Multi-class Graph Clustering via Approximated Effective p-Resistance

Shota Saito [1] **Mark Herbster** [1]

Abstract

This paper develops an approximation to the (effective) p-resistance and applies it to multi-class clustering. Spectral methods based on the graph Laplacian and its generalization to the graph p-Laplacian have been a backbone of non-euclidean clustering techniques. The advantage of the p-Laplacian is that the parameter p induces a controllable bias on cluster structure. The drawback of p-Laplacian eigenvector based methods is that the third and higher eigenvectors are difficult to compute. Thus, instead, we are motivated to use the p-resistance induced by the p-Laplacian for clustering. For p-resistance, small p biases towards clusters with high internal connectivity while large p biases towards clusters of small "extent," that is a preference for smaller shortest-path distances between vertices in the cluster. However, the p-resistance is expensive to compute. We overcome this by developing an approximation to the p-resistance. We prove upper and lower bounds on this approximation and observe that it is exact when the graph is a tree. We also provide theoretical justification for the use of p-resistance for clustering. Finally, we provide experiments comparing our approximated p-resistance clustering to other p-Laplacian based methods.

1. Introduction

Graphs are widely used data structures representing a pairwise relationship (Shi & Malik, 1997). In machine learning, various graph methods are considered, such as clustering and semi-supervised learning (von Luxburg, 2007; Zhu et al., 2003). Common to these methods, graph 2-seminorm, 2-seminorm induced from the graph Laplacian, is actively used. Its generalization to the graph p-seminorm is known to exhibit performance improvement (Bühler & Hein, 2009; Slepcev & Thorpe, 2019).

This paper considers multi-class clustering over a graph using the graph p-seminorm. For this purpose, spectral clustering is the most popular. In the 2-seminorm based (i.e., standard) spectral clustering, we use the first k eigenvectors of the graph Laplacian for k-class clustering (von Luxburg, 2007). This use of the first k eigenvectors is theoretically supported (Lee et al., 2014). Using the p-seminorm, this graph Laplacian is extended to the graph p-Laplacian (Bühler & Hein, 2009). Similar to the standard case, using the first k eigenvectors of this graph p-Laplacian for k-class clustering is also theoretically supported (Tudisco & Hein, 2018). However, there is not yet known an exact identification for the third or higher eigenpairs of p-Laplacian (Lindqvist, 2008), and hence in practice, it is difficult to obtain them. Due to this limitation, the existing methods using p-Laplacian propose an ad-hoc resolution of this limitation for multi-class clustering (Bühler & Hein, 2009; Ding et al., 2019; Luo et al., 2010). On the other hand, this limitation makes the p-Laplacian difficult to use in practice to leverage the full potential of graph p-seminorm for multi-class clustering purposes.

Thus, in order to aim to exploit the graph p-seminorm more for multi-class clustering, we explore an alternative way to spectral clustering; in this paper, we propose multi-class clustering via approximated effective p-resistance. The p-resistance is also induced by the graph p-seminorm. The use of p-resistance[1] for clustering is motivated in the following way. Looking back to the 2-seminorm case, the 2-resistance is considered in the context of the graph analog to electric circuit (Doyle & Snell, 1984). The 2-resistance is defined as an inverse of the constrained optimization problem using the graph 2-seminorm. This 2-resistance is known to be a metric over a graph (Klein & Randić, 1993). Moreover, 2-resistance is characterized by a semi-supervised learning problem of the graph 2-seminorm regularization (Alamgir & Luxburg, 2011). Given these properties, the 2-resistance is used for the multi-class graph clustering (Yen et al., 2005; Alev et al., 2017). However, in the large graph setting, the 2-resistance converges to a meaningless limit function (Nadler et al., 2009; von Luxburg

[1]Department of Computer Science, University College London, London, United Kingdom. Correspondence to: Shota Saito <ssaito@cs.ucl.ac.uk>.

Proceedings of the 40th International Conference on Machine Learning, Honolulu, Hawaii, USA. PMLR 202, 2023. Copyright 2023 by the author(s).

[1]In the following, we abbreviate effective p-resistance as p-resistance.

et al., 2010). Using the graph p-seminorm, the 2-resistance is generalized to the p-resistance (Herbster & Lever, 2009), which overcomes this problem (Slepcev & Thorpe, 2019). The $1/(p-1)$-th power of the p-resistance is also shown to be a metric (Herbster, 2010; Kalman & Krauthgamer, 2021). Furthermore, since different p of p-resistance captures a different characteristic of a graph (Alamgir & Luxburg, 2011), we expect that the parameter p serves as a tuning parameter for the clustering result. Thus, the natural idea for the multi-class clustering is to use the $1/(p-1)$-th power of p-resistance.

While the discussion above motivates us to use the $1/(p-1)$-th power of the p-resistance to multi-class clustering, there remain two issues; i) computational cost of p-resistances for many pairs ii) lack of theoretical justification for using p-resistance for clustering other than the metric property. In this paper, we address these in the following way. For i), it is computationally expensive to compute p-resistances for many pairs. The reason is that we need to solve the constrained optimization problem for many pairs. Looking back at the 2-resistance, we can compute the 2-resistance efficiently in the following way. Recall that we can compute 2-resistance as

$$r_{G,2}(i,j) = \|L^+\mathbf{e}_i - L^+\mathbf{e}_j\|_{G,2}^2, \tag{1}$$

where $r_{G,p}(i,j)$ is p-resistance for a graph G, i and j are vertices, L^+ is a pseudoinverse of the graph Laplacian L for G, $\mathbf{e}_i$ is the i-th coordinate vector of $\mathbb{R}^n$, and $\|\cdot\|_{G,2}$ is a 2-seminorm induced from the graph Laplacian L. By this representation, once we compute L^+, we can "reuse" L^+ to compute 2-resistance for different pairs. This reuse makes the computation of 2-resistances for many pairs faster than naively solving the optimization problem for each pair. However, we do not know such representation for p-resistance. Thus, to obtain p-resistance for many pairs, we need to solve many constrained optimization problems. The significant result of this work is that in Thm. 3.3, we give a theoretical guarantee for the approximation of p-resistance as

$$r_{G,p}(i,j) \approx \|L^+\mathbf{e}_i - L^+\mathbf{e}_j\|_{G,q}^p, \tag{2}$$

where q satisfies $1/p + 1/q = 1$, and $\|\cdot\|_{G,q}$ is a graph q-seminorm whose formal definition is given later. We also show that for a tree, the approximation of Eq. (2) becomes exact (Thm. 3.4). By this approximation, we can compute the approximated p-resistance efficiently, similar to the p=2 case. For ii), we do not have a theoretical justification for using p-resistance for clustering other than the metric property. While the p-resistance has the metric property, this property itself does not support the clustering quality. For spectral clustering and 2-resistance, we have theoretical justifications for clustering. For spectral clustering, using the first k eigenvectors of the graph p-Laplacian is theoretically justified (Lee et al., 2014; Tudisco & Hein, 2018). The 2-resistance has a theoretical connection to a semi-supervised learning problem of graph 2-seminorm regularization (Alamgir & Luxburg, 2011). For p-resistance, we show that p-resistance is characterized by the semi-supervised learning problem of p-seminorm regularization. This resolves the open problem stated in (Alamgir & Luxburg, 2011). This gives a theoretical foundation for using p-resistance for clustering from a view of the semi-supervised learning problem. Addressing the two issues above, as a multi-class clustering algorithm, we propose to apply the k-medoids algorithm to the distance matrix obtained from the approximated p-resistance. With these two results, our algorithm can be said to be more theoretically supported than existing multi-class spectral clusterings via graph p-Laplacian. Our experiment demonstrates that our algorithm outperforms the existing multi-class clustering using graph p-Laplacian and 2-resistance-based methods.

Our contributions are as follows: i) We give a guarantee for the approximated representation of p-resistance using the q-seminorm. ii) We show that the p-resistance characterizes the solution of semi-supervised learning of p-seminorm regularization of a graph. iii) We provide graph p-seminorm-based multi-class clustering. iv) We numerically show that our method outperforms the existing and standard methods. *All proofs are in Appendix.*

2. Preliminaries

We define a graph $G = (V, E)$, where V is a set of *vertices* and E is a set of *edges*. Throughout this paper, we use $n := |V|$ and $m := |E|$. An edge connects two vertices, and we do not consider the direction of the edge (*undirected*). A graph is *connected* if there is a path for every pair of vertices. In the following we assume that the graph is connected. We associate an ℓ-th edge (ℓ=$1,\dots,m$) with a positive value w_ℓ, which we refer to as a *weight*. We define a *weight vector* $\mathbf{w} \in \mathbb{R}^m$, whose ℓ-th element is w_ℓ. We define a *weight matrix* as a diagonal matrix $W \in \mathbb{R}^{m\times m}$ whose ℓ-th diagonal element is w_ℓ. We define an *incidence matrix* $C \in \mathbb{R}^{m\times n}$, where $c_{\ell i} := 1$ and $c_{\ell j} := -1$ when ℓ-th edge connects vertices i to j ($i > j$) otherwise 0. We represent a graph by an *adjacency matrix* $A \in \mathbb{R}^{n\times n}$; the ij-th element and ji-th element of A are w_ℓ if ℓ-th edge connects vertices i to j ($i > j$), i.e., $a_{ij} = a_{ji} := w_\ell$, and we define $a_{ij} = a_{ji} := 0$ if there is no edge between vertices i and j. By construction, the adjacency matrix A is symmetric. A *degree* d_i for a vertex i is defined as $d_i := \sum_j a_{ij}$. We define a degree matrix D, a diagonal matrix whose diagonal elements are $D_{ii} := d_i$. We define a matrix called *graph Laplacian*, as $L := D - A$. The graph Laplacian can also be written as $L = C^\top W C$. For more details, see (Bapat, 2010). A graph Laplacian induces a seminorm from a inner product $\langle \mathbf{u}, \mathbf{x}\rangle_L := \mathbf{u}^\top L\mathbf{x}$. We now

define $\mathcal{V}(L)$, a set of $L^+\mathbf{e}_i$, a *coordinate spanning set* of L, and we compute as

$$\mathcal{V}(L) := \{\mathbf{v}_i = L^+\mathbf{e}_i : i = 1, \ldots, n\},$$
$$\langle \mathbf{u}, \mathbf{v}_i \rangle_L = \mathbf{u}^\top L L^+ \mathbf{e}_i = u_i, \forall \mathbf{v}_i \in \mathcal{V}(L),\ \mathbf{u} \in \mathcal{H}(L), \quad (3)$$

where $\mathcal{H}(L) := \mathrm{span}(\mathcal{V}(L))$. This is a reproducing kernel property for a kernel whose gram matrix is L^+ (Aronszajn, 1950; Shawe-Taylor & Cristianini, 2004). Thus, $\mathbf{v}_i = L^+\mathbf{e}_i$ works as coordinate for the space $\mathcal{H}(L)$. Note that although we call $\mathcal{V}(L)$ as "coordinate", the vectors in $\mathcal{V}(L)$ are not necessarily orthonormal to each other.

An analog is established between graph and electric circuit (Doyle & Snell, 1984). In this analog, the energy $S_2(\mathbf{x})$ for a vector over vertices $\mathbf{x}\in\mathbb{R}^n$ and effective 2-resistance $r_{G,2}(i,j)$ between two vertices $i,j\in V$ are defined as

$$S_{G,2}(\mathbf{x}) := \sum_{i,j\in V} a_{ij}(x_i - x_j)^2 = \mathbf{x}^\top L \mathbf{x}, \quad (4)$$
$$r_{G,2}(i,j) := (\min_{\mathbf{x}}\{S_{G,2}(\mathbf{x}) \text{ s.t. } x_i - x_j = 1\})^{-1} \quad (5)$$

Using the inner product $\langle \cdot, \cdot \rangle_L$ and its induced seminorm $\|\cdot\|_L$, we can rewrite 2-resistance as

$$r_{G,2}(i,j) = \|L^+\mathbf{e}_i - L^+\mathbf{e}_j\|_L^2 = \|\mathbf{v}_i - \mathbf{v}_j\|_L^2, \quad (6)$$

where $\mathbf{v}_i, \mathbf{v}_j \in \mathcal{V}(L)$. From the definition, $r_{G,2}(i,i)=0$ and $r_{G,2}(i,j)=r_{G,2}(j,i)$. These energy and effective resistance are extended to p-energy $S_{G,p}$ and p-resistance $r_{G,p}$ for $p > 1$ as

$$S_{G,p}(\mathbf{x}) := \sum_{i,j\in V} a_{ij}|x_i - x_j|^p, \quad (7)$$
$$r_{G,p}(i,j) := (\min_{\mathbf{x}}\{S_{G,p}(\mathbf{x}) \text{ s.t. } x_i - x_j = 1\})^{-1}. \quad (8)$$

This p-resistance shows the triangle inequality (Herbster, 2010), that is for $a, b, c \in V$

$$r_{G,p}^{1/(p-1)}(a,b) \le r_{G,p}^{1/(p-1)}(a,c) + r_{G,p}^{1/(p-1)}(c,b). \quad (9)$$

With $r_{G,p}^{1/(p-1)}$, the graph G is a metric space. Particularly, when p=2, 2-resistance defines a metric between $\mathbf{v}_i, \mathbf{v}_j \in \mathcal{V}(L)$. More properties on p-energy and p-resistance, see (Alamgir & Luxburg, 2011; Herbster & Lever, 2009).

Lastly, we review several notions from linear algebra. We refer to (Horn & Johnson, 2012) for the details. First, we recall the weighted p-norm. Given positive weights $\mathbf{r} \in \mathbb{R}^{n_1}$ where $r_i > 0$, for a vector $\mathbf{x} \in \mathbb{R}^{n_1}$ we define the weighted p-norm $\|\mathbf{x}\|_{\mathbf{r},p}$, and its inner product $\langle \mathbf{x}, \mathbf{y} \rangle_{\mathbf{r}}$ as

$$\|\mathbf{x}\|_{\mathbf{r},p} := \left(\sum_{i=1}^{n_1} r_i |x_i|^p\right)^{1/p}, \langle \mathbf{x}, \mathbf{y} \rangle_{\mathbf{r}} := \sum_{i=1}^{n_1} r_i x_i y_i. \quad (10)$$

For this weighted p-norm and inner product, we have Hölder's inequality as follows;

Lemma 2.1 (Hölder's inequality)**.** *For $p, q > 1$ such that $1/p + 1/q = 1$, $\langle \mathbf{x}, \mathbf{y} \rangle_{\mathbf{r}} \le \|\mathbf{x}\|_{\mathbf{r},p}\|\mathbf{y}\|_{\mathbf{r},q}$.*

For a matrix $M\in\mathbb{R}^{n_1\times n_2}$, we define an *image* of M as $\mathrm{Im}(M)$:=$\{\mathbf{y}|\mathbf{y}=M\mathbf{x}, \mathbf{x}\in\mathbb{R}^{n_2}\}\subseteq\mathbb{R}^{n_1}$, that is a space spanned by the matrix M. Note that $MM^+\mathbf{y}$ is an orthogonal projection of $\mathbf{y}$ onto $\mathrm{Im}(M)$, where M^+ is a pseudoinverse of M. We introduce a *matrix operator p-norm* $|||M|||_p$ for a matrix M as

$$|||M|||_p := \sup_{\mathbf{x}\in\mathbb{R}^n} \|M\mathbf{x}\|_p / \|\mathbf{x}\|_p. \quad (11)$$

3. Graph p-Seminorm and Approximating p-Resistance

This section defines a graph p-seminorm, which is a foundation of our discussion. We then discuss several properties of the graph p-seminorm. Using these properties, we provide the approximation of p-resistance.

3.1. Graph p-Seminorm

In this section, we define a graph p-seminorm and discuss its characteristics. For a vector over vertices $\mathbf{x} \in \mathbb{R}^n$, we define a graph p-seminorm over a graph using a weighted p-norm for a graph weight vector $\mathbf{w} \in \mathbb{R}^m$. We define a graph p-seminorm $\|\mathbf{x}\|_{G,p}$ for $\mathbf{x} \in \mathbb{R}^n$ as

$$\|\mathbf{x}\|_{G,p} := \|C\mathbf{x}\|_{\mathbf{w},p} = \left(\sum_{i\in E} w_i |(C\mathbf{x})_i|^p\right)^{1/p} = \left(\sum_{i,j\in V} a_{ij}|x_i - x_j|^p\right)^{1/p}. \quad (12)$$

From the definition of p-energy Eq. (7), $S_{G,p}(\mathbf{x}) = \|\mathbf{x}\|_{G,p}^p$. Also, we immediately know that this norm is induced by the inner product $\langle C\mathbf{x}, C\mathbf{y} \rangle_{\mathbf{w}}$ from the definition of the graph p-seminorm. We now see that this graph seminorm can also be induced from the inner product $\langle \mathbf{x}, \mathbf{y} \rangle_L$, because

$$\langle C\mathbf{x}, C\mathbf{y} \rangle_{\mathbf{w}} = \mathbf{x}^\top C^\top W C \mathbf{y} = \mathbf{x}^\top L \mathbf{y} = \langle \mathbf{x}, \mathbf{y} \rangle_L. \quad (13)$$

From this observation, we see that $\|\mathbf{x}\|_{G,2} = \|\mathbf{x}\|_L$. Also, we can restrict graph p-seminorm to a norm if we consider $\mathbf{x} \in \mathrm{Im}(L)$. Note that this graph p-seminorm is same as the graph p-seminorm defined in (Herbster & Lever, 2009). For this graph p-seminorm, using Lemma 2.1, the Hölder's inequality holds;

$$\langle \mathbf{x}, \mathbf{y} \rangle_L \le \|\mathbf{x}\|_{G,p}\|\mathbf{y}\|_{G,q}, 1/p + 1/q = 1. \quad (14)$$

When $p = 2$ Hölder's inequality plays a fundamental role to show the representation of 2-resistance by Eq. (6) in the following way. Using the equality condition of the Hölder's inequality Eq. (14) for $p = 2$, we have a lemma.

Lemma 3.1 (Classical, e.g., Herbster & Pontil (2006)). *For* $\mathbf{y} \in \mathbb{R}^n$, $\|\mathbf{y}\|_{G,2}^{-2} = \min_{\mathbf{x}}\{\|\mathbf{x}\|_{G,2}^2 \text{ s.t. } \langle \mathbf{x}, \mathbf{y}\rangle_L = 1\}$.

This lemma is a classical result rewritten with our notation of graph p-seminorm. By substituting $\mathbf{y}{:=}L^+\mathbf{e}_i - L^+\mathbf{e}_j$, the right hand side of Lemma 3.1 becomes the inverse of 2-resistance (see Appendix C). Thus, we obtain $r_{G,2}(i,j){=}\|L^+\mathbf{e}_i - L^+\mathbf{e}_j\|_L^2{=}\|L^+\mathbf{e}_i - L^+\mathbf{e}_j\|_{G,2}^2$. For p-resistance, the question is how is the coordinate spanning set $\mathcal{V}(L^+)$ related to the p-resistance? Can we derive such relation using Hölder's inequality Eq. (14), similarly to the $p{=}2$ case? Next section will show such connection between p-resistance and the coordinate spanning set using Eq. (14).

3.2. Approximating p-Resistance via Coordinate Spanning Set

This section discusses approximation of p-resistance via the coordinate spanning set $\mathcal{V}(L^+)$. Looking at the $p{=}2$ case, we see that $L^+\mathbf{e}_i{\in}\mathcal{V}(L^+)$ can be regarded as coordinate, and $r_{G,2}(i,j){=}\|L^+\mathbf{e}_i - L^+\mathbf{e}_j\|_{G,2}^2$. This expression aids us to compute all the pairs of 2-resistance much faster than naively obtaining 2-resistance. For p-resistance, a natural question to ask is that does there exist some norm $\|\cdot\|^\ddagger$ such that $r_{G,p}(i,j) = \|L^+\mathbf{e}_i - L^+\mathbf{e}_j\|^\ddagger$? If not, how can we approximate as $r_{G,p}(i,j) \approx \|L^+\mathbf{e}_i - L^+\mathbf{e}_j\|^\ddagger$? If we can write p-resistance by such expression, we expect to obtain all the pairs of approximated p-resistance much faster than naively computing all the pairs of p-resistance. This section addresses this problem.

As we see in Sec. 3.1, Lemma 3.1 is a key to show that $r_{G,2}(i,j) = \|L^+\mathbf{e}_i - L^+\mathbf{e}_j\|_{G,2}^2$. In the following, we now extend Lemma 3.1 from the case of $p = 2$ to the general p.

Proposition 3.2. *For a graph G and $p, q > 1$ such that $1/p + 1/q = 1$, we have*

$$\|\mathbf{y}\|_{G,q}^{-p} \le \min_{\mathbf{x}}\{\|\mathbf{x}\|_{G,p}^p \text{ s.t. } \langle \mathbf{y}, \mathbf{x}\rangle_L = 1\} \le \|\mathbf{z}\|_{G,p}^p \quad (15)$$

where

$$\mathbf{z} := C^+\frac{f_{q/p}(C\mathbf{y})}{\|\mathbf{y}\|_{G,q}^q}, \quad (f_\theta(\mathbf{x}))_i := \mathrm{sgn}(x_i)|x_i|^\theta. \quad (16)$$

When $f_{q/p}(C\mathbf{y}) \in \mathrm{Im}(C)$, we have

$$\|\mathbf{y}\|_{G,q}^{-p} = \min_{\mathbf{x}}\{\|\mathbf{x}\|_{G,p}^p \text{ s.t. } \langle \mathbf{y}, \mathbf{x}\rangle_L = 1\} = \|\mathbf{z}\|_{G,p}^p \quad (17)$$

We first note that the minimization problem of Eq. (15) is the inverse of p-resistance Eq. (7). The left hand side of inequality Eq. (15) immediately follows from Hölder's inequality (Eq. (14)) with $\langle \mathbf{y}, \mathbf{x}\rangle_L = 1$. We now turn our attention to the right hand side. Recall that when $p = 2$ we always have $f_{q/p}(C\mathbf{x}) \in \mathrm{Im}(C)$ and $\|\mathbf{y}\|_{G,2}^{-1} = \|\mathbf{z}\|_{G,2}$, which matches Lemma 3.1. In the general p case, $f_{q/p}(C\mathbf{x}) \notin \mathrm{Im}(C)$ and $\|\mathbf{y}\|_{G,q}^{-1}{\neq}\|\mathbf{z}\|_{G,p}$. Thus, neither $\|\mathbf{y}\|_{G,q}^{-p}$ nor $\|\mathbf{z}\|_{G,p}^p$ gives the solution to the minimization problem. However, this theorem tells us that we can upper bound the solution to the minimization problem by $\|\mathbf{z}\|_{G,p}^p$.

Applying Prop. 3.2, we obtain the bound for p-resistance as follows;

Theorem 3.3. *For a graph G and $p, q > 1$ such that $1/p + 1/q = 1$, the p-resistance can be bounded as*

$$\frac{1}{\alpha_{G,p}^p}\|L^+\mathbf{e}_i - L^+\mathbf{e}_j\|_{G,q}^p \le r_{G,p}(i,j) \le \|L^+\mathbf{e}_i - L^+\mathbf{e}_j\|_{G,q}^p,$$
$$\text{where } \alpha_{G,p} := \|\!|W^{1/p}CC^+W^{-1/p}|\!\|_p,$$

Theorem 3.4. *For a tree G and $p, q > 1$ such that $1/p + 1/q = 1$, the p-resistance can be written as*

$$r_{G,p}(i,j) = \|L^+\mathbf{e}_i - L^+\mathbf{e}_j\|_{G,q}^p \quad (18)$$

Thm. 3.3 and Thm. 3.4 show the relationship between p-resistance and $\|L^+\mathbf{e}_i - L^+\mathbf{e}_j\|_{G,q}^p$. For general graphs, we do not obtain the exact representation of p-resistance. However, Thm. 3.3 guarantees the quality of approximation as

$$r_{G,p}(i,j) \approx \|L^+\mathbf{e}_i - L^+\mathbf{e}_j\|_{G,q}^p = \|\mathbf{v}_i - \mathbf{v}_j\|_{G,q}^p, \quad (19)$$

where $\mathbf{v}_i{,}\mathbf{v}_j{\in}\mathcal{V}(L)$. By this approximation, we obtain the similar representation of p-resistance to the $p = 2$ case Eq. (6). The term $\alpha_{G,p}$ is a p-norm of the orthogonal projector to $\mathrm{Im}(W^{1/p}C)$. Note that we always have $\alpha_{G,p}{\ge}1$. For a tree graph, Thm. 3.4 shows that $\|L^+\mathbf{e}_i - L^+\mathbf{e}_j\|_{G,q}^p$ becomes the exact representation of p-resistance.

The next question is what is $\alpha_{G,p}$. We bound $\alpha_{G,p}$ as follows;

Proposition 3.5. *For a general graph G and $p > 1$, we have $\alpha_{G,p} \le m^{|1/2-1/p|}$.*

This proposition gives the guarantee for the approximation in Thm. 3.3. Although Prop. 3.5 gives the quality guarantee, we expect this upper bound to be loose, i.e., we expect that the actual approximation value is closer to the exact value than this bound. The reason why we expect in this way is that to prove the bound we only use the general technique that holds for any matrix and we do not use any graph structural information. In the real dataset, we observe that the approximation of p-resistance and $\alpha_{G,p}$ is far better than this guarantee, see Appendix I.3. We give more discussion on this $\alpha_{G,p}$ in Appendix J.

Finally, we discuss computational times of the p-resistance. To compute Eq. (19), it takes $O(m)$, given L^+. Also, in general it takes $O(n^3)$ to compute L^+. Note that we can reuse L^+ to compute p-resistance for different pairs. We now consider to obtain the p-resistance by naively solving the optimization problem. We can rewrite the constrained

problem Eq. (8) as unconstrained problem, which is solvable by gradient descent. In each step of the gradient descent, we compute $\nabla_{\mathbf{x}}\|\mathbf{x}\|_{G,p}^{p}$, which takes almost same time as Eq. (19). Moreover, we cannot reuse the result of a single pair to compute for other pairs, while we can reuse L^{+}. Thus, to compute p-resistance for a single pair, our approximation is expected to be faster than naively solving the optimization problem. Moreover, if we compute for many pairs, our approximation is much faster by reusing L^{+}.

Algorithm 1 Clustering Algorithm via p-Resistance

Require: Graph $G = (V, E)$ and p

1: Compute pseudoinverse of the graph Laplacian L^{+}.
2: Compute all the pairs of the p-resistance metrics $r_{G,p}^{1/(p-1)}$ using Eq. (20) and obtain a distance matrix.
3: Apply k-medoids to the distance matrix.

Ensure: The clustering result.

4. Clustering via p-Resistance

This section considers using the p-resistance for the clustering algorithm. Firstly, we propose a clustering algorithm using the approximated p-resistance. We next characterize our clustering algorithm from the semi-supervised problem point of view. From this characterization, we can see that our clustering algorithm inherits properties from semi-supervised learning.

4.1. Proposed Clustering Algorithm via p-Resistance

This section proposes an algorithm using p-resistance. The triangle inequality Eq. (9) gives a metric property to $r_{G,p}^{1/(p-1)}(i,j)$. We call this $1/(p-1)$-th power of p-resistance as *p-resistance metric*. This metric property motivates us to use p-resistance for clustering algorithms.

Furthermore, the parameter p serves as a tuning parameter of the clustering result. The general p of p-resistance captures the graph structure somewhere between the cut and shortest path. Using this characteristic, we expect varying p tunes the clustering result somewhere suitable between cut-based and path-based. When p is small, the clustering result biases towards clusters with high internal connectivity, like a min-cut. When p is large, the clustering result focus more on path-based topology, that is a preference for smaller shortest-path distances between vertices in the cluster. We illustrate this with examples of the two-class clustering in Fig. 1. In these examples, we conduct clustering with k-center algorithm using p-resistance. The left example is intuitively "symmetric"; for this kind, $p \to \infty$, which looks at the path-based topology, gives more natural result. The more natural clustering of the right example is "cut"; for this kind, $p \to 1$, where we focus on the graph cut, gives the more natural result. More details are in Appendix G.

While the discussion above motivates us to use the p-resistance metric for clustering, computing the p-resistance metric for all pairs is costly. Thus, we approximate this metric by Thm. 3.3, and we obtain

$$\begin{aligned} r_{G,p}^{1/p-1}(i,j) &\approx \|L^{+}\mathbf{e}_i - L^{+}\mathbf{e}_j\|_{G,q}^{p/(p-1)} \\ &= \|L^{+}\mathbf{e}_i - L^{+}\mathbf{e}_j\|_{G,q}^{q} = \|\mathbf{v}_i - \mathbf{v}_j\|_{G,q}^{q}, \end{aligned} \quad (20)$$

where $\mathbf{v}_i, \mathbf{v}_j \in \mathcal{V}(L)$. We then apply k-medoids to the distance matrix obtained by Eq. (20). The overall proposed algorithm is summarized in Alg. 1

We discuss the choice of k-medoids over the other distance based method, such as k-means (Bishop & Nasrabadi, 2006) and k-center (Gonzalez, 1985). Although the main emphasis of our algorithm does not comes from the choice of k-median but from the approximation of p-resistance metric, k-median has some advantages. Since p-resistance metric cannot define a distance between other than the data points defined as $\mathcal{V}(L^{+})$, we cannot define distance for the some "mean", which is outside of the data points. Therefore, the mean-based method such as k-means is not appropriate for this setting. Instead, k-medoids is similar to the k-means (Kaufman & Rousseeuw, 1990) but more appropriate since k-medoids assigns the centers to the actual data points. The other potential choice is k-center algorithm. The k-center algorithm also assigns the center to the actual data point, and is known to be faster than k-medoids. Also, k-center algorithm is approximated by the fast greedy farthest first algorithm (Gonzalez, 1985; Herbster, 2010). However, the k-medoids is more robust to the outliers than k-center. Thus, we propose to use k-medoids. The overall computational time for Alg. 1 is dominated by the computation of the all the pairs of the approximated p-resistance, $O(mn^2)$. If we use the farthest first algorithm instead of k-medoids the algorithm is dominated either by the computation of L^{+}, $O(n^3)$, or farthest first $O(kmn)$. Thus, farthest first is faster since in general $m \gg n$ but less robust than k-medoids.

4.2. Connection between Semi-supervised Learning and p-Resistance

This section explores connection from the p-resistance to the semi-supervised learning (SSL) via graph p-seminorm. As we saw in Sec. 2, Herbster (2010) shows the metric property of p-resistance. While the metric property itself can motivates us to use p-resistance for our clustering, we do not know how much p-resistance shows connectivity of a graph. This section shows that p-resistance can be seen from as an SSL perspective. This connection assures us to use p-resistance for the clustering problem. In the following we explain the connection by taking the following steps; i) SSL problem in the clustering context ii) the connection between the SSL and p-resistance.

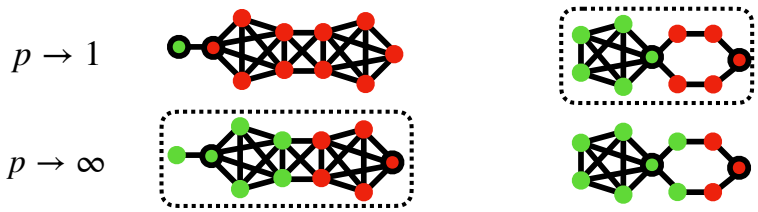


Figure 1: The illustrative examples where p changes the clustering results. These examples conduct clustering with k-center algorithm using p-resistance as a metric. The red and green colors show the clustering result. Also, the vertices with borders show the obtained centers. The dotted boxes exhibit natural clustering results. These examples show varying p tunes the clustering result; the left example gives a more natural clustering result when $p \to \infty$ whereas for right $p \to 1$ gives more natural result. Details are in Sec. 4.1 and Appendix G.

We first consider an SSL problem for two known labels as

$$\min_{\mathbf{x}}\{S_{G,p}(\mathbf{x}) \text{ s.t. } x_i - x_j = 1\} = \min_{\mathbf{x}}\{S_{G,p}(\mathbf{x}) \text{ s.t. } x_i = 1, x_j = 0\}. \tag{21}$$

The equality holds since $S_{G,p}(\mathbf{x}){=}S_{G,p}(\mathbf{x}{+}c\mathbf{1}), \forall c{\in}\mathbb{R}$. We first note that Eq. (21) is an inverse of the p-resistance and we use the optimal value of this problem to p-resistance. This learning problem for $p = 2$ case has been considered in many literature, such as (Zhu et al., 2003), and extended to the p-seminorm setting (Herbster & Lever, 2009; Alamgir & Luxburg, 2011; Slepcev & Thorpe, 2019).

We now put Eq. (21) into clustering context; the solution of Eq. (21) tells us the graph structural information on clustering. We recognize that Eq. (21) is two fixed-label problem. Let $\mathbf{x}^{*ij}$ be a solution of the problem Eq. (21). It is straightforward to interpret Eq. (21) if i and j is in different binary classes; we see which clusters the third point ℓ belongs to, the cluster which i or j is in. More specifically, by comparing $x_\ell^{*ij} - x_j^{*ij}$ and $x_i^{*ij} - x_\ell^{*ij}$ we know which cluster the third point ℓ belongs to. If we take the a pair of vertices (i, j) arbitrarily, the assumption that "i and j in different binary classes" is not always appropriate. In this case, rather than assuming i and j in different binary classes, it is more natural to interpret in the following way; the two-pole binary SSL problem tells us that which of i and j the third point ℓ is close to in a graph. From this observation, if we look at $\mathbf{x}^{*ij}$ for all pairs, we know "graph structural information" from the SSL point of view.

We next show the connection between p-resistance and the solution of Eq. (21), $\mathbf{x}^{*ij}$.

Theorem 4.1. *Let $\mathbf{x}^{*ij}$ be the solution of the problem Eq. (21), and $\ell \in V$ be the third unlabeled point. Then we have*

$$x_\ell^{*ij} - x_j^{*ij} \geq x_i^{*ij} - x_\ell^{*ij} \iff r_{G,p}(j, \ell) \geq r_{G,p}(\ell, i).$$

First note that Alamgir & Luxburg (2011) proved Thm. 4.1 only for the $p = 2$ case in a different context than clustering (See Appendix H.2), and posed the case of general p as an open problem. We resolve this open problem.

Thm. 4.1 means that the p-resistance has a good property inherited from the SSL problem Eq. (21) in a following sense. Thm. 4.1 tells us that the "graph structural information", which can be obtained by comparing $x_\ell^{*ij} - x_i^{*ij}$ and $x_j^{*ij} - x_\ell^{*ij}$, is equivalent to comparing p-resistances $r_{G,p}(i, \ell)$ and $r_{G,p}(\ell, j)$. Henceforth, Thm. 4.1 further translates the intuition about $\mathbf{x}^{*ij}$ into p-resistance.

Thus, combining the two observations above, looking at the distance matrix computed from p-resistances can be interpreted as follows. Each distance shows how close the pair is in terms of two-pole binary SSL problem. Doing clustering with this distance matrix assigns a cluster by looking at all the graph structural information of two-pole binary SSL problems, which tells us that "which the third point ℓ is close to, i or j?"

From the observations above, we see that Thm. 4.1 motivates us to use p-resistance metrics for multi-class problem. Without Thm. 4.1, our algorithm is somewhat naive; even though p-resistance has a metric property, we do not know how much p-resistance contains the structural information.

5. Related Work

This section reviews the related work to the clustering via graph p-seminorm. Since our work uses graph p-seminorm for the clustering purpose, spectral clustering using graph p-Laplacian is relevant. The graph p-Laplacian is induced from graph p-seminorm and used for the clustering purpose (Bühler & Hein, 2009). Tudisco & Hein (2018) showed a theoretical guarantee for the use of the first k variational eigenvectors (i.e., eigenvectors obtained by variational principle) of p-Laplacian for k-class clustering. While we know the exact identification for the second eigenvectors of p-Laplacian, we do not know how to obtain the third or higher eigenvectors (Lindqvist, 2008). Thus, it is practically diffi-

cult to use spectra of p-Laplacian for multi-class clustering. To bypass this limitation, Bühler & Hein (2009) applied two-class clustering method to multi-class by recursively bisectioning a subgraph into two subgraphs. However, even in the p=2 case, recursive bisectioning produces suboptimal results (Simon & Teng, 1997). The earlier works (Luo et al., 2010; Ding et al., 2019; Pasadakis et al., 2022) used approximated orthogonality between eigenvectors of p-Laplacian for multi-class clustering. However, we do not have theoretical supports that this approximated k eigenvectors are the approximation of *the first* k variational eigenvectors. Thus, we need to say that these methods rely on the "ad-hoc bypasses" and do not fully exploit the graph p-seminorm. For more details, see Appendix A.

Another relevant approach is resistance-based clustering. In (Yen et al., 2005), k-medoids algorithm is applied to the square of 2-resistance. For clustering purpose, similar distances to the 2-resistance is proposed (Fouss et al., 2007; Nguyen & Mamitsuka, 2016; Yen et al., 2008) The most relevant approach in this category is the k-center algorithm for the "distance" matrix obtained from the *exact* p-resistance in (Herbster, 2010). Herbster (2010) did not numerically verify the algorithm. Our work uses p-resistance metric instead of p-resistance since without the $1/(p-1)$-th power operation p-resistance does not satisfy the metric property (Eq. (9)). However, if we use k-center algorithm to the *exact* p-resistance metric, we obtain the same result as (Herbster, 2010). The reason is that the k-center algorithm only matters the order of the distance, and the $1/(p-1)$-th power operation does not change the order of the p-resistance. On the other hand, we emphasize that the most significant difference between our work and (Herbster, 2010) is that while we use the *approximated* p-resistance (Herbster, 2010) uses exact p-resistance. We also mention that the work (Nguyen & Mamitsuka, 2016) proposed a distance from the p-seminorm flow point of view. However, this distance does not have characterization from the learning problem (Thm. 4.1).

Also, the graph p-seminorm is actively used in semi-supervised learning (SSL). The SSL problem using graph p-seminorm is relevant to p-resistance since the p-resistance can be seen as SSL for two known labels. Earlier, the SSL using graph 2-seminorm is considered (Zhou et al., 2003; Zhu et al., 2003; Calder et al., 2020). The SSL via graph 2-seminorm and effective resistance is known to be "ill-posed" when the size of the unlabeled data points is asymptotically large (Nadler et al., 2009). To overcome this problem, graph p-seminorm based SSL and p-resistance are considered (Alamgir & Luxburg, 2011; Bridle & Zhu, 2013; El Alaoui et al., 2016; Slepcev & Thorpe, 2019), where the p-resistance is shown to be meaningful when p is large. Finally, the graph p-seminorm is widely used in the machine learning community, such as online learning (Herbster & Lever, 2009) and the local graph clustering task, where we find a cluster which the given vertices belong to (Veldt et al., 2019; Fountoulakis et al., 2020; Liu & Gleich, 2020).

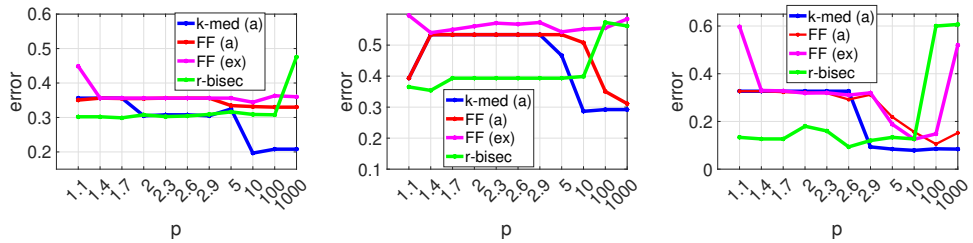
6. Preliminary Experiments

This section numerically demonstrates the performance of our Alg. 1 using approximated p-resistance. The purpose of this preliminary experiments is to evaluate if our algorithm on two-class and multi-class clustering problem improves the existing p-seminorm based graph clustering algorithm. Thus, we compared with existing resistance based algorithms and spectral clustering algorithms using graph p-seminorm and its p=2 setting. For the resistance based method, we compared with the farthest first algorithm on our approximated p-resistance. Additionally, we compared with existing methods; the farthest first using *exact* p-resistance (Herbster, 2010), a p-seminorm flow based method (Nguyen & Mamitsuka, 2016), and 2-resistance based method (Yen et al., 2005). We especially note that for the farthest first (Herbster, 2010), we computed the exact p-resistance by the gradient descent as discussed in Sec. 3.2. We also apply this exact p-resistance to k-medoids. Note that k-medoids is more costly than k-center as discussed in Sec. 4.1. For spectral clustering methods, we compared with a recursively bisection method (Bühler & Hein, 2009) and a method using the approximated orthogonality (Luo et al., 2010). Since the p-resistance is related to the unnormalized graph Laplacian, we use unnormalized graph Laplacian for the spectral methods. Our experiments were performed on classification datasets, ionosphere, iris and wine from UCI repository. We also used Hopkins155 dataset (Tron & Vidal, 2007), which contains 120 two-class motion segmentation datasets and 35 three-class ones. We created a graph with the following procedure. First, we built a k-NN graph, where we choose k=μn ($0<\mu\leq 1$). Then, we computed the edge weight with a Gaussian kernel ($\kappa(\mathbf{x}_i,\mathbf{x}_j)$=$\exp(-\sigma\|\mathbf{x}_i-\mathbf{x}_j\|^2)$) for two real-valued vectors $\mathbf{x}_i,\mathbf{x}_j$. We used free parameters $\mu\in\{0.04, 0.06, 0.08, 0.1, 1\}$, $\sigma\in\{10^{-3},\dots,10^2\}$ and $p\in\{1.1, 1.4,\dots,2.9, 5, 10, 100, 1000\}$. For comparisons, we followed the original parameters other than above μ,σ, and p. We evaluated the performance by error rate, similarly to the previous study (Bühler & Hein, 2009). Since the Hopkins155 dataset contains multiple two-class and three-class tasks, we take an average of error rates among a set of two-class tasks and three-class tasks and report both. The implementation of our method is available at `https://github.com/ShotaSAITO/approximated-presistance`.

The results are summarized as follows. We see that ours outperforms the others except for iris. As we expected, seeing the deviation, k-medoids offers more robust performance than the farthest-first algorithms. Moreover, our approxi-

Table 1: Experimental Results. The "type" shows the type of methods; (ER) for effective resistance based methods and (SC) for spectral clustering methods. The "Hop" stands for Hopkins 155 dataset. In method of ER, "(a)" shows that the method uses the approximation by (Eq. (20)) and "(ex)" computes the exact p-resistance by gradient descent. Also, "k-med" is k-medoids, and "FF" is the farthest first. Thus, the method "k-med (a) p" is our proposed algorithm, and "FF (ex) p" and "FF $p = 2$" is a method proposed by (Herbster, 2010). The "p-Flow" is (Nguyen & Mamitsuka, 2016), "ECT" is (Yen et al., 2005), "Rec-bi p" is (Bühler & Hein, 2009), and "p-orth" is (Luo et al., 2010). Since "Rec-bi p" is a deterministic method, we only report error. Also, since Hop contains multiple datasets, we only show the average. Due to the significant computational time, we were unable to finish some of the experiments, which are shown as "–".

		2 clsss		multi-class		
Type	Method	ionosphere	Hop 2 cls	iris	wine	Hop 3 cls
ER	k-med (a) p	**0.196 ± 0.000**	**0.056**	0.078 ± 0.013	**0.287 ± 0.000**	**0.144**
ER	k-med (ex) p	–	–	**0.075 ± 0.000**	0.427 ± 0.000	–
ER	k-med $p = 2$	0.305 ± 0.000	0.236	0.331 ± 0.000	0.534 ± 0.000	0.306
ER	FF (a) p	0.330 ± 0.023	0.109	0.108 ± 0.045	0.339 ± 0.054	0.313
ER	FF (ex) p (Herbster, 2010)	0.344 ± 0.020	–	0.109 ± 0.019	0.524 ± 0.046	–
ER	FF $p = 2$ (Herbster, 2010)	0.355 ± 0.035	0.274	0.320 ± 0.000	0.530 ± 0.000	0.357
ER	p-Flow (Nguyen & Mamitsuka, 2016)	0.291 ± 0.000	0.231	0.247 ± 0.000	0.543 ± 0.043	0.243
ER	ECT (Yen et al., 2005)	0.376 ± 0.000	0.155	0.247 ± 0.000	0.534 ± 0.000	0.310
SC	Rec-bi p (Bühler & Hein, 2009)	0.225	0.200	0.089	0.354	0.237
SC	SC p-orth (Luo et al., 2010)	0.215 ± 0.123	0.237	0.087 ± 0.089	0.327 ± 0.116	0.221
SC	SC $p = 2$	0.308 ± 0.000	0.216	0.093 ± 0.000	0.438 ± 0.000	0.251



(a) Ion (b) Wine (c) Iris

Figure 2: Plots of the error verse p.

Table 2: Computational time for approximated vs exact p-resistance. (a) denotes approximation and (e) denotes exact. In "r" we reuse L^+. In "et" we compute L^+ each time. All time is in second.

	ionosphere	iris	wine
(a) + r	0.08 ± 0.04	0.07 ± 0.03	0.01 ± 0.00
(a) + et	0.39 ± 0.00	0.32 ± 0.00	0.05 ± 0.00
(e)	1.11 ± 0.00	1.03 ± 0.04	0.36 ± 0.00

mation provides faster computation than the exact method since we could not finish some of the experiments using the exact p-resistance even for the farthest first. Seeing Fig. 2, in k-medoids large p offers better performance. Also, if we look at $p = 2$, the k-medoids with 2-resistance is not always better than spectral clustering. However, for general p the k-medoids with p-resistance performs better. Thus, the k-medoids with p-resistance can be said to be more benefited by p than spectral clustering. These correspond to the existing theoretical indication; p-resistance with large p becomes meaningful function while 2-resistance is not (Alamgir & Luxburg, 2011). Comparing exact and approximation p-resistance in Fig. 2, while we observe similar performance in the middle range of p, we observe the better performance for approximation at the very small p or very large ps. This might come from the numerical computation of the gradient $\nabla_{\mathbf{x}} \|\mathbf{x}\|_{G,p}^p$ as follows. For the exact solution of the very small p, the gradient of each step tends to be very small. For the very large p, there is a risk of amplifying round-off numerical error at each step of optimization by taking the power of large p. On the other hand, the approximation offers a robust computation, especially for important large ps, because instead we compute by taking the power of small q in Eq. (20), by which we can avoid the risk discussed.

We next compare times to compute the exact and approximated p-resistance for randomly chosen 100 pairs of vertices. We made a graph for ionosphere, iris, and wine by

choosing the best-performing parameters in Table. 1. We measure time for exact p-resistance by naively optimizing Eq. (7) by the gradient descent. For approximation, since we can reuse L^+ for our approximation, we report the time in two ways; i) we compute L^+ each time and compute the approximation ii) we reuse L^+. The "each time" scenario reports the computational time of p-resistance for a single pair. In the reuse scenario, we measure time t to compute L^+ and p-resistance 100 pairs using L^+. Then, we report the time $t/100$. By this, the reuse scenario is much faster than the each time scenario. Table 2 summarizes the computing time. We observe that the approximation method for a single pair is faster than the exact method by comparing the "each time" and exact. As expected, the reuse scenario provides much faster computation than the exact p-resistance For more details and results, including the computing time of Table 1 and the comparison between values of the exact and approximated p-resistance, see Appendix I.

7. Conclusion

We have proposed the multi-class clustering algorithm using the approximated p-resistance. For this purpose, we have shown the guarantee for the approximation of p-resistance. We also have shown that p-resistance characterizes the solution of the semi-supervised learning problem. Our algorithm has outperformed the existing clustering methods using the graph p-seminorm.

The limitation of this work is that we cannot exploit the sparse structures of graphs. It is because we use L^+, which becomes dense even if the graph is sparse. For future work, there remains an ample opportunity to further speed up the procedures involving pseudoinverse of graph Laplacian, such as sparsification techniques (Spielman & Srivastava, 2008; 2011; Spielman & Teng, 2014) or numerical linear-algebraic methods (Saad, 2003). Another direction to consider is connection from p-resistance to p-Laplacian and nonlinear modularity such as (Tudisco et al., 2018). However, unfortunately, the current state of knowledge in the field offers very limited theoretical understanding, even for the standard case; we do not know the connections between 2-resistance and the standard existing clustering methods using the standard modularity and also the standard Laplacian. Nevertheless, this potential connection highlights interesting open challenges. Additionally, since we can compute 2-resistance by the simple size n norm $\|(L^+)^{1/2}\mathbf{e}_i - (L^+)^{1/2}\mathbf{e}_j\|$, it is interesting to see if we can further approximate $\|L^+\mathbf{e}_i - L^+\mathbf{e}_j\|_{G,p}$ by the size n norm instead of the size m graph p-seminorm. This further speeds up the approximation from $O(m)$ to $O(n)$. Moreover, instead of our approximated representation of p-resistance approach, the exciting approach is to obtain exact representation of p-resistance. We leave some discussion on the difficulty of this approach in Appendix. K. Furthermore, the experiments using network datasets would be a nice to confirm. Finally, it would be also interesting if we apply this p-resistance framework to the recently growing space of hypergraph clustering using hypergraph p-Laplacians (Hein et al., 2013; Saito et al., 2018; Li & Milenkovic, 2018; Saito & Herbster, 2023).

Acknowledgments

This research has been made possible thanks to Huawei, who are supporting Shota Saito's PhD study at UCL.

References

Alamgir, M. and Luxburg, U. Phase transition in the family of p-resistances. In *Proc. NIPS*, pp. 379–387, 2011.

Alev, V. L., Anari, N., Lau, L. C., and Gharan, S. O. Graph clustering using effective resistance. *arXiv preprint arXiv:1711.06530*, 2017.

Amghibech, S. Eigenvalues of the discrete p-Laplacian for graphs. *Ars Comb.*, 2003.

Aronszajn, N. Theory of reproducing kernels. *Trans. Am. Math. Soc*, 68(3):337–404, 1950.

Azimi, A. and Bapat, R. B. Moore–penrose inverse of the incidence matrix of a distance regular graph. *Linear Algebra Appl.*, 551:92–103, 2018.

Azimi, A., Bapat, R. B., and Estaji, E. Moore–penrose inverse of incidence matrix of graphs with complete and cyclic blocks. *Discrete Math.*, 342(1):10–17, 2019.

Bapat, R. B. Moore-penrose inverse of the incidence matrix of a tree. *Linear Multilinear Algebra*, 42(2):159–167, 1997.

Bapat, R. B. *Graphs and matrices*. Springer, 2010.

Ben-Israel, A. and Greville, T. N. *Generalized inverses: theory and applications*. Springer, 2003.

Bishop, C. M. and Nasrabadi, N. M. *Pattern recognition and machine learning*. Springer, 2006.

Bougleux, S., Elmoataz, A., and Melkemi, M. Discrete regularization on weighted graphs for image and mesh filtering. In *Proc. SSVM*, pp. 128–139, 2007.

Bougleux, S., Elmoataz, A., and Melkemi, M. Local and nonlocal discrete regularization on weighted graphs for image and mesh processing. *Int. J. Comput. Vision*, 84 (2):220–236, 2009.

Bridle, N. and Zhu, X. p-voltages: Laplacian regularization for semi-supervised learning on high-dimensional data. In *Proc. MLG*, 2013.

Bühler, T. and Hein, M. Spectral clustering based on the graph p-Laplacian. In *Proc. ICML*, pp. 81–88, 2009.

Calder, J. The game theoretic p-Laplacian and semi-supervised learning with few labels. *Nonlinearity*, 32 (1):301, 2018.

Calder, J., Cook, B., Thorpe, M., and Slepcev, D. Poisson learning: Graph based semi-supervised learning at very low label rates. In *Proc. ICML*, pp. 1306–1316, 2020.

Deidda, P., Putti, M., and Tudisco, F. Nodal domain count for the generalized graph p-laplacian. *arXiv preprint arXiv:2201.01248*, 2022.

Ding, S., Cong, L., Hu, Q., Jia, H., and Shi, Z. A multiway p-spectral clustering algorithm. *Knowl. Based Syst.*, 164: 371–377, 2019.

Doyle, P. G. and Snell, J. L. *Random walks and electric networks*, volume 22. American Mathematical Society, 1984.

El Alaoui, A., Cheng, X., Ramdas, A., Wainwright, M. J., and Jordan, M. I. Asymptotic behavior of ℓ_p-based Laplacian regularization in semi-supervised learning. In *Proc. COLT*, pp. 879–906, 2016.

Elmoataz, A., Lezoray, O., and Bougleux, S. Nonlocal discrete regularization on weighted graphs: a framework for image and manifold processing. *IEEE Trans. Image Process.*, 17(7):1047–1060, 2008.

Fountoulakis, K., Wang, D., and Yang, S. p-norm flow diffusion for local graph clustering. In *Proc. ICML*, pp. 3222–3232, 2020.

Fouss, F., Pirotte, A., Renders, J.-M., and Saerens, M. Random-walk computation of similarities between nodes of a graph with application to collaborative recommendation. *IEEE Trans. Knowl. Data Eng.*, 19(3):355–369, 2007.

Gonzalez, T. F. Clustering to minimize the maximum intercluster distance. *Theor. Comput. Sci.*, 38:293–306, 1985.

Hein, M., Setzer, S., Jost, L., and Rangapuram, S. S. The total variation on hypergraphs - learning on hypergraphs revisited. In *Proc. NIPS*, pp. 2427–2435, 2013.

Herbster, M. A triangle inequality for p-resistance. Technical Note. https://discovery.ucl.ac.uk/id/eprint/1311162/1/triangle.pdf, 2010.

Herbster, M. and Lever, G. Predicting the labelling of a graph via minimum p-seminorm interpolation. In *Proc. COLT*, 2009.

Herbster, M. and Pontil, M. Prediction on a graph with a perceptron. In *Proc. NIPS*, pp. 577–584, 2006.

Higham, N. J. Estimating the matrix p-norm. *Numer. Math.*, 62(1):539–555, 1992.

Horn, R. A. and Johnson, C. R. *Matrix analysis*. Cambridge University Press, 2012.

Kalman, L. and Krauthgamer, R. Flow metrics on graphs. *arXiv preprint arXiv:2112.06916*, 2021.

Kaufman, L. and Rousseeuw, P. J. Partitioning around medoids (program pam). *Finding groups in data: an introduction to cluster analysis*, 344:68–125, 1990.

Klein, D. J. and Randić, M. Resistance distance. *J. Math. Chem.*, 12(1):81–95, 1993.

Lee, J. R., Gharan, S. O., and Trevisan, L. Multiway spectral partitioning and higher-order cheeger inequalities. *J. ACM*, 61(6):37, 2014.

Li, P. and Milenkovic, O. Submodular hypergraphs: p-Laplacians, cheeger inequalities and spectral clustering. In *Proc. ICML*, pp. 3020–3029, 2018.

Lindqvist, P. A nonlinear eigenvalue problem. In *Topics in Mathematical Analysis*, pp. 175–203. World Scientific, 2008.

Liu, M. and Gleich, D. F. Strongly local p-norm-cut algorithms for semi-supervised learning and local graph clustering. In *Proc. NeurIPS*, pp. 5023–5035, 2020.

Luo, D., Huang, H., Ding, C., and Nie, F. On the eigenvectors of p-Laplacian. *Mach. Learn.*, 81(1):37–51, 2010.

Nadler, B., Srebro, N., and Zhou, X. Semi-supervised learning with the graph Laplacian: The limit of infinite unlabelled data. In *Proc. NIPS*, pp. 1330–1338, 2009.

Nguyen, C. H. and Mamitsuka, H. New resistance distances with global information on large graphs. In *Proc. AISTATS*, pp. 639–647, 2016.

Pasadakis, D., Alappat, C. L., Schenk, O., and Wellein, G. Multiway p-spectral graph cuts on grassmann manifolds. *Mach. Learn.*, 111(2):791–829, 2022.

Saad, Y. *Iterative methods for sparse linear systems*. SIAM, 2003.

Saito, S. and Herbster, M. Generalizing p-Laplacian: spectral hypergraph theory and a partitioning algorithm. *Mach. Learn.*, 112(1):241–280, 2023.

Saito, S., Mandic, D. P., and Suzuki, H. Hypergraph p-Laplacian: A differential geometry view. In *Proc. AAAI*, pp. 3984–3991, 2018.

Shawe-Taylor, J. and Cristianini, N. *Kernel methods for pattern analysis*. Cambridge University Press, 2004.

Shi, J. and Malik, J. Normalized cuts and image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell*, 22:888–905, 1997.

Simon, H. D. and Teng, S.-H. How good is recursive bisection? *SIAM J. Sci. Comput.*, 18(5):1436–1445, 1997.

Singaraju, D., Grady, L., and Vidal, R. P-brush: Continuous valued mrfs with normed pairwise distributions for image segmentation. In *Proc. CVPR*, pp. 1303–1310, 2009.

Slepcev, D. and Thorpe, M. Analysis of p-Laplacian regularization in semisupervised learning. *SIAM J. Math. Anal.*, 51(3):2085–2120, 2019.

Spielman, D. A. and Srivastava, N. Graph sparsification by effective resistances. In *Proc. STOC*, pp. 563–568, 2008.

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

Spielman, D. A. and Teng, S.-H. Nearly linear time algorithms for preconditioning and solving symmetric, diagonally dominant linear systems. *SIAM J. Matrix Anal. Appl.*, 35(3):835–885, 2014.

Tron, R. and Vidal, R. A benchmark for the comparison of 3-d motion segmentation algorithms. In *Proc. CVPR*, pp. 1–8, 2007.

Tudisco, F. and Hein, M. A nodal domain theorem and a higher-order cheeger inequality for the graph p-Laplacian. *J. Spectr. Theory*, 8(3):883–908, 2018.

Tudisco, F., Mercado, P., and Hein, M. Community detection in networks via nonlinear modularity eigenvectors. *SIAM J. Appl. Math.*, 78(5):2393–2419, 2018.

Veldt, N., Klymko, C., and Gleich, D. F. Flow-based local graph clustering with better seed set inclusion. In *Proc. SDM*, pp. 378–386. SIAM, 2019.

von Luxburg, U. A tutorial on spectral clustering. *Stat. Comput.*, 17(4):395–416, 2007.

von Luxburg, U., Radl, A., and Hein, M. Getting lost in space: Large sample analysis of the resistance distance. In *Proc. NIPS*, pp. 2622–2630, 2010.

Yen, L., Vanvyve, D., Wouters, F., Fouss, F., Verleysen, M., Saerens, M., et al. Clustering using a random walk based distance measure. In *Proc. ESANN*, pp. 317–324, 2005.

Yen, L., Saerens, M., Mantrach, A., and Shimbo, M. A family of dissimilarity measures between nodes generalizing both the shortest-path and the commute-time distances. In *Proc. KDD*, pp. 785–793, 2008.

Zhang, D. Homological eigenvalues of graph p-Laplacians. *arXiv preprint arXiv:2110.06054*, 2021.

Zhou, D. and Schölkopf, B. Discrete regularization. In *Semi-supervised Learning*. MIT Press, 2006.

Zhou, D., Bousquet, O., Lal, T., Weston, J., and Schölkopf, B. Learning with local and global consistency. In *Proc. NIPS*, pp. 321–328, 2003.

Zhu, X., Ghahramani, Z., and Lafferty, J. D. Semi-supervised learning using gaussian fields and harmonic functions. In *Proc. ICML*, pp. 912–919, 2003.

A. On Limitation to Multi-class Spectral Clustering using Graph p-Laplacian

In Sec. 1 and Sec. 5, we claim that it is difficult to obtain the third or higher eigenpairs of graph p-Laplacian. This claim is a key motivation on why we take an alternate approach to the graph p-Laplacian. Thus, for completeness, this section briefly reviews graph p-Laplacian, and its difficulties to apply for multi-class clustering.

A.1. Eigenpairs of p-Laplacian

To start, following (Bühler & Hein, 2009) we introduce the graph p-Laplacian and its eigenpairs. The graph p-Laplacian Δ_p is defined as

$$(\Delta_p \mathbf{x})_i := \sum_{i,j \in V} a_{ij} |x_i - x_j|^{p-1} \mathrm{sgn}(x_i - x_j). \tag{22}$$

By this construction, we have $S_{G,p}(\mathbf{x}) = \langle \mathbf{x}, \Delta_p \mathbf{x} \rangle$. The k-th eigenpair $(\lambda_k, \mathbf{x}_k)$ of the p-Laplacian Δ_p satisfies

$$(\Delta_p \mathbf{x}_k)_i = \lambda_k |(\mathbf{x}_k)_i|^{p-1} \mathrm{sgn}((\mathbf{x}_k)_i), \forall i \in V. \tag{23}$$

Note that the first eigenpair is $(0, \mathbf{1})$. The eigenpairs are characterized by the critical values of the Rayleigh quotient.

Proposition A.1 (Tudisco & Hein (2018))**.** *The eigenpairs of graph p-Laplacian is a critical value and point of the Rayleigh quotient defined as*

$$R_{G,p}(\mathbf{x}) := \frac{S_{G,p}(\mathbf{x})}{\|\mathbf{x}\|_p^p}. \tag{24}$$

From this proposition, we can know that $R_{G,p}(a\mathbf{x}) = R_{G,p}(\mathbf{x})$, for $a \in \mathbb{R}$. Therefore, to consider the eigenpairs of the graph p-Laplacian, we can limit our interest to $\mathcal{S}_p := \{\mathbf{x} \mid \|\mathbf{x}\|_p^p = 1\}$, seeing the Rayleigh quotient in Eq. (24).

In the sequel, we briefly explain why we can obtain the second eigenpair and why it is difficult to obtain the third or higher eigenpairs of the graph p-Laplacian. We now define the following quotient,

$$R_{G,p}^{(2)}(\mathbf{x}) := \frac{S_{G,p}(\mathbf{x})}{\min_\eta \|\mathbf{x} - \eta \mathbf{1}\|_p^p}. \tag{25}$$

This quotient gives the second eigenpair of p-Laplacian.

Proposition A.2 (Bühler & Hein (2009))**.** *The global solution to Eq. (25) is given by $\mathbf{x}^* = \mathbf{x}_2 + \eta^* \mathbf{1}$, where $\eta^* = \arg\min_\eta \|\mathbf{x}_2 - \eta \mathbf{1}\|_p^p$, and $\mathbf{x}_2$ is the second p-eigenvector.*

This proposition shows that we have an exact identification for the second p-eigenpair; minimizing Eq. (25) gives the second p-eigenpair of Δ_p. However, we have not known yet the exact identification for third or higher eigenpair of p-Laplacian (Lindqvist, 2008).

While we do not know the identification such as Eq. (25) for the higher order eigenpairs, the next question is if there is characterization of eigenpairs of the graph p-Laplacian, such as "orthognality" for the $p = 2$ case. When $p = 2$, the eigenvectors of the graph Laplacian are further characterized from Prop. A.1; the eigenvectors of graph Laplacian are orthogonal to each other. By using orthogonality and Rayleigh-Ritz theorem, we can obtain the full eigenvectors of the graph 2-Laplacian as

Proposition A.3 (Classical, e.g.,von Luxburg (2007))**.** *Let $\mathbf{x}_1, \dots \mathbf{x}_{k-1}$ be eigenvectors of the graph Laplacian L. Then the k-th eigenvector $\mathbf{x}_k$ is given as*

$$\mathbf{x}_k = \arg\max_{\mathbf{x}} R_{G,2}(\mathbf{x}) \quad \text{s.t. } \mathbf{x}_k \perp \mathbf{x}_1, \dots, \mathbf{x}_{k-1}. \tag{26}$$

This proposition is the simplest form of the Courant's min-max theorem, and also called as *variational principle*. By this proposition and the sequence Eq. (26), we can easily obtain the higher order eigenvectors. This orthogonality constraint of eigenvectors comes from the nature of L^2 space induced from the graph 2-seminorm. However, we lose this sense of orthogonality if we expand to p-seminorm, since we lose the inner product structure in the L^p space.

In this context, the following further generalizes the "orthogonality" to graph p-Laplacian. To characterize the eigenpairs of graph p-Laplacian, we use *Krasnoselskii genus* γ for a set B;

$$\gamma(B) = \begin{cases} 0 \text{ if } B = \emptyset \\ \inf\{k \in \mathbf{Z}^+ \mid \exists \text{odd continuous } h : B \to \mathbf{R}^k \backslash \{0\}\} \\ \infty \text{ when no such } h \text{ exists } \forall j \in \mathbf{Z}^+ \end{cases} \tag{27}$$

This genus is a generalized concept of dimension of B. Using this genus, we can characterize the eigenpairs;

Proposition A.4 (Tudisco & Hein (2018)). *Consider the set of subsets $\mathcal{F}_k(\mathcal{S}_p) := \{B \subset \mathcal{S}_p \mid B = -B, \text{ closed }, \gamma(B) \geq k\}$. The sequence defined as*

$$\lambda_k = \min_{B \subset \mathcal{F}_k(\mathcal{S}_p)} \max_{\mathbf{x} \in B} R_p(\mathbf{x}) \tag{28}$$

gives a critical point of $R_p(\mathbf{x})$, whose corresponding $\mathbf{x}$ and λ_k constitute an eigenpair of Δ_p.

This proposition is the generalized Courant's min-max theorem Prop. A.3; when $p = 2$, this proposition corresponds to Prop. A.3. This proposition is also called as *variational principle*. The eigenvectors obtained by the sequence Eq. (28) is called as *variational eigenvectors*.

In this proposition, the space $\mathcal{F}_k(\mathcal{S}_p)$ serves as a generalized orthogonal k-dimensional space. Moreover, the sequence Eq. (28) may serve as a method to obtain eigenpairs in the sequential way. However, we have two issues for the practical use of this sequence. First problem is that due to the abstract characterization of Krasnoselskii genus, we do not know how we can *numerically* apply this genus to obtain the higher eigenvectors. When $p = 2$, this abstract characterization can be translated into the concrete and "numerically computable" characterization, "orthogonality". However, in the current form of Krasnoselskii genus given as Eq. (27), at this point we do not know how to numerically obtain this genus. Secondly, similarly to the continuous p-Laplacian theory (Lindqvist, 2008), we do not know in which condition this sequence yields exhaustive eigenpairs. For the tree (and the disconnected forest) case the sequence Eq. (28) exhausts all the spectra (Deidda et al., 2022; Zhang, 2021). On the other hand, for the complete graph case, it is shown that there are other eigenpairs than ones yielded by the sequence Eq. (28) (Amghibech, 2003). Despite these extensive studies, we are yet to understand in which conditions this sequence exhausts all the spectra of p-Laplacian. Thus, for a general graph we do not know if the variational eigenvalues are the same eigenvalues as the Rayleigh quotient would do in Prop. A.1.

To conclude the discussion above, while we know the identification for the second eigenpair of the graph p-Laplacian (Eq. (25)), we have three open problems related to the multi-class spectral clustering as follows;

1. We do not know the identification of the third or higher eigenpairs.
2. We do not know if the sequence Eq. (28) exhausts the spectra of the graph p-Laplacian for a general graph.
3. We do not know how to numerically obtain Krasnoselskii genus.

A.2. Cheeger Inequalities for graph p-Laplacian

This section discusses Cheeger inequalities for the graph p-Laplacian. The Cheeger inequality theoretically supports the use of the variational eigenvectors of p-Laplacian from the Cheeger cut point of view.

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

$$C(U) := \frac{\mathrm{Cut}(U, \overline{U})}{\min(|U|, |\overline{U}|)}, \quad \mathrm{Cut}(U, \overline{U}) := \sum_{i,j \in V} a_{ij} \tag{29}$$

We call the optimal cut $h_2 := \min_{U \subset V} C(U)$ as *Cheeger constant*. By recursively using this Cheeger cut, we define the multi-class Cheeger cut, which we call *k-way Cheeger constant* as

$$h_k := \min_{\{V_i\}_{i=1,\cdots,k}} \max_{j \in \{1,\ldots,k\}} C(V_j). \tag{30}$$

This k-way Cheeger constant can be seen as the smallest k-way Cheeger cut. To obtain the k-way Cheeger cut is known to be NP hard. However, relaxing into the real-value would ease this problem; this Cheeger cut can be approximated by the variational eigenvalues of the graph p-Laplacian by Cheeger inequality.

Before we discuss the Cheeger inequality, we need a setup of the nodal domain. A nodal domain is a maximally connected subgraph A of a graph G such that for $\mathbf{x} \in \mathbb{R}^n$ where A is either $\{i \mid x_i > 0\}$ or $\{i \mid x_i < 0\}$. With this idea, a nodal domain can be seen as a "partition" of the graph by the sign function. The number of the nodal domains of the variational eigenvectors is bounded, see (Tudisco & Hein, 2018).

This nodal domain is used in the Cheeger inequality for the variational eigenvectors of graph p-Laplacian as follows.

Proposition A.5 ((Tudisco & Hein, 2018))**.** *Let $(\lambda_k, \mathbf{x}_k)$ be a k-th eigenpair of Δ_p, obtained by the sequence Eq. (28). Let also ω_k be a number of nodal domains of $\mathbf{x}_k$. Then,*

$$\left(\max_i \frac{d_i}{2}\right)^{-(p-1)} \left(\frac{h_{\omega_k}}{p}\right)^p \leq \lambda_k \leq 2^{p-1} h_k$$

This proposition theoretically supports the use of the higher variational eigenvectors of the graph p-Laplacian for multi-way Cheeger Cut, since the eigenvalues can serve as an approximation of the k-way Cheeger constant.

For two-class spectral clustering, we are ready to use the second eigenvector of the graph p-Laplacian because we have theoretical supports (Prop. A.5) and we can numerically obtain by Prop. A.2. On the other hand, for multi-class spectral clustering, while we still have a theoretical supports (Prop. A.5), at this point we do not have a numerical way to obtain the higher eigenpairs due to open problems.

A.3. Existing Work for Multi-class Spectral Clustering Using Graph p-Laplacian

So far, we discuss limitations for spectral clustering using the graph p-Laplacian. This section discusses how the existing works "bypass" this limitation. In a rough sense, there are two ways to materialize the multi-class clustering using graph p-Laplacian; i) recursively bisectioning and ii) the use of the approximated orthogonality.

For i), the work (Bühler & Hein, 2009) proposed a multi-class clustering, which recursively bisections a graph by using Prop. A.2. Thus, the work (Bühler & Hein, 2009) partitions a subgraph when we partition further than two, which does not exploit the full structure of a graph. In fact, this way is known to lead to the suboptimal cut even when $p = 2$ (Simon & Teng, 1997).

The methods in line with ii), such as (Ding et al., 2019; Luo et al., 2010; Pasadakis et al., 2022), assume that the k eigenvectors of the graph p-Laplacian are close to ones of the graph 2-Laplacian, since the Rayleigh quotients $R_{G,p}$ and $R_{G,2}$ are similar. Using this assumption, these methods use optimization methods for the Rayleigh quotients $R_{G,p}$ with the initial conditions as the first k-eigenvectors of the graph 2-Laplacian. By these initial conditions, we expect that the obtained k-eigenvectors are "close" to the first k-eigenvectors of the graph 2-Laplacian, and thus we expect that the obtained k-eigenvectors are the first k-eigenvectors from the Rayleigh quotient $R_{G,p}$. These methods exploit approximated orthogonality proven in (Luo et al., 2010) of eigenvectors of graph p-Laplacian in order to achieve better algorithms. However, this assumption might be too strong, especially for the very large p or very small p, i.e., p close to 1. Moreover, even if this assumption may be reasonable, we do not know if the first k eigenvectors from Rayleigh quotient are the same set of vectors with the first k-eigenvectors obtained by the sequence Eq. (28) as we discussed in Sec. A.1. Now, recall that the Cheeger inequality guarantees the quality of the cut for the latter, the eigenvectors obtained by Eq. (28). Thus, at this point, we do not know if this assumption is suitable for multi-class clustering.

As a side, we remark that we have a different way to define graph p-energy and corresponding graph p-seminorm in many literature (Bougleux et al., 2007; 2009; Calder, 2018; Elmoataz et al., 2008; Singaraju et al., 2009; Zhou & Schölkopf, 2006). In this line, the p-energy is defined in "vertex-wise" way, which is written as

$$S_{G,p}^{VW}(\mathbf{x}) = \left(\sum_{i \in V} \left(\sum_{i \in V} a_{ij} |x_i - x_j|^2\right)^{1/2}\right)^p, \tag{31}$$

This definition sums the vertex-wise energy. For more details of the difference, see (Saito et al., 2018). For the corresponding p-Laplacian, we do not have an exact identification of higher eigenpairs either. Moreover, for the corresponding seminorm,

we have not theoretical characteristics yet, such as Cheeger inequality. Thus, for this graph p-Laplacian, we have less understanding than the graph p-Laplacian induced from the p-energy we used in the main text.

To conclude, there are three important open problems for the graph p-Laplacian. As we see in this section, these open problems are a key if we want to apply the graph p-Laplacian to multi-class clustering. Hence, without solving these open problems, we cannot say that it is theoretically guaranteed to use the graph p-Laplacian for multi-class clustering. However, these problems remain open not only in the graph domain, but also in the continuous domain, which has a longer history and wider research communities. Thus, in this study, for the purpose of multi-class clustering, we take an alternative approach to spectral clustering using the graph p-Laplacian. For more on the continuous p-Laplacian, see (Lindqvist, 2008), and on the theory of the graph p-Laplacian, refer to (Tudisco & Hein, 2018).

B. Additional Preliminary Definitions

First, we make an additional note for an intuition behind the analog between graph and electric circuit. In this analog, a vertex is a point at a circuit, and an edge is a resistor with resistance $1/a_{ij}$. A flow over a graph mapped to a current, and a distribution over V as $\mathbf{x}$ is seen as a potential at each vertex point. For the equations Eq. (4), the energy is defined as a sum of the inverse of resistance times square of the difference of the potential. The effective resistance between i and j is computed as follows; we inverse the energy that is minimized with the constraint that the difference of potential between i and j is unit. Given an electrical network the effective resistance between two vertices is the voltage difference needed to induce a unit "current" flow between the vertices i.e., it is resistance measured across the vertices.

Next, on top of the image for a matrix $M \in \mathbb{R}^{n_1 \times n_2}$, $\mathrm{Im}(M)$, we also define a *kernel*[2] of M, which is a subclass of $\mathbb{R}^n$, as

$$\mathrm{Ker}(M) := \{\mathbf{x} | M\mathbf{x} = 0, \mathbf{x} \in \mathbb{R}^n\}. \tag{32}$$

From the elementary result in the linear algebra area, we note that

$$\mathrm{Im}(M)^{\perp} = \mathrm{Ker}(M^{\top}), \tag{33}$$

where $\mathrm{Im}(M)^{\perp}$ is an orthogonal space to $\mathrm{Im}(M)$.

The matrix norm is *submultiplicative*, i.e., $|||M_1 M_2|||_p \leq |||M_1|||_p |||M_2|||_p$ whenever a product of matrices $M_1 M_2$ can be defined. A matrix norm is shown to be bounded as follows;

Lemma B.1 (Higham (1992)). *For a square matrix $M \in \mathbb{R}^{n_1 \times n_1}$, $\|M\|_p \leq n_1^{|1/2 - 1/p|} \|M\|_2$.*

Lemma B.2 (Higham (1992)). *For a matrix $M \in \mathbb{R}^{n_1 \times n_2}$, $\|M\|_p \leq \max(\|M\|_1, \|M\|_\infty)$.*

We elaborate more on Lemma B.2. For a symmetric matrix, since we have

$$|||M|||_1 = |||M|||_\infty = \max_j \sum_i |m_{ij}|. \tag{34}$$

From the Lemma B.2 and Eq.(34), for a symmetric matrix M, we have

$$|||M|||_p \leq |||M|||_1 = |||M|||_\infty. \tag{35}$$

By this we can bound $|||M|||_p$ by 1 or infinity norm of the matrix M.

An operator weighted matrix norm is defined for any matrix $M \in \mathbb{R}^{n_1 \times n_2}$ and weights $\mathbf{r}$ as

$$|||M|||_{\mathbf{r},p} := \sup_{\mathbf{x} \in \mathbb{R}^{n_2}} \frac{\|M\mathbf{x}\|_{\mathbf{r},p}}{\|\mathbf{x}\|_{\mathbf{r},p}}. \tag{36}$$

Recall that

$$\|\mathbf{x}\|_{\mathbf{r},p} = \|R^{1/p}\mathbf{x}\|_p, \tag{37}$$

where R is a diagonal matrix whose diagonal element is a weight of the norm.

[2]This kernel is a linear algebraic kernel, not a kernel function which often appears in the machine learning context.

From this definition, we can rewrite $|||M|||_{\mathbf{r},p}$ as

$$
\begin{aligned}
|||M|||_{\mathbf{r},p} &= \sup_{\mathbf{x}\in\mathbb{R}^{n_2}} \frac{\|M\mathbf{x}\|_{\mathbf{r},p}}{\|\mathbf{x}\|_{\mathbf{r},p}} \\
&= \sup_{\mathbf{x}\in\mathbb{R}^{n_2}} \frac{\|R^{1/p}M\mathbf{x}\|_p}{\|R^{1/p}\mathbf{x}\|_p} && (38)\\
&= \sup_{\mathbf{x}':=R^{1/p}\mathbf{x},\mathbf{x}\in\mathbb{R}^{n_2}} \frac{\|R^{1/p}MR^{-1/p}\mathbf{x}'\|_p}{\|\mathbf{x}'\|_p} && (39)\\
&= \sup_{\mathbf{x}'\in\mathbb{R}^{n_2}} \frac{\|R^{1/p}MR^{-1/p}\mathbf{x}'\|_p}{\|\mathbf{x}'\|_p} && (40)\\
&= \sup_{\mathbf{x}\in\mathbb{R}^{n_2}} \frac{\|R^{1/p}MR^{-1/p}\mathbf{x}\|_p}{\|\mathbf{x}\|_p} && (41)\\
&= |||R^{1/p}MR^{-1/p}|||_p, && (42)
\end{aligned}
$$

C. Details of Lemma 3.1

This section provides detailed explanation on Lemma 3.1. We first note that the trick in this transformation is as same as done in (Herbster & Pontil, 2006; Klein & Randić, 1993). The elaboration here follows these earlier works.

Using the reproducing property Eq. (3) as done in (Herbster & Pontil, 2006; Klein & Randić, 1993), the constraints of 2-resistance (and also for p-resistance Eq. (7)) can be rewritten as

$$
1 = x_i - x_j = \langle \mathbf{x}, L^+\mathbf{e}_i - L^+\mathbf{e}_j\rangle_L. \tag{43}
$$

Now, since $L\mathbf{1} = 0$, there exists $c \in \mathbb{R}$ such that $\mathbf{x} - c\mathbf{1} \in \mathcal{H}(L)$. We now define as $\mathbf{x}' := \mathbf{x} - c\mathbf{1}$. We then compute

$$
\begin{aligned}
\langle \mathbf{x}', L^+\mathbf{e}_i\rangle_L &= \mathbf{x}'^\top LL^+\mathbf{e}_i && (44)\\
&= \mathbf{x}'^\top \mathbf{e}_i && (45)\\
&= x'_i. && (46)
\end{aligned}
$$

The second line follows since we have $LL^+\mathbf{x}' = \mathbf{x}'$ for $\mathbf{x}' \in \mathcal{H}(L)$. Note that this computation is same as the reproducing kernel property characteristics Eq. (3). Also, the definition of $\mathbf{x}'$ immediately leads to

$$
L\mathbf{x}' = L(\mathbf{x} - c\mathbf{1}) = L\mathbf{x} - cL\mathbf{1} = L\mathbf{x}, \tag{47}
$$

and thus for $\mathbf{u} \in \mathbb{R}^n$ we have

$$
\begin{aligned}
\langle \mathbf{x}', \mathbf{u}\rangle_L &= \mathbf{x}'^\top L\mathbf{u} && (48)\\
&= \mathbf{x}^\top L\mathbf{u} && (49)\\
&= \langle \mathbf{x}, \mathbf{u}\rangle_L && (50)
\end{aligned}
$$

From these discussions, we obtain

$$
\begin{aligned}
1 &= x_i - x_j && (51)\\
&= (x_i - c) - (x_j - c) && (52)\\
&= x'_i - x'_j && (53)\\
&= \langle \mathbf{x}', L^+\mathbf{e}_i\rangle_L - \langle \mathbf{x}', L^+\mathbf{e}_j\rangle_L && (54)\\
&= \langle \mathbf{x}, L^+\mathbf{e}_i\rangle_L - \langle \mathbf{x}, L^+\mathbf{e}_j\rangle_L && (55)\\
&= \langle \mathbf{x}, L^+\mathbf{e}_i - L^+\mathbf{e}_j\rangle_L. && (56)
\end{aligned}
$$

The third line follows from the definition of $\mathbf{x}'$. The fourth line follows from Eq. (46) and the fifth line follows from Eq. (50). Thus, we obtain Eq. (43)

D. Proof for Proposition 3.2

For the proof we divide the proof into lower bound, upper bound and equal condition.

D.1. Lower Bound

In the following we give a proof for this Proposition. From Hölder's inequality, we have

$$\langle \mathbf{x}, \mathbf{y} \rangle_L \leq \|\mathbf{x}\|_{G,p} \|\mathbf{y}\|_{G,q} \tag{57}$$

Assuming $\langle \mathbf{x}, \mathbf{y} \rangle_L = 1$, we have

$$1 \leq \|\mathbf{x}\|_{G,p} \|\mathbf{y}\|_{G,q}, \tag{58}$$

and hence

$$\|\mathbf{y}\|_{G,q}^{-p} \leq \|\mathbf{x}\|_{G,p}^{p}, \tag{59}$$

which proves the lower bound.

D.2. Upper Bound

This section proves the upper bound of Prop. 3.2.

Recall the variable of the minimization problem in Eq. (15) is $\mathbf{x}$. If we prove that when $\mathbf{x} = \mathbf{z}$, this $\mathbf{z}$ satisfies the condition in the minimization problem $\langle \mathbf{z}, \mathbf{y} \rangle_L = 1$, from the minimization problem nature we can prove the upper bound. For this strategy, we use the following lemma.

Lemma D.1. *For $\boldsymbol{\alpha} \in \mathbb{R}^n$ and $\boldsymbol{\beta} \in \mathbb{R}^m$, we have*

$$\langle C\boldsymbol{\alpha}, \boldsymbol{\beta} \rangle_{\mathbf{w}} = \langle C\boldsymbol{\alpha}, \boldsymbol{\beta}' \rangle_{\mathbf{w}} \tag{60}$$

where

$$\boldsymbol{\beta} := \boldsymbol{\beta}' + \boldsymbol{\beta}'', \quad \boldsymbol{\beta}' := CC^{+}\boldsymbol{\beta}, \boldsymbol{\beta}'' := (I - CC^{+})\boldsymbol{\beta} \tag{61}$$

For readability, we move the proof of Lemma D.1 to Sec. D.4.2.

By using Lemma D.1, we now prove the upper bound.

Note that

$$\frac{f_{q/p}(C\mathbf{y})}{\|C\mathbf{y}\|_{\mathbf{w},q}^{q}} = CC^{+}\frac{f_{q/p}(C\mathbf{y})}{\|C\mathbf{y}\|_{\mathbf{w},q}^{q}} + (I - CC^{+})\frac{f_{q/p}(C\mathbf{y})}{\|C\mathbf{y}\|_{\mathbf{w},q}^{q}}. \tag{62}$$

Recall that we define as

$$\mathbf{z} := C^{+}\frac{f_{q/p}(C\mathbf{y})}{\|\mathbf{y}\|_{G,q}^{q}} = C^{+}\frac{f_{q/p}(C\mathbf{y})}{\|C\mathbf{y}\|_{\mathbf{w},q}^{q}}, \quad (f_{\theta}(\mathbf{x}))_i := \mathrm{sgn}(x_i)|x_i|^{\theta}. \tag{63}$$

By using this relation and Lemma D.1, we have

$$\begin{aligned}
\langle \mathbf{z}, \mathbf{y} \rangle_L &= \langle C\mathbf{z}, C\mathbf{y} \rangle_{\mathbf{w}} && (64)\\
&= \langle C\mathbf{y}, C\mathbf{z} \rangle_{\mathbf{w}} && (65)\\
&= \langle C\mathbf{y}, CC^{+}\frac{f_{q/p}(C\mathbf{y})}{\|C\mathbf{y}\|_{\mathbf{w},q}^{q}} \rangle_{\mathbf{w}} && (66)\\
&= \langle C\mathbf{y}, \frac{f_{q/p}(C\mathbf{y})}{\|C\mathbf{y}\|_{\mathbf{w},q}^{q}} \rangle_{\mathbf{w}} && (67)\\
&= 1. && (68)
\end{aligned}$$

From Eq. (66) to Eq. (67) we apply Lemma D.1. The last equality comes from the same nature of Eq. (81) in Prop. D.3, which we will discuss in Sec. D.4.1.

From the discussion above, $\mathbf{z}$ satisfies the condition of the minimization problem of Eq. (15). Therefore, we obtain

$$\min_{\mathbf{x}}\{\|\mathbf{x}\|_{G,p}^p \text{ s.t. } \langle \mathbf{y}, \mathbf{x}\rangle_L = 1\} \leq \|\mathbf{z}\|_{G,p}^p \tag{69}$$

D.3. Proof for the Equal Condition

Finally, we turn into the equality condition. We obtain the following lemma.

Lemma D.2. *For any p, q such that $1/p + 1/q = 1$, we have*

$$\min_{\mathbf{x}}\{\|\mathbf{x}\|_{G,p}^p \text{s.t. } \langle \mathbf{x}, \mathbf{y}\rangle_L = 1\} = \|\mathbf{z}\|_{G,q}^p = \|\mathbf{y}\|_{G,q}^{-p}, \tag{70}$$

where

$$\mathbf{z} := C^+ \frac{f_{q/p}(C\mathbf{y})}{\|\mathbf{y}\|_{G,q}^q} = C^+ \frac{f_{q/p}(C\mathbf{y})}{\|C\mathbf{y}\|_{\mathbf{w},q}^q}, \tag{71}$$

when $f_{q/p}(C\mathbf{y}) \in \mathrm{Im}(C)$

The proof of Lemma D.2 is given in Sec. D.4.3

D.4. Proof for Lemma D.1 and Lemma D.2

This section provides proofs for Lemma D.1 and Lemma D.2. These lemmas are critical components for the proof for Prop. 3.2. In order to enhance the readability, we gather proofs for these claims in this section. We first give auxiliary lemmas, that hold for the general setting. Then, using these auxiliary lemmas, we provide the proofs for Lemma D.1 and Lemma D.2.

D.4.1. Auxiliary Lemmas

This section provides auxiliary lemmas. We start with the following claim.

Proposition D.3. *For any $p, q > 1$ such that $1/p + 1/q = 1$, we have*

$$\min_{\mathbf{x}}\{\|\mathbf{x}\|_{\mathbf{r},p}^p \text{ s.t. } \langle \mathbf{y}, \mathbf{x}\rangle_{\mathbf{r}} = 1\} = \|\mathbf{y}\|_{\mathbf{r},q}^{-p} \tag{72}$$

Proof. Using the Hölder's inequality, we get

$$\|\mathbf{y}\|_{\mathbf{r},q}\|\mathbf{x}\|_{\mathbf{r},p} \geq \langle \mathbf{x}, \mathbf{y}\rangle_{\mathbf{r}} \tag{73}$$

Assuming $\langle \mathbf{x}, \mathbf{y}\rangle_{\mathbf{r}} = 1$, we can rearrange as

$$\|\mathbf{x}\|_{\mathbf{r},p} \geq \|\mathbf{y}\|_{\mathbf{r},q}^{-1}. \tag{74}$$

Now we consider when the minimum of the right hand side of Eq. (74). The minimum with the assumption $\langle \mathbf{y}, \mathbf{x}\rangle_{\mathbf{r}} = 1$ is achieved when $\mathbf{x} = \boldsymbol{\zeta}$ such that

$$\boldsymbol{\zeta} := \frac{f_{q/p}(\mathbf{y})}{\|\mathbf{y}\|_{\mathbf{r},q}^q}, \quad (f_\theta(\mathbf{y}))_i := \mathrm{sgn}(y_i)|y_i|^\theta \tag{75}$$

which means

$$\zeta_i = \frac{\mathrm{sgn}(y_i)|y_i|^{q/p}}{\|\mathbf{y}\|_{\mathbf{r},q}^q}. \tag{76}$$

For this $\boldsymbol{\zeta}$, we compute

$$\begin{aligned}
\langle \mathbf{y}, \boldsymbol{\zeta} \rangle_{\mathbf{r}} &= \sum_{i=1}^{n} r_i y_i \zeta_i && (77)\\
&= \sum_{i=1}^{n} r_i y_i \frac{\mathrm{sgn}(y_i)|y_i|^{q/p}}{\|\mathbf{y}\|_{\mathbf{r},q}^{q}} && (78)\\
&= \frac{\sum_{i=1}^{n} r_i |y_i|^{q/p+1}}{\|\mathbf{y}\|_{\mathbf{r},q}^{q}} && (79)\\
&= \frac{\sum_{i=1}^{n} r_i |y_i|^{q}}{\|\mathbf{y}\|_{\mathbf{r},q}^{q}} && (80)\\
&= \frac{\|\mathbf{y}\|_{\mathbf{r},q}^{q}}{\|\mathbf{y}\|_{\mathbf{r},q}^{q}} = 1. && (81)
\end{aligned}$$

The transition from Eq. (79) to Eq. (80) comes from $q/p + 1 = q$. Also, we have

$$\begin{aligned}
\|\boldsymbol{\zeta}\|_{\mathbf{r},p} &= \left\| \frac{f_{q/p}(\mathbf{y})}{\|\mathbf{y}\|_{\mathbf{r},q}^{q}} \right\|_{\mathbf{r},p} && (82)\\
&= \left(\sum_{i=1}^{n} r_i \left| \frac{\mathrm{sgn}(y_i)|y_i|^{q/p}}{\|\mathbf{y}\|_{\mathbf{r},q}^{q}} \right|^{p} \right)^{1/p} && (83)\\
&= \frac{1}{\|\mathbf{y}\|_{\mathbf{r},q}^{q}} \left(\sum_{i=1}^{n} r_i \left| \mathrm{sgn}(y_i)|y_i|^{q/p} \right|^{p} \right)^{1/p} && (84)\\
&= \frac{1}{\|\mathbf{y}\|_{\mathbf{r},q}^{q}} \left(\sum_{i=1}^{n} r_i \, |y_i|^{q} \right)^{1/p} && (85)\\
&= \frac{1}{\|\mathbf{y}\|_{\mathbf{r},q}^{q}} \|\mathbf{y}\|_{\mathbf{r},q}^{q/p} && (86)\\
&= \|\mathbf{y}\|_{\mathbf{r},q}^{q/p-q} = \|\mathbf{y}\|_{\mathbf{r},q}^{-1}. && (87)
\end{aligned}$$

By substituting $\mathbf{x} = \boldsymbol{\zeta}$ in Eq. (74), the assumption $\langle \mathbf{y}, \mathbf{x} \rangle_{\mathbf{r}} = 1$ is satisfied and the equality holds. Thus, we obtain

$$\|\mathbf{x}\|_{\mathbf{r},p} \geq \|\mathbf{y}\|_{\mathbf{r},q}^{-1} \iff \|\mathbf{x}\|_{\mathbf{r},p}^{p} \geq \|\mathbf{y}\|_{\mathbf{r},q}^{-p}, \tag{88}$$

where the equality holds when $\mathbf{x} = \boldsymbol{\zeta}$. □

We bring another lemma about spaces spanned by matrices.

Lemma D.4 ((Ben-Israel & Greville, 2003) Ex.9, §1.3, p.43 & §2.6, p.71)**.** *For a matrix $M \in \mathbb{R}^{n_1 \times n_2}$, we define a generalized inverse of matrix M denoted by $M^{\dagger} \in \mathbb{R}^{n_2 \times n_1}$, satisfying that*

$$MM^{\dagger}M = M. \tag{89}$$

Then,

$$\mathrm{Im}(M) = \mathrm{Im}(MM^{\dagger}). \tag{90}$$

Also,

$$S = \{\mathbf{y} : \mathbf{y} = (I - MM^{\dagger})\mathbf{x}, \mathbf{x} \in \mathbb{R}^{n_1}\} \subseteq \mathrm{Im}(M)^{\perp}. \tag{91}$$

Note that the generalized inverse $M^{\dagger}$ is not unique. However, the pseudoinverse M^{+} is unique, and also be one of generalized inverses $M^{\dagger}$. From this lemma, we can write as

$$\mathrm{Im}(M) = \{\mathbf{a} : \mathbf{a} = MM^{\dagger}\mathbf{b}, \mathbf{b} \in \mathbb{R}^{n_1}\}, \quad \mathrm{Im}(M)^{\perp} \supseteq \{\mathbf{a} : \mathbf{y} = (I - MM^{\dagger})\mathbf{b}, \mathbf{b} \in \mathbb{R}^{n_1}\}. \tag{92}$$

D.4.2. Proof for Lemma D.1

This section provides a proof for Lemma D.1.

For the illustrative purpose, we start with the $\mathbf{w} = \mathbf{1}$ case. If $\mathbf{w} = \mathbf{1}$, then

$$\begin{aligned}\langle C\boldsymbol{\alpha}, \boldsymbol{\beta}\rangle_{\mathbf{w}} &= \langle C\boldsymbol{\alpha}, \boldsymbol{\beta}' + \boldsymbol{\beta}''\rangle_{\mathbf{w}} && (93)\\ &= \langle C\boldsymbol{\alpha}, \boldsymbol{\beta}'\rangle_{\mathbf{w}} + \langle C\boldsymbol{\alpha}, \boldsymbol{\beta}''\rangle_{\mathbf{w}} && (94)\\ &= \langle C\boldsymbol{\alpha}, \boldsymbol{\beta}'\rangle_{\mathbf{w}}. && (95)\end{aligned}$$

The last equality follows for the following reason. By composition, $C\boldsymbol{\alpha} \in \mathrm{Im}(C)$. Also, from Lemma D.4, $\boldsymbol{\beta}'' \in \mathrm{Im}(C)^{\perp}$. Hence, $C\boldsymbol{\alpha}$ and $\boldsymbol{\beta}''$ are orthogonal to each other and we get $\langle C\boldsymbol{\alpha}, \boldsymbol{\beta}''\rangle_{\mathbf{w}} = 0$.

We now turn into the case where $\mathbf{w}$ is arbitrary. Things are less trivial when we introduce the weight. Thus, we further analyze the weighted inner product.

Since the matrix W is a full rank diagonal matrix, we obtain

$$W^{1/2}C(C^{+}W^{-1/2})W^{1/2}C = W^{1/2}CC^{+}C = W^{1/2}C, \tag{96}$$

and thus $C^{+}W^{-1/2}$ is a generalized inverse of $W^{1/2}C$, i.e.,

$$(W^{1/2}C)^{\dagger} = C^{+}W^{-1/2}. \tag{97}$$

Also, since W is a full rank diagonal matrix,

$$\{\mathbf{b} : \mathbf{b} = W^{-1/2}\mathbf{a}, \mathbf{a} \in \mathbb{R}^m\} = \{\mathbf{b}' : \mathbf{b}' \in \mathbb{R}^m\} = \mathbb{R}^m. \tag{98}$$

Using these relations and Lemma D.4, we get

$$\begin{aligned}\mathrm{Im}(W^{1/2}C) &= \{\mathbf{b} : \mathbf{b} = W^{1/2}C(W^{1/2}C)^{\dagger}\mathbf{a}, \mathbf{a} \in \mathbb{R}^m\} && (99)\\ &= \{\mathbf{b} : \mathbf{b} = W^{1/2}CC^{+}W^{-1/2}\mathbf{a}, \mathbf{a} \in \mathbb{R}^m\} && (100)\\ &= \{\mathbf{b} : \mathbf{b} = W^{1/2}CC^{+}\mathbf{a}', \mathbf{a}' \in \mathbb{R}^m\}, && (101)\end{aligned}$$

where we use Eq. (97) for Eq. (100) and we use Eq. (98) for Eq. (101). Moreover, we have

$$\begin{aligned}\mathrm{Im}(W^{1/2}C)^{\perp} &\supseteq \{\mathbf{b} : \mathbf{b} = (I - W^{1/2}C(W^{1/2}C)^{\dagger})\mathbf{a}, \mathbf{a} \in \mathbb{R}^m\} && (102)\\ &= \{\mathbf{b} : \mathbf{b} = (I - W^{1/2}CC^{+}W^{-1/2})\mathbf{a}, \mathbf{a} \in \mathbb{R}^m\} && (103)\\ &= \{\mathbf{b} : \mathbf{b} = W^{1/2}(I - CC^{+})W^{-1/2}\mathbf{a}, \mathbf{a} \in \mathbb{R}^m\} && (104)\\ &= \{\mathbf{b} : \mathbf{b} = W^{1/2}(I - CC^{+})\mathbf{a}', \mathbf{a}' \in \mathbb{R}^m\}, && (105)\end{aligned}$$

where we use Eq. (97) for Eq. (103) and we use Eq. (98) for Eq. (105). Therefore,

$$\begin{aligned}\langle C\boldsymbol{\alpha}, \boldsymbol{\beta}\rangle_{\mathbf{w}} &= \langle W^{1/2}C\boldsymbol{\alpha}, W^{1/2}\boldsymbol{\beta}\rangle && (106)\\ &= \langle W^{1/2}C\boldsymbol{\alpha}, W^{1/2}\boldsymbol{\beta}'\rangle + \langle W^{1/2}C\boldsymbol{\alpha}, W^{1/2}\boldsymbol{\beta}''\rangle && (107)\\ &= \langle W^{1/2}C\boldsymbol{\alpha}, W^{1/2}CC^{+}\boldsymbol{\beta}\rangle + \langle W^{1/2}C\boldsymbol{\alpha}, W^{1/2}(I - CC^{+})\boldsymbol{\beta}\rangle && (108)\\ &= \langle W^{1/2}C\boldsymbol{\alpha}, W^{1/2}CC^{+}\boldsymbol{\beta}\rangle. && (109)\\ &= \langle W^{1/2}C\boldsymbol{\alpha}, W^{1/2}\boldsymbol{\beta}'\rangle. && (110)\\ &= \langle C\boldsymbol{\alpha}, \boldsymbol{\beta}'\rangle_{\mathbf{w}} && (111)\end{aligned}$$

The line Eq. (109) follows because from Eq. (105) the $W^{1/2}(I - CC^{+})\boldsymbol{\beta} \in \mathrm{Im}(W^{1/2}C)^{\perp}$ and therefore $W^{1/2}(I - CC^{+})\boldsymbol{\beta}$ is orthogonal to $W^{1/2}C\boldsymbol{\alpha}$, which induces $\langle W^{1/2}C\boldsymbol{\alpha}, W^{1/2}(I - CC^{+})\boldsymbol{\beta}\rangle = 0$.

Eq. (111) concludes the proof.

D.4.3. PROOF FOR LEMMA D.2

This section proves Lemma D.2.

If $f_{q/p}(C\mathbf{y}) \in \mathrm{Im}(C)$, then

$$C\mathbf{z} = CC^{+}\frac{f_{q/p}(C\mathbf{y})}{\|\mathbf{y}\|_{G,q}^{q}} \tag{112}$$

$$= \frac{f_{q/p}(C\mathbf{y})}{\|\mathbf{y}\|_{G,q}^{q}} \tag{113}$$

$$= \frac{f_{q/p}(C\mathbf{y})}{\|C\mathbf{y}\|_{\mathbf{w},q}^{q}} \tag{114}$$

From Eq. (112) to Eq. (113), we use the following relation; for a vector $\mathbf{a}\in\mathrm{Im}(C)$ we have $\mathbf{a} = CC^{+}\mathbf{a}$ since CC^{+} is an orthogonal projection onto the space $\mathrm{Im}(C)$. Eq. (114) is a form of Eq. (75), and thus from Prop. D.3, Eq. (114) satisfies the equality condition of the Hölder's inequality as

$$\|C\mathbf{z}\|_{\mathbf{w},p}^{p} = \|C\mathbf{y}\|_{\mathbf{w},q}^{-p} \iff \|\mathbf{z}\|_{G,p}^{p} = \|\mathbf{y}\|_{G,q}^{-p}, \tag{115}$$

where we use the definition of the graph p-seminorm. Thus, we obtain the claim.

E. Proofs for Theorem 3.4 and Theorem 3.3

In this section we prove Thm. 3.4 and Thm. 3.3. The general strategy is applying Prop. 3.2. We first prove the general case of Thm. 3.3.

E.1. Proof for Theorem 3.3

By definition,

$$r_{G,p}(i,j) = \frac{1}{\min_{\mathbf{x}} \|\mathbf{x}\|_{G,p}^{p} \text{ s.t. } x_i - x_j = 1}. \tag{116}$$

First, we rewrite the condition of the minimization problem. Using Eq. (43), we observe that the denominator of Eq.(8) can be written as

$$\min_{\mathbf{x}}\{\|\mathbf{x}\|_{G,p}^{p} \text{ s.t. } x_i - x_j = 1\} = \min_{\mathbf{x}}\{\|\mathbf{x}\|_{G,p}^{p} \text{ s.t. } \langle L^{+}\mathbf{e}_i - L^{+}\mathbf{e}_j, \mathbf{x}\rangle_L = 1\} \tag{117}$$

From this rewrite, we see that Eq. (117) is exactly same as the minimization problem of Prop. 3.2 if we substitute $\mathbf{y} := L^{+}(\mathbf{e}_i - \mathbf{e}_j)$. Thus, we apply Prop. 3.2 to this problem in order to obtain lower and upper bounds of Eq. (117).

Lower Bound of Eq. (117). Now, we come to the lower bound of this problem Eq. (117). By applying the lower bound of Prop. 3.2 with substituting $\mathbf{y} := L^{+}(\mathbf{e}_i - \mathbf{e}_j)$, we obtain

$$\|L^{+}(\mathbf{e}_i - \mathbf{e}_j)\|_{G,q}^{-p} \leq \min_{\mathbf{x}}\{\|\mathbf{x}\|_{G,p}^{p} \text{ s.t. } \langle L^{+}(\mathbf{e}_i - \mathbf{e}_j), \mathbf{x}\rangle_L = 1\}. \tag{118}$$

This conclude the lower bound.

Upper Bound of Eq. (117). Next, we turn to the upper bound of this problem Eq. (117).

We first compute

$$\|\mathbf{z}\|_{G,p} = \|C\mathbf{z}\|_{\mathbf{w},p} = \left\| \frac{CC^+ f_{q/p}(C\mathbf{y})}{\|C\mathbf{y}\|_{\mathbf{w},q}^q} \right\|_{\mathbf{w},p} \tag{119}$$

$$= \frac{\left\|CC^+ f_{q/p}(C\mathbf{y})\right\|_{\mathbf{w},p}}{\|C\mathbf{y}\|_{\mathbf{w},q}^q} \tag{120}$$

$$\leq |||CC^+|||_{\mathbf{w},p} \frac{\|f_{q/p}(C\mathbf{y})\|_{\mathbf{w},p}}{\|C\mathbf{y}\|_{\mathbf{w},q}^q} \tag{121}$$

$$= |||CC^+|||_{\mathbf{w},p} \|C\mathbf{y}\|_{\mathbf{w},q}^{-1} \tag{122}$$

$$= |||W^{1/p}CC^+W^{-1/p}|||_p \|C\mathbf{y}\|_{\mathbf{w},q}^{-1} \tag{123}$$

$$= |||W^{1/p}CC^+W^{-1/p}|||_p \|\mathbf{y}\|_{G,q}^{-1} \tag{124}$$

$$= \alpha_{G,p} \|\mathbf{y}\|_{G,q}^{-1}, \tag{125}$$

where we recall that we defined as $\alpha_{G,p} := |||W^{1/p}CC^+W^{-1/p}|||_p$. The transformation from Eq. (120) to Eq. (121) follows from the submultiplicative characteristics of the matrix norm discussed in Sec. 2. The equality from Eq. (121) to Eq. (122) holds due to the same discussion as Eq. (87) in Prop. D.3, which we discussed in Sec. D.4.1. The transformation from Eq. (122) to Eq. (123) follows from a characteristics of the weighted matrix norm discussed in Eq. (40). Hence, by taking the p-th power of the inequality Eq. (125) and observing that we substitute $\mathbf{y} := L^+(\mathbf{e}_i - \mathbf{e}_j)$, we obtain

$$\|\mathbf{z}\|_{G,p}^p \leq \alpha_{G,p}^p \|L^+(\mathbf{e}_i - \mathbf{e}_j)\|_{G,q}^{-p} \tag{126}$$

Thus, from Prop. 3.2 and the inequality Eq. (125) we get

$$\min_{\mathbf{x}}\{\|\mathbf{x}\|_{G,p}^p \text{ s.t. } \langle L^+(\mathbf{e}_i - \mathbf{e}_j), \mathbf{x}\rangle_L = 1\} \leq \|\mathbf{z}\|_{G,p}^p \leq \alpha_{G,p}^p \|L^+(\mathbf{e}_i - \mathbf{e}_j)\|_{G,q}^{-p}. \tag{127}$$

Combining Lower and Upper Bounds of Eq. (117). We now combine the lower bound Eq. (118) and the upper bound Eq. (127). By combining these two and using Eq. (117), we get

$$\begin{aligned} &\|L^+\mathbf{e}_i - L^+\mathbf{e}_j\|_{G,q}^{-p} \leq \min_{\mathbf{x}}\{\|\mathbf{x}\|_{G,p}^p \text{ s.t. } \langle L^+(\mathbf{e}_i - \mathbf{e}_j), \mathbf{x}\rangle_L = 1\} \leq \alpha_{G,p}^p \|L^+\mathbf{e}_i - L^+\mathbf{e}_j\|_{G,q}^{-p} \\ \iff &\|L^+\mathbf{e}_i - L^+\mathbf{e}_j\|_{G,q}^{-p} \leq \min_{\mathbf{x}}\{\|\mathbf{x}\|_{G,p}^p \text{ s.t. } x_i - x_j = 1\} \leq \alpha_{G,p}^p \|L^+\mathbf{e}_i - L^+\mathbf{e}_j\|_{G,q}^{-p} \end{aligned} \tag{128}$$

For the p-effective resistance, taking the inverse we obtain

$$\frac{1}{\alpha_{G,p}^p} \|L^+\mathbf{e}_i - L^+\mathbf{e}_j\|_{G,q}^p \leq r_{G,p}(i,j) \leq \|L^+\mathbf{e}_i - L^+\mathbf{e}_j\|_{G,q}^p. \tag{129}$$

E.2. Proof for Theorem 3.4

For the incidence matrix of tree, $\mathrm{rank}(C) = n-1$ (Bapat, 2010). Hence $\mathrm{Im}(C) = \mathbb{R}^{n-1}$. Thus, $f_{q/p}(C\mathbf{y}) \in \mathbb{R}^{n-1} = \mathrm{Im}(C)$. Using the Lemma D.2 and substituting $\mathbf{y} = L^+\mathbf{e}_i - L^+\mathbf{e}_j$,

$$\min_{\mathbf{x}}\{\|\mathbf{x}\|_{G,p}^p \text{s.t. } \langle \mathbf{x}, L^+\mathbf{e}_i - L^+\mathbf{e}_j\rangle_L = 1\} = \|L^+\mathbf{e}_i - L^+\mathbf{e}_j\|_{G,q}^{-p}. \tag{130}$$

Recall that the minimization problem of Eq. (130) is the inverse of the p-resistance. Therefore, Eq. (130) leads to the claim.

F. Proof for Proposition. 3.5

We recall that by definition of pseudoinverse, we have

$$|||CC^+|||_2 = 1, \tag{131}$$

since the eigenvalues of CC^+ is either 0 or 1. Also, for any matrix M and any invertible matrix P, PMP^{-1} and M share the same eigenvalues. By construction, W is also an invertible matrix. Thus, using Lemma B.1, we obtain

$$\alpha_{G,p} = |||W^{1/p}CC^+W^{-1/p}|||_p \tag{132}$$
$$\leq m^{|1/2-1/p|}|||W^{1/p}CC^+W^{-1/p}|||_2 \tag{133}$$
$$= m^{|1/2-1/p|}. \tag{134}$$

G. Illustrative Examples of Clustering via p-resistance Fig. 1

This section explains illustrative examples of clustering via p-resistance where p plays a role.

G.1. Preliminaries for Illustrative Examples

Before we discuss the details of the clustering, we setup preliminaries. We now setup the notions on the graph metrics. First, a *st-mincut* is defined as the minimum cut between the vertices s and t, i.e.,

$$\min_{V'} \mathrm{Cut}(s,t) := \min_{V'} \sum_{i\in V', j\in V'\setminus V | s\in V', t\in V'\setminus V} a_{ij}. \tag{135}$$

The act of the "cut" of the edges is defined to divide into two graphs so that the vertex s belongs to one and the vertex t belongs to the other. The minimum cut is that we want such a cut so that the sum of the weight of the edges to be cut is minimized.

Now, we also define the shortest path between vertices s and t is defined as

$$\min_{\mathbf{i}} \sum_{\ell\in E} w_\ell i_\ell \text{ s.t. } \mathbf{i} = (i_\ell)_{\ell\in E} \text{ unit flow from } i \text{ to } j, \tag{136}$$

where $\mathbf{i} \in \{0,1\}^m$. The shortest path problem is to finding the path with smallest sum of the weights of edges between s and t.

In the following, we show that p-resistance is connection with st-mincut and the shortest path. We recall the theorem in (Alamgir & Luxburg, 2011) as

Proposition G.1 (Alamgir & Luxburg (2011))**.** *Consider a p-flow problem as*

$$F_{G,p}(i,j) := \min_{\mathbf{i}} \sum_{\ell\in E} w_\ell^{1-p} i_\ell^p \text{ s.t. } \mathbf{i} = (i_\ell)_{\ell\in E} \text{ unit flow from } i \text{ to } j, \tag{137}$$

where $\mathbf{i} \in \mathbb{R}^{+m}$ is a current at edges. Then, for $1/p + 1/q = 1$, we have

$$r_{G,p}^{1/(p-1)}(i,j) = F_{G,q}(i,j). \tag{138}$$

We first remark that $\mathbf{i}$ in q-flow problem is non-negative real value whereas $\mathbf{i}$ for the shortest path is either 0 or 1. We remark that when $p \to \infty$, q goes to 1 and q-flow problem is a simple shortest path flow problem.

This proposition means that the $1/(p-1)$-th power of p-resistance is equivalent to the q-flow. From this proposition, we now see the connection between p-resistance, and st-mincut and shortest path as follows.

- When $p \to 1$, p-resistance between s and t is 1/st-mincut.
- When $p \to \infty$, $1/(p-1)$-th power of the p-resistance is the discrete shortest path of the *unweighted* graph.

Thus, we intuitively characterize the p-resistance as

- When p is small, p-resistance more focus on a minimum cut.
- When p is large, p-resistance more focus on the "path", and also more focus on the "unweighted topology".

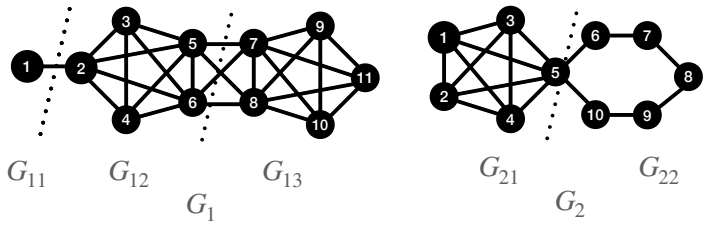


Figure 3: The notations of illustrative example graphs. In the graph G_2 the vertex 5 is in both G_{21} and G_{22}.

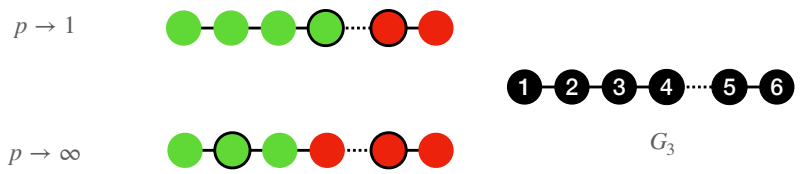


Figure 4: The illustrative example of a weighted graph and its notations. The weights of edge drawn in the line are 1, whereas weight of the dotted line is $\epsilon \ll 1$. The other drawing rule follows Fig. 1. In the example, we observe that we focus on the difference of the weight when $p \to 1$, while we ignore the weight when $p \to \infty$. For this example, "more natural result" depends on the perspective. If we look at the cut, the more natural result is obtained when $p \to 1$. If we look at the path-based topology, we obtain the natural result when $p \to \infty$. Details in Appendix G.

We next formulate the clustering problem as follows. We use the k-center algorithm using p-resistance as a metric as

$$C^*_{G,p} := \min_{v_1^*, v_2^* \in V} \max_{v \in V} \min_{i \in \{1,2\}} r_{G,p}^{1/(p-1)}(v, v_i^*), \tag{139}$$

where $\{v_1^*, v_2^*\}$ is a minimizer. Since when $p \to 1$ and $r_{G,p}^{1/(p-1)} > 0$, then $r_{G,p}^{1/(p-1)} \to \infty$ and therefore Eq. (139) cannot be used. In this case, we note that the following relation that is

$$x < y \iff x^{1/(p-1)} < y^{1/(p-1)} \tag{140}$$

we have

$$C^{*p-1}_{G,p} := \min_{v_1^*, v_2^* \in V} \max_{v \in V} \min_{i \in \{1,2\}} r_{G,p}(v, v_i^*). \tag{141}$$

Thus, we simply use the comparison of $r_{G,p}$ instead of $r_{G,p}^{(1/(p-1))}$ when $p \to 1$. We finally remark that Herbster (2010) showed that when $p \to 1$ the triangle inequality still holds, i.e.,

$$r_{G,p\to 1}(i,j) \leq r_{G,p\to 1}(i,\ell) + r_{G,p\to 1}(\ell,j). \tag{142}$$

G.2. Illustrative Examples of Clustering via p-Resistance

Now, we discuss the examples in Fig. 1. We give notations as in Fig. 3. We denote by (V_{ij}, E_{ij}) the vertices and edges of the graph G_{ij}. We also give the example where the weight matters and its notation in Fig. 4.

G.2.1. THE CASE OF G_1

For the case of $p \to 1$, since p-resistance is the 1 over min-cut, we have for $j > i$

$$r_{G,p}(i,j) = \begin{cases} 1 & i = 1 \text{ and } j \in V \backslash \{1\} \\ 1/5 & i,j \in V_{12} \text{ or } i,j \in V_{13} \\ 1/4 & i \in \{5,6\}, j \in \{7,8\} \end{cases} \tag{143}$$

Note that $r_{G,p}(i,j) = r_{G,p}(j,i)$. By using this p-resistance, the set satisfying Eq. (141) is $v_1^* = 1$ and $v_2^* \in V_{12} \cup V_{13}$. This is because if we do not take $v_1^* = 1$,

$$\min_{v_1^*, v_2^* \in V} \max_{v \in V} \min_{i \in \{1,2\}} r_{G,p}(v, v_i^*) = 1, \tag{144}$$

which is the maximum of the weight of edges of G.

For $p \to \infty$, since p-resistance is a shortest path, we have for $j > i$

$$r_{G,p}^{1/(p-1)}(i,j) = \begin{cases} 1 & i = 1, j = 2 \\ 2 & i = 1, j \in \{3,4,5,6\} \\ 3 & i = 1, j \in \{7,8\} \\ 4 & i = 1, j \in \{9,10,11\} \\ 1 & i,j \in V_{12} \text{ or } i,j \in V_{13} \\ 2 & i \in V_{12}, j \in V_{13}. \end{cases} \tag{145}$$

Then if we set $v_1^* = 2$ and $v_2^* \in V_{13}$, we have

$$\min_{v_1^*, v_2^* \in V} \max_{v \in V} \min_{i \in \{1,2\}} r_{G,p}^{1/(p-1)}(v, v_i^*) = 1. \tag{146}$$

Since this is the minimum of the weight of the edge, it is clear that this set is optimal.

Coloring the vertices in the same color if the vertices are closer to the same center than the others, we obtain Fig. 1.

G.2.2. THE CASE OF G_2

For the case of $p \to 1$, we have for $j > i$

$$r_{G,p}(i,j) = \begin{cases} 1/5 & i,j \in V_{21} \\ 1/2 & i \in V_{21}, j \in V_{22} \\ 1/2 & i,j \in V_{22}. \end{cases} \tag{147}$$

Then if we set $v_1^* \in V_{21}$ and $v_2^* \in V_{22}$, we have

$$\min_{v_1^*, v_2^* \in V} \max_{v \in V} \min_{i \in \{1,2\}} r_{G,p}^{1/(p-1)}(v, v_i^*) = 1/2. \tag{148}$$

Since $\min_{i \in V_{22},} r_{G,p}(i,j) = 1/2$, this is the best possible minimum.

For the case of $p \to \infty$, we have for $j > i$

$$r_{G,p}(i,j)^{1/(p-1)} = \begin{cases} 1 & i,j \in V_{21} \\ 2 & i \in V_{21} \backslash \{5\}, j \in \{6,10\} \\ 3 & i \in V_{21} \backslash \{5\}, j \in \{7,9\} \\ 4 & i \in V_{21} \backslash \{5\}, j = 8 \\ \min\{j-i, 6-(j-i)\} & i,j \in V_{22} \end{cases} \tag{149}$$

Then if we set $v_1^* = 5$ and $v_2^* = 8$, we have

$$\min_{v_1^*, v_2^* \in V} \max_{v \in V} \min_{i \in \{1,2\}} r_{G,p}^{1/(p-1)}(v, v_i^*) = 1. \tag{150}$$

Since the minimum of p-resistance is 1, this is the best possible minimum.

Coloring the vertices in the same color if the vertices are closer to the same center than the others, we obtain Fig. 1.

G.2.3. THE CASE OF G_3

For the case of $p \to 1$, we have for $j > i$

$$r_{G,p}(i,j) = \begin{cases} 1 & i,j \in \{1,\dots,4\} \\ 1 & i,j \in \{5,6\} \\ 1/\epsilon & i \in \{1,\dots,4\}, j \in \{5,6\}. \end{cases} \tag{151}$$

Then if we set $v_1^* \in \{1,\dots,4\}$ and $v_2^* \in \{5,6\}$, we have

$$\min_{v_1^*,v_2^* \in V} \max_{v \in V} \min_{i \in \{1,2\}} r_{G,p}^{1/(p-1)}(v, v_i^*) = 1. \tag{152}$$

Since the minimum of p-resistance is 1, this is the best possible minimum.

For the case of $p \to \infty$, we have for $j > i$

$$r_{G,p}^{1/(p-1)}(i,j) = j - i \text{ if } j > i \tag{153}$$

Then if we set $v_1^* = 2$ and $v_2^* = 5$, we have

$$\min_{v_1^*,v_2^* \in V} \max_{v \in V} \min_{i \in \{1,2\}} r_{G,p}^{1/(p-1)}(v, v_i^*) = 1. \tag{154}$$

Since the minimum of p-resistance is 1, this is the best possible minimum.

Coloring the vertices in the same color if the vertices are closer to the same center than the others, we obtain Fig. 1.

H. On Theorem 4.1

This section discusses Thm. 4.1, including proof and some existing claim on Thm. 4.1.

H.1. Proof for Theorem 4.1

We use the following characteristics of p-Laplacian, defined as Eq. (22).

Proposition H.1 ((Bühler & Hein, 2009))**.**

$$S_{G,p}(\mathbf{x}) = \langle \mathbf{x}, \Delta_p \mathbf{x} \rangle_{\mathcal{H}(V)}, \tag{155}$$

$$\left(\frac{\partial S_p(\mathbf{x})}{\partial \mathbf{x}} \right)_i = p(\Delta_p \mathbf{x})_i. \tag{156}$$

Before we prove the main argument, we now explore a matrix expression of the p-Laplacian Δ_p. We define a matrix $A_{p,\mathbf{x}}$ as

$$A_{p,\mathbf{x}}(i,j) := a_{ij}|x_i - x_i|^{p-2}, \tag{157}$$

and its degree-like matrix $D_{p,\mathbf{x}}$ as

$$D_{p,\mathbf{x}}(i,j) = \begin{cases} \sum_{j=1}^n A_{p,\mathbf{x}}(i,j) & \text{if } l = i \\ 0 & \text{if } l \neq i \end{cases} \tag{158}$$

Define the matrix $L_{p,\mathbf{x}}$ as

$$L_{p,\mathbf{x}} := D_{p,\mathbf{x}} - A_{p,\mathbf{x}}. \tag{159}$$

We note that these matrix expressions depend on $\mathbf{x}$ except the $p = 2$ case. Now,

$$(L_{p,\mathbf{x}}\mathbf{x})_i = D_{p,\mathbf{x}}(i,i)x_i - \sum_{j=1}^{n} A_{p,\mathbf{x}}(i,j)x_j \tag{160}$$

$$= \sum_{j=1}^{n} A_{p,\mathbf{x}}(i,j)x_i - \sum_{j=1}^{n} A_{p,\mathbf{x}}(i,j)x_j \tag{161}$$

$$= \sum_{j=1}^{n} A_{p,\mathbf{x}}(i,j)(x_i - x_j) \tag{162}$$

$$= \sum_{j=1}^{n} a_{ij}|x_i - x_i|^{p-1}\mathrm{sgn}(x_i - x_j) \tag{163}$$

$$= (\Delta_p \mathbf{x})_i. \tag{164}$$

Thus, we can say that $L_{p,\mathbf{x}}$ is a matrix expression of the p-Laplacian, satisfying

$$L_{p,\mathbf{x}}\mathbf{x} = \Delta_p \mathbf{x}. \tag{165}$$

Then, by Prop. H.1, we can prove that

$$\mathbf{x}^\top L_{p,\mathbf{x}}\mathbf{x} = S_{G,p}(\mathbf{x}). \tag{166}$$

Now we turn to the optimization problem Eq. (21). By using the Lagrangian multiplier method, the optimal solution satisfies the following:

$$F(\mathbf{x},\lambda) := (S_{G,p}(\mathbf{x})) - \lambda(x_i - x_j - 1) \tag{167}$$

$$\frac{\partial F}{\partial \mathbf{x}} = pL_{p,\mathbf{x}}\mathbf{x} - \lambda(\mathbf{e}_i - \mathbf{e}_j) = 0 \tag{168}$$

$$\frac{\partial F}{\partial \lambda} = x_i - x_j - 1 = 0. \tag{169}$$

From Eq. (168), we have

$$\mathbf{x}^{*_{ij}} = \frac{\lambda}{p} L^+_{p,\mathbf{x}^{*_{ij}}}(\mathbf{e}_i - \mathbf{e}_j). \tag{170}$$

From Eq. (170) and Eq. (169), we have

$$\frac{\lambda}{p}\left((L^+_{p,\mathbf{x}^{*_{ij}}}(i,i) - L^+_{p,\mathbf{x}^{*_{ij}}}(i,j)) - (L^+_{p,\mathbf{x}^{*_{ij}}}(j,i) - L^+_{p,\mathbf{x}^{*_{ij}}}(j,j))\right) = 1. \tag{171}$$

Following Eq. (170), we substitute λ/p from Eq. (171) into Eq. (170), and we have

$$\mathbf{x}^{*_{ij}} = \frac{L^+_{p,\mathbf{x}^*_{ij}}}{L^+_{p,\mathbf{x}^{*_{ij}}}(i,i) + L^+_{p,\mathbf{x}^{*_{ij}}}(j,j) - 2L^+_{p,\mathbf{x}^{*_{ij}}}(i,j)}(\mathbf{e}_i - \mathbf{e}_j). \tag{172}$$

Since p-resistance is an inverse of the energy, we obtain

$$r_{G,p}(i,j) = (\mathbf{x}^{*_{ij}\top} L^+_{p,\mathbf{x}^{*_{ij}}}\mathbf{x}^{*_{ij}})^{-1} \tag{173}$$

$$= L^+_{p,\mathbf{x}^{*_{ij}}}(i,i) + L^+_{p,\mathbf{x}^{*_{ij}}}(j,j) - 2L^+_{p,\mathbf{x}_{*_{ij}}}(i,j) \tag{174}$$

$$= (\mathbf{e}_i - \mathbf{e}_j)^\top L^+_{p,\mathbf{x}^{*_{ij}}}(\mathbf{e}_i - \mathbf{e}_j) \tag{175}$$

The rest of the proof is same as the original proof in Thm. 6 in (Alamgir & Luxburg, 2011). The trick is that we do not have to the exact form of $L_{p,\mathbf{x}^{*_{ij}}}$. Only this expression is enough to prove Theorem 4.1.

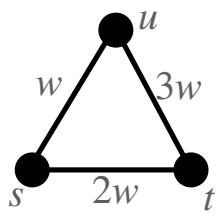


Figure 5: The example discussed in (Bridle & Zhu, 2013).

H.2. Original Context of Theorem 4.1

Originally in Sec. 5 (Alamgir & Luxburg, 2011), Thm. 4.1 when $p = 2$ has a different interpretation. Nadler et al. (2009) proves that the semi-supervised learning problem of $p = 2$ case is meaningless if the number of vertices are infinite. Thm. 4.1 for $p = 2$ supports this claim in (Nadler et al., 2009) for two-pole semi-supervised leaning problem for the following way. Since the equivalent 2-resistance is known to converge to a meaningless function, the solution of the semi-supervised problem is equivalently characterized by this meaningless function. Thus, the semi-supervised learning does not make sense, if the number of the vertices are large. If the conjecture for $p > 1$ case were proven, Thm. 4.1 can be interpreted that for some range of $p > 1$ two-pole semi-supervised learning problem is not meaningless, since the equivalent p-resistance is shown not to converge to a meaningless one. Later year, independent of p-resistance, the statement "for some range of $p > 1$ semi-supervised learning problem is not meaningles" is proven by (Slepcev & Thorpe, 2019). Thm. 4.1 now supports (Slepcev & Thorpe, 2019) from a p-resistance view.

H.3. Remark on the Existing Claims on Theorem 4.1

Finally, we discuss several existing claims on this theorem. First, we need to mention a small *fixable* mistake in the original proof in (Alamgir & Luxburg, 2011) for the $p = 2$ case. The original proof assumes that the solution to the semi-supervised learning Eq. (21) when $p = 2$ is that

$$\mathbf{x}^{*ij} = L^{+}(\mathbf{e}_i - \mathbf{e}_j). \tag{176}$$

However, this is not true since this does not satisfy the constraint

$$x_i^{*ij} - x_j^{*ij} = (\mathbf{e}_i - \mathbf{e}_j)^{\top} L^{+}(\mathbf{e}_i - \mathbf{e}_j) \neq 1. \tag{177}$$

Instead, the solution is given as

$$\mathbf{x}^{*ij} = \frac{L^{+}(\mathbf{e}_i - \mathbf{e}_j)}{(\mathbf{e}_i - \mathbf{e}_j)^{\top} L^{+}(\mathbf{e}_i - \mathbf{e}_j)}. \tag{178}$$

Note that this corresponds to Eq. (172). However, this does not affected the rest of the proof, since the proof exploits only $\mathbf{x}^{*ij} = \rho L^{+}(\mathbf{e}_i - \mathbf{e}_j)$ for $\rho \in \mathbb{R}$, and ρ does not matter. Thus, the validity of the original claim still remains.

Next, since Thm. 4.1 resolves the open problem in (Alamgir & Luxburg, 2011), there is an existing discussion on if this statement is true or not. The work (Bridle & Zhu, 2013) claims that there is a counterexample to Thm. 4.1 in the general p case. In the following, we argue that the discussion on the example in (Bridle & Zhu, 2013) does not work as a counterexample.

The "counterexample" given in (Bridle & Zhu, 2013) is based on the example shown as Fig. 5. However, unfortunately, we believe that there is invalidity in the discussion on this example. Firstly, we recall that

$$\min_{\mathbf{x}}\{S_{G,p}(\mathbf{x}) \text{ s.t. } x_s - x_t = 1\} \neq \min_{\mathbf{x}} S_{G,p}(\mathbf{x}), \tag{179}$$

since $\min_{\mathbf{x}} S_{G,p}(\mathbf{x}) = 0$ when $\mathbf{x} = c\mathbf{1}$, $\forall c \in \mathbb{R}$, while $c\mathbf{1}$ does not satisfy the constraint of the left hand side. However, the work (Bridle & Zhu, 2013) assumes the equality of Eq. (179), see the the first equality at the top of the left column in p.3

Table 3: Dataset Summary. Since Hopkins 155 contains 155 different videos, we report the sum of the data points and sum of the dimensions of videos. Also, Hopkins 155 dataset contains 120 2-class datasets and 35 3-class datasets.

	ionosphere	hop 155 2cls	iris	wine	hop 155 3cls
# of class	2	2	3	3	3
size	351	31981	150	178	13983
dimension	34	3542	4	13	999

Table 4: Computational Time for the Experiment (unit:sec). Here we use E notation, e.g., E-6= 10^{-6} or E1 =10^{1}. For methods we use the same abbriviation as Table 1. Since "Rec-bi p" is a deterministic method, we only report time. Also, since Hop contains multiple datasets, we only show the average.

		2 clsss		multi-class		
Type	Method	ionosphere	Hop 2 cls	iris	wine	Hop 3 cls
ER	k-med (a) p	1.37E1 $\pm$ 0.01E1	1.21E1	4.93E0 $\pm$ 0.01E0	1.30E-1 $\pm$ 0.13E-1	1.78E1
ER	k-med $p = 2$	3.76E0 $\pm$ 0.01E0	1.01E1	1.36E0 $\pm$ 0.00E0	9.07E-2 $\pm$ 1.05E-2	1.34E1
ER	FF (a) p	9.71E-2 $\pm$ 0.32E-1	4.88E-1	5.45E-1 $\pm$ 0.33E-1	4.72E-2 $\pm$ 0.18E-2	8.01E0
ER	FF $p = 2$	8.71E-2 $\pm$ 0.12E-1	3.56E-1	4.21E-1 $\pm$ 0.12E-1	3.82E-2 $\pm$ 0.06E-2	6.45E0
ER	p-Flow	1.46E1 $\pm$ 0.01E1	1.33E1	5.21 E0 $\pm$ 0.02E0	1.60E-1 $\pm$ 0.15E-1	1.81E1
ER	ECT	1.01E-1 $\pm$ 0.10E-1	8.12E-1	2.03E-2 $\pm$ 0.42E-2	2.05E-2 $\pm$ 0.49E-2	9.26E-1
SC	Rec-bi p	1.18E-1	8.11E-1	5.35E-1	9.07E-2	1.01E0
SC	SC p-orth	8.60E-2 $\pm$ 0.01E-2	6.31E-1	4.78E-1 $\pm$ 1.65E-1	3.02E-2 $\pm$ 0.57E-2	8.10E-1
SC	SC $p = 2$	1.60E-2 $\pm$ 0.01E-2	3.43E-2	1.28E-1 $\pm$ 0.00E-1	8.78E-3 $\pm$ 0.00E-3	6.12E-2

of (Bridle & Zhu, 2013). Moreover, we note that

$$\frac{\partial S_{G,p}(\mathbf{x})}{\partial x_u} = \frac{\partial}{\partial x_u}\left(w|x_s - x_u|^p + 2w|x_u - x_v|^p + 3w|x_s - x_t|^p\right) \tag{180}$$

$$= p\left(w|x_s - x_u|^{p-1} + 2w|x_u - x_v|^{p-1}\right), \tag{181}$$

and therefore

$$\frac{\partial S_{G,p}(\mathbf{x})}{\partial x_u} \neq p\left(-w|x_s - x_u|^{p-1} + 2w|x_u - x_v|^{p-1}\right), \tag{182}$$

where the difference between Eq. (181) and Eq. (182) is the sign of the term $w|x_s - x_u|^{p-1}$. [3] However, the work (Bridle & Zhu, 2013) assumes the equality of Eq. (182), see the the third equality at the top of the left column in p.3 of (Bridle & Zhu, 2013). In (Bridle & Zhu, 2013), these invalid equality assumptions of Eq. (179) and Eq. (182) derive the fundamental relationship in order to bring a counterexample. The rest of the analysis in (Bridle & Zhu, 2013) is carried with this relationship. Due to this invalidity, we believe that there are serious flaws in the claim that the example Fig. 5 leads to a counterexample to Thm. 4.1. Hence, we claim that Thm. 4.1 holds with the proof in this section.

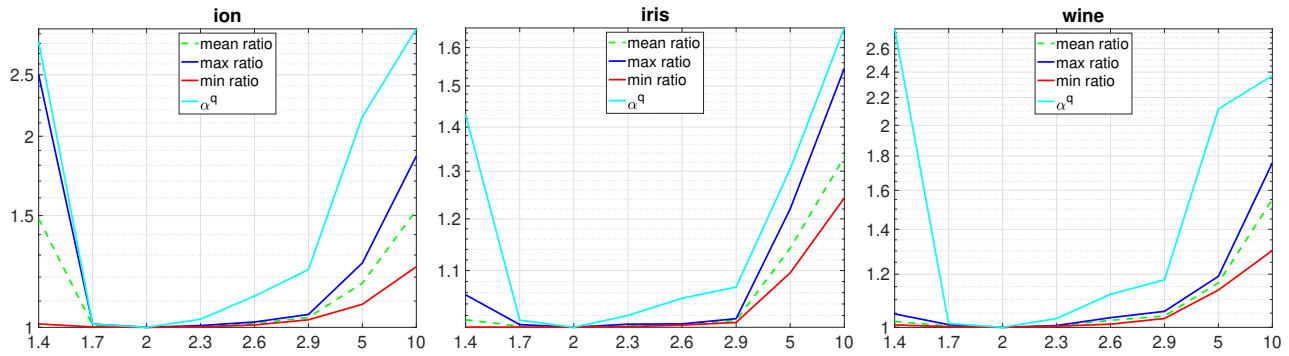
I. Details of the Preliminary Experiments

This section discusses missing details of preliminary experiments and additional experimental results.

I.1. Details of Preliminary Experiments

To begin with, we further explain the details of the experimental setting. We report the size of the dataset in Table. 3. Our experiment was conducted on Mac Studio with M1 Max Processor and 32GiB RAM. Also, we use an Intel binary Matlab

[3] There is a slight difference between the definition of the p-resistance between ours and (Bridle & Zhu, 2013). We follow the definition of (Herbster & Lever, 2009) and (Bridle & Zhu, 2013) follows the definition of (Alamgir & Luxburg, 2011). However, these two have almost same properties. Moreover, while we write the equations in our form, this difference does not affect the discussion here. For more details of the difference, see §6.1 in (Alamgir & Luxburg, 2011) or Sec. I.



(a) ion (b) iris (c) wine

Figure 6: The ratio of the approximated value of p-resistance to the exact p-resistance, i.e., $\|L^+\mathbf{e}_i - L^+\mathbf{e}_j\|_{G,q}^q / r_{G,p}^{1/(p-1)}(i,j)$. Also, the factor of the bound $\alpha_{G,p}^q$.

translated by Rosetta, which is a standard use in MacOS with Apple Silicon environment. We need to note that for the comparison method Rec-bi (Bühler & Hein, 2009), originally, the algorithm is defined for $p \leq 2$. Thus, we apply the same technique for $p \leq 3$. In (Bühler & Hein, 2009), in order to avoid the conversion to too close or too far local optimum, at the step t (Bühler & Hein, 2009) minimizes Eq. (25) via the gradient descent method using the initial condition as the obtained eigenvector of the previous step $p_t = 0.9p_{t-1}$. If we increase the p, we use the same technique; $p_t = p_{t-1}/0.9$ until p reaches 5. Beyond 5, we use $p_t = 2p_{t-1}$. When $p = 2$, we know that 2-resistance is further computed as

$$r_{G,2}(i,j) = \|L^+\mathbf{e}_i - L^+\mathbf{e}_j\|_{G,2}^2 = \|(L^+)^{1/2}\mathbf{e}_i - (L^+)^{1/2}\mathbf{e}_j\|^2. \tag{183}$$

The graph p-seminorm is the size m norm, while the latter is based on the size n norm. However, we use graph 2-seminorm even for the $p = 2$ case. The reason is that we needed to be consistent in the experiment since we observed numerical round-off errors in the other methods. On the other hand, ECT (Yen et al., 2005) uses the square of 2-resistance. For this method, we use $\|(L^+)^{1/2}\mathbf{e}_i - (L^+)^{1/2}\mathbf{e}_j\|^2$ since the original paper (Yen et al., 2005) uses this.

Next, we discuss the computational time for the experiment. Due to its significant computational time, we parallelized the distance computation for the exact method, while we did not use such a technique for the others. Thus, we first compare the approximation of the p-reistance and the exact p-resistance. Next, we compare the computational time among the methods except for the exact methods.

I.2. Computational Times of Experiments

In this section we discuss computational times of the experiment.

In Table 2, we compare the approximation by Eq. (20) and the exact computation of p-resistance by naively optimizing Eq. (7) by the gradient descent. For ionosphere, iris, and wine, we made a graph for the best performing parameters in Table. 1.

In Table 4, we compare the computing time for the the best performing parameters in Table. 1. We can see that ours are slower than spectral clustering methods. This slowness is because ours takes $O(mn^2)$ while spectral clustering methods using p-Laplacian are $O(n^3)$-based convergence methods. Looking at the computational time for ECT, the time is similar to the spectral methods since ECT is also the $O(n^3)$ method. The computing time on ECT further motivates this future direction.

I.3. Comparison of the Values of Approximated and Exact p-Resistance

This section discusses the comparison of approximated and exact p-resistance. This experiment shows that the value is approximation is tighter than Prop. 3.5.

This preliminary experiment aims to evaluate the quality of the approximation comparing to Thm. 3.3. To do so, we would like to compute the ratio of approximation to the exact p-resistance as $\|L^+\mathbf{e}_i - L^+\mathbf{e}_j\|_{G,q}^q / r_{G,p}^{1/(p-1)}(i,j)$. Using Thm. 3.3,

this ratio can be theoretically evaluated as

$$1 \leq \frac{\|L^+\mathbf{e}_i - L^+\mathbf{e}_j\|_{G,q}^q}{r_{G,p}^{1/(p-1)}(i,j)} \leq \alpha_{G,p}^q, \alpha_{G,p} := \|\!|W^{1/p}CC^+W^{-1/p}|\!\|_p. \tag{184}$$

This experiment also aims to evaluate this inequality. We numerically computed the approximated p-resistance, exact p-resistance, and $\alpha_{G,p}$. To compute $\alpha_{G,p}$, we use the same algorithm for the Table. 8, which is discussed in Sec. J.3 To compute the approximated and exact p-resistance, we conducted with the following procedure. To create a graph, we used $\mu = 0.1$, in order to make k-nn graph. This means that we use $k = \lfloor 0.1n \rfloor$. To make the comparison simple, we use an unweighted graph. This is because if we incorporate weights, it is not trivial how $\alpha_{G,p}$ behaves, and thus, the results might not be a consistent of a comparison among different ps. The rest of the analysis was carried in the same procedure as Table 4.

The result is summarized in Fig. 6. We can see that the all the ratios are in the bound of Thm. 3.3. We also remark that using Prop. 3.5 we have the bound as

$$\alpha_{G,p}^q \leq m^{|q/2-1|}, \tag{185}$$

all of the plots of $\|L^+\mathbf{e}_i - L^+\mathbf{e}_j\|_{G,q}^q / r_{G,p}^{1/(p-1)}(i,j)$ in Fig. 6 is obviously far lower than this bound. For this result we observe the looser bound for larger p, since $\|\!|CC^+|\!\|_p \leq \|\!|CC^+|\!\|_\infty$ assuming an unweighted graph. By incorporating the weight, we might observe $\alpha_{G,p}$ differently, since we do not know which is larger $\|\!|W^{1/p}CC^+W^{-1/p}|\!\|_p$ and $\|\!|W^{1/\infty}CC^+W^{-1/\infty}|\!\|_\infty$. Further, we have

$$\alpha_{G,p}^q = \|\!|W^{1/p}CC^+W^{-1/p}|\!\|_p^q \leq \left(\frac{w_{\max}}{w_{\min}}\right)^{q/p=q-1} \|\!|CC^+|\!\|_p^q. \tag{186}$$

Seeing the current derived bound Eq. (186), the bound may be looser if we involve the weights and p is small and hence q is large. A tighter bound particularly for smaller p is a possible future direction, but this might be a lower priority due to the low performance at the smaller p. The reason is that, for small p, it is known that $r_{G,p}(i,j)$ converges to a meaningless function (Alamgir & Luxburg, 2011; Slepcev & Thorpe, 2019) under certain graph building conditions. Also, possibly due to this, our method performs better for larger p.

We remark on the large p observation. In the main text, we argue that "these correspond to the existing theoretical insight; p-resistance with large p becomes meaningful function while 2-resistance is not (Alamgir & Luxburg, 2011; Slepcev & Thorpe, 2019)." At a first glance, (Alamgir & Luxburg, 2011) seems to argue that "smaller p works". However, this is due to the difference of the definition. While we follow the p-resistance definition in (Herbster & Lever, 2009), in (Alamgir & Luxburg, 2011) their p-resistance $r_{G,p}^A$ is given as

$$r_{G,p}^A(i,j) = \frac{1}{\min_{\mathbf{x}} \sum_{ij} a_{ij}^{1/(p-1)} |x_i - x_j|^{p/(p-1)} \text{ s.t. } x_i - x_j = 1} \tag{187}$$

$$= \frac{1}{\min_{\mathbf{x}} \sum_{ij} a_{ij}^{q-1} |x_i - x_j|^{q} \text{ s.t. } x_i - x_j = 1}. \tag{188}$$

Hence, in (Alamgir & Luxburg, 2011) the parameter p works in the opposite way; if we mean large, p (Alamgir & Luxburg, 2011) means smaller p and vice-versa. Despite this slight change of the definition, p-resistance in (Herbster & Lever, 2009) and (Alamgir & Luxburg, 2011) shares almost the same properties. More discussion can be seen in §6.1 in (Alamgir & Luxburg, 2011).

I.4. Code for the Experiments

We leave a remark on our code in the supplemental material. Although we include an implementation of Alg. 1, at this stage we include a part of the whole codebase due to the copyright of the library reasons. In the final version, we plan to reduce the blockers to publish the code as much as possible. Moreover, we plan to publish our implementation at Github, an online codebase repository service.

J. More Discussion on $\alpha_{G,p}$

This section gives more observations on $\alpha_{G,p}$. In practice, we want to know how close to the exact value and how far from this upper bound the value of $\|\!|W^{1/p}CC^+W^{-1/p}|\!\|_p$ is. In the following, we argue that in the general case $\alpha_{G,p}$ is far less

than the bound given in Prop. 3.5.

Before we get into the detail, we give a brief overview of an interpretation of $\alpha_{G,p}$. From the definition of $\mathbf{z}$, $\mathbf{z}$ is a mapping of $f_{q/p}(C\mathbf{y})/\|\mathbf{y}\|_{G,q}^q$ from $\mathbb{R}^m \to \mathrm{Im}(C)$. Comparing the equality condition Eq. (114), we observe that if $f_{q/p}(C\mathbf{y}) \in \mathrm{Im}(C)$, we obtain the tightest bound since $\|\mathbf{z}\|_{G,p} = \|\mathbf{y}\|_{G,q}^{-1}$. By looking at this, we observe that the $\alpha_{G,p}$ is the worst possible "overflow" of the mapping from $\mathrm{Im}(C)$ from $\mathbb{R}^m$, in a sense of the weighted p-norm.

J.1. Condition Number Point of View

To prove the bound of Prop. 3.5, we only use $|||MM^+|||_2 = 1$ and Lemma B.1, which holds for *any* matrix M. Hence, we can say that this is the "worst" bound and we expect a far lower value of $|||CC^+|||_p$ for a general incidence matrix of graph. To gain some qualitative observation on how close between the exact and approximation, we further decompose $\alpha_{G,p}$. By using the submultiplicity and Lemma B.2,

$$\alpha_{G,p} = |||W^{1/p}CC^+W^{-1/p}|||_p \leq |||W^{1/p}|||_p |||CC^+|||_p |||W^{-1/p}|||_p \leq |||CC^+|||_p w_{\max}^{1/p}/w_{\min}^{1/p} \tag{189}$$

$$\leq |||C|||_p |||C^+|||_p w_{\max}^{1/p}/w_{\min}^{1/p} \tag{190}$$

where $w_{\max}{:=}\max_\ell w_\ell$ and $w_{\min} := \min_\ell w_\ell$. In numerical analysis, the term $|||C|||_p|||C^+|||_p$ is called as a *condition number* of the matrix C (Saad, 2003). A condition number is related to the "difficulty" to numerically solve the linear equation $C\mathbf{x} = \mathbf{y}$. The larger the condition number gets, the more difficult to solve the linear equation. The linear equation is difficult to solve if we can make one or more pairs of column or row of C close to parallel by elementary operations. However, by construction of incidence matrix, no pairs of column or row of the incidence matrix are close to parallel. Thus, we expect that the condition number of C will not be large, and hence we expect a smaller value of $\alpha_{G,p}$ than Prop. 3.5 in general. For the specific graphs, we theoretically show this in the next section.

J.2. Bound of $\alpha_{G,p}$ for Some Specific Graphs

In this section we give a constant bound of $\alpha_{G,p}$ for some specific graphs. By this we can independently bound $\alpha_{G,p}$ of m.

Now, for the specific cases, we have the following.

Proposition J.1. *If a graph is complete or cyclic, then* $|||CC^+|||_p \leq 4$ *and hence* $\alpha_{G,p} \leq 4w_{\max}^{1/p}/w_{\min}^{1/p}$.

For these specific graphs, we can bound the p-resistance (Thm. 3.3) by a constant. We now divide the proof into the complete case and the cyclic case.

J.2.1. Complete Case

First, we obtain the pseudoinverse of C of a complete graph.

Lemma J.2. *For an incidence matrix C' for a complete graph,*

$$C'^+ = \frac{1}{n}C'^\top \tag{191}$$

Proof. For a graph Laplacian L of unweighted graph can be written as

$$L = nI - \mathbf{1}^\top\mathbf{1}, \tag{192}$$

and thus

$$L_{ij} = \begin{cases} n-1 & \text{if } i = j \\ -1 & \text{if } i \neq j \end{cases}. \tag{193}$$

Also, we know that

$$L = C'^\top C'. \tag{194}$$

Now we consider the the vector $\mathbf{x}_{ij} \in \mathbb{R}^n$ as

$$\mathbf{x}_{ij}^\top := \underbrace{(0, \ldots, 0, \overbrace{1}^{i\text{th element}}, 0, \ldots, 0, \overbrace{-1}^{j\text{th element}}, 0, \ldots, 0)}_{\text{size } n}. \tag{195}$$

Note that this $\mathbf{x}_{ij}$ is one row of the incidence matrix C'. Now we get

$$(L\mathbf{x}_{ij})_l = \begin{cases} (n-1) \times 1 + (-1) \times (-1) = n & \text{if } l = i \\ 1 \times (-1) + (n-1) \times (-1) = -n & \text{if } l = j \\ (-1) \times 1 + (-1) \times (-1) = 0 & \text{otherwise} \end{cases} \tag{196}$$

$$= n(L\mathbf{x}_{ij})_l. \tag{197}$$

Since $\mathbf{x}_{ij}$ is one column of the transpose of the incidence matrix $\mathbf{C}^{'\top}$,

$$LC = C^{'\top} C' C^{'\top} = nC^{'\top} \iff \left(\frac{1}{n} C^{'\top}\right) C' \left(\frac{1}{n} C^{'\top}\right) = \frac{1}{n} C^{'\top} \tag{198}$$

Also,

$$(LC)^\top = C' C^{'\top} C' = nC' \iff C' \left(\frac{1}{n} C^{'\top}\right) C' = C' \tag{199}$$

From Eq. (198) and Eq. (199), the matrix $1/nC'$ satisfies the definition of C^+, which leads to the claim. □

Note that $|||CC^+|||_1 = |||CC^+|||_\infty$ due to the symmetricity of CC^+.

$$|||CC^+|||_p \leq |||CC^+|||_\infty \leq |||C|||_\infty |||C^+|||_\infty = 4\frac{n+1}{n} \leq 4. \tag{200}$$

J.2.2. Cyclic Case

In the cyclic graph, $m = n$, i.e., the number of vertices is equal to the number of edges. Thus, the incidence matrix C is square. However, in order to avoid confusion, in the following we use m and n. Now, we define the incidence matrix $C \in \mathbb{R}^{m \times n}$of the cyclic graph as

$$c_{i1} = \begin{cases} -1 & \text{when } i = 1 \\ 1 & \text{when } i = 2 \\ 0 & \text{otherwise} \end{cases} \tag{201}$$

$$c_{i2} = \begin{cases} -1 & \text{when } i = 1 \\ 1 & \text{when } i = n \\ 0 & \text{otherwise} \end{cases} \tag{202}$$

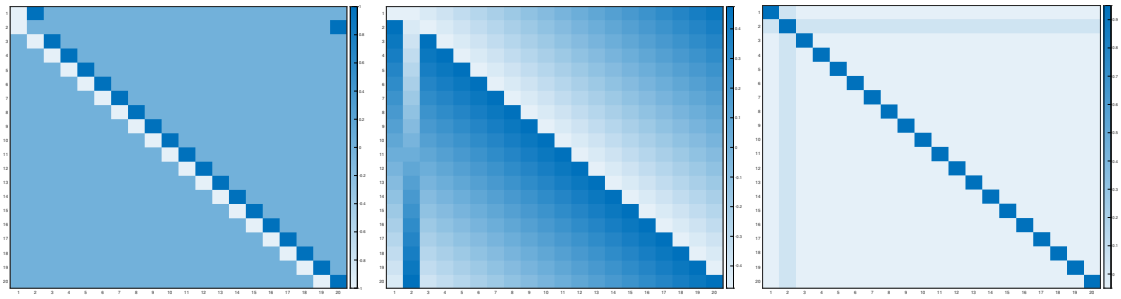
$$c_{ij} = \begin{cases} -1 & \text{when } i = j-1 \\ 1 & \text{when } i = j \\ 0 & \text{otherwise for} \end{cases} \quad \text{for } j \geq 3. \tag{203}$$

Before we explore C^+, we introduce *cyclic shift operator* of the vector. Given the vector $\mathbf{a}$, the shift operator (l) "cyclic shifts" the element, as

$$\mathbf{a}^{(l)} = (a_{n-l+1}, a_{n-l+2}, \ldots, a_n, a_1, \ldots, a_{n-l})^\top. \tag{204}$$

Thus, $\mathbf{a}^{(0)} = \mathbf{a}$. Also, we define the reverse operator rev for a vector $\mathbf{a}$ as

$$\mathrm{rev}(\mathbf{a}) = (a_n, a_{n-1}, \ldots, a_1). \tag{205}$$



(a) C (b) $B = C^+$ (c) CC^+

Figure 7: Heatmap plot for the matrices C, $B = C^+$ and CC^+ of the cyclic graph for $n = 20$.

We also define the vector $\boldsymbol{\xi} \in \mathbb{R}^n$ as

$$\boldsymbol{\xi} = (1/2 - 1/2n, 1/2 - 3/2n, \ldots, 1/2 - (2i-1)/2n, \ldots, -1/2 + 1/n). \tag{206}$$

Now, we define a matrix B as

$$B_{1\cdot} = \boldsymbol{\xi}^{(1)} \tag{207}$$

$$B_{2\cdot} = \mathrm{rev}(\boldsymbol{\xi}^{(0)}) = \mathrm{rev}(\boldsymbol{\xi}) \tag{208}$$

$$B_{j\cdot} = \boldsymbol{\xi}^{(j-1)} \text{for } j \geq 3, \tag{209}$$

where $B_{i\cdot}$ denotes i-th column of B. We plot a heatmap of C and B for the illustrative purpose.

Now we prove that $C^+ = B$. To claim that, it is enough to prove that $BC = I - \mathbf{1}^\top \mathbf{1}/n$ (Bapat, 2010). From the construction,

$$(BC)_{ii} = \xi_1 - \xi_n = 1/2 - 1/2n - (-1/2 - 1/2n) = 1 - 1/n. \tag{210}$$

Also, when $i \neq j$,

$$(BC)_{ij} = \begin{cases} -\xi_{i-1} - \xi_{n-i+1} & \text{when when } j = 1 \\ \xi_{i+1}^{(j)} - \xi_i^{(j)} & \text{when when } 2 \leq j < n, \\ \xi_{n-i+1} + \xi_{i+1} & \text{when when } j = n, \end{cases} \tag{211}$$

$$= -1/n. \tag{212}$$

Thus, we can say that $B = C^+$. By doing a similar computation, we get

$$CC^+ = \begin{cases} 1 - 1/n & \text{when } i = j \\ 1/n & \text{when } i = 2 \text{ or } j = 2, i \neq j \\ -1/n & \text{otherwise} \end{cases} \tag{213}$$

We also plot a heatmap for CC^+ for the illustrative purpose in Fig. 7(c). Thus, applying Lemma B.2, we get

$$|||CC^+|||_p \leq |||CC^+|||_1 = \max_i \sum_{j=1}^{n} |(CC^+)_{ij}| = 2 - 1/n \leq 4. \tag{214}$$

We leave a brief note for other concrete examples. Several attempts are made to obtain the concrete form of C^+ for the specific graph (Azimi & Bapat, 2018; Azimi et al., 2019). However, due to their abstract ways to characterize the graph such as distance or cut, we think that it is hard to immediately obtain a non-trivial bound from these results. Also, C^+ for tree is studied (Bapat, 1997). However, since we know the exact representation of p-resistance for tree in Thm. 3.4, we do not have to discuss the tree case.

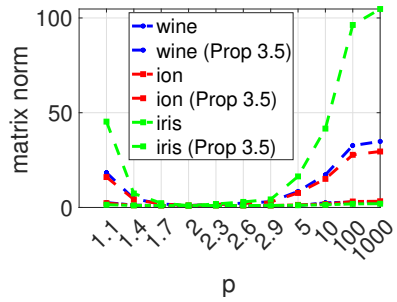


Figure 8: Matrix norm $\|CC^+\|$ vs p.

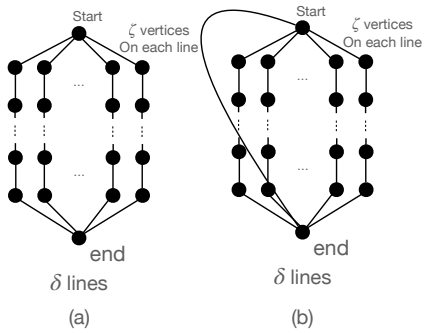


Figure 9: The example where the approximated value is far lower than the exact value. See J.3 for details.

Table 5: The values of approximated 1-Resistance for (a). The exact 1-resistance for this graph is $1/\delta$.

		ζ				
		5	10	20	40	80
	5	0.2	0.1	0.05	0.025	0.0125
	10	0.2	0.1	0.05	0.025	0.0125
δ	20	0.2	0.1	0.05	0.025	0.0125
	40	0.2	0.1	0.05	0.025	0.0125
	80	0.2	0.1	0.05	0.025	0.0125

Table 6: The values of approximated p-Resistance for (b). The exact 1-resistance for this graph is $1/(\delta+1)$.

		ζ				
		5	10	20	40	80
	5	0.46	0.33	0.21	0.13	0.07
	10	0.61	0.48	0.33	0.21	0.12
δ	20	0.75	0.64	0.49	0.33	0.20
	40	0.85	0.77	0.65	0.49	0.33
	80	0.92	0.87	0.79	0.66	0.50

J.3. Additional Experiments for Bound of Approximation

This section discusses some additional experiments for bound of approximation in Prop. 3.5. For the real datasets we confirmed that the approximation over exact $\|L^+\mathbf{e}_i - L^+\mathbf{e}_j\|_{G,q}^q / r_{G,p}^{1/(p-1)}(i,j)$ is far below the bound and its worst case, in Fig. 6. In this section, we further investigate the $\alpha_{G,p}$ of the real datasets. Moreover, we further investigate artificial dataset that the ratio is close to the worst dataset, and we discuss why it happens.

J.3.1. Plot of $|||CC^+|||_p$.

As we discussed in Sec. I.3, since $\alpha_{G,p} = |||W^{1/p}CC^+W^{-1/p}|||_p$ involves p in the matrix as well as norm, it is somewhat difficult to how $\alpha_{G,p}$ behaves by changing p. Thus, we focus on the unweighted graph; we numerically investigate the unweighted $\|CC^+\|_p$ that is a matrix norm evaluated in Eq. (186). We plot $\|CC^+\|_p$ for wine, ion, and iris with the μ when k-medoids performs the best in Fig. 8. We use the matrix p-norm estimation algorithm proposed by (Higham, 1992). We plot $p \to 1$ and $p \to \infty$ as the exact value. We remark on the estimation of the matrix norm by (Higham, 1992); let ξ be the output by the estimation of the matrix norm of CC^+, then $|||CC^+|||_p / m^{|1/2-p|} \leq \xi \leq |||CC^+|||_p$. We note that using Lemma B.2 and symmetricity of CC^+, we have

$$|||CC^+|||_p \leq |||CC^+|||_1 = |||CC^+|||_\infty, \tag{215}$$

whose exact value is computable. Thus, although we need to use estimation algorithm for $|||CC^+|||_p$, we can exactly compute the bound of this norm.

Fig. 8 shows that $\|CC^+\|_p$ is far lower than the worst bound in Prop. 3.5. Moreover, Fig. 8 shows that the estimation algorithm proposed by Higham (1992) outputs is reliable results on this problem since this follows theory in terms of $|||CC^+|||_{p'} \leq |||CC^+|||_p$ if $2 < p' < p$ or $p < p' < 2$, also Eq. (215). Also, even only looking at $|||CC^+|||_1 = |||CC^+|||_\infty$, which is exactly computable the bound of $|||CC^+|||_p$, we have far lower value than the worst bound, especially in larger p where our clustering algorithm works better.

Table 7: The values of $\|\|CC^+\|\|_1/m^{1/2}$ for the graph (a). If this value is close to 1, we have a looser bound.

		ζ				
		5	10	20	40	80
	5	0.51	0.33	0.23	0.16	0.11
	10	0.41	0.27	0.19	0.13	0.09
δ	20	0.31	0.2	0.14	0.1	0.07
	40	0.23	0.15	0.1	0.07	0.05
	80	0.16	0.11	0.07	0.05	0.04

Table 8: The values of $\|\|CC^+\|\|_1/m^{1/2}$ for the graph (b). If this value is close to 1, we have a looser bound.

		ζ				
		5	10	20	40	80
	5	0.70	0.55	0.43	0.33	0.24
	10	0.68	0.59	0.51	0.41	0.32
δ	20	0.59	0.56	0.54	0.49	0.41
	40	0.47	0.48	0.51	0.52	0.48
	80	0.36	0.38	0.44	0.49	0.51

Table 9: The values of condition number $\|\|C\|\|_1\|\|C^+\|\|_1$ for (a).

		ζ				
		5	10	20	40	80
	5	20.4	45.2	95.1	195.1	395
	10	40.4	90.2	190.1	390.1	790
δ	20	80.4	180.2	380.1	780.1	1580
	40	160.4	360.2	760.1	1560.1	3160
	80	320.4	720.2	1520.1	3120.1	6320

Table 10: The values of condition number $\|\|C\|\|_1\|\|C^+\|\|_1$ for (b).

		ζ				
		5	10	20	40	80
	5	24.4	49.7	101.8	219.3	457.7
	10	51	136.7	346.2	812.8	1790
δ	20	120.2	362.3	1035.1	2702.6	6432.7
	40	266.2	871.6	2768.2	8119.5	21449
	80	563.8	1943.9	6670.3	21713	64438

J.3.2. Example where Approximation is Far Lower than the Exact Value

Lastly, we discuss an example where the approximation is far lower than the exact value and how this happens. We consider a graph depicted in Fig. 9. First, we see a graph in Fig. 9 (a). To build this graph, first consider the line graph, where the ζ vertices are in line. This graph is constructed with δ lines of diameter ζ each lines start vertex is glued to each other lines "start vertex" and similar to the "end vertices". For Fig. 9 (b), we add one edge to the graph in Fig. 9 between the start vertex and end vertex.

We now compare with approximation and the exact value of 1-resistance between the start vertex and end vertex. As we discussed in Sec. G, we compute the exact 1-resistance between i and j as the minimum cut's inverse. Thus, for (a) $r_{G,1}(\text{start}, \text{end}) = 1/\delta$ and for (b) we have $r_{G,1}(\text{start}, \text{end}) = 1/(\delta + 1)$. We then compute the approximated values and $\|\|CC^+\|\|_1$ for Fig. 9. We give a result in Tables 5–8. From Tables 5 and 6, we observe that we have a far less accurate approximation for graph (b) than that for graph (a). In Tables 7 and 8, we observe that a far larger value of $\|\|CC^+\|\|_1$ for the graph (b) than that for the graph (a). We also observe that comparing with $\|\|CC^+\|\|_1$ of the graph constructed from the real dataset in Fig. 8, we see a far larger value of $\|\|CC^+\|\|_1$ for the graph (b). The larger value of $\|\|CC^+\|\|_1$ might be the reason why the approximation of the 1-resistance of graph (b) is far worse than the graph (a).

We now discuss why $\|\|CC^+\|\|_1$ for graph (b) is far larger than that for graph (a). We now revisit the condition number argument in Sec. J.1. The condition number is the stableness of the linear equation of the matrix. The stable linear equation is even if we add small value ϵ to the linear equation, i.e., $C\mathbf{x} = \mathbf{y} + \epsilon$, the solution $\mathbf{x}$ is almost unchanged. If we add perturbation on each edge in graph (a), the graph can absorb the perturbation since each line graph is almost independent. However, on the graph (b), each line graph becomes dependent due to the additional edge. Moreover, the start and the end vertex are like "pivots" of the graph. The perturbations might be widely spread over the graph by connecting two pivots. By this spread, graph (b) becomes unstable, while graph (a), where we do not connect the pivots is more stable. In Tables 9 and 10 we see that graph (b) is far more unstable than graph (a).

Finally, we argue that we do not observe this phenomenon in the real setting. In the example of graph (b), the unstableness comes from the sparse connection over the graph and connection of the "pivots" over such a spares graph. In a dense graph such as a complete graph, we saw far lower $\|\|CC^+\|\|_1$ as we observe in Sec J.2. As we saw in the real dataset case, we can assume that there is a denser connection over the graph, even between the clusters.

K. On Difficulties of The Exact Solution

In this section, we briefly explain the difficulties to obtain the exact solution of the resistance. Again, we consider the minimization problem Eq. (21). The Lagrangian multiplier method gives Eq. (168) and Eq. (168). From Eq. (168), the

optimal solution $\mathbf{x}$ satisfies

$$0 = p\Delta_p \mathbf{x} - \lambda(\mathbf{e}_i - \mathbf{e}_j) \tag{216}$$

To solve this problem, we want to consider Δ_p^+, which is "generalized inverse" function Δ_p, defined as

$$\Delta_p^+(\Delta_p(\Delta_p^+(\mathbf{x}))) = \Delta_p^+ \mathbf{x} \tag{217}$$

Recall that we can write as

$$\Delta_p = C^\top W f_{p-1}(C\mathbf{x}). \tag{218}$$

For the convenience of notation, we write

$$f_{\mathbf{w},p} = W f_p(C\mathbf{x}). \tag{219}$$

If there exists $\boldsymbol{\alpha} \in \mathrm{Ker}(C)$ s.t.

$$f_{\mathbf{w},p-1}^{-1}(C^{+\top}\mathbf{x} - \boldsymbol{\alpha}) \in \mathrm{Im}(C), \tag{220}$$

the Δ_p^+ is given as

$$\Delta_p^+(\mathbf{x}) := C^+ f_{\mathbf{w},p-1}^{-1}(C^{+\top}\mathbf{x} - \boldsymbol{\alpha}), \tag{221}$$

The reason is as follows. We get

$$\begin{aligned}
\Delta_p(\Delta_p^+(\mathbf{x})) &= C^\top f_{\mathbf{w},p-1}(CC^+ f_{\mathbf{w},p-1}^{-1}(C^{+\top}\mathbf{x} - \boldsymbol{\alpha})) && (222)\\
&= C^\top f_{\mathbf{w},p-1}(f_{\mathbf{w},p-1}^{-1}(C^{+\top}\mathbf{x} - \boldsymbol{\alpha})) && (223)\\
&= C^\top (C^{+\top}\mathbf{x} - \boldsymbol{\alpha})) && (224)\\
&= C^\top C^{+\top}\mathbf{x}. && (225)
\end{aligned}$$

The second line follows from the assumption that $f_{\mathbf{w},p-1}^{-1}(C^{+\top}\mathbf{x} - \boldsymbol{\alpha}) \in \mathrm{Im}(C)$. Thus,

$$\begin{aligned}
\Delta_p^+(\Delta_p(\Delta_p^+(\mathbf{x}))) &= C^+ f_{\mathbf{w},p-1}^{-1}(C^{+\top}C^\top C^{+\top}\mathbf{x} - \boldsymbol{\alpha}) && (226)\\
&= C^+ f_{\mathbf{w},p-1}^{-1}(C^{+\top}\mathbf{x} - \boldsymbol{\alpha}) && (227)\\
&= \Delta_p^+ \mathbf{x}. && (228)
\end{aligned}$$

From this property, if we substitute

$$\mathbf{x} = \Delta_p^+ \left(\frac{\lambda}{p}(\mathbf{e}_i - \mathbf{e}_j)\right) \tag{229}$$

the Eq. (216) satisfied. Therefore, the next question is what $\boldsymbol{\alpha}$ is. However, we do not know even if such $\boldsymbol{\alpha}$ satisfying Eq. (220) exists or not.