Spectrally Approximating Large Graphs with Smaller Graphs

Andreas Loukas [1] **Pierre Vandergheynst** [1]

Abstract

How does coarsening affect the spectrum of a general graph? We provide conditions such that the principal eigenvalues and eigenspaces of a coarsened and original graph Laplacian matrices are close. The achieved approximation is shown to depend on standard graph-theoretic properties, such as the degree and eigenvalue distributions, as well as on the ratio between the coarsened and actual graph sizes. Our results carry implications for learning methods that utilize coarsening. For the particular case of spectral clustering, they imply that coarse eigenvectors can be used to derive good quality assignments even without refinement—this phenomenon was previously observed, but lacked formal justification.

1. Introduction

One of the most wide-spread techniques for sketching graph-structured data is coarsening. As with most sketching methods, instead of solving a large graph problem in its native domain, coarsening involves solving an akin problem of reduced size at a lower cost; the solution can then be inexpensively lifted and refined in the native domain.

The benefits of coarsening are well known both in the algorithmic and machine learning communities. There exists a long list of algorithms that utilize it for partitioning (Hendrickson & Leland, 1995; Karypis & Kumar, 1998a; Kushnir et al., 2006; Dhillon et al., 2007; Wang et al., 2014) and visualizing (Koren, 2002; Walshaw, 2006) large graphs in a computationally efficient manner. In addition, it has been frequently used to create multi-scale representations of graph-structured data, such as coarse-grained diffusion maps (Lafon & Lee, 2006), multi-scale wavelets (Gavish et al., 2010) and pyramids (Shuman et al., 2016).

More recently, coarsening is employed as a component of graph convolutional networks analogous to pooling (Bruna et al., 2014; Defferrard et al., 2016; Bronstein et al., 2017; Simonovsky & Komodakis, 2017). Combining the values of adjacent vertices reduces the spatial size of each layer's output, prevents overfitting, and encourages a hierarchical scaling of representations.

Yet, much remains to be understood about the properties and limitations of graph coarsening.

The majority of theoretical work has so far focused on constructing fast linear solvers using multigrid techniques. These methods are especially relevant for approximating the solution of differential equations on grids and finite-element meshes. Multigrids were also adapted to arbitrary graphs by Koutis et al. (2011) and later on by Livne and Brandt (2012). Based on an optimized version of the Galerkin coarsening, the authors demonstrate an algebraic multi-level approximation scheme that is numerically shown to solve symmetric diagonally dominant linear systems in almost linear time. Similar techniques have also been applied for approximating the Fiedler vector (Urschel et al., 2014; Gandhi, 2016) and solving least-squares problems (Hirani et al., 2015; Colley et al., 2017).

Despite this progress, with the exception of certain interlacing results (Chung, 1997; Chen et al., 2004), it is currently an open question how coarsening affects the spectrum of a general graph. As a consequence, there is no rigorous way of determining to what extend one may coarsen a graph without significantly affecting the performance of spectral methods for graph partitioning and visualization. The absence of a fundamental understanding of what and how much information is lost also hinders our ability to design efficient learning algorithms for graph-structured data: e.g., coarsening is the least studied (and less optimized) component of graph convolutional networks.

This paper sheds light into some of these questions. Specifically, we consider a one-shot coarsening operation and ask how much it affects the eigenvalues and eigenvectors of the graph Laplacian. Key to our argument is the introduced *restricted spectral similarity* (RSS) property, asserting that the Laplacian of the coarsened and actual graphs behave similarly (up to some constants) with respect to an appropriate set of vectors. The RSS property is shown to hold for coarsenings constructed by contracting the edges contained in a randomized matching. Moreover, the attained

[1]École Polytechnique Fédérale Lausanne, Switzerland. Correspondence to: Andreas Loukas <andreas.loukas@epfl.ch>.

Proceedings of the 35^{th} International Conference on Machine Learning, Stockholm, Sweden, PMLR 80, 2018.

constants depend on the degree distribution and can be controlled by the ratio of the coarsened and actual graph sizes, i.e., the extend of dimensionality reduction.

We utilize the RSS property to provide spectrum approximation guarantees. It is proven that the principal eigenvalues and eigenspaces of the coarsened and actual Laplacian matrices are close when the RSS constants are not too large. Our results carry implications for non-linear methods for data clustering (Von Luxburg, 2007) and dimensionality reduction (Belkin & Niyogi, 2003). A case in point is spectral clustering: we show that lifted eigenvectors can be used to produce clustering assignments of good quality even without refinement. This phenomenon has been observed experimentally (Karypis & Kumar, 1998a; Dhillon et al., 2007), but up to now lacked formal justification.

Paper organization. After introducing the RSS property in Section 2, we demonstrate in Section 3 how to generate coarsenings featuring small RSS constants. Sections 4 and 5 then link our results to spectrum preservation and spectral clustering, respectively. The paper concludes by briefly discussing the limitations of our analysis. The proofs can be found in a supplementary document.

2. Graph coarsening

Consider a weighted graph $G = (\mathcal{V}, \mathcal{E}, W)$ of $N = |\mathcal{V}|$ vertices and $M = |\mathcal{E}|$ edges, with the edge e_{ij} between vertices v_i and v_j weighed by $w_{ij} \leq 1$. As usual, we denote by L the combinatorial Laplacian of G defined as

$$L(i,j) = \begin{cases} d_i & \text{if } i = j \\ -w_{ij} & \text{if } e_{ij} \in \mathcal{E} \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

and d_i the weighted degree of v_i. Moreover, let λ_k be the k-th eigenvalue of L and u_k the associated eigenvector.

2.1. How to coarsen a graph?

At the heart of a coarsening lies a surjective (and therefore dimension reducing) mapping $\varphi : \mathcal{V} \to \mathcal{V}_\mathrm{c}$ between the original vertex set $\mathcal{V} = \{v_1, \ldots, v_N\}$ and the smaller vertex set $\mathcal{V}_\mathrm{c} = \{v'_1, \ldots, v'_n\}$. In other words, the coarse graph $G_\mathrm{c} = (\mathcal{V}_\mathrm{c}, \mathcal{E}_\mathrm{c})$ has $m = |\mathcal{E}_\mathrm{c}|$ and contains every edge $(i,j) \in \mathcal{E}$ for which $\varphi(v_i) \neq \varphi(v_j)$. We define the *coarsened Laplacian* as

$$L_\mathrm{c} = CLC^\top, \tag{2}$$

where the fat $n \times N$ coarsening matrix C describes how different $v \in \mathcal{V}$ are mapped onto the vertex set $\mathcal{V}_\mathrm{c}$. Similarly, we may *downsample* a vector $x \in \mathbb{R}^N$ supported on $\mathcal{V}$ by the linear transformation

$$x_\mathrm{c} = Cx, \tag{3}$$

where now $x_\mathrm{c} \in \mathbb{R}^n$. We here focus on coarsenings where each vertex v_i is mapped into a single v'_j. This is equivalent to only considering coarsening matrices with block-diagonal form $C = \mathrm{blkdiag}\left(c_1^\top, \ldots, c_n^\top\right)$, where each $c_j^\top = [c_j(1), \ldots, c_j(n_j)]$ is the length n_j coarsening weight vector associated with the j-th vertex v'_j of $\mathcal{V}_\mathrm{c}$. In addition, we restrict our attention to constant coarsening weight vectors of unit norm $\|c_j\|_2 = 1$ with entries equal to $n_j^{-1/2}$.

Though L_c is not a proper combinatorial Laplacian matrix (e.g., $L_\mathrm{c}1 \neq 0$ for 1 being the all ones vector), it can take the proper form using the simple re-normalization $QL_\mathrm{c}Q$, where $Q = \mathrm{diag}(C1)$. This might seem inconvenient at a first glance. We argue that it is not: it should not be the action of L_c in itself that matters, but its effect when combined with downsampling. When acting on x_c, the desired nullspace property is regained since $L_\mathrm{c}C1 = 0$. Alternatively, one could define $L_{c'} = QCLC^\top Q$ and $x_{c'} = Q^{-1}Cx$, where now $L_{c'}$ has the proper combinatorial Laplacian form. This construction however is equivalent to the one we consider here since $x_\mathrm{c}^\top L_\mathrm{c} x_\mathrm{c} = x^\top C^\top Q^{-1} QCLC^\top QQ^{-1}Cx = x_{c'}^\top L_{c'} x_{c'}$.

We will also utilize the notion of a coarsening frame:

Definition 1 (Coarsening frame)**.** *The coarsening frame $G_F = (\mathcal{V}_F, \mathcal{E}_F, W_F)$ is the subgraph of G induced by set $\mathcal{V}_F = \{v_i \mid \exists v_j$ with $\varphi(v_i) = \varphi(v_j)\}$.*

Informally, G_F is the subgraph of G that is coarsened (see Figure 1c). We say that the coarsening corresponds to an *edge contraction* if no two edges of the coarsening frame are themselves adjacent—in other words, $\mathcal{E}_F$ forms a matching on G.

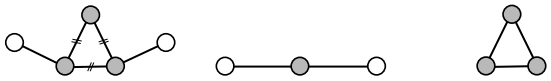
Lifting. We write $\widetilde{x} = C^\top x_\mathrm{c}$ to do an approximate inverse mapping from $\mathcal{V}_\mathrm{c}$ to $\mathcal{V}$, effectively *lifting* the dimension from $\mathbb{R}^n$ back to $\mathbb{R}^N$. To motivate this choice notice that, even though $\Pi = C^\top C$ is not an identity matrix, it is block diagonal $\Pi = \mathrm{blkdiag}\left(c_1 c_1^\top, \ldots, c_n c_n^\top\right)$. Moreover, Π is an identity mapping for all vectors in its range.

Property 1. *$\Pi = C^\top C$ is a rank n projection matrix.*

Proof. For each block Π_j in the diagonal of Π, we have $\Pi_j^2 = \Pi_j \Pi_j = c_j c_j^\top c_j c_j^\top = c_j c_j^\top \|c_j\|^2 = \Pi_j$. The rank of Π is n because each diagonal block Π_j is of rank one. □

Therefore, if x is a vector in $\mathbb{R}^N$ and $x_\mathrm{c} = Cx$ is its coarsened counterpart, then $\widetilde{x} = C^\top Cx = \Pi x$ is a locality-preserving approximation of x w.r.t. graph G.

A toy example. Consider the example graph shown in Figure 1a and suppose that we want to coarsen the $n_1 = 3$



(a) G. (b) G_c. (c) G_F.

Figure 1: A toy coarsening example.

gray vertices $\mathcal{V}_F = \{v_1, v_2, v_3\}$ of G into vertex v_1', as shown in Figure 1b. Matrices C and L_c take the form:

$$C = \begin{bmatrix} 1/\sqrt{3} & 1/\sqrt{3} & 1/\sqrt{3} & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} c_1^\top & 0 \\ 0 & I_2 \end{bmatrix}$$

$$L_c = CLC^\top = \begin{bmatrix} 2/3 & -1/\sqrt{3} & -1/\sqrt{3} \\ -1/\sqrt{3} & 1 & 0 \\ -1/\sqrt{3} & 0 & 1 \end{bmatrix}$$

Above, the 2×2 identity matrix I_2 preserves the neighborhood of all vertices not in $\mathcal{V}_F$. The coarsening frame is shown in Figure 1c.

2.2. Restricted spectral similarity

The objective of coarsening is dual. First, we aim to attain computational acceleration by reducing the dimensionality of our problem. On the other hand, we must ensure that we do not loose too much valuable information, in the sense that the structure of the reduced and original problems should be as close as possible.

Spectral similarity. One way to define how close a matrix B approximates the action of matrix A is to establish a spectral similarity relation of the form:

$$(1-\epsilon)\, x^\top Ax \leq x^\top Bx \leq (1+\epsilon)\, x^\top Ax, \qquad (4)$$

for all $\forall x \in \mathbb{R}^N$ and with ϵ a positive constant. Stated in our context, (4) can be rewritten as:

$$(1-\epsilon)\, x^\top Lx \leq x_c^\top L_c x_c \leq (1+\epsilon)\, x^\top Lx \qquad (5)$$

for all $x \in \mathbb{R}^N$ and with $x_c = Cx$. If the equation holds, we say that matrix L_c is an ϵ-spectral approximation of L. In graph theory, the objective of constructing sparse spectrally similar graphs is the main idea of spectral graph sparsifiers, a popular method for accelerating the solution of linear systems involving the Laplacian, initially proposed by Spielman and co-authors (Spielman & Srivastava, 2011; Spielman & Teng, 2011).

In contrast to the sparsification literature however, here the dimension of the space changes and one needs to take into account both the Laplacian coarsening (L becomes L_c) and the vector downsampling operation (x becomes x_c) in the similarity relation [1]. Yet, from an analysis standpoint, an alternative interpretation is possible. Defining $\widetilde{L} = \Pi L \Pi$, we re-write

$$x_c^\top L_c x_c = x^\top (C^\top C) L (C^\top C) x = x^\top \Pi L \Pi x = x^\top \widetilde{L} x.$$

Remembering that $C^\top$ acts as an approximate inverse of C, we interpret $\widetilde{L} \in \mathbb{R}^{N \times N}$ as an approximation of L that contains the same information as $L_c \in \mathbb{R}^{n \times n}$.

Restricted spectral similarity (RSS). Equation (5) thus states that the rank $n-1$ matrix $\widetilde{L}$ is an ϵ-spectral approximation of L, a matrix of rank $N-1$. Since the two matrices have different rank, the relation cannot hold for every $x \in \mathbb{R}^N$. To carry out a meaningful analysis, we focus on an appropriate subset of vectors.

More specifically, we restrict our attention to the first K eigenvectors of L and introduce the following property:

Definition 2 (Restricted spectral similarity)**.** *Suppose that there exists an integer K and positive constants ϵ_k, such that for every $k \leq K$,*

$$(1-\epsilon_k)\, \lambda_k \leq u_k^\top \widetilde{L} u_k \leq (1+\epsilon_k)\, \lambda_k. \qquad (6)$$

Then L_c is said to satisfy the restricted spectral similarity (RSS) property with RSS constants $\{\epsilon_k\}_{k=1}^K$.

The relation to spectral similarity is exposed by substituting $u_k^\top L u_k = \lambda_k$.

For every k, inequality (6) should intuitively capture how close is Cu_k to being an eigenvector of L_c: When $\epsilon_k = 0$, vector Cu_k is an eigenvector of L_c with eigenvalue λ_k. On the hand, for $\epsilon_k > 0$, Cu_k is not an eigenvector of L_c, but matrices L and $\widetilde{L}$ alter the length of vectors in the span of u_k in a similar manner (up to $1 \pm \epsilon_k$).

This intuition turns out to be valid. In the following we will demonstrate that the RSS property is a key ingredient in characterizing the relation between the first K eigenvalues and principal eigenspaces of the coarsened and actual Laplacian matrices. In particular, we will prove that the spectrum of L_c approximates that of L (up to lifting) when the constants ϵ_k are sufficiently small. This line of thinking will be developed in Section 4.

Remark 1. *A uniform RSS constant $\epsilon = \max_{k \leq K} \epsilon_k$ is sufficient to guarantee spectrum preservation, however, individual constants $\{\epsilon_k\}_{k=1}^K$ lead to tighter bounds.*

3. A randomized edge contraction algorithm

This section proposes an algorithm for coarsening a graph that provably produces coarsenings with bounded RSS con-

[1] Coarsening could perhaps be interpreted as combining edge and vertex sparsification (Moitra, 2011).

Algorithm 1 Randomized Edge Contraction (REC)

1: **input**: $G = (\mathcal{V}, \mathcal{E})$, T, ϕ
2: **output**: $G_{\mathrm{c}} = (\mathcal{V}_{\mathrm{c}}, \mathcal{E}_{\mathrm{c}})$
3: $\mathcal{C} \leftarrow \mathcal{E}$, $G_{\mathrm{c}} \leftarrow G$
4: $\Phi \leftarrow \sum_{e_{ij} \in \mathcal{E}} \phi_{ij}$, $t \leftarrow 0$.
5: **while** $|\mathcal{C}| > 0$ and $t < T$ **do**
6: $\quad t \leftarrow t + 1$.
7: $\quad$ Select each e_{ij} from $\mathcal{C}$ with prob. $p_{ij} = \phi_{ij}/\Phi$ or **continue** with prob. $1 - \sum_{e_{ij} \in \mathcal{C}} p_{ij}$.
8: $\quad \mathcal{C} \leftarrow \mathcal{C} \setminus \mathcal{N}_{ij}$
9: $\quad G_{\mathrm{c}} \leftarrow$ contract(G_{c}, e_{ij}) as in (2)
10: **end while**

stants ϵ_k.

The method, which we refer to as REC, is described in Algorithm 1. REC resembles the common greedy procedure of generating maximal matchings, in that it maintains a candidate set $\mathcal{C}$ containing all edges that can be added to the matching. At each iteration, a new edge e_{ij} is added and set $\mathcal{C}$ is updated by removing from it all edges in the edge neighborhood set $\mathcal{N}_{ij}$ defined as follows:

$$\mathcal{N}_{ij} = \{e_{pq} \mid e_{pq} \in \mathcal{E} \text{ and } p = i \text{ or } q = j\},$$

which also includes the edge e_{ij}.

Yet, REC features two main differences. First, instead of selecting each new edge added to the matching uniformly at random, it utilizes a potential function ϕ defined on the edge set, i.e., $\phi : \mathcal{E} \to \mathbb{R}_+$ with which it controls the probability p_{ij} that every edge is contracted. The second difference is that, at each iteration, REC does not select a valid edge with probability $1 - \sum_{e_{ij} \in \mathcal{C}} p_{ij}$. This choice is not driven by computational concerns, but facilitates the analysis, as it alleviates the need for updating the total potential Φ after every iteration.

Remark 2. *REC is equivalent to the $O(M)$ complexity algorithm that samples from $\mathcal{C}$ directly in line 7 by updating Φ at every iteration such that its value is $\sum_{e_{ij} \in \mathcal{C}} \phi_{ij}$. Though we suggest to use this latter algorithm in practice, it is easier to express our results using the number of iterations T of Algorithm 1.*

REC returns a *maximal* matching when T is sufficiently large. As we will see in the following, it is sufficient to consider $T = O(N)$. The exact number of iterations will be chosen in order to balance the trade-off between the expected dimensionality reduction ratio

$$r \triangleq \mathbf{E}\left[\frac{N-n}{N}\right]$$

and the size of the RSS constants.

3.1. Analysis of REC

The following theorem characterizes an individual RSS constant of an L_{c} generated by REC. As exemplified in Corollary 5.1, similar argumenrs can also be used to derive a uniform bound over all ϵ_k for $k \leq K$.

Theorem 3.1. *Let L_c be the coarsened Laplacian produced by REC and further suppose that*

$$\lambda_k \leq 0.5 \min_{e_{ij} \in \mathcal{E}} \left\{ \frac{d_i + d_j}{2} + w_{ij} \right\}.$$

For any $\epsilon_k \geq 0$, the relation $\lambda_k \leq u_k^\top \widetilde{L} u_k \leq \lambda_k(1 + \epsilon_k)$ holds with probability at least

$$1 - c_2 \frac{1 - e^{-c_1 T/N}}{4\,\epsilon_k} \max_{e_{ij} \in \mathcal{E}} \chi_{ij} \left(\sum_{e_{pq} \in \mathcal{N}_{ij}} \frac{w_{pq}}{w_{ij}} + 3 - \frac{4\lambda_k}{w_{ij}} \right)$$

where $c_1 = N \max_{e_{ij} \in \mathcal{E}} \sum_{e_{pq} \in \mathcal{N}_{ij}} p_{pq}$,

$$c_2 = \frac{c_1/N}{1 - e^{-c_1/N}} \quad \textit{and} \quad \chi_{ij} = \frac{\phi_{ij}}{\sum_{e_{pq} \in \mathcal{N}_{ij}} \phi_{pq}}.$$

The theorem reveals that the dependency of ϵ_k to some extremal properties implied by the potential function ϕ and the graph structure. It is noteworthy that

$$c_1 = O(1) \quad \text{implies} \quad \lim_{N \to \infty} c_2 = 1. \tag{7}$$

These asymptotics can be taken at face value even for finite size problems: coarsening typically becomes computationally relevant for large N (typically $N > 10^3$), for which c_2 has effectively converged to its limit.

The assumption that c_1 is independent of N can be satisfied either by assuming that G is a *bounded degree graph*, such that $|\mathcal{N}_{ij}| \ll N$ for every $e_{ij} \in \mathcal{E}$, or by choosing potential functions ϕ_{ij} that are inversely proportional to $|\mathcal{N}_{ij}|$.

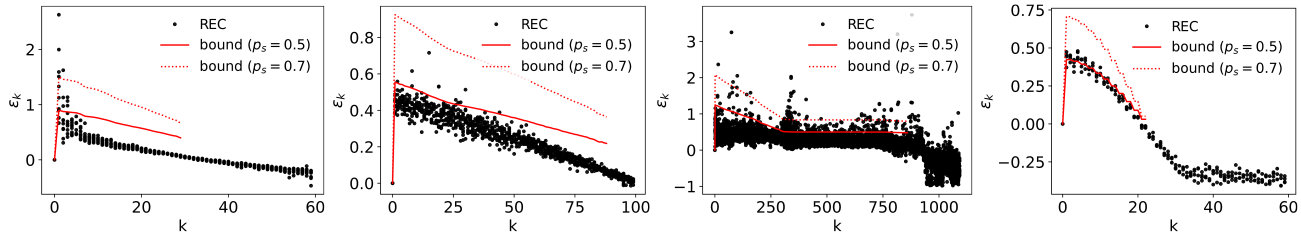
We can also incorporate the expected reduction ratio r in the bound, by noting that

$$rN = \sum_{e_{ij} \in \mathcal{E}} P(e_{ij} \in \mathcal{E}_F) \geq \sum_{e_{ij} \in \mathcal{E}} p_{ij} \frac{1 - e^{-TP_{ij}}}{P_{ij}} \geq \frac{1 - e^{-TP_{\max}}}{P_{\max}} = \frac{1 - e^{-c_1 T/N}}{c_1/N}, \tag{8}$$

(see proof of Theorem 3.1 for definitions of P_{ij} and $P_{\max}$) implying

$$1 - e^{-c_1 T/N} \leq r c_1, \tag{9}$$

as well as that $T = \frac{N}{c_1} \log\left(\frac{1}{1 - r c_1}\right)$ iterations suffice to achieve any $r < 1/c_1$. Nevertheless, this latter estimate is more pessimistic than the one presented in Theorem 3.1.



(a) Bunny (point cloud) (b) Swiss roll (manifold) (c) Yeast (protein network) (d) Regular graph

Figure 2: The proposed bounds follow the behavior of the RSS constants ϵ_k, especially for regular graphs or graphs with small degree variance. The two red lines plot the bounds of Theorem 3.1 for a success probability of $p_s = 0.5$ and $p_s = 0.7$.

The norm $\|\Pi^{\perp} u_k\|_2^2$. For all k, one has

$$P\left(\|\Pi^{\perp} u_k\|_2^2 \geq \epsilon\, \lambda_k\right) \leq c_2 \frac{1 - e^{-c_1 T/N}}{2\,\epsilon} \max_{e_{ij} \in \mathcal{E}} \frac{\chi_{ij}}{w_{ij}},$$

with constants defined as before (the derivation is not included as it resembles the one employed in the proof of Theorem 3.1). Thus, $\|\Pi^{\perp} u_k\|_2^2$ depends on the ratio r (through (9)) and is smaller for small k (due to λ_k). This is reasonable: by definition, eigenvectors corresponding to small eigenvalues are smooth functions on G; averaging some of their entries locally on the graph is unlikely to alter their values significantly.

3.2. The heavy-edge potential function

Let us examine how the achieved results behave for a specific potential function. Setting $\phi_{ij} = w_{ij}$ is a simple way to give preference to heavy frames—indeed, heavy-edge matchings have been utilized as a heuristic for coarsening (e.g., in combination with graph partitioning (Karypis & Kumar, 1998b)). It is perhaps interesting to note this particular potential function can be derived naturally from Theorem 3.1 if we require that

$$\chi_{ij} \frac{\sum_{e_{pq} \in \mathcal{N}_{ij}} w_{pq}}{w_{ij}} = 1 \quad \text{for all } e_{ij}. \tag{10}$$

It will be useful to denote respectively by $\varrho_{\min}$ and $\varrho_{\max}$ the minimum and maximum of $(d_i + d_j - w_{ij})/2d_{\text{avg}}$ over all e_{ij}, with d_{avg} being the average degree. It is straightforward to calculate that in this case

$$c_1 = 2\left(\frac{d_i + d_j - w_{ij}}{d_{\text{avg}}}\right) = 4\varrho_{\max}.$$

Therefore, $c_1 = O(1)$ for all graphs in which $\Omega(1) = \varrho_{\min} \leq \varrho_{\max} = O(1)$, and given sufficiently large N and some manipulation the probability estimate of Theorem 3.1 reduces to

$$1 - \frac{1 - e^{-4\varrho_{\max} T/N}}{4\,\epsilon_k}\left(1 + \frac{1.5 - 2\lambda_k}{d_{\text{avg}}\,\varrho_{\min}}\right). \tag{11}$$

In addition, $P\left(\|\Pi^{\perp} u_k\|_2^2 > \epsilon\, \lambda_k\right) \leq \frac{1 - e^{-4\varrho_{\max} T/N}}{2\,\epsilon\,\varrho_{\min}\, d_{\text{avg}}}$.

The heavy-edge potential function is therefore more efficient for graphs with small degree variations. Such graphs are especially common in machine learning, where often the connecticity of each vertex is explicitly constructed such that all degrees are close to some target value (e.g., using a k-nearest neighbor graph construction (Muja & Lowe, 2014)).

As a proof of concept, Figure 2 compares the actual constants ϵ_k with the bound of Theorem 3.1 when utilizing REC with a heavy-edge potential to coarsen the following benchmark graphs: (*i*) a point cloud representing a bunny obtained by re-sampling the Stanford bunny 3D-mesh (Turk & Levoy, 1994) and applying a k-nn construction ($N = 1000, r = 0.4, k = 30$), (*ii*) a k-nn similarity graph capturing the geometry of a 2D manifold usually referred to as Swiss roll ($N = 1000, r = 0.4, k = 10$), (*iii*) A network describing the interaction of yeast proteins (Rossi & Ahmed, 2015) ($N = 1458, r = 0.25, d_{\text{avg}} = 2, d_{\max} = 56$), and (*iv*) a d-regular graph ($N = 400, r = 0.4, d = 20$). To derive the bounds, we started from Theorem 3.1 and identified for each k the smallest ϵ_k such that the success probability is at least $p_s = \{0.5, 0.7\}$. As predicted by our analysis, ϵ_k decrease with k (the decrease is close to linear in λ_k) and with the variance of the degree distribution. The heavy-tailed yeast network and the regular graph constitute two extreme examples, with the latter featuring much smaller constants.

3.3. Regular graphs

For regular graphs, (9) becomes asymptotically tight, leading to the following Corollary:

Corollary 3.1. *If G is a regular graph with combinatorial degree d and equal edge weights $w_{ij} = w$, then for any k such that $\lambda_k \leq (d+1)/2$ and for sufficiently large N, the relation $\lambda_k \leq u_k^{\top} \widetilde{L} u_k \leq \lambda_k(1 + \epsilon_k)$ holds with probability at least*

$$\geq 1 - r\,\frac{1 - (2d)^{-1}}{\epsilon_k}\left(1 + \frac{1.5 - \lambda_k}{d - 0.5}\right) \overset{d \gg 1}{\approx} 1 - \frac{r}{\epsilon_k}, \tag{12}$$

inequality $\left\|\Pi^{\top} u_k\right\|_2^2 \geq \epsilon r \lambda_2$ holds for all k with proba-

bility at most $2/(d\epsilon)$ *and* $T = \frac{N}{2(2-1/d)} \log\left(\frac{1}{1-2(2-1/d)r}\right)$ *iterations of* **REC** *suffice in expectation to achieve reduction* r.

An other way to read Corollary 3.1 is that, for a sufficiently dense regular graph, there exists[2] an edge contraction for which L_c satisfies the RSS property with constants bounded by r.

4. The spectrum of the coarsened Laplacian

This section links the RSS property with spectrum preservation. Our results demonstrate that the distance between the spectrum of a coarsened Laplacian and of the combinatorial Laplacian it approximates is directly a function of RSS constants. This relation also extends to eigenspaces.

4.1. Basic facts about the spectrum

Before delving into our main results, let us first consider the spectrum of a coarsened Laplacian which does not (necessarily) meet the RSS property.

W.l.o.g., let G be connected and sort its eigenvalues as $0 = \lambda_1 < \lambda_2 \leq \ldots \leq \lambda_N$. Similarly, let $\widetilde{\lambda}_k$ be the k-th largest eigenvalue of the coarsened Laplacian L_c and name $\widetilde{u}_k$ the associated eigenvector. As the following theorem shows, there is a direct relation between the eigenvalues $\widetilde{\lambda}$ and λ.

Theorem 4.1. *Inequality* $\lambda_k \leq \widetilde{\lambda}_k$ *holds for all* $k \leq n$.

We remark the similarity of the above to a known result in spectral graph theory (Chung, 1997) (Lemma 1.15) assering that, if ν_k is the k-th eigenvalue of the normalized Laplacian of G and $\widetilde{\nu}_k$ is the k-th eigenvalue of the normalized Laplacian of a graph G_c obtained by edge contraction, then $\nu_k \leq \widetilde{\nu}_k$ for all $k = 1, 2, \ldots, n$. Despite this similarity however, Theorem 4.1 deals with the eigenvalues of the combinatorial Laplacian matrix and its coarsened counterpart $L_c = CLC^\top$.

We also notice that, when the weight vectors c are chosen to be constant over each connected component of G_F (as we assume in this work) the nullspace of L_c spans the downsampled constant vector implying that

$$C^\top \widetilde{u}_1 = u_1 \quad \text{and} \quad 0 = \widetilde{\lambda}_1 < \widetilde{\lambda}_2. \tag{13}$$

The above relations constitute the main reason why we utilize constant coarsening weights in our construction.

4.2. From the RSS property to spectrum preservation

For eigenvalues, the RSS property implies an upper bound:

[2]The existence is implied by the probabilistic method.

Theorem 4.2. *If* L_c *satisfies the RSS property, then*

$$\lambda_k \leq \widetilde{\lambda}_k \leq \max\left\{\widetilde{\lambda}_{k-1},\ \frac{(1+\epsilon_k)}{\sum_{i\geq k}(\widetilde{u}_i^\top C u_k)^2}\,\lambda_k\right\} \tag{14}$$

for all $k \leq K$, *where* ϵ_k *is the* k*-th RSS constant.*

The term $\sum_{i\geq k}(\widetilde{u}_i^\top C u_k)^2$ depends on the orientation of the eigenvectors of L_c with respect to those of L. We expect:

$$\lambda_k \leq \widetilde{\lambda}_k \leq \frac{(1+\epsilon_k)}{\sum_{i\geq k}(\widetilde{u}_i^\top C u_k)^2}\,\lambda_k \approx \frac{(1+\epsilon_k)}{\|\Pi u_k\|_2^2}\,\lambda_k.$$

Indeed, for λ_2 the above becomes an equality as

$$\sum_{i\geq 2}(\widetilde{u}_i^\top C u_2)^2 = \|\Pi u_2\| - (\widetilde{u}_1^\top C u_2)^2 = \|\Pi u_2\|,$$

where the last equality follows from (13). In this case, the above results combined with the analysis presented in Section 3.1 imply the following corollary:

Corollary 4.1. *Consider a bounded degree graph with* $\lambda_2 \leq 0.5 \min_{e_{ij}\in\mathcal{E}}\left\{\frac{d_i+d_j}{2} + w_{ij}\right\}$ *and suppose that it is coarsened by* **REC** *using a heavy-edge potential. For any feasible expected dimensionality reduction ratio* r, *sufficiently large* N *and any* $\epsilon > 0$

$$\widetilde{\lambda}_2 \leq \frac{1 + r\epsilon}{1 - \lambda_2\, r\epsilon}\,\lambda_2,$$

with probability at least $1 - \frac{c_3}{4\,\epsilon}\left(1 + \frac{1.5 w_{max} + 2(1-\lambda_2)}{d_{avg}\ \varrho_{min}}\right)$ *where* $c_3 = r\,(1 - e^{-4\varrho_{max} T/N})$. *For a* d*-regular graph this probability is at least* $1 - \frac{1}{\epsilon}(1 + \frac{3-\lambda_2}{d})$.

The statement can be proved by taking a union bound with respect to the events $\{\|\Pi^\perp u_2\|_2^2 > \lambda_2\, r\epsilon\}$ and $\{u_2^\top \widetilde{L} u_2 > (1 + r\epsilon)\lambda_2\}$, whose probabilities can be easily obtained from the results of Section 3.

Eigenspaces. We also analyze the angle between principal eigenspaces of L and L_c. We follow Li (1994) and split the (lifted) eigendecompositions of L and L_c as

$$L = U\Lambda U^\top = (U_k, U_{k^\perp}) \begin{pmatrix} \Lambda_k & \\ & \Lambda_{k^\perp} \end{pmatrix} \begin{pmatrix} U_k^\top \\ U_{k^\perp}^\top \end{pmatrix}$$

$$\begin{aligned} C^\top L_c C &= (C^\top \widetilde{U})\widetilde{\Lambda}(\widetilde{U}^\top C) \\ &= (C^\top \widetilde{U}_k, C^\top \widetilde{U}_{k^\perp}) \begin{pmatrix} \widetilde{\Lambda}_k & \\ & \widetilde{\Lambda}_{k^\perp} \end{pmatrix} \begin{pmatrix} \widetilde{U}_k^\top C \\ \widetilde{U}_{k^\perp}^\top C \end{pmatrix}, \end{aligned}$$

where $\Lambda_k = \operatorname{diag}(\lambda_1, \ldots, \lambda_k)$ and $U_1 = (u_1, \ldots, u_k)$ (analogously for $\widetilde{\Lambda}_k$ and $\widetilde{U}_k$). The *canonical angles* (Davis & Kahan, 1970; Stewart, 1990) between the eigenspaces

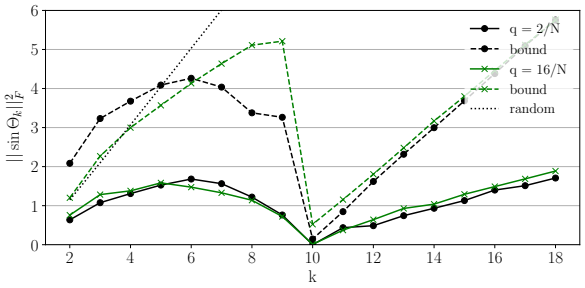


Figure 3: The extend to how coarsening preserves eigenspace alignment is a function of eigenvalue distribution.

spanned by U_k and $C^\top \widetilde{U}_k$ are the singlular values of the matrix

$$\Theta(U_k, C^\top \widetilde{U}_k) \triangleq \arccos(U_k^\top C^\top \widetilde{U}_k \widetilde{U}_k^\top C U_k)^{-1/2} \quad (15)$$

and moreover, the smaller the sinus of the canonical angles are, the closer the two subspaces lie.

The following theorem characterizes $\vartheta_k = \|\sin\Theta\big(U_k, C^\top \widetilde{U}_k\big)\|_F^2$, a measure of the miss-alignment of the eigenspaces spanned by U_k and $C^\top \widetilde{U}_k$.

Theorem 4.3. *If L_c satisfies the RSS property, then*

$$\vartheta_k \leq \min\left\{ \sum_{2\leq i\leq k} \frac{\epsilon_i \lambda_i + \lambda_k \|\Pi^\perp u_i\|_2^2}{\widetilde{\lambda}_{k+1} - \lambda_k}, \right.$$
$$\left. \sum_{2\leq i\leq k} \frac{(1+\epsilon_i)\lambda_i - \lambda_2 \|\Pi u_i\|_2^2}{\widetilde{\lambda}_{k+1} - \lambda_2} \right\},$$

for every $k \leq K$.

Both bounds have something to offer: The first is applicable to situations where there is a significant eigenvalue separation between the subspace of interest and neighboring spaces (this condition also appears in classic perturbation analysis (Davis & Kahan, 1970)) and has the benefit of vanishing when $n = N$. The second bound on the other hand does not depend on the minimum eigengap between λ_k and λ_{k+1}, but on the gap between every eigenvalue λ_i in the subspace of interest and λ_{k+1}, which can be significantly smaller.

We obtain an end-to-end analysis of coarsening by combining Theorem 4.3 with Theorem 3.1 and taking a union bound over all $k \leq K$. However, the reader is urged to consider the proof of Corollary 5.1 for a more careful analysis with significantly improved probability estimates.

The importance of the eigenvalue distribution can be seen in Figure 3, where we examine the alignment of U_k and $C^\top \widetilde{U}_k$ for different k when $r = 0.4$. The figure summarizes the results for 10 stochastic block model graphs, each consisting of $N = 1000$ vertices. These graphs were built by uniformly assigning vertices into $K = 10$ communities and connecting any two vertices with probability p or q depending on whether they belong in the same or different communities, respectively. Such constructions are well known to produce eigenvalue distributions that feature a large gap between the K and $K+1$ eigenvalues and small gaps everywhere else.

Below K the eigenspaces are poorly aligned and not much better than chance (dotted line). As soon as the size of the subspace becomes equal to K however we observe a significant drop, signifying good alignment. This matches the prediction offered by our bounds (dashed line). The phenomenon is replicated for two parametrizations of the stochastic block model, one featuring a low q (and thus a large gap) and one with larger q. Due to the smaller gap, in the latter case the trend is slightly less exaggerated.

5. Implications for spectral clustering

Spectral clustering is a non-linear method for partitioning N points $z_1, z_2, \ldots, z_N \in \mathbb{R}^D$ into K sets $S = \{S_1, S_2, \ldots, S_K\}$. There exist many versions of the algorithm. We consider the "unnormalized spectral clustering" (Von Luxburg, 2007):

1. Construct a similarity graph with $w_{ij} = e^{-\|z_i - z_j\|_2^2/\sigma^2}$ between vertices v_i and v_j. Let L be the combinatorial Laplacian of the graph and write $\Psi = U_K \in \mathbb{R}^{N\times K}$ to denote the matrix of its first K eigenvectors.
2. Among all cluster assignments $\mathcal{S}$, search for the assignment S^* that minimizes the k-means cost:

$$\mathcal{F}_K(\Psi, S) = \sum_{k=1}^{K} \sum_{v_i, v_j \in S_k} \frac{\|\Psi(i,:) - \Psi(j,:)\|_2^2}{2\,|S_k|}$$

Though a naive implementation of the above algorithm scales with $O(N^3)$, the acceleration of spectral clustering has been an active topic of research. A wide-range of sketching techniques have been proposed(Boutsidis et al., 2015a; Tremblay et al., 2016), arguably one of the fastest known algorithms utilizes coarsening. Roughly, the algorithm involves: (*i*) hierarchically coarsening the input graph (using edge contractions) until the latter reaches a target size; (*ii*) solving the clustering problem in the small dimension; (*iii*) lifting the solution back to the original domain; and (*iv*) performing some fast refinement to attain the final clustering.

In the following, we provide theoretical guarantees on the solution quality of the aforementioned scheme for a single coarsening level. To the extend of our knowledge, this is the first time that such an analysis has been carried out.

To perform the analysis, we suppose that

$$S^* = \arg\min_{S\in\mathcal{S}} \mathcal{F}_K(\Psi, S) \text{ and } \widetilde{S}^* = \arg\min_{S\in\mathcal{S}} \mathcal{F}_K(\widetilde{\Psi}, S)$$

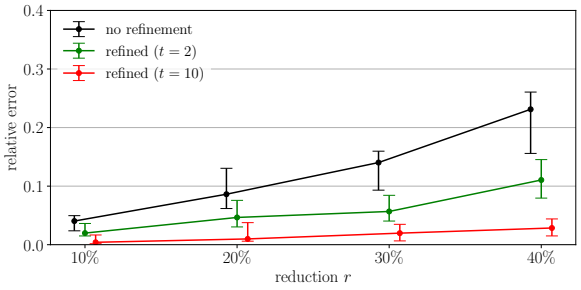


Figure 4: The relative k-means error induced by coarsening as a function of dimensionality reduction r (in percentage) for clustering 5 MNIST digits. The three lines correspond to the lifted eigenvectors with and without refinement. The errorbars span one standard deviation. A small horizontal offset has been inserted in order to diminish overlap.

are the (optimal) clustering assignments obtained by solving the k-means using as input the original eigenvectors U_K and the lifted eigenvectors $\widetilde{\Psi} = C^\top U_K$ of L_c, respectively. We then measure the quality of $\widetilde{S}^*$ by examining how far the correct minimizer $\mathcal{F}_K(\Psi, S^*)$ is to $\mathcal{F}_K(\Psi, \widetilde{S}^*)$. Note that the latter quantity utilizes the correct eigenvectors as points and necessarily $\mathcal{F}_K(\Psi, S^*) \leq \mathcal{F}_K(\Psi, \widetilde{S}^*)$. Boutsidis et al. (2015a) noted that, if the two quantities are close then, despite the assignments themselves possibly being different, they both feature the same quality with respect to the k-means objective.

We prove the following approximation result:

Corollary 5.1. *Consider a bounded degree graph with* $\lambda_K \leq 0.5 \min_{e_{ij} \in \mathcal{E}} \left\{ \frac{d_i + d_j}{2} + w_{ij} \right\}$ *and suppose that it is coarsened by* REC *using a heavy-edge potential. For sufficiently large* N*, any feasible ratio* r*, and* $\epsilon > 0$*,*

$$\left[\mathcal{F}_K(\Psi, \widetilde{S}^*)^{1/2} - \mathcal{F}_K(\Psi, S^*)^{1/2} \right]^2 \leq \sum_{k=2}^{K} \frac{8\epsilon r \lambda_k}{\delta_K}$$

with probability at least $1 - \frac{\varrho_{max}}{\epsilon} \left(1 + \frac{6 + 4\lambda_K - 8\, c_3}{d_{avg}\, \varrho_{min}} \right)$*, where* $\delta_K = \lambda_{K+1} - \lambda_K$ *and* $c_3 = \sum_{k=2}^{K} \lambda_k^2 / \sum_{k=2}^{K} \lambda_k$.

The corollary therefore provides conditions such that the clustering assignment produced with the aid of coarsening has quality that is close to that of the original in terms of absolute error, even without refinement. Practically, our result states that $\widetilde{S}^*$ is a good candidate for the final solution as long as the graph has almost constant degree (such that $\varrho_{\text{min}} \approx 1 \approx \varrho_{\text{max}}$) and it is K clusterable (i.e., the gap $\delta_K = \lambda_{K+1} - \lambda_K$ is large).

We are unaware of any technique that provides meaningful lower bounds on $\mathcal{F}_K(\Psi, S^*)$ and therefore cannot transform the bound to a relative error statement. However, from the algebraic formulation of the k-means cost it follows that, when the number of clusters is $\kappa < K$ whereas the feature matrix remains Ψ, then $\mathcal{F}_\kappa(\Psi, S^*) \geq K - \kappa$ (this is because Ψ has exactly K unit singular values, whereas the k-means clustering cannot do better than a rank κ approximation of Ψ (Ding & He, 2004; Boutsidis et al., 2015b)). Under the conditions of Corollary 5.1 and with the same probability:

$$\left[\frac{\mathcal{F}_K(\Psi, \widetilde{S}^*)^{\frac{1}{2}} - \mathcal{F}_K(\Psi, S^*)^{\frac{1}{2}}}{\mathcal{F}_\kappa(\Psi, S^*)^{1/2}} \right]^2 \leq \sum_{k=2}^{K} \frac{8\epsilon r \lambda_k}{\delta_K (K - \kappa)},$$

which is a relaxed relative error guarantee.

Figure 4 depicts the growth of the actual relative error with r. The particular experiment corresponds to a clustering problem involving $N = 1000$ images, each depicting a selected digit between 0 and 4 from the MNIST database (i.e., $K = 5$). We constructed a 12-nearest neighbor similarity graph and repeated the experiment 10 times, each time using a different image set, selected uniformly at random. This setting produces a simple, but non-trivial, clustering problem featuring some overlaps between clusters (see Figure 5 in the supplementary material).

We observed that most remaining error occurred at coarsened vertices lying at cluster boundaries. This can be eliminated by only a few iterations of local smoothing. Though more advanced techniques might be preferable from a computational perspective, such as Chebychev or ARMA graph filters (Shuman et al., 2011; Isufi et al., 2017), for illustration purposes, we here additionally perform $t = \{2, 10\}$ steps of a simple power iteration scheme, yielding an $O(tM(K-1))$ overhead. The experiment confirms that most errors are removed after few iterations.

6. Discussion

The main message of our work is the following: coarsening locally perturbs individual eigenvectors; however, if carefully constructed, it can leave well-separated principal eigenspaces relatively untouched.

The main limitation of our analysis is that the concentration estimates given by Theorem 3.1 are conservative. We are currently considering methods to improve our bounds by taking into account the dependency structure of the binomial random variables in the sum. In addition, we are investigating how to generalize our analysis to the multi-level setting, where n can be as small as $O(\log N)$. Finally, we are considering the implications of our results to supervised methods for learning from graph-structured data in general, and graph convolutional neural networks in particular.

References

Belkin, M. and Niyogi, P. Laplacian eigenmaps for dimensionality reduction and data representation. *Neural computation*, 15(6):1373–1396, 2003.

Boutsidis, C., Kambadur, P., and Gittens, A. Spectral clustering via the power method-provably. In *International Conference on Machine Learning*, pp. 40–48, 2015a.

Boutsidis, C., Zouzias, A., Mahoney, M. W., and Drineas, P. Randomized dimensionality reduction for k-means clustering. *IEEE Transactions on Information Theory*, 61(2):1045–1062, 2015b.

Bronstein, M. M., Bruna, J., LeCun, Y., Szlam, A., and Vandergheynst, P. Geometric deep learning: Going beyond euclidean data. *IEEE Signal Processing Magazine*, 34(4):18–42, July 2017. ISSN 1053-5888. doi: 10.1109/MSP.2017.2693418.

Bruna, J., Zaremba, W., Szlam, A., and Lecun, Y. Spectral networks and locally connected networks on graphs. In *International Conference on Learning Representations (ICLR2014), CBLS, April 2014*, 2014.

Chen, G., Davis, G., Hall, F., Li, Z., Patel, K., and Stewart, M. An interlacing result on normalized laplacians. *SIAM Journal on Discrete Mathematics*, 18(2):353–361, 2004.

Chung, F. R. *Spectral graph theory*. Number 92. American Mathematical Soc., 1997.

Colley, C., Lin, J., Hu, X., and Aeron, S. Algebraic multigrid for least squares problems on graphs with applications to hodgerank. In *Parallel and Distributed Processing Symposium Workshops (IPDPSW), 2017 IEEE International*, pp. 627–636. IEEE, 2017.

Davis, C. and Kahan, W. M. The rotation of eigenvectors by a perturbation. iii. *SIAM Journal on Numerical Analysis*, 7(1):1–46, 1970.

Defferrard, M., Bresson, X., and Vandergheynst, P. Convolutional neural networks on graphs with fast localized spectral filtering. In *Advances in Neural Information Processing Systems*, pp. 3844–3852, 2016.

Dhillon, I. S., Guan, Y., and Kulis, B. Weighted graph cuts without eigenvectors a multilevel approach. *IEEE transactions on pattern analysis and machine intelligence*, 29 (11), 2007.

Ding, C. and He, X. K-means clustering via principal component analysis. In *Proceedings of the twenty-first international conference on Machine learning*, pp. 29. ACM, 2004.

Gandhi, S. Improvement of the cascadic multigrid algorithm with a gauss seidel smoother to efficiently compute the fiedler vector of a graph laplacian. *arXiv preprint arXiv:1602.04386*, 2016.

Gavish, M., Nadler, B., and Coifman, R. R. Multiscale wavelets on trees, graphs and high dimensional data: Theory and applications to semi supervised learning. In *ICML*, pp. 367–374, 2010.

Hendrickson, B. and Leland, R. W. A multi-level algorithm for partitioning graphs. *SC*, 95(28):1–14, 1995.

Hirani, A. N., Kalyanaraman, K., and Watts, S. Graph laplacians and least squares on graphs. In *Parallel and Distributed Processing Symposium Workshop (IPDPSW), 2015 IEEE International*, pp. 812–821. IEEE, 2015.

Isufi, E., Loukas, A., Simonetto, A., and Leus, G. Autoregressive moving average graph filtering. *IEEE Transactions on Signal Processing*, 65(2):274–288, 2017.

Karypis, G. and Kumar, V. A fast and high quality multilevel scheme for partitioning irregular graphs. *SIAM Journal on scientific Computing*, 20(1):359–392, 1998a.

Karypis, G. and Kumar, V. Multilevelk-way partitioning scheme for irregular graphs. *Journal of Parallel and Distributed computing*, 48(1):96–129, 1998b.

Koren, D. H. Y. A fast multi-scale method for drawing large graphs. *Journal of graph algorithms and applications*, 6 (3):179–202, 2002.

Koutis, I., Miller, G. L., and Tolliver, D. Combinatorial preconditioners and multilevel solvers for problems in computer vision and image processing. *Computer Vision and Image Understanding*, 115(12):1638–1646, 2011.

Kushnir, D., Galun, M., and Brandt, A. Fast multiscale clustering and manifold identification. *Pattern Recognition*, 39(10):1876–1891, 2006.

Lafon, S. and Lee, A. B. Diffusion maps and coarse-graining: A unified framework for dimensionality reduction, graph partitioning, and data set parameterization. *IEEE transactions on pattern analysis and machine intelligence*, 28(9):1393–1403, 2006.

Li, R.-C. Relative perturbation theory:(ii) eigenspace variations. Technical report, 1994.

Livne, O. E. and Brandt, A. Lean algebraic multigrid (lamg): Fast graph laplacian linear solver. *SIAM Journal on Scientific Computing*, 34(4):B499–B522, 2012.

Moitra, A. *Vertex sparsification and universal rounding algorithms*. PhD thesis, Massachusetts Institute of Technology, 2011.

Muja, M. and Lowe, D. G. Scalable nearest neighbor algorithms for high dimensional data. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 36(11): 2227–2240, 2014.

Rossi, R. A. and Ahmed, N. K. The network data repository with interactive graph analytics and visualization. In *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence*, 2015. URL `http://networkrepository.com`.

Shuman, D. I., Vandergheynst, P., and Frossard, P. Chebyshev polynomial approximation for distributed signal processing. In *Distributed Computing in Sensor Systems and Workshops (DCOSS), 2011 International Conference on*, pp. 1–8. IEEE, 2011.

Shuman, D. I., Faraji, M. J., and Vandergheynst, P. A multiscale pyramid transform for graph signals. *IEEE Transactions on Signal Processing*, 64(8):2119–2134, 2016.

Simonovsky, M. and Komodakis, N. Dynamic edge-conditioned filters in convolutional neural networks on graphs. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017.

Spielman, D. A. and Srivastava, N. Graph sparsification by effective resistances. *SIAM Journal on Computing*, 40 (6):1913–1926, 2011.

Spielman, D. A. and Teng, S.-H. Spectral sparsification of graphs. *SIAM Journal on Computing*, 40(4):981–1025, 2011.

Stewart, G. W. Matrix perturbation theory. 1990.

Tremblay, N., Puy, G., Gribonval, R., and Vandergheynst, P. Compressive spectral clustering. In *International Conference on Machine Learning*, pp. 1002–1011, 2016.

Turk, G. and Levoy, M. Zippered polygon meshes from range images. In *Proceedings of the 21st annual conference on Computer graphics and interactive techniques*, pp. 311–318. ACM, 1994.

Urschel, J. C., Hu, X., Xu, J., and Zikatanov, L. T. A cascadic multigrid algorithm for computing the fiedler vector of graph laplacians. *arXiv preprint arXiv:1412.0565*, 2014.

Von Luxburg, U. A tutorial on spectral clustering. *Statistics and computing*, 17(4):395–416, 2007.

Walshaw, C. A multilevel algorithm for force-directed graph-drawing. *Journal of Graph Algorithms and Applications*, 7(3):253–285, 2006.

Wang, L., Xiao, Y., Shao, B., and Wang, H. How to partition a billion-node graph. In *Data Engineering (ICDE), 2014 IEEE 30th International Conference on*, pp. 568–579. IEEE, 2014.