Submodular Hypergraphs: p-Laplacians, Cheeger Inequalities and Spectral Clustering

Pan Li [1] **Olgica Milenkovic** [1]

Abstract

We introduce submodular hypergraphs, a family of hypergraphs that have different submodular weights associated with different cuts of hyperedges. Submodular hypergraphs arise in clustering applications in which higher-order structures carry relevant information. For such hypergraphs, we define the notion of p-Laplacians and derive corresponding nodal domain theorems and k-way Cheeger inequalities. We conclude with the description of algorithms for computing the spectra of 1- and 2-Laplacians that constitute the basis of new spectral hypergraph clustering methods.

1. Introduction

Spectral clustering algorithms are designed to solve a relaxation of the graph cut problem based on graph Laplacians that capture pairwise dependencies between vertices, and produce sets with small conductance that represent clusters. Due to their scalability and provable performance guarantees, spectral methods represent one of the most prevalent graph clustering approaches (Chung, 1997; Ng et al., 2002).

Many relevant problems in clustering, semisupervised learning and MAP inference (Zhou et al., 2007; Hein et al., 2013; Zhang et al., 2017) involve higher-order vertex dependencies that require one to consider hypergraphs instead of graphs. To address spectral hypergraph clustering problems, several approaches have been proposed that typically operate by first projecting the hypergraph onto a graph via *clique expansion* and then performing spectral clustering on graphs (Zhou et al., 2007). Clique expansion involves transforming a weighted hyperedge into a weighted clique such that the graph cut weights approximately preserve the cut weights of the hyperedge. Almost exclusively, these approximations have been based on the assumption that each hyperedge cut has the same weight, in which case the underlying hypergraph is termed *homogeneous*.

However, in image segmentation, MAP inference on Markov random fields (Arora et al., 2012; Shanu et al., 2016), network motif studies (Li & Milenkovic, 2017; Benson et al., 2016; Tsourakakis et al., 2017) and rank learning (Li & Milenkovic, 2017), higher order relations between vertices captured by hypergraphs are typically associated with different cut weights. In (Li & Milenkovic, 2017), Li and Milenkovic generalized the notion of hyperedge cut weights by assuming that different hyperedge cuts have different weights, and that consequently, each hyperedge is associated with a vector of weights rather than a single scalar weight. If the weights of the hyperedge cuts are submodular, then one can use a graph with nonnegative edge weights to efficiently approximate the hypergraph, provided that the largest size of a hyperedge is a relatively small constant. This property of the projected hypergraphs allows one to leverage spectral hypergraph clustering algorithms based on clique expansions with provable performance guarantees. Unfortunately, the clique expansion method in general has two drawbacks: The spectral clustering algorithm for graphs used in the second step is merely quadratically optimal, while the projection step can cause a large distortion.

To address the quadratic optimality issue in graph clustering, Amghibech (Amghibech, 2003) introduced the notion of p-Laplacians of graphs and derived Cheeger-type inequalities for the second smallest eigenvalue of a p-Laplacian, $p > 1$, of a graph. These results motivated Bühler and Hein's work (Bühler & Hein, 2009) on spectral clustering based on p-Laplacians that provided tighter approximations of the Cheeger constant. Szlam and Bresson (Szlam & Bresson, 2010) showed that the 1-Laplacian allows one to exactly compute the Cheeger constant, but at the cost of computational hardness (Chang, 2016). Very little is known about the use of p-Laplacians for hypergraph clustering and their spectral properties.

To address the clique expansion problem, Hein et al. (Hein et al., 2013) introduced a clustering method for homogeneous hypergraphs that avoids expansions and works directly with the total variation of homogeneous hypergraphs,

[1]Department of Electrical and Computer Engineering, University of Illinois Urbana-Champaign, USA. Correspondence to: Pan Li <panli2@illinois.edu>, Olgica Milenkovic <milenkov@illinois.edu>.

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

without investigating the spectral properties of the operator. The only other line of work trying to mitigate the projection problem is due to Louis (Louis, 2015), who used a natural extension of 2-Laplacians for homogeneous hypergraphs, derived quadratically-optimal Cheeger-type inequalities and proposed a semidefinite programing (SDP) based algorithm whose complexity scales with the size of the largest hyperedge in the hypergraph.

Our contributions are threefold. First, we introduce submodular hypergraphs. Submodular hypergraphs allow one to perform hyperedge partitionings that depend on the subsets of elements involved in each part, thereby respecting higher-order and other constraints in graphs (see (Li & Milenkovic, 2017; Arora et al., 2012; Fix et al., 2013) for applications in food network analysis, learning to rank, subspace clustering and image segmentation). Second, we define p-Laplacians for submodular hypergraphs and generalize the corresponding discrete nodal domain theorems (Tudisco & Hein, 2016; Chang et al., 2017) and higher-order Cheeger inequalities. Even for homogeneous hypergraphs, nodal domain theorems were not known and only one low-order Cheeger inequality for 2-Laplacians was established by Louis (Louis, 2015). An analytical obstacle in the development of such a theory is the fact that p-Laplacians of hypergraphs are operators that act on vectors and produce *sets of values*. Consequently, operators and eigenvalues have to be defined in a set-theoretic manner. Third, based on the newly established spectral hypergraph theory, we propose two spectral clustering methods that learn the second smallest eigenvalues of 2- and 1-Laplacians. The algorithm for 2-Laplacian eigenvalue computation is based on an SDP framework and can provably achieve quadratic optimality with an $O(\sqrt{\zeta(E)})$ approximation constant, where $\zeta(E)$ denotes the size of the largest hyperedge in the hypergraph. The algorithm for 1-Laplacian eigenvalue computation is based on the inverse power method (IPM) (Hein & Bühler, 2010) that only has convergence guarantees. The key novelty of the IPM-based method is that the critical inner-loop optimization problem of the IPM is efficiently solved by algorithms recently developed for decomposable submodular minimization (Jegelka et al., 2013; Ene & Nguyen, 2015; Li & Milenkovic, 2018). Although without performance guarantees, given that the 1-Laplacian provides the tightest approximation guarantees, the IPM-based algorithm – as opposed to the clique expansion method (Li & Milenkovic, 2017) – performs very well empirically even when the size of the hyperedges is large. This fact is illustrated on several UC Irvine machine learning datasets available from (Asuncion & Newman, 2007).

The paper is organized as follows. Section 2 contains an overview of graph Laplacians and introduces the notion of submodular hypergraphs. The section also contains a description of hypergraph Laplacians, and relevant concepts in submodular function theory. Section 3 presents the fundamental results in the spectral theory of p-Laplacians, while Section 4 introduces two algorithms for evaluating the second largest eigenvalue of p-Laplacians needed for 2-way clustering. Section 5 presents experimental results. All proofs are relegated to the Supplementary Material.

2. Mathematical Preliminaries

A weighted graph $G = (V, E, w)$ is an ordered pair of two sets, the vertex set $V = [N] = \{1, 2, \dots, N\}$ and the edge set $E \subseteq V \times V$, equipped with a weight function $w : E \to \mathbb{R}^+$.

A cut $C = (S, \bar{S})$ is a bipartition of the set V, while the cut-set (boundary) of the cut C is defined as the set of edges that have one endpoint in S and one in the complement of S, $\bar{S}$, i.e., $\partial S = \{(u, v) \in E \mid u \in S, v \in \bar{S}\}$. The weight of the cut induced by S equals $\text{vol}(\partial S) = \sum_{u\in S,\, v\in \bar{S}} w_{uv}$, while the conductance of the cut is defined as

$$c(S) = \frac{\text{vol}(\partial S)}{\min\{\text{vol}(S), \text{vol}(\bar{S})\}},$$

where $\text{vol}(S) = \sum_{u\in S} \mu_u$, and $\mu_u = \sum_{v\in V} w_{uv}$. Whenever clear from the context, for $e = (uv)$, we write w_e instead of w_{uv}. Note that in this setting, the vertex weight values μ_u are determined based on the weights of edges w_e incident to u. Clearly, one can use a different choice for these weights and make them independent from the edge weights, which is a generalization we pursue in the context of submodular hypergraphs. The smallest conductance of any bipartition of a graph G is denoted by h_2 and referred to as the Cheeger constant of the graph.

A generalization of the Cheeger constant is the $k-$*way Cheeger constant* of a graph G. Let P_k denote the set of all partitions of V into k-disjoint nonempty subsets, i.e., $P_k = \{(S_1, S_2, ..., S_k) | S_i \subset V, S_i \neq \emptyset, S_i \cap S_j = \emptyset, \forall i, j \in [k], i \neq j\}$. The $k-$way Cheeger constant is defined as

$$h_k = \min_{(S_1,S_2,\dots,S_k)\in P_k} \max_{i\in[k]} c(S_i).$$

Spectral graph theory provides a means for bounding the Cheeger constant using the (normalized) Laplacian matrix of the graph, defined as $L = D - A$ and $\mathcal{L} = I - D^{-1/2}AD^{-1/2}$, respectively. Here, A stands for the adjacency matrix of the graph, D denotes the diagonal degree matrix, while I stands for the identity matrix. The graph Laplacian is an operator $\triangle_2^{(g)}$ (Chung, 1997) that satisfies

$$\langle x, \triangle_2^{(g)}(x)\rangle = \sum_{(uv)\in E} w_{uv}(x_u - x_v)^2.$$

A generalization of the above operator termed the p-Laplacian operator of a graph $\triangle_p^{(g)}$ was introduced by

Amghibech in (Amghibech, 2003), where

$$\langle x, \triangle_p^{(g)}(x)\rangle = \sum_{(uv)\in E} w_{uv}|x_u - x_v|^p.$$

The well known Cheeger inequality asserts the following relationship between h_2 and λ, the second smallest eigenvalue of the normalized Laplacian $\triangle_2^{(g)}$ of a graph:

$$h_2 \leq \sqrt{2\lambda} \leq 2\sqrt{h_2}.$$

It can be shown that the cut $\hat{h}_2$ dictated by the elements of the eigenvector associated with λ satisfies $\hat{h}_2 \leq \sqrt{2\lambda}$, which implies $\hat{h}_2 \leq 2\sqrt{h_2}$. Hence, spectral clustering provides a quadratically optimal graph partition.

2.1. Submodular Hypergraphs

A weighted hypergraph $G = (V, E, w)$ is an ordered pair of two sets, the vertex set $V = [N]$ and the hyperedge set $E \subseteq 2^V$, equipped with a weight function $w : E \to \mathbb{R}^+$. The relevant notions of cuts, boundaries and volumes for hypergraphs can be defined in a similar manner as for graphs. If each cut of a hyperedge e has the same weight w_e, we refer to the cut as a homogeneous cut and the corresponding hypergraph as a homogeneous hypergraph.

For a ground set Ω, a set function $f : 2^\Omega \to \mathbb{R}$ is termed submodular if for all $S, T \subseteq \Omega$, one has $f(S) + f(T) \geq f(S \cup T) + f(S \cap T)$.

A weighted hypergraph $G = (V, E, \boldsymbol{\mu}, \mathbf{w})$ is termed a *submodular hypergraph* with vertex set V, hyperedge set E and positive vertex weight vector $\boldsymbol{\mu} \triangleq \{\mu_v\}_{v\in V}$, if each hyperedge $e \in E$ is associated with a submodular weight function $w_e(\cdot) : 2^e \to [0, 1]$. In addition, we require the weight function $w_e(\cdot)$ to be:

1) Normalized, so that $w_e(\emptyset) = 0$, and all cut weights corresponding to a hyperedge e are normalized by $\vartheta_e = \max_{S\subseteq e} w_e(S)$. In this case, $w_e(\cdot) \in [0, 1]$;

2) Symmetric, so that $w_e(S) = w_e(e\backslash S)$ for any $S \subseteq e$;

The submodular hyperedge weight functions are summarized in the vector $\mathbf{w} \triangleq \{(w_e, \vartheta_e)\}_{e\in E}$. If $w_e(S) = 1$ for all $S \in 2^e\backslash\{\emptyset, e\}$, submodular hypergraphs reduce to homogeneous hypergraphs. We omit the designation homogeneous whenever there is no context ambiguity.

Clearly, a vertex v is in e if and only if $w_e(\{v\}) > 0$: If $w_e(\{v\}) = 0$, the submodularity property implies that v is not *incident* to e, as for any $S \subseteq e\backslash\{v\}$, $|w_e(S \cup \{v\}) - w_e(S)| \leq w_e(\{v\}) = 0$.

We define the degree of a vertex v as $d_v = \sum_{e\in E:\, v\in e} \vartheta_e$, i.e., as the sum of the max weights of edges incident to the vertex v. Furthermore, for any vector $y \in \mathbb{R}^N$, we define the projection weight of y onto any subset $S \subseteq V$ as $y(S) = \sum_{v\in S} y_v$. The volume of a subset of vertices $S \subseteq V$ equals $\mathrm{vol}(S) = \sum_{v\in S} \mu_v$.

For any $S \subseteq V$, we generalize the notions of the boundary of S and the volume of the boundary of S according to $\partial S = \{e \in E | e \cap S \neq \emptyset, e \cap \bar{S} \neq \emptyset\}$, and

$$\mathrm{vol}(\partial S) = \sum_{e\in\partial S} \vartheta_e w_e(S) = \sum_{e\in E} \vartheta_e w_e(S), \qquad (1)$$

respectively. Then, the normalized cut induced by S, the Cheeger constant and the k-way Cheeger constant for hypergraphs are defined in an analogous manner as for graphs.

2.2. Laplacian Operators for Hypergraphs

We introduce next p-Laplacians of hypergraphs and a number of relevant notions associated with Laplacian operators.

Hein et al.(Hein et al., 2013) connected p-Laplacians $\triangle_p^{(h)}$ for homogeneous hypergraphs with the total variation via

$$\langle x, \triangle_p^{(h)}(x)\rangle = \sum_{e\in E} w_e \max_{u,v\in e} |x_u - x_v|^p,$$

where w_e denotes the weight of a homogeneous hyperedge e. They also introduced the Inverse Power Method (IPM) to evaluate the spectrum of the hypergraph 1-Laplacian $\triangle_1^{(h)}$ (Hein et al., 2013), but did not establish any performance guarantees. In an independent line of work, Louis (Louis, 2015) introduced a quadratic variant of a hypergraph Laplacian

$$\langle x, \triangle_2^{(h)}(x)\rangle = \sum_{e\in E} w_e \max_{u,v\in e} (x_u - x_v)^2.$$

He also derived a Cheeger-type inequality relating the second smallest eigenvalue λ of $\triangle_2^{(h)}$ and the Cheeger constant of the hypergraph h_2 that reads as $\hat{h}_2 \leq O(\sqrt{\log \zeta(E)})\sqrt{\lambda} \leq O(\sqrt{\log \zeta(E)})\sqrt{h_2}$. Compared to the result of graph (3.12), for homogeneous hypergraphs, $\log \zeta(E)$ plays as some additional difficulty to approximate h_2. Learning the spectrum of generalizations of hypergraph Laplacians can be an even more challenging task.

2.3. Relevant Background on Submodular Functions

Given an arbitrary set function $F : 2^V \to \mathbb{R}$ satisfying $F(V) = 0$, the *Lovász extension* (Lovász, 1983) $f : \mathbb{R}^N \to \mathbb{R}$ of F is defined as follows: For any vector $x \in \mathbb{R}^N$, we order its entries in nonincreasing order $x_{i_1} \geq x_{i_2} \geq \cdots \geq x_{i_n}$ while breaking the ties arbitrarily, and set

$$f(x) = \sum_{j=1}^{N-1} F(S_j)(x_{i_j} - x_{i_{j+1}}), \qquad (2)$$

with $S_j = \{i_1, i_2, ..., i_j\}$. For submodular F, the Lovász extension is a convex function (Lovász, 1983).

Let $\mathbf{1}_S \in \mathbb{R}^N$ be the indicator vector of the set S. Hence, for any $S \subseteq V$, one has $F(S) = f(\mathbf{1}_S)$. For a submodular F, we define a convex set termed the *base polytope*

$$\mathcal{B} \triangleq \{y \in \mathbb{R}^N | y(S) \leq F(S), \text{ for all } S \subseteq V, \text{and such that } y(V) = F(V) = 0\}.$$

According to the defining property of submodular functions (Lovász, 1983), we may write $f(x) = \max_{y\in\mathcal{B}}\langle y, x\rangle$.

The subdifferential $\nabla f(x)$ of f is defined as

$$\{y \in \mathbb{R}^N \mid f(x') - f(x) \geq \langle y, x' - x\rangle, \forall x' \in \mathbb{R}^N\}.$$

An important result from (Bach et al., 2013) characterizes the subdifferentials $\nabla f(x)$: If $f(x)$ is the Lovász extension of a submodular function F with base polytope $\mathcal{B}$, then

$$\nabla f(x) = \arg\max_{y\in\mathcal{B}}\langle y, x\rangle. \tag{3}$$

Observe that $\nabla f(x)$ is a set and that the right hand side of the definition represents a set of maximizers of the objective function. If $f(x)$ is the Lovász extension of a submodular function, then $\langle q, x\rangle = f(x)$ for all $q \in \nabla f(x)$.

For each hyperedge $e \in E$ of a submodular hypergraph, following the above notations, we let $\mathcal{B}_e$, $\mathcal{E}(\mathcal{B}_e)$, f_e denote the base polytope, the set of extreme points of the base polytope, and the Lovász extension of the submodular hyperedge weight function w_e, respectively. Note that for any $S \subseteq V$, $w_e(S) = w_e(S \cap e)$. Consequently, for any $y \in \mathcal{B}_e$, $y_v = 0$ for $v \notin e$. Since $\nabla f_e \subseteq \mathcal{B}_e$, it also holds that $(\nabla f_e)_v = 0$ for $v \notin e$. When using formula (2) to explicitly describe the Lovász extension f_e, we can either use a vector x of dimension N or only those of its components that lie in e. Furthermore, in the later case, $|\mathcal{E}(\mathcal{B}_e)| = |e|!$.

3. p-Laplacians of Submodular Hypergraphs

We start our discussion by defining the notion of a p-Laplacian operator for submodular hypergraphs. We find the following definitions useful for our subsequent exposition.

Let $\text{sgn}(\cdot)$ be the sign function defined as $\text{sgn}(a) = 1$, for $a > 0$, $\text{sgn}(a) = -1$, for $a < 0$, and $\text{sgn}(a) = [-1, 1]$, for $a = 0$. For all $v \in V$, define the entries of a vector φ_p over $\mathbb{R}^N$ according to $(\varphi_p(x))_v = |x_v|^{p-1}\text{sgn}(x_v)$. Furthermore, let U be a $N \times N$ diagonal matrix such that $U_{vv} = \mu_v$ for all $v \in V$.

Let $\|x\|_{\ell_p,\mu} = (\sum_{v\in V}\mu_v|x_v|^p)^{1/p}$ and $\mathcal{S}_{p,\mu} \triangleq \{x \in \mathbb{R}^N | \|x\|_{\ell_p,\mu} = 1\}$. For a function Φ over $\mathbb{R}^N$, let $\Phi|_{\mathcal{S}_{p,\mu}}$ stand for Φ restricted to $\mathcal{S}_{p,\mu}$.

Definition 3.1. The p-Laplacian operator of a submodular hypergraph, denoted by $\triangle_p$ ($p \geq 1$), is defined for all $x \in \mathbb{R}^N$ according to

$$\langle x, \triangle_p(x)\rangle \triangleq Q_p(x) = \sum_{e\in E}\vartheta_e f_e(x)^p. \tag{4}$$

Hence, $\triangle_p(x)$ may also be specified directly as an operator over $\mathbb{R}^N$ that reads as

$$\triangle_p(x) = \begin{cases} \sum_{e\in E}\vartheta_e f_e(x)^{p-1}\nabla f_e(x) & p > 1, \\ \sum_{e\in E}\vartheta_e \nabla f_e(x) & p = 1. \end{cases}$$

Definition 3.2. A pair $(\lambda, x) \in \mathbb{R} \times \mathbb{R}^N/\{\mathbf{0}\}$ is called an eigenpair of the p-Laplacian $\triangle_p$ if $\triangle_p(x) \cap \lambda U\, \varphi_p(x) \neq \emptyset$.

As $f_e(\mathbf{1}) = 0$, we have $\triangle_p(\mathbf{1}) = 0$, so that $(0, \mathbf{1})$ is an eigenpair of the operator $\triangle_p$. A p-Laplacian operates on vectors and produces sets. In addition, since for any $t > 0$, $\triangle_p(tx) = t^{p-1}\triangle_p(x)$ and $\varphi_p(tx) = t^{p-1}\varphi_p(x)$, (tx, λ) is an eigenpair if and only if (x, λ) is an eigenpair. Hence, one only needs to consider normalized eigenpairs: In our setting, we choose eigenpairs that lie in $\mathcal{S}_{p,\mu}$ for a suitable choice for the dimension of the space.

For linear operators, the Rayleigh-Ritz method (Gould, 1966) allows for determining approximate solutions to eigenproblems and provides a variational characterization of eigenpairs based on the critical points of functionals. To generalize the method, we introduce two even functions,

$$\tilde{Q}_p(x) \triangleq Q_p(x)|_{\mathcal{S}_{p,\mu}}, \quad R_p(x) \triangleq \frac{Q_p(x)}{\|x\|^p_{\ell_p,\mu}}.$$

Definition 3.3. A point $x \in \mathcal{S}_{p,\mu}$ is termed a *critical point* of $R_p(x)$ if $0 \in \nabla R_p(x)$. Correspondingly, $R_p(x)$ is termed a *critical value* of $R_p(x)$. Similarly, x is termed a *critical point* of $\tilde{Q}_p$ if there exists a $\sigma \in \nabla Q_p(x)$ such that $P(x)\sigma = 0$, where $P(x)\sigma$ stands for the projection of σ onto the tangent space of $\mathcal{S}_{p,\mu}$ at the point x. Correspondingly, $\tilde{Q}_p(x)$ is termed a *critical value* of $\tilde{Q}_p$.

The relationships between the critical points of $\tilde{Q}_p(x)$ and $R_p(x)$ and the eigenpairs of $\triangle_p$ relevant to our subsequent derivations are listed in Theorem 3.4.

Theorem 3.4. A pair (λ, x) ($x \in \mathcal{S}_{p,\mu}$) is an eigenpair of the operator $\triangle_p$
1) if and only if x is a critical point of $\tilde{Q}_p$ with critical value λ, and provided that $p \geq 1$.
2) if and only if x is a critical point of R_p with critical value λ, and provided that $p > 1$.
3) if x is a critical point of R_p with critical value λ, and provided that $p = 1$.

The critical points of $\tilde{Q}_p$ bijectively characterize eigenpairs for all choices of $p \geq 1$. However, R_p has the same property only if $p > 1$. This is a consequence of the nonsmoothness of the set $\mathcal{S}_{1,\mu}$, which has been observed for graphs as well (See the examples in Section 2.2 of (Chang, 2016)).

3.1. Discrete Nodal Domain Theorem for $p-$Laplacians

Nodal domain theorems are essential for understanding the structure of eigenvectors of operators and they have been the subject of intense study in geometry and graph theory alike (Bıyıkoglu et al., 2007). The eigenfunctions of a Laplacian operator may take positive and negative values. The signs of the values induce a partition of the vertices in V into maximal connected components on which the sign of the eigenfunction does not change: These components represent the nodal domains of the eigenfunction and approximate the clusters of the graphs.

Davies et al. (BrianDavies et al., 2001) derived the first discrete nodal domain theorem for the $\triangle_2^{(g)}$ operator. Chang et al. (Chang et al., 2017) and Tudisco et al. (Tudisco & Hein, 2016) generalized these theorem for $\triangle_1^{(g)}$ and $\triangle_p^{(g)}$ $(p > 1)$ of graphs. In what follows, we prove that the discrete nodal domain theorem applies to $\triangle_p$ of submodular hypergraphs.

As every nodal domain theorem depends on some underlying notion of connectivity, we first define the relevant notion of connectivity for submodular hypergraphs. In a graph or a homogeneous hypergraph, vertices on the same edge or hyperedge are considered to be connected. However, this property does not generalize to submodular hypergraphs, as one can merge two nonoverlapping hyperedges into one without changing the connectivity of the hyperedges. To see why this is the case, consider two hyperedges e_1 and e_2 that are nonintersecting. One may transform the submodular hypergraph so that it includes a hyperedge $e = e_1 \cup e_2$ with weight $w_e = w_{e_1} + w_{e_2}$. This transformation essentially does not change the submodular hypergraph, but in the newly obtained hypergraph, according to the standard definition of connectivity, the vertices in e_1 and e_2 are connected. This problem may be avoided by defining connectivity based on the volume of the boundary set.

Definition 3.5. Two distinct vertices $u, v \in V$ are said to be *connected* if for any S such that $u \in S$ and $v \notin S$, $\text{vol}(\partial S) > 0$. A submodular hypergraph is *connected* if for any non-empty $S \subset V$, one has $\text{vol}(\partial S) > 0$.

According to the following lemma, it is always possible to transform the weight functions of submodular hypergraph in such a way as to preserve connectivity.

Lemma 3.6. Any submodular hypergraph $G = (V, E, \mathbf{w}, \boldsymbol{\mu})$ can be reduced to another submodular hypergraph $G' = (V, E', \mathbf{w}', \boldsymbol{\mu})$ without changing $\text{vol}(\partial S)$ for any $S \subseteq V$ and ensuring that for any $e \in E'$, and $u, v \in e$, u and v are connected.

Definition 3.7. Let $x \in \mathbb{R}^N$. A *positive (respectively, negative) strong nodal domain* is the set of vertices of a maximally connected induced subgraph of G such that $\{v \in V | x_v > 0\}$ (respectively, $\{v \in V | x_v < 0\}$). A *positive (respectively, negative) weak nodal domain* is defined in the same manner, except for changing the strict inequalities as $\{v \in V | x_v \geq 0\}$ (*respectively*, $\{v \in V | x_v \leq 0\}$).

The following lemma establishes that for a connected submodular hypergraph G, all nonconstant eigenvectors of the operator $\triangle_p$ correspond to nonzero eigenvalues.

Lemma 3.8. If G is connected, then all eigenvectors associated with the zero eigenvalue have constant entries.

We next state new nodal domain theorems for submodular hypergraph $p-$Laplacians. The results imply the bounds for the numbers of nodal domains induced from eigenvectors of p-Laplacian do not essentially change compared to those for graphs (Tudisco & Hein, 2016). We do not consider the case $p = 1$, although it is possible to adapt the methods for analyzing the $\triangle_1^{(g)}$ operators of graphs to $\triangle_1$ operators of submodular hypergraphs. Such a generalization requires extensions of the critical-point theory to piecewise linear manifolds (Chang, 2016).

Theorem 3.9. Let $p > 1$ and assume that G is a connected submodular hypergraph. Furthermore, let the eigenvalues of $\triangle_p$ be ordered as $0 = \lambda_1^{(p)} < \lambda_2^{(p)} \leq \cdots \leq \lambda_{k-1}^{(p)} < \lambda_k^{(p)} = \cdots = \lambda_{k+r-1}^{(p)} < \lambda_{k+r}^{(p)} \leq \cdots \leq \lambda_n^{(p)}$, with $\lambda_k^{(p)}$ having multiplicity r. Let x be an arbitrary eigenvector associated with $\lambda_k^{(p)}$. Then x induces at most $k + r - 1$ strong and at most k weak nodal domains.

Lemma 3.10. Let G be a connected submodular hypergraph. For $p > 1$, any nonconstant eigenvector has at least two weak (strong) nodal domains. Hence, the eigenvectors associated with the second smallest eigenvalue $\lambda_2^{(p)}$ have exactly two weak (strong) nodal domains. For $p = 1$, the eigenvectors associated with the second smallest eigenvalue $\lambda_2^{(1)}$ may have only one single weak (strong) nodal domain.

We define next the following three functions: $\mu_p^+(x) \triangleq \sum_{v \in V: x_v > 0} \mu_v |x_v|^{p-1}$, $\mu^0(x) \triangleq \sum_{v \in V: x_v = 0} \mu_v$, and $\mu_p^-(x) \triangleq \sum_{v \in V: x_v < 0} \mu_v |x_v|^{p-1}$.

Lemma 3.11. Let G be a connected submodular hypergraph. Then, for any nonconstant eigenvector x of $\triangle_p$, one has $\mu_p^+(x) - \mu_p^-(x) = 0$ for $p > 1$, and $|\mu_1^+(x) - \mu_1^-(x)| \leq \mu^0(x)$ for $p = 1$. Consequently, $0 \in \arg\min_{c \in \mathbb{R}} \|x - c\mathbf{1}\|_{\ell_p, \mu}^p$ for any $p \geq 1$.

The nodal domain theorem characterizes the structure of the eigenvectors of the operator, and the number of nodal domains determines the approximation guarantees in Cheeger-type inequalities relating the spectra of graphs and hypergraphs and the Cheeger constant. These observations are rigorously formalized in the next section.

3.2. Higher-Order Cheeger Inequalities

In what follows, we analytically characterize the relationship between the Cheeger constants and the eigenvalues

$\lambda_k^{(p)}$ of $\triangle_p$ for submodular hypergraphs.

Theorem 3.12. Suppose that $p \geq 1$ and let $(\lambda_k^{(p)}, x_k)$ be the $k-$th eigenpair of the operator $\triangle_p$, with m_k denoting the number of strong nodal domains of x_k. Then,

$$\left(\frac{1}{\tau}\right)^{p-1} \left(\frac{h_{m_k}}{p}\right)^p \leq \lambda_k^{(p)} \leq (\min\{\zeta(E), k\})^{p-1}\, h_k,$$

where $\tau = \max_v\, d_v/\mu_v$. For homogeneous hypergraphs, a tighter bound holds that reads as

$$\left(\frac{2}{\tau}\right)^{p-1} \left(\frac{h_{m_k}}{p}\right)^p \leq \lambda_k^{(p)} \leq 2^{p-1}\, h_k.$$

It is straightforward to see that setting $p = 1$ produces the tightest bounds on the eigenvalues, while the case $p = 2$ reduces to the classical Cheeger inequality. This motivates an in depth study of algorithms for evaluating the spectrum of $p = 1, 2$-Laplacians, described next.

4. Spectral Clustering Algorithms for Submodular Hypergraphs

The Cheeger constant is frequently used as an objective function for (balanced) graph and hypergraph partitioning (Zhou et al., 2007; Bühler & Hein, 2009; Szlam & Bresson, 2010; Hein & Bühler, 2010; Hein et al., 2013; Li & Milenkovic, 2017). Theorem 3.12 implies that $\lambda_k^{(p)}$ is a good approximation for the k-way Cheeger constant of submodular graphs. Hence, to perform accurate hypergraph clustering, one has to be able to efficiently learn $\lambda_k^{(p)}$ (Ng et al., 2002; Von Luxburg, 2007). We outline next how to do so for $k = 2$.

In Theorem 4.1, we describe an objective function that allows us to characterize $\lambda_2^{(p)}$ in a computationally tractable manner; the choice of the objective function is related to the objective developed for graphs in (Bühler & Hein, 2009; Szlam & Bresson, 2010). Minimizing the proposed objective function produces a real-valued output vector $x \in \mathbb{R}^N$. Theorem 4.3 describes how to round the vector x and obtain a partition which provably upper bounds $c(S)$. Based on the theorems, we propose two algorithms for evaluating $\lambda_2^{(2)}$ and $\lambda_2^{(1)}$. Since $\lambda_2^{(1)} = h_2$, the corresponding partition corresponds to the tightest approximation of the 2-way Cheeger constant. The eigenvalue $\lambda_2^{(2)}$ can be evaluated in polynomial time with provable performance guarantees. The problem of devising good approximations for values $\lambda_k^{(p)}$, $k \neq 2$, is still open.

Let $Z_{p,\mu}(x, c) \triangleq \|x - c\mathbf{1}\|_{\ell_p,\mu}^p$ and $Z_{p,\mu}(x) \triangleq \min_{c\in\mathbb{R}} Z_{p,\mu}(x, c)$, and define

$$\mathcal{R}_p(x) \triangleq \frac{Q_p(x)}{Z_{p,\mu}(x)}. \tag{5}$$

Theorem 4.1. For $p > 1$, $\lambda_2^{(p)} = \inf_{x\in\mathbb{R}^N} \mathcal{R}_p(x)$. Moreover, $\lambda_2^{(1)} = \inf_{x\in\mathbb{R}^N} \mathcal{R}_1(x) = h_2$.

Definition 4.2. Given a nonconstant vector $x \in \mathbb{R}^N$, and a threshold θ, set $\Theta(x, \theta) = \{v : x_v > \theta\}$. The optimal conductance obtained from thresholding vector x equals

$$c(x) = \inf_{\theta\in[x_{\min}, x_{\max})} \frac{\mathrm{vol}(\partial\Theta(x, \theta))}{\min\{\mathrm{vol}(\Theta(x, \theta)), \mathrm{vol}(V/\Theta(x, \theta))\}}.$$

Theorem 4.3. For any $x \in \mathbb{R}^N$ that satisfies $0 \in \arg\min_c Z_{p,\mu}(x, c)$, i.e., such that $Z_{p,\mu}(x, 0) = Z_{p,\mu}(x)$, one has $c(x) \leq p\,\tau^{(p-1)/p}\, \mathcal{R}_p(x)^{1/p}$, where $\tau = \max_{v\in V}\, d_v/\mu_v$.

In what follows, we present two algorithms. The first algorithm describes how to minimize $\mathcal{R}_2(x)$, and hence provides a polynomial-time solution for submodular hypergraph partitioning with provable approximation guarantees, given that the size of the largest hyperedge is a constant. The result is concluded in Theorem 4.5. The algorithm is based on an SDP, and may be computationally too intensive for practical applications involving large hypergrpahs of even moderately large hyperedges. The second algorithm is based on IPM (Hein & Bühler, 2010) and aims to minimize $\mathcal{R}_1(x)$. Although this algorithm does not come with performance guarantees, it provably converges (see Theorem 4.6) and has good heuristic performance. Moreover, the inner loop of the IPM involves solving a version of the proximal-type decomposable submodular minimization problem (see Theorem 4.7), which can be efficiently performed using a number of different algorithms (Kolmogorov, 2012; Jegelka et al., 2013; Nishihara et al., 2014; Ene & Nguyen, 2015; Li & Milenkovic, 2018).

4.1. An SDP Method for Minimizing $\mathcal{R}_2(x)$

The $\mathcal{R}_2(x)$ minimization problem introduced in Equation (5) may be rewritten as

$$\min_{x:Ux\perp\mathbf{1}} \frac{Q_2(x)}{\|x\|_{\ell_2,\mu}^2}, \tag{6}$$

where we observe that $Q_2(x) = \sum_{e\in E} \vartheta_e f_e^2(x) = \sum_{e\in E} \vartheta_e \max_{y\in\mathcal{E}(\mathcal{B}_e)} \langle y, x\rangle^2$. This problem is, in turn, equivalent to the nonconvex optimization problem

$$\min_{x\in\mathbb{R}^N} \sum_e \vartheta_e \left(\max_{y\in\mathcal{E}(\mathcal{B}_e)} \langle y, x\rangle\right)^2 \tag{7}$$
$$\text{s.t.} \sum_{v\in V} \mu_v x_v^2 = 1,\ \sum_{v\in V} \mu_v x_v = 0.$$

Following an approach proposed for homogeneous hypergraphs (Louis, 2015), one may try to solve an SDP relaxation of (7) instead. To describe the relaxation, let each vertex v of the graph be associated with a vector $x'_v \in \mathbb{R}^n$,

$n \geq \zeta(E)$. The assigned vectors are collected into a matrix of the form $X = (x'_1, .., x'_N)$. The SDP relaxation reads as

$$\min_{X \in \mathbb{R}^{n\times N},\, \eta\in\mathbb{R}^{|E|}} \sum_e \vartheta_e \eta_e^2 \tag{8}$$
$$\text{s.t.} \quad \|Xy\|_2^2 \leq \eta_e^2 \quad \forall y \in \mathcal{E}(\mathcal{B}_e), e \in E$$
$$\sum_{v\in V} \mu_v \|x'_v\|_2^2 = 1, \sum_{v\in V} \mu_v x'_v = 0.$$

Note that $\mathcal{E}(\mathcal{B}_e)$ is of size $O(|e|!)$, and the above problem can be solved efficiently if $\zeta(E)$ is small.

Algorithm 1 lists the steps of an SDP-based algorithm for minimizing $\mathcal{R}_2(x)$, and it comes with approximation guarantees stated in Lemma 4.4. In contrast to homogeneous hypergraphs (Louis, 2015), for which the approximation factor equals $O(\log \zeta(E))$, the guarantees for general submodular hypergraphs are $O(\zeta(E))$. This is due to the fact that the underlying base polytope $\mathcal{B}_e$ for a submodular function is significantly more complex than the corresponding polytope for the homogeneous case. We conjecture that this approximation guarantee is optimal for SDP methods.

Algorithm 1: Minimization of $\mathcal{R}_2(x)$ using SDP

Input: A submodular hypergraph $G = (V, E, \mathbf{w}, \boldsymbol{\mu})$
1: Solve the SDP (8).
2: Generate a random Gaussian vector $g \sim N(0, I_n)$, where I_n denotes the identity matrix of order n.
3: Output $x = X^T g$.

Lemma 4.4. Let x be as in Algorithm 1, and let the optimal value of (8) be SDPopt. Then, with high probability, $\mathcal{R}_2(x) \leq O(\zeta(E))\,\text{SDPopt} \leq O(\zeta(E))\, \min \mathcal{R}_2$.

This result immediately leads to the following theorem.

Theorem 4.5. Suppose that x is the output of Algorithm 1. Then, $c(x) \leq O(\sqrt{\zeta(E)\tau\, h_2})$ with high probability.

We describe next Algorithm 2 for optimizing $\mathcal{R}_1(x)$ which has guaranteed convergence properties.

Theorem 4.6. The sequence $\{x^k\}$ generated by Algorithm 2 satisfies $\mathcal{R}_1(x^{k+1}) \leq \mathcal{R}_1(x^k)$.

The computationally demanding part of Algorithm 2 is the optimization procedure in Step 3. The optimization problem is closely related to the problem of submodular function minimization (SFM) due to the defining properties of the Lovász extension. Theorem 4.7 describes different equivalent formulations of the optimization problem in Step 3.

Theorem 4.7. If the norm of the vector z in Step 3 is $\|z\|_2$, the underlying optimization problem is the dual of the following ℓ_2 minimization problem

$$\min_{y_e} \|\sum_{e\in E} y_e - \hat{\lambda}^k g^k\|_2^2, \quad y_e \in \vartheta_e \mathcal{B}_e, \forall\, e \in E, \tag{9}$$

Algorithm 2: IPM-based minimization of $\mathcal{R}_1(x)$

Input: A submodular hypergraph $G = (V, E, \mathbf{w}, \boldsymbol{\mu})$
Find nonconstant $x^0 \in \mathbb{R}^N$ s.t. $0 \in \arg\min_c \|x^0 - c\mathbf{1}\|_{\ell_1,\mu}$
initialize $\hat{\lambda}^0 \leftarrow \mathcal{R}_1(x^0)$, $k \leftarrow 0$
1: **Repeat**:
2: For $v \in V$, $g_v^k \leftarrow \begin{cases} \text{sgn}(x_v^k)\mu_v, & \text{if } x_v^k \neq 0 \\ -\frac{\mu_1^+(x^k) - \mu_1^-(x^k)}{\mu^0(x^k)}\mu_v, & \text{if } x_v^k = 0 \end{cases}$
3: $z^{k+1} \leftarrow \arg\min_{z:\|z\|\leq 1} Q_1(z) - \hat{\lambda}^k \langle z, g^k\rangle$
4: $c^{k+1} \leftarrow \arg\min_c \|z^{k+1} - c\mathbf{1}\|_{\ell_1,\mu}$
5: $x^{k+1} \leftarrow z^{k+1} - c^{k+1}\mathbf{1}$
6: $\hat{\lambda}^{k+1} \leftarrow \mathcal{R}_1(x^{k+1})$
7: Until $|\hat{\lambda}^{k+1} - \hat{\lambda}^k|/\hat{\lambda}^k < \epsilon$
8. **Output** x^{k+1}

where the primal and dual variables are related according to $z = \frac{\hat{\lambda}^k g^k - \sum_{e\in E} y_e}{\|\hat{\lambda}^k g^k - \sum_{e\in E} y_e\|_2}$.

If the norm of the vector z in Step 3 is $\|z\|_\infty$, the underlying optimization problem is equivalent to the following SFM problem

$$\min_{S\subseteq V} \sum_e \vartheta_e w_e(S) - \hat{\lambda}^k g^k(S), \tag{10}$$

where the the primal and dual variables are related according to $z_v = 1$ if $v \in S$, and $z_v = -1$ if $v \notin S$.

For special forms of submodular weights, different algorithms for the optimization problems in Theorem 4.7 may be used instead. For graphs and homogeneous hypergraphs with hyperedges of small size, the min-cut algorithms by Karger et al. and Chekuri et al. (Karger, 1993; Chekuri & Xu, 2017) allow one to efficiently solve the discrete problem (10). Continuous optimization methods such as alternating projections (AP) (Nishihara et al., 2014) and coordinate descend methods (CDM) (Ene & Nguyen, 2015) can be used to solve (9) by "tracking" minimum norm points of base polytopes corresponding to individual hyperedges, where for general submodular weights, the Wolfe's Algorithm (Wolfe, 1976) can be used. When the submodular weights have some special properties, such as that they depend only on the cardinality of the input, there exist algorithms that operate efficiently even when $|e|$ is extremely large (Jegelka et al., 2013).

In our experimental evaluations, we use a random coordinate descent method (RCDM) (Ene & Nguyen, 2015), which ensures an expected $(1+\epsilon)-$approximation by solving an expected number of $O(|V|^2|E|\log\frac{1}{\epsilon})$ min-norm-point problems. Note that when performing continuous optimization, one does not need to solve the inner-loop optimization problem exactly and is allowed to exit the loop as long as the objective function value decreases. The underlying algorithm – Algorithm 3 – is described in the Supplement.

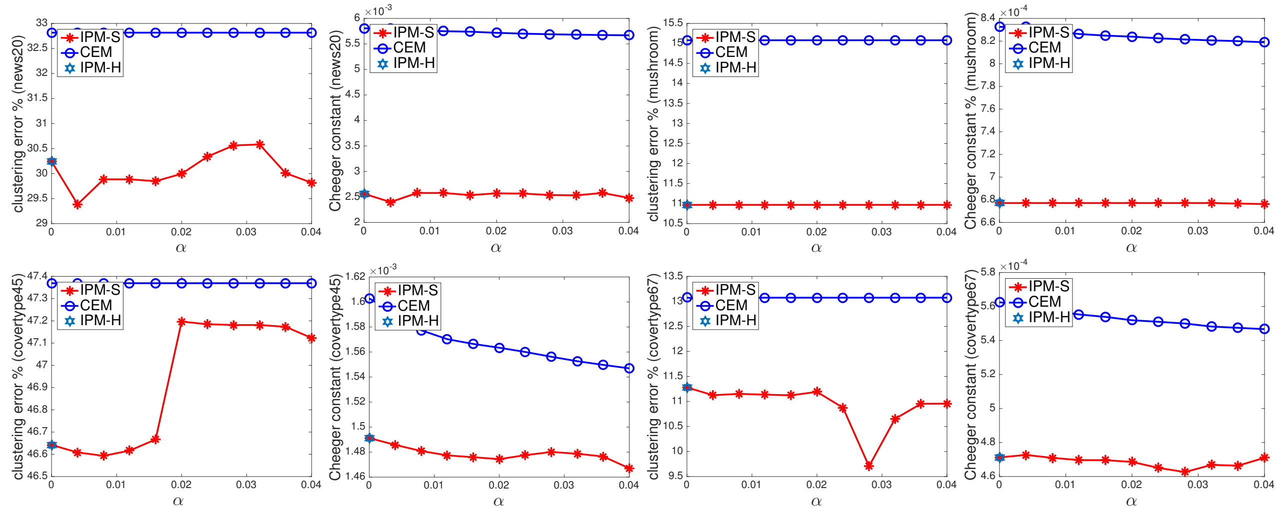

Figure 1. Experimental clustering results for four UCI datasets, displayed in pairs of figures depicting the Clustering error and the Cheeger constant versus α. Fine tuning the parameter α may produce significant performance improvements in several datasets - for example, on the Covertype67 dataset, choosing $\alpha = 0.028$ results in visible drops of the clustering error and the Cheeger constant. *Both the use of* 1-*Laplacians and submodular weights may be credited for improving clustering performance.*

5. Experiments

In what follows, we compare the algorithms for submodular hypergraph clustering described in the previous section to two methods: The IPM for homogeneous hypergraph clustering (Hein et al., 2013) and the clique expansion method (CEM) for submodular hypergraph clustering (Li & Milenkovic, 2017). We focus on 2-way graph partitioning problems related to the University of California Irvine (UCI) datasets selected for analysis in (Hein et al., 2013), described in Table 1 of the Supplementary Material. The datasets include 20Newsgroups, Mushrooms, Covertype. In all datasets, $\zeta(E)$ was roughly 10^3, and each of these datasets describes multiple clusters. Since we are interested in 2-way partitioning, we focused on two pairs of clusters in Covertype, denoted by $(4, 5)$ and $(6, 7)$, and paired the four 20Newsgroups clusters, one of which includes *Comp.* and *Sci*, and another one which includes *Rec.* and *Talk*. The Mushrooms and 20Newsgroups datasets contain only categorical features, while Covertype also includes numerical features. We adopt the same approach as the one described in (Hein et al., 2013) to construct hyperedges: Each feature corresponds to one hyperedge; hence, each categorical feature is captured by one hyperedge, while numerical features are first quantized into 10 bins of equal size, and then mapped to hyperedges. To describe the submodular weights, we fix $\vartheta_e = 1$ for all hyperedges and parametrize w_e using a variable $\alpha \in (0, 0.5]$

$$w_e(S;\alpha) = \frac{1}{2} + \frac{1}{2}\min\left\{1, \frac{|S|}{\lceil \alpha|e|\rceil}, \frac{|e/S|}{\lceil \alpha|e|\rceil}\right\}, \ \forall S \subseteq e.$$

The intuitive explanation behind our choice of weights is that it allows one to accommodate categorization errors and outliers: In contrast to the homogeneous case in which any partition of a hyperedge has weight one, the chosen submodular weights allow a smaller weight to be used when the hyperedge is partitioned into small parts, i.e., when $\min\{|S|, |e/S|\} < \lceil \alpha|e|\rceil$. In practice, α is chosen to be relatively small – in all experiments, we set $\alpha \leq 0.04$, with α close to zero producing homogeneous hyperedge weights.

The results are shown in Figure 1. As may be observed, both in terms of the Clustering error (i.e., the total number of erroneously classified vertices) and the values of the Cheeger constant, IPM-based methods outperform CEM. This is due to the fact that for large hyperedge sizes, CEM incurs a high distortion when approximating the submodular weights ($O(\zeta(E))$ (Li & Milenkovic, 2017)). Moreover, as $w_e(S)$ depends merely on $|S|$, the submodular hypergraph CEM reduces to the homogeneous hypergraph CEM (Zhou et al., 2007), which is an issue that the IPM-based method does not face. Comparing the performance of IPM on submodular hypergraphs (IPM-S) with that on homogeneous hypergraphs (IPM-H), we see that IPM-S achieves better clustering performance on both 20Newsgroups and Covertypes, and offers the same performance as IPM-H on the Mushrooms dataset. This indicates that it is practically useful to use submodular hyperedge weights for clustering purposes. A somewhat unexpected finding is that for certain cases, one observes that when α increases (and thus, when w_e decreases), the corresponding Cheeger constant increases. This may be caused by the fact that the IPM algorithm can get trapped in a local optima.

References

Amghibech, S. Eigenvalues of the discrete p-Laplacian for graphs. *Ars Combinatoria*, 67:283–302, 2003.

Arora, Chetan, Banerjee, Subhashis, Kalra, Prem, and Maheshwari, SN. Generic cuts: An efficient algorithm for optimal inference in higher order MRF-MAP. In *Proceedings of the European Conference on Computer Vision*, pp. 17–30. Springer, 2012.

Asuncion, Arthur and Newman, David. UCI machine learning repository, 2007.

Bach, Francis et al. Learning with submodular functions: A convex optimization perspective. *Foundations and Trends® in Machine Learning*, 6(2-3):145–373, 2013.

Benson, Austin R, Gleich, David F, and Leskovec, Jure. Higher-order organization of complex networks. *Science*, 353(6295):163–166, 2016.

Bıyıkoglu, Türker, Leydold, Josef, and Stadler, Peter F. Laplacian eigenvectors of graphs. *Lecture notes in mathematics*, 1915, 2007.

BrianDavies, E, Gladwell, GrahamM L, Leydold, Josef, and Stadler, Peter F. Discrete nodal domain theorems. *Linear Algebra and its Applications*, 336(1-3):51–60, 2001.

Bühler, Thomas and Hein, Matthias. Spectral clustering based on the graph p-Laplacian. In *Proceedings of the International Conference on Machine Learning*, pp. 81–88. ACM, 2009.

Chang, KC, Shao, Sihong, and Zhang, Dong. Nodal domains of eigenvectors for 1-Laplacian on graphs. *Advances in Mathematics*, 308:529–574, 2017.

Chang, Kung Ching. Spectrum of the 1-Laplacian and Cheeger's constant on graphs. *Journal of Graph Theory*, 81(2):167–207, 2016.

Chekuri, Chandra and Xu, Chao. Computing minimum cuts in hypergraphs. In *Proceedings of the ACM-SIAM Symposium on Discrete Algorithms*, pp. 1085–1100. Society for Industrial and Applied Mathematics, 2017.

Chung, Fan RK. *Spectral graph theory*. Number 92. American Mathematical Soc., 1997.

Ene, Alina and Nguyen, Huy. Random coordinate descent methods for minimizing decomposable submodular functions. In *Proceedings of the International Conference on Machine Learning*, pp. 787–795, 2015.

Fix, Alexander, Joachims, Thorsten, Park, Sung Min, and Zabih, Ramin. Structured learning of sum-of-submodular higher order energy functions. In *Proceedings of the IEEE International Conference on Computer Vision*, pp. 3104–3111. IEEE, 2013.

Gould, Sydney Henry. *Variational methods for eigenvalue problems*, volume 22. University of Toronto Press Toronto, 1966.

Hein, Matthias and Bühler, Thomas. An inverse power method for nonlinear eigenproblems with applications in 1-spectral clustering and sparse pca. In *Advances in Neural Information Processing Systems*, pp. 847–855, 2010.

Hein, Matthias, Setzer, Simon, Jost, Leonardo, and Rangapuram, Syama Sundar. The total variation on hypergraphs-learning on hypergraphs revisited. In *Advances in Neural Information Processing Systems*, pp. 2427–2435, 2013.

Jegelka, Stefanie, Bach, Francis, and Sra, Suvrit. Reflection methods for user-friendly submodular optimization. In *Advances in Neural Information Processing Systems*, pp. 1313–1321, 2013.

Karger, David R. Global min-cuts in RNC, and other ramifications of a simple min-cut algorithm. In *Proceedings of the ACM-SIAM Symposium on Discrete Algorithms*, volume 93, pp. 21–30, 1993.

Kolmogorov, Vladimir. Minimizing a sum of submodular functions. *Discrete Applied Mathematics*, 160(15):2246–2258, 2012.

Li, Pan and Milenkovic, Olgica. Inhomogeneous hypergraph clustering with applications. In *Advances in Neural Information Processing Systems*, pp. 2305–2315, 2017.

Li, Pan and Milenkovic, Olgica. Revisiting decomposable submodular function minimization with incidence relations. *arXiv preprint arXiv:1803.03851*, 2018.

Louis, Anand. Hypergraph markov operators, eigenvalues and approximation algorithms. In *Proceedings of the ACM symposium on Theory of computing*, pp. 713–722. ACM, 2015.

Lovász, László. Submodular functions and convexity. In *Mathematical Programming The State of the Art*, pp. 235–257. Springer, 1983.

Ng, Andrew Y, Jordan, Michael I, and Weiss, Yair. On spectral clustering: Analysis and an algorithm. In *Advances in Neural Information Processing Systems*, pp. 849–856, 2002.

Nishihara, Robert, Jegelka, Stefanie, and Jordan, Michael I. On the convergence rate of decomposable submodular function minimization. In *Advances in Neural Information Processing Systems*, pp. 640–648, 2014.

Shanu, Ishant, Arora, Chetan, and Singla, Parag. Min norm point algorithm for higher order MRF-MAP inference. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 5365–5374, 2016.

Szlam, Arthur and Bresson, Xavier. Total variation and cheeger cuts. In *Proceedings of the International Conference on Machine Learning*, pp. 1039–1046, 2010.

Tsourakakis, Charalampos E, Pachocki, Jakub, and Mitzenmacher, Michael. Scalable motif-aware graph clustering. In *Proceedings of the 26th International Conference on World Wide Web*, pp. 1451–1460. International World Wide Web Conferences Steering Committee, 2017.

Tudisco, Francesco and Hein, Matthias. A nodal domain theorem and a higher-order cheeger inequality for the graph p-Laplacian. *arXiv preprint arXiv:1602.05567*, 2016.

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

Wolfe, Philip. Finding the nearest point in a polytope. *Mathematical Programming*, 11(1):128–149, 1976.

Zhang, Chenzi, Hu, Shuguang, Tang, Zhihao Gavin, and Chan, T-H. Hubert. Re-revisiting learning on hypergraphs: Confidence interval and subgradient method. In *Proceedings of the International Conference on Machine Learning*, volume 70, pp. 4026–4034, 2017.

Zhou, Denny, Huang, Jiayuan, and Schölkopf, Bernhard. Learning with hypergraphs: Clustering, classification, and embedding. In *Advances in Neural Information Processing Systems*, pp. 1601–1608, 2007.