Multi-class Graph Clustering via Approximated Effective p-Resistance

Shota Saito and Mark Herbster, Dept. of Computer Science at University College London

Link to the paper

TL; DR

Proposed a multi-class graph clustering (MGC) algorithm using approximated p-resistance

Motivation

Want to exploit p-seminorm more

The most standard way: graph p-Laplacian

Graph p-seminorm for a graph $G = (V, E)$

$$\|\mathbf{x}\|_{G,p} := \|C\mathbf{x}\|_{\mathbf{w},p} = \Big(\sum_{ij \in V} a_{ij} |x_i - x_j|^p\Big)^{1/p}$$

C: incidence matrix, $\mathbf{w}$: weight vector, a_{ij}: weight between i and j

Spectral Clustering via graph p-Laplacian $\Delta_{G,p}\mathbf{x}$

$$(\Delta_p \mathbf{x})_i := \sum_{j \in V} a_{ij} |x_i - x_j|^{p-1} \operatorname{sgn}(x_i - x_j), \ \langle \mathbf{x}, \Delta_p \mathbf{x} \rangle = \|\mathbf{x}\|^p_{G,p}$$

But practically difficult to obtain higher order eigenpairs of p-Laplacian -> How can we exploit p-seminorm more for MGC?

Proposed Alternative: p-Resistance

(Effective) p-resistance

$$r_{G,p}(i,j) := \big(\min_{\mathbf{x}}\{\|\mathbf{x}\|^p_{G,p} \ \text{s.t.}\ x_i - x_j = 1\}\big)^{-1}.$$

Advantage: Triangle Inequality [Herbster '10]

$$r^{1/(p-1)}_{G,p}(i,j) \le r^{1/(p-1)}_{G,p}(i,\ell) + r^{1/(p-1)}_{G,p}(\ell,j)$$

-> motivates us to use this for MGC

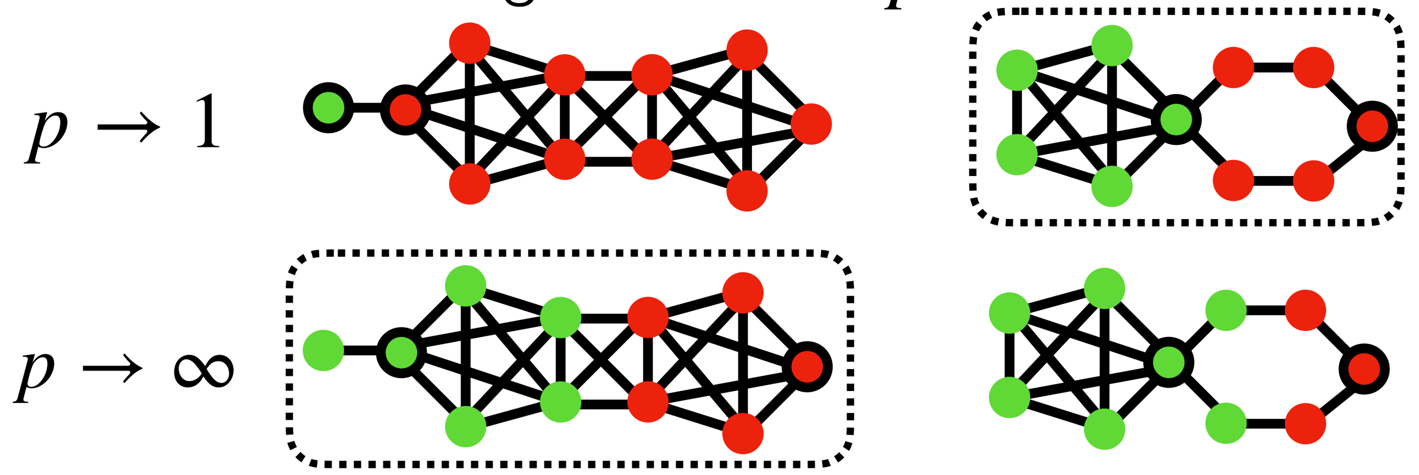
Difficulties of p-resistance

- Expensive to compute for all pairs
- Triangle inequality fully motivates the use for MGC

-> This work addressed these

How p works

The results of k-center algorithm with p-resistance



$p \to 1$: behave like a min cut

$p \to \infty$: preference for smaller shortest-path distances between vertices in the cluster.

Approximated p-resistance

Preliminaries: $p = 2$ case

$$r_{G,p}(i,j) = \|L^+\mathbf{e}_i - L^+\mathbf{e}_j\|^2_{G,2} \qquad L\text{: Laplacian}$$

Proposed Approximation

$$r_{G,p}(i,j) \approx \|L^+\mathbf{e}_i - L^+\mathbf{e}_j\|^q_{G,p} \qquad \frac{1}{p} + \frac{1}{q} = 1$$

Approximation Guarantee (Thm 3.3-Prop 3.5)

$$\frac{1}{\alpha^p_{G,p}} \|L^+\mathbf{e}_i - L^+\mathbf{e}_j\|^p_{G,q} \le r_{G,p}(i,j) \le \|L^+\mathbf{e}_i - L^+\mathbf{e}_j\|^p_{G,q}$$

$$\alpha_{G,p} := \|W^{1/p} C C^+ W^{-1/p}\|_p, \quad W := \operatorname{diag}(\mathbf{w})$$

We have the bound for $\alpha_{G,p}$ $\qquad \alpha_{G,p} \le m^{|1/2 - 1/p|}$

Note 1: Empirically $\alpha_{G,p}$ is far smaller than this bound

Note 2 (**attn!**): $r_{G,p}(i,j) = \|L^+\mathbf{e}_i - L^+\mathbf{e}_j\|^p_{G,p}$ for tree

Justification for the use of p-resistance from SSL

Semi-supervised learning (SSL) problem

$$\mathbf{x}^{*ij} := \arg\min_{\mathbf{x}}\{\|\mathbf{x}\|^p_{G,p} \ \text{s.t.}\ x_i - x_j = 1\}$$

From the SSL view, the third point x^{*ij}_ℓ shows which is closer, i or j

"Translation" of this SSL into p-resistance

$$x^{*ij}_\ell - x^{*ij}_j \ge x^{*ij}_i - x^{*ij}_\ell \iff r_{G,p}(j,\ell) \ge r_{G,p}(\ell,i).$$

The p-resistance is a "translation" of the SSL view

Note: This was an open problem since [Alamgir+ '11]. Now we resolved this.

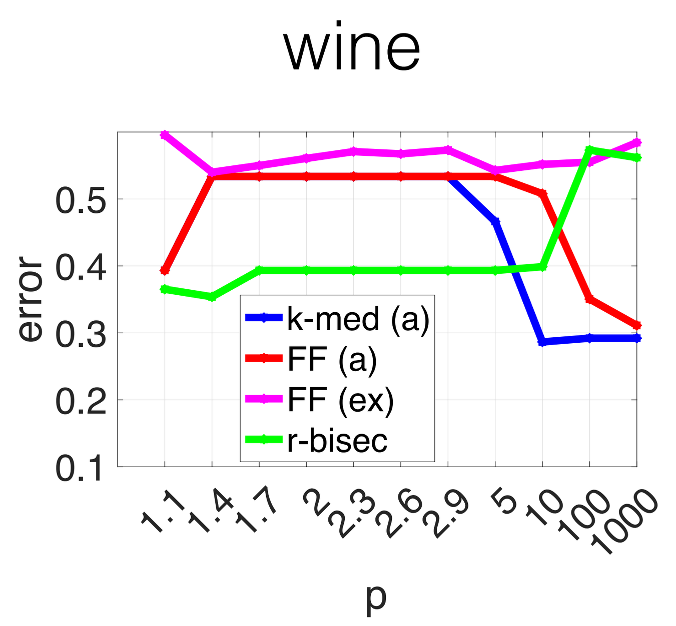
Proposed Algorithm

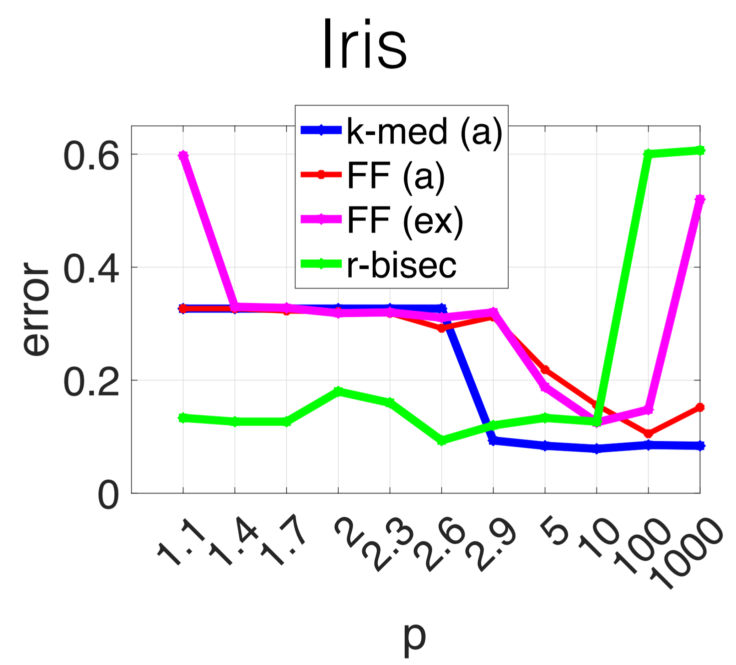
1. Compute the approx. p-res $r^{1/(p-1)}_{G,p}(i,j)$ for all pairs
2. Apply k-meodoids to this approx. p-res.

Experiments

Obj: To confirm if ours can outperform the existing methods

Comparison: Farthest first with approx. p-res
Farthest first with exact p-res
Spec. clus. Using p-Laplacian [Bühler+ 09]





Ours outperforms the existing ones