# A Triangle Inequality for $p$-Resistance

**Mark Herbster**
Department of Computer Science
University College London
Gower Street, London WC1E 6BT, England, UK
m.herbster@cs.ucl.ac.uk

## Abstract

The geodesic distance (path length) and effective resistance are both metrics defined on the vertices of a graph. The effective resistance is a more refined measure of connectivity than the geodesic distance. For example if there are $k$ edge disjoint paths of geodesic distance $d$ between two vertices, then the effective resistance is no more than $\frac{d}{k}$. Thus, the more paths, the closer the vertices. We continue the study of the recently introduced *$p$-effective resistance* [9]. The main technical contribution of this note is to prove that the $p$-effective resistance is a metric for $p \in (1, 2]$ and obeys a strong triangle inequality. Given an efficient method to compute the $p$-effective resistance then an easy consequence of this inequality is that we may efficiently find a $k$-center clustering within a factor of $2^{p-1}$ from the optimal clustering with respect to $p$-effective resistance.

## 1 Introduction

Learning a function defined on a graph has received considerable attention in machine learning. A common approach is to represent functions defined on a graph by a Hilbert space associated with the graph Laplacian. The norm induced by the graph Laplacian is a natural measure of smoothness of these functions. If we are given a partial labeling of the graph this set-up is often referred to as semi-supervised learning [2, 14, 18, 17, 10]. The unsupervised learning of a labeling is often referred to as clustering (community detection) see for example [15] and [6, Section VII]. In machine learning and machine vision, recently, a generalization of the graph Laplacian to a $p$-(graph) Laplacian has been discussed in [16, 4, 9]. The dual norm associated with the $p$-Laplacian induces a metric between vertices which measures connectivity. In [9] the properties of the $p$th power of the dual norm were found to be analogous to the electrical network concept of effective resistance.

A $p$-resistive network is an undirected graph with an edge-resistance (a positive scalar) associated with each edge. This may be viewed as a generalization of both an electrical network ($p = 2$) and of an undirected flow ("pipe") network ($p = 1$). Such networks (graphs) are commonly used in graph-based semi-supervised learning. For semi-supervised learning, we are given a fixed set of objects, some of which are labeled and some of which are unlabeled, and we wish to predict the unlabeled objects. A graph is then defined where an edge between objects indicates similarity between objects. If the graph is weighted then the weights indicate the degree of similarity (inverse resistance). These include the min-cut method of [3] ($p = 1$) and the harmonic energy (power) minimization ($p = 2$) procedure of [18] (also [1]). We interpret these methods as specific instances of the minimization of a $p$-power [9]. When $p = 2$ the analogy is that the graph is an electrical network [5]; the edges are now resistors whose edge-resistance is the reciprocal of the similarity. The fixed labels from $\{-1, 1\}$ now correspond to voltage constraints and the algorithm for labeling the graph is then to find the set of consistent voltages which minimize the power and then to predict with the "sign" of the voltages. In the case $p = 1$ this is equivalent to finding the label-consistent min-cut.

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[1]. In fact the effective resistance induces a metric on the vertices of the graph, see for example [13]. Specifically it obeys the the triangle inequality, that is, given vertices $v_a, v_b$, and $v_c$

$$r_{\mathcal{G},2}(a,c) \leq r_{\mathcal{G},2}(a,b) + r_{\mathcal{G},2}(b,c)\,,$$

where $r_{\mathcal{G},2}(s,t)$ denotes the effective resistance between vertex $v_s$ and $v_t$ on the electric network as determined by the graph $\mathcal{G}$ and the associated set of edge resistances. For a flow network the 1-*effective resistance* denoted $r_{\mathcal{G},1}(s,t)$ may be defined to be the reciprocal of the minimum value of a separating-cut of $v_s$ and $v_t$ where a cut is a set of edges and a cut separates two vertices in a graph if after removal of the cut edges there is no path between the two vertices. Then the value of the cut is the sum of the reciprocals of the edge-resistances constituting the cut. Gomory and Hu [7] observed that the following stronger triangular inequality,

$$r_{\mathcal{G},1}(a,c) \leq \max(r_{\mathcal{G},1}(a,b), r_{\mathcal{G},1}(b,c))\,,$$

holds for flow networks. The key technical contribution of this note is then to prove the following triangular inequality for the *p-effective resistance* (see Definition 1)

$$r_{\mathcal{G},p}(a,c) \leq \left(r_{\mathcal{G},p}(a,b)^{\frac{1}{p-1}} + r_{\mathcal{G},p}(b,c)^{\frac{1}{p-1}}\right)^{p-1}$$

which smoothly interpolates from the triangular inequalities at $p = 1$ to $p = 2$.

In the following section we provide the formal definition of $p$-effective resistance. Then we recall the results of [9] in Theorem 1 which characterize the sense in which $p$-effective resistance is a measure of (inverse) connectivity. In Section 3, in Theorem 2, we prove the triangle inequality for $p$-resistance. We conclude in Section 4 with an observation about the farthest-first heuristic for $k$-center clustering in $p$-resistance.

## 2 $p$-Resistive networks

Let $\mathbb{N}$ be the set of natural numbers and $\mathbb{N}_\ell := \{1,\ldots,\ell\}$. If $\boldsymbol{z} \in \mathbb{R}^n$ then let $\|\boldsymbol{z}\|_p := \sqrt[p]{\sum_{i=1}^n |z_i|^p}$ denote the $p$-norm when $p \in [1,\infty)$. Given a seminorm $\|\cdot\| : \mathbb{R}^n \to \mathbb{R}$ the *dual* seminorm $\|\cdot\|^* : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is defined on the vector space of linear functionals $Z : \mathbb{R}^n \to \mathbb{R}$ as

$$\|Z\|^* := \sup_{\boldsymbol{w}\in\mathbb{R}^n}\left\{\frac{|Z(\boldsymbol{w})|}{\|\boldsymbol{w}\|}\right\} = \left[\inf_{\boldsymbol{w}\in\mathbb{R}^n}\{\|\boldsymbol{w}\| : Z(\boldsymbol{w}) = 1\}\right]^{-1}.$$

The canonical basis vectors of $\mathbb{R}^n$ we denote as $\boldsymbol{e}_1,\ldots,\boldsymbol{e}_n$ with corresponding functionals $E_i(\boldsymbol{w}) := \boldsymbol{e}_i^\top \boldsymbol{w}$.

A weighted graph $\mathcal{G} = (V, E, \boldsymbol{A})$ is a collection of vertices $V = \{v_1,\ldots,v_n\}$ joined by connecting (possibly weighted) edges. Denote $i \sim j$ whenever $v_i$ and $v_j$ are connected by an edge. We consider *undirected* weighted graphs so that $E := \{(i,j)|i \sim j\}$ is the set of unordered pairs of adjacent vertex indexes. A graph is *connected* if there does not exist partitioning vertex sets $V_a, V_b \subset V, V_a \cup V_b = V$ such that for every pair of vertices $v_a \in V_a$ and $v_b \in V_b$ there does not exist any edge $(a,b) \in E$. Associated with each edge $(i,j) \in E$ is a weight $A_{ij} > 0$ and $A_{ij} = 0$ if $(i,j) \notin E$, so that $\boldsymbol{A}$ is the weighted symmetric *adjacency matrix*. For compactness in discussion we will now refer to a weighted graph $\mathcal{G} = (V, E, \boldsymbol{A})$ as a graph $\mathcal{G} = (V, E)$ where the implicit adjacency matrix $\boldsymbol{A}$ is understood. In this paper we always assume that graphs are connected.

A *labelling* $\boldsymbol{u} \in \mathbb{R}^n$ of an $n$-vertex graph $\mathcal{G}$ is viewed as a function $\boldsymbol{u} : V_{\mathcal{G}} \to \mathbb{R}$ defined on the vertices of $\mathcal{G}$ whereby $u_i$ corresponds to the label of $v_i$. We introduce a class of *Laplacian p-seminorms* defined on the space of graph labellings: if $\boldsymbol{u} \in \mathbb{R}^n$ then

$$\|\boldsymbol{u}\|_{\mathcal{G},p} := \left(\sum_{(i,j)\in E_{\mathcal{G}}} A_{ij}|u_i - u_j|^p\right)^{\frac{1}{p}}. \tag{1}$$

[1]This is distinct and strictly smaller than the edge-resistance associated to an edge connecting the vertices unless removing this edge separate the networks into distinct components.

These $p$-seminorms generalize the commonly used "smoothness functional" $\boldsymbol{u}^T L \boldsymbol{u}$ [1, 18] where $L$ is the graph Laplacian, and as such measure the complexity of graph labellings.

When $p = 2$ there is an established natural connection [5] between graphs and resistive networks where each edge $(i, j) \in E_{\mathcal{G}}$ is viewed as a resistor with resistance $\pi_{ij} := \frac{1}{A_{ij}}$. We exploit this analogy so that a set of label constraints $\{(v_1, y_1), \ldots, (v_\ell, y_\ell)\} \in (V_{\mathcal{G}} \times \mathbb{R})^\ell$ are interpreted as (the effect of) voltage sources applied to the relevant vertices. This leads to following definition of the *power* for a network with voltage constraints,

$$\min_{\boldsymbol{u} \in \mathbb{R}^n} \{\|\boldsymbol{u}\|_{\mathcal{G},p}^p : u_1 = y_1, \ldots, u_\ell = y_\ell)\},\ (p \geq 1)\,.$$

Since the graph is assumed connected there is a unique minimizer if the set of constraints is nonempty. The *effective resistance* is the voltage difference needed to induce a unit "current" flow between $v_i$ and $v_j$. With the above definition of power it is natural to generalize the effective resistance as follows,

**Definition 1.** *The* $p$-(effective) resistance *between vertex* $v_i$ *and* $v_j$ *is*

$$r_{\mathcal{G},p}(i,j) := \left(\|E_i - E_j\|_{\mathcal{G},p}^*\right)^p \tag{2}$$

$$= \left(\min_{\boldsymbol{u} \in \mathbb{R}^n} \left\{\|\boldsymbol{u}\|_{\mathcal{G},p}^p : u_i = 1, u_j = 0\right\}\right)^{-1}, \tag{3}$$

where (3) follows as $\|\boldsymbol{u}\|_{\mathcal{G},p}^p = \|\boldsymbol{u} + k\mathbf{1}\|_{\mathcal{G},p}^p$ for $k \in \mathbb{R}$. We will now abbreviate $p$-effective resistance to $p$-resistance. In practice the $p$-resistance between all pairs of vertices may be efficiently computed in the case $p = 1$ by construction of a Gomory-Hu tree [7] and for $p = 2$ by computing the pseudoinverse of the graph Laplacian (see [13]). For other values of $p$ we may apply conjugate-gradient directly to the unconstrained optimization given by (3).

The following theorem summarizes some of the characteristics of the $p$-resistance.

**Theorem 1** ([9, Section 4.1.2])**.** *For $p \in (1, \infty)$ we have the following properties.*

1. *(Resistors in series) Consider a path graph $\mathcal{G}$, with $V_{\mathcal{G}} = \{v_1, v_2, \ldots, v_n\}, E_{\mathcal{G}} = \{(1,2), (2,3)...(n-1,n)\}$ and edge resistances $\{\pi_{12}, \pi_{23}, \ldots, \pi_{n-1,n}\}$. Then*

$$r_{\mathcal{G},p}(1,n) = \left(\sum_{i=1}^{n-1} \pi_{i,i+1}^{\frac{1}{p-1}}\right)^{p-1}.$$

2. *(Resistors in parallel) Consider a multigraph $\mathcal{G}$ with two vertices $V_{\mathcal{G}} = \{v_a, v_b\}$ joined by $m$ edges with edge resistances $\{\pi_k\}_{k=1}^m$. Then*

$$r_{\mathcal{G},p}(a,b) = \left(\sum_{k=1}^{m} \frac{1}{\pi_k}\right)^{-1}.$$

3. *(2-Port black box principle) Given a subgraph $\mathcal{G}' \subseteq \mathcal{G}$ with only "2 ports" at $v_a$ and $v_b$, that is if $(i,j) \in E_{\mathcal{G}}$ and $v_i \in V_{\mathcal{G}} \setminus V_{\mathcal{G}'}$ and $v_j \in V_{\mathcal{G}'}$ then $v_j \in \{v_a, v_b\}$. We may then construct a graph $\mathcal{G}''$ that replaces the subgraph $\mathcal{G}'$ with a single edge. So that $\mathcal{G}''$ is "electrically" equivalent to $\mathcal{G}$ if there are no voltages constraints on $V_{\mathcal{G}'} \setminus \{v_a, v_b\}$. Thus if $\mathcal{G}'' = (V_{\mathcal{G}''}, E_{\mathcal{G}''})$ is constructed so that*

$$V_{\mathcal{G}''} := (V_{\mathcal{G}} \setminus V_{\mathcal{G}'}) \cup \{v_a, v_b\},\ \ E_{\mathcal{G}''} := \{(i,j) \in E_{\mathcal{G}} : v_i, v_j \in V_{\mathcal{G}''}\} \cup \{(a,b)\}\,,$$

*new edge resistance $\pi_{ab} = r_{\mathcal{G}',p}(a,b)$ and if $\boldsymbol{z} \in \mathbb{R}^m$ then*

$$\|\boldsymbol{z}\|_{\mathcal{G}'',p}^p = \operatorname*{argmin}_{\boldsymbol{u} \in \mathbb{R}^n} \{\|\boldsymbol{u}\|_{\mathcal{G},p}^p : u_1 = z_1, \ldots, u_m = z_m\}\,,$$

*with $m = |V_{\mathcal{G}''}|$ and $n = |V_{\mathcal{G}}|$.*

4. *(Rayleigh's monotonicity principle) Given $\mathcal{G}$ with adjacency matrix $\boldsymbol{A}$. Let $\mathcal{G}'$, with adjacency $\boldsymbol{A}'$, be identical to $\mathcal{G}$ except for the increase in the weight of one arbitrary edge $(a,b)$, so that $A'_{ab} = A'_{ba} = A_{ab} + \delta$ for $\delta > 0$. Then for arbitrary vertices $v_i$ and $v_j$,*

$$r_{\mathcal{G},p}(i,j) \geq r_{\mathcal{G}',p}(i,j)\,.$$

5. *("p"-monotonicity) Given $\mathcal{G}$ and vertices $v_i$ and $v_j$ if $p \leq s$ then*

$$r_{\mathcal{G},p}(i,j) \leq r_{\mathcal{G},s}(i,j)\,.$$

We observe that the "resistors in parallel" law is unchanged as a function of $p$ while the "serial" law generalizes by becoming a $(\frac{1}{p-1})$-norm on the edge resistances. Combining these laws with the 2-port black box and Rayleigh's monotonicity principle demonstrates that the p-resistance between two vertices is bounded above by $\frac{d}{k}$ where $k$ is the number of edge disjoint paths and $d$ is the minimum path length as determined by the serial law.

## 3 Triangle Inequality

We first need a straightforward generalization of the well-known maximum principle for electric networks (see for example [5]).

**Lemma 1** (Maximum principle)**.** *Given a network with voltage constraints, the minimizing voltages are in the interior of the constraints. Thus given a connected graph $\mathcal{G}$, a constant $p \geq 1$, and $\boldsymbol{y} \in \mathbb{R}^{\ell}$ then if*

$$\boldsymbol{u}^* := \operatorname*{argmin}_{\boldsymbol{u}\in\mathbb{R}^n}\{\|\boldsymbol{u}\|^p_{\mathcal{G},p} : u_1 = y_1, \ldots, u_\ell = y_\ell)\} \tag{4}$$

*then*

$$\max_{i\in\mathbb{N}_n} u^*_i = \max_{i\in\mathbb{N}_\ell} y_i \quad \text{and} \quad \min_{i\in\mathbb{N}_n} u^*_i = \min_{i\in\mathbb{N}_\ell} y_i\,. \tag{5}$$

*Proof.* Suppose (5) is false then without loss of generality, define

$$m := \max_{i\in\mathbb{N}_n} u^*_i > \max_{i\in\mathbb{N}_\ell} y_i$$

and let $m' < m$ be the value of the second largest component of component of $\boldsymbol{u}^*$ ($m' = \max_{i\in\mathbb{N}_n}\{u^*_i : u^*_i \neq m\}$). Now construct $\boldsymbol{u}'$ component-wise via

$$u'_i := \begin{cases} m' & u^*_i = m \\ u^*_i & u^*_i \neq m \end{cases} \quad (i = 1,\ldots,n)\,.$$

The vector $\boldsymbol{u}'$ is a feasible solution of the objective of (4), but $\|\boldsymbol{u}'\|^p_{\mathcal{G},p} < \|\boldsymbol{u}^*\|^p_{\mathcal{G},p}$ and this is a contradiction. □

The following is our triangular inequality for $p$-resistance. We may obtain an equality, for example, if we have a simple path graph with $a \sim b \sim c$ (and more generally if every path from $v_a$ to $v_c$ must contain $v_b$) by Theorem 1 (series law). The inequality (6) also implies that the "usual" triangle inequality holds for the $p$-resistance if $p \in (1,2]$ and cannot for $p \in (2,\infty)$ because of the equality on the path graph. With respect to (7) the fact that $\|\cdot\|^*_{\mathcal{G},p}$ is a semi-norm and $|\cdot|^s$ $(0 \leq s \leq 1)$ is a subadditive function implies the inequality is a triviality for $q \in (0,1]$ thus the "interesting" range is $q \in (1, \frac{p}{p-1})$.

**Theorem 2** (Triangle Inequality)**.** *Given a graph $\mathcal{G}$ and vertices $v_a$, $v_b$, and $v_c$ then*

$$r_{\mathcal{G},p}(a,c) \leq \left(r_{\mathcal{G},p}(a,b)^{\frac{1}{p-1}} + r_{\mathcal{G},p}(b,c)^{\frac{1}{p-1}}\right)^{p-1} \qquad p \in (1,\infty) \tag{6}$$

*and thus for all $0 < q \leq \frac{p}{p-1}$ we also have,*

$$\left(\|E_a - E_c\|^*_{\mathcal{G},p}\right)^q \leq \left(\|E_a - E_b\|^*_{\mathcal{G},p}\right)^q + \left(\|E_b - E_c\|^*_{\mathcal{G},p}\right)^q \qquad p \in (1,\infty) \tag{7}$$

*Proof.* Construct a graph $\tilde{\mathcal{G}}$, see figure 1, which consists of two duplicates $\mathcal{G}'$, $\mathcal{G}''$ of $\mathcal{G}$ joined together at the vertices $v'_b$ and $v''_b$ which are now identified as a single vertex. Thus

$$V_{\tilde{\mathcal{G}}} := \{v'_1,\ldots,v'_n,v''_1,\ldots,v''_{b-1},v''_{b+1},\ldots,v''_n\}$$

and

$$E_{\tilde{\mathcal{G}}} := \{(v'_i,v'_j) : (v_i,v_j) \in E_{\mathcal{G}}\} \cup \{(v''_i,v''_j) : (v_i,v_j) \in E_{\mathcal{G}} \text{ and } v_b \notin \{v_i,v_j\}\} \cup \{(v''_i,v'_b) : (v_i,v_b) \in E_{\mathcal{G}}\}$$

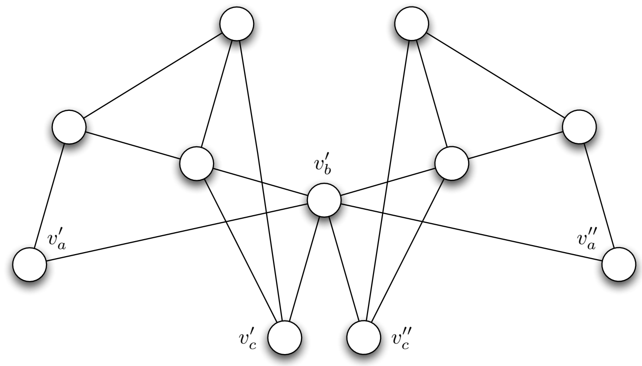


Figure 1: The graph $\tilde{\mathcal{G}}$

and the edge weights of $\tilde{\mathcal{G}}$ correspond to those of $\mathcal{G}$. We now argue that

$$r_{\tilde{\mathcal{G}},p}(a',c'') = \left(r_{\mathcal{G},p}(a,b)^{\frac{1}{p-1}} + r_{\mathcal{G},p}(b,c)^{\frac{1}{p-1}}\right)^{p-1} \quad p \in (1,\infty). \tag{8}$$

First observe that $r_{\mathcal{G},p}(a,b) = r_{\tilde{\mathcal{G}},p}(a',b')$ for if we define the power minimizer

$$\boldsymbol{w} := \underset{\boldsymbol{u}\in\mathbb{R}^n}{\operatorname{argmin}}\{\|\boldsymbol{u}\|^p_{\mathcal{G},p} : u_a = 1, u_b = 0\}$$

and the power minimizer

$$\tilde{\boldsymbol{w}} := \underset{\boldsymbol{u}\in\mathbb{R}^{2n-1}}{\operatorname{argmin}}\{\|\boldsymbol{u}\|^p_{\tilde{\mathcal{G}},p} : u_a' = 1, u_b' = 0\}$$

is "decoupled" as $\tilde{\boldsymbol{w}} = (\boldsymbol{w}' = \boldsymbol{w}, \boldsymbol{w}'' = \boldsymbol{0})$, and similarly $r_{\mathcal{G},p}(b,c) = r_{\tilde{\mathcal{G}},p}(b',c'')$ We may compute $r_{\tilde{\mathcal{G}},p}(a',c'')$ as follows

$$\begin{aligned}
r_{\tilde{\mathcal{G}},p}(a',c'') &= \left(\min_{\boldsymbol{u}\in\mathbb{R}^{2n-1}} \left\{\|\boldsymbol{u}\|^p_{\tilde{\mathcal{G}},p} : u_a' = 1, u_c'' = 0\right\}\right)^{-1} && (9)\\
&= \left(\min_{\lambda\in\mathbb{R}} \left[\begin{array}{r} \displaystyle\min_{\boldsymbol{u}\in\mathbb{R}^{2n-1}} \left\{\|\boldsymbol{u}\|^p_{\tilde{\mathcal{G}},p} : u_a' = 1, u_b' = \lambda\right\} \;+ \\ \displaystyle\min_{\boldsymbol{u}\in\mathbb{R}^{2n-1}} \left\{\|\boldsymbol{u}\|^p_{\tilde{\mathcal{G}},p} : u_b' = \lambda, u_c'' = 0\right\}\end{array}\right]\right)^{-1} && (10)\\
&= \left(\min_{\lambda\in\mathbb{R}} \left[\frac{|1-\lambda|^p}{r_{\mathcal{G},p}(a,b)} + \frac{|\lambda|^p}{r_{\mathcal{G},p}(b,c)}\right]\right)^{-1} && (11)
\end{aligned}$$

where (9) follows from (3). This optimization is split into separate optimizations coupled only via $\lambda$ in (10). Then since

$$|\alpha|^p\|\boldsymbol{u}\|^p_{\mathcal{G},p} = \|\alpha\boldsymbol{u} + k\boldsymbol{1}\|^p_{\mathcal{G},p}$$

and $r_{\mathcal{G},p}(a,b) = r_{\tilde{\mathcal{G}},p}(a',b')$ as well as $r_{\mathcal{G},p}(b,c) = r_{\tilde{\mathcal{G}},p}(b',c'')$ this gives (11). We observe that the minimizing $\lambda$ of (11) is

$$\lambda^* = \frac{r_{\mathcal{G},p}(b,c)^{\frac{1}{p-1}}}{r_{\mathcal{G},p}(a,b)^{\frac{1}{p-1}} + r_{\mathcal{G},p}(b,c)^{\frac{1}{p-1}}}$$

after substituting $\lambda = \lambda^*$ into the minimand of (11) then (8) follows immediately.

We now proceed to prove

$$r_{\mathcal{G},p}(a,c) \le r_{\tilde{\mathcal{G}},p}(a',c''), \tag{12}$$

which is equivalent to

$$\min_{\boldsymbol{u}\in\mathbb{R}^{2n-1}} \left\{\|\boldsymbol{u}\|^p_{\tilde{\mathcal{G}},p} : u_a' = 1, u_c'' = 0\right\} \le \|\boldsymbol{u}^*\|^p_{\mathcal{G},p}, \tag{13}$$

with

$$\boldsymbol{u}^* := \operatorname*{argmin}_{\boldsymbol{u}\in\mathbb{R}^n} \left\{ \|\boldsymbol{u}\|_{\mathcal{G},p}^p : u_a = 1, u_c = 0 \right\} .$$

We construct the vector $\tilde{\boldsymbol{u}} := (\tilde{\boldsymbol{u}}', \tilde{\boldsymbol{u}}'') \in \mathbb{R}^{2n-1}$ as

$$\tilde{u}'_i := \begin{cases} u^*_i & u^*_i > u^*_b \\ u^*_b & u^*_i \le u^*_b \end{cases} \quad (i = 1, \dots, n)\,; \tilde{u}''_i := \begin{cases} u^*_i & u^*_i < u^*_b \\ u^*_b & u^*_i \ge u^*_b \end{cases} \quad (i = 1, \dots, n-1)\,. \tag{14}$$

We infer that $u^*_b \in [0,1]$ from Lemma 1 thus $\tilde{u}'_a = 1$ and $\tilde{u}''_c = 0$ and therefore the vector $\tilde{\boldsymbol{u}}$ is a feasible solution to the objective of the left-hand-side of (13). We now define the three index sets,

$$L := \{i \in \mathbb{N}_n : u^*_i < u^*_b\}\,, M := \{i \in \mathbb{N}_n : u^*_i = u^*_b\} \text{ and } H := \{i \in \mathbb{N}_n : u^*_i > u^*_b\}$$

Which we use to compute $\|\tilde{\boldsymbol{u}}\|_{\tilde{\mathcal{G}},p}^p$ (where $\boldsymbol{A}$ is the adjacency matrix of $\mathcal{G}$),

$$\begin{aligned} \|\tilde{\boldsymbol{u}}\|_{\tilde{\mathcal{G}},p}^p = &\sum_{\{(i,j)\in L^2 : i<j\}} A_{ij}|\tilde{u}'_i - \tilde{u}'_j|^p + \sum_{\{(i,j)\in H^2 : i<j\}} A_{ij}|\tilde{u}'_i - \tilde{u}'_j|^p + \sum_{\{(i,j)\in M^2 : i<j\}} A_{ij}|\tilde{u}'_b - \tilde{u}'_j|^p \\ &+ \sum_{\{(i,j)\in L\times H\}} A_{ij}|\tilde{u}'_i - \tilde{u}'_j|^p + \sum_{\{(i,j)\in M\times L\cup H\}} A_{ij}|\tilde{u}'_b - \tilde{u}'_j|^p \\ &+ \sum_{\{(i,j)\in L^2 : i<j\}} A_{ij}|\tilde{u}''_i - \tilde{u}''_j|^p + \sum_{\{(i,j)\in H^2 : i<j\}} A_{ij}|\tilde{u}''_i - \tilde{u}''_j|^p + \sum_{\{(i,j)\in M^2 : i<j\}} A_{ij}|\tilde{u}'_b - \tilde{u}''_j|^p \\ &+ \sum_{\{(i,j)\in L\times H\}} A_{ij}|\tilde{u}''_i - \tilde{u}''_j|^p + \sum_{\{(i,j)\in M\times L\cup H\}} A_{ij}|\tilde{u}'_b - \tilde{u}''_j|^p \end{aligned} \tag{15}$$

eliminating "zero" terms we have,

$$\begin{aligned} \|\tilde{\boldsymbol{u}}\|_{\tilde{\mathcal{G}},p}^p = &\sum_{\{(i,j)\in H^2 : i<j\}} A_{ij}|\tilde{u}'_i - \tilde{u}'_j|^p + \sum_{\{(i,j)\in L\times H\}} A_{ij}|\tilde{u}'_i - \tilde{u}'_j|^p + \sum_{\{(i,j)\in M\times H\}} A_{ij}|\tilde{u}'_b - \tilde{u}'_j|^p \\ &+ \sum_{\{(i,j)\in L^2 : i<j\}} A_{ij}|\tilde{u}''_i - \tilde{u}''_j|^p + \sum_{\{(i,j)\in L\times H\}} A_{ij}|\tilde{u}''_i - \tilde{u}''_j|^p + \sum_{\{(i,j)\in M\times L\}} A_{ij}|\tilde{u}'_b - \tilde{u}''_j|^p \end{aligned} \tag{16}$$

rewriting using the definition of $\tilde{\boldsymbol{u}}$ in (14)

$$\begin{aligned} \|\tilde{\boldsymbol{u}}\|_{\tilde{\mathcal{G}},p}^p = &\sum_{\{(i,j)\in H^2 : i<j\}} A_{ij}|u^*_i - u^*_j|^p + \sum_{\{(i,j)\in L\times H\}} A_{ij}|u^*_b - u^*_j|^p + \sum_{\{(i,j)\in M\times H\}} A_{ij}|u^*_b - u^*_j|^p \\ &+ \sum_{\{(i,j)\in L^2 : i<j\}} A_{ij}|u^*_i - u^*_j|^p + \sum_{\{(i,j)\in L\times H\}} A_{ij}|u^*_i - u^*_b|^p + \sum_{\{(i,j)\in M\times L\}} A_{ij}|u^*_b - u^*_j|^p . \end{aligned} \tag{17}$$

We now compute

$$\begin{aligned} \|\boldsymbol{u}^*\|_{\mathcal{G},p}^p = &\sum_{\{(i,j)\in H^2 : i<j\}} A_{ij}|u^*_i - u^*_j|^p + \sum_{\{(i,j)\in L\times H\}} A_{ij}|u^*_i - u^*_j|^p + \sum_{\{(i,j)\in M\times H\}} A_{ij}|u^*_b - u^*_j|^p \\ &+ \sum_{\{(i,j)\in L^2 : i<j\}} A_{ij}|u^*_i - u^*_j|^p + \sum_{\{(i,j)\in M\times L\}} A_{ij}|u^*_b - u^*_j|^p . \end{aligned} \tag{18}$$

Now subtracting we have

$$\|\boldsymbol{u}^*\|_{\mathcal{G},p}^p - \|\tilde{\boldsymbol{u}}\|_{\tilde{\mathcal{G}},p}^p = \sum_{\{(i,j)\in L\times H\}} A_{ij}|u^*_i - u^*_j|^p - \sum_{\{(i,j)\in L\times H\}} A_{ij}\left(|u^*_i - u^*_b|^p + |u^*_b - u^*_j|^p\right) \tag{19}$$

therefore since

$$(|r| + |s|)^p \ge |r|^p + |s|^p \text{ for } p \ge 1\,, \tag{20}$$

we have that $\|\tilde{\boldsymbol{u}}\|_{\tilde{\mathcal{G}},p}^p \le \|\boldsymbol{u}^*\|_{\mathcal{G},p}^p$ and since $\tilde{\boldsymbol{u}}$ is a feasible solution for the minimand of the left-hand-side of (13) this proves (12).

Finally substituting (8) into (12) proves (6) from which (7) follows immediately. $\square$

## 4 Clustering with $p$-resistance

The metric $k$-center clustering problem is to find the solution to the following objective,

$$\min_{v_1^*,\ldots,v_k^* \in V} \max_{v \in V} \min_{i \in \mathbb{N}_k} d(v, v_i^*) \,. \tag{21}$$

Thus the goal is to find $k$ centers $v_1^*, \ldots, v_k^*$ such that maximum distance from any center is minimized where $d(\cdot,\cdot)$ is a metric on the set $V$. The "farthest-first" heuristic for this problem is known to give 2-opt clustering [8, 11] for this problem, which is matched by the result that there is no polynomial-time $(2-\epsilon)$-opt approximation algorithm [8, 12] unless $\mathbb{P} = \mathbb{NP}$. Given the strong triangle inequality proved for $p$-resistance we now argue that the "farthest-first" heuristic gives a $2^{p-1}$-opt algorithm for clustering the vertices of a graph by $p$-resistance by a simple modification of the original proofs.

**Input:** A set $V = v_1, \ldots, v_n$, an positive integer $k$ and a metric $d(V,V) \to \mathbb{R}$
**Initialization:** $\tilde{v}_1 = v_1$
**for** $t = 2, \ldots, k$ **do**
  $\tilde{v}_t = \operatorname{argmax}_{v \in V} \min_{i \in \mathbb{N}_{t-1}} d(v, \tilde{v}_i)$
**end for**[2]
**return** $\{\tilde{v}_1, \ldots, \tilde{v}_k\}$

Figure 2: Farthest-first clustering

**Theorem 3.** *Given a graph $\mathcal{G}$ the farthest first algorithm gives a $2^{p-1}$-opt $k$-center clustering with respect to the $p$-resistance for $p > 1$.*

*Proof.* Let,

$$C^* := \min_{v_1^*,\ldots,v_k^* \in V} \max_{v \in V} \min_{i \in \mathbb{N}_k} r_{\mathcal{G},p}(v, v_i^*)$$

where $\{v_1^*, \ldots, v_k^*\}$ is a minimizer and let

$$\tilde{C} := \max_{v \in V} \min_{i \in \mathbb{N}_k} r_{\mathcal{G},p}(v, \tilde{v}_i)$$

where $\{\tilde{v}_1, \ldots, \tilde{v}_k\}$ is the approximate solution returned by farthest-first algorithm, thus we prove

$$\tilde{C} \leq 2^{p-1} C^* \,.$$

Consider the construction of $\tilde{v}_1, \ldots, \tilde{v}_k$ each of the points must be separated from each other by at least $\tilde{C}$ further there must exist one additional point $\tilde{v}_{k+1}$ which is also separated by $\tilde{C}$ otherwise the farthest-first clustering would cost less than $\tilde{C}$. Now given these $k+1$ points $\tilde{v}_1, \ldots, \tilde{v}_{k+1}$ by the pigeonhole principle two of these points $\tilde{v}', \tilde{v}''$ must share a center $v^* \in \{v_1^*, \ldots, v_k^*\}$ such that $r_{\mathcal{G},p}(\tilde{v}', v^*) \leq C^*$ and $r_{\mathcal{G},p}(\tilde{v}'', v^*) \leq C^*$. An application of the $p$-resistance triangle inequality (6) gives,

$$\tilde{C} \leq r_{\mathcal{G},p}(\tilde{v}', \tilde{v}'') \leq 2^{p-1} C^* \,.$$

□

We observe that the farthest-first algorithm is optimal as $p \to 1$.

**Acknowledgements.** I would like to thank Massimiliano Pontil and Guy Lever for valuable discussions. Finally, I would like to thank the PASCAL 2 European network of excellence for supporting this work through the grant, "Data-Dependent Geometries and Structures : Analyses and Algorithms for Machine Learning."

[2]Ties may be resolved arbitrarily.

## References


[1] M. Belkin, I. Matveeva, and P. Niyogi. Regularization and semi-supervised learning on large graphs. In *Proc. of the 17–th Annual Conf. on Learning Theory (COLT'04)*, Banff, Alberta, 2004.

[2] M. Belkin and P. Niyogi. Semi-supervised learning on riemannian manifolds. *Machine Learning*, 56:209–239, 2004.

[3] A. Blum and S. Chawla. Learning from labeled and unlabeled data using graph mincuts. In *Proc. 18th International Conf. on Machine Learning*, pages 19–26. Morgan Kaufmann, San Francisco, CA, 2001.

[4] T. Bühler and M. Hein. Spectral clustering based on the graph $p$-laplacian. In *ICML*, 2009.

[5] P. G. Doyle and J. L. Snell. Random walks and electric networks, 2000.

[6] S. Fortunato. Community detection in graphs. *Physics Reports*, 486(3-5):75 – 174, 2010.

[7] R. E. Gomory and T. C. Hu. Multi-terminal network flows. *Journal of the Society for Industrial and Applied Mathematics*, 9(4):pp. 551–570, 1961.

[8] T. F. Gonzalez. Clustering to minimize the maximum intercluster distance. *Theoretical Computer Science*, 38:293 – 306, 1985.

[9] M. Herbster and G. Lever. Predicting the labelling of a graph via minimum p-seminorm interpolation. In *Proceedings of the 22nd Annual Conference on Learning Theory (COLT'09)*, 2009.

[10] M. Herbster, M. Pontil, and L. Wainer. Online learning over graphs. In *ICML '05: Proceedings of the 22nd international conference on Machine learning*, pages 305–312, New York, NY, USA, 2005. ACM.

[11] D. S. Hochbaum and D. B. Shmoys. A best possible heuristic for the k-center problem. *Mathematics of Operations Research*, 10(2):pp. 180–184, 1985.

[12] D. S. Hochbaum and D. B. Shmoys. A unified approach to approximation algorithms for bottleneck problems. *J. ACM*, 33(3):533–550, 1986.

[13] D. Klein and M. Randić. Resistance distance. *Journal of Mathematical Chemistry*, 12(1):81–95, 1993.

[14] R. I. Kondor and J. Lafferty. Diffusion kernels on graphs and other discrete input spaces. In *ICML 2002*, 2002.

[15] U. Luxburg. A tutorial on spectral clustering. *Statistics and Computing*, 17(4):395–416, 2007.

[16] D. Singaraju, L. Grady, and R. Vidal. P-brush: Continuous valued mrfs with normed pairwise distributions for image segmentation. In *CVPR 2009*.

[17] A. Smola and R. Kondor. Kernels and regularization on graphs. In *COLT 2003*, 2003.

[18] X. Zhu, Z. Ghahramani, and J. Lafferty. Semi-supervised learning using gaussian fields and harmonic functions. In *20-th International Conference on Machine Learning (ICML-2003)*, pages 912–919, 2003.