Cheeger Inequalities for General Edge-Weighted Directed Graphs

T-H. Hubert Chan, Zhihao Gavin Tang, and Chenzi Zhang

The University of Hong Kong
{hubert,zhtang,czzhang}@cs.hku.hk

Abstract. We consider Cheeger Inequalities for general edge-weighted directed graphs. Previously the directed case was considered by Chung for a probability transition matrix corresponding to a strongly connected graph with weights induced by a stationary distribution. An Eulerian property of these special weights reduces these instances to the undirected case, for which recent results on multi-way spectral partitioning and higher-order Cheeger Inequalities can be applied.
We extend Chung's approach to general directed graphs. In particular, we obtain higher-order Cheeger Inequalities for the following scenarios:
(1) The underlying graph needs not be strongly connected.
(2) The weights can deviate (slightly) from a stationary distribution.

1 Introduction

There have been numerous works relating the expansion properties of an undirected graph with the eigenvalues of its Laplacian [3, 1, 9]. Given an undirected graph with non-negative edge weights, the weight of a vertex is the sum of the weights of its incident edges. Then, the *expansion* $\rho(S)$ of a subset S of vertices is the ratio of the sum of the weights of edges having only one end-point in S to the sum of the weights of vertices in S. The celebrated Cheeger's Inequality [6, 1] relates the smallest expansion of a subset of vertices having at most half the sum of vertex weights with the second smallest eigenvalue of the corresponding normalized Laplacian. Recently, there have been extensions to the case where the expansions of k disjoint subsets are related to the k-th smallest eigenvalue [13].

The notion of expansion can be extended to directed graphs, where the weight of a vertex is the sum of the weights of its out-going edges. Then, the expansion of a subset S is defined with respect to the sum of the weights of edges going out of S. Chung [8] considered the special case for a probability transition matrix whose non-zero entries correspond to the edges of a **strongly connected graph**. The weights of the vertices are chosen according to the (unique) stationary distribution, and the weight of an edge is the probability mass going along the edge under this stationary distribution. Under this specific choice of weights, Chung has proved an analogous Cheeger's Inequality [8] for directed graphs.

In this paper, we explore how this relationship between expansion and spectral properties can be extended to more general cases for directed graphs. In particular, we consider the following cases.

1. The directed graph is not strongly connected.

2. The weights of vertices deviate (slightly) from the stationary distribution.

As we shall explain, each of these cases violates the technical assumptions that are used by Chung to derive the Cheeger Inequality for directed graphs. We explore what expansion notions are relevant in these scenarios, and how to define Laplacians whose eigenvalues can capture these notions.

1.1 Overview of Chung's Approach [8].

All spectral arguments rely on some symmetric matrix, which has the desirable properties of having real eigenvalues and an orthonormal basis of eigenvectors. For an undirected graph (with non-negative edge weights), its normalized Laplacian is a symmetric matrix. To apply spectral analysis on directed graphs, one should consider what the natural candidates for symmetric matrices should be and whether they have any significance. We explain the importance of the technical assumptions made by Chung in the analysis of the transition matrix $\mathbf{P}$ associated with the random walk on some directed graph $G(V, E)$.

(1) Choice of Weights. Suppose $\boldsymbol{\phi} : V \to \mathbb{R}_+$ is a stationary distribution of the transition matrix $\mathbf{P}$. Then, the weights are chosen such that each vertex u has weight $\phi(u)$, and each (directed) edge (u, v) has weight $\phi(u) \cdot P(u, v)$, which is the probability mass going from u to v in one step of the random walk starting from distribution $\boldsymbol{\phi}$.

Suppose the starting vertex u of a random walk is chosen according to distribution $\boldsymbol{\phi}$. The expansion of a subset S has the following meaning: conditioning on the event that u is in S, it is the probability that the next step of the random walk goes out of S.

This notion of expansion can be defined with respect to any distribution on the vertex set V, but the edge weights induced by a stationary distribution has the following *Eulerian* property: for any subset S of vertices, the sum of weights of edges going out of S is the same as that of edges going into S.

Hence, one can consider the underlying undirected graph such that each undirected edge has weight that is the average of those for the corresponding directed edges in each direction. Then, because of the Eulerian property, for any subset S, its expansion in the directed graph defined with respect to the out-going edges is exactly the same as its expansion defined with respect to the undirected graph (with edge weights defined above). Therefore, it suffices to consider the normalized Laplacian of the undirected graph to analyze the expansion properties of the directed graph.

(2) Irreducibility of Transition Matrix. This means that the underlying directed graph with edges corresponding to transitions with non-zero probabilities is strongly connected. Under this assumption, the stationary distribution is unique, and every vertex has a positive mass.

If the directed graph is not strongly connected, a strongly connected component is known as a *sink* if there is no edge going out of it. If there is more than one sink, the stationary distribution is not unique. Moreover, under any stationary distribution, any vertex in a non-sink has probability mass zero. Hence, Chung's method essentially deletes all non-sinks before considering the expansion properties of the remaining graph.

In this paper, we explore ways to consider expansion properties that involve the non-sinks of a directed graph that is not strongly connected.

1.2 Our Contribution

The contribution of this paper is mainly conceptual, and offers an approach to extend Chung's spectral analysis of transition matrices to the scenarios when the underlying directed graph is not strongly connected, or when the vertex weights do not follow a stationary distribution. On a high level, our technique to handle both issues is to add a new vertex to the graph and define additional transition probabilities involving the new vertex such that the new underlying graph is strongly connected, and the expansion properties for the old vertices are also preserved in the new graph. Therefore, Chung's technique can be applied after the transformation. We outline our approaches and results as follows.

(1) Transition matrix whose directed graph is not necessarily strongly connected. Given a transition matrix $\mathbf{P}$ corresponding to a random walk on a graph $G(V, E)$ and a subset $S \subseteq V$ of vertices, we denote by $\mathbf{P}|_S$ the submatrix defined by restricting $\mathbf{P}$ only to the rows and the columns corresponding to S.

It is known [7] that the eigenvalues of $\mathbf{P}$ are the union of the eigenvalues of $\mathbf{P}|_C$ over all strongly connected components C in directed graph G. An important observation is that as long as the strongly connected components and the transition probabilities within a component remain the same, the eigenvalues of $\mathbf{P}$ are independent of the transition probabilities between different strongly connected components. This suggests that it might be difficult to use spectral properties to analyze expansion properties involving edges between different strongly connected components.

Therefore, we propose that it makes sense to consider the expansion properties for each strongly connected component separately. If C is a sink (i.e., there is no edge leaving C), then $\mathbf{P}|_C$ itself is a probability transition matrix, for which Chung's approach can be applied by using the (unique) stationary distribution on C.

However, if C is a non-sink, then there is no stationary distribution for $\mathbf{P}|_C$, because there is non-zero probability mass leaking out of C in every step of the random walk. For the non-trivial case when $|C| \geq 2$, by the Perron-Frobenius Theorem [10], there exists some maximal eigenvalue $\lambda > 0$ with respect to the complex norm, and unique (left) eigenvector $\boldsymbol{\phi}$ with strictly positive coordinates such that $\boldsymbol{\phi}^{\mathrm{T}}\mathbf{P}|_C = \lambda\boldsymbol{\phi}^{\mathrm{T}}$. When $\boldsymbol{\phi}$ is normalized such that all coordinates sum to 1, we say that $\boldsymbol{\phi}$ is the *diluted stationary distribution* of $\mathbf{P}|_C$. It is stationary in the sense that if we start the random walk with distribution $\boldsymbol{\phi}$, then conditioning on the event that the next step remains in C (which has probability λ), we have the same distribution $\boldsymbol{\phi}$ on C.

Hence, we can define the expansion of a subset S in C with respect to the diluted stationary distribution $\boldsymbol{\phi}$. Given a vertex $u \in C$ and a vertex $v \in V$ (that could be outside C), the weight of the edge (u, v) is $\phi(u) \cdot P(u, v)$. Observe that the sum of weights of edges going out of u is $\phi(u)$. Hence, the expansion of a subset S in C are due to edges leaving S that can either stay in or out of the component C.

In order to analyze this notion of expansion using Chung's approach, we construct a strongly connected graph on the component C together with a new vertex v_0, which absorbs all the probabilities leaking out of C, and returns them to C according to the diluted stationary distribution $\boldsymbol{\phi}$. This defines a probability transition matrix $\widehat{\mathbf{P}}$ on the new graph that is strongly connected with various nice properties. For instance, $\widehat{\mathbf{P}}$ has 1 as the maximal eigenvalue with the corresponding left eigenvector formed from the diluted stationary distribution $\boldsymbol{\phi}$ by appending an extra coordinate corresponding to the new vertex with value $1-\lambda$.

One interesting technical result (Lemma 1) is that the new transition matrix $\widehat{\mathbf{P}}$ preserves the spectral properties of $\mathbf{P}|_C$ in the sense that the eigenvalues of $\widehat{\mathbf{P}}$ can be obtained by removing λ from the multi-set of eigenvalues of $\mathbf{P}|_C$ and including 1 and $\lambda - 1$. In other words, other than the removal of λ and the inclusion of 1 and $\lambda - 1$, all other eigenvalues are preserved, even up to their algebraic and geometric multiplicities.

Hence, we can use Chung's approach to define a symmetric Laplacian for $\widehat{\mathbf{P}}$, and use the recent results from Lee et al. [13] on higher-order Cheeger Inequalities to achieve an analogous result for a strongly connected component in a directed graph. In particular, multi-way partition expansion is considered. For a subset C of vertices, we denote:

$\rho_k(C) := \min\{\max_{i\in[k]} \rho(S_i) : S_1, S_2, \ldots, S_k$ are disjoint subsets of C$\}$.

Theorem 1. *Suppose C is a strongly connected component of size n associated with some probability transition matrix, and the expansion $\rho(S)$ of a subset S of vertices within C is defined with respect to the diluted stationary distribution $\boldsymbol{\phi}$ as described above.*

Then, one can define a Laplacian matrix with dimension $(n+1) \times (n+1)$ having eigenvalues $0 = \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_{n+1}$ such that for $1 \leq k \leq n$, we have $\frac{\lambda_k}{2} \leq \rho_k(C) \leq O(k^2) \cdot \sqrt{\lambda_{k+1}}$.

(2) Vertex weights deviate from stationary distribution. Given a transition matrix $\mathbf{P}$, recall that in Chung's approach, the expansion is defined with the careful choice of setting each vertex's weight according to a stationary distribution. We consider the case when the vertex weights $\boldsymbol{\phi} : V \to \mathbb{R}_+$ can deviate from a stationary distribution of $\mathbf{P}$.

Suppose each vertex is assigned a positive weight according to $\boldsymbol{\phi}$. Then, the following parameter measures how much $\boldsymbol{\phi}$ deviates from a stationary distribution:

$$\varepsilon := 1 - \min_{u \in V} \frac{\phi(u)}{\sum_{v \in V} \phi(v) \cdot P(v,u)}.$$

A smaller value of ε means that $\boldsymbol{\phi}$ is closer to a stationary distribution. In particular, if $\varepsilon = 0$, then $\boldsymbol{\phi}$ is a stationary distribution.

Our idea is to first scale down all probabilities in $\mathbf{P}$ by a factor of $(1-\varepsilon)$. We add a new vertex v_0 to absorb the extra ε probability from each existing vertex. Then, we define the transition probabilities from v_0 to the original vertices carefully such that each original vertex u receives the same weight $\phi(u)$ after one step. In other words, we can append a new coordinate corresponding to v_0 to $\boldsymbol{\phi}$ to obtain a stationary distribution $\widehat{\boldsymbol{\phi}}$ for the transition matrix $\widehat{\mathbf{P}}$ of the augmented random walk. Moreover, for each subset $S \subset V$ of the original vertices, the new expansion $\widehat{\rho}(S)$ with respect to $\widehat{\boldsymbol{\phi}}$ and $\widehat{\mathbf{P}}$ can be related to the old expansion $\rho(S)$ as follows: $\widehat{\rho}(S) = (1-\varepsilon)\cdot\rho(S) + \varepsilon$.

Therefore, we can apply Chung's approach to $\widehat{\mathbf{P}}$ and $\widehat{\boldsymbol{\phi}}$ to construct a symmetric Laplacian matrix, whose eigenvalues are related to the expansion properties using the results by Lee et al. [13].

Theorem 2. *Suppose n vertices have positive weights defined by $\phi : V \to \mathbb{R}_+$, and $\varepsilon \geq 0$ is the parameter defined above. Then, there exists a symmetric Laplacian matrix with dimension $(n+1)\times(n+1)$ and eigenvalues $0 = \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_{n+1}$ such that for $1 \leq k \leq n$, we have $\frac{\lambda_k}{2} \leq (1-\varepsilon)\cdot\rho_k(V) + \varepsilon \leq O(k^2)\cdot\sqrt{\lambda_{k+1}}$.*

In Appendix A, we show that if we allow a self-loop at the new vertex v_0 with negative weight, we can slighlty improve the left hand side of the inequality in Theorem 2.

1.3 Related Work

Since the Cheeger's Inequality [6] was introduced in the context of Riemannian geometry, analogous results have been achieved by Alon et al. [1, 3] to relate the expansion of an undirected graph with the smallest positive eigenvalue of the associated Laplacian matrix. The reader is referred to the standard textbook by Chung [9] on spectral graph theory for a more comprehensive introduction of the subject.

As far as we know, the only previous attempt to apply spectral analysis to directed graphs was by Chung [8], who reduced the special case of directed instances induced by stationary distributions into undirected instances. On a high level, our approach is to reduce general directed instances into instances induced by stationary distributions.

Recently, for undirected instances, Lee et al. [13] extended Cheeger's Inequality to relate higher order eigenvalues with multi-way spectral partition. This result was further improved by Kwok et al. [12]. Since Chung's approach [8] made use of the Laplacian induced by an undirected instance, the higher order Cheeger Inequalities can be directly applied to the cases considered by Chung.

The reader can refer to the survey on spectral partitioning by Shewchuck [15], who also mentioned expansion optimization problems with negative edge weights. Other applications of spectral analysis include graph coloring [4, 2], web search [11, 5] clustering [14], image segmentation [16, 17], etc.

2 Preliminaries

We consider a graph $G(V, E)$ with non-negative edge weights $w : E \to \mathbb{R}_+$. In most cases, we consider directed graphs, but we will also use results for undirected graphs. Note that a vertex might have a self-loop (even in an undirected graph).

For a subset $S \subseteq V$, $\partial(S)$ is the set of edges leaving S in a directed graph, whereas in an undirected graph, it includes those edges having exactly one end-point in S (excluding self-loops). Given the edge weights, vertex weights are defined as follows. For each $u \in V$, its weight $w(u)$ is the sum of the weights of its out-going edges (including its self-loop) in a directed graph, whereas in an undirected graph, the edges incident on u are considered.

The *expansion* $\rho(S)$ of a subset S (with respect to w) is defined as:

$$\frac{w(\partial(S))}{w(S)} = \frac{\sum_{e \in \partial(S)} w(e)}{\sum_{u \in S} w(u)}.$$

In this paper, we use bold capital letters (such as $\mathbf{A}$) to denote matrices and bold small letters (such as $\boldsymbol{\phi}$) to denote column vectors. The transpose of a matrix $\mathbf{A}$ is denoted as $\mathbf{A}^{\mathrm{T}}$. For a positive integer n, $\mathbf{I}_n$ is the $n \times n$ identity matrix, and $\mathbf{0}_n$ and $\mathbf{1}_n$ are the all zero's and all one's column vectors, respectively, of dimension n, where the subscript n is omitted if the dimension is clear from context.

Undirected graphs. Suppose $\mathbf{W}$ is the symmetric matrix indicating the edge weights w of an undirected graph of size n, and $\mathbf{\Phi}$ is the diagonal matrix whose diagonal entries correspond to the vertex weights induced by w as described above. Then, the *normalized Laplacian* of $\mathbf{W}$ is $\boldsymbol{\mathcal{L}} := \mathbf{I}_n - \mathbf{\Phi}^{-\frac{1}{2}} \mathbf{W} \mathbf{\Phi}^{-\frac{1}{2}}$.

Multi-way partition expansion is considered in Lee et al. [13] by considering the following parameter. For $C \subseteq V$ and positive integer k, denote

$\rho_k(C) := \min\{\max_{i \in [k]} \rho(S_i) : S_1, S_2, \ldots, S_k$ are disjoint subsets of C$\}$.

The following fact relates the eigenvalues of $\boldsymbol{\mathcal{L}}$ with the multi-way partition expansion with respect to $\mathbf{W}$, which may contain self-loops.

Fact 1. *(Higher Order Cheeger's Inequality [13]) Given a symmetric matrix* $\mathbf{W}$ *indicating the non-negative edge weights of an undirected graph, suppose its normalized Laplacian* $\boldsymbol{\mathcal{L}}$ *as defined above has eigenvalues* $0 = \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$*. Then, for* $1 \leq k \leq n$*, we have:* $\frac{\lambda_k}{2} \leq \rho_k(V) \leq O(k^2) \cdot \sqrt{\lambda_k}$.

Chung's approach [8] to transition matrices. Given a probability transition matrix $\mathbf{P}$ (which is a square matrix with non-negative entries such that every row sums to 1) corresponding to a random walk on vertex set V, and non-negative vertex weight $\boldsymbol{\phi}$, one can define edge weights $w : V \times V \to \mathbb{R}_+$ as $w(u,v) := \phi(u) \cdot P(u,v)$. (Observe that these edge weights induce vertex weights that are consistent with $\boldsymbol{\phi}$.)

One interpretation of Chung's approach is that the vertex weights $\boldsymbol{\phi}$ are chosen to be a stationary distribution of $\mathbf{P}$, i.e., $\boldsymbol{\phi}^{\mathrm{T}} \mathbf{P} = \boldsymbol{\phi}^{\mathrm{T}}$. Hence, the edge weights w satisfy the following *Eulerian property*: for any subset $S \subseteq V$, we have $w(\partial(S)) = w(\partial(\overline{S}))$, where $\overline{S} := V \setminus S$.

We can define edge weights $\widehat{w}$ for the (complete) undirected graph with vertex set V such that for $u \neq v$, $\widehat{w}(u,v) = \frac{1}{2}(w(u,v) + w(v,u))$, and each self-loop has the same weight in $\widehat{w}$ and w.

Because of the Eulerian property of w, it is immediate that for all $S \subseteq V$, $w(\partial(S)) = \widehat{w}(\partial(S))$, where $\partial(S)$ is interpreted according to the directed case on the left and to the undirected case on the right. Moreover, for all $u \in V$,

$w(u) = \widehat{w}(u)$. Hence, as far as expansion is concerned, it is equivalent to consider the undirected graph with edge weights given by the matrix $\widehat{\mathbf{W}}$, for which the (higher-order) Cheeger Inequalities (as in Fact 1) can be readily applied.

Chung's approach can be applied to any stationary distribution $\boldsymbol{\phi}$ of $\mathbf{P}$, but special attention is paid to the case when $\mathbf{P}$ is irreducible, i.e., the edges corresponding to non-zero transition probabilities form a strongly connected graph on V. The advantages is that in this case, the stationary distribution is unique, and every vertex has non-zero probability.

In Section 3, we consider how to extend Chung's approach to the case where the underlying directed is not strongly connected. In Section 4, we consider the case when the edge weights $\boldsymbol{\phi}$ deviate (slightly) from a stationary distribution.

3 Directed Graphs with Multiple Strongly Connected Components

In a directed graph, we say that a strongly connected component C is a *sink*, if there is no edge leaving C. Otherwise, we say that it is a non-sink.

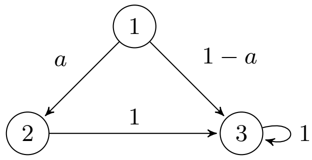
Even if the underlying directed graph of a given transition matrix is not strongly connected, Chung's approach [8] can still be applied if one chooses the vertex weights according to some stationary distribution.

However, under any stationary distribution, the weight on any vertex in a non-sink component must be zero. If we consider expansion using weights induced by a stationary distribution, essentially we are considering only the sink components. In this section, we explore if there is any meaningful way to consider expansion properties involving the non-sink components. As we shall see, it makes sense to consider the expansion properties of each strongly connected component separately.

3.1 Motivation for Considering Components Separately

Suppose $\mathbf{P}$ is a probability transition matrix corresponding to a random walk on some direted graph $G(V, E)$. Given a subset $C \subseteq V$, let $\mathbf{P}|_C$ be the square matrix restricting to the columns and the rows corresponding to C.

It is known [7, Theorem 3.22] that the eigenvalues of $\mathbf{P}$ are the union of the eigenvalues of $\mathbf{P}|_C$ over all strongly connected components C in a directed graph G. An important observation is that as long as the strongly connected components and the transition probabilities within a component remain the same, the eigenvalues of $\mathbf{P}$ are independent of the transition probabilities between different strongly connected components. For instance, the figure below depicts a directed graph, where the edges are labeled with the transition probabilities. Observe that each vertex is its own strongly connected component.



The transition matrix is:

$$\mathbf{P} = \begin{pmatrix} 0 & a & 1-a \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{pmatrix},$$

whose eigenvalues are $\{0, 0, 1\}$, which are independent of the parameter a. This suggests that it might be difficult to use spectral properties to analyze expansion properties involving edges between different strongly connected components. Hence, we propose that the expansion of each connected component should be analyzed separately.

3.2 Defining Expansion via Diluted Stationary Distribution

Observe that if a strongly connected component C is a sink, then the transition matrix $\mathbf{P}|_C$ has a stationary distribution, and we can apply Chung's approach. However, if C is a non-sink, then not every row of $\mathbf{P}|_C$ sums to 1, and so $\mathbf{P}|_C$ has no stationary distribution.

However, since C is a strongly connected component, by the Perron-Frobenius Theorem [10], $\mathbf{P}|_C$ has a unique maximum eigenvalue $\lambda_{\max} \geq 0$ with algebraic and geometric multiplicity 1 such that every other eigenvalue (which might be complex) has magnitude at most $\lambda_{\max}$. Moreover,the associated eigenvector of $\lambda_{\max}$ has positive coordinates and is unique up to scaling. Suppose $\boldsymbol{\phi}$ is the (left) eigenvector which is normalized such that the coordinates sum to 1, i.e., $\sum_{u \in V} \phi(u) = 1$ and $\boldsymbol{\phi}^{\mathrm{T}} \mathbf{P}|_C = \lambda_{\max} \boldsymbol{\phi}$. We say that $\boldsymbol{\phi}$ is the *diluted stationary distribution* of $\mathbf{P}|_C$, because the distribution on vertices in C is diluted by a factor of $\lambda_{\max}$ after one step of the random walk.

We use the diluted stationary distribution $\boldsymbol{\phi}$ as vertex weights to define expansion $\rho(S)$ for $S \subseteq C$. Observe that the weight of edges leaving component C also contributes to the expansion.

3.3 Augmenting Graph to Achieve Stationary Distribution

Suppose the component C has size n, and has $\lambda_{\max} < 1$. In order to use Chung's approach, we construct an augmented graph $\widehat{G}$ consisting of the component C and an extra vertex v_0. For each $u \in C$, all the original probabilities leaking out of C from u are now directed to v_0. For the new vertex v_0, the transition probabilities from v_0 to vertices in C are given by the diluted stationary distribution $\boldsymbol{\phi}$. Hence, the augmented graph $\widehat{G}$ is strongly connected. We write $\mathbf{A} = \mathbf{P}|_C \in \mathbb{R}^{n\times n}$, and $\boldsymbol{\mu} = \mathbf{1}_n - \mathbf{A}\mathbf{1}_n \in \mathbb{R}^n$. The new transition matrix is

$$\mathbf{B} = \begin{pmatrix} \mathbf{A} & \boldsymbol{\mu} \\ \boldsymbol{\phi}^{\mathrm{T}} & 0 \end{pmatrix} \in \mathbb{R}^{(n+1)\times(n+1)}.$$

Given a square matrix $\mathbf{M}$, its determinant is denoted as $|\mathbf{M}|$, and $eig(\mathbf{M})$ is the multi-set of its eigenvalues, which are roots of the polynomial $|\lambda\mathbf{I} - \mathbf{M}|$ in λ. We first show that the matrix $\mathbf{B}$ preserves the spectral properties of $\mathbf{A}$.

Lemma 1 (Spectral Preservation). *We have $eig(\mathbf{B}) = eig(\mathbf{A}) - \{\lambda_{max}\} + \{1, \lambda_{max} - 1\}$. Furthermore, if an eigenvalue $\lambda \in eig(\mathbf{A}) - \{\lambda_{max}, \lambda_{max} - 1\}$, it has the same geometric multiplicity in $\mathbf{A}$ and $\mathbf{B}$.*

Proof. To prove the first part, it suffices to show that for all $\lambda \in \mathbb{R}$,

$$\left|\lambda\mathbf{I}_n - \mathbf{A}\right| (\lambda - 1)(\lambda - (\lambda_{\max} - 1)) = \left|\lambda\mathbf{I}_{n+1} - \mathbf{B}\right| (\lambda - \lambda_{\max}),$$

because both sides are polynomials in λ of degree $n + 2$. Hence, they must be equivalent polynomials if they are equal for more than $n + 2$ values of λ.

If $\lambda = 1$, then the right hand side is zero because $\mathbf{B}$ has eigenvalue 1; similarly, if $\lambda = \lambda_{\max}$, the left hand side is zero because $\mathbf{A}$ has eigenvalue $\lambda_{\max}$.

For $\lambda \neq 1, \lambda_{\max}$, we have:

$$\begin{aligned}
\left|\lambda\mathbf{I}_{n+1} - \mathbf{B}\right| &= \begin{vmatrix} \lambda\mathbf{I}_n - \mathbf{A} & -\boldsymbol{\mu} \\ -\boldsymbol{\phi}^{\mathrm{T}} & \lambda \end{vmatrix} \cdot \begin{vmatrix} \mathbf{I}_n & \mathbf{1}_n \\ \mathbf{0}_n^{\mathrm{T}} & 1 \end{vmatrix} && (1) \\
&= \begin{vmatrix} \lambda\mathbf{I}_n - \mathbf{A} & (\lambda - 1)\mathbf{1}_n \\ -\boldsymbol{\phi}^{\mathrm{T}} & \lambda - 1 \end{vmatrix} && (2) \\
&= (\lambda - 1)\left|\lambda\mathbf{I}_n - \mathbf{A} - (\lambda - 1)\mathbf{1}_n(\lambda - 1)^{-1}(-\boldsymbol{\phi}^{\mathrm{T}})\right| && (3) \\
&= (\lambda - 1)\left|\lambda\mathbf{I}_n - \mathbf{A} + \mathbf{1}_n\boldsymbol{\phi}^{\mathrm{T}}\right|, && (4)
\end{aligned}$$

where (1) follows because the second determinant is 1. Moreover, (2) follows from $|\mathbf{X}| \cdot |\mathbf{Y}| = |\mathbf{XY}|$. Equation (3) follows from the identity that for invertible $\mathbf{H}$,

$$\begin{vmatrix} \mathbf{E} & \mathbf{F} \\ \mathbf{G} & \mathbf{H} \end{vmatrix} = |\mathbf{H}| \cdot |\mathbf{E} - \mathbf{F}\mathbf{H}^{-1}\mathbf{G}|.$$

Similarly, using $|\mathbf{X}| \cdot |\mathbf{Y}| = |\mathbf{X}\mathbf{Y}|$ repeatedly, we have

$$\begin{aligned}
\left|\lambda\mathbf{I}_n - \mathbf{A} + \mathbf{1}_n\boldsymbol{\phi}^{\mathrm{T}}\right|(\lambda - \lambda_{\max}) &= \begin{vmatrix} \mathbf{I}_n & -\mathbf{1}_n \\ \mathbf{0}_n^{\mathrm{T}} & 1 \end{vmatrix} \cdot \begin{vmatrix} \lambda\mathbf{I}_n - \mathbf{A} + \mathbf{1}_n\boldsymbol{\phi}^{\mathrm{T}} & \mathbf{0}_n \\ \boldsymbol{\phi}^{\mathrm{T}} & \lambda - \lambda_{\max} \end{vmatrix} \\
&= \begin{vmatrix} \mathbf{I}_n & \mathbf{0}_n \\ \frac{-\boldsymbol{\phi}^{\mathrm{T}}}{\lambda - \lambda_{\max}} & 1 \end{vmatrix} \cdot \begin{vmatrix} \lambda\mathbf{I}_n - \mathbf{A} & -(\lambda - \lambda_{\max})\mathbf{1}_n \\ \boldsymbol{\phi}^{\mathrm{T}} & \lambda - \lambda_{\max} \end{vmatrix} \\
&= \begin{vmatrix} \lambda\mathbf{I}_n - \mathbf{A} & -(\lambda - \lambda_{\max})\mathbf{1}_n \\ \mathbf{0}_n & \lambda - \lambda_{\max} + 1 \end{vmatrix} \\
&= \left|\lambda\mathbf{I}_n - \mathbf{A}\right|(\lambda - \lambda_{\max} + 1).
\end{aligned}$$

This completes the proof of the first part. To show that the geometry multiplicities of a common eigenvalue $\lambda \neq \lambda_{\max}, \lambda_{\max} - 1$ are equal, we show that $\mathbf{x} \longleftrightarrow \binom{\mathbf{x}}{0}$ is a bijection between $\mathbf{A}$ and $\mathbf{B}$'s corresponding right eigenvectors.

Since $\lambda \neq \lambda_{\max}$ is an eigenvalue of $\mathbf{A}$, if $\mathbf{x}$ is a corresponding eigenvector, then $\boldsymbol{\phi}^{\mathrm{T}}\mathbf{x} = 0$, because $\lambda_{\max}\boldsymbol{\phi}^{\mathrm{T}}\mathbf{x} = \boldsymbol{\phi}^{\mathrm{T}}\mathbf{A}\mathbf{x} = \lambda\boldsymbol{\phi}^{\mathrm{T}}\mathbf{x}$. Hence, $B\binom{\mathbf{x}}{0} = \lambda\binom{\mathbf{x}}{0}$.

Conversely, suppose $\binom{\mathbf{x}}{y}$ is an eigenvector of $\mathbf{B}$ with eigenvalue λ, where $\mathbf{x} \in \mathbb{R}^n$ and $y \in \mathbb{R}$. We have

$$\begin{cases} \mathbf{A}\mathbf{x} + \boldsymbol{\mu}y = \lambda\mathbf{x} \\ \boldsymbol{\phi}^{\mathrm{T}}\mathbf{x} = \lambda y \end{cases} \quad \text{or} \quad \begin{cases} (\lambda\mathbf{I} - \mathbf{A})\mathbf{x} = \boldsymbol{\mu}y \\ \boldsymbol{\phi}^{\mathrm{T}}\mathbf{x} = \lambda y. \end{cases}$$

Then, $\boldsymbol{\phi}^{\mathrm{T}}(\lambda\mathbf{I} - \mathbf{A})\mathbf{x} = (\lambda - \lambda_{\max})\boldsymbol{\phi}^{\mathrm{T}}\mathbf{x} = (\lambda - \lambda_{\max})\lambda y$.

But we also have $\boldsymbol{\phi}^{\mathrm{T}}\boldsymbol{\mu}y = \boldsymbol{\phi}^{\mathrm{T}}(\mathbf{1} - \mathbf{A}\mathbf{1})y = (1 - \lambda_{\max})y$. Since the two quantities are equal, this implies that $\lambda = 1, \lambda_{\max} - 1$ or $y = 0$. By assumption, $\lambda \neq 1, \lambda_{\max} - 1$, and so the only possibility is $y = 0$, then $\mathbf{A}\mathbf{x} = \lambda\mathbf{x}$. This shows that λ has the same geometric multiplicity in $\mathbf{A}$ and $\mathbf{B}$. □

3.4 Higher-Order Cheeger Inequalities for Component

Given a non-sink component C, we have described how to add an extra vertex v_0 to construct an augmented graph $\widehat{G}$ with transition matrix $\mathbf{B}$. Observe that $\mathbf{B}$ has stationary distribution $\widehat{\boldsymbol{\phi}} = (\boldsymbol{\phi}, 1 - \lambda_{\max})$, where $\boldsymbol{\phi}$ is the diluted stationary distribution of $\mathbf{A}$. Hence, it follows that for all $S \subseteq C$, the old expansion $\rho(S)$ is the same as the new expansion $\widehat{\rho}(S)$ in the augmented graph.

Therefore, we can apply Chung's approach [8] and the spectral analysis by Lee et al. [13] to obtain the following lemma, which is a restatement of Theorem 1.

Lemma 2 (Cheeger Inequalities for Component C). *Suppose $\widehat{\mathbf{\Phi}}$ is the diagonal matrix whose diagonal entries are coordinates of the stationary distribution $\widehat{\boldsymbol{\phi}}$ of $\mathbf{B}$. Moreover, suppose the normalized Laplacian $\boldsymbol{\mathcal{L}}$ of the symmetric matrix $\frac{1}{2}(\widehat{\mathbf{\Phi}}\mathbf{B} + \mathbf{B}^T\widehat{\mathbf{\Phi}})$ has eigenvalues $0 = \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_{n+1}$. Then, for all $1 \leq k \leq n$, $\frac{\lambda_k}{2} \leq \rho_k(C) \leq O(k^2) \cdot \sqrt{\lambda_{k+1}}$.*

Proof. We use the following inequality from [13]:
$\frac{\lambda_k}{2} \le \widehat{\rho}_k(\widehat{G}) \le O(k^2) \cdot \sqrt{\lambda_k}$.

Observe that if $S \subseteq C$ does not contain the new vertex v_0, then S has the same expansion $\rho(S) = \widehat{\rho}(S)$ in both graphs.

From $k+1$ disjoint subsets in the augmented graph $\widehat{G}$, we can get at least k subsets of C by removing the one containing v_0. Hence, we have
$\rho_k(C) \le \widehat{\rho}_{k+1}(\widehat{G}) \le O((k+1)^2)\sqrt{\lambda_{k+1}}$.

On the other hand, k disjoint subsets in C are also disjoint in $\widehat{G}$. Therefore, we have $\rho_k(C) \ge \rho_k(\widehat{G}) \ge \frac{\lambda_k}{2}$, as required. □

4 Vertex Weights Deviate from Stationary Distribution

In this section, we consider a transition matrix $\mathbf{P}$ whose underlying directed graph $G(V, E)$ (where $n = |V|$) is not necessarily strongly connected. Moreover, each vertex has a positive weight given by a vector $\boldsymbol{\phi} \in \mathbb{R}^n$ that is not necessarily a stationary distribution of $\mathbf{P}$. We wish to analyze the expansion with respect to $\mathbf{P}$ and $\boldsymbol{\phi}$ using spectral techniques. As in Section 3, we shall add an extra vertex v_0 to form an augmented graph $\widehat{G}$.

We measure how much $\boldsymbol{\phi}$ deviates from a stationary distribution by the following parameter:

$$\varepsilon := 1 - \min_{u \in V} \frac{\phi(u)}{\sum_{v \in V} \phi(v) \cdot P(v, u)}.$$

A smaller value of ε means that $\boldsymbol{\phi}$ is closer to a stationary distribution. In particular, if $\varepsilon = 0$, then $\boldsymbol{\phi}$ is a stationary distribution.

Our idea is to first scale down all probabilities in $\mathbf{P}$ by a factor of $(1-\varepsilon)$. We add a new vertex v_0 to absorb the extra ε probability from each existing vertex. Then, we define the transition probabilities from v_0 to the original vertices carefully such that each original vertex u receives the same weight $\phi(u)$ after one step.

For each vertex u, the weight mass it obtains from vertices V through the scaled-down $\mathbf{P}$ is $(1-\varepsilon)\sum_{v \in V} \phi(v) P(v, u)$. Hence, the new vertex v_0 needs to return mass weights to vertices in V given by the vector $\mathbf{m} := \boldsymbol{\phi} - (1-\varepsilon)\mathbf{P}^{\mathrm{T}}\boldsymbol{\phi}$, whose coordinates are non-negative by the choice of ε. Normalizing by $\mathbf{m}^{\mathrm{T}}\mathbf{1}_n = \varepsilon\boldsymbol{\phi}^{\mathrm{T}}\mathbf{1}_n$, we have the vector $\boldsymbol{\mu} := \frac{\mathbf{m}}{\varepsilon\boldsymbol{\phi}^{\mathrm{T}}\mathbf{1}_n}$ of transition probabilities from v_0 to vertices in V.

The transition matrix of the augmented graph $\widehat{G}$ is

$$\widehat{\mathbf{P}} = \begin{pmatrix} (1-\varepsilon)\mathbf{P} & \varepsilon\mathbf{1}_n \\ \boldsymbol{\mu}^{\mathrm{T}} & 0 \end{pmatrix}.$$

Observe that $\widehat{G}$ is strongly connected, and its stationary distribution can be obtained by normalizing the vector $\widehat{\boldsymbol{\phi}} = (\boldsymbol{\phi}, \varepsilon\boldsymbol{\phi}^{\mathrm{T}}\mathbf{1}_n)$. In other words, we can append a new coordinate corresponding to v_0 to $\boldsymbol{\phi}$ to obtain a left eigenvector $\widehat{\boldsymbol{\phi}}$ with eigenvalue 1 for matrix $\widehat{\mathbf{P}}$.

Moreover, for each subset $S \subseteq V$ of the original vertices, the new expansion $\widehat{\rho}(S)$ with respect to $\widehat{\boldsymbol{\phi}}$ and $\widehat{\mathbf{P}}$ can be related to the old expansion $\rho(S)$ as follows: $\widehat{\rho}(S) = (1-\varepsilon) \cdot \rho(S) + \varepsilon$.

Hence, we can apply Chung's approach to $\widehat{\mathbf{P}}$ and $\widehat{\boldsymbol{\phi}}$ to construct a symmetric Laplacian matrix, whose eigenvalues are related to the expansion properties using the results by Lee et al. [13]. The following lemma is a restatement of Theorem 2, and its proof uses the same argument as in the proof of Lemma 2.

Lemma 3. *Suppose $\widehat{\mathbf{\Phi}}$ is the diagonal matrix whose diagonal entries are coordinates of $\widehat{\boldsymbol{\phi}}$ as defined above. Moreover, suppose the normalized Laplacian $\boldsymbol{\mathcal{L}}$ of the symmetric matrix $\frac{1}{2}(\widehat{\mathbf{\Phi}}\widehat{\mathbf{P}} + \widehat{\mathbf{P}}^T\widehat{\mathbf{\Phi}})$ has eigenvalues $0 = \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_{n+1}$. Then, for all $1 \leq k \leq n$, $\frac{\lambda_k}{2} \leq (1-\varepsilon) \cdot \rho_k(V) + \varepsilon \leq O(k^2) \cdot \sqrt{\lambda_{k+1}}$.*

References

[1] Noga Alon. Eigenvalues and expanders. *Combinatorica*, 6(2):83–96, 1986.

[2] Noga Alon and Nabil Kahale. A spectral technique for coloring random 3-colorable graphs. *SIAM J. Comput.*, 26(6):1733–1748, 1997.

[3] Noga Alon and V. D. Milman. lambda_1, isoperimetric inequalities for graphs, and superconcentrators. *J. Comb. Theory, Ser. B*, 38(1):73–88, 1985.

[4] Bengt Aspvall and John R Gilbert. Graph coloring using eigenvalue decomposition. *SIAM Journal on Algebraic Discrete Methods*, 5(4):526–538, 1984.

[5] Sergey Brin and Lawrence Page. Reprint of: The anatomy of a large-scale hypertextual web search engine. *Computer Networks*, 56(18):3825–3833, 2012.

[6] Jeff Cheeger. A lower bound for the smallest eigenvalue of the laplacian. *Problems in analysis*, 625:195–199, 1970.

[7] W.K. Chen, H.A. Lauwerier, and W.T. Koiter. *Applied Graph Theory: Graphs and Electrical Networks*. North-Holland Series in Applied Mathematics and Mechanics. Elsevier Science, 2014.

[8] Fan Chung. Laplacians and the cheeger inequality for directed graphs. *Annals of Combinatorics*, 9(1):1–19, 2005.

[9] Fan RK Chung. *Spectral graph theory*, volume 92. American Mathematical Soc., 1997.

[10] Roger A. Horn and Charles R. Johnson. *Matrix analysis*. Cambridge University Press, 1990.

[11] Jon M. Kleinberg. Authoritative sources in a hyperlinked environment. *J. ACM*, 46(5):604–632, 1999.

[12] Tsz Chiu Kwok, Lap Chi Lau, Yin Tat Lee, Shayan Oveis Gharan, and Luca Trevisan. Improved cheeger's inequality: analysis of spectral partitioning algorithms through higher order spectral gap. In *STOC*, pages 11–20, 2013.

[13] James R. Lee, Shayan Oveis Gharan, and Luca Trevisan. Multiway spectral partitioning and higher-order cheeger inequalities. *J. ACM*, 61(6):37, 2014.

[14] Andrew Y. Ng, Michael I. Jordan, and Yair Weiss. On spectral clustering: Analysis and an algorithm. In *NIPS*, pages 849–856, 2001.

[15] Jonathan Richard Shewchuk. Ladies and gentlemen, allow me to introduce spectral and isoperimetric graph partitioning. 2011.

[16] Jianbo Shi and Jitendra Malik. Normalized cuts and image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.*, 22(8):888–905, 2000.

[17] David Tolliver and Gary L. Miller. Graph partitioning by spectral rounding: Applications in image segmentation and clustering. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 1053–1060, 2006.

A Lower Bound Improvement for Theorem 2

We show that the lower bound in the inequality in Theorem 2 can be slightly improved if we allow a self-loop with negative weight at the new vertex v_0 in the augmented graph described in Section 4.

Following the notation from Section 4, for $b \geq 0$, we define the following transition matrix:

$$\widehat{\mathbf{P}}_b = \begin{pmatrix} \mathbf{P}(1-\varepsilon) & \varepsilon \mathbf{1}_n \\ \boldsymbol{\mu}^{\mathrm{T}}(1+b) & -b \end{pmatrix} \in \mathbb{R}^{(n+1)\times(n+1)}.$$

In particular, $\widehat{\mathbf{P}}_0 = \widehat{\mathbf{P}}$. The corresponding left eigenvector with eigenvalue 1 is $\widehat{\boldsymbol{\phi}}_b = (\boldsymbol{\phi}, \frac{\varepsilon \boldsymbol{\phi}^{\mathrm{T}} \mathbf{1}_n}{1+b})$. Without loss of generality, we can normalize $\boldsymbol{\phi}$ and assume $\boldsymbol{\phi}^{\mathrm{T}} \mathbf{1}_n = 1$. Denoting $\widehat{\mathbf{\Phi}}_b$ as the diagonal matrix with entries derived from $\widehat{\boldsymbol{\phi}}_b$, the Laplacian is

$$\boldsymbol{\mathcal{L}}_b := \mathbf{I} - \frac{\widehat{\mathbf{\Phi}}_b^{\frac{1}{2}} \widehat{\mathbf{P}}_b \widehat{\mathbf{\Phi}}_b^{-\frac{1}{2}} + \widehat{\mathbf{\Phi}}_b^{-\frac{1}{2}} \widehat{\mathbf{P}}_b^{\mathrm{T}} \widehat{\mathbf{\Phi}}_b^{\frac{1}{2}}}{2}.$$

We use $\lambda_k(b)$ to denote the k-th smallest eigenvalue of $\boldsymbol{\mathcal{L}}_b$. Observe that for any subset S of original vertices, the expansion $\widehat{\rho}(S)$ is independent of b, and we still have $\widehat{\rho}(S) = (1-\varepsilon)\cdot\rho(S) + \varepsilon$.

It can be checked that the lower bound in the inequality of Fact 1 still holds when vertices can have negative-weighted self-loops, as long as the weight of a vertex remains non-negative. Therefore, we still have the following modified inequality from Lemma 3:

$$\frac{\lambda_k(b)}{2} \leq (1-\varepsilon)\rho_k(V) + \varepsilon.$$

We next show that for each $1 \leq k \leq n$, the function $b \mapsto \lambda_k(b)$ is increasing. Moreover, as b tends to infinity, we can show that these eigenvalues converge to the eigenvalues of some $n \times n$ matrix $\boldsymbol{\mathcal{L}}_\infty$.

Lemma 4 (Converging Eigenvalues). *For $1 \leq k \leq n$, the eigenvalue $\lambda_k(b)$ is an increasing function of b, and as b tends to infinity, it converges to the k-th eigenvalue $\lambda_k(\infty)$ of following $n \times n$ matrix:*

$$\boldsymbol{\mathcal{L}}_\infty = \mathbf{I} - \frac{1}{2}\widehat{\mathbf{\Phi}}^{-\frac{1}{2}}[(1-\varepsilon)(\widehat{\mathbf{\Phi}}\mathbf{P} + \mathbf{P}^T\widehat{\mathbf{\Phi}}) + \varepsilon(\frac{(\boldsymbol{\phi}+\boldsymbol{\mu})(\boldsymbol{\phi}^T + \boldsymbol{\mu}^T)}{2})]\widehat{\mathbf{\Phi}}^{-\frac{1}{2}}.$$

Proof. The Rayleigh quotient of $\mathcal{L}_b$ for $f : V \cup \{v_0\} \to \mathbb{R}^{n+1}$ is:

$$\begin{aligned}\mathcal{R}_b(f) &= \frac{\sum_{u,v} \widehat{\phi}_b(u)\widehat{P}_b(u,v)|f(u)-f(v)|^2}{\sum_u \widehat{\phi}_b(u)|f(u)|^2} \\ &= \frac{\sum_{u,v} \phi(u)\widehat{P}(u,v)|f(u)-f(v)|^2}{\sum_{u\neq v_0} \phi(u)|f(u)|^2 + \frac{\varepsilon}{1+b}|f(v_0)|^2},\end{aligned}$$

which is an increasing function of b.

Recall that $\lambda_k(b) = \min_{f_1,\ldots,f_k} \max_{f\in\text{span}\{f_1,\ldots,f_k\}} \mathcal{R}_b(f)$, where the minimum is over all mutually orthogonal $f_1, \ldots, f_k$ with respect to the inner product $\langle f, g\rangle_w := \sum_u w(u)f(u)g(u)$. Since for each f, $\mathcal{R}_b(f)$ is an increasing function of b, it follows that $\lambda_k(b)$ is also an increasing function of b.

Next, we show the convergence of $\lambda_k(b)$ as b tends to infinity. Consider the following:

$$
\begin{aligned}
\lim_{b\to\infty} \mathcal{R}_b(f) &= \lim_{b\to\infty} \frac{\sum_{u,v} \phi(u)\widehat{P}(u,v)|f(u)-f(v)|^2}{\sum_{u\neq v_0}\phi(u)|f(u)|^2 + \frac{\varepsilon}{1+b}|f(v_0)|^2} \\
&= \frac{\sum_{u,v} \phi(u)\widehat{P}(u,v)|f(u)-f(v)|^2}{\sum_{u\neq v_0}\phi(u)|f(u)|^2} \\
&= \frac{\sum_{u,v\neq v_0} \phi(u)\widehat{P}(u,v)|f(u)-f(v)|^2 + \sum_{u\neq v_0}(\phi(u)\varepsilon + \varepsilon\mu(u))|f(u)-f(v_0)|^2}{\sum_{u\neq v_0}\phi(u)|f(u)|^2} \\
&= \frac{\sum_{u,v\neq v_0} \phi(u)\widehat{P}(u,v)|f(u)-f(v)|^2}{\sum_{u\neq v_0}\phi(u)|f(u)|^2} + \frac{\varepsilon\sum_{u\neq v_0}(\phi(u)+\mu(u))|f(u)-f(v_0)|^2}{\sum_{u\neq v_0}\phi(u)|f(u)|^2}
\end{aligned}
$$

The first term is

$$
\begin{aligned}
&\frac{\sum_{u,v\neq v_0} \phi(u)\widehat{P}(u,v)|f(u)-f(v)|^2}{\sum_{u\neq v_0}\phi(u)|f(u)|^2} \\
=&\frac{\sum_{u,v\neq v_0}(\phi(u)\widehat{P}(u,v)+\phi(v)\widehat{P}(v,u))f(u)^2 - \sum_{u,v\neq v_0}(\phi(u)\widehat{P}(u,v)+\phi(v)\widehat{P}(v,u))f(u)f(v)}{\sum_{u\neq v_0}\phi(u)|f(u)|^2} \\
=&(1-\varepsilon)\frac{\langle f, \boldsymbol{\mathcal{J}} f\rangle}{\langle f, \widehat{\mathbf{\Phi}} f\rangle},
\end{aligned}
$$

where

$$
\boldsymbol{\mathcal{J}} = \widehat{\mathbf{\Phi}} + diag(\boldsymbol{\phi}^{\mathrm{T}}\mathbf{P}) - (\widehat{\mathbf{\Phi}}\mathbf{P} + \mathbf{P}^{\mathrm{T}}\widehat{\mathbf{\Phi}}).
$$

Note that in the second term,

$$
\begin{aligned}
&\sum_{u\neq v_0}(\phi(u)+\mu(u))|f(u)-f(v_0)|^2 \\
=&\sum_{u\neq v_0}(\phi(u)+\mu(u))f(v_0)^2 - 2f(v_0)\sum_{u\neq v_0}(\phi(u)+\mu(u))f(u) + \sum_{u\neq v_0}(\phi(u)+\mu(u))f(u)^2 \\
=&2(f(v_0) - \frac{\sum_{u\neq v_0}(\phi(u)+\mu(u))f(u)}{2})^2 - \frac{(\sum_{u\neq v_0}(\phi(u)+\mu(u))f(u))^2}{2} + \sum_{u\neq v_0}(\phi(u)+\mu(u))f(u)^2 \\
\geq&\sum_{u\neq v_0}(\phi(u)+\mu(u))f(u)^2 - \frac{(\sum_{u\neq v_0}(\phi(u)+\mu(u))f(u))^2}{2} \\
=&\langle f, \boldsymbol{\mathcal{H}} f\rangle
\end{aligned}
$$

where

$$\boldsymbol{\mathcal{H}} = \widehat{\boldsymbol{\Phi}} + \mathbf{M} - \frac{(\boldsymbol{\phi} + \boldsymbol{\mu})(\boldsymbol{\phi}^{\mathrm{T}} + \boldsymbol{\mu}^{\mathrm{T}})}{2}.$$

Hence,

$$\lim_{b\to\infty} \mathcal{R}_b(f) \geq \frac{\langle f, ((1-\varepsilon)\boldsymbol{\mathcal{J}} + \varepsilon\boldsymbol{\mathcal{H}})f\rangle}{\langle f, \widehat{\boldsymbol{\Phi}} f\rangle}.$$

By taking $g = \widehat{\boldsymbol{\Phi}}^{\frac{1}{2}} f$, we have

$$\lim_{b\to\infty} \mathcal{R}_b(f) \geq 2\frac{\langle g, \boldsymbol{\mathcal{L}}_\infty g\rangle}{\langle g, g\rangle},$$

where

$$\begin{aligned} \boldsymbol{\mathcal{L}}_\infty &= \frac{1}{2}\widehat{\boldsymbol{\Phi}}^{-\frac{1}{2}}((1-\varepsilon)\mathbf{J} + \varepsilon\mathbf{H})\widehat{\boldsymbol{\Phi}}^{-\frac{1}{2}} \\ &= \mathbf{I} - \frac{1}{2}\widehat{\boldsymbol{\Phi}}^{-\frac{1}{2}}[(1-\varepsilon)(\widehat{\boldsymbol{\Phi}}\mathbf{P} + \mathbf{P}^{\mathrm{T}}\widehat{\boldsymbol{\Phi}}) + \varepsilon(\frac{(\boldsymbol{\phi} + \boldsymbol{\mu})(\boldsymbol{\phi}^{\mathrm{T}} + \boldsymbol{\mu}^{\mathrm{T}})}{2})]\widehat{\boldsymbol{\Phi}}^{-\frac{1}{2}}. \end{aligned}$$

This directly implies that $\lim_{b\to\infty} \lambda_k(b) \geq \lambda_k(\infty)$. On the other hand, we choose $f_1, \ldots, f_k \in \mathbb{R}^{V-v_0}$ such that

$$\lambda_k(\infty) = \max_{f\in\mathrm{span}\{f_1,\ldots,f_k\}} \mathcal{R}_\infty(f).$$

We define:

$$f_i' = \begin{cases} f_i(u) & u \neq v_0 \\ \frac{\sum(\phi(u)+\mu(u))f_i(u)}{2} & u = v_0 \end{cases}$$

We have $\lambda_k(\infty) \geq \max_{f\in\mathrm{span}\{f_1,\ldots,f_k\}} \mathcal{R}_b(f') \geq \lambda_k(b)$. As b tends to infinity, we have $\lambda_k(\infty) = \lim_{b\to\infty} \lambda_k(b)$. □

The following corollary is an improvement of Lemma 3, because the eigenvalues of $\boldsymbol{\mathcal{L}}_\infty$ are greater than those of $\boldsymbol{\mathcal{L}}_0$.

Corollary 1. *For the original graph $G(V, E)$ with transition matrix $\mathbf{P}$ and vertex weight $\boldsymbol{\phi}$, suppose ε is defined as in Section 4. Then, $(1-\varepsilon)\rho_k(V)+\varepsilon \geq \frac{\lambda_k(\infty)}{2}$, where $\lambda_k(\infty)$ is the k-th smallest eigenvalue of matrix $\boldsymbol{\mathcal{L}}_\infty$ as defined in Lemma 4*