Graph Summarization and Graph Embeddings: Towards A Spectral Connection

Arpit Merchant
University of Helsinki
Finland
arpit.merchant@helsinki.fi

Michael Mathioudakis
University of Helsinki
Finland
michael.mathioudakis@helsinki.fi

ABSTRACT

The graph summarization problem is to define a compressed data structure that can concisely describe the original graph. A standard class of techniques for summarization involves grouping nodes into supernodes via aggregation or clustering such that the l_p-reconstruction error, i.e. the p-norm between the original adjacency matrix and the adjacency matrix recovered from the compressed summary, is minimized. Our main result shows that graph summarization can be reformulated as a trace maximization problem, the relaxed version of which can be solved exactly by all the eigenvectors of the adjacency matrix. We also prove a lower bound on the optimal solution which uses k eigenvectors for a summary with k supernodes. Our results motivate a simple spectral clustering algorithm that can yield excellent summaries. Our experiments validate the quality of the resultant summaries.

KEYWORDS

graph summarization, graph embeddings, spectral graph theory

ACM Reference Format:
Arpit Merchant and Michael Mathioudakis. 2020. Graph Summarization and Graph Embeddings: Towards A Spectral Connection. In *Workshop on Mining and Learning with Graphs '20: ACM Conference on Knowledge Discovery and Data Mining, August 24, 2020, San Diego, CA, USA.* ACM, New York, NY, USA, 7 pages.

1 INTRODUCTION

Visualization and analysis of graphs with millions of vertices and billions of edges can be intractable due to memory and time constraints. Graph summaries facilitate preservation of the original graph topology and efficient processing of queries while requiring significantly less storage space [4]. However, extracting important and interesting aspects of the topology is a subjective question that depends on the context and domain of the graph(s) in question [1, 11]. Some applications may seek summaries that improve the efficiency of querying basic graph properties such as node adjacency, degree, PageRank, number of triangles, etc. [10, 13]; while others may seek summaries that preserve patterns such as reachability, stars, bipartite-cores, and cliques [9, 12].

MLG '20, August 24, 2020, San Diego, CA, USA

Node-grouping techniques are commonly used for constructing summaries for such purposes [11]. As the name suggests, the goal is to group nodes into supernodes based on application-specific error measures. A natural way of obtaining supernodes is by partitioning the original graph into (say) k components via clustering or community detection since structurally similar nodes tend to be connected or close to each other. Thus, summarization, partitioning and clustering are closely related problems [14].

Our work is motivated by spectral graph theory which relates combinatorial properties of graph structures to algebraic properties of associated matrices. Spectra of graphs play a crucial role in graph invariants, sparsification, partitioning and cuts [5]. The main contributions of this paper are:

- We show that minimizing the l_2-reconstruction error of the hypergraph summary can be rewritten as an equivalent trace maximization problem whose relaxation can be solved exactly by all the eigenvectors of the adjacency matrix.
- We show that eigenvectors corresponding to the k largest (in magnitude) eigenvalues of the adjacency matrix provide a lower bound for the relaxed problem of dimension k.
- Experiments on real-world data show that the summary obtained via spectral clustering has low reconstruction error and compares favourably to related methods.

2 RELATED WORK

In this section, we present two broad classes of techniques for graph summarization via node grouping. We refer the reader to an excellent survey by Liu et al. [11] for summarization via edge-grouping, bit-compression, simplification and influence.

Node clustering based approaches. Riondato et al. [14] propose algorithms to construct summaries that minimize the l_p-reconstruction error with a constant-factor approximation guarantee. They define a low-dimensional representation of the adjacency matrix (by reducing the number of rows and columns) and apply k-means clustering to obtain supernodes. Xu et al. [20] use incremental static clustering to identify structurally connected nodes while other heuristics use modularity [8] and minimum cut trees [15].

Node aggregation based approaches. LeFevre and Terzi [10] present GraSS that constructs a lossy representation from which an expected adjacency matrix can be inferred (where expectation is over the set of all graphs compatible with a specific summary). This can efficiently and accurately handle degree, adjacency and eigenvector centrality queries. Koutra et al. [9] propose VoG, an efficient method based on the MDL principle for characterizing important

candidate subgraphs (vocabulary) such as stars and cliques. Dunne and Schneiderman [7] use motif simplification to identify common subgraphs with compact glyphs to improve graph visualization. CoSum by Zhu et al. [21] creates supernodes and superedges with the objective to improve the accuracy of entity resolution.

3 PROBLEM STATEMENT

3.1 Preliminary Definitions

We consider a simple, static, undirected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ with $|\mathcal{V}| = n$. A k-summary S of $\mathcal{G}$ is a weighted, undirected hypergraph defined by a k-partition $(\{V_1, \ldots, V_k\})$ of $\mathcal{V}$. Each component V_i of the partition is referred to as a supernode and there is a superedge between every pair of supernodes V_i and V_j whose weight is the density of edges between them.

$$A_{\mathrm{S}}\left(V_i, V_j\right) = \frac{\sum_{u \in V_i, v \in \mathcal{V}_j} A_{\mathcal{G}}\left(u, v\right)}{|V_i| \cdot |V_j|}$$

where $A_{\mathcal{G}}$ is the adjacency matrix of $\mathcal{G}$ and A_{S} is the weighted adjacency matrix of the summary S.

Following [14], we use the lifted matrix $A_{\mathrm{S}}^{\uparrow} \in \mathbb{R}^{n \times n}$ to denote the adjacency matrix reconstructed from the summary:

$$A_{\mathrm{S}}^{\uparrow}\left(u, v\right) = A_{\mathrm{S}}\left(\mathrm{S}\left(u\right), \mathrm{S}\left(v\right)\right)$$

where $\mathrm{S}(u)$ represents the supernode of S that vertex u belongs to. We define l_2-reconstruction error as the entry-wise 2-norm between $A_{\mathcal{G}}$ and $A_{\mathrm{S}}^{\uparrow}$ for a particular summary S:

$$\mathrm{err}_2\left(A_{\mathcal{G}}, A_{\mathrm{S}}^{\uparrow}\right) = \left\|A_{\mathcal{G}} - A_{\mathrm{S}}^{\uparrow}\right\|_2^2 = \sum_{u=1}^{|\mathcal{V}|} \sum_{v=1}^{|\mathcal{V}|} |A_{\mathcal{G}}\left(u, v\right) - A_{\mathrm{S}}^{\uparrow}\left(u, v\right)|^2 \tag{1}$$

Problem 1: *Given adjacency matrix $A_{\mathcal{G}}$ of graph $\mathcal{G}$ and $k \in \mathbb{N}$, find a k-summary S^* from all possible k-summaries S such that*

$$\mathrm{S}^* = \arg\min_{\mathrm{S}} \mathrm{err}_2\left(A_{\mathcal{G}},\, A_{\mathrm{S}}^{\uparrow}\right) \tag{2}$$

Let $X_{\mathrm{S}} \in \{0,1\}^{n \times k}$ represent the one-hot encoding or membership matrix where the (i,j)-th entry denotes that vertex i belongs to supernode V_j of summary S. Denote $Z_{\mathrm{S}} = X_{\mathrm{S}}\left(X_{\mathrm{S}}^{\top} X_{\mathrm{S}}\right)^{-1/2} = [z_1\, z_2\, \ldots\, z_k]$ as the normalized membership matrix where the column vectors $\{z_i\}_{i \in [k]}$ are orthonormal. Define the smoothing projection matrix $P_{\mathrm{S}} = Z_{\mathrm{S}} Z_{\mathrm{S}}^{\top} \in \mathbb{R}^{n \times n}$ to be the orthogonal projection onto the subspace generated by partition vectors $\{z_i\}_{i \in [k]}$. Note that $P_{\mathrm{S}}^2 = P_{\mathrm{S}}$, $P_{\mathrm{S}}^{\top} = P_{\mathrm{S}}$ and $P_{\mathrm{S}} z_i = z_i$.

The following result connects the lifted matrix with the adjacency and smoothing projection matrices.

Lemma 1 (Lemma 3 - Riondato et al. [14]): $A_{\mathrm{S}}^{\uparrow} = P_{\mathrm{S}} A_{\mathcal{G}} P_{\mathrm{S}}$

3.2 Graph Summarization as Trace Maximization

At first glance, spectral clustering and graph summarization appear to be unrelated problems. Spectral clustering uses eigenvectors to determine partitions and eigenvectors are absent from Equation 1. In this section, we show that graph summarization can be expressed as a trace maximization problem whose relaxed version can be solved exactly by the eigenvectors of the adjacency matrix.

Using the facts that $\|L\|_2^2 = \mathbf{tr}\left[L^{\top} L\right]$, $\mathbf{tr}\left[L + M\right] = \mathbf{tr}\left[L\right] + \mathbf{tr}\left[M\right]$, $\mathbf{tr}\left[cL\right] = c \cdot \mathbf{tr}\left[L\right]$, and from trace invariance under cyclic permutation, $\mathbf{tr}\left[LMN\right] = \mathbf{tr}\left[MNL\right] = \mathbf{tr}\left[NLM\right]$, we get that[1]:

$$\begin{aligned}
\mathrm{err}_2\left(A_{\mathcal{G}}\, A_{\mathrm{S}}^{\uparrow}\right)^2 &= \left\|A_{\mathcal{G}}, -A_{\mathrm{S}}^{\uparrow}\right\|_2^2 \\
&= \mathbf{tr}\left[\left(A_{\mathcal{G}} - A_{\mathrm{S}}^{\uparrow}\right)^{\top}\left(A_{\mathcal{G}} - A_{\mathrm{S}}^{\uparrow}\right)\right] \\
&= \mathbf{tr}\left[\left(A - PAP\right)^{\top}\left(A - PAP\right)\right] \\
&= \mathbf{tr}\left[\left(A - PAP\right)\left(A - PAP\right)\right] \\
&= \mathbf{tr}\left[A^2 - APAP - PAPA + PAPPAP\right] \\
&= \mathbf{tr}\left[A^2 - APAP\right] - \mathbf{tr}\left[PAPA\right] + \mathbf{tr}\left[PAPPAP\right] \\
&= \mathbf{tr}\left[A^2 - APAP\right] - \mathbf{tr}\left[PAPA\right] + \mathbf{tr}\left[PPAPPA\right] \\
&= \mathbf{tr}\left[A^2 - APAP\right] - \cancel{\mathbf{tr}\left[PAPA\right]} + \cancel{\mathbf{tr}\left[PAPA\right]} \\
&= \mathbf{tr}\left[A^2\right] - \mathbf{tr}\left[APAP\right] \\
&= \mathbf{tr}\left[A^2\right] - \mathbf{tr}\left[AZZ^{\top}AZZ^{\top}\right] \\
&= \mathbf{tr}\left[A^2\right] - \mathbf{tr}\left[Z^{\top}AZZ^{\top}AZ\right] \\
&= \mathbf{tr}\left[A^2\right] - \mathbf{tr}\left[\left(Z_{\mathrm{S}}^{\top} A Z_{\mathrm{S}}\right)^2\right]
\end{aligned} \tag{3}$$

where the last equation follows from substituting $P = ZZ^{\top}$ and again using the invariance under cyclic permutation property.

We can now rewrite norm minimization in Problem 1 equivalently as trace minimization.

Problem 2: *Given adjacency matrix $A_{\mathcal{G}}$ of graph $\mathcal{G}$ and $k \in \mathbb{N}$, find a k-summary S^* from all possible k-summaries S such that*

$$\begin{aligned}
\mathrm{S}^* = \underset{\mathrm{S}}{\mathrm{argmin}} \quad & \mathbf{tr}\left[A_{\mathcal{G}}^2\right] - \mathbf{tr}\left[\left(Z_{\mathrm{S}}^{\top} A_{\mathcal{G}} Z_{\mathrm{S}}\right)^2\right] \\
\text{s.t.} \quad & Z_{\mathrm{S}}^{\top} Z_{\mathrm{S}} = I, \quad Z_{\mathrm{S}} = X_{\mathrm{S}}\left(X_{\mathrm{S}}^{\top} X_{\mathrm{S}}\right)^{-1/2} \\
& X_{\mathrm{S}} \in \{0,1\}^{n \times k}
\end{aligned} \tag{4}$$

Since $\mathbf{tr}\left[A^2\right]$ is a constant, to minimize the objective function in Problem 2 is to maximize $\mathbf{tr}\left[\left(Z_{\mathrm{S}}^{\top} A Z_{\mathrm{S}}\right)^2\right]$. That is, Problems 1, 2, and 3 are equivalent.

Problem 3: *Given adjacency matrix $A_{\mathcal{G}}$ of graph $\mathcal{G}$ and $k \in \mathbb{N}$, find a k-summary S^* from all possible k-summaries S such that*

$$\begin{aligned}
\mathrm{S}^* = \underset{\mathrm{S}}{\mathrm{argmax}} \quad & \mathbf{tr}\left[\left(Z_{\mathrm{S}}^{\top} A_{\mathcal{G}} Z_{\mathrm{S}}\right)^2\right] \\
\text{s.t.} \quad & Z_{\mathrm{S}}^{\top} Z_{\mathrm{S}} = I, \quad Z_{\mathrm{S}} = X_{\mathrm{S}}\left(X_{\mathrm{S}}^{\top} X_{\mathrm{S}}\right)^{-1/2} \\
& X_{\mathrm{S}} \in \{0,1\}^{n \times k}
\end{aligned} \tag{5}$$

[1] We drop $\mathcal{G}$ and S from the subscript for notational convenience.

4 METHOD

We now present our solution to Problem 3. Our approach is similar in flavour to the one for computing normalized cuts [18] which also can be expressed as a trace optimization (different than ours) problem. We first relax the binary constraint on the entries of X_S and allow Z_S to be any arbitrary orthonormal matrix. We obtain candidate solutions to the relaxed problem. Finally, we round the candidate solution to its binary form via k-means clustering to obtain the summary. For our analysis, we consider three cases based on the dimension of the problem: (1) $k = n$, (2), $k = 1$, and (3) for any general $k \in 2, \ldots, n-1$.

4.1 Case 1: $k = n$

Consider the case where Z_S comprises of n orthonormal column vectors, i.e. $Z_S \in \mathbb{R}^{n \times n}$. The following result shows that all eigenvectors of $A_{\mathcal{G}}$ exactly solve the relaxed version of Problem 2.

Theorem 1: Let $A_{\mathcal{G}} = B\Lambda B^\top$ be the eigendecomposition of $A_{\mathcal{G}}$ where columns of B are it's eigenvectors and B is orthonormal. Let Z^* be the optimal solution to Equation 4. That is,

$$Z^* = \arg\min_{Z} \mathbf{tr}\left[A^2\right] - \mathbf{tr}\left[\left(Z^\top A Z\right)^2\right] \text{ subject to } Z^\top Z = I, Z \in \mathbb{R}^{n \times n} \tag{6}$$

Then, $Z^* = B$.

Proof. Since Equation 1 is the 2-norm and is equivalent to Equation 4, it's minimum value must be non-negative. Substituting $Z = B$ into Equation 4, we get

$$\begin{aligned}
\mathbf{tr}\left[A^2 - \left(Z^\top A Z\right)^2\right] &= \mathbf{tr}\left[A^2 - B\Lambda B^\top Z Z^\top B\Lambda B^\top Z Z^\top\right] \\
&= \mathbf{tr}\left[A^2 - \left(B\Lambda\left(B^\top B\right)B^\top\right)\left(B\Lambda\left(B^\top B\right)B^\top\right)\right] \\
&= \mathbf{tr}\left[A^2 - B\Lambda B^\top B\Lambda B^\top\right] \\
&= \mathbf{tr}\left[A^2 - A^2\right] \\
&= 0
\end{aligned}$$

□

4.2 Case 2: $k = 1$

Lets consider the case when $Z = z \in \mathbb{R}^n$ is a single column vector of unit norm. Then $\mathbf{tr}\left[\left(Z^\top A Z\right)^2\right] = \mathbf{tr}\left[\left(z^\top A z\right)^2\right] = \left(z^\top A z\right)^2$. The following result shows that Problem 3 is optimized by the eigenvector corresponding to the largest (in magnitude) eigenvalue of A.

Lemma 2. *Let $z = z^*$ be a non-zero vector that maximizes the following quantity with respect to a fixed A:*

$$R(z) = \frac{\left(z^\top A z\right)^2}{\left(z^\top z\right)^2}$$

Then, $Az^* = \lambda_{max} z^*$, where λ_{max} is the largest (in magnitude) eigenvalue of A and z^* is the corresponding eigenvector.

Proof. The proof follows similarly to Theorem 2.2.1 [17]. Since the Rayleigh quotient is homogeneous[2], the square of the Rayleigh quotient is also homogeneous. And so it suffices to consider unit vectors z. Since the set of unit vectors is closed and compact, the function has a maximum value. So we have,

$$\nabla \frac{\left(z^\top A z\right)^2}{\left(z^\top z\right)^2} = \frac{4\left(z^\top A z\right)(Az)\left(z^\top z\right)^2 - 4\left(z^\top A z\right)^2 \cdot \left(z^\top z\right) \cdot (z)}{\left(\left(z^\top z\right)^2\right)^2}$$

Let z^* be a non-zero vector that maximizes $R(z)$. The gradient of a function at it's maximum value must equal the zero vector. Therefore,

$$\begin{aligned}
\nabla \frac{\left(\left(z^*\right)^\top A z^*\right)^2}{\left(\left(z^*\right)^\top z^*\right)^2} &= 0 \\
Az^* &= \left(\frac{\left(z^*\right)^\top A z^*}{\left(z^*\right)^\top z^*}\right) \cdot z^*
\end{aligned}$$

This implies, that z^* maximizes $R(z)$ if and only if z^* is an eigenvector of A with eigenvalue equal to the Rayleigh quotient. And therefore the maximum value of $R(z) = \lambda_{\max}^2$ where z^* is the corresponding eigenvector.

□

4.3 Case 3: $k \in \{2, \ldots, n-1\}$

Lets first write Equation 5 in vector form for any $k \in \{2, \ldots, n-1\}$ where $Z_S = [z_1, z_2, \ldots, z_k]$. We have,

$$\begin{aligned}
\mathbf{tr}\left[\left(Z_S^\top A Z_S\right)^2\right] &= \mathbf{tr}\left[\left(\left[z_1, \ldots, z_k\right]^\top A \left[z_1, \ldots, z_k\right]\right)^2\right] \\
&= \mathbf{tr}\left[\begin{pmatrix} z_1^\top A z_1 & z_1^\top A z_2 & \cdots & z_1^\top A z_k \\ z_2^\top A z_1 & z_2^\top A z_2 & \cdots & z_2^\top A z_k \\ \vdots & \vdots & \ddots & \vdots \\ z_k^\top A z_1 & z_k^\top A z_2 & \cdots & z_k^\top A z_k \end{pmatrix}^2\right] \\
&= \sum_{j=1}^{k}\left(z_j^\top A z_j\right)^2 + \sum_{j=1}^{k}\sum_{i \in [k] \setminus \{j\}} z_j^\top A z_i z_i^\top A z_j
\end{aligned} \tag{7}$$

Problem 4: *Given adjacency matrix $A_{\mathcal{G}}$ of graph $\mathcal{G}$, find k orthonormal vectors $z_1, \ldots, z_k$ that optimize the following:*

$$\begin{aligned}
\arg\max_{z_1, \ldots, z_k} \quad & \sum_{j=1}^{k}\left(z_j^\top A z_j\right)^2 + \sum_{j=1}^{k}\sum_{i \in [k], \setminus \{j\}} z_j^\top A z_i z_i^\top A z_j \\
\text{subject to} \quad & \forall i \in [k]\, z_i^\top z_i = 1 \\
& \forall i, j \in [k],\, i \neq j,\, z_i^\top z_j = 0
\end{aligned} \tag{8}$$

Since A is symmetric, $\forall i \neq j, z_j^\top A z_i = z_i^\top A z_j$. Each term in the objective function of Equation 8 is the square of a scalar and is thus nonnegative. Therefore, the optimal value of Problem 4 is at least as much as the maximum value of $\sum_{j=1}^{k}\left(z_j^\top A z_j\right)^2$. Consider the subspace orthogonal to the subspace defined by the first (say) m largest (in magnitude) eigenvectors of $A_{\mathcal{G}}$. The following result shows that the unit vector z from this orthogonal subspace that maximizes $\left(z^\top A z\right)^2$ is the $m+1$-th largest eigenvector of $A_{\mathcal{G}}$. Subsequently, the maximum value of $\sum_{j=1}^{k}\left(z_j^\top A z_j\right)^2$ is achieved by eigenvectors

[2] A function is called homogeneous with degree k, if it satisfies the condition $f(\alpha x, \alpha y) = \alpha^k f(x, y)$.

corresponding to the k largest (in magnitude) eigenvalues of $A_{\mathcal{G}}$. Moreover, it's value is the sum of the corresponding eigenvalues.

Theorem 2. *Given adjacency matrix A of graph $\mathcal{G}$, there exist numbers $\mu_1, \ldots, \mu_k$ and orthonormal vectors $\psi_1, \ldots, \psi_k$ such that $\forall i \in [n]$, $A\psi_i = \mu_i \psi_i$. And, for $2 \leq i \leq k$,*

$$\psi_i \in \underset{\substack{\|z\|=1 \\ z^\top \psi_j = 0, \text{for } j<i}}{\arg\max} \left(z^\top A z\right)^2 \tag{9}$$

Proof. Similarly to the Spectral Theorem, we show this result by induction. Let $\lambda^2_{\min}$ be the minimum value of $R(z)$ for some vector $z_{\min}$. Matrices A and $\tilde{A} = A + \left(1 - \mu^2_{\min}\right)\mathbf{I}$ have the same eigenvectors. For all unit norm vectors z, $\tilde{A}$ is positive definite because $z^\top \tilde{A} z = z^\top A z + 1 - \mu^2_{\min} \geq 1$. So it suffices to prove the result for positive definite matrices.

The base case is true for ψ_1 due to Lemma 2. Assume that Equation 9 holds for the first m eigenvectors $\psi_1, \ldots, \psi_m$. We now show that the result is valid for $i = m+1$ and ψ_{m+1}. Define,

$$A_m = A - \sum_{i=1}^{m} \mu_i \psi_i \psi_i^\top.$$

For all $j \leq m$, due to the orthogonality of eigenvectors, we have

$$\begin{aligned} A_m \psi_j &= A\psi_j - \sum_{i=1}^{m} \mu_i \psi_i \psi_i^\top \psi_j \\ &= A\psi_j - \mu_j \psi_j \\ &= 0 \end{aligned} \tag{10}$$

For all vectors z orthogonal to $\psi_1, \ldots, \psi_m$, we have

$$\begin{aligned} A_m z &= Az \\ \left(z^\top A_m z\right)^2 &= \left(z^\top A z\right)^2 \\ \underset{\substack{\|z\|=1 \\ z^\top \psi_j = 0, j \leq m}}{\arg\max} \left(z^\top A z\right)^2 &= \underset{\substack{\|z\|=1 \\ z^\top \psi_j = 0, j \leq m}}{\arg\max} \left(z^\top A_m z\right)^2 \subseteq \underset{\|z\|=1}{\arg\max} \left(z^\top A_m z\right)^2 \end{aligned} \tag{11}$$

Consider a unit vector $\mathbf{u}$ that maximizes $\left(z^\top A_m z\right)^2$. Since A_m is a symmetric matrix, according to Lemma 2, $\mathbf{u}$ must be an eigenvector of A_m. If we show that $\mathbf{u}$ is orthogonal to $\psi_1, \ldots, \psi_m$, then from Equation 11, we know that $\mathbf{u}$ is also an eigenvector of A. Define the projection of $\mathbf{u}$ orthogonal to $\psi_1, \ldots, \psi_m$.

$$\tilde{\mathbf{u}} = \mathbf{u} - \sum_{j=1}^{m} \psi_j \left(\psi_j^\top \mathbf{u}\right)$$

Algorithm 1: Spectral Summarization with l_2-reconstruction error

```
Input: G = (V, E); k, l ∈ ℕ.
Output: k-summary of G under l2-reconstruction error.
/* Compute top-l (in magnitude) eigenvectors */
Z ← computeEigenvectors(A_G, l)
/* Cluster rows of Z ∈ ℝ^{n×l} */
S ← getClusteringPartition(Z, k)
/* Compute densities for the summary */
A_S ← computeDdensities(S, A_G)
return (S, A_S)
```

If $\tilde{\mathbf{u}} = \mathbf{u}$, then we are done. We show this by contradition. Say that there exists some $\left(\psi_i^\top \mathbf{u}\right) \neq 0$. This implies, $\|\tilde{\mathbf{u}}\| < \|\mathbf{u}\|$. We have

$$\begin{aligned} \tilde{\mathbf{u}}^\top A_m \tilde{\mathbf{u}} &= \tilde{\mathbf{u}}^\top A_m \left(\mathbf{u} - \sum_{j=1}^{m} \psi_j \left(\psi_j^\top \mathbf{u}\right)\right) \\ &= \tilde{\mathbf{u}}^\top A_m \mathbf{u} - \tilde{\mathbf{u}}^\top \left(\sum_{j=1}^{m} \cancel{\left(A_m \psi_j\right)} \left(\psi_j^\top \mathbf{u}\right)\right) \\ &= \tilde{\mathbf{u}}^\top A_m \mathbf{u} \\ &= \left(\mathbf{u} - \sum_{j=1}^{m} \psi_j \left(\psi_j^\top \mathbf{u}\right)\right)^\top A_m \mathbf{u} \\ &= \mathbf{u}^\top A_m \mathbf{u} \end{aligned} \tag{12}$$

So $\left(\tilde{\mathbf{u}}^\top A_m \tilde{\mathbf{u}}\right)^2 = \left(\mathbf{u}^\top A_m \mathbf{u}\right)^2$.

Define $\hat{\mathbf{u}} = \tilde{\mathbf{u}} / \|\tilde{\mathbf{u}}\|$. Substituting into Equation 12, we get

$$\begin{aligned} \left(\tilde{\mathbf{u}}^\top A_m \tilde{\mathbf{u}}\right)^2 &= \left(\mathbf{u}^\top A_m \mathbf{u}\right)^2 \\ \left(\left(\|\tilde{\mathbf{u}}\|\, \hat{\mathbf{u}}\right)^\top A_m \left(\|\tilde{\mathbf{u}}\|\, \hat{\mathbf{u}}\right)\right)^2 &= \left(\|\mathbf{u}\|\, \mathbf{u}^\top A_m \mathbf{u}\, \|\mathbf{u}\|\right)^2 \\ \left(\frac{\|\tilde{\mathbf{u}}\|^2}{\|\mathbf{u}\|^2}\right)^2 \left(\hat{\mathbf{u}}^\top A_m \hat{\mathbf{u}}\right)^2 &= \left(\mathbf{u}^\top A_m \mathbf{u}\right)^2 \end{aligned} \tag{13}$$

where the equality holds because $\mathbf{u}$ is a unit vector. But $\|\tilde{\mathbf{u}}\|^2 / \|\mathbf{u}\|^2 < 1$ and therefore $\left(\hat{\mathbf{u}}^\top A_m \hat{\mathbf{u}}\right)^2 > \left(\mathbf{u}^\top A_m \mathbf{u}\right)^2$. This is a contradiction because by definition, $\mathbf{u}$ maximizes $\left(z^\top A_m z\right)^2$ for all unit vectors z. Therefore $\tilde{\mathbf{u}} = \mathbf{u}$ and $\mathbf{u}$ is orthogonal to $\psi_1, \ldots, \psi_m$. We can thus set $\mathbf{u} = \psi_{m+1}$ and this completes the proof. □

4.4 Algorithm

With these results, we motivate a spectral algorithm for graph summarization. Consider the eigenvector coordinates as a spectral embedding of the nodes of a graph. These represent good candidate solutions to the relaxed problem. By clustering these points, we round the relaxed solution to obtain supernodes for our summary. Algorithm 1 presents the pseudocode. We define two parameters, k represents the size of the summary and l represents the number of eigenvectors of $A_{\mathcal{G}}$ used for clustering.

Time Complexity. In Step 1, we compute top l eigenvectors[3] using fast iterative methods such as Lanczos and subspace embeddings [6] that compute each single eigenvector in $\tilde{O}\left(m + \frac{n \cdot s(A_{\mathcal{G}})}{\max\{\text{gap}^{2.5}\epsilon, \epsilon^{2.5}\}}\right)$ (where ϵ is the accuracy parameter, $s\left(A_{\mathcal{G}}\right)$ is the stable rank and gap is the relative eigengap) with $\tilde{O}(\cdot)$ hiding log factors in n and gap. For Step 2, we use the lightweight coreset algorithm by Bachem, et al [2] that runs independent of the number of clusters k. The running time of this algorithm is $O(nl)$. In Step 3, densities can be computed in time $O\left(m + k^2\right)$. If we choose $l = O(\log n)$, this gives an overall runtime of $\tilde{O}\left(m + n + k^2\right)$ upto log factors.

Discussion. The primary distinction between the solutions to normalized cut and graph summarization is that the former uses eigenvectors corresponding to k *algebraically* largest eigenvalues of $A_{\mathcal{G}}$ while the latter uses eigenvectors corresponding to k largest *in magnitude* eigenvalues. Moreover, we only provide a lower bound on the optimal solution for the relaxed problem. But we conjecture that this lower bound is tight. In our future work, we aim to resolve this conjecture by better understanding Problem 5. Lastly, we note that our relaxation approach may not be able to provide a constant-factor approximation guarantee to the integer problem. In general, approximate balanced graph cut problems can be NP-hard themselves [3].

5 EXPERIMENTS

In this section, we empirically evaluate the quality of the summaries built by our algorithm in terms of the l_2-reconstruction error. We compare the performance of our algorithm with that of Riondato, et al [14].

Setup. We use `Cora` from Sen, et al. [16] and `Facebook`, and `Enron` graphs from the SNAP repository[4]. All graphs are simple, undirected, and unweighted. $A_{\mathcal{G}}$ thus has binary entries and $\text{err}_2\left(A_{\mathcal{G}}, A_S^{\uparrow}\right) \in [0, n]$. We report results for the err_2 divided by n for normalization (which may result in differences between our reported values). We compute the l_2 norm directly without approximations. We extend the C++11 implementation[5] [14] for our experiments. Our implementation of Algorithm 1 does not use the lightweight coreset construction by Bachem, et al [2] since it would not affect the performance for the sparse, small graphs we consider. For the algorithm by Riondato, et al [14], we similarly do not use sketching and approximate distance computations. We use Lloyd's iterative procedure with `kmeans++` initialization for k-means clustering in both algorithms. Our experiments are performed on a 4-core AMD processor with 32GB RAM running Linux CentOS 7.6.

Algorithm 1 (SS) takes k (number of clusters) and l (number of top eigenvectors) as input while the algorithm by Riondato, et al [14] (S2L) takes only k as input. Table 1 presents the results averaged over 5 runs for each combination of parameters.

[3]The full eigendecomposition of $A_{\mathcal{G}}$ can be computed in $O(n^{\omega})$ where $\omega < 2.373$ is the matrix multiplication constant [19].
[4]http://snap.stanford.edu/data/
[5]https://github.com/rionda/graphsumm

Size of summary. For $k = 1$ and for other smaller values of k, we find that the standard deviation is relatively large. This is especially true for the small, sparse graphs we study in our experiments. So choose the values of k to be a significant fraction of the total number of nodes in the graph. As Theorem 1 showed, l_2 error is minimized for the trivial summary when $k = n$. This implies, summaries with low error can be achieved for large values of k. Note, setting $k = n$ will simply recover the original adjacency matrix. The runtime of both, SS and S2L, increases linearly as k increases.

Number of eigenvectors. We choose l, the number of eigenvectors, to range from 1 (only the top eigenvector) to $l = n$ (full eigendecomposition). We also choose 2 intermediate values for l to match the size of the summary. As expected, the runtime increases as l increases. Since we do not use a fast subspace embedding implementation, computing the eigenvectors is expensive. Our algorithm therefore, runs slower than S2L.

l_2-reconstruction error. As the size of the summary increases, the reconstruction error decreases for both algorithms. Similarly, as the number of eigenvectors increases the l_2 error reduces. For all three graphs, our algorithm compares favourably against S2L for an appropriate choice of l.

6 CONCLUSION

In this paper, we demonstrated a novel spectral connection to graph summarization by rewriting it as an equivalent trace maximization problem. This has a similar (but not equivalent) form to the k-way normalized cut problem. Motivated by this, we investigate its relaxed version. We find that clustering nodes using eigenvectors corresponding to the top-k largest (in magnitude) eigenvalues of the adjacency matrix effectively captures the community structure and leads to useful summaries with low reconstruction error.

REFERENCES

[1] Charu C Aggarwal and Haixun Wang. 2010. A survey of clustering algorithms for graph data. In *Managing and mining graph data*. Springer, 275–301.
[2] Olivier Bachem, Mario Lucic, and Andreas Krause. 2018. Scalable k-means clustering via lightweight coresets. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. 1119–1127.
[3] Thang Nguyen Bui and Curt Jones. 1992. Finding good approximate vertex and edge partitions is NP-hard. *Inform. Process. Lett.* 42, 3 (1992), 153–159.
[4] Flavio Chierichetti, Ravi Kumar, Silvio Lattanzi, Michael Mitzenmacher, Alessandro Panconesi, and Prabhakar Raghavan. 2009. On compressing social networks. In *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*. 219–228.
[5] Fan RK Chung and Fan Chung Graham. 1997. *Spectral graph theory*. Number 92. American Mathematical Soc.
[6] Kenneth L Clarkson and David P Woodruff. 2017. Low-rank approximation and regression in input sparsity time. *Journal of the ACM (JACM)* 63, 6 (2017), 1–45.
[7] Cody Dunne and Ben Shneiderman. 2013. Motif simplification: improving network visualization readability with fan, connector, and clique glyphs. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 3247–3256.
[8] Robert Görke, Pascal Maillard, Christian Staudt, and Dorothea Wagner. 2010. Modularity-driven clustering of dynamic graphs. In *International Symposium on Experimental Algorithms*. Springer, 436–448.
[9] Danai Koutra, U Kang, Jilles Vreeken, and Christos Faloutsos. 2014. Vog: Summarizing and understanding large graphs. In *Proceedings of the 2014 SIAM international conference on data mining*. SIAM, 91–99.
[10] Kristen LeFevre and Evimaria Terzi. 2010. GraSS: Graph structure summarization. In *Proceedings of the 2010 SIAM International Conference on Data Mining*. SIAM, 454–465.
[11] Yike Liu, Tara Safavi, Abhilash Dighe, and Danai Koutra. 2018. Graph summarization methods and applications: A survey. *ACM Computing Surveys (CSUR)* 51, 3 (2018), 1–34.

Dataset	k	Algorithm	l (for SS)	l_2-reconstr. err.		Runtime(s)
				avg ($\times 10^{-2}$)	stdev ($\times 10^{-3}$)	
Cora $\lvert\mathcal{V}\rvert = 2708$ $\lvert\mathcal{E}\rvert = 5429$	100	SS	1	0.29	0.1	0.01
			100	0.14	0.1	0.11
			500	0.09	0.1	0.2
			2708	0.08	0.1	0.9
		S2L		0.13	0.2	0.07
	250	SS	1	0.28	0.1	0.03
			100	0.09	0.2	0.22
			500	0.08	0.2	0.31
			2708	0.06	0.1	1.1
		S2L		0.11	0.2	0.15
	500	SS	1	0.26	0.1	0.09
			100	0.11	0.1	0.31
			500	0.06	0.1	0.43
			2708	0.04	0.2	2.7
		S2L		0.09	0.2	0.32
Facebook $\lvert\mathcal{V}\rvert = 4039$ $\lvert\mathcal{E}\rvert = 88,329$	500	SS	1	0.97	0.3	0.1
			500	0.49	0.2	2.8
			1000	0.38	0.1	3.9
			4039	0.34	0.1	6.4
		S2L		0.43	0.1	1.1
	750	SS	1	0.91	0.3	0.4
			500	0.42	0.2	3.5
			1000	0.36	0.1	4.5
			4039	0.33	0.1	6.9
		S2L		0.38	0.1	1.8
	1000	SS	1	0.90	0.3	0.2
			500	0.41	0.2	3.1
			1000	0.31	0.1	4.7
			4039	0.27	0.1	8.2
		S2L		0.33	0.1	2.5
Enron $\lvert\mathcal{V}\rvert = 36,692$ $\lvert\mathcal{E}\rvert = 183,831$	6000	SS	1	0.2031	0.04	17
			6000	0.0118	0.01	131
			10000	0.0094	0.01	153
			36,692	0.0091	0.01	334
		S2L		0.0093	0.01	110
	8000	SS	1	0.1877	0.04	26
			6000	0.0094	0.01	183
			10000	0.0071	0.01	221
			36,692	0.0065	0.01	405
		S2L		0.0070	0.01	170
	10000	SS	1	0.1625	0.05	53
			6000	0.0063	0.01	256
			10000	0.0051	0.01	311
			36,692	0.0034	0.01	487
		S2L		0.0053	0.01	250

Table 1: l_2-reconstruction error comparison between Spectral Summarization (SS) and Riondato, et al [14] (S2L) on `Cora`, `Facebook`, and `Enron`. The reported results are averaged over five runs.

[12] Antonio Maccioni and Daniel J Abadi. 2016. Scalable pattern matching over compressed graphs via dedensification. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 1755–1764.

[13] Hossein Maserrat and Jian Pei. 2010. Neighbor query friendly compression of social networks. In *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining*. 533–542.

[14] Matteo Riondato, David García-Soriano, and Francesco Bonchi. 2017. Graph summarization with quality guarantees. *Data mining and knowledge discovery* 31, 2 (2017), 314–349.

[15] Barna Saha and Pabitra Mitra. 2007. Dynamic algorithm for graph clustering using minimum cut tree. In *Proceedings of the 2007 SIAM International Conference on Data Mining*. SIAM, 581–586.

[16] Prithviraj Sen, Galileo Namata, Mustafa Bilgic, Lise Getoor, Brian Galligher, and Tina Eliassi-Rad. 2008. Collective classification in network data. *AI magazine* 29, 3 (2008), 93–93.

[17] Daniel A Spielman. 2007. Spectral graph theory and its applications. In *48th Annual IEEE Symposium on Foundations of Computer Science (FOCS'07)*. IEEE, 29–38.

[18] X Yu Stella and Jianbo Shi. 2003. Multiclass spectral clustering. In *null*. IEEE, 313.

[19] Virginia Vassilevska Williams. 2012. Multiplying matrices faster than Coppersmith-Winograd. In *Proceedings of the forty-fourth annual ACM symposium on Theory of computing*. 887–898.

[20] Zhiqiang Xu, Yiping Ke, Yi Wang, Hong Cheng, and James Cheng. 2012. A model-based approach to attributed graph clustering. In *Proceedings of the 2012 ACM SIGMOD international conference on management of data*. 505–516.

[21] Linhong Zhu, Majid Ghasemi-Gol, Pedro Szekely, Aram Galstyan, and Craig A Knoblock. 2016. Unsupervised entity resolution on multi-type graphs. In *International semantic web conference*. Springer, 649–667.