



Generalizing p -Laplacian: spectral hypergraph theory and a partitioning algorithm

Shota Saito¹ · Mark Herbster¹

Received: 16 February 2022 / Revised: 20 July 2022 / Accepted: 29 September 2022
© The Author(s) 2022

Abstract

For hypergraph clustering, various methods have been proposed to define hypergraph p -Laplacians in the literature. This work proposes a general framework for an abstract class of hypergraph p -Laplacians from a differential-geometric view. This class includes previously proposed hypergraph p -Laplacians and also includes previously unstudied novel generalizations. For this abstract class, we extend current spectral theory by providing an extension of nodal domain theory for the eigenvectors of our hypergraph p -Laplacian. We use this nodal domain theory to provide bounds on the eigenvalues via a higher-order Cheeger inequality. Following our extension of spectral theory, we propose a novel hypergraph partitioning algorithm for our generalized p -Laplacian. Our empirical study shows that our algorithm outperforms spectral methods based on existing p -Laplacians.

Keywords Spectral clustering · Hypergraph learning · Hypergraph p -Laplacian · Cheeger inequality

1 Introduction

Graphs are one of the most widely used data representations for structured data, such as social networks (Newman, 2006), web pages (Brin and Page, 1998) and images (Shi and Malik, 1997). The graph Laplacian is a linear operator that characterizes the graph. A natural discrete optimization problem whose solution characterizes a balanced clustering is solved in its *relaxed* form by finding the second eigenvector of the graph Laplacian (Fiedler, 1973; Alon and Milman, 1985; Hagen and Kahng, 1992; Shi and Malik, 1997; von Luxburg, 2007). The Laplacian has been previously generalized to a nonlinear p -Laplacian in the context of machine learning. Improved performance as a function of p was previously demonstrated in various special cases such as (Bühler and Hein,

Editors: Krzysztof Dembczynski and Emilie Devijver

✉ Shota Saito
ssaito@cs.ucl.ac.uk

¹ Department of Computer Science, University College London, Gower Street, London WC1E 6BT, UK

Table 1 Comparison table for existing methods and ours. STAR is studied in Zhou et al. (2006), and unnormalized CLIQUE is studied in Rodriguez (2002) and edge-normalized CLIQUE is in Saito et al. (2018)

	STAR	CLIQUE	TV	Submodular	Ours
$\langle \psi, \Delta_p \psi \rangle = S_p(\psi)$	✓	✓	✓	✓	✓
p -Laplacian		✓	✓	✓	✓
Nodal domain theorem				✓	✓
Cheeger inequality				✓	✓
Clustering algorithm for fixed p	✓	✓	✓	✓	✓
Clustering algorithm for any $p > 1$		✓			✓

The TV is first proposed in Hein et al. (2013) and generalized to a submodular hypergraph (Li and Milenkovic, 2018). Ours incorporates all of these hypergraph Laplacians. All of these Laplacians have a similar structure, stated as $\langle \psi, \Delta_p \psi \rangle = S_p(\psi)$ (Proposition 4, for STAR $p = 2$ since p -Laplacian in a STAR form has not been studied), which gives a foundation of hypergraph cut partitioning algorithm. However, some Laplacians have missed some key characteristics in the table. One of the advantages of our abstract class of hypergraph Laplacian gives all the features of the existing methods. See the main text for details

2009; Bougleux et al., 2009). The p -Laplacian has previously been motivated from the perspective of differential geometry (Zhou and Schölkopf, 2005), as well as from a Cheeger inequality perspective (Bühler and Hein, 2009).

Hypergraphs generalize graphs and serve as a natural representation of multi-relational data. The hypergraph representation has been used to model videos (Huang et al., 2009), web browsing histories (Mobasher et al., 2000), recommender engines (Gharahighehi et al., 2021) and cell molecules (Klamt et al., 2009). However, although a hypergraph is a natural data representation, generalization from graph Laplacian to hypergraph Laplacian is not straightforward. Thus, multiple such generalizations have been proposed in the literature (Agarwal et al., 2006; Saito et al., 2018; Hein et al., 2013; Li and Milenkovic, 2018). Furthermore, the extension from Laplacian to p -Laplacian may take different forms (Saito et al., 2018; Li and Milenkovic, 2018). However, as shown in Table 1, although they have a similar structure, some Laplacians miss some key features. The objective of this work is to construct a theoretical structure to bring these similar but disparate models into one unified framework.

In our unified framework, we define an abstract class of hypergraph p -Laplacians that incorporates a number of previously proposed hypergraph p -Laplacians as well as previously unstudied novel hypergraph p -Laplacians. This framework builds on a limited special case previously proposed in Saito et al. (2018). The overall framework is inspired by a differential-geometric analogy from the continuous to the discrete domain. Exploiting the differential-geometric connection, we provide a generalized nodal domain theorem (see Theorem 9) and a generalized Cheeger inequality (see Theorem 10 and Corollary 11). These provide a theoretical justification and bounds for using the eigenvectors of a hypergraph p -Laplacian to perform partitioning. Exploiting these theoretical results, we provide an algorithm for finding an approximation to the second eigenvector. We provide an empirical study of this algorithm which shows that our algorithm outperforms a variety of existing hypergraph p -Laplacian based methods.

We highlight five salient contributions of this work.

1. From a differential-geometric perspective, we define an abstract class of p -Laplacians of hypergraphs that can incorporate previously proposed p -Laplacians as well as novel unstudied p -Laplacians.
2. We provide theoretical results for our abstract class of p -Laplacians, such as the Nodal domain theorem, the Cheeger inequality, and a bound on the relationship between the minimum Cheeger cut and the second p -eigenvalue of the p -Laplacian.
3. Exploiting these theoretical results, we propose a convergent hypergraph partitioning algorithm with respect to our abstract class of hypergraph p -Laplacians.
4. We demonstrate empirically that our method can improve the performance of the existing p -Laplacians.
5. Based on our theoretical and empirical observations, we provide guidance on the choice of p -Laplacian.

We remark that in the literature, there are a number of special case results (Zhou et al., 2006; Agarwal et al., 2006; Saito et al., 2018; Hein et al., 2013). These prior results derive a patchwork of nodal domain theorems, Cheeger inequalities, as well as partitioning algorithms for some particular cases of hypergraph p -Laplacians, as shown in Table 1. The advantage of the approach here is that we define an abstract class of hypergraph p -Laplacians, and both our theory and our partitioning algorithm apply to the complete class. Finally, we provide guidance on how to select a particular value of p for hypergraph p -Laplacians. *All proofs are in appendix sections.*

2 Preliminaries

This section reviews the notions of hypergraph and Cheeger inequality.

2.1 Hypergraph notions

We begin with standard definitions and notations of a hypergraph. A *hypergraph* $G = (V, E, \mathbf{w}, \boldsymbol{\mu})$, where V is a *vertex* set, an E is an *edge* set, \mathbf{w} is a vector $\{w(e)\}_{e \in E}$ where $w : E \rightarrow \mathbb{R}^+$ maps each edge with a *weight*, and $\boldsymbol{\mu}$ is a vector $\{\mu(v)\}_{v \in V}$, where $\mu : V \rightarrow \mathbb{R}^+$ maps each edge with a *edge weight*. The edge set E is a subset of all possible permutation of vertices, i.e., $E \subseteq \bigcup_{k=1}^{|V|} \bigcup_{\{v_1, \dots, v_k\} \subseteq V} \{[v_{\sigma(1)}, \dots, v_{\sigma(k)}] \mid \sigma \in \mathcal{P}_k\}$, where \mathcal{P}_k denotes the set of permutations σ on $\{1, \dots, k\}$. A hypergraph is *connected* if there is a path for every pair of vertices. A hypergraph is *undirected* when the set of edges are symmetric; defining a relation \sim between two edges as $[v_{\sigma(1)}, \dots, v_{\sigma(k)}] \sim [v_{\sigma'(1)}, \dots, v_{\sigma'(k)}]$ where $\sigma, \sigma' \in \mathcal{P}_k (k = 1, \dots, n)$ and we denote a set of undirected edges as $E_u = E / \sim$. In what follows, we assume that the hypergraph G is connected and undirected unless noted. We define the *degree of a vertex* $v \in V$ as $d(v) = \sum_{e \in E: v \in e} w(e)$, while *the degree of an edge* $e \in E$ is defined as $|e|$. For the benefit of the representation of hypergraph, we define various matrices. *Degree matrices* D_v and D_e are diagonal matrices whose diagonal elements are degree of vertices and edges, respectively. Let W_e be a diagonal $|E| \times |E|$ matrix, whose diagonal elements are weight of

edge e . We denote by $|V| \times |E_u|$ matrix H is an *incidence matrix*, whose element $h(v, e) = 1$ if vertex v is connected to the edge e , and 0 otherwise. For more details, see Berge (1984).

2.2 Cheeger inequality

This section reviews Cheeger inequality for the standard 2-Laplacian case. Influenced by the inequality of the eigenvalue of continuous Laplacian called *Cheeger inequality*, there is existing research on the Cheeger inequality in the discrete domain (Alon and Milman, 1985; Lee et al., 2014; Tudisco and Hein, 2016). This inequality shows the connection between the eigenproblem of Laplacian and a graph cut called as *Cheeger cut* Alon and Milman (1985). This inequality motivates us to use eigenvectors to partition in the following way. While the discrete optimization problem of finding a subset of V that minimizes the Cheeger cut is NP-hard (von Luxburg, 2007), the eigenproblem of the graph Laplacian is not NP-hard. Since the Cheeger inequality gives bounds between eigenvalues of graph Laplacian and optimal Cheeger cut, the Cheeger inequality can be said to guarantee the performance of the eigenproblem compared to the ground truth from the original cut problem. This performance guarantee enables us to use eigenvectors obtained by less computationally expensive eigenproblem instead of the costly ground truth from the discrete cut problem. In other words, the Cheeger inequality “connects” Cheeger cut and eigenproblem; the Cheeger inequality shows how much we approximate the original graph cut problem by relaxing this into the real-valued eigenproblem of Laplacian.

We observe the “connection” as follows. Let $A \subset V$ be a set and \bar{A} be a complement of A . The Cheeger cut may be defined as

$$C(A) := \frac{\partial V(A, \bar{A})}{\min(\text{vol}(A), \text{vol}(\bar{A}))}, \text{ where } \partial V(A, \bar{A}) := \sum_{u \in A, v \in \bar{A}} w(\{u, v\}),$$

where $\text{vol}(A) = \sum_{v \in A} d(v)$. The *Cheeger constant* $h_2 := \min_{A \subset V} C(A)$ is the value of the optimal cut. The Cheeger inequality shows a connection between the eigenvalue of Laplacian and the Cheeger constant as

$$h_2 \leq \sqrt{2\lambda_2} \leq 2\sqrt{h_2}. \quad (1)$$

Equation (1) shows how we approximate the Cheeger constant by relaxing the original discrete cut problem into the real-valued eigenproblem of the graph Laplacian. The Cheeger inequality guarantees the performance of the cut resulting from algorithms using the second eigenvector of Laplacian as follows. Let (B, \bar{B}) be the cut found by the second eigenvector of the Laplacian ψ , such that $(\{v : \psi(v) \geq t\}, \{v : \psi(v) < t\})$ minimizing Cheeger cut. Chung (2007) showed that $C(B) \sqrt{2\lambda_2}$. Therefore, by the upper bound of Eq. (1) we observe

$$C(B) < 2\sqrt{h_2}, \quad (2)$$

which gives a guarantee for the worst case of performance of spectral clustering. This motivates us to use spectral methods for graph partitioning problems.

3 Hypergraph p -Laplacian

This section defines and discusses a hypergraph p -Laplacian and associated p -eigenpairs.

3.1 Differential operators: gradient ∇ , divergence div and p -Laplacian Δ_p

In this section, we aim to extend various differential operators proposed in Saito et al. (2018) to an abstract class of p -Laplacians. We firstly introduce the following two inner product spaces $\mathcal{H}(V)$ and $\mathcal{H}(E)$ of real-valued functions over the vertex set and the (directed) edge set respectively,

$$\langle f, g \rangle_{\mathcal{H}(V)} := \sum_{v \in V} f(v)g(v) \quad (3)$$

$$\langle r, s \rangle_{\mathcal{H}(E)} := \sum_{e \in E} \frac{r(e)s(e)}{|e|!}, \quad (4)$$

We next define three operators on these spaces; *gradient* $\nabla: \mathcal{H}(V) \rightarrow \mathcal{H}(E)$, *divergence* $\text{div}: \mathcal{H}(E) \rightarrow \mathcal{H}(V)$, and *p -Laplacian* $\Delta_p: \mathcal{H}(V) \rightarrow \mathcal{H}(E)$. These operators are discrete geometric analogs to the comparable operators in the continuous differential geometry. In the continuous domain, for the second differentiable function f , the p -Laplace operator is defined as $\Delta_p^{(c)} f := \text{div}^{(c)}(\|f\|^{p-2} \nabla^{(c)} f)$, where operators with superscripted by (c) are the standard continuous calculus ones. In the following, we would like to establish a differential-geometric framework in a generalized discrete setting analogous to the continuous one to define an abstract class of p -Laplacians. The operators div , Δ_p were introduced in the graph setting (Zhou and Schölkopf, 2005; Grady, 2006) and generalized to the hypergraph setting in (Saito et al., 2018), whereas a similar formulation of ∇ was given graph and hypergraph settings (Zhou and Schölkopf, 2005; Saito et al., 2018). The definition that we propose below broadly generalizes all previous definitions. We define and discuss its interpretation below.

We propose to define the hypergraph-gradient as follows. The definition below is the generalization of the definition of gradient over hypergraphs proposed in Saito et al. (2018).

Definition 1 Let ∇ be an operator $\nabla: \mathcal{H}(V) \rightarrow \mathcal{H}(E)$. A hypergraph-gradient ∇ is

$$(\nabla \psi)(e) := \sum_{u, v \in e} w_p^{\frac{1}{p}}(e) c^{\frac{1}{p}}(u, v, e, \psi) \left(\frac{\psi(u)}{\mu^{1/p}(u)} - \frac{\psi(v)}{\mu^{1/p}(v)} \right),$$

where the operator ∇ and the function $c(u, v, e, \psi)$ satisfies the following three conditions for all $e \in E$ and vertices $u, v \in e$;

$$|\nabla(\alpha \psi)| = |\alpha \nabla(\alpha \psi)| \quad (5)$$

$$\sum_{u, v \in e} c(u, v, e, \psi) = c(e) \quad (6)$$

Table 2 The relationship between a function $c(u, v, e, \psi)$ in hypergraph-gradients and Laplacians. We denote $e_{[1]}$ by the first vertex of an edge e . Also, $F : 2^{|e|} \rightarrow [0, 1]$ is a submodular function, and we use rearranged vertices v_i so that $\psi(v_{|e|}) \geq \dots \geq \psi(v_1)$. See Sect. 3.4 for the details and all the notations

Type	$c(u, v, e, \psi)$
CLIQUE E-N	$1/(e - 1)$ when $e_{[1]} = v$ otherwise 0
CLIQUE E-UN	1 when $e_{[1]} = v$ otherwise 0
STAR	$1/ e $ when $e_{[1]} = v$ otherwise 0
TV	1 when $(u, v) = \operatorname{argmax}_{u,v}$ $ \psi(u)/\mu^{1/p}(u) - \psi(v)/\mu^{1/p}(v) $ otherwise 0
SUB	$F(S_i)(\max_{S \subseteq e} (F(S)))^{1/p}$ when $u = u_i$ and $v = u_{i+1}$ otherwise 0

$$\left. \frac{\partial c^{1/p}(u, v, e, \psi)}{\partial \psi} \right|_v = 0, \forall u \in e \quad \text{and} \quad \left. \frac{\partial c^{1/p}(u, v, e, \psi)}{\partial \psi} \right|_u = 0, \forall v \in e \quad (7)$$

This hypergraph-gradient can be intuitively interpreted as follows. The term $\psi(u)/\mu^{1/p}(u) - \psi(v)/\mu^{1/p}(v)$ can be interpreted as “roughness” (normalized by μ) between two vertices. The hypergraph-gradient is a sum of all possible combinations of this term for the edge e . Hence, the hypergraph-gradient can be intuitively seen as the roughness in one edge, similar to the continuous gradient $\nabla^{(c)}$.

The definition of the hypergraph-gradient function depends on a “weighting” function $c(u, v, e, \psi)$. This weighting function can be seen as a coefficient of the difference between every pair of vertices. Varying $c(u, v, e, \psi)$ allows us to model different types of hypergraph expansions including but not limited to the star (Zhou et al., 2006) or clique expansions (Saito et al., 2018) (see Table 2 for details), i.e., the function c enables the following our generalized p -Laplacian framework to be abstract.

We leave a few remarks on equations of gradient Definition 1. First, the gradient operator ∇ and the function c has three conditions described as Eqs. (5), (6) and (7). Equation (5) requires the operator ∇ to be either homogeneous or absolute homogeneous. Equation (6) requires that the summation of the function over all the pairs of vertices at an edge e is independent of ψ . Equation (7) enforces that the function $c^{1/p}$ is independent of ψ once we fix one edge and one vertex in the edge. In the following, when c is not differentiable, we consider subdifferential instead of derivative. For a more detailed discussion on these conditions, see Appendix 1. We normalize ψ by vertex weights μ . We call the vertex weights *unnormalized* when $\mu_v = 1$ and *normalized* when $\mu(v) = d(v)$, $\forall v \in V$. We observe that the existing unnormalized p -Laplacian such as Hein et al. (2013) $\mu(v) = 1$ and normalized 2-Laplacian (Zhou et al., 2006; Saito et al., 2018) when $\mu(v) = d(v)$.

The following definition of a divergence operator is inspired by an analogy to the continuous setting.

Definition 2 A hypergraph divergence is an operator $\operatorname{div} : \mathcal{H}(E) \rightarrow \mathcal{H}(V)$ which satisfies $\forall \psi \in \mathcal{H}(V), \forall \phi \in \mathcal{H}(E) \langle \phi, \nabla \psi \rangle_{\mathcal{H}(E)} = \langle \psi, -\operatorname{div} \phi \rangle_{\mathcal{H}(V)}$.

Note that Definition 2 is an analog to the continuous Stoke’s Theorem. Also, we can check that div is unique. Intuitively, divergence counts the net flow defined by ϕ on the vertex, similarly to the intuition in the continuous domain.

Finally, we propose to define p -Laplacian.

Definition 3 An operator $\Delta_p: \mathcal{H}(V) \rightarrow \mathcal{H}(E)$ is a hypergraph p -Laplacian if $\Delta_p \psi := -\operatorname{div}(\|\nabla \psi\|_p^{p-2} \nabla \psi)$.

3.2 p -Dirichlet sum and p -Laplacian

This section defines the p -Dirichlet sum, which can be interpreted as energy over the hypergraph. Also, we discuss relations between the p -Dirichlet sum and the p -Laplacian. Lastly, we discuss how these relations are the foundation of graph partitioning.

Using the norm defined by the Hilbert space in Eq. (4), we define p -Dirichlet sum of $\psi \in H(V)$ as

$$S_p(\psi) := \|\nabla \psi\|_p^p = \sum_{e \in E} \frac{|\nabla \psi(e)|^p}{|e|^p}, \quad (8)$$

which measures roughness of ψ over the hypergraph. Hence, it is natural to interpret the p -Dirichlet sum as an energy over hypergraph. Later we will use this energy serves as the objective function of the hypergraph partitioning.

For the p -Dirichlet sum and p -Laplacian, the following relationships hold;

Proposition 4 $S_p(\psi) = \langle \psi, \Delta_p \psi \rangle_{\mathcal{H}(V)}$.

Proposition 5 $\left. \frac{\partial S_p(\psi)}{\partial \psi} \right|_v = p \Delta_p \psi(v)$.

These relations are important both in the continuous and discrete domains. In the continuous domain, the analog of these relations is fundamental for an important problem on p -Laplacian, called *Dirichlet Principle* (Courant and Hilbert, 1962); the Dirichlet energy is minimized when the Laplace equation is satisfied. For the clustering in the discrete domain, we minimize the p -Dirichlet sum. To do so, we consider a problem similar to the Laplace equation, which is the eigenproblem of Laplacian. This is how we see an analogy between continuous differential geometry and this discrete geometry. In the following, we illustrate this with an example of the standard graph 2-Laplacian case. Let L be a standard normalized graph 2-Laplacian $L = D^{-1/2}(D - A)D^{-1/2}$ where A is an adjacency matrix, and D is a diagonal degree matrix. This Laplacian is computed through our framework if we set hypergraph-gradient as

$$[4](\nabla \psi)(\{u, v\}) = w^{1/2}(\{u, v\}) \left(\frac{\psi(u)}{\sqrt{d(u)}} - \frac{\psi(v)}{\sqrt{d(v)}} \right). \quad (9)$$

Note that we put $\mu = \{d(v)\}_{v \in V}$. We then obtain

$$\Delta_2 \psi(v) = \psi(v) - \sum_{\{u, v\} \in E} \frac{w(\{u, v\})}{d^{1/2}(u)d^{1/2}}, \quad (10)$$

which corresponds to $L\psi$. From Eq. (8), the energy is defined as

$$S_2(\psi) = \|\nabla \psi\|_2^2 = \sum_{u, v} w(\{u, v\}) \left(\frac{\psi(u)}{\sqrt{d(u)}} - \frac{\psi(v)}{\sqrt{d(v)}} \right)^2 = \psi^\top L \psi. \quad (11)$$

This corresponds to Proposition 4. Differentiating $S_2(\psi)$ by ψ , we observe

$$\frac{\partial S_2(\psi)}{\partial \psi} = 2L\psi, \quad (12)$$

which corresponds to Proposition 5. More detailed steps are explained in Appendix 2. In the graph partitioning context, Eqs. (11) and (12) serve as a foundation of the relationship between balanced cuts and graph Laplacian. Using these properties and Courtant's min-max theorem, we can connect the discrete cut problem into the eigenproblem of Laplacian by the Nodal domain theorem and the Cheeger inequality, as we will see in the next sections.

We often see the properties of Propositions 4 and 5 in the graph p -Laplacian (Bühler and Hein, 2009; Bougleux et al., 2009). Moreover, also in the hypergraph context, without a defining differential geometric setup, we see these properties in the existing hypergraph Laplacians, as seen in Table 1 (Zhou et al., 2006; Saito et al., 2018; Hein et al., 2013; Li and Milenkovic, 2018). Hence, it is natural to expect that all of the hypergraph Laplacians have a similar structure in this sense. However, as we see in Table 1, some Laplacians miss some features; particularly, some Laplacians miss the useful nodal domain theorem and Cheeger inequality (discussed in Sect. 4). Note that we "borrow" these results from the continuous differential geometry. One of the benefits of our abstract Laplacian is to give comprehensive analyses to all Laplacians defined from the gradient Definition 1.

3.3 p -Eigenproblem of p -Laplacian

Next, we discuss the eigenproblem of this p -Laplacian. Since a p -Laplace operator is non-linear, we introduce the standard generalization of eigenpair for p -Laplacian (see for examples of Tudisco and Hein (2016)).

Definition 6 [4] Let $\xi_p(x) := |x|^{p-1} \operatorname{sgn}(x)$. For $p \geq 1$, a hypergraph p -eigenpair, which is a pair of p -eigenvalue $\lambda \in \mathbb{R}$ and p -eigenvector $\psi \in \mathcal{H}(V)$ of Δ_p , is defined by

$$(\Delta_p \psi)(v) = \lambda \xi_p(\psi(v)), \forall v \in V. \quad (13)$$

In the standard Laplacian, one can show the connection between its eigenpair and Rayleigh quotient from the matrix theory as well as the continuous analysis. To obtain p -eigenpair, we consider the following Rayleigh quotient:

Proposition 7 Consider the Rayleigh quotient for our abstract class of p -Laplacians as

$$R_p(\psi) := \frac{S_p(\psi)}{\|\psi\|_p^p}, \text{ where } \|\psi\|_p := \left(\sum_v \psi^p(v) \right)^{\frac{1}{p}},$$

The function R_p has a critical point at ψ^* if and only if ψ^* is p -eigenvector of Δ_p . The corresponding p -eigenvalue λ^* is given as $\lambda_p^* = R_p(\psi^*)$. Moreover, the first p -eigenvalue is 0, whose p -eigenvector is $M^{1/p} \mathbf{1}$, where M is a $|V| \times |V|$ diagonal matrix whose diagonal element is μ_v .

For the standard Laplacians, the first p -eigenvector is equal to $c\mathbf{1}$ for unnormalized and $cD_v^{1/2}\mathbf{1}$ for normalized case. For more properties of p -eigenpair, see Appendix 5.

3.4 p -Laplacians and related regularizers

This section shows that various hypergraph Laplacians and related regularizers can be seen as a special case of our framework. We discuss a clique expansion way (Rodriguez, 2002; Saito et al., 2018), a star expansion way (Zhou et al., 2006), a total variation way (Hein et al., 2013), and a submodular hypergraph way (Li and Milenkovic, 2018).

3.4.1 Clique expansion hypergraph Laplacian (CLIQUE)

This approach constructs a graph where a clique replaces every pair of vertices in an original edge of a hypergraph. Hypergraph clique 2-Laplacian normalized by a degree of edge (CLIQUE E-N) is proposed by Saito et al. (2018), and clique 2-Laplacian but edge-unnormalized (CLIQUE E-UN) is proposed Rodriguez (2002). For those 2-Laplacians, clique Laplacian is $L = I - D^{-1/2} W D^{-1/2}$, where for the edge-normalized setting W is a matrix whose element is $w(u, v) = \sum_{uv \in e} w(e) / (|e| - 1)$ and $D = D_v$ and for an edge-unnormalized setting $w(u, v) = \sum_{u, v \in e} w(e)$ and D is a diagonal elements are $d(v, v) = \sum_{v \in e} w(e)$. Note that this can be seen as hypergraph contraction to graph, represented by W , and L is a standard 2-Laplacian induced by W .

3.4.2 Star expansion hypergraph Laplacian (STAR)

This way constructs a graph by making a new vertex for every edge to form a star. This Hypergraph 2-Laplacian can be written as $L_{2,s} = I - D_v^{-1/2} H W_e D_e^{-1} H^T D_v^{-1/2}$ (Zhou et al., 2006). We remark that this view is also hypergraph contraction to graph, represented by adjacency matrix $H W_e D_e^{-1} H^T$. Note that this Laplacian can be seen as the standard Laplacian if we consider hypergraph as a graph, except for the coefficient 1/2. This coefficient difference comes from the nature of this view, as discussed in Saito et al. (2018).

3.4.3 Total variation and submodular p -Laplacian (TV/SUB)

The total variation (TV) approach for hypergraph has been considered in a different context than the other two (Hein et al., 2013). The TV regularizer is defined as

$$S_p(\psi) = \sum_{e \in E} w(e) \max_{v, u \in e} |\psi(v) - \psi(u)|^p, \quad (14)$$

which is not normalized by a degree of vertex (TV V-UN). We here propose *normalized total variation* (TV V-N) p -Laplacian, whose regularizer we define as $S_p(\psi) = \sum_e w(e) \max_{v, u \in e} |\psi(v)/d^{1/p}(v) - \psi(u)/d^{1/p}(u)|^p$. This TV p -Laplacian is incorporated by the submodular p -Laplacian (Li and Milenkovic, 2018). The extensive study by Li and Milenkovic (2018) considers hypergraph p -Laplacian in the context of a submodular function, which we refer to as SUB. For a submodular function $F : 2^{|e|} \rightarrow [0, 1]$, associated with edge e , the submodular p -Laplacian is associated with the following energy;

$$S_p(\psi) = \sum_{e \in E} (w(e) \max_{S \subset V} (F(S)) \left(\sum_{i=1}^{|e|-1} F(S_i) \left(\frac{\psi(v_{i+1})}{\mu^{1/p}(v_{i+1})} - \frac{\psi(v_i)}{\mu^{1/p}(v_i)} \right) \right)^p, \quad (15)$$

by reordering v in e as $\psi(v_{|e|}) \geq \psi(v_{|e|-1}) \geq \dots \geq \psi(v_1)$, where $S_i := \{v_j\}_{j=1}^i$. Note that this form is one form of Lovász extension. By taking $F(S_i) = 1$ for all i , we obtain the p -Laplacian.

3.4.4 Connections between our p -Laplacian and existing Laplacians

These Laplacians can be seen as a special case of our abstract Laplacian, defined by Definition 3, followed by hypergraph-gradient (Definition 1) and hypergraph-divergence (Definition 2). Table 2 summarizes the corresponding function $c(u, v, e, \psi)$ in the definition of hypergraph-gradient. Edge normalized and unnormalized clique 2-Laplacians in Table 2 are 2-Laplacians proposed by Saito et al. (2018) and Rodriguez (2002), respectively. Star 2-Laplacian in Table 2 is equal to the Laplacian proposed by Zhou et al. (2006). The regularizer of unnormalized TV p -Laplacian in Table 2 corresponds one by Hein et al. (2013). We also note that all the functions of $c(u, v, e, \psi)$ satisfy the condition of Definition 1 (see Appendix 1). For more discussion, see Appendix 6.

4 Properties of p -Eigenpair of p -Laplacian

This section discusses the properties of the p -eigenproblem of our hypergraph p -Laplacian. Hence, we aim to establish the theoretical background of spectral clustering using p -Laplacian, such as the nodal domain theorem and the Cheeger inequality. The nodal domain theorem is about the bounds of the number of nodal domains, which can be seen as a “division”. Using this nodal domain, the Cheeger inequality shows how much p -eigenproblem can approximate a minimal graph cut.

4.1 Nodal domain theorem of the p -Laplacian

This section aims to extend the classical nodal domain theorem to our framework. The nodal domain theorem in the discrete domain is developed analogously from the Courant’s nodal domain theorem in the continuous domain (Courant and Hilbert, 1962). In the continuous case, a *nodal domain* is defined as a region for a function where a sign does not change. Therefore, a nodal domain marks the natural division of regions of real values. The nodal domain theorem shows a connection between eigenvectors of Laplacian and nodal domains; the theorem describes the bounds of the number of nodal domains of eigenvectors of Laplacian (Lindqvist, 2008). The same idea can be established in the discrete domain, i.e., a nodal domain is a connected sub-hypergraph where a sign of p -eigenvector does not change. This nodal domain can be seen as a “partition” by the p -eigenvector in the discrete domain. A next question is “can we obtain a similar bound to the number of this nodal domain?”

We begin with the definition of a nodal domain for a hypergraph.

Definition 8 A nodal domain is a maximally connected subgraph A of hypergraph G such that for $\psi \in H(V)$ A is either $\{v \mid \psi(v) > 0\}$ or $\{v \mid \psi(v) < 0\}$.

Next, with this definition, we discuss the nodal domain theorem for our hypergraph p -Laplacian. The nodal domain theorem for graph Laplacian has been proven in Fiedler (1975), generalized to graph p -Laplacian by Tudisco and Hein (2016), and extended to a particular type of hypergraph p -Laplacian by Li and Milenkovic (2018). In this line of research, we extend these nodal domain theorems to our abstract class of hypergraph p -Laplacians as follows;

Theorem 9 *Let $0 = \lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_{k-1} \leq \lambda_k = \dots = \lambda_{k+r-1} \leq \lambda_{k+r} \leq \dots$, be eigenvalues of Δ_p , and ψ_k is an associated eigenvector with λ_k . Then ψ_k induces at most $k + r - 1$ nodal domains.*

As seen in Theorem 9, the nodal domain theorem studies the structure of p -eigenvectors of p -Laplacian; Theorem 9 shows the bound on the number of nodal domains of p -eigenvectors. The number of nodal domains matters to Cheeger inequality, which is a theoretical justification for spectral methods via our p -Laplacian. We will discuss this Cheeger inequality next.

4.2 k -way cheeger inequality

This section establishes the k -way Cheeger inequality for our hypergraph p -Laplacian. As we saw in Sect. 2.2, the 2-way Cheeger inequality serves as the connection between Cheeger cut and eigenproblem. Moreover, the inequality gives a performance guarantee of the relaxed graph partitioning problem. We want to establish such a connection between the Cheeger cut and p -eigenproblem of our hypergraph p -Laplacian. For this purpose, we aim to generalize this Cheeger inequality to our hypergraph p -Laplacians to achieve spectral partitioning via our p -Laplacian.

We start our discussion from a 2-way Cheeger cut. Let $A \subset V$ be a set and \bar{A} be a complement of A . The generalized Cheeger cut may be defined as

$$\begin{aligned}
 C(A) &:= \frac{\partial V(A, \bar{A})}{\min(\text{vol}(A), \text{vol}(\bar{A}))}, \quad \text{where } \text{vol}(A) = \sum_{u \in A} \mu(v) \\
 \partial V(A, \bar{A}) &:= \sum_{e: u, v \in e, u \in A, v \in \bar{A}} w(e)c(e), \quad c(e) := \sum_{u, v \in e} c(u, v, e, \psi).
 \end{aligned}
 \tag{16}$$

We call the optimal cut $h_2 := \min_{A \subset V} C(A)$ as *Cheeger constant*. Considering the standard graph, this generalized Cheeger cut becomes the standard Cheeger cut discussed in Sect. 2.2. We shall extend this generalized 2-way Cheeger cut to k -way Cheeger cut. Consider disjoint partitioning of V into k sets $\{V_j\}_{j=1, \dots, k}$. Then, we define the k -way Cheeger constant as

$$h_k := \min_{\{V_j\}_{j=1, \dots, k}} \max_{J \in \{1, \dots, k\}} C(V_j).
 \tag{17}$$

Similarly to the previous studies (Tudisco and Hein, 2016; Li and Milenkovic, 2018), we establish k -way Cheeger inequality for our p -Laplacian as follows.

Theorem 10 *Let (λ_k, ψ_k) be a p -eigenpair of Δ_p , m_k be the number of nodal domains of ψ_k . Then,*

$$\left(\max_v \frac{d(v)}{\mu(v)} \right)^{-(p-1)} \left(\frac{h_{m_k}}{p} \right)^p \leq \lambda_k \leq \min(k, \max_{e \in E} |e|)^{p-1} h_k.$$

Corollary 11 Let (B, \bar{B}) be the cut found by the second eigenvector of the p -Laplacian ψ , such that $(\{v : \psi(v) \geq t\}, \{v : \psi(v) < t\})$ minimizing Cheeger cut. Then,

$$\left(\frac{1}{\max_v d(v)/\mu(v)} \right)^{p-1} \left(\frac{C(B)}{p} \right)^p < 2h_2 \quad (18)$$

Theorem 10 is an extension of the graph Cheeger inequality in terms of three perspectives; graph to hypergraph, 2-way to k -way, and the standard 2-Laplacian to our abstract class of p -Laplacians. Following Theorem 10, Corollary 11 is the bound of the relationship between the cut obtained by the second p -eigenvector of our abstract class of p -Laplacians and the generalized Cheeger constant. Similarly to the classical case in Sect. 2.2, Theorem 10 shows how we approximate the k -way Cheeger constant by relaxing discrete k -way cut problem into p -eigenproblem of Δ_p ; Theorem 10 gives the upper and lower bounds of the optimal cut using k -th p -eigenvalue. Moreover, Corollary 11 gives a guarantee for the worst case of a 2-way cut obtained by p -eigenvector. These bounds can be said to guarantee the performance of the cut resulting from spectral methods via p -eigenvectors of our p -Laplacian. Hence, Theorem 10 and Corollary 11 motivate us to use spectral methods via our p -Laplacian for the hypergraph partitioning problem instead of the costly discrete original cut problem. This inequality gives the tightest bound when $p \rightarrow 1$. Since the original cut problem is NP-hard, the eigenproblem is also an NP-hard problem in this asymptotic case. Moreover, by considering the case of the standard graph 2-Laplacian, this inequality can be reduced to the classical Cheeger inequality. Also, when $k = 2$, this inequality is for h_2 , which is a 2-way Cheeger cut. Therefore, in the next section, we focus on constructing a spectral algorithm for a 2-way partitioning.

Finally, we remark that the discussion on k -way Cheeger cut is a generalization of the standard graph 2-way Cheeger inequality of 2-Laplacian (Alon and Milman, 1985; Alon, 1986), k -way Cheeger inequality of 2-Laplacian (Lee et al., 2014), k -way Cheeger inequality of graph p -Laplacian (Tudisco and Hein, 2016), and k -way Cheeger inequality of p -Laplacian of submodular hypergraph (Li and Milenkovic, 2018) cases. We also note that the proofs for the nodal domain theorem (Theorem 9) and the Cheeger inequality (Theorem 10) are a natural generalization of the previous studies such as Tudisco and Hein (2016) and Li and Milenkovic (2018). Rather than introducing new techniques in the proofs, the focus of this work is that we generalize the hypergraph p -Laplacian as much as possible where these structures preserve in order to provide a unified framework.

5 Hypergraph partitioning via p -Laplacian

Sect. 4 shows the guarantee of performance of eigenproblem instead of the NP-hard discrete Cheeger cut problem. Therefore, this section establishes our partitioning algorithm, exploiting p -eigenpairs of our hypergraph p -Laplacian.

We firstly discuss a property of p -eigenvectors of Δ_p . For the p -Laplacian eigenproblem, since the p -Laplacian is nonlinear, p -eigenvectors are not necessarily be orthogonal to each other. However, we still want a relationship between p -eigenvectors. For this motivation, instead of the orthogonality, Luo et al. (2010) proposed p -orthogonality as follows.

Definition 12 Let $\Xi_p(\psi)$ be a vector, whose v -th element is $\xi_p(\psi(v))$. We call $\psi \neq 0$ and $\omega \neq 0$ as p -orthogonal if $\Xi_p(\psi)^\top \Xi_p(\omega) = 0$.

In order to analyze this p -orthogonality of our abstract class of p -Laplacians, we recall the Taylor expansion, which is often used for approximating functions in physics. For example, in the motion of a pendulum, if we approximate functions with respect to the amplitude of the angular of the pendulum by Taylor expansion, the motion equation is approximated by a simple harmonic motion (Courant and Hilbert, 1962). The Taylor expansion leads an infinite differentiable function $f(x)$ to write as $f(x) = f(a_{f,x}) + \sum_{n=1}^{\infty} \frac{f^{(n)}(x - a_{f,x})}{n!}$, where $a_{f,x}$ is a constant, and $f^{(n)}$ is a n -th derivative of f . This Taylor expansion is often used to approximate the function. If we consider approximating the function by the first order, the remainder (the second or higher terms) can be seen as the approximation error. For two functions f and g , if the error term can be written as the sum of the second or higher terms, i.e., $f(x) = g(x) + o_2$, where $o_2 = \sum_{n_f+n_g \geq 2, n_f, n_g \in \mathbb{N}^+} \beta_{f,g, n_f, n_g} (x - a_{f,x})^{n_f} (x - a_{g,x})^{n_g}$, and β_{f,g, n_f, n_g} is a coefficient, then we call the function f is equal to g up to the second order of Taylor expansion. Using this notion and p -orthogonality, we obtain the following;

Theorem 13 Let (ψ, λ^ψ) and (ω, λ^ω) be the p -eigenpairs of Δ_p . The p -eigenvectors ψ and ω are p -orthogonal up to the second order of Taylor expansion with the vertex if λ^ψ and λ^ω are not equal up to the second order of Taylor expansion.

Theorem 13 tells us that two p -eigenvectors are approximated p -orthogonal, up to the second-order of Taylor expansion.

We move our discussion to the second p -eigenpair by considering the Rayleigh quotient. In the graph p -Laplacian (Bühler and Hein, 2009) and the clique p -Laplacian case (Saito et al., 2018) and also the continuous case (Lindqvist, 2008), the global minimum of a variant of Rayleigh quotient gives the second p -eigenpair. Similarly to these works, we propose to define the following quotient as

$$R_p^{(2)}(\psi) := \frac{S_p(\psi)}{\min_{\eta} \|\psi - \eta\psi_1\|_p^p}, \quad (19)$$

where ψ_1 is the first p -eigenvector. This quotient is supported by the following theorem.

Theorem 14 The global solution of Eq. (19) is given by $\psi^* = \psi_2 + \eta^*\psi_1$, where $\eta^* = \arg\min_{\eta} \|\psi_2 - \eta\psi_1\|_p^p$, and ψ_2 is the second p -eigenvector.

This theorem shows that we have an exact identification for the second p -eigenpair; minimizing Eq. (19) gives the second p -eigenpair of Δ_p . However, the major disadvantage is that Eq. (19) is not convex and hence difficult to obtain the global optimum; optimization algorithms applied to Eq. (19) would give the local optimum instead of the global optimum.

Therefore, we next consider a strategy to get a better local optimum for a 2-way hypergraph partitioning. The idea to obtain a better optimum is using the exact p -orthogonality as a constraint, instead of the constraint “ p -orthogonal up the second order”, which each pair of p -eigenvectors must obey (Theorem 13). The reason why we use this strategy is as follows. Due to the non-convexity of Eq. (19), the solution obtained by an optimization

algorithm can be a local optimum. However, this local optimum is not guaranteed to be p -orthogonal up to the second-order to the first p -eigenvector ψ_1 while ψ_2 is so. To gain a better optimal solution, we exploit Theorem 13, and we want a constraint that enforces that the solution to be p -orthogonal up to the second-order to ψ_1 . However, it is difficult to work directly with this constraint “ p -orthogonal up to the second order”. To ease this difficulty, we propose to use an exact p -orthogonal condition as a constraint. Thanks to Theorem 13, this exact constraint can be seen as an approximated condition by the second order of Taylor expansion. We borrow this approximation idea from physics; it is common to approximate physical phenomena by the second order of Taylor expansion, such as the explained motion of a pendulum case. Following this discussion, we incorporate the exact p -orthogonality as a constraint. Then, we consider the optimization problem as,

$$\min_{\psi} J(\psi) = R_p^{(2)}(\psi) \quad \text{s.t.} \quad \Xi_p(\psi)^\top \Xi_p(\psi_1) = 0. \quad (20)$$

Since $R_p^{(2)}(\alpha\psi) = R_p^{(2)}(\psi)$ for $\alpha \neq 0$, we can arbitrarily change the scale of ψ to $R_p^{(2)}(\psi)$. Hence, we add the scale constraints to Eq. (20) as

$$J'(\psi) = R_p^{(2)}(\Xi_p^{-1}(\psi)) \quad \text{s.t.} \quad \psi^\top \psi_1 = 0, \|\psi\|_2^2 = 1, \quad (21)$$

which gives the same global minimum solution as Eq. (20). To solve Eq. (21), we propose to apply natural gradient algorithm (Amari, 1998) as shown in Algorithm 1, similarly to Luo et al. (2010). If we use a simple gradient method as $\partial J' / \partial \psi$, the orthogonal condition does not hold for each update. Instead of using this for the update of Algorithm 1, we use $\frac{\partial J'}{\partial \psi} - \psi \left(\frac{\partial J'}{\partial \psi} \right)^\top \psi$ so that we can preserve the orthogonal condition in Eq.(21) (Luo et al., 2010). The convergence of this algorithm is also guaranteed (Luo et al., 2010).

Algorithm 1 Natural Gradient Algorithms for Eigenvectors of p -Laplacian.

Input: hypergraph-gradient, and p

Initialize ψ

while the change is not sufficiently small **do**

$$\Omega \leftarrow \frac{\partial J'}{\partial \psi} - \psi \left(\frac{\partial J'}{\partial \psi} \right)^\top \psi$$

$$\psi \leftarrow \psi - \alpha \Omega$$

end while

Output: Second p -eigenvector ψ of Δ_p

6 Related work

This section compares related hypergraph 2-Laplacians and p -Laplacians and partitioning algorithms. This section is complementary to Sect. 3.4. While Sect. 3.4 defines the related p -Laplacians, this section focuses on discussing the context and explaining the difference between ours and existing ones.

One major hypergraph Laplacian is from a clique expansion way (CLIQUE). The unweighted setting edge-unnormalized 2-Laplacian was proposed in Rodriguez (2002) (CLIQUE E-UN). This 2-Laplacian and Laplacians proposed in other studies (Zien et al., 1999; Bolla, 1993; Gibson et al., 2000) are theoretically equivalent (Agarwal et al., 2006). In

this line of research, 2-Laplacian from a differential geometry viewpoint is proposed (Saito et al., 2018). When $p = 2$, this Laplacian also can be explained by the clique expansion way but normalized by a degree of edge (CLIQUE E-N). Moreover, this p -Laplacian is proposed based on forming vertex-wise energy (CLIQUE E-N-VW) (Saito et al., 2018), while ours is edge-wise energy. In Saito et al. (2018), the p -energy $S_p^{VW}(\psi)$ is defined using the norm of the hypergraph-gradient is defined $\nabla\psi$ at vertex v as

$$S_p^{VW}(\psi) := \sum_{v \in V} \|\nabla\psi(v)\|^p, \text{ where } \|\nabla\psi(v)\| := \left(\sum_{e \in E: e_{(1)}=v} \frac{|\nabla\psi(e)|^2}{|e|!} \right)^{\frac{1}{2}}. \quad (22)$$

This idea comes from the definition of the energy around vertex as $\|\nabla\psi(v)\|$ and obtains total energy by summing up those energies over all vertices. Note that if we assume the standard graph, p -Laplacian in Saito et al. (2018) corresponds to a series of graph studies (Zhou and Schölkopf, 2005; Bougleux et al., 2009), which also assume vertex-wise energy. On the other hand, ours corresponds to the p -Laplacian, which assumes edge-wise energy (Bühler and Hein, 2009; Tudisco and Hein, 2016). Hence, our work does not incorporate p -Laplacian proposed in Saito et al. (2018) since the p -Dirichlet sum setting is different. Remark that when $p = 2$, our model incorporates CLIQUE E-N-VW by using c in Table 2. However, Saito et al. (2018) did not give theoretical analyses such as the nodal domain theorem or the Cheeger inequality. Moreover, Saito et al. (2018) did not give a specific partitioning algorithm exploiting characteristics of p -Laplacian such as p -orthogonality. Hence, we need to use a general-purpose optimization method for the p -eigenproblem. However, such methods do not always leverage the characteristics of p -Laplacian, which would possibly lead to better performance in terms of space, time, and accuracy.

Another line of research is in a star expansion way, shown in Sect. 3.4. Zhou et al. (2006) proposed 2-Laplacian based on a lazy random walk view. Agarwal et al. (2006) shows that this 2-Laplacian is theoretically equivalent to Laplacians by studies of (Zien et al., 1999; Li and Solé, 1996), also further discussed in Ghoshdastidar and Dukkipati (2017a).

Other Laplacian is from a total variation way and subsequent submodular way (TV/SUB). A regularization framework for $p \geq 1$ is proposed in Hein et al. (2013) with hypergraph partitioning algorithm for $p = 1$, and further explored in Chan et al. (2018). This idea is extended to a submodular hypergraph (Li and Milenkovic, 2018). A submodular hypergraph has an objective function of energy using one form of Lovász expansion of a submodular function. Moreover, SUB incorporates the inhomogeneous cut proposed by Li and Milenkovic (2017), where weights can vary when we partition the edge. Along with this new class of hypergraph cut, Li and Milenkovic (2018) proposed partitioning algorithms for $p = 1$ and $p = 2$. Seeing the definition (Eq. (15)), submodular p -Laplacian describes a broad class hypergraph p -Laplacian using submodular function. We also mention the $p = 2$ case for submodular cut objective functions are discussed in Yoshida (2019) using the general form of Lovász extension. Moreover, a series of research (Veldt et al., 2020; Benson et al., 2020) directly defines objective function using submodular function, instead of Lovász extension. While submodular models seem flexible, ours are more versatile since we do not assume submodularity. The submodular p -Laplacian is a special case of ours as long as the conditions in Definition 1 are satisfied. Additionally, our algorithm can address arbitrary p , while algorithms from Hein et al. (2013) and Li and Milenkovic (2018) focused on the specific p ($p = 1$ or $p = 2$).

We remark that our framework can address existing 2-Laplacian from `CLIQUE` and `STAR`, and `TV/SUB` p -Laplacian. Moreover, our partitioning algorithm can work for arbitrary $p \geq 1$, while those existing algorithms focus on specific p or use a general-purpose optimization algorithm without theoretical analyses. We also note that our framework can define a new p -Laplacian, which is (but not limited to) normalized TV, shown in Sect. 3.4. However, we need to recognize that it is out of the scope of our work to incorporate `CLIQUE E-N-VW` p -Laplacian. Moreover, since our framework based on the relationships of Propositions 4 and 5, our framework does not incorporate a tensor modeled Laplacian for uniform hypergraph, where all edges connect the same number of vertices (Cooper and Dutle, 2012; Hu and Qi, 2012; Qi, 2013; Hu and Qi, 2015; Chen et al., 2017; Ghoshdastidar and Dukkipati, 2017b; Chang et al., 2020). The reason why we cannot incorporate `CLIQUE E-N-VW` p -Laplacian and tensors into our model is that our model is based upon the energy formed as Eq. (8), while energies for those two are differently defined. Particularly, we note that the difference in the aims between tensor modeled Laplacians and our framework is as follows; while the tensor modeled Laplacians are the tensor operation, our framework focuses on the contraction made by the energy Eq. (8).

Lastly, we comment on p -Laplacians in the continuous domain. The continuous p -Laplacian has a longer history than in the discrete domain. The Dirichlet energy is defined similarly to Eq. (8), and the variation of the energy would give the Laplace equation (Courant and Hilbert, 1962). The energy is minimized when the Laplace equation is satisfied. This framework extended to arbitrary p -norm, such as Binding and Rynne (2008), and was theoretically analyzed in many ways, such as nodal domain theorem and Cheeger inequality. We remark that in the continuous case, we can identify the second p -eigenpair similarly to Eq. (19), but no exact identification for the third or higher has been found yet (Lindqvist, 2008). For more comprehensive study, we refer to Lindqvist (2008) and Struwe (2000).

7 Preliminary experiments

Our experiments aim to evaluate our approximation algorithm (Algorithm 1) as a function of p and the particular type of hypergraph Laplacians (`STAR`, `CLIQUE`, and `TV/SUB`).

Primary Objective of the Experiments. The objective of the experiments is to see if our algorithm (Algorithm 1) improves on the existing methods introduced in Sect. 6. Algorithm 1 has two key “levers”; the choice of the parameter p and the choice of hypergraph Laplacian, i.e., the function c in the gradient (Definition 1). On the one hand, in the previous works discussed in Sect. 6 the algorithms for hypergraph p -Laplacians were designed for a particular p (e.g., $p = 1, 2$) or applied to all $p > 1$ without theoretical justifications. On the contrary, Algorithm 1 for our abstract class of hypergraph p -Laplacians works for all $p > 1$ with theoretical justification. Therefore, we provide experiments for a wide range of hypergraph Laplacians for $p > 1$ in comparison to existing algorithms. We thus apply Algorithm 1 to five existing hypergraph Laplacians (`CLIQUE E-N`, `CLIQUE E-UN`, `STAR`, `TV V-UN`, and `TV V-N`) for $p > 1$. We compare these to the existing fixed p algorithms for particular type of Laplacians. Moreover, since `CLIQUE E-N` has a partitioning algorithm using a particular hypergraph p -Laplacian (`CLIQUE E-N-VW` by Saito et al. (2018), see Sect. 6 for the definition), we also compare to this. Hence, we compare five instantiations of ours with six previous algorithms as;

- Algorithm 1 for all $p > 1$ is applied to the five geometries:

Table 3 Dataset summary. All the dataset has two classes. The parameter δ is the average edge degree parameter $\delta := \sum_{e \in E} |e|/|E|$, and $\tau := \sum_{e \in E} |e| / |V||E|$ is the average ratio of the number of vertices connected by each edge to the total number of vertices, which we can recognize as "density" of hypergraph

	Mushroom	Chess	Cancer	Congress	News(1,2)	News(3,4)
$ V $	8124	3196	699	435	8124	8188
$ E $	112	73	90	48	100	100
$\sum_{e \in E} e $	170604	115056	6291	6960	31066	34382
δ	1523.25	1576.11	69.9	145	310.66	343.82
τ	0.18	0.49	0.10	0.33	0.038	0.042

1. CLIQUE E-N: $p > 1$
 2. CLIQUE E-UN: $p > 1$
 3. STAR: $p > 1$
 4. TV V-UN: $p > 1$
 5. TV V-N: $p > 1$
- Comparison as six existing algorithms:
 1. CLIQUE E-N: $p = 2$ (Saito et al., 2018)
 2. CLIQUE E-N-VW: $p > 1$ (Saito et al., 2018)
 3. CLIQUE E-UN: $p = 2$ (Rodriguez, 2002)
 4. STAR: $p = 2$ (Zhou et al., 2006)
 5. TV V-UN: $p = 1$ (Hein et al., 2013)
 6. TV V-N: $p = 1$ (Hein et al., 2013)

Note that there are variety of submodular function for SUB can be considered, but we made TV by Hein et al. (2013) as a representative of the SUB group.

Experimental Setup. We build a hypergraph using the method for categorical datasets introduced in Zhou et al. (2006). Each instance in the dataset consists of $|E|$ categories. The vertices of the hypergraph are the instances. The edges are defined by the attribute values. Each attribute value within a given category defines an edge where each vertex in the edge corresponds to those instances that share the same attribute value. All edges are given weight one. Our experiment is performed on the datasets mushroom, cancer, chess, and congress from the UCI repository (Dua and Graff, 2022), and two datasets created from 20newsgroups¹ (for short "news") with two classes (1,2) and (3,4). All of these were used in the previous studies (Zhou et al., 2006; Hein et al., 2013; Saito et al., 2018). We summarize the datasets in Table 3. The value $\delta = \sum_{e \in E} |e|/|E|$ is the average edge degree. Furthermore, $\tau = \sum_{e \in E} |e| / |V||E|$ is the average ratio of the number of vertices connected by each edge to the total number of vertices, which we can recognize as "density" of a hypergraph. In Table 4 we compare 11 instantiations of hypergraph p -Laplacians as discussed above. For CLIQUE E-N-VW ($p \in [1, 3]$) we conducted experiments using the same setting as Saito et al. (2018) as the setting matches

¹ We used the tiny version of the original 20newsgroups available at https://cs.nyu.edu/~roweis/data/20news_w100.mat.

Table 4 The experimental results for hypergraph partitioning for our methods and existing ones. We applied our algorithm 1 for $p > 1$ to five geometry of the hypergraph Laplacians (CLIQUE E-N, CLIQUE E-UN, STAR, TV V-UN, TV V-N). We compared these to the existing fixed p algorithms for the five hypergraph Laplacians; CLIQUE E-N for $p = 2$ by Saito et al. (2018), CLIQUE E-UN for $p = 2$ is by Rodriguez (2002), STAR $p = 2$ is by Zhou et al. (2006), and TV V-UN and TV V-N for $p = 1$ is by Hein et al. (2013). Moreover, for CLIQUE E-N, we also compared with the algorithm for $p > 1$ (CLIQUE E-N-VW) by Saito et al. (2018). Thus, we compare five instantiations of ours with six existing ones. We compare the performance by error. Performance with ours is marked with #. For free-parameter $p > 1$, we give the value of p giving the smallest error in the column p next to the dataset. The superscripted * means fixed-parameter. See the main text for more discussion

	Mushroom	p	Chess	p	Cancer	p	Congress	p	News(1,2)	p	News(3,4)	p
# CLIQUE E-N p	0.1118	1.1	0.4765	2.1	0.0243	1.7	0.1172	1.6	0.2390	1.1	0.2146	1.1
$p = 2$	0.3156	2.0*	0.4781	2.0*	0.0286	2.0*	0.1241	2.0*	0.2906	2.0*	0.2173	2.0*
CLIQUE E-N-VW p	0.2329	1.0	0.2847	2.6	0.0243	1.5	0.1195	2.5	0.2442	3.0	0.2167	1.1
# CLIQUE E-UN p	0.1211	1.1	0.4768	2.6	0.0358	2.2	0.1195	1.4	0.2493	1.7	0.3272	1.1
$p = 2$	0.4791	2.0*	0.4772	2.0*	0.1659	2.0*	0.1379	2.0*	0.2511	2.0*	0.3447	2.0*
# STAR p	0.1113	1.1	0.4759	1.3	0.0286	1.1	0.1195	2.3	0.2411	2.4	0.2162	1.7
$p = 2$	0.3156	2.0*	0.4781	2.0*	0.0300	2.0*	0.1333	2.0*	0.2477	2.0*	0.2169	2.0*
# TV V-UN p	0.1083	1.1	0.4659	2.6	0.1960	1.1	0.1862	1.2	0.2701	1.1	0.3274	1.7
$p = 1$	0.1349	1.0*	0.4778	1.0*	0.3362	1.0*	0.3034	1.0*	0.3072	1.0*	0.4313	1.0*
# TV V-N p	0.1083	1.1	0.4643	2.7	0.2489	1.6	0.1839	1.5	0.2672	1.1	0.3275	1.4
$p = 1$	0.1088	1.0*	0.4997	1.0*	0.4692	1.0*	0.3034	1.0*	0.2983	1.0*	0.4544	1.0*

to ours. For our methods we used $p \in \{1.1, 1.2, \dots, 3.0\}$; we limited ourselves to $p \leq 3$ since the Cheeger Inequality (Theorem 10), is progressively looser for larger p values. For the free-parameter experiments, we set the starting condition of our algorithm to the solution of the corresponding fixed-parameter Laplacian. We used the step size $\alpha = 0.01 \|\psi\|_1 / \|\Omega\|_1$ as done in Luo et al. (2010). For all methods, a second eigenvector was hence computed, and we used the k -means objective to determine the “split point” on the eigenvector (as was also done in (Zhou et al., 2006; Saito et al., 2018)). We evaluated the performance of our algorithms via their error rate, i.e., (# of errors)/(# of data), as used in (Zhou et al., 2006; Hein et al., 2013; Saito et al., 2018).

Overall Results The results are summarized in Table 4. First, looking into our algorithm (Algorithm 1) vs. fixed-parameter algorithms (existing methods, see the performances associated with in Table 3) for five geometries, we see that our methods consistently demonstrate improved performance from existing fixed-parameter methods. We also remark that among for CLIQUE E-N ours consistently outperforms CLIQUE E-N-VW, except chess.

Further Discussion A natural question to ask for our algorithm is “which hypergraph Laplacian and which p is suitable?”. A further look into our abstract class of p -Laplacians can answer this question; the experimental result reveals how the choice of p and type of the hypergraph Laplacian are connected to the underlying parameters δ (average edge degree) and τ (density) of the datasets. Although the experiments are preliminary, there seem to be consistent trends that provide guidance on a range of p and the type of Laplacian to consider. Further, the experimental guidance is supported by the theory given earlier in this manuscript.

Our observation is that the density parameter (τ) is related to the range of p while the the average edge degree parameter (δ) is connected to the hypergraph Laplacian. The density parameter (τ) indicates the natural range for p . The dataset chess is significantly more dense (large τ) than the other datasets. The table indicates that while large p tends to work better for the chess dataset, the tendency is that small p improves on large p for the non-chess datasets. To understand this, we consider the trade-off between the Cheeger inequality (see Theorem (10)) and the p -Dirichlet sum. The Cheeger inequality is tighter for smaller p ; hence, the relaxed objective becomes closer to the discrete objective. On the other hand, if we examine the p -Dirichlet sum (see Eq. (8)), one may observe that it is a p -norm to p -th power of the hypergraph-gradient. The dimensionality of hypergraph-gradient scales with the graph density (τ). Hence in the dense case, a relatively larger p is needed to induce the same magnitude of change in the p -Dirichlet sum, which is connected to the second p -eigenvector via Rayleigh quotient (see Eq. (14)). The analogous phenomena connecting the choice of p to density have been observed in a standard graph such as online graph transduction (Herbster and Lever, 2009). Turning to the average edge degree parameter (δ), we observe the following preliminary indications that suggest how to choose the Laplacian as a function of δ . There we see on the large δ dataset (chess and mushroom) that all TV methods out perform STAR and CLIQUE methods of our p -Laplacian whereas for the other smaller δ datasets all STAR and clique methods outperform all TV methods. We have provided some guidance on the choice of Laplacian and the range of p based on the density τ and average edge

degree δ of the graph. For more detailed experimental results and further discussion, see Appendix 12.

We further observe a different behavior than the semi-supervised learning in (Alamgir and Luxburg, 2011; Slepcev and Thorpe, 2019) using the same energy Eq. (8) in the standard graph setting. These works deal with the case of semi-supervised learning using p -Laplacians of a graph with an asymptotically large number of vertices. In this case, the problem does not degenerate into the trivial one when p is large, while the problem does so when p is small. However, from these experimental results, we observed a different behavior; small p also works when τ is small, as we discussed. This might be because there is a structural difference in the use of the p -Laplacian between semi-supervised learning and unsupervised learning.

8 Conclusion

This work has considered hypergraph spectral clustering. We have proposed a general framework for hypergraph p -Laplacian and provided theoretical results for our p -Laplacian. We also have proposed a convergent hypergraph partitioning algorithm with respect to our abstract class of p -Laplacian exploiting theoretical results. Our experiment has shown that our algorithm outperforms the existing spectral clustering algorithms for hypergraph Laplacians. Also, we have shown practical guidance on the choice of p -Laplacian.

There are several future directions. A fruitful direction would be to explore if our p -Laplacian can converge to the continuous p -Laplace operator in the limit of infinite data, similarly to the graph case (Belkin and Niyogi, 2003) and the hypergraph case (Saito, 2022). Moreover, similarly to the previous studies (Hein et al., 2013; Saito et al., 2018), semi-supervised learning using S_p as a regularizer would be valuable for a future study. Furthermore, while we conduct our experiment on a real dataset, it would be interesting to conceive an illustrative toy dataset where some hypergraph Laplacian works better than the others or where some p works while $p = 2$ does not. It would also be valuable to study multi-class clustering for arbitrary p using higher-order eigenvectors similarly to the standard graph 2-Laplacian case (von Luxburg, 2007), as opposed to the methods using recursive one-vs-rest two-class partitioning (Bühler and Hein, 2009; Hein et al., 2013). However, unlike the 2-Laplacian matrix, where those can be easily obtained, it would be difficult to obtain the third or higher p -eigenpairs of p -Laplacian. The reason is that while we know the algebraic identification for the second p -eigenpair (Eq. (19)), there have not been such identifications for the higher eigenpairs both in the discrete and continuous domain (Lindqvist, 2008).

Appendix

Discussion on the conditions of definition 1

In this section, we discuss the conditions of the operator ∇ and the function $c(u, v, e, \psi)$ by drawing examples. We use the examples listed in Table 2, which is mainly discussed in Sect. 3.4.

The first condition

$$\nabla(\alpha\psi) = \alpha\nabla\psi \quad \text{or} \quad \nabla(\alpha\psi) = |\alpha|\nabla\psi \tag{23}$$

forces the operator ∇ to be homogeneous or absolute homogeneous. For the examples in Table 2, this condition for cliques and STAR is also obviously satisfied since the functions c for cliques and STAR are independent of ψ . For the total variation, we obtain

$$\begin{aligned} c(u, v, e, \alpha\psi) &= \begin{cases} 1 & ((u, v) = \operatorname{argmax}_{u,v} \left| \alpha \frac{\psi(u)}{\mu(u)} - \alpha \frac{\psi(v)}{\mu(v)} \right|) \\ 0 & (\text{otherwise}) \end{cases} \\ &= \begin{cases} 1 & ((u, v) = \operatorname{argmax}_{u,v} \left| \alpha \left| \frac{\psi(u)}{\mu(u)} - \frac{\psi(v)}{\mu(v)} \right| \right) \\ 0 & (\text{otherwise}) \end{cases} \\ &= \begin{cases} 1 & ((u, v) = \operatorname{argmax}_{u,v} \left| \frac{\psi(u)}{\mu(u)} - \frac{\psi(v)}{\mu(v)} \right|) \\ 0 & (\text{otherwise}) \end{cases} \\ &= c(u, v, e, \psi), \end{aligned} \tag{24}$$

which therefore satisfies the condition Eq. (23). For SUB, we compute

$$\begin{aligned} (\nabla\alpha\psi)(e) &= \sum_{u,v \in e} w^{\frac{1}{p}}(e) c^{\frac{1}{p}}(u, v, e, \alpha\psi) \alpha \left(\frac{\psi(u)}{\mu^{1/p}(u)} - \frac{\psi(v)}{\mu^{1/p}(v)} \right) \\ &= \sum_{i=1}^{|e|} \alpha F(S_i) \max_{S \subseteq e} (F(S)) \left(-\frac{\psi(u_{i+1})}{\mu^{1/p}(u_{i+1})} + \frac{\psi(u_i)}{\mu^{1/p}(u_i)} \right) \\ &= -\alpha \nabla\psi(e), \end{aligned} \tag{25}$$

and therefore this satisfies the condition.

Secondly, we discuss the first condition, which is

$$\sum_{u,v \in e} c(u, v, e, \psi) = c(e). \tag{26}$$

The condition Eq. (26) wants the summation of the function over all the pairs of vertices at an edge e to be independent of ψ . For the examples in Table 2, this condition for cliques and STAR is satisfied since the functions c for cliques and STAR are independent of ψ . For the total variation, the function c depends on ψ . However, since the function c for total variation can be written as

$$c(u, v, e, \psi) = \begin{cases} 1 & ((u, v) = \operatorname{argmax}_{u,v} |\psi(u) - \psi(v)|) \\ 0 & (\text{otherwise}), \end{cases} \tag{27}$$

then the summation is

$$\sum_{u,v \in e} c(u, v, e, \psi) = 1. \tag{28}$$

Therefore, the summation of $c(u, v, e, \psi)$ over u, v is independent of ψ although $c(u, v, e, \psi)$ is dependent on ψ . To see SUB, we observe that

$$\sum_{u,v \in V} c(u, v, e, \psi) = \sum_{i=1}^{|e|} \max_{S \subseteq e} (F(S)) F(S_i) = c(e) \tag{29}$$

which is independent of the order of ψ and the particular vertices u, v .

The third condition is

$$\partial c^{\frac{1}{p}}(u, v, e, \psi) / \partial \psi \Big|_{v: v \in e} = 0, \forall e \in E, u \in e, \quad (30)$$

$$\partial c^{\frac{1}{p}}(u, v, e, \psi) / \partial \psi \Big|_{u: u \in e} = 0, \forall e \in E, v \in e \quad (31)$$

which means the function $c(u, v, e, \psi)$ is constant once we fix $e \in E$ and one vertex in the edge. From this condition, we obtain

$$\begin{aligned} & \frac{\partial}{\partial \psi} c^{\frac{1}{p}}(u, v, e, \psi) \Big|_{v: v \in e} \\ &= \frac{\partial}{\partial \psi} c^{\frac{1}{p}}(u, v, e, \psi) \Big|_{v: v \in e} \psi + c^{1/p}(u, v, e, \psi) \\ &= c^{1/p}(u, v, e, \psi), \forall e \in E, u \in e. \end{aligned} \quad (32)$$

Similarly, we can prove the condition of u . This implies that the function c works as a coefficient for ψ once we fix v and e , and c is independent of ψ . Note that if c is not differentiable, then we simply change to subdifferentiation as

$$\partial c^{\frac{1}{p}}(u, v, e, \psi) \Big|_{v: v \in e} \ni 0, \forall e \in E, u \in e, \quad (33)$$

$$\partial c^{\frac{1}{p}}(u, v, e, \psi) \Big|_{u: u \in e} \ni 0, \forall e \in E, v \in e. \quad (34)$$

For examples in Table 2, this condition is also obviously satisfied, since the functions c for cliques and star are independent of ψ . Although the function c for total variation depends on ψ , the function c is constant once we fix one vertex and one edge e . Moreover, this implies that the function c satisfies the condition Eq. (31).

More discussion on the p -Laplacian definition in Sect. 3.1

Proof of propositions 4 and 5

For the convenience of the other proofs, we start our discussion from Eqs. (4) and (5). Equation (4) can be shown by

$$\begin{aligned} \langle \psi, \Delta_p \psi \rangle_{\mathcal{H}(V)} &= \langle \psi, -\operatorname{div} \|\nabla \psi\|^{p-2} \nabla \psi \rangle_{\mathcal{H}(V)} \\ &= \langle \nabla \psi, \|\nabla \psi\|^{p-2} \nabla \psi \rangle_{\mathcal{H}(E)} \\ &= \sum_{e \in E: v \in e} \|\nabla \psi\|^{p-2} \frac{(\nabla \psi)^2(e)}{|e|!} \\ &= \|\nabla \psi(e)\|_p^p = S_p(\psi). \end{aligned} \quad (35)$$

Equation (5) can be shown by the following. By differentiating $S_p(\psi)$ by ψ at the vertex v , we obtain

$$\begin{aligned}
 \left. \frac{\partial}{\partial \psi} S_p(\psi) \right|_v &= \left. \frac{\partial}{\partial \psi} \sum_{e \in E} \frac{1}{|e|!} |(\nabla \psi)(e)|^p \right|_v \\
 &= \sum_{e: v \in e} p \frac{1}{|e|!} |(\nabla \psi)(e)|^{p-1} \left. \frac{\partial}{\partial \psi} |(\nabla \psi)(e)| \right|_v \\
 &= p \sum_{e: v \in e} \frac{1}{|e|!} |(\nabla \psi)(e)|^{p-1} \operatorname{sgn}((\nabla \psi)(e)) \frac{w^{1/p}(e)}{\mu^{1/p}(v)} \\
 &\quad \times \left(\sum_{u \in e; \{v, u\} \subseteq e} c^{1/p}(v, u, e, \psi) - \sum_{u \in e; \{u, v\} \subseteq e} c^{1/p}(u, v, e, \psi) \right)
 \end{aligned} \tag{36}$$

By using Eq. (36), we consider the following equation.

$$\begin{aligned}
 \langle \psi, \frac{1}{p} \frac{\partial}{\partial \psi} S_p(\psi) \rangle_{\mathcal{H}(V)} &= \sum_{v \in V} \psi(v) \sum_{e: v \in e} \frac{1}{|e|!} |(\nabla \psi)(e)|^{p-1} \operatorname{sgn}(\nabla \psi(e)) \frac{w^{1/p}(e)}{\mu^{1/p}(v)} \\
 &\quad \times \left(\sum_{u \in e; \{v, u\} \subseteq e} c^{1/p}(v, u, e, \psi) - \sum_{u \in e; \{u, v\} \subseteq e} c^{1/p}(u, v, e, \psi) \right)
 \end{aligned} \tag{37}$$

As the summation in Eq. (37) runs over all vertices $v \in V$, we can reconstruct all pairs of vertices in each edge. Therefore,

$$\begin{aligned}
 &\sum_{v \in V} \psi(v) \sum_{e: v \in e} \operatorname{sgn}(\nabla \psi(e)) \frac{w^{1/p}(e)}{\mu^{1/p}(v)} \\
 &\times \left(\sum_{u \in e; \{v, u\} \subseteq e} c^{1/p}(v, u, e, \psi) - \sum_{u \in e; \{u, v\} \subseteq e} c^{1/p}(u, v, e, \psi) \right) \\
 &= \sum_e |\nabla \psi(e)|
 \end{aligned} \tag{38}$$

From this and Eq. (37), we obtain

$$\begin{aligned}
 \langle \psi, \frac{1}{p} \frac{\partial}{\partial \psi} S_p(\psi) \rangle_{\mathcal{H}(V)} &= \sum_e \frac{1}{|e|!} |(\nabla \psi)(e)|^p \\
 &= \|\nabla \psi\|_p^p \\
 &= S_p(\psi)
 \end{aligned} \tag{39}$$

From Eq. (4) and Proposition (5), we can show that

$$\frac{1}{p} \frac{\partial}{\partial \psi} S_p(\psi) = \Delta_p \psi \tag{40}$$

Basic properties for the definition of the hypergraph-gradient

The following basic properties of hypergraph-gradient easily follow from the definition.

Proposition 15 *The hypergraph-gradient has the following properties.*

$$(\nabla cM^{1/p})\mathbf{1}(e) = \mathbf{0}, \forall e \in E, \quad (41)$$

$$(\nabla \mathbf{0})(e) = \mathbf{0}, \forall e \in E, \quad (42)$$

$$\partial^2 \nabla \psi / \partial^2 \psi \Big|_{\psi} = 0, \forall e \in E \quad (43)$$

Also, hypergraph-gradient is not 0 except Eqs. (41) and (42).

These properties directly follow from the definition of the hypergraph-gradient.

Proof of proposition 7

By differentiating Eq. (14) by ψ , we can obtain the condition for critical points of Eq. (14) as follows;

$$\Delta_p \psi - \frac{S_p(\psi)}{\|\psi\|_p^p} \xi_p(\psi) = 0 \quad (44)$$

By Definition 6, we can immediately show that ψ is an eigenvector of Δ_p . Moreover the eigenvalue λ can be obtained by $S_p(\psi)/\|\psi\|_p^p$. The last statement can be shown immediately by the definition.

By the definition of Rayleigh quotient, we immediately have the following property.

Corollary 16 *We have $R_p(\alpha\psi) = R_p(\psi)$ for $\alpha \in \mathbb{R}, \alpha \neq 0$.*

For the first p -eigenvector, we compute p -Laplacian by differentiating by ψ , that is

$$p\Delta_p \psi = \frac{\partial}{\partial \psi} S_p. \quad (45)$$

Then, we obtain

$$\begin{aligned} \frac{\partial}{\partial \psi} p\Delta_p \psi &= \frac{\partial}{\partial \psi} S_p(\psi) \\ &= \frac{\partial}{\partial \psi} \|\nabla \psi(e)\|_p^p \\ &= \frac{\partial}{\partial \psi} \left(\sum_e \frac{|\nabla \psi(e)|^p}{|e|!} \right) \\ &= p \sum_e \frac{|\nabla \psi(e)|^{p-1}}{|e|!} \times \frac{1}{|e|!} \frac{\partial}{\partial \psi} \nabla \psi(e) \end{aligned} \quad (46)$$

From Eq. (43), the derivative of hypergraph-gradient is independent of ψ . Therefore, from Eq. (46), the p -Laplacian $\Delta_p \psi$ equals to 0 if $\nabla \psi(e) = 0, \forall e \in E$. This means that $\Delta_p \mathbf{0} = 0$. Also, $\Delta_p cM^{1/p} \mathbf{1} = 0$.

As the p -eigenvalue is equal or greater than 0 from Proposition 7, the first p -eigenvector is $M^{1/p}\mathbf{1}$, associated with the first p -eigenvalue 0.

The following corollary follows.

Corollary 17 $\frac{\partial}{\partial \psi} \Delta_p \psi |_{\psi=cM^{1/p}\mathbf{1}} = 0$.

Proof Similarly to the proof of p -eigenvector, we compute

$$\begin{aligned}
 p \frac{\partial}{\partial \psi} \Delta_p \psi &= \frac{\partial^2}{\partial^2 \psi} S_p(\psi) \\
 &= \frac{\partial}{\partial \psi} p \sum_e \frac{|\nabla \psi(e)|^{p-1}}{|e|!} \times \frac{1}{|e|!} \frac{\partial}{\partial \psi} \nabla \psi(e) \\
 &= p(p-1) \sum_e \frac{|\nabla \psi(e)|^{p-2}}{|e|!} \times \left(\frac{1}{|e|!} \frac{\partial}{\partial \psi} \nabla \psi(e) \right)^2 \\
 &\quad + p \sum_e \frac{|\nabla \psi(e)|^{p-1}}{|e|!} \times \frac{1}{|e|!} \frac{\partial^2}{\partial^2 \psi} \nabla \psi(e)
 \end{aligned} \tag{47}$$

As the second derivative of hypergraph-gradient is 0 from Eq.(43), $\Delta_p \psi = 0$ if $\nabla \psi(e) = 0, \forall e \in E$. □

Properties of p -Eigenpair of p -Laplacian

In this section, we remark that interesting properties of p -Eigenpair of p -Laplacian.

We firstly remark that the definition of p -eigenpair (Definition 6) leads to existing an infinite number of p -eigenpairs, similarly to the continuous case (Binding and Rynne, 2008).

We move our discussion to a property of multiplicity of first p -eigenvalues.

Proposition 18 *Suppose that hypergraph G is a union of k independent and connected hypergraphs G_i ($i = 1, \dots, k$), i.e., $G = \bigcup_{i=1}^k G_i$ where $G_j \cap G_l = \emptyset$, for, $j, l \neq i$. Then, k equals to the multiplicity of eigenvalue 0 of Δ_p .*

The following corollary follows from this proposition.

Corollary 19 *The second p -eigenvalue of Δ_p is greater than 0, if a hypergraph G is connected.*

To analyze critical point of Eq. (14), Index theory (Struwe, 2000) is useful. We use Krasnoselskii genus (Struwe, 2000) γ for a set A , that is defined as $\gamma(A) := 0$ if $A = \emptyset$, $\gamma(A) := \inf\{k \in \mathbf{Z}^+ \mid \exists \text{ odd continuous } h : A \rightarrow \mathbf{R}^k \setminus \{0\}, \text{ and } \gamma := \infty \text{ when no such } h \text{ exists } \forall j \in \mathbf{Z}^+$. In this context, this genus is a generalized concept of dimension of A . Since $R_p(\alpha\psi) = R_p(\psi)$ by Corollary 16, to consider the p -eigenpair of p -Laplacian, we can limit our interest to $S_p := \{\psi \mid \|\psi\|_p^p = 1\}$. From the results in discrete case (Chang, 2016; Tudisco and Hein, 2016; Li and Milenkovic, 2018) and continuous case (Lindqvist, 2008), we obtain the following proposition, which is a generalized Rayleigh min-max theorem.

Proposition 20 Consider the set of subsets $\zeta_k(\mathcal{S}_p) = \{A \subset \mathcal{S}_p \mid A = -A, \text{ closed}, \gamma(A) \geq k\}$. The sequence defined as

$$\lambda_k = \min_{A \subset \zeta_k(\mathcal{S}_p)} \max_{\psi \in A} R_p(\psi) \quad (48)$$

gives a critical point of $R_p(\psi)$, whose corresponding ψ and λ_k constitute a p -eigenpair of Δ_p .

Similarly to the Rayleigh min max theorem, this proposition yields the sequence of p -eigenpairs. Moreover, for a standard Laplacian of standard graph, this reduces into Rayleigh min-max theorem. However, similarly to the continuous p -Laplacian theory (Lindqvist, 2008), we do not know if this sequence yields exhaustive p -eigenpairs.

Proof of proposition 18

As we observe $R_p(\psi) \geq 0$ from the definition, we can show that all the p -eigenvalues are non-negative. We denote by $\mathbf{1}_{G_i}$ a vector whose size of vector is the number of vertices of G and fill 1 to the elements corresponds to G_i and else 0. By using this notation, we show that $\Delta_p(cM^{-1/p}\mathbf{1}_{G_i}) = 0$ for all $i = 1, \dots, k$, which shows that those vectors are p -eigenvector and corresponding p -eigenvalues are 0. From the definition of p -Laplacian, those are the only p -eigenvectors whose p -eigenvalues are 0. The above shows that the multiplicity of p -eigenvalues of 0 equals to the number of independent components.

Proof of proposition 20

We start the proof by introducing a classical notion of *locally Lipschitz*.

Definition 21 A function $g : \mathcal{S}_p \rightarrow \mathbb{R}$ is locally Lipschitz when for each $x \in \mathcal{S}_p$, there exists a neighborhood \mathcal{N}_x of x and a constant C depending on \mathcal{N}_x such that $|g(x') - g(x)| \leq C\|x' - x\|_2 \forall x' \in \mathcal{S}_p \cap \mathcal{N}_x$.

Here we obtain the following observation.

Lemma 22 $R_p(\psi)|_{\mathcal{S}_p}$ is locally Lipschitz.

$$\begin{aligned} \left| \frac{S_p(\psi')}{\|\psi'\|_p^p} - \frac{S_p(\psi)}{\|\psi\|_p^p} \right| &= \left| \frac{S_p(\psi')\|\psi\|_p^p - S_p(\psi)\|\psi'\|_p^p}{\|\psi'\|_p^p\|\psi\|_p^p} \right| \\ &\leq \frac{\sup_{\psi' \in \mathcal{N}_x \cap \mathcal{S}_p} S_p(\psi')}{\inf_{\psi' \in \mathcal{N}_x \cap \mathcal{S}_p} \|\psi'\|_p^{2p}} \left| \|\psi'\|_p^p - \|\psi\|_p^p \right| \\ &\leq \frac{\sup_{\psi' \in \mathcal{N}_x \cap \mathcal{S}_p} S_p(\psi')}{\inf_{\psi' \in \mathcal{N}_x \cap \mathcal{S}_p} \|\psi'\|_p^{2p}} \left(\sup_{\psi' \in \mathcal{N}_x \cap \mathcal{S}_p} p\|\psi'\|_p^{p-1} \|\psi'\|_p - \|\psi\|_p \right) \\ &\leq \frac{\sup_{\psi' \in \mathcal{N}_x \cap \mathcal{S}_p} S_p(\psi')}{\inf_{\psi' \in \mathcal{N}_x \cap \mathcal{S}_p} \|\psi'\|_p^{2p}} \left(\sup_{\psi' \in \mathcal{N}_x \cap \mathcal{S}_p} p\|\psi'\|_p^{p-1} \|\psi' - \psi\|_p \right) \end{aligned} \quad (49)$$

Proof

From Eq.(49), if we choose \mathcal{N}_x as the space where $\|\psi' - \psi\|_p < \|\psi' - \psi\|_2$, e.g., $|\psi' - \psi| < 1$, then we can conclude $R_p(\psi)|_{\mathcal{S}_p}$ is locally Lipschitz. \square

Next, we introduce the classical result of Lunsternik–Schinirelman theorem.

Theorem 23 [Struwe (2000)] *Suppose function $g : \mathcal{S}_p \rightarrow \mathbb{R}$ is a locally Lipschitz, then*

$$\lambda_k = \min_{A \subset \zeta_k(\mathcal{S}_p^m)} \max_{x \in A} g(x) \quad (50)$$

yields a sequence of critical values of g .

By applying Theorem 23 to R_p , we can show that the sequence in Eq. (20) yields a critical values of R_p , which are p -eigenvalues of p -Laplacian.

Discussion on table 2

We discuss how Table. 2 connects to the existing Laplacians, by splitting the discussion by clique, star, and total variation Laplacians.

We note that all the functions in Table 2 satisfies the conditions of Definition 1, which is discussed in Sect. A.

Clique Laplacians

The hypergraph-gradient for edge-normalized Laplacian is given as

$$(\nabla\psi)(e) = \frac{w^{1/p}(e)}{(|e| - 1)^{1/p}} \sum_{i=1}^{|e|} \left(\frac{\psi(v_i)}{\mu^{1/p}(v_i)} - \frac{\psi(v_1)}{\mu^{1/p}(v_1)} \right), \quad (51)$$

and edge-unnormalized Laplacian can be written as

$$(\nabla\psi)(e) = w^{1/p}(e) \sum_{i=1}^{|e|} \left(\frac{\psi(v_i)}{\mu^{1/p}(v_i)} - \frac{\psi(v_1)}{\mu^{1/p}(v_1)} \right). \quad (52)$$

To show that the function c for clique can induce existing clique Laplacian, we only consider when $p = 2$, and $\mu(v) = d(v)$. The following proposition directly shows that Saito's 2-Laplacian is our 2-Laplacian for clique setting.

Proposition 24 *Let e be $e = \{v_1, \dots, v_{|e|}\}$. Then if we choose hypergraph-gradient operator $\nabla : \mathcal{H}(V) \rightarrow \mathcal{H}(E)$ for Saito et al. (2018) as*

$$(\nabla\psi)(e) := \frac{\sqrt{w(e)}}{\sqrt{|e| - 1}} \sum_{i=1}^{|e|} \left(\frac{\psi(v_i)}{\sqrt{d(v_i)}} - \frac{\psi(v_1)}{\sqrt{d(v_1)}} \right). \quad (53)$$

The induced 2-Laplacian correspond to 2-Laplacians proposed by Saito et al. (2018). If we choose the same hypergraph-gradient but omitted denominator $\sqrt{|e| - 1}$, then the induced 2-Laplacian corresponds to Rodriguez's Laplacian.

This also shows that Rodriguez 2-Laplacian is our edge-unnormalized clique 2-Laplacian.

The 2-Laplacians L are given as

$$L = I - D^{-1/2}WD^{-1/2}, \quad (54)$$

where for an edge-normalized setting W is a matrix whose element is $w(u, v) = \sum_{uv \in e} w(e) / (|e| - 1)$ and $D = D_v$, and for an edge-unnormalized setting $w(u, v) = \sum_{u, v \in e} w(e)$ and D is a diagonal matrix whose element $d(v, v) = \sum_{v \in e} w(e)$.

In Saito et al. (2018), they used the different energy setup for p -Laplacian, as discussed in Sect. 6. However, when $p = 2$, Saito's 2-Laplacian matches our clique normalized 2-Laplacian. Actually, in the case of $p = 2$, we obtain

$$\begin{aligned} S_2^S \psi &= \sum_{v \in V} \sum_{e \in E: e[1]=v} \frac{1}{|e|!} |(\nabla \psi)(e)|^2 \\ &= \sum_{e \in E} \frac{1}{|e|!} |(\nabla \psi)(e)|^2 \\ &= S_2(\psi) \end{aligned} \quad (55)$$

Therefore, given Saito et al. (2018) has the same structure of different geometry, we can directly apply their proof to our setting in the case of $p = 2$.

Star Laplacian

The given hypergraph-gradient for star Laplacian can be written as

$$(\nabla \psi)(e) = \frac{w^{1/p}(e)}{(|e|)^{1/p}} \sum_{i=1}^{|e|} \left(\frac{\psi(v_i)}{\mu^{1/p}(v_i)} - \frac{\psi(v_1)}{\mu^{1/p}(v_1)} \right). \quad (56)$$

Here, we show that this hypergraph 2-Laplacian also can be seen from the same framework. Similarly to the clique Laplacians, the following proposition follows.

Proposition 25 *Let e be $\{v_1, \dots, v_{|e|}\} = e$. Then if we choose hypergraph-gradient operator $\nabla : \mathcal{H}(V) \rightarrow \mathcal{H}(E)$ for as*

$$(\nabla \psi)(e) := \frac{\sqrt{w(e)}}{\sqrt{|e|}} \sum_{i=1}^{|e|} \left(\frac{\psi(v_i)}{\sqrt{d(v_i)}} - \frac{\psi(v_1)}{\sqrt{d(v_1)}} \right), \quad (57)$$

this induces star expansion 2-Laplacian.

We can compute 2-Laplacian in the same manner as Saito et al. (2018), with a slight change of denominator of hypergraph-gradient from $\sqrt{|e| - 1}$ to $\sqrt{|e|}$. The 2-Laplacian induced from the hypergraph-gradient Eq. (57) can be computed as

$$L = I - D_v^{-1/2}W_s D_v^{-1/2}, \quad (58)$$

where W_s is a matrix whose element $w_s(u, v) = \sum_{u, v \in e} w(e) / |e|$. We can show that Eq. (58) satisfies the condition of Laplacian, in the same manner as the proof for Proposition 9 in Saito et al. (2018).

Total variation and submodular Laplacian

The hypergraph-gradient for total variation is written as

$$(\nabla\psi)(e) = w^{1/p}(e) \max_{v,u \in e} \left(\frac{\psi(u)}{\mu^{1/p}(u)} - \frac{\psi(v)}{\mu^{1/p}(v)} \right). \quad (59)$$

We show that total variation method in Hein et al. (2013) can be seen as a special case of our framework.

Proposition 26 *Let $\mu(v) = 1, \forall v \in V$. The Total Variation Regularizer defined as*

$$S_p(\psi) = w(e) \left(\max_{u,v \in e} \left| \frac{\psi(v)}{\mu^{1/p}(v)} - \frac{\psi(u)}{\mu^{1/p}(u)} \right| \right)^p \quad (60)$$

is p -Dirichlet Sum, if we choose hypergraph-gradient as Eq. (59).

This is obvious from the definition of p -Dirichlet energy by Eq. (8), which is called *regularizer* in Hein et al. (2013).

The hypergraph-gradient for SUB is for

$$\nabla\psi(e) = (w(e) \max_{S \subseteq e} (F(S)))^{1/p} \sum_{i=1}^{|e|} F(S_i) \left(\frac{\psi(u_{i+1})}{\mu^{1/p}(u_{i+1})} - \frac{\psi(u_i)}{\mu^{1/p}(u_i)} \right). \quad (61)$$

By definition, the energy can be written as

$$S_p(\psi) = \sum_{e \in \mathcal{E}} w(e) \max_{S \subseteq e} (F(S)) \left(\sum_{i=1}^{|e|} F(S_i) \left(\frac{\psi(u_{i+1})}{\mu^{1/p}(u_{i+1})} - \frac{\psi(u_i)}{\mu^{1/p}(u_i)} \right) \right)^p, \quad (62)$$

which is Eq. (15).

Proof of theorem 9

Most of the proof can be done in the similar manner as graph (Tudisco and Hein, 2016; Li and Milenkovic, 2018), while we need to change from the the graph p -Laplacian in (Tudisco and Hein, 2016) and hypergraph p -Laplacian in (Li and Milenkovic, 2018) to our framework p -Laplacian. We firstly denote by $\psi|_A(v)$ for a set $A \subset V$ as

$$\psi|_A(v) = \begin{cases} \psi(v) & v \in A \\ 0 & v \notin A \end{cases} \quad (63)$$

We start to prove the following lemma to prove Theorem 9.

Lemma 27 *For a set $A \subset V$,*

$$\left\langle \frac{\partial}{\partial \psi} \nabla\psi(e), \psi|_A \right\rangle = \nabla\psi|_A(e). \quad (64)$$

Proof Since hypergraph-gradient is a first degree polynomial of ψ from Definition 2, for $c_v \in \mathbb{R}$ we can write hypergraph-gradient as

$$\nabla\psi(e) = \sum_v c_v \psi(v). \quad (65)$$

By Eq. (65),

$$\left\langle \frac{\partial}{\partial\psi} \nabla\psi(e), \psi|_A \right\rangle = \sum_v c_{v,\psi} \psi|_A(v) = \nabla\psi_A(e), \quad (66)$$

which ends the proof. \square

We move to prove the following lemma.

Lemma 28 Denote λ, ψ be a p -eigenpair of p -Laplacian Δ_p . Let $A_1(\psi), \dots, A_m(\psi)$ is a nodal domains induced from $\psi \in H(V)$, and ψ' be the vector $\psi' \in F(\psi)$, where $F(\psi)$ is a nodal space induced from ψ . Then, $S(\psi') \leq \lambda \|\psi'\|_p^p$

Proof We consider the vector $f = \sum_{i=1}^m \alpha_i \psi|_{A_i}$, where α_i is a constant. From the definition of nodal domains, each edge e intersects at most two nodal domains with different signs. Therefore, $\psi|_{e \cap A_i} = \psi|_e$ for any $e \in E$ and for any nodal domain A_i , and $\psi|_{e \cap A_i} = \text{sgn}(\alpha_i) \psi|_e$ for any $e \in E$ and for any nodal domain A_i . We divide edges into two classes according to the number of nodal domains intersected by each edge as follows.

$$E_1 = \{e : |\{e \cap A_i\}| \leq 1\} \quad (67)$$

$$E_2 = \{e : |\{e \cap A_i\}| = 2\} \quad (68)$$

Note that $E_1 \cup E_2 = E$. Then, since $\nabla\psi(e) = \nabla\psi|_{A_i}(e)$ if $A_i \cap e = \emptyset$ and $\nabla\psi(e) = 0$ for those i such that $A_i \cap e = \emptyset$ and simpler version of Hölder's inequality ($\sum_{i=1}^n |x_i|^p \leq n^{p-1} \sum_{i=1}^n |x_i|^p$), we have

$$\begin{aligned} S_p(f) &= \sum_{e \in E} |\nabla f(e)|^p \\ &= \sum_{e \in E_1} \sum_i |\alpha_i|^p |\nabla\psi|_{A_i}(e)| |\nabla\psi(e)|^{p-1} + \sum_{e \in E_2} \left(\sum_i |\alpha_i| |\nabla\psi|_{A_i}(e) \right)^p \end{aligned} \quad (69)$$

Moreover, we have

$$\begin{aligned} \lambda \|f\|_p^p &= \sum_i |\alpha_i|^p \lambda \|\psi|_{A_i}\|_p^p \\ &= \sum_i |\alpha_i|^p \langle \psi|_{A_i}, \Delta_p \psi \rangle_{H(V)} \\ &= \sum_i |\alpha_i|^p \sum_v \psi|_{A_i}(v) \Delta_p \psi(v) \\ &= \sum_i |\alpha_i|^p \sum_v \psi|_{A_i}(v) \sum_{e:v \in e} \frac{1}{|e|!} |\nabla\psi(e)|^{p-1} \frac{\partial}{\partial\psi} \nabla\psi(e) \\ &= \sum_i |\alpha_i|^p \sum_v \sum_{e:v \in e} \frac{1}{|e|!} |\nabla\psi(e)|^{p-1} \frac{\partial}{\partial\psi} \nabla\psi(e) \psi|_{A_i}(v) \\ &= \sum_i |\alpha_i|^p \sum_e \nabla\psi|_{A_i}(e) |\nabla\psi(e)|^{p-1} \end{aligned} \quad (70)$$

From Eqs. (69) and (70), we have

$$\begin{aligned}
 S_p(f) - \lambda \|f\|_p^p &= \sum_{e \in E_2} \left(\left(\sum_i |\alpha| |\nabla \psi|_{A_i}(e) \right)^p - \left(\sum_i |\alpha_i|^p \sum_e |\nabla \psi|_{A_i}(e) |\nabla \psi(e)|^{p-1} \right) \right) \\
 &= \sum_{e \in E_2} \left(|\alpha_{i_1}| |\nabla \psi|_{A_{i_1}}(e) + |\alpha_{i_2}| |\nabla \psi|_{A_{i_2}}(e) \right)^p \\
 &\quad + \left(|\alpha_{i_1}|^p |\nabla \psi|_{A_{i_1}}(e) + |\alpha_{i_2}|^p |\nabla \psi|_{A_{i_2}}(e) \right) \\
 &\quad \times \left(|\nabla \psi|_{A_{i_1}} + |\nabla \psi|_{A_{i_2}} \right) \geq 0.
 \end{aligned}
 \tag{71}$$

The last equality follows from Lemma 3.7 in Tudisco and Hein (2016). □

Now we prove Theorem 9. Suppose that λ_k has multiplicity r and associated eigenvector ψ . As functions $\psi|_{A_1}, \dots, \psi|_{A_m}$ are linear independent of the definition of a nodal domain, $\gamma(F \cap S_p) \leq m$. From Lemma 28, $\forall \psi' \in F \cap S_p$ we have $R_p(\psi') \leq R_p(\psi) = \lambda_k$. Also, $F \cap S_p \in \mathcal{F}_m(S_p)$. Hence,

$$\lambda_m = \min_{A \in \mathcal{F}_m(S_p)} \max_{\psi' \in A} R_p(\psi') \leq \max_{\psi \in F \cap S_p} R_p(\psi') \leq \lambda_k.
 \tag{72}$$

From Eq. (72) $\lambda_m \geq \lambda_k$ and $m \geq k + r - 1$. This concludes the proof.

Proof of theorem 10

We begin our discussion by the following lemma.

Lemma 29 *Let $\mathcal{A} = \text{Span}(\mathbf{1}_{V_1}, \dots, \mathbf{1}_{V_k})$. Choose a vector $\psi \in \mathcal{A} \cap S_p$ and suppose that it can be written as $\|\psi\|_p^p = 1$ and $\psi = \sum_i \alpha_i M^{1/p} \mathbf{1}_{V_i}$. Then,*

$$\sum_i |\alpha_i|^p \text{vol}(V_i) = 1.
 \tag{73}$$

Proof As \mathcal{A} is a k -way partition, $V_i \cap V_j = \emptyset$. Therefore, we obtain

$$\begin{aligned}
 1 &= \|\psi\|_p^p \\
 &= \sum_i \|\alpha_i M^{1/p} \mathbf{1}_{V_i}\|_p^p \\
 &= \sum_i |\alpha_i|^p \text{vol}(V_i)
 \end{aligned}
 \tag{74}$$

□

Firstly, we prove the upper bound of the inequality. Without loss of generality, we limit our interest to $\|\psi\|_p^p = 1$. If we set $\psi = \sum_i \alpha_i M^{1/p} \mathbf{1}_{V_i}$, then we obtain

$$\begin{aligned}
R_p(\psi) &= \sum_e \|\nabla\psi\|_p^p \\
&= \sum_e \sum_i \alpha_i^p \|\nabla(M^{1/p}\mathbf{1}_{V_i})\|_p^p \\
&\leq \sum_e \min(k, |e|)^{p-1} \sum_i |\alpha|^p \|\nabla(M^{1/p}\mathbf{1}_{V_i})\|_p^p \\
&\leq \min(\max |e|, k) \sum_i |\alpha|^p \sum_e \|\nabla(M^{1/p}\mathbf{1}_{V_i})\|_p^p
\end{aligned} \tag{75}$$

From $\|\nabla M^{1/p}\mathbf{1}_{V_i}\|_p^p \leq \sum_{e \in V_i, \bar{V}_i} c(e)w(e) = \partial\text{vol}(V_i, \bar{V}_i)$ and Lemma 29, Eq. (75) can be tightened as

$$\begin{aligned}
R_p(\psi) &\leq \min(\max |e|, k) \sum_i |\alpha|^p \partial\text{vol}(V_i, \bar{V}_i) \\
&\leq \min(\max |e|, k) \frac{\sum_i |\alpha|^p \partial\text{vol}(V_i, \bar{V}_i)}{\sum_i |\alpha|^p \text{vol}(V_i)} \\
&\leq \min(\max |e|, k) h_k.
\end{aligned} \tag{76}$$

Next, we prove the lower bound. The proof of the lower bound depends on the following lemma.

Lemma 30 For any vectors $\psi \in H(V)$, and $p \geq 1$, there exists $\theta \geq 0$ such that $\Theta(x, \theta) = \{u : \psi(u) > \theta\}$ satisfies

$$R_p(M^{1/p}\psi) \geq (\rho)^{p-1} \left(\frac{c(V)}{p} \right)^p, \tag{77}$$

where $\rho := \max_v d(v)/\mu(v)$.

Proof We denote by ψ^p element-wise p -th power. Then, we obtain

$$\begin{aligned}
S_1(M\psi^p) &= \sum_e \|\nabla(M\psi^p)\| \\
&= \sum_e \sum_{u,v \in e} w(e)c(u, v, e, \psi) |\psi^p(u) - \psi^p(v)| \\
&\leq \sum_e \sum_{u,v} w(e)pc(u, v, e, \psi) |\psi(u) - \psi(v)| \\
&\quad \times \max(\psi(u), \psi(v))^{p-1} \\
&\leq p \left(\left(\sum_e \sum_{u,v} w(e)pc(u, v, e, \psi) |\psi(u) - \psi(v)| \right)^p \right)^{\frac{1}{p}} \\
&\quad \times \left(\sum_e \sum_{u,v \in e} \max(\psi(u), \psi(v))^p \right)^{\frac{p-1}{p}} \\
&\leq p S_p^{\frac{1}{p}}(M^{1/p}\psi) \left(\sum_v d_v \psi^p(v) \right)^{\frac{p-1}{p}} \\
&\leq p \rho^{1-1/p} R_p^{\frac{1}{p}}(\psi) \|M^{1/p}\psi\|_p^{p-1}
\end{aligned} \tag{78}$$

Moreover, we get

$$\begin{aligned}
 S_1(M\psi^p) &= \sum_{e \in E} \sum_{u, v \in e} w(e)c(u, v, e, \psi)|\psi^p(u) - \psi^p(v)| \\
 &= \sum_{e \in E} \sum_{u, v \in e; \psi^p(u) - \psi^p(v) > 0} w(e)c(u, v, e, \psi) \int_{\psi(v)}^{\psi(u)} d\theta \\
 &= \int_0^\infty \sum_{e \in E} w(e) \sum_{u, v \in e; \psi^p(u) - \psi^p(v) > \theta} c(u, v, e, \psi) d\theta.
 \end{aligned} \tag{79}$$

From Eq.(79), the left hand side of Eq. (77) can be rewritten as

$$\begin{aligned}
 &\frac{S_1(M\psi^p)}{\|M\psi\|_p^p} \\
 &= \frac{\int_0^\infty \sum_{e \in E} \sum_{u, v \in e; \psi^p(u) - \psi^p(v) > \theta} w(e)c(u, v, e, \psi) d\theta}{\int_0^\infty \sum_{\psi^p(v) > \theta} \mu(v : \psi^p(v) > \theta) d\theta} \\
 &\geq \inf_{\theta > 0} \frac{\sum_{e \in E} w(e) \sum_{u, v \in e; \psi^p(u) - \psi^p(v) > \theta} c(u, v, e, \psi)}{\mu(v : \psi^p(v) > \theta)} \\
 &= \inf_{\theta > 0} \frac{\partial \text{vol}(\{v : \psi^p(v) > \theta\}, \{v : \psi^p(v) > \theta\})}{\text{vol}(\{v : \psi^p(v) > \theta\})} \\
 &= \inf_{\theta > 0} C(\{v : \psi^p(v) > \theta\})
 \end{aligned} \tag{80}$$

Hence, the following inequality holds for the set $\Theta^* = \{v : \psi^p(v) > \theta^*\}$ where θ^* is the minimizer.

$$\begin{aligned}
 R_p(M^{1/p}\psi) &= \frac{S_p(M^{1/p}\psi)}{\|M^{1/p}\psi\|_p^p} \\
 &\geq \left(\frac{S_1(M\psi^p)}{\|M^{1/p}\psi\|_p^p} \frac{1}{p^p \rho^{p-1}} \right) \\
 &= \left(\frac{1}{\rho} \right)^{p-1} \left(\frac{c(\Theta^*)}{p} \right)^p
 \end{aligned} \tag{81}$$

This concludes Lemma 30. □

Suppose λ_k has a corresponding eigenvector ψ that induces the strong nodal domains A_1, A_2, \dots, A_m . From Lemma 28, we have $\lambda \leq R_p(\mathbf{M}^{1/p}\mathbf{1}_{A_i})$. Moreover, from Lemma 30, for any $i(i = 1, \dots, m)$, there exists a set $B_i \subset A_i$ such that

$$R_p(M^{1/p}\mathbf{1}_{A_i}) \leq \left(\frac{1}{\rho} \right)^{p-1} \left(\frac{c(B_i)}{p} \right)^p \tag{82}$$

Therefore,

$$\begin{aligned}
\lambda_k &\geq R_p(M^{1/p} \mathbf{1}_{A_i}) \\
&\geq \max_i \left(\frac{1}{\rho} \right)^{p-1} \left(\frac{c(B_i)}{p} \right)^p \\
&\geq \min_{B_i} \max_i \left(\frac{1}{\rho} \right)^{p-1} \left(\frac{c(B_i)}{p} \right)^p \\
&\geq \left(\frac{1}{\rho} \right)^{p-1} \left(\frac{h_m}{p} \right)^p
\end{aligned} \tag{83}$$

Proof of corollary 11

Let us start with the following lemma.

Lemma 31 *Let $\psi \in H(V)$ be orthogonal to $\mathbf{1}$. Then there is $\psi' \in H(V)$ and $\forall v \in V \psi'(v) \geq 0$ with at most $|V|/2$ non-zero entries such that*

$$R_p(M^{1/p} \psi') \leq R_p(M^{1/p} \psi). \tag{84}$$

Moreover, $\forall t$ satisfying $0 < t \leq \max_v \psi'(v)$, the set $B = \{v : \psi'(v) \geq t\}$ is one of the set obtained by ψ , such that $(\{v : \psi(v) \geq t\}, \{v : \psi(v) < t\})$ minimizing Cheeger cut.

Proof Firstly, we observe that

$$R(M^{1/p}(\psi + c\mathbf{1})) \geq R(\psi), \tag{85}$$

since

$$S_p(M^{1/p}(\psi + c\mathbf{1})) = S_p(M^{1/p}\psi) \tag{86}$$

and

$$\|M^{1/p}(\psi + c\mathbf{1})\| = \|M^{1/p}(\psi)\| + \|cM^{1/p}\mathbf{1}\| \geq \|M^{1/p}\psi\|.$$

Let m be the median value of ψ , and set $\psi_m := \psi - m\mathbf{1}$. Then $R(\psi_m) < R(\psi)$, and median of ψ_m is zero, which means ψ_m has at most $|V|/2$ positive entities and at most $|V|/2$ negative entities. We decompose ψ_m as follows;

$$\psi_m = \psi_{m+} - \psi_{m-}, \tag{86}$$

where $\psi_{m+}(v) = \psi_m(v)$ if $\psi_m(v)$ is positive and otherwise set 0, and $\psi_{m-}(v) = -\psi_m(v)$ if $\psi_m(v)$ is negative and otherwise set 0. We remark that ψ_{m+} and ψ_{m-} are non-negative, orthogonal to each other, and have at most $|V|/2$ non-zero entities. The cut defined by the set $\{v : \psi_{m+} \geq t\}$, $\forall t$ is one of the cut obtained by ψ , such that $(\{v : \psi(v) \geq t\}, \{v : \psi(v) < t\})$ minimizing Cheeger cut, since we can obtain the same cut by considering

$$(\{v : \psi(v) \geq t + m\}, \{v : \psi(v) < t + m\}). \tag{87}$$

Similarly, cut defined by the set $\{v : \psi_{m-} \geq t\}$, $\forall t$ is one of the cut obtained by ψ , such that $(\{v : \psi(v) \geq t\}, \{v : \psi(v) < t\})$ minimizing Cheeger cut.

We move on to show that at least one of ψ_{m+} or ψ_{m-} has Rayleigh quotient equal to or smaller than Rayleigh quotient of ψ_m , by showing the following

$$\begin{aligned}
 R_p(M^{1/p}\psi_m) &= \frac{S_p(M^{1/p}\psi_m)}{\|M^{1/p}\psi_m\|^p} \\
 &\geq \frac{S_p(M^{1/p}\psi_{m+}) + S_p(M^{1/p}\psi_{m-})}{\|M^{1/p}\psi_{m+}\|^p + \|M^{1/p}\psi_{m-}\|^p} \\
 &= \frac{R_p(M^{1/p}\psi_{m+})\|M^{1/p}\psi_{m+}\| + R_p(M^{1/p}\psi_{m-})\|M^{1/p}\psi_{m-}\|}{\|M^{1/p}\psi_{m+}\|^p + \|M^{1/p}\psi_{m-}\|^p} \\
 &\geq \min(R_p(M^{1/p}\psi_{m+}), R_p(M^{1/p}\psi_{m-}))
 \end{aligned}
 \tag{88}$$

This concludes the proof. □

By combining lemma 30 and lemma 31, we can say a stronger statement than Corollary 11.

Corollary 32 *Let $\psi \in H(V)$ be orthogonal to $M^{1/p}\mathbf{1}$, and let (B, \bar{B}) be the cut found by ψ , such that $(\{v : \psi(v) \geq t\}, \{v : \psi(v) < t\})$ minimizing Cheeger cut. Then*

$$R_p(M^{1/p}\psi) \geq (\rho)^{p-1} \left(\frac{c(B)}{p} \right)^p
 \tag{89}$$

This concludes the proof of Corollary 11.

Proof of theorem 13

By definition, for all $v \in V$

$$\Delta_p \psi(v) = \lambda^\psi \xi_p(\psi(v))
 \tag{90}$$

$$\Delta_p \omega(v) = \lambda^\omega \xi_p(\omega(v)),
 \tag{91}$$

where λ_ψ and λ_ω are eigenvalues associated with ψ and ω , respectively. Then, for all $v \in V$ we obtain

$$\Delta_p \psi(v) \xi_p(\omega(v)) = \lambda^\psi \xi_p(\psi(v)) \xi_p(\omega(v))
 \tag{92}$$

$$\Delta_p \omega(v) \xi_p(\psi(v)) = \lambda^\omega \xi_p(\omega(v)) \xi_p(\psi(v)).
 \tag{93}$$

By summing up over all $v \in V$ and taking difference of both side of Eqs. (92) and (93), we compute

$$\begin{aligned}
 &(\lambda^\psi - \lambda^\omega) \sum_{v \in V} \xi_p(\psi(v)) \xi_p(\omega(v)) \\
 &= \sum_{v \in V} (\Delta_p \psi(v) \xi_p(\omega) - \Delta_p \omega(v) \xi_p(\psi(v)))
 \end{aligned}
 \tag{94}$$

By applying Taylor expansion at $a_{\Delta_p \psi} M^{1/p} \mathbf{1}$ to $\Delta_p \psi$, at $a_{\Delta_p \omega} M^{1/p} \mathbf{1}$ to $\Delta_p \omega$ at $a_{\xi(\psi)}$ to $\xi_p(\psi)$, and at $a_{\xi(\omega)}$ to $\xi_p(\omega)$ in right hand side of Eq. (94), we obtain

$$\begin{aligned}
& (\lambda^\psi - \lambda^\omega) \sum_{v \in V} \xi_p(\psi(v)) \xi_p(\omega(v)) = \\
& \sum_{v \in V} (\Delta_p(a_{\Delta_p \psi} M^{1/p} \mathbf{1})(v) \\
& + \Delta'_p(a_{\Delta_p \psi} M^{1/p} \mathbf{1})(v)(\psi(v) - a_{\Delta_p \psi} \mu^{1/p}(v)) + o_{2, \Delta_p \psi}) \\
& \times (\xi_p(a_{\xi(\omega)}) + \xi'_p(a_{\xi(\omega)})(\psi(v) - a_{\xi_p(\omega)}) + o_{2, \xi_p(\omega)}) \\
& - \sum_{v \in V} (\Delta_p(a_{\Delta_p \omega} M^{1/p} \mathbf{1})(v) \\
& + \Delta'_p(a_{\Delta_p \omega} M^{1/p} \mathbf{1})(v)(\psi(v) - c_{\Delta_p \omega}) + o_{2, \Delta_p \omega}) \\
& \times (\xi_p(a_{\xi(\psi)}) + \xi'_p(a_{\xi(\psi)})(\psi(v) - a_{\xi_p(\psi)}) + o_{2, \xi_p(\psi)})
\end{aligned} \tag{95}$$

By Proposition 7 and Corollary 17, Eq. (95) is

$$\begin{aligned}
& (\lambda_\psi - \lambda_\omega) \sum_{v \in V} \xi_p(\psi(v)) \xi_p(\omega(v)) \\
& = o_2,
\end{aligned}$$

which concludes the proof.

Proof of theorem 14

Before we start a proof, let us motivate the discussion by considering $p = 2$. Let ψ_1 be a first eigenvector either $M^{1/p} \mathbf{1}$, then $\langle \psi_1, \psi_2 \rangle = 0$ for the second eigenvector ψ_2 , and we observe

$$\|\psi_2\|_2^2 = \|\psi_2 - (\langle \psi_1, \psi_2 \rangle / |\psi_1|) \psi_1\|_2^2 = \min_{\eta \in \mathbb{R}} \|\psi_2 - \eta \psi_1\|_2^2$$

Using this equation, the Rayleigh quotient to get the second eigenvector ψ_2 can be written as

$$\psi_2 = \operatorname{argmin}_{\psi \in H(V)} \frac{S_2(\psi)}{\min_{\eta} \|\psi - \eta \psi_1\|_2^2}. \tag{96}$$

This inspires Eq. 19.

Let us start our proof by proving $\lambda_2 \geq \inf R_p(\psi)$. Let ψ_2 be a p -eigenvector corresponding to the second p -eigenvalue λ_2 . As $\sum_v \xi_p(\psi_2(v)) = \mathbf{1}^\top \delta \psi_2 = 0$, the norm $\|\psi_2 - c \mathbf{1}\|$ is convex in c is minimized when $c = 0$. Moreover, $\lambda_2 = R_p(\psi_2) = S_p(\psi_2) / \|\psi_2\|_p^p = S_p(\psi_2) / \min_c \|\psi_2 - c\|_p^p = R_p^{(2)}(\psi_2)$. Hence, $\lambda_2 \geq \inf R_p^{(2)}(\psi_2)$.

Second, we prove $\lambda_2 \leq \inf R_p(\psi)$. From the definition of $R_p^{(2)}$, we can easily check that $R_p^{(2)}(a\psi + b) = R_p^{(2)}(\psi)$. Let ψ^* be $\psi^* = \operatorname{argmin}_p R_p^{(2)}(\psi)$, and consider the space $A = \{(a\psi^* + b)\}$. As $\psi^* \neq c \mathbf{1}$, $\gamma(A \cap \mathcal{S}_p) = 2$. From Proposition 7, we obtain

$$\begin{aligned}
 \lambda_2 &\leq \max_{\psi \in A \cap S_p} S_p(\psi) \\
 &= \max_{a,b} \frac{S_p(a\psi^* + b)}{\|a\psi^* + b\|_p^p} \\
 &= \max_{a,b} \frac{S_p(a\psi^*)}{\|a\psi^* + b\|_p^p} \\
 &= \max_a \frac{S_p(a\psi^*)}{\min_b \|a\psi^* + b\|_p^p} \\
 &= R_p^{(2)}(\psi^*)
 \end{aligned}
 \tag{97}$$

As we have $\lambda_2 \geq \inf R_p(\psi)$ and $\lambda_2 \leq \inf R_p(\psi)$, we obtain $\lambda_2 = R_p(\psi)$.
 Since the global minimum is λ_2 , then

Table 5 The detailed results of hypergraph partitioning for various p

		best	$p = 1.1$	$p = 1.5$	$p = 2$	$p = 2.5$	$p = 3.0$
mushroom	CLIQUE E-N	0.1188 ($p = 1.1$)	0.1188	0.4719	0.4791	0.4796	0.4796
	CLIQUE E-UN	0.1211 ($p = 1.1$)	0.1211	0.1967	0.1460	0.4793	0.4788
	STAR	0.1113 ($p = 1.1$)	0.1113	0.1278	0.2189	0.4796	0.4791
	TV V-UN	0.1083 ($p = 1.1$)	0.1083	0.4796	0.4796	0.4796	0.4798
	TV V-N	0.1083 ($p = 1.1$)	0.1083	0.4798	0.4796	0.4446	0.4796
	cancer	CLIQUE E-N	0.0243 ($p = 1.7$)	0.0272	0.0253	0.0286	0.1288
CLIQUE E-UN		0.0358 ($p = 2.2$)	0.0801	0.0930	0.2117	0.3462	0.3433
STAR		0.0286 ($p = 1.1$)	0.0286	0.0286	0.0401	0.3433	0.3448
TV V-UN		0.1960 ($p = 1.1$)	0.1960	0.2632	0.2546	0.3419	0.3433
TV V-N		0.2489 ($p = 1.6$)	0.4612	0.3090	0.2661	0.3433	0.3462
chess		CLIQUE E-N	0.4765 ($p = 2.1$)	0.4768	0.4981	0.4765	0.4778
	CLIQUE E-UN	0.4768 ($p = 2.2$)	0.4950	0.4950	0.4772	0.4775	0.4775
	STAR	0.4759 ($p = 1.3$)	0.4865	0.4781	0.4781	0.4781	0.4778
	TV V-UN	0.4659 ($p = 2.6$)	0.4966	0.4740	0.4775	0.4778	0.4784
	TV V-N	0.4643 ($p = 2.7$)	0.4912	0.4734	0.4778	0.4775	0.4671
	congress	CLIQUE E-N	0.1172 ($p = 1.5$)	0.1264	0.1172	0.1241	0.1448
CLIQUE E-UN		0.1195 ($p = 1.5$)	0.1218	0.1287	0.1379	0.1241	0.3885
STAR		0.1195 ($p = 2.3$)	0.1287	0.1310	0.1333	0.1241	0.1333
TV V-UN		0.1862 ($p = 1.2$)	0.1839	0.2598	0.2828	0.3218	0.3241
TV V-N		0.1839 ($p = 1.5$)	0.1908	0.1839	0.2736	0.2966	0.2529
news(1,2)		CLIQUE E-N	0.2390 ($p = 1.1$)	0.2390	0.2587	0.2906	0.2400
	CLIQUE E-UN	0.2493 ($p = 1.7$)	0.2507	0.2516	0.2511	0.4330	0.4330
	STAR	0.2411 ($p = 2.4$)	0.2477	0.4324	0.2477	0.4323	0.4330
	TV V-UN	0.2701 ($p = 1.1$)	0.2701	0.4356	0.3015	0.3016	0.3017
	TV V-N	0.2672 ($p = 1.5$)	0.2672	0.4334	0.3029	0.3015	0.3028
	news(3,4)	CLIQUE E-N	0.2146 ($p = 1.1$)	0.2146	0.2167	0.2173	0.2186
CLIQUE E-UN		0.3272 ($p = 1.1$)	0.3272	0.3507	0.3447	0.3481	0.3474
STAR		0.2162 ($p = 1.7$)	0.2183	0.2356	0.2169	0.2170	0.2170
TV V-UN		0.2701 ($p = 1.4$)	0.4378	0.3274	0.3274	0.4305	0.4340
TV V-N		0.3275 ($p = 1.4$)	0.4544	0.3274	0.3274	0.4340	0.4305

$$\begin{aligned}
\lambda_2 &= R_p(\psi_2) \\
&= \frac{S_p(\psi_2)}{\|\psi_2\|_p^p} \\
&= \frac{S_p(\psi^* + \eta^* \psi_1)}{\min_{\eta} \|\psi^* - \eta \psi_1\|} \\
&= \frac{S_p(\psi_2)}{\min_{\eta} \|\psi - \eta \psi_1\|} \\
&= R_p^{(2)}(\psi^*).
\end{aligned} \tag{98}$$

Note that we use $S_p(\psi + \psi_1) = S_p(\psi)$. Therefore, we can see $\psi^* = \psi_2 + \eta^* \psi_1$, where $\eta = \operatorname{argmin}_{\eta} \|\psi_2 - \eta \psi_1\|_p^p$.

Additional experimental results

We first mention that our experiment was run on mac mini with Intel i7 and 32GiB RAM. Table 5 shows the detailed results of the hypergraph partitioning experiment. This is a further evidence of a trade-off between Cheeger inequality and natural intuition of p from Eq. (8), discussed in Sect. 7. However, the difference of the performance between the best p and the others varies; sometimes the contribution of p is small while sometimes p makes the large difference. Also, you can see the further evidence that E-N-VW outperforms the other method in chess, while looking at the other datasets, there we see on the large δ dataset (mushroom) that all TV methods outperform STAR and clique methods whereas for smaller δ all STAR and clique methods outperform all TV methods.

Author contributions SS conceived the framework, conducted theoretical analyses, and carried out the experiments. SS and MH wrote the manuscript.

Funding This research has been made possible thanks partially to Huawei, who are generously supporting Shota Saito's PhD study at University College London.

Data availability We used publicly available data, and we cite them appropriately.

Code availability Part of our code depends on the others' packages, which are not available to publish due to copyright reasons. We will endeavor to reduce the dependency on these restricted packages by the time of publication. Also, full code is available upon request to the corresponding author.

Declarations

Conflict of interest The authors have declared that no competing interests exist.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Agarwal, S., Branson, K., Belongie, S. (2006). Higher order learning with graphs. *Proc. ICML* (pp. 17–24).
- Alamgir, M., & Luxburg, U.V. (2011). Phase transition in the family of p -resistances. In: *Proceedings NIPS* (pp. 379–387).
- Alon, N. (1986). Eigenvalues and expanders. *Combinatorica*, 6(2), 83–96.
- Alon, N., & Milman, V. D. (1985). λ_1 , isoperimetric inequalities for graphs and superconcentrators. *J. Comb. Theory Series B.*, 38(1), 73–88.
- Amari, S.-I. (1998). Natural gradient works efficiently in learning. *Neural Comput.*, 10(2), 251–276.
- Belkin, M., & Niyogi, P. (2003). Laplacian eigenmaps for dimensionality reduction and data representation. *Neural Computer*, 15(6), 1373–1396.
- Benson, A.R., Kleinberg, J., Veldt, N. (2020). Augmented sparsifiers for generalized hypergraph cuts. arXiv preprint [arXiv:2007.08075](https://arxiv.org/abs/2007.08075).
- Berge, C. (1984). *Hypergraphs: combinatorics of finite sets* (Vol. 45). Elsevier.
- Binding, P. A., & Rynne, B. P. (2008). Variational and non-variational eigenvalues of the p -Laplacian. *Journal of Differential Equations*, 244(1), 24–39.
- Bolla, M. (1993). Spectra, euclidean representations and clusterings of hypergraphs. *Discrete Mathematics*, 117(1–3), 19–39.
- Bougleux, S., Elmoataz, A., & Melkemi, M. (2009). Local and nonlocal discrete regularization on weighted graphs for image and mesh processing. *International Journal of Computer Vision*, 84(2), 220–236.
- Brin, S., & Page, L. (1998). The anatomy of a large-scale hypertextual web search engine. *Computer Network ISDN*, 30(1–7), 107–117.
- Bühler, T., & Hein, M. (2009). Spectral clustering based on the graph p -Laplacian. In: *Proceedings of ICML* pp 81–88.
- Chan, T.-H.H., Louis, A., Tang, Z. G., & Zhang, C. (2018). Spectral properties of hypergraph laplacian and approximation algorithms. *Journal of the ACM*, 6(5), 31–48.
- Chang, J., Chen, Y., Qi, L., & Yan, H. (2020). Hypergraph clustering using a new laplacian tensor with applications in image processing. *SIAM Journal on Imaging Sciences*, 13(3), 1157–1178.
- Chang, K. C. (2016). Spectrum of the 1-Laplacian and cheeger's constant on graphs. *Journal of Graph Theory*, 81(2), 167–207.
- Chen, Y., Qi, L., & Zhang, X. (2017). The fiedler vector of a laplacian tensor for hypergraph partitioning. *SIAM Journal on Science Computer*, 39(6), A2508–A2537.
- Chung, F. (2007). Four proofs for the cheeger inequality and graph partition algorithms. In: *Proceeding of ICCM* (Vol. 2, p. 751–772).
- Cooper, J., & Dutle, A. (2012). Spectra of uniform hypergraphs. *Linear Algebra and its Applications*, 436(9), 3268–3292.
- Courant, R., & Hilbert, D. (1962). *Methods of mathematical physics*. Interscience Publishers.
- Dua, D., & Graff, C. (2022). *UCI machine learning repository*. Retrieved from <http://archive.ics.uci.edu/ml>
- Fiedler, M. (1973). Algebraic connectivity of graphs. *Czechoslovak Mathematical Journal*, 23(2), 298–305.
- Fiedler, M. (1975). A property of eigenvectors of nonnegative symmetric matrices and its application to graph theory. *Czechoslovak Mathematical Journal*, 25(4), 619–633.
- Gharahighi, A., Vens, C., & Pliakos, K. (2021). Fair multi-stakeholder news recommender system with hypergraph ranking. *Information Processing & Management*, 58(5), 102663.
- Ghoshdastidar, D., & Dukkipati, A. (2017). Consistency of spectral hypergraph partitioning under planted partition model. *The Annals of Statistics*, 45(1), 289–315.
- Ghoshdastidar, D., & Dukkipati, A. (2017). Uniform hypergraph partitioning: Provable tensor methods and sampling techniques. *The Journal of Machine Learning Research*, 18(1), 1638–1678.
- Gibson, D., Kleinberg, J., & Raghavan, P. (2000). Clustering categorical data: An approach based on dynamical systems. *The VLDB Journal*, 8(3), 222–236.
- Grady, L. (2006). Random walks for image segmentation. *Transactions on Pattern Analysis and Machine Intelligence*, 28(11), 1768–1783.
- Hagen, L., & Kahng, A. B. (1992). New spectral methods for ratio cut partitioning and clustering. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 11(9), 1074–1085.
- Hein, M., Setzer, S., Jost, L., Rangapuram, S.S. (2013). The total variation on hypergraphs - learning on hypergraphs revisited. In: *Proceedings NIPS* pp. 2427–2435.

- Herbster, M., & Lever, G. (2009). *Predicting the labelling of a graph via minimum- p -seminorm interpolation*. COLT: Proceedings.
- Hu, S., & Qi, L. (2012). Algebraic connectivity of an even uniform hypergraph. *Journal of Combinatorial Optimization*, 24(4), 564–579.
- Hu, S., & Qi, L. (2015). The Laplacian of a uniform hypergraph. *Journal of Combinatorial Optimization*, 29(2), 331–366.
- Huang, Y., Liu, Q., Metaxas, D. (2009). Video object segmentation by hypergraph cut. In: *Proceeding CVPR* pp. 1738–1745.
- Klamt, S., Haus, U.-U., & Theis, F. (2009). Hypergraphs and Cellular Networks. *PLoS Computer Biology*, 5(5), e1000385.
- Lee, J. R., Gharan, S. O., & Trevisan, L. (2014). Multiway spectral partitioning and higher-order cheeger inequalities. *Journal of the ACM*, 6, 1637.
- Li, P., & Milenkovic, O. (2017). Inhomogeneous hypergraph clustering with applications. In: *Proceedings NIPS* pp. 2305–2315.
- Li, P., & Milenkovic, O. (2018). Submodular hypergraphs: p -Laplacians, cheeger inequalities and spectral clustering. In: *Proceedings ICML* pp. 3020–3029.
- Li, W.-C.W., & Solé, P. (1996). Spectra of regular graphs and hypergraphs and orthogonal polynomials. *European Journal of Combinatorics*, 17(5), 461–477.
- Lindqvist, P. (2008). A nonlinear eigenvalue problem. *Topics in mathematical analysis* (pp 175–203). World Scientific.
- Luo, D., Huang, H., Ding, C., & Nie, F. (2010). On the eigenvectors of p -Laplacian. *Machine Learning*, 81(1), 37–51.
- Mobasher, B., Cooley, R., & Srivastava, J. (2000). Automatic personalization based on web usage mining. *Communication of ACM*, 43(8), 142–151.
- Newman, M. E. (2006). Modularity and community structure in networks. *Proceedings of the National Academy of Sciences*, 103(23), 8577–8582.
- Qi, L. (2013). h^+ -eigenvalues of Laplacian and signless Laplacian tensors. arXiv preprint [arXiv:1303.2186](https://arxiv.org/abs/1303.2186).
- Rodriguez, J. A. (2002). On the Laplacian eigenvalues and metric parameters of hypergraphs. *Linear and Multilinear Algebra*, 50(1), 1–4.
- Saito, S. (2022). *Hypergraph modeling via spectral embedding connection: Hypergraph cut, weighted kernel-means, and heat kernel*. AAAI (to appear): Proceedings.
- Saito, S., Mandic, D.P., Suzuki, H. (2018). Hypergraph p -Laplacian: A differential geometry view. *Proceeding AAAI* pp. 3984–3991.
- Shi, J., & Malik, J. (1997). Normalized cuts and image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2, 888–905.
- Slepcev, D., & Thorpe, M. (2019). Analysis of p -laplacian regularization in semisupervised learning. *SIAM Journal on Mathematical Analysis*, 51(3), 2085–2120.
- Struwe, M. (2000). *Variational methods: Applications to nonlinear partial differential equations and hamiltonian systems, third edition*. Springer.
- Tudisco, F., & Hein, M. (2016). A nodal domain theorem and a higher-order Cheeger inequality for the graph p -Laplacian. [arXiv:1602.05567](https://arxiv.org/abs/1602.05567).
- Veldt, N., Benson, A.R., Kleinberg, J. (2020). Hypergraph cuts with general splitting functions. arXiv preprint [arXiv:2001.02817](https://arxiv.org/abs/2001.02817).
- von Luxburg, U. (2007). A tutorial on spectral clustering. *Statistics and computing*, 17(4), 395–416.
- Yoshida, Y. (2019). Cheeger inequalities for submodular transformations. In: *Proceedings SODA* pp. 2582–2601.
- Zhou, D., Huang, J., Schölkopf, B. (2006). Learning with hypergraphs: Clustering, classification, and embedding. In: *Proceeding NIPS* pp. 1601–1608.
- Zhou, D., & Schölkopf, B. (2005). Regularization on discrete spaces. *Pattern recognition* (pp. 361–368). Springer.
- Zien, J. Y., Schlag, M. D. F., & Chan, P. K. (1999). Multilevel spectral hypergraph partitioning with arbitrary vertex sizes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(9), 1389–1399.