A - \lambda D(\varepsilon)).$$

Since $\lambda_2^{A \cup b} < 1$, it is enough to consider the smaller $k \times k$ matrices $D(\varepsilon)$ and W^A . Obviously, $\mathbf{x} \neq \mathbf{0}$ is an eigenvector of $D(\varepsilon)^{-1}W^A$ corresponding to an eigenvalue λ if and only if

$$(6.1) \quad W^A \mathbf{x} = \lambda D(\varepsilon) \mathbf{x}.$$

When (6.1) holds, we say that λ is a generalized eigenvalue of the matrix pair $(W^A, D(\varepsilon))$. Clearly, the solutions to (6.1) are $\lambda_2 \equiv \lambda_2^{A \cup b} \geq \dots \geq \lambda_{k+1}^{A \cup b}$.

6.2. Lower bound for *Ncut* (A, B). The lower bound (3.2) follows from a min-max characterization of generalized eigenvalues [18, part VI, Corollary 1.16, p. 281], whereby the largest eigenvalue of the matrix pair $(W^A, D(\varepsilon))$ is given by

$$(6.2) \quad \lambda_2^{A \cup b} = \max_{\mathbf{x} \neq \mathbf{0}} \frac{\mathbf{x}^\top W^A \mathbf{x}}{\mathbf{x}^\top D(\varepsilon) \mathbf{x}} \geq \frac{\mathbf{1}^\top W^A \mathbf{1}}{\mathbf{1}^\top D(\varepsilon) \mathbf{1}} = \frac{d_{in}}{d_{in} + \varepsilon d_{out}} = 1 - Ncut(A, B).$$

We note that this simple inequality can also be derived by other methods. For example, from a lemma by Ky Fan [10, 2] it follows that $Ncut(A, B) \geq 2 - \lambda_1^{A \cup b} - \lambda_2^{A \cup b} = 1 - \lambda_2^{A \cup b}$.

6.3. Upper bound. The proof of the upper bound relies on perturbation theory of generalized eigenvalues [18, Chap. 6], where we treat the matrix pair above as a perturbed pair:

$$(W^A, D(\varepsilon)) = (W^A, D_{in}) + \varepsilon(0, D_{out}).$$

Two facts simplify the perturbation analysis to follow. First, the perturbation does not affect the left matrix W^A . Second, the right-hand side matrices on both the perturbed and unperturbed pair are diagonal.

We will need the following key lemma, which is proved in the appendix. Recall the notation $\mathbb{Q}_{A \rightarrow B} = \varepsilon^{-1}\mathbb{P}_{A \rightarrow B}$, as defined in (4.2).

LEMMA 6.1. *If $\varepsilon < \frac{\Gamma^A}{8}$, then there exists $\beta \in \mathbb{R}$ with $|\beta| \leq \frac{8}{\Gamma^A} \min \{2\mathbb{Q}_{A \rightarrow B}, 1\}$ such that*

$$(6.3) \quad \Lambda = \frac{1}{1 + \varepsilon \frac{d_{out}}{d_{in}} + \varepsilon^2 \beta}$$

is a generalized eigenvalue of the matrix pair $(W^A, D(\varepsilon))$.

Given our assumption that ε satisfies (3.3), Lemma 6.1 holds. Hence, let β be such that $\Lambda = (1 + \varepsilon \frac{d_{out}}{d_{in}} + \varepsilon^2 \beta)^{-1}$ is a generalized eigenvalue of the matrix pair $(W^A, D(\varepsilon))$. We now conclude the proof in two steps, from which the theorem readily follows:

1. We show that $Ncut(A, B) = 1 - \Lambda + \varepsilon^2 \mathcal{R}(\varepsilon)$, with a suitable bound on $\mathcal{R}(\varepsilon)$.
2. We then show that Λ is this pair's largest eigenvalue, so that necessarily $\Lambda = \lambda_2$.

For the first step, we rewrite (6.3) as

$$(6.4) \quad \Lambda = \frac{d_{in}}{d_{in} + \varepsilon d_{out}} \left(\frac{1}{1 + \varepsilon^2 \beta \cdot \frac{d_{in}}{d_{in} + \varepsilon d_{out}}} \right).$$

Now, for any $\delta \in [-\frac{1}{2}, \frac{1}{2}]$ we have $\frac{1}{1+\delta} \leq 1 + 2|\delta|$. As $\varepsilon < \frac{\Gamma^A}{8}$ and $|\beta| \leq \frac{8}{\Gamma^A}$, obviously

$$\varepsilon^2 |\beta| \cdot \frac{d_{in}}{d_{in} + \varepsilon d_{out}} \leq \left(\frac{\Gamma^A}{8} \right)^2 \cdot |\beta| \leq \frac{1}{2}.$$

Hence,

$$(6.5) \quad \frac{1}{1 + \varepsilon^2 \beta \cdot \frac{d_{in}}{d_{in} + \varepsilon d_{out}}} \leq 1 + 2\varepsilon^2 |\beta| \frac{d_{in}}{d_{in} + \varepsilon d_{out}} \leq 1 + 2\varepsilon^2 |\beta|.$$

Combining (6.4) and (6.5) gives that

$$\Lambda = \frac{1}{1 + \varepsilon \frac{d_{out}}{d_{in}} + \varepsilon^2 \beta} = \frac{d_{in}}{d_{in} + \varepsilon d_{out}} + \varepsilon^2 \mathcal{R}(\varepsilon) = 1 - Ncut(A, B) + \varepsilon^2 \mathcal{R}(\varepsilon),$$

with

$$|\mathcal{R}(\varepsilon)| \leq \frac{16}{\Gamma^A} \cdot \min \{2\mathbb{Q}_{A \rightarrow B}, 1\}$$

as required.

For the second step, we need to show that if $\varepsilon < \frac{\Gamma^A}{8}$, then Λ is indeed the largest eigenvalue, λ_2 . This is achieved by proving that $\lambda_3^{A \cup b}$, the next largest eigenvalue of $(W^A, D(\varepsilon))$, is strictly smaller than Λ . To this end, we appeal to [18, part VI, Theorem 3.2, p. 314] whereby if

$$(6.6) \quad \zeta = \varepsilon \cdot \max_{\|\mathbf{x}\|_2=1} \frac{\mathbf{x}^\top D_{out} \mathbf{x}}{\sqrt{(\mathbf{x}^\top W^A \mathbf{x})^2 + (\mathbf{x}^\top D_{in} \mathbf{x})^2}} < 1,$$

then

$$(6.7) \quad \frac{|\lambda_3^{A \cup b} - \lambda_2^A|}{\sqrt{(\lambda_2^A)^2 + 1} \cdot \sqrt{(\lambda_3^{A \cup b})^2 + 1}} \leq \zeta.$$

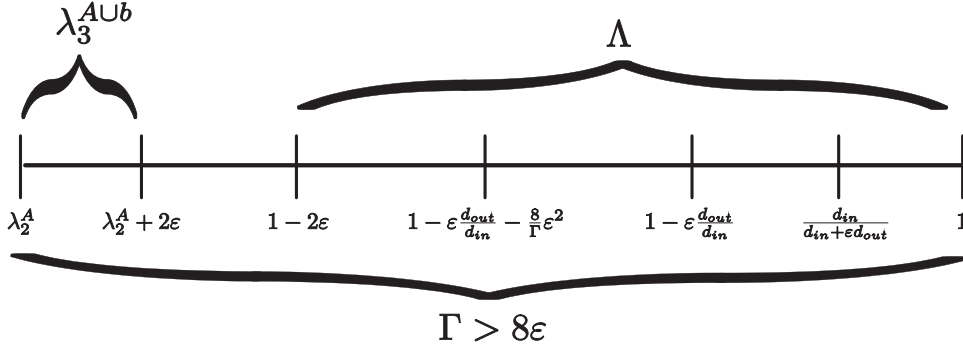


FIG. 6.1. Localization intervals of various eigenvalues.

Here as before, λ_2^A is the second eigenvalue of $(W^A, D(0))$. To see that the condition $\zeta < 1$ holds, observe that from (6.6) we have

$$\zeta \leq \varepsilon \cdot \max_{\|\mathbf{x}\|_2=1} \frac{\mathbf{x}^\top D_{out} \mathbf{x}}{\mathbf{x}^\top D_{in} \mathbf{x}} = \varepsilon \cdot \max_i \frac{\mathbf{d}_{out}(i)}{\mathbf{d}_{in}(i)} = \varepsilon < 1.$$

We thus deduce from (6.7) that

$$\frac{|\lambda_3^{A\cup b} - \lambda_2^A|}{2} \leq \frac{|\lambda_3^{A\cup b} - \lambda_2^A|}{\sqrt{(\lambda_2^A)^2 + 1} \cdot \sqrt{(\lambda_3^{A\cup b})^2 + 1}} \leq \zeta \leq \varepsilon,$$

and in particular,

$$(6.8) \quad \lambda_3^{A\cup b} \leq \lambda_2^A + 2\varepsilon.$$

To complete the second step, as illustrated in Figure 6.1, we now localize Λ and $\lambda_3^{A\cup b}$ in two disjoint intervals, which implies that necessarily $\lambda_3^{A\cup b} < \Lambda$. To this end, first observe that for all $\delta \in [-\frac{1}{2}, \frac{1}{2}]$, we have $1 - |\delta| \leq \frac{1}{1+\delta}$. Similarly to (6.5), this implies that

$$\Lambda = \frac{d_{in}}{d_{in} + \varepsilon d_{out}} \left(\frac{1}{1 + \varepsilon^2 \beta \cdot \frac{d_{in}}{d_{in} + \varepsilon d_{out}}} \right) \geq \frac{d_{in}}{d_{in} + \varepsilon d_{out}} \left(1 - \varepsilon^2 |\beta| \cdot \frac{d_{in}}{d_{in} + \varepsilon d_{out}} \right),$$

so that $\Lambda \geq \frac{d_{in}}{d_{in} + \varepsilon d_{out}} - \varepsilon^2 |\beta|$. Now, clearly

$$\frac{d_{in}}{d_{in} + \varepsilon d_{out}} = \frac{1}{1 + \varepsilon \frac{d_{out}}{d_{in}}} \geq 1 - \varepsilon \frac{d_{out}}{d_{in}} \geq 1 - \varepsilon.$$

Finally, since $|\beta| \leq \frac{8}{\Gamma^A}$ and $\varepsilon < \frac{\Gamma^A}{8}$,

$$(6.9) \quad \Lambda \geq \frac{d_{in}}{d_{in} + \varepsilon d_{out}} - \varepsilon^2 |\beta| \geq 1 - \varepsilon - \varepsilon = 1 - 2\varepsilon.$$

Combining (6.9), (6.8), and the fact that $\Gamma^A = 1 - \lambda_2^A \geq 8\varepsilon$, we obtain

$$\lambda_3^{A\cup b} \leq 1 - \Gamma^A + 2\varepsilon \leq 1 - 8\varepsilon + 2\varepsilon = 1 - 6\varepsilon < 1 - 2\varepsilon \leq \Lambda.$$

Hence, the eigenvalue Λ is strictly larger than $\lambda_3^{A \cup b}$, namely, $\Lambda = \lambda_2^{A \cup b}$. The requested upper bound and Theorem 3.1 follow. \square

Appendix. Lemma 6.1 finds an eigenvalue for the perturbed matrix $(W^A, D(\varepsilon)) = (W^A, D_{in}) + \varepsilon(0, D_{out})$. To prove the lemma, following [18], we introduce a tool for studying the effect of a diagonal perturbation on generalized eigenvalues.

Let the spectrum of the reversible Markov matrix $D_{in}^{-1}W^A$ be

$$(A.1) \quad 1 = \lambda_1^A > \lambda_2^A \geq \lambda_3^A \geq \dots \geq \lambda_k^A,$$

and define

$$(A.2) \quad A_2 = \begin{pmatrix} \lambda_2^A & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \lambda_k^A \end{pmatrix}.$$

Equip the product space $\mathbb{R}^{k-1} \times \mathbb{R}^{k-1}$ with the norm

$$\|(\mathbf{v}, \mathbf{w})\|_{\mathcal{F}} = \max \{ \|\mathbf{v}\|_2, \|\mathbf{w}\|_2 \},$$

where $\mathbf{v}, \mathbf{w} \in \mathbb{R}^{k-1}$ and $\|\cdot\|_2$ is the usual Euclidean norm.

We now define an operator $T : \mathbb{R}^{k-1} \times \mathbb{R}^{k-1} \rightarrow \mathbb{R}^{k-1} \times \mathbb{R}^{k-1}$ by

$$T(\mathbf{v}, \mathbf{w}) = (\mathbf{w} + A_2\mathbf{v}, \mathbf{w} + \mathbf{v}),$$

and its associated quantity $dif[A_2]$ by

$$(A.3) \quad dif[A_2] = \inf_{\|(\mathbf{v}, \mathbf{w})\|_{\mathcal{F}}=1} \|T(\mathbf{v}, \mathbf{w})\|_{\mathcal{F}}.$$

The following auxiliary lemma relates this quantity to the spectral gap of A .

LEMMA A.1. Assume that $\lambda_2^A = \max_{j \geq 2} |\lambda_j^A|$. Then,

$$dif[A_2] = \frac{1 - \lambda_2^A}{2} = \frac{\Gamma^A}{2}.$$

Proof. Let (\mathbf{v}, \mathbf{w}) be a pair with $\|(\mathbf{v}, \mathbf{w})\|_{\mathcal{F}} = 1$. Therefore, either $\|\mathbf{w}\|_2 = 1$ or $\|\mathbf{v}\|_2 = 1$. When $\|\mathbf{w}\|_2 = 1$,

$$\|T(\mathbf{v}, \mathbf{w})\|_{\mathcal{F}} \geq \|\mathbf{w} + A_2\mathbf{v}\|_2 \geq \|\mathbf{w}\|_2 - \|A_2\mathbf{v}\|_2.$$

Since $\|\mathbf{v}\|_2 \leq 1$ and $\|A_2\| = \lambda_2^A$, it follows that

$$\|T(\mathbf{v}, \mathbf{w})\|_{\mathcal{F}} \geq 1 - \lambda_2^A = \Gamma^A.$$

Next, consider the case where $\|\mathbf{v}\|_2 = 1$. Then by definition $\|A_2\mathbf{v}\|_2 \leq \lambda_2^A < 1$. It is easy to see that inside the unit ball $\|\mathbf{w}\|_2 \leq 1$, the minimizer of $\|T(\mathbf{v}, \mathbf{w})\|$ is simply the average $\mathbf{w} = -\frac{1}{2}(\mathbf{v} + A_2\mathbf{v})$, which gives

$$\|T(\mathbf{v}, \mathbf{w})\|_{\mathcal{F}} = \|\mathbf{v} - A_2\mathbf{v}\|_2/2.$$

To conclude the proof we note that since A_2 is a diagonal matrix with largest entry λ_2^A in position $(1, 1)$, the quantity $\|\mathbf{v} - A_2\mathbf{v}\|$ is minimized at the vector $\mathbf{v} = \mathbf{e}_1$. In this case,

$$\frac{\|\mathbf{e}_1 - A_2\mathbf{e}_1\|_2}{2} = \frac{1 - \lambda_2^A}{2} = \frac{\Gamma^A}{2}. \quad \square$$

Proof of Lemma 6.1. Recall that the matrix pair $(W^A, D(\varepsilon))$ is a sum of the pair (W^A, D_{in}) and a diagonal perturbation:

$$(W^A, D(\varepsilon)) = (W^A, D_{in}) + \varepsilon(0, D_{out}).$$

For the unperturbed pair, observe that its generalized eigenvalues are all real, since the matrix $D_{in}^{-1}W^A$ is similar to the symmetric matrix $D_{in}^{\frac{1}{2}}(D_{in}^{-1}W^A)D_{in}^{-\frac{1}{2}} = D_{in}^{-\frac{1}{2}}W^AD_{in}^{-\frac{1}{2}}$. Furthermore, since $D_{in}^{-1}W^A$ is the transition matrix of an irreducible Markov chain, the largest eigenvalue in (A.1) is $\lambda_1^A = 1$, and the remaining eigenvalues are all strictly smaller than 1 in absolute value.

Now, the vector of length k ,

$$\mathbf{x} = \frac{1}{\sqrt{d_{in}}} \cdot D_{in}^{\frac{1}{2}}(1, \dots, 1)^\top,$$

is an eigenvector of $D_{in}^{-\frac{1}{2}}W^AD_{in}^{-\frac{1}{2}}$ corresponding to the eigenvalue $\lambda_1^A = 1$. Let

$$U = \begin{pmatrix} | & | & \cdots & | \\ \mathbf{x} & \mathbf{u}_1 & \cdots & \mathbf{u}_{k-1} \\ | & | & \cdots & | \end{pmatrix}$$

be the $k \times k$ orthogonal matrix that diagonalizes $D_{in}^{-\frac{1}{2}}W^AD_{in}^{-\frac{1}{2}}$, namely,

$$D_{in}^{-\frac{1}{2}}W^AD_{in}^{-\frac{1}{2}} = U \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \lambda_2^A & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & \lambda_k^A \end{pmatrix} U^\top.$$

Now define $X = D_{in}^{-\frac{1}{2}}U$ and decompose X to blocks as $X = (X_1 X_2)$, where

$$X_1 = \frac{1}{\sqrt{d_{in}}}(1, \dots, 1)^\top$$

is a $k \times 1$ vector and X_2 is the remaining $k \times (k - 1)$ matrix,

$$X_2 = D_{in}^{-\frac{1}{2}} \begin{pmatrix} | & & | \\ \mathbf{u}_1 & \cdots & \mathbf{u}_{k-1} \\ | & & | \end{pmatrix}.$$

This gives a *simultaneous* diagonalization of the two matrices W^A and D_{in} , whereby in matrix block notation,

$$\begin{pmatrix} X_1^\top \\ X_2^\top \end{pmatrix} \cdot W^A \cdot (X_1 X_2) = \begin{pmatrix} 1 & 0 \\ 0 & A_2 \end{pmatrix}$$

and

$$\begin{pmatrix} X_1^\top \\ X_2^\top \end{pmatrix} \cdot D_{in} \cdot (X_1 X_2) = \begin{pmatrix} 1 & 0 \\ 0 & \mathbf{I}_{k-1} \end{pmatrix}.$$

Next, we turn to the perturbation pair, $(0, D_{out})$. Again using matrix block notation, we have

$$\begin{pmatrix} X_1^\top \\ X_2^\top \end{pmatrix} \cdot D_{out} \cdot (X_1 X_2) = \begin{pmatrix} \frac{d_{out}}{d_{in}} & \mathbf{d}_{out}^{12} \\ \mathbf{d}_{out}^{21} & D_{out}^{22} \end{pmatrix},$$

where \mathbf{d}_{out}^{12} is a $1 \times (k - 1)$ vector, $\mathbf{d}_{out}^{21} = (\mathbf{d}_{out}^{12})^\top$, and D_{out}^{22} is a $(k - 1) \times (k - 1)$ matrix. It is straightforward to verify that

$$(A.4) \quad \mathbf{d}_{out}^{12} = \frac{1}{\sqrt{d_{in}}} \left(\frac{\mathbf{d}_{out}(1)}{\sqrt{d_{in}(1)}}, \dots, \frac{\mathbf{d}_{out}(k)}{\sqrt{d_{in}(k)}} \right) \begin{pmatrix} | & & | \\ \mathbf{u}_1 & \cdots & \mathbf{u}_{k-1} \\ | & & | \end{pmatrix}.$$

The amount by which the eigenvalue λ_2^A is perturbed as ε departs from 0 depends primarily on the norms of \mathbf{d}_{out}^{12} and D_{out}^{22} . For \mathbf{d}_{out}^{12} we have that

$$\|\mathbf{d}_{out}^{12}\|_2^2 \leq \frac{1}{d_{in}} \cdot \sum_{i=1}^k \frac{\mathbf{d}_{out}^2(i)}{d_{in}(i)} = \sum_{i=1}^k \frac{\mathbf{d}_{out}^2(i)}{d_{in}^2(i)} \cdot \frac{d_{in}(i)}{d_{in}} = \sum_{i=1}^k \frac{\mathbf{d}_{out}^2(i)}{d_{in}^2(i)} \pi^A(i),$$

where $\pi^A(i) = \frac{d_{in}(i)}{d_{in}}$ is the stationary distribution of a random walk inside A . Furthermore, since $\max_i \frac{\mathbf{d}_{out}(i)}{d_{in}(i)} = 1$, it follows that for all $i \in A$, $d_{in}(i) \geq \frac{1}{2}[d_{in}(i) + \varepsilon d_{out}(i)]$. Hence,

$$\|\mathbf{d}_{out}^{12}\|_2^2 \leq \sum_{i=1}^k \frac{\mathbf{d}_{out}^2(i)}{d_{in}^2(i)} \cdot \pi^A(i) \leq \sum_{i=1}^k \frac{2\mathbf{d}_{out}(i)}{d_{in}(i) + \varepsilon d_{out}(i)} \cdot \pi^A(i) = 2\mathbb{Q}_{A \rightarrow B}.$$

Since by (A.4), clearly $\|\mathbf{d}_{out}^{12}\|_2^2 \leq 1$, we can write

$$(A.5) \quad \|\mathbf{d}_{out}^{12}\|_2^2 \leq \min \{2\mathbb{Q}_{A \rightarrow B}, 1\}.$$

As for D_{out}^{22} , it is easy to verify that since U is orthogonal, $\|D_{out}^{22}\|_F \leq 1$.

Now, as the generalized eigenvalues of the pair (A_2, \mathbf{I}_{k-1}) are strictly smaller than 1 in absolute value, the column X_1 spans a simple eigenspace of the unperturbed pair (W^A, D_{in}) . We can thus appeal to the main instrument in our proof, a theorem from matrix perturbation theory for generalized eigenvalue problems [18, part VI, Theorem 2.15, p. 310].

In our setting and our notation, this theorem states that if $\text{dif}[A_2] > 0$ and if

$$(A.6) \quad \frac{\varepsilon\gamma}{\text{dif}[A_2] - \varepsilon\rho} < \frac{1}{2},$$

where

$$\gamma = \|\mathbf{d}_{out}^{12}\|_2 = \|\mathbf{d}_{out}^{21}\|_2, \quad \rho = \frac{d_{out}}{d_{in}} + \|D_{out}^{22}\|_F,$$

then there exists $\mathbf{p} \in \mathbb{R}^{k-1}$ with

$$(A.7) \quad \|\mathbf{p}\|_2 \leq 2 \cdot \frac{\gamma}{\text{dif}[A_2] - \varepsilon\rho}$$

such that the pair

$$\left(1, \left(1 + \varepsilon \frac{d_{out}}{d_{in}} + \varepsilon^2 \mathbf{d}_{out}^{12} \cdot \mathbf{p} \right)^{-1} \right)$$

is a generalized eigenvalue of the matrix pair $(W^A, D(\varepsilon))$.

The first condition in this theorem, namely, $\text{dif}[A_2] > 0$, follows directly from Lemma A.1. For the second condition, we need to bound the constants γ and ρ as follows. According to (A.5) we have that $\gamma = \|\mathbf{d}_{out}^{21}\|_2 = \|\mathbf{d}_{out}^{12}\|_2 \leq 1$. Similarly, $\rho = \frac{d_{out}}{d_{in}} + \|D_{out}^{22}\|_F \leq 2$.

Suppose now that $\varepsilon < \frac{\Gamma^A}{8}$. Then we have $\varepsilon\rho < \frac{\Gamma^A}{4}$. Invoking Lemma A.1, whereby $\text{dif}[A_2] = \frac{\Gamma^A}{2}$, we get that

$$(A.8) \quad \text{dif}[A_2] - \varepsilon\rho > \frac{\Gamma^A}{2} - \frac{\Gamma^A}{4} = \frac{\Gamma^A}{4},$$

and therefore

$$\frac{\varepsilon\gamma}{\text{dif}[A_2] - \varepsilon\rho} < \frac{\varepsilon}{\frac{\Gamma^A}{4}} < \frac{\frac{\Gamma^A}{8}}{\frac{\Gamma^A}{4}} = \frac{1}{2},$$

so the second condition of the theorem, condition (A.6), is satisfied.

In conclusion, the theorem above holds, so a vector \mathbf{p} as above exists, with

$$(A.9) \quad \|\mathbf{p}\|_2 \leq \frac{2\gamma}{\text{dif}[A_2] - \varepsilon\rho}.$$

Finally, using (A.9), (A.8), (A.5), and the definition $\gamma = \|\mathbf{d}_{out}^{12}\|_2$, we have

$$\begin{aligned} |\mathbf{d}_{out}^{12} \cdot \mathbf{p}| &\leq \|\mathbf{p}\|_2 \|\mathbf{d}_{out}^{12}\|_2 \leq \frac{2\gamma}{\text{dif}[A_2] - \varepsilon\rho} \|\mathbf{d}_{out}^{12}\|_2 = \frac{2\|\mathbf{d}_{out}^{12}\|_2^2}{\text{dif}[A_2] - \varepsilon\rho} \\ &\leq \frac{2\|\mathbf{d}_{out}^{12}\|_2^2}{\frac{\Gamma^A}{4}} \leq \frac{8}{\Gamma^A} \min\{2\mathbb{Q}_{A \rightarrow B}, 1\}. \end{aligned}$$

This completes the proof of Lemma 6.1. □

Acknowledgments. We thank the anonymous referees for their helpful comments and suggestions.

REFERENCES

- [1] K. E. AVRACHENKOV, J. A. FILA, AND M. HAVIV, *Singular perturbations of Markov chains and decision processes. A survey*, in Markov Decision Processes: Models, Methods, Directions and Open Problems, E. Feinberg and A. Schwartz, eds., Kluwer, Dordrecht, The Netherlands, 2002, pp. 113–152.
- [2] F. BACH AND M. I. JORDAN, *Learning spectral clustering with applications to speech separation*, J. Mach. Learn. Res., 7 (2006), pp. 1963–2001.
- [3] P. DEUFLHARD, W. HUISINGA, A. FISCHER, AND CH. SCHUTTE, *Identification of almost invariant aggregates in reversible nearly uncoupled Markov chains*, Linear Algebra Appl., 315 (2000), pp. 39–59.
- [4] I. S. DHILLON, Y. GUAN, AND B. KULIS, *Kernel k-means: Spectral clustering and normalized cuts*, in Proceedings of the Tenth ACM SIGKDD Conference on Knowledge Discovery and Data Mining, ACM, New York, 2004, pp. 551–556.
- [5] C. DING, X. HE, AND H. D. SIMON, *On the equivalence of nonnegative matrix factorization and spectral clustering*, in Proceedings of the Fifth SIAM International Conference on Data Mining, SIAM, Philadelphia, 2005, pp. 606–610.
- [6] M. HAVIV AND Y. RITOV, *On series expansions and stochastic matrices*, SIAM J. Matrix Anal. Appl., 14 (1993), pp. 670–676.
- [7] R. KANNAN, S. VEMPALA, AND A. VETTA, *On clusterings: good, bad and spectral*, J. ACM, 51 (2004), pp. 497–515.

- [8] B. J. MATKOWSKY AND Z. SCHUSS, *Eigenvalues of the Fokker-Planck operator and the approach to equilibrium for diffusions in potential fields*, SIAM J. Appl. Math., 40 (1981), pp. 242–254.
- [9] M. MEILA AND J. SHI, *A random walks view of spectral segmentation*, in Proceedings of the 8th International Workshop on Artificial Intelligence and Statistics (AISTATS), 2001.
- [10] M. MEILA, S. SHORTREED, AND L. XU, *Regularized spectral learning*, in Proceedings of the 10th International Workshop on Artificial Intelligence and Statistics (AISTATS), 2005.
- [11] C. D. MEYER, *Stochastic complementation, uncoupling Markov chains, and the theory of nearly reducible systems*, SIAM Rev., 31 (1989), pp. 240–272.
- [12] B. NADLER AND M. GALUN, *Fundamental limitations of spectral clustering*, in Advances in Neural Information Processing Systems 19, B. Schölkopf, J. Platt, and T. Hoffman, eds., MIT Press, Cambridge, MA, 2007, pp. 1017–1024.
- [13] B. NADLER, M. GALUN, AND T. GROSS, *Clustering and multi-scale structure of graphs: A scale-invariant diffusion based approach*, in preparation.
- [14] B. NADLER, S. LAFON, R. COIFMAN, AND I. KEVREKIDIS, *Diffusion maps, spectral clustering and eigenfunctions of Fokker-Planck operators*, in Advances in Neural Information Processing Systems 18, Y. Weiss, B. Schölkopf, and J. Platt, eds., MIT Press, Cambridge, MA, 2006, pp. 955–962.
- [15] A. Y. NG, M. I. JORDAN, AND Y. WEISS, *On spectral clustering: Analysis and an algorithm*, in Advances in Neural Information Processing Systems 14, MIT Press, Cambridge, MA, 2002, pp. 849–856.
- [16] P. J. SCHWEITZER, *A survey of aggregation/disaggregation in large Markov chains*, in Numerical Solution of Markov Chain, W. J. Stewart, ed., Marcel Dekker, New York, 1991, pp. 63–88.
- [17] J. SHI AND J. MALIK, *Normalized cuts and image segmentation*, IEEE Trans. Pattern Anal. Mach. Intell., 22 (2000), pp. 888–905.
- [18] G. W. STEWART AND J.-G. SUN, *Matrix Perturbation Theory*, Academic Press, New York, 1990.
- [19] S. YU AND J. SHI, *Multiclass spectral clustering*, in Proceedings of the 9th IEEE International Conference on Computer Vision, IEEE Press, Piscataway, NJ, 2003, pp. 313–319.