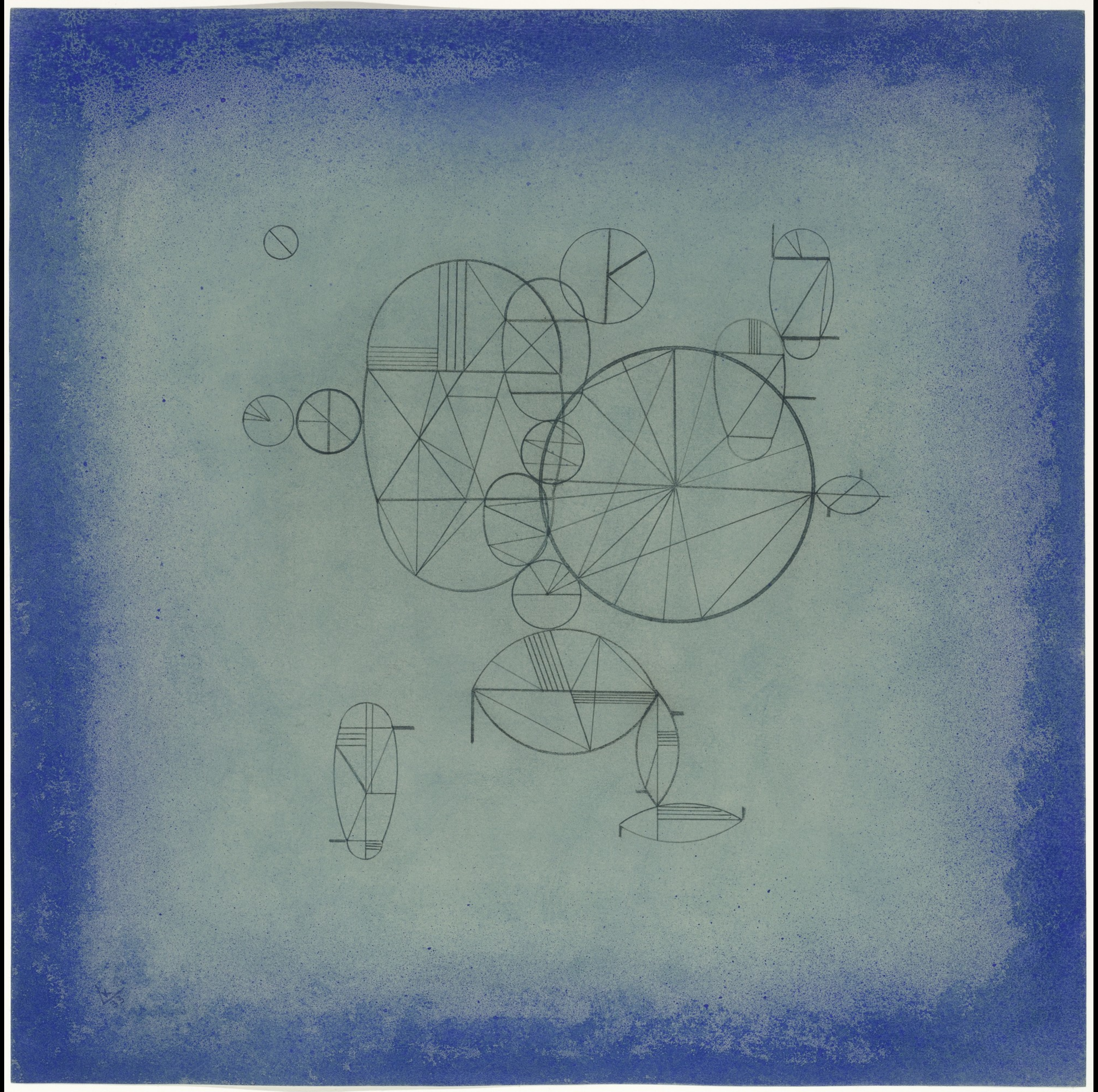


Image Credits:
Wassily Kandinsky -
Round poetry 1933

Multiway p-Spectral Clustering on Grassmann Manifolds

D. Pasadakis, C. L. Alappat,
O. Schenk, G. Wellein

Multiway p-Spectral Clustering on Grassmann Manifolds

D. Pasadakis[1], C.L. Alappat[2], O. Schenk[1], G. Wellein[2]

[1] Università della Svizzera italiana
Institute of Computing

Università della Svizzera italiana
Institute of Computing CI

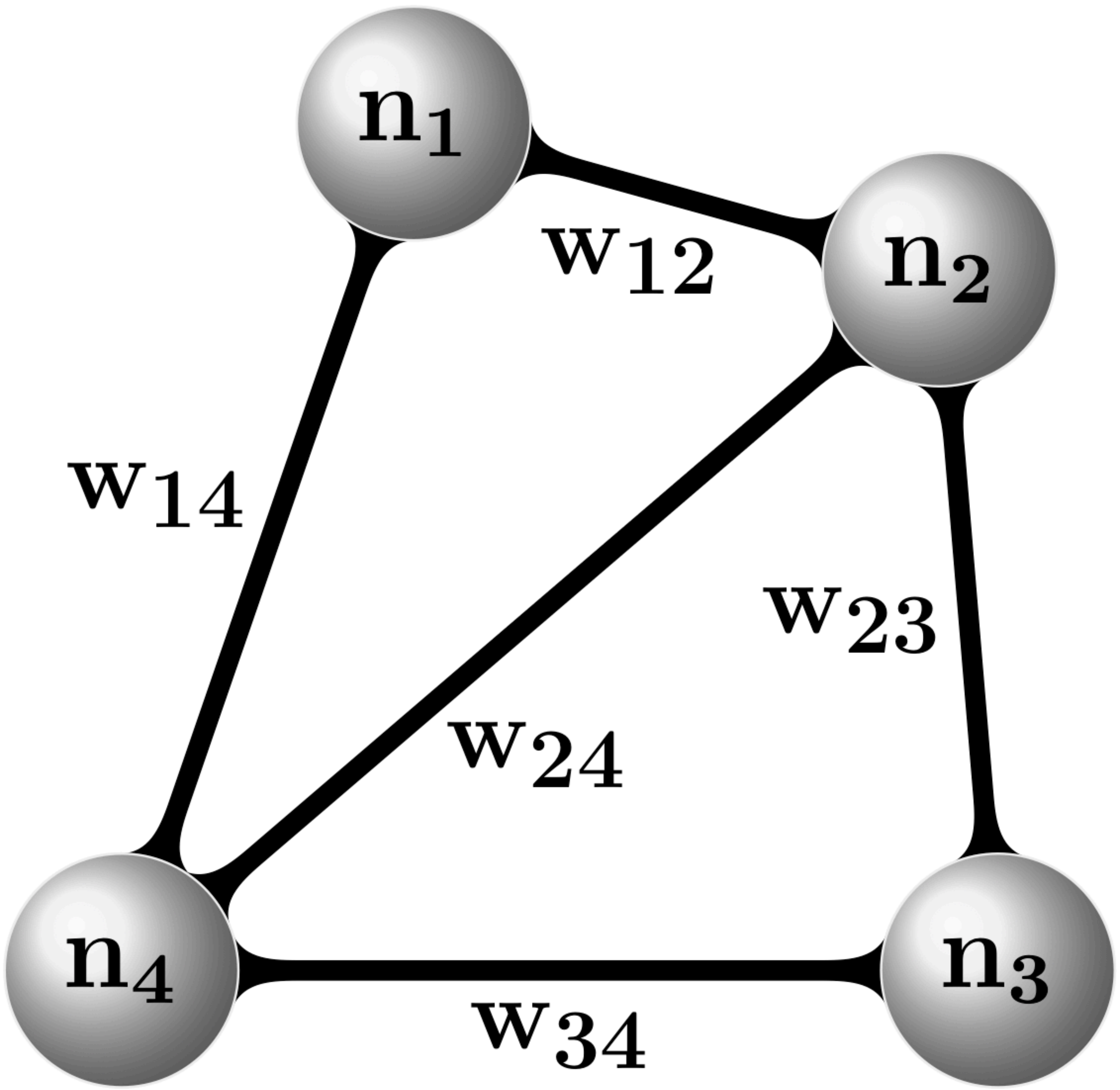
[2] Department of Computer Science
Friedrich-Alexander-Universität Erlangen-Nürnberg

May 17, 2021

2-norm Definitions

For a graph $\mathcal{G}(V, E, W)$

- Adjacency: $\boldsymbol{W} \in \mathbb{R}^{n\times n}$, Degree: $\boldsymbol{D} \in \mathbb{R}^{n\times n}$.
- Graph Laplacian: $\boldsymbol{\Delta}_2 \in \mathbb{R}^{n\times n}$.



$$\boldsymbol{W} = \begin{bmatrix} 0 & w_{12} & 0 & w_{14} \\ w_{12} & 0 & w_{23} & w_{24} \\ 0 & w_{23} & 0 & w_{34} \\ w_{14} & w_{24} & w_{34} & 0 \end{bmatrix}, \quad d_{ii} = \begin{bmatrix} \sum_j w_{1j} \\ \sum_j w_{2j} \\ \sum_j w_{3j} \\ \sum_j w_{4j} \end{bmatrix}.$$

$$\boldsymbol{\Delta}_2 = \boldsymbol{D} - \boldsymbol{W},$$

$$\lambda_1 \le \lambda_2 \le \cdots \le \lambda_n, \lambda_1 = 0, \boldsymbol{v}^{(1)} = c \cdot \boldsymbol{e}.$$

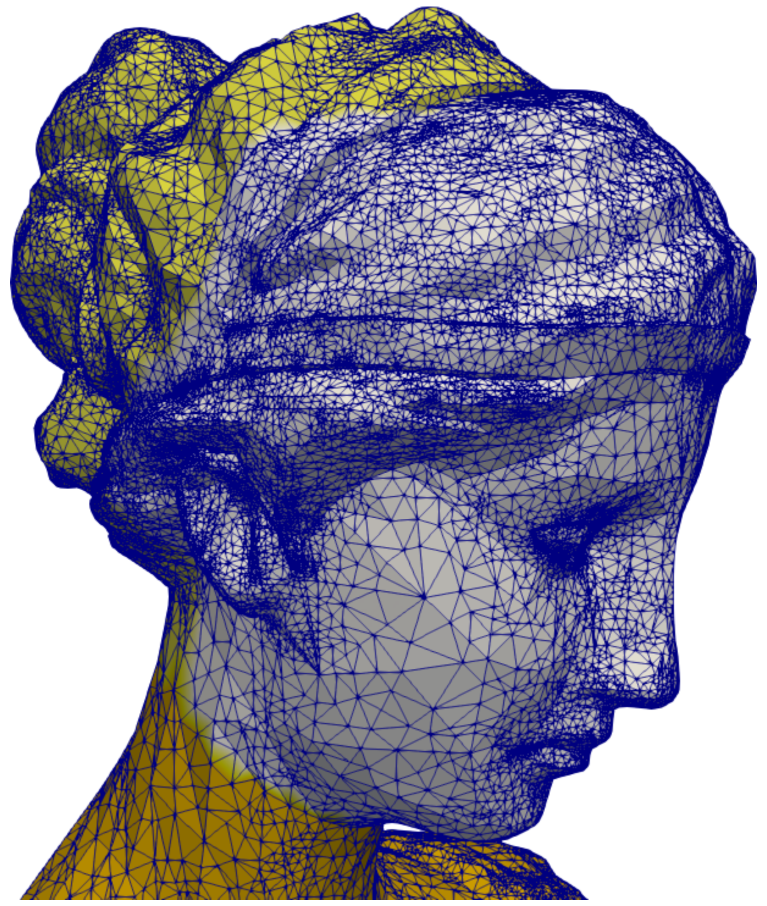
Balanced Cut Metrics - Multiway

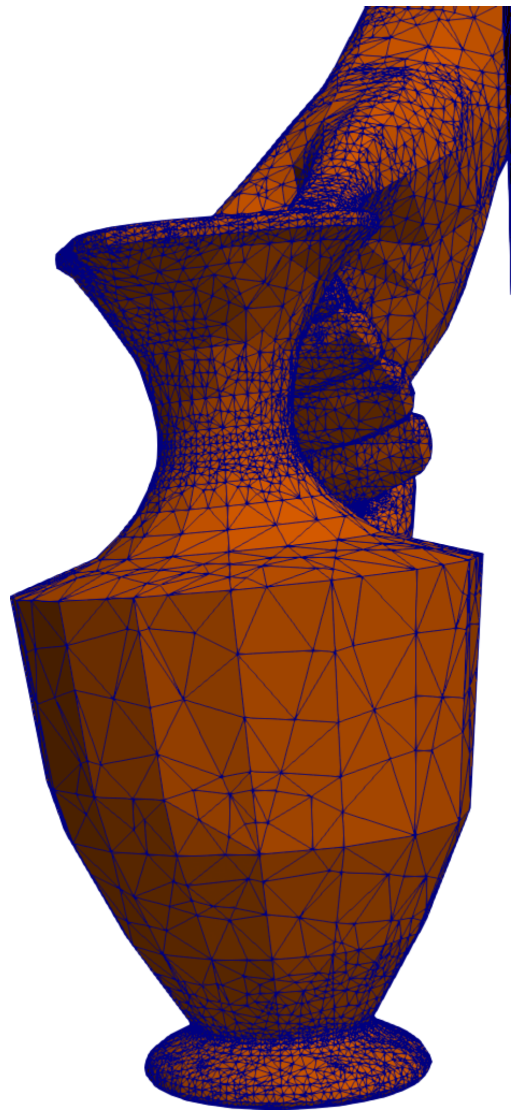
For subsets $C_1, \dots, C_k$

- $\mathrm{cut}(C, \overline{C}) = \sum_{i \in C, j \in \overline{C}} w_{ij}$
  $\mathrm{vol}(C) = \sum_{i \in C} d_{ii}$
- Minimize the balanced cut criteria

$$\mathrm{RCut}(C_1, \dots, C_k) = \sum_{i=1}^{k} \frac{\mathrm{cut}(C_i, \overline{C_i})}{|C_i|}$$

$$\mathrm{NCut}(C_1, \dots, C_k) = \sum_{i=1}^{k} \frac{\mathrm{cut}(C_i, \overline{C_i})}{\mathrm{vol}(C_i)}$$





p-norm Definitions

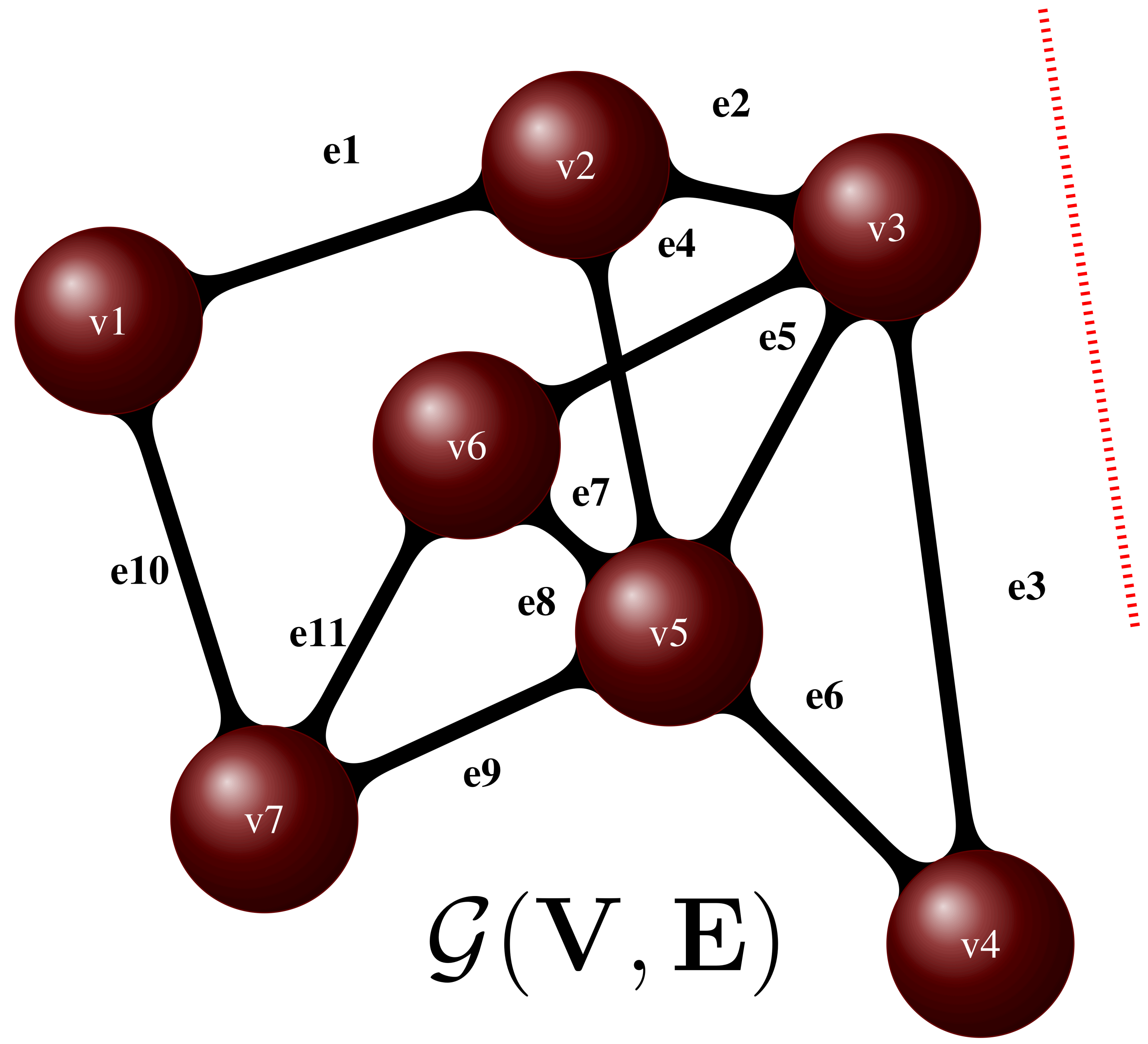
For a graph $\mathcal{G}(V, E, W)$, $p \in (1, 2]$

- $\phi_p(x) = |x|^{p-1}\mathrm{sign}(x)$,
- p-norm: $\|\boldsymbol{u}\|_p = \sqrt[p]{\sum_{i=1}^{n} |u_i|^p}$.

* p-eigenspectrum: $(\Delta_p \boldsymbol{v})_i = \lambda_p \phi_p(v_i)$,
* $\lambda_p^{(1)} = 0 \rightarrow \#$ of conn. components,
* $\boldsymbol{v}_p^{(1)} = c \cdot \boldsymbol{e}$,
* $\mathrm{cut} \leq \mathrm{pcut} \leq p(\max_{i \in V} d_{ii})^{\frac{p-1}{p}} (\mathrm{cut})^{\frac{1}{p}}$.

For a node $i \in V$

$$(\Delta_p \boldsymbol{u})_i = \sum_{j \in V} w_{ij} \phi_p(u_i - u_j)$$



p-norm Definitions

For a graph $\mathcal{G}(V, E, W)$, $p \in (1, 2]$

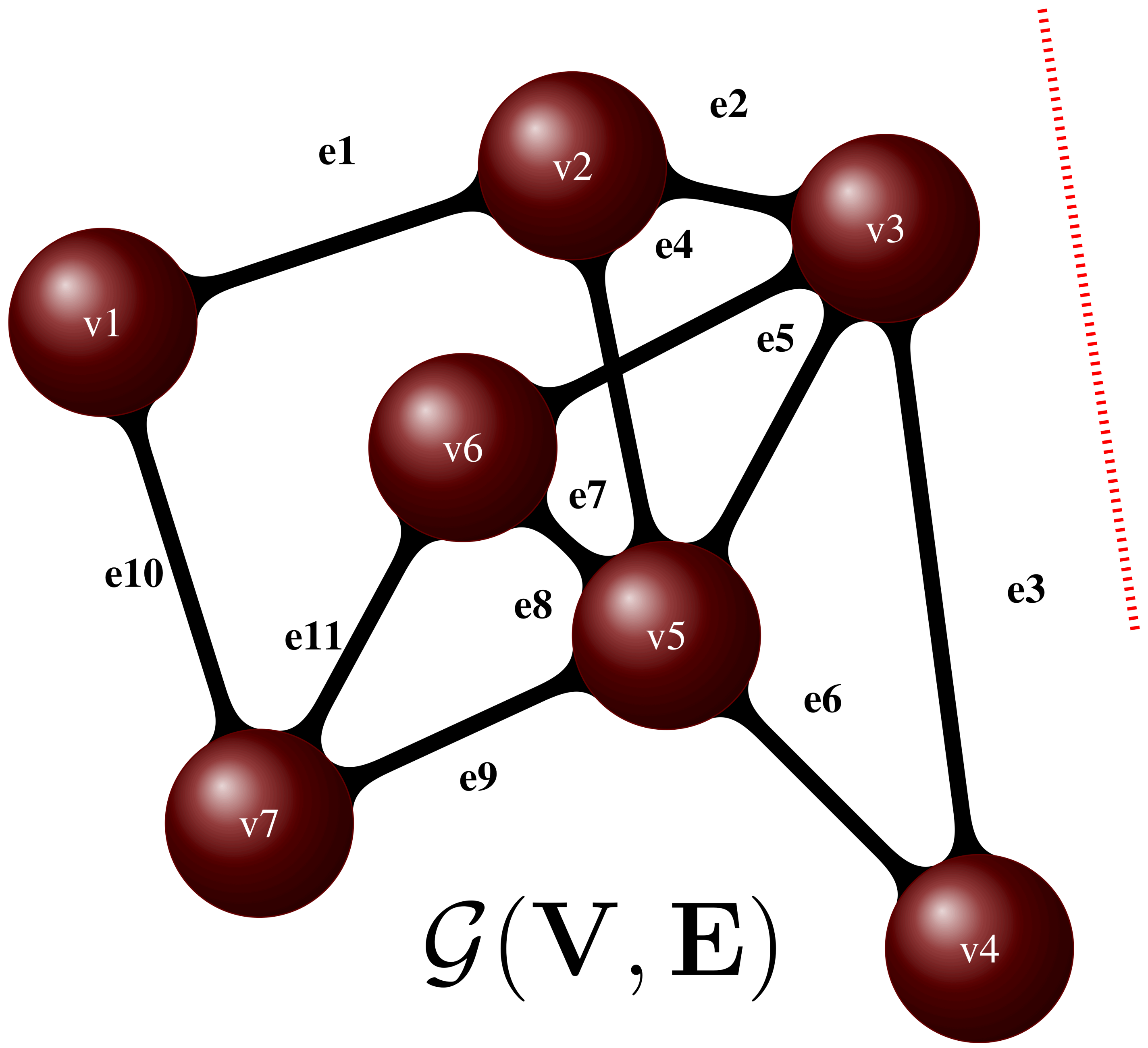
- $\phi_p(x) = |x|^{p-1}\mathrm{sign}(x)$,
- p-norm: $\|\boldsymbol{u}\|_p = \sqrt[p]{\sum_{i=1}^{n} |u_i|^p}$.

* p-eigenspectrum: $(\Delta_p \boldsymbol{v})_i = \lambda_p \phi_p(v_i)$,
* $\lambda_p^{(1)} = 0 \rightarrow \#$ of conn. components,
* $\boldsymbol{v}_p^{(1)} = c \cdot \boldsymbol{e}$,
* $\mathrm{cut} \leq \mathrm{pcut} \leq p \left(\max_{i \in V} d_{ii}\right)^{\frac{p-1}{p}} (\mathrm{cut})^{\frac{1}{p}}$.

For a node $i \in V$

$$(\Delta_p \boldsymbol{u})_i = \sum_{j \in V} w_{ij} \phi_p(u_i - u_j)$$

Gajewski & Gärtner, 2001
Amghibech, 2006
Bühler & Hein, 2009



Spectral Direct Multiway Clustering

Avoid

- lack of global information, and
- dependency on first recursive steps.

2-Laplacian

$$\min_{\boldsymbol{U}\in\mathbb{R}^{n\times k}} F_2(\boldsymbol{U}) = \operatorname{Tr}\left(\boldsymbol{U}^\top \Delta_2 \boldsymbol{U}\right),$$

$$\text{s.t.}\ \ \boldsymbol{U}^\top \boldsymbol{U} = \boldsymbol{I}.$$

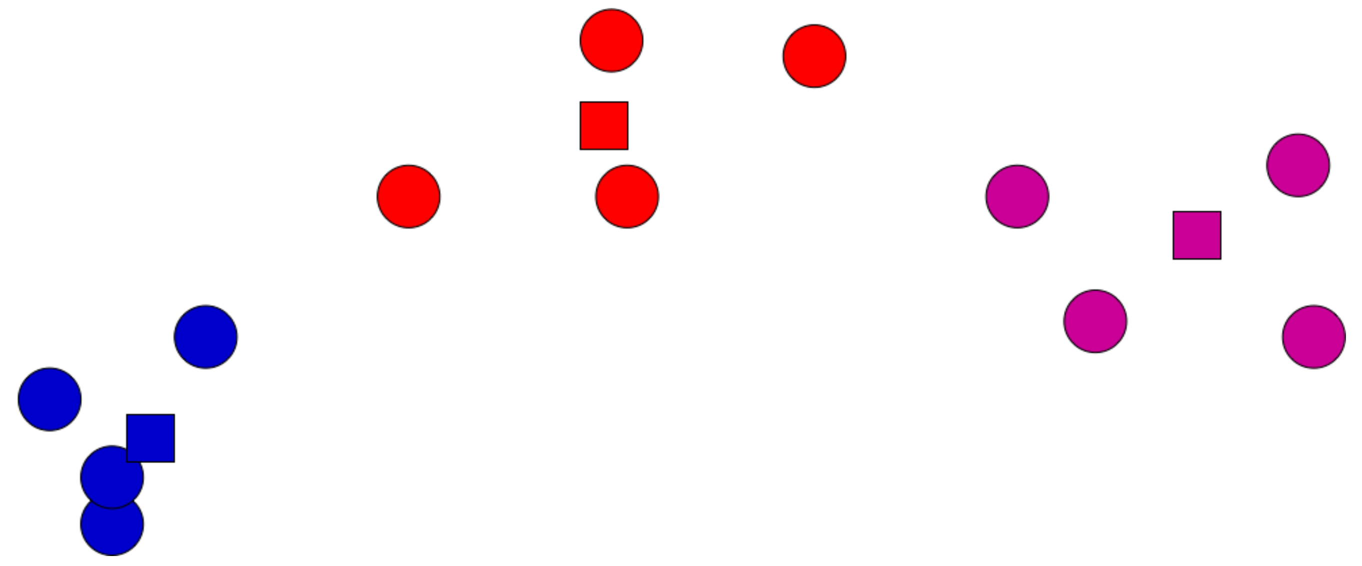
	$\boldsymbol{u}_1$	$\dots$	$\boldsymbol{u}_k$
$\boldsymbol{U}_1$	u_{11}	$\dots$	u_{1k}
$\vdots$	$\vdots$	$\dots$	$\vdots$
$\boldsymbol{U}_n$	u_{n1}	$\dots$	u_{nk}

$\Rightarrow$ Dimensionality reduction: $n \times n \rightarrow n \times k$

Spectral Direct Multiway Clustering

Avoid

- lack of global information, and
- dependency on first recursive steps.



2-Laplacian

$$\min_{\boldsymbol{U}\in\mathbb{R}^{n\times k}} F_2(\boldsymbol{U}) = \operatorname{Tr}\left(\boldsymbol{U}^\top \Delta_2 \boldsymbol{U}\right),$$

$$\text{s.t.} \quad \boldsymbol{U}^\top \boldsymbol{U} = \boldsymbol{I}.$$

	$\boldsymbol{u}_1$	$\dots$	$\boldsymbol{u}_k$
$\boldsymbol{U}_1$	u_{11}	$\dots$	u_{1k}
$\vdots$	$\vdots$	$\dots$	$\vdots$
$\boldsymbol{U}_n$	u_{n1}	$\dots$	u_{nk}

$\Rightarrow$ Dimensionality reduction: $n \times n \to n \times k$

p-Spectral Direct Multiway Clustering

Combine benefits from

- sparse solution vectors,
- proven optimal cuts (unweighted path graphs),
- global information.

p-Laplacian, $p \in (1, 2]$

$$\min_{\boldsymbol{U} \in \mathbb{R}^{n \times k}} F_p(\boldsymbol{U}) = \sum_{l=1}^{k} \sum_{ij}^{n} \frac{w_{ij} |u_i^l - u_j^l|^p}{2 \|\boldsymbol{u}^l\|_p^p}$$

$$\text{s.t.} \quad \sum_{i=1}^{n} \phi_p(u_i^l) \phi_p(u_i^m) = 0 \;\; \forall\, l \neq m, \;\; p \in (1, 2], \;\; l \in [1, k], \;\; m \in [1, k].$$

	$\boldsymbol{u}_1^p$	$\dots$	$\boldsymbol{u}_k^p$
$\boldsymbol{U}_1$	u_{11}^p	$\dots$	u_{1k}^p
$\vdots$	$\vdots$	$\dots$	$\vdots$
$\boldsymbol{U}_n$	u_{n1}^p	$\dots$	u_{nk}^p

p-Spectral Direct Multiway Clustering

Combine benefits from

- sparse solution vectors,
- proven optimal cuts (unweighted path graphs),
- global information.

Bresson et al., 2013

Rangapuram et al., 2014

Tudisco & Hein, 2017

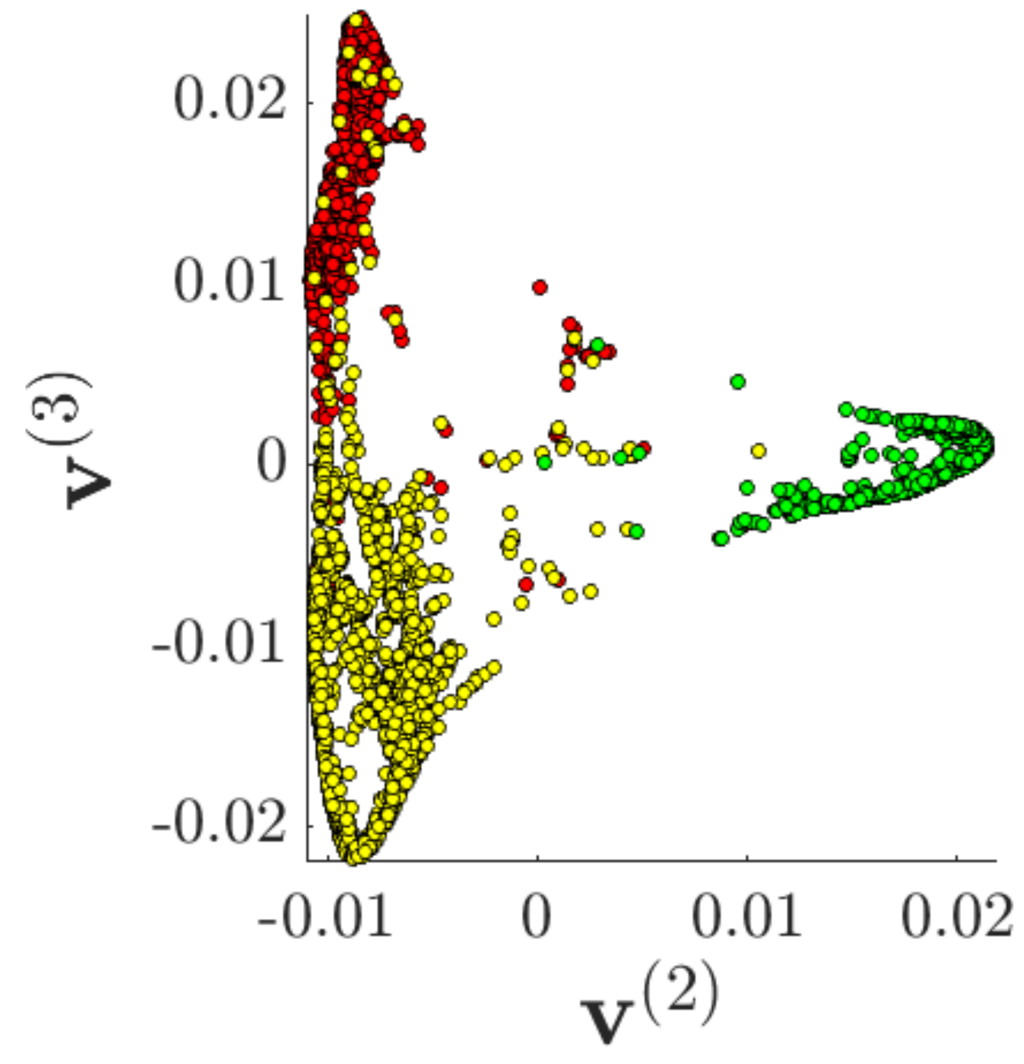
p-Laplacian, $p \in (1, 2]$

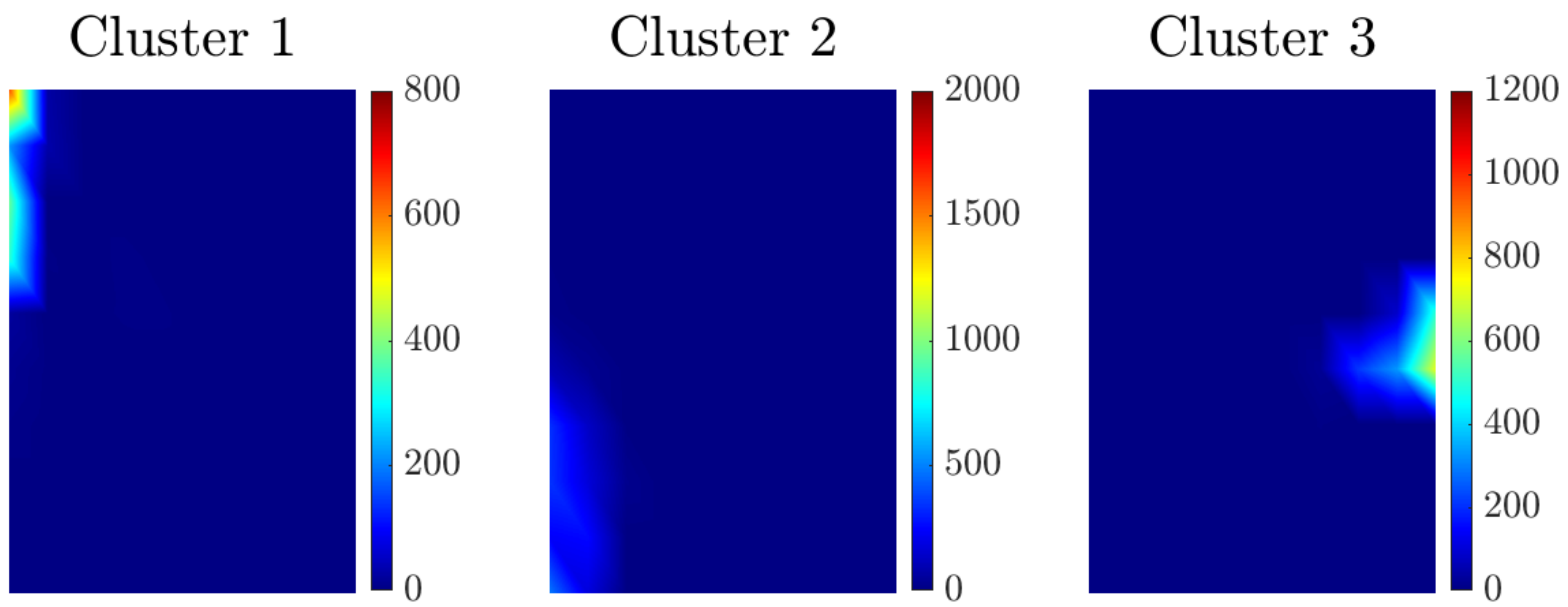
$$\min_{\boldsymbol{U} \in \mathbb{R}^{n \times k}} F_p(\boldsymbol{U}) = \sum_{l=1}^{k} \sum_{ij}^{n} \frac{w_{ij} |u_i^l - u_j^l|^p}{2 \|\boldsymbol{u}^l\|_p^p}$$

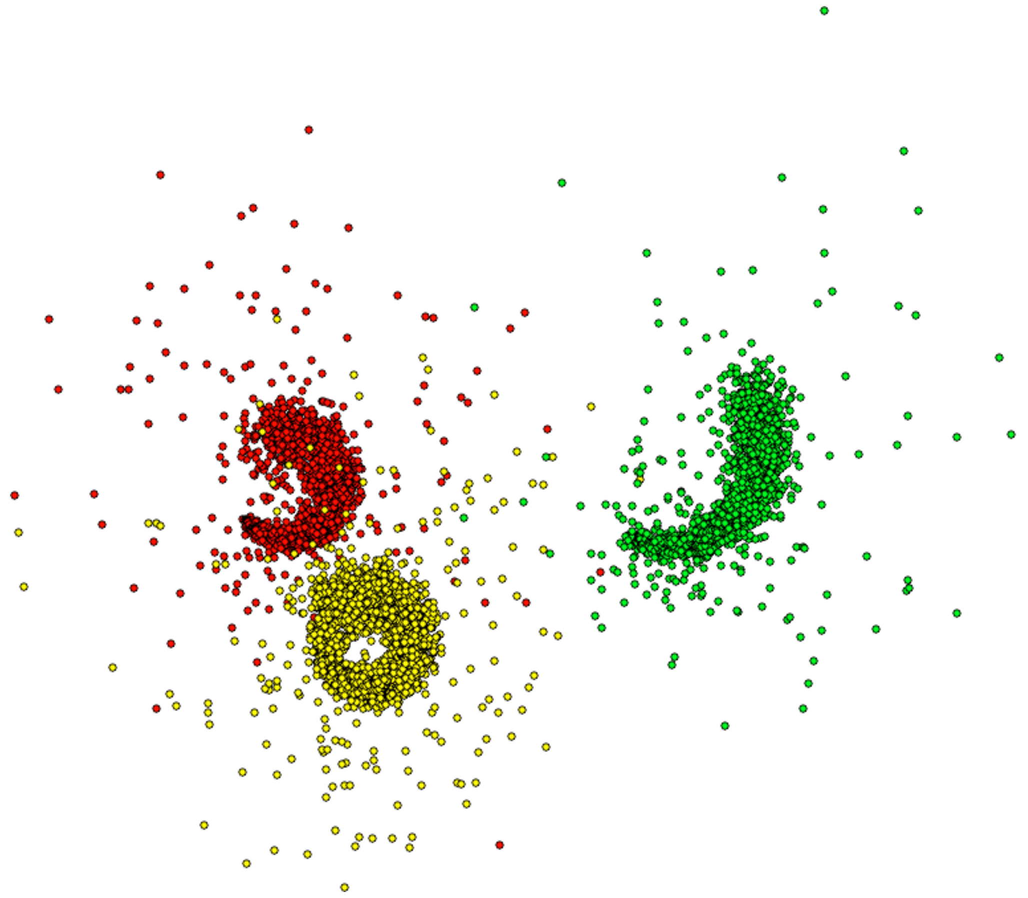
$$\text{s.t.} \quad \sum_{i=1}^{n} \phi_p(u_i^l) \phi_p(u_i^m) = 0 \ \ \forall\, l \neq m,\ p \in (1, 2],\ l \in [1, k],\ m \in [1, k].$$

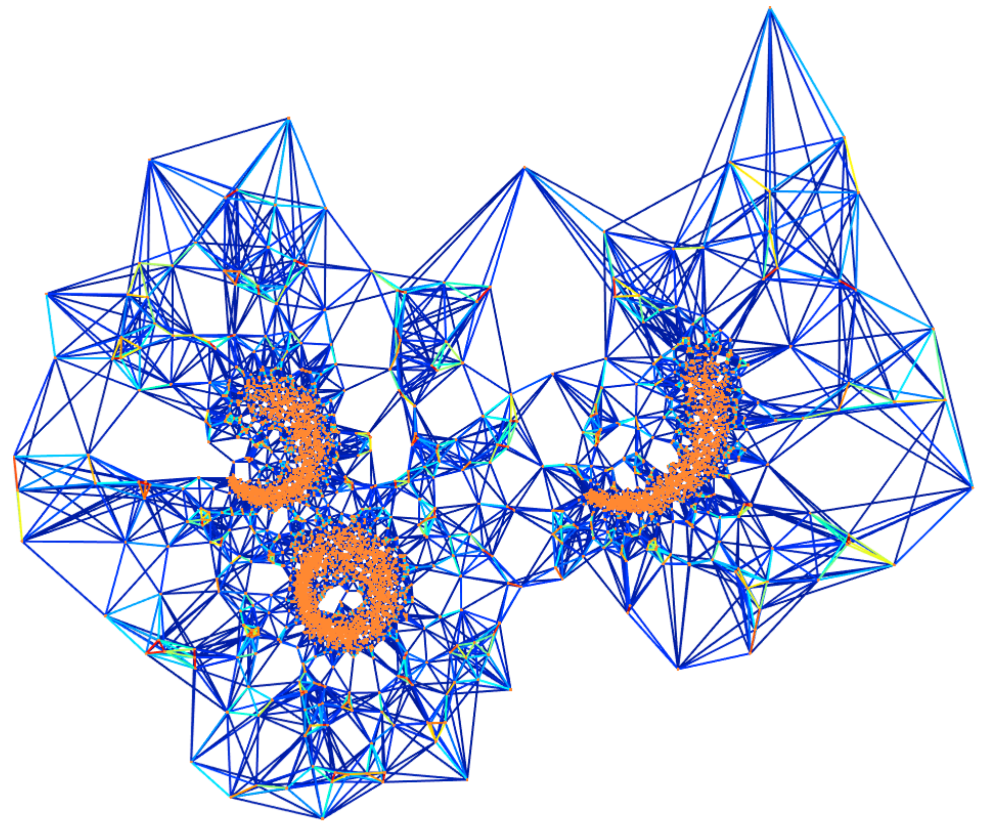
	$\boldsymbol{u}_1^p$	$\dots$	$\boldsymbol{u}_k^p$
$\boldsymbol{U}_1$	u_{11}^p	$\dots$	u_{1k}^p
$\vdots$	$\vdots$	$\dots$	$\vdots$
$\boldsymbol{U}_n$	u_{n1}^p	$\dots$	u_{nk}^p

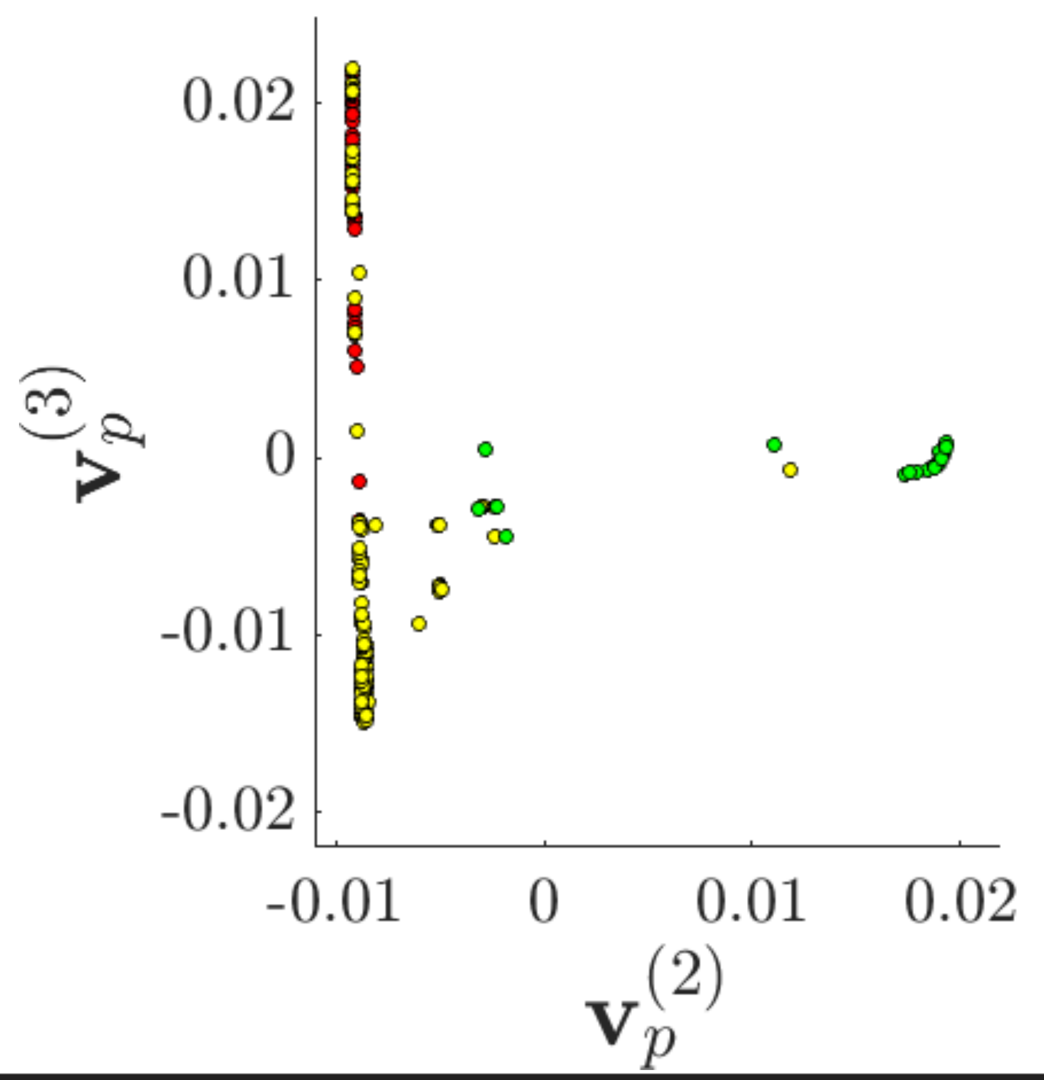
p-spectral embeddings

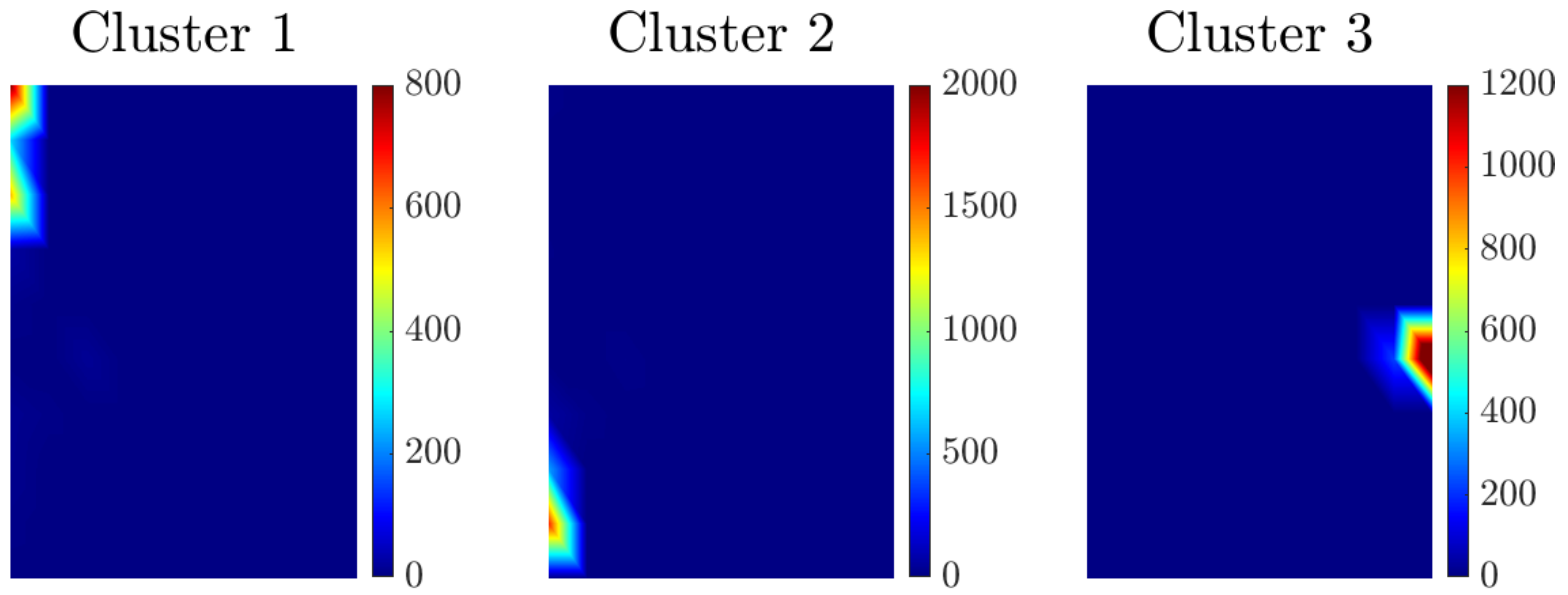












worms dataset

$p = 1.2$

nodes = 3200

edges = 19319

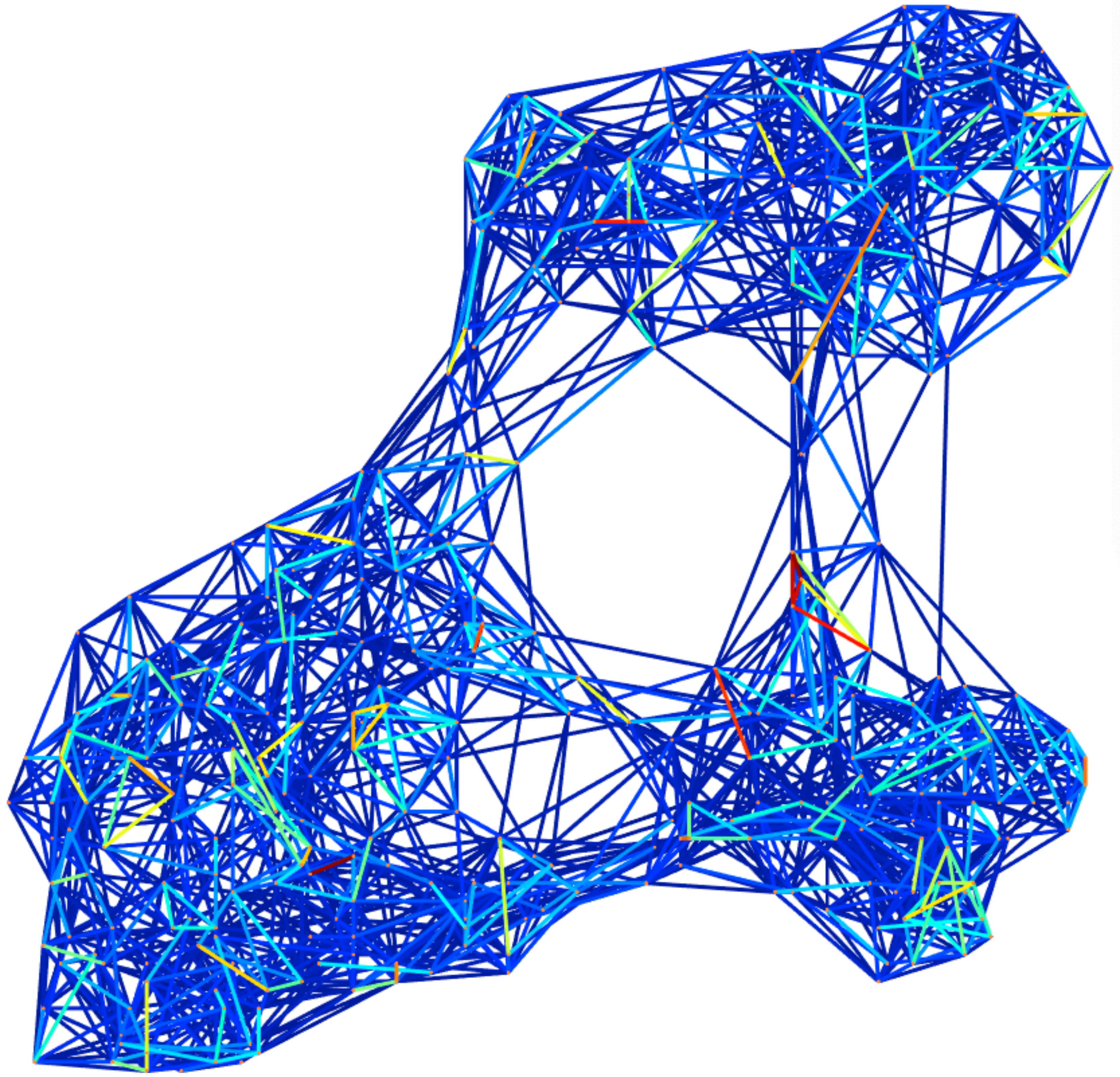
Utilizing the Manifold

- p-orthogonality $\Rightarrow$ intractable optimization problem $\Rightarrow$ consider $\boldsymbol{U}^\top \boldsymbol{U} = \boldsymbol{I}$. **Luo et al., 2010**

Preserving mutual orthogonality

$$\mathcal{S}t(k,n) = \{\boldsymbol{U} \in \mathbb{R}^{n\times k} \mid \boldsymbol{U}^\top \boldsymbol{U} = \boldsymbol{I}\},$$
$$\mathcal{G}r(k,n) \simeq \mathcal{S}t(k,n)/\mathcal{O}(k)$$
$$= \{\mathrm{span}(\boldsymbol{U}) : \boldsymbol{U} \in \mathbb{R}^{n\times k}, \boldsymbol{U}^\top \boldsymbol{U} = \boldsymbol{I}\}.$$

* $\mathcal{S}t(k,n) \Rightarrow$ unique choice of $\boldsymbol{U} \Rightarrow$ identifiability issue, local minima problem.



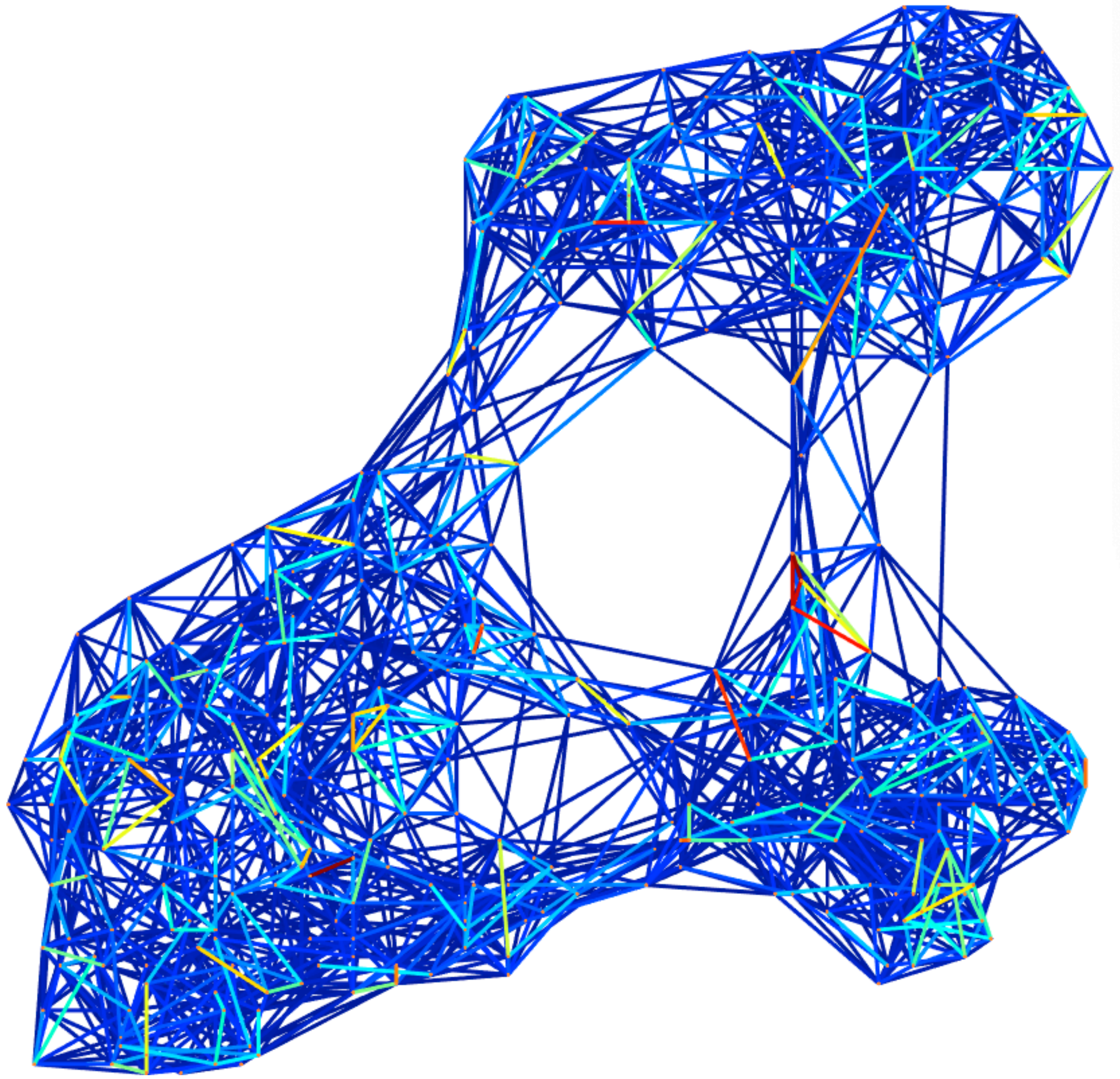
Utilizing the Manifold

- p-orthogonality $\Rightarrow$ intractable optimization problem

Preserving mutual orthogonality

$$\mathcal{S}t(k,n) = \{\boldsymbol{U} \in \mathbb{R}^{n\times k} \mid \boldsymbol{U}^\top \boldsymbol{U} = \boldsymbol{I}\},$$
$$\mathcal{G}r(k,n) \simeq \mathcal{S}t(k,n)/\mathcal{O}(k)$$
$$= \{\mathrm{span}(\boldsymbol{U}) : \boldsymbol{U} \in \mathbb{R}^{n\times k}, \boldsymbol{U}^\top \boldsymbol{U} = \boldsymbol{I}\}.$$

* $\mathcal{S}t(k,n) \Rightarrow$ unique choice of $\boldsymbol{U} \Rightarrow$ identifiability issue, local minima problem.



Ecoli graph

nodes = 336
edges = 2280

Utilizing the Manifold

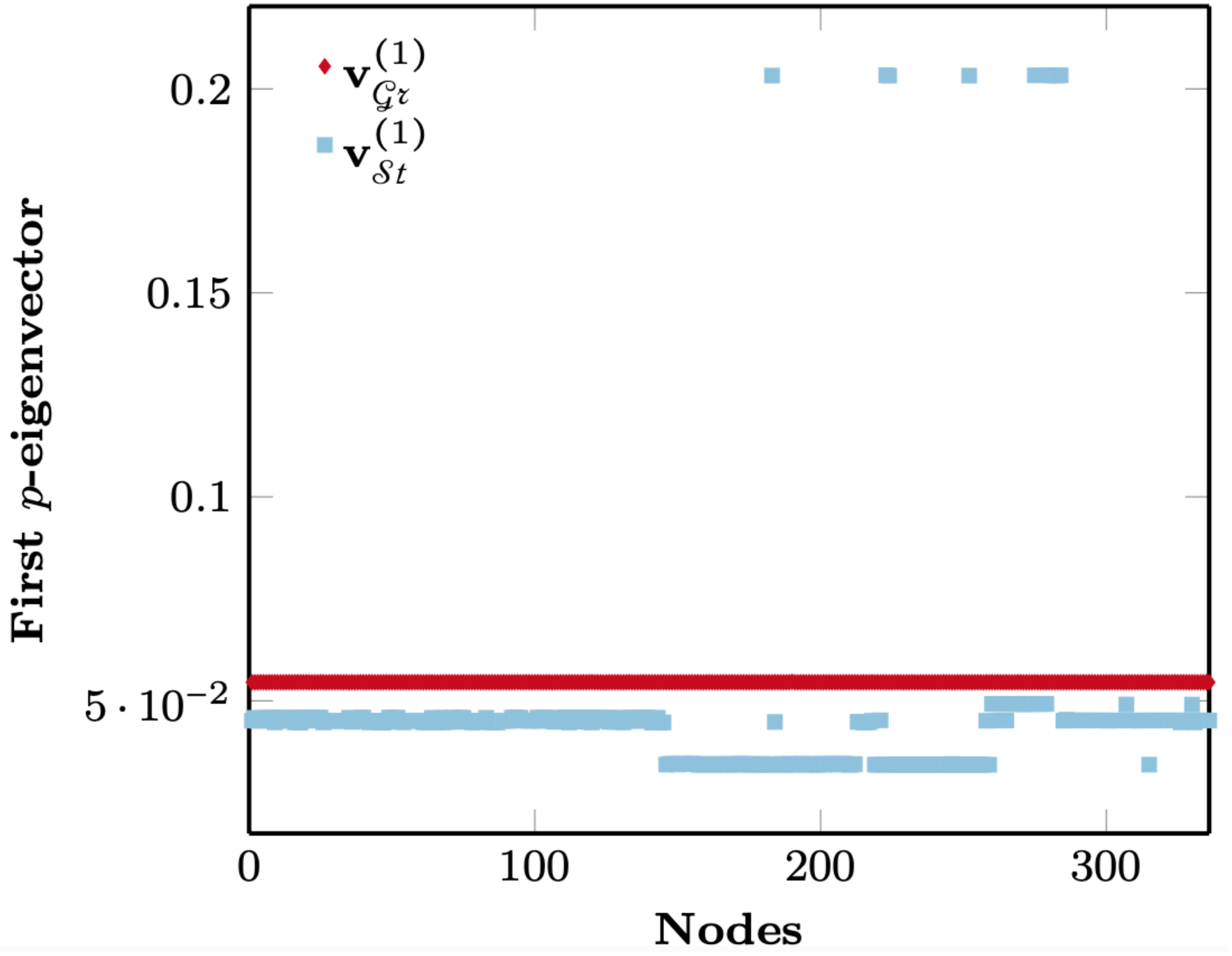
- p-orthogonality $\Rightarrow$ intractable optimization problem $\Rightarrow$ consider $\boldsymbol{U}^\top \boldsymbol{U} = \boldsymbol{I}$.

Preserving mutual orthogonality

$$\mathcal{S}t(k,n) = \{\boldsymbol{U} \in \mathbb{R}^{n\times k} \mid \boldsymbol{U}^\top \boldsymbol{U} = \boldsymbol{I}\},$$
$$\mathcal{G}r(k,n) \simeq \mathcal{S}t(k,n)/\mathcal{O}(k)$$
$$= \{\mathrm{span}(\boldsymbol{U}) : \boldsymbol{U} \in \mathbb{R}^{n\times k}, \boldsymbol{U}^\top \boldsymbol{U} = \boldsymbol{I}\}.$$

∗ $\mathcal{G}r(k,n) \Rightarrow$ non unique choice of $\boldsymbol{U}$,

$$\boldsymbol{U}^{\mathcal{G}r} = \{\boldsymbol{UQ} \mid \forall\, \boldsymbol{Q} \in \mathcal{O}(k)\}, \quad \boldsymbol{U} \in \mathbb{R}^{n\times k}, \; n \gg k.$$



Ecoli 1st eigenvectors

Utilizing the Manifold

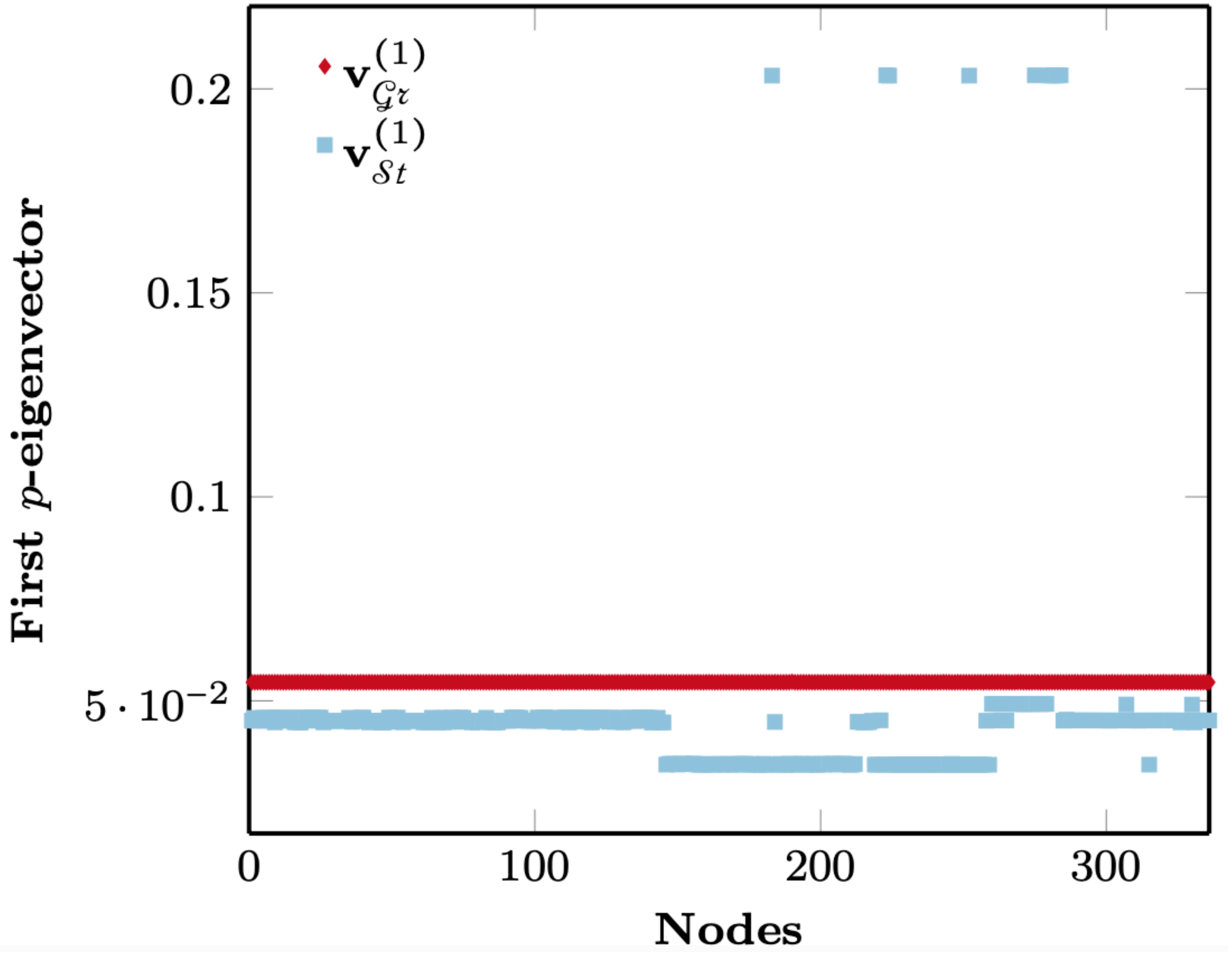
- p-orthogonality $\Rightarrow$ intractable optimization problem $\Rightarrow$ consider $\boldsymbol{U}^\top \boldsymbol{U} = \boldsymbol{I}$.

Preserving mutual orthogonality

$$\mathcal{S}t(k,n) = \{\boldsymbol{U} \in \mathbb{R}^{n\times k} \mid \boldsymbol{U}^\top \boldsymbol{U} = \boldsymbol{I}\},$$
$$\mathcal{G}r(k,n) \simeq \mathcal{S}t(k,n)/\mathcal{O}(k)$$
$$= \{\mathrm{span}(\boldsymbol{U}) : \boldsymbol{U} \in \mathbb{R}^{n\times k}, \boldsymbol{U}^\top \boldsymbol{U} = \boldsymbol{I}\}.$$

$*$ $\mathcal{G}r(k,n) \Rightarrow$ non unique choice of $\boldsymbol{U}$,

$$\boldsymbol{U}^{\mathcal{G}r} = \{\boldsymbol{U}\boldsymbol{Q} \mid \forall\, \boldsymbol{Q} \in \mathcal{O}(k)\}, \quad \boldsymbol{U} \in \mathbb{R}^{n\times k}, \; n \gg k.$$



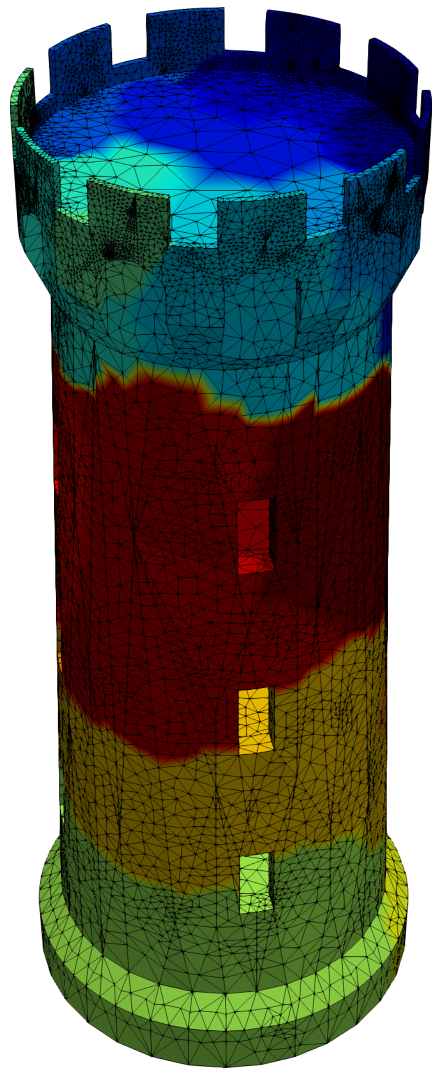
Ecoli 1st eigenvectors

pGrassmann Spectral Clustering

Unconstrained optimization problem

$$\min_{\boldsymbol{U} \in \mathcal{G}r(k,n)} F_p(\boldsymbol{U}) = \sum_{l}^{k} \sum_{ij}^{N} \frac{w_{ij} |u_i^l - u_j^l|^p}{2\|\boldsymbol{u}^l\|_p^p}, \quad p \in (1, 2].$$

- cluster indices $l, m = 1, 2, \ldots, k$,
- k predetermined for this work.



ALGORITHM: main pGrass loop

```
Initialize: c, r_{new,old,best} = Cut(c)          ▷ p = 2
while p ≥ p_w && r_new ≤ 1.05 · r_old do
    Reduce p
    Find U: minimize_{U∈Gr(k,n)} F_p(U) using W
    c = discretize(U)
    r_old = r_new
    r_new = Cut(c)
    if r_new < r_best then
        r_best = r_new
        c_best = c
    end if
end while
```

Key Algorithmic Components

Minimization on the manifold

- Software package ROPTLIB.
- Newton's method, truncated CG for the linear steps.
- Inputs: Euclidean gradient ($\mathbf{g}^k$) and Hessian.
- Converges if: $\|\mathbf{g}_m^k\|/\|\mathbf{g}_0^k\| < 10^{-6}$.

https://github.com/whuang08/ROPTLIB

Huang et al., 2018

ALGORITHM: main pGrass loop

Initialize: $\boldsymbol{c}, \mathrm{r}_{\mathrm{new,old,best}} = \mathrm{Cut}(\boldsymbol{c})$ ▷ $p = 2$

```
1  while p ≥ p_w && r_new ≤ 1.05 · r_old do
2      Reduce p
3      Find U: minimize_{U∈Gr(k,n)} F_p(U) using W
4      c = discretize(U)
5      r_old = r_new
6      r_new = Cut(c)
7      if r_new < r_best then
8          r_best = r_new
9          c_best = c
10     end if
11 end while
```

Key Algorithmic Components

Monitor monotonic descent

- Discrete objective (RCut, NCut).
- Experiments on synthetic datasets.

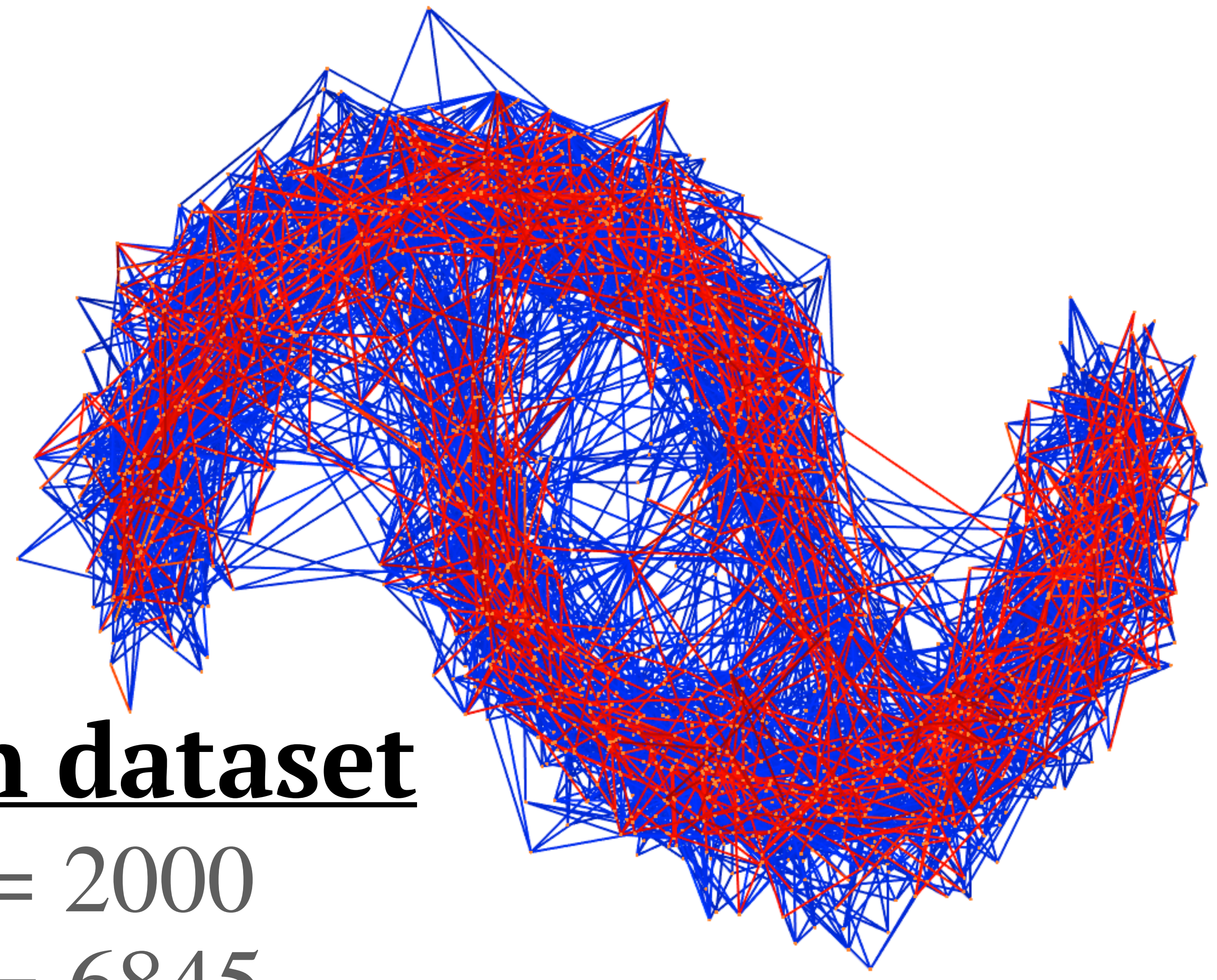
$$\mathrm{RCut}(C_1, \ldots, C_k) = \sum_{i=1}^{k} \frac{\mathrm{cut}(C_i, \overline{C_i})}{|C_i|}$$

$$\mathrm{NCut}(C_1, \ldots, C_k) = \sum_{i=1}^{k} \frac{\mathrm{cut}(C_i, \overline{C_i})}{\mathrm{vol}(C_i)|}$$

ALGORITHM: main pGrass loop

Initialize: $\boldsymbol{c}, \mathrm{r}_{\mathrm{new,old,best}} = \mathrm{Cut}(\boldsymbol{c})$ $\quad \triangleright\ p = 2$
while $p \geq p_w$ && $\mathrm{r}_{\mathrm{new}} \leq 1.05 \cdot \mathrm{r}_{\mathrm{old}}$ **do**
  Reduce p
  Find $\boldsymbol{U}$: $\underset{\boldsymbol{U} \in \mathcal{G}r(k,n)}{\mathrm{minimize}}\ F_p(\boldsymbol{U})$ using $\boldsymbol{W}$
  $\boldsymbol{c} = \mathrm{discretize}(\boldsymbol{U})$
  $\mathrm{r}_{\mathrm{old}} = \mathrm{r}_{\mathrm{new}}$
  $\mathrm{r}_{\mathrm{new}} = \mathrm{Cut}(\boldsymbol{c})$
  if $\mathrm{r}_{\mathrm{new}} < \mathrm{r}_{\mathrm{best}}$ **then**
    $\mathrm{r}_{\mathrm{best}} = \mathrm{r}_{\mathrm{new}}$
    $\boldsymbol{c}_{\mathrm{best}} = \boldsymbol{c}$
  end if
end while

Key Algorithmic Components

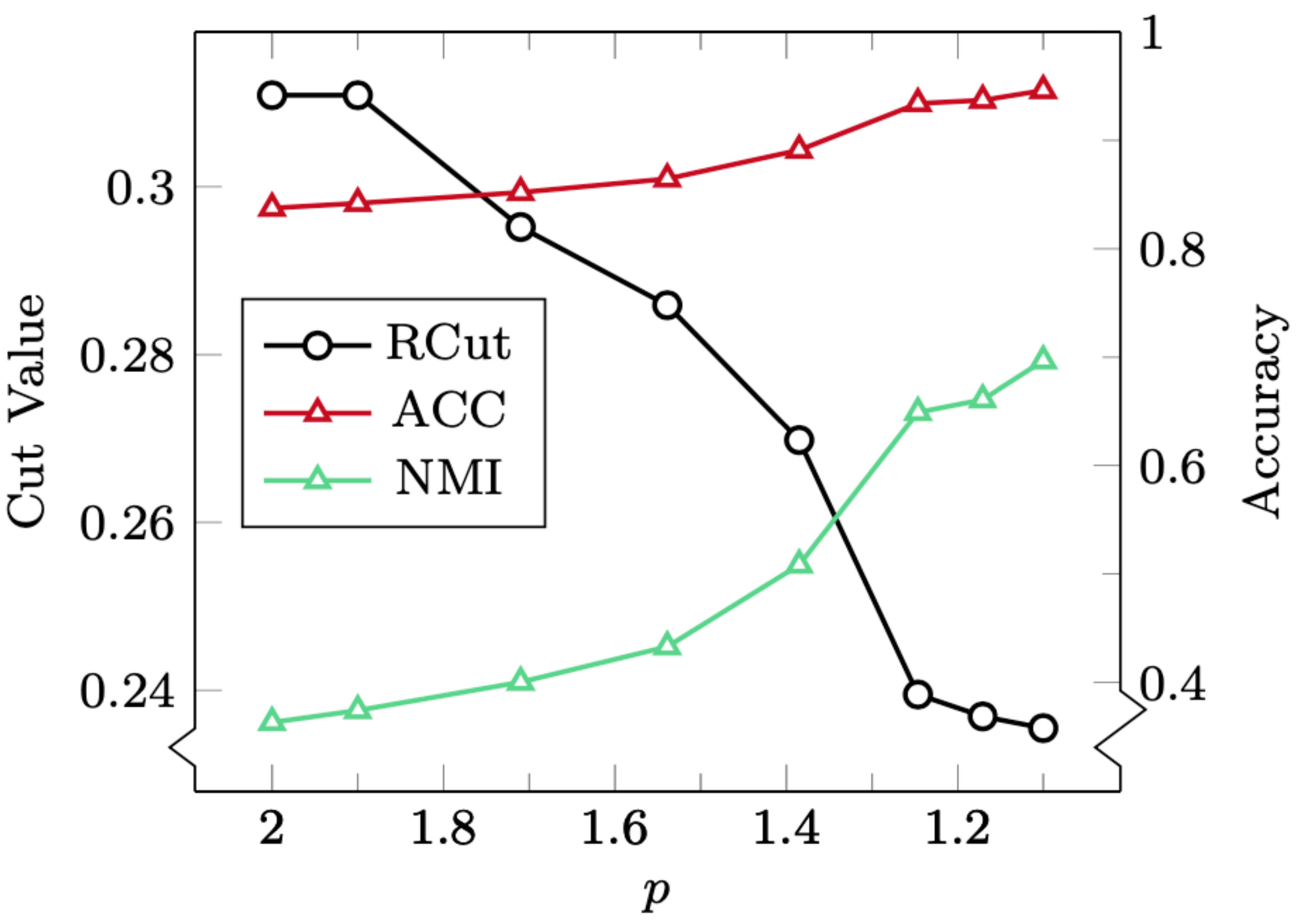


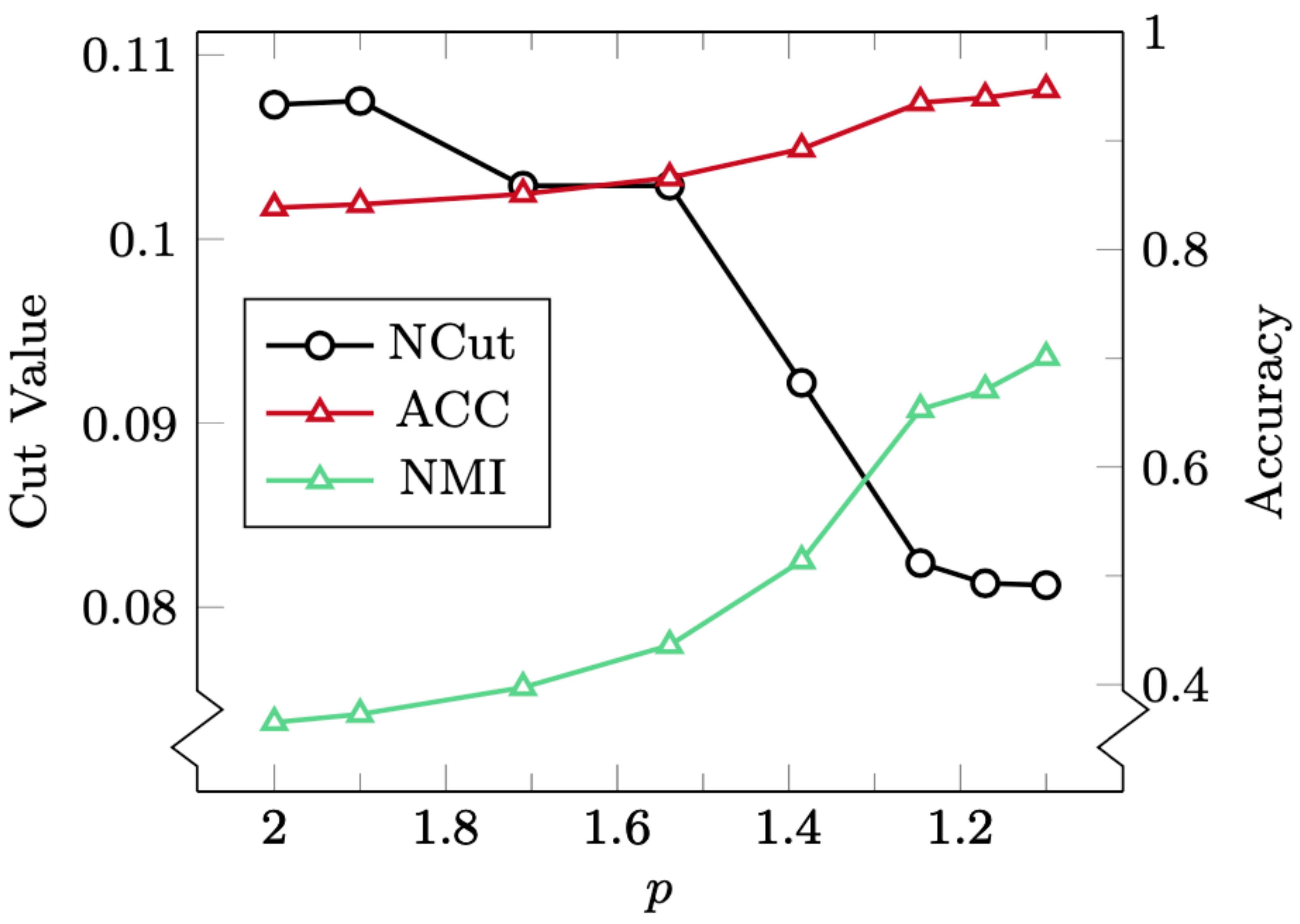
HighMoon dataset

nodes = 2000
edges = 6845

Monitor monotonic descent

- Discrete objective (RCut, NCut).
- Experiments on synthetic datasets.





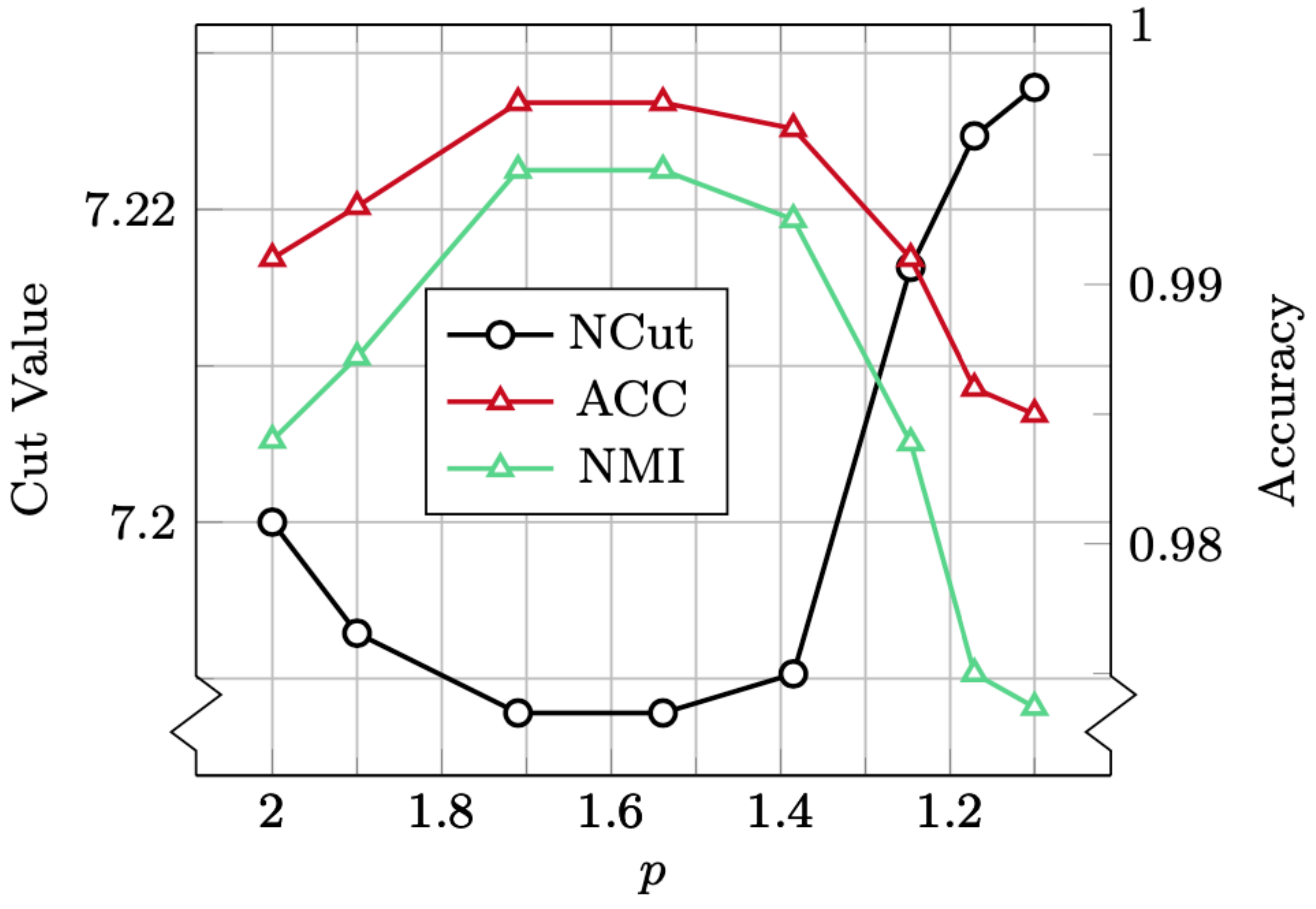
Key Algorithmic Components

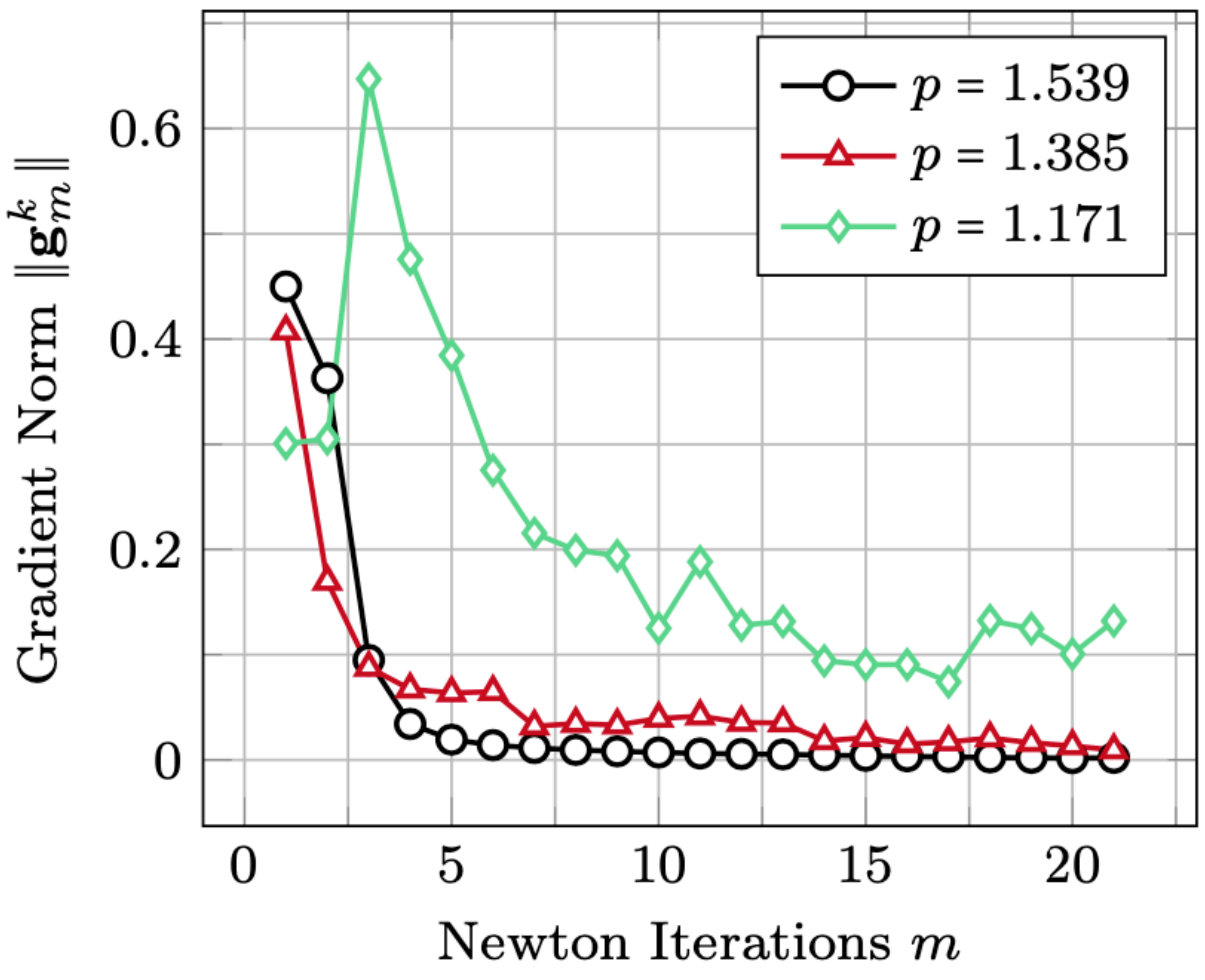
LFR dataset

nodes = 1000
edges = 2280
$k = 19, \ \mu = 0.38$

Monitor monotonic descent

- Discrete objective (RCut, NCut).
- Experiments on synthetic datasets.

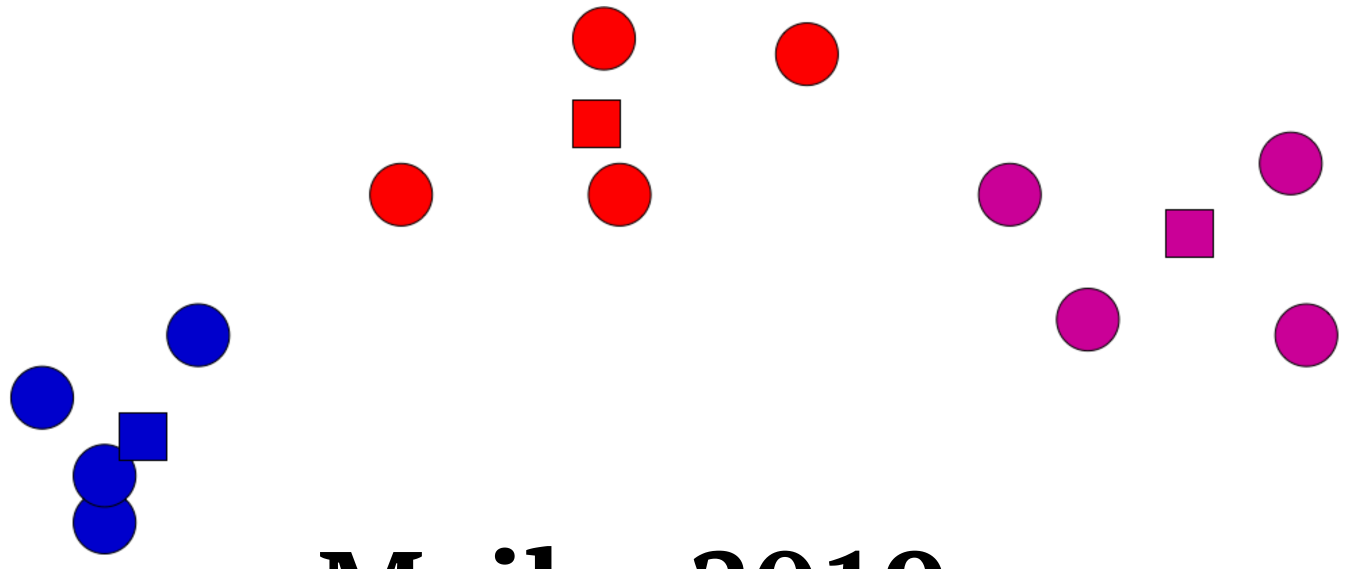




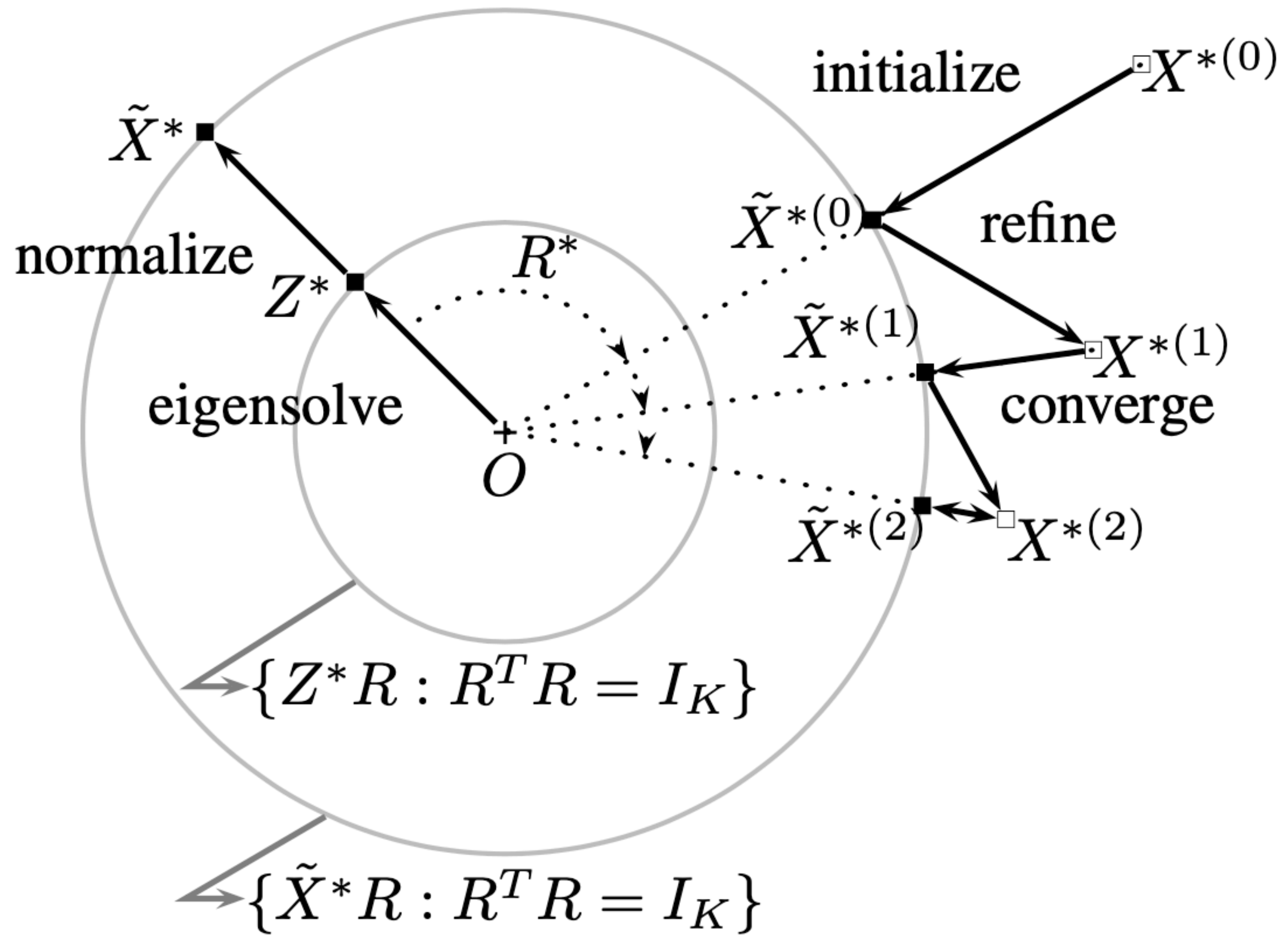
Key Algorithmic Components

Discretize the $p-$eigenvectors

1. k-means orthogonal $\Rightarrow$ pGrass-kmeans.
2. Rotate the normalized eigenvectors to obtain an optimal clustering $\Rightarrow$ pGrass-disc.



Meila, 2019



Yu & Shi, 2003

ALGORITHM: main pGrass loop

Initialize: $\boldsymbol{c}, \mathrm{r}_{\mathrm{new,old,best}} = \mathrm{Cut}(\boldsymbol{c})$ ▷ $p = 2$

```
while p ≥ p_w && r_new ≤ 1.05 · r_old do
    Reduce p
    Find U: minimize_{U∈Gr(k,n)} F_p(U) using W
    c = discretize(U)
    r_old = r_new
    r_new = Cut(c)
    if r_new < r_best then
        r_best = r_new
        c_best = c
    end if
end while
```

Numerical Experiments

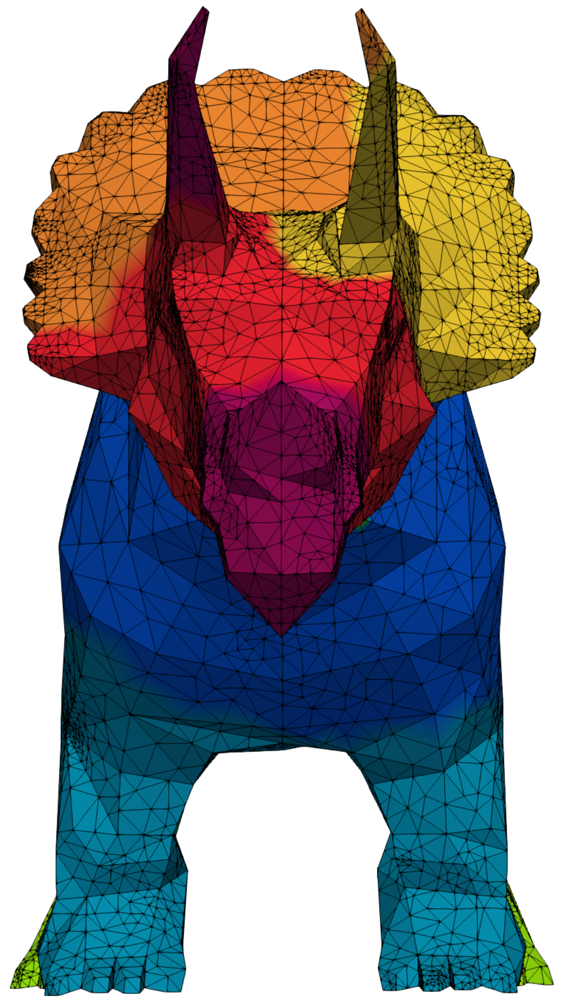
Methods under consideration

1. **Spec**: Traditional direct multiway spectral clustering.
   ▸ https://github.com/panji530/Ncut9 **Yu & Shi, 2003; Luxburg, 2007**
2. **pSpec**: Recursive bi-partitioning with the *p*-Laplacian.
   ▸ https://www.ml.uni-saarland.de/code/pSpectralClustering **Bühler & Hein, 2009**
3. **kCuts**: A tight continuous relaxation for the balanced direct *k*-cut problem.
   ▸ https://www.ml.uni-saarland.de/code/cfsp **Rangapuram et al., 2014**
4. **Graclus**: A multilevel algorithm using a weighted kernel k-means objective, thus eliminating the need for eigenvector computations.
   ▸ https://www.cs.utexas.edu/users/dml/Software/graclus.html **Dhillon et al., 2007**
5. **pMulti**: The first full eigenvector analysis of *p*-Laplacian leading to direct multiway clustering. **Luo et al., 2010**
   ▷ We implement this method in MATLAB R2020a.

Experimental Setup

Graph construction

- $\boldsymbol{G} \in \mathbb{R}^{n \times n} \rightarrow$ k-NN routine.
  $\boldsymbol{S} \in \mathbb{R}^{n \times n} \rightarrow$ Gaussian similarity kernel.
  $\boldsymbol{W} = \boldsymbol{G} \odot \boldsymbol{S}$.



Labelling accuracy metrics

✓ Unsupervised clustering accuracy

$$\mathrm{ACC} = \frac{1}{n}\sum_{i}^{n} \delta(l_i, c_i) \in [0, 1],$$

l_i: true class label, c_i: inferred cluster label of u_i, $\delta(\cdot)$: Dirac delta function.

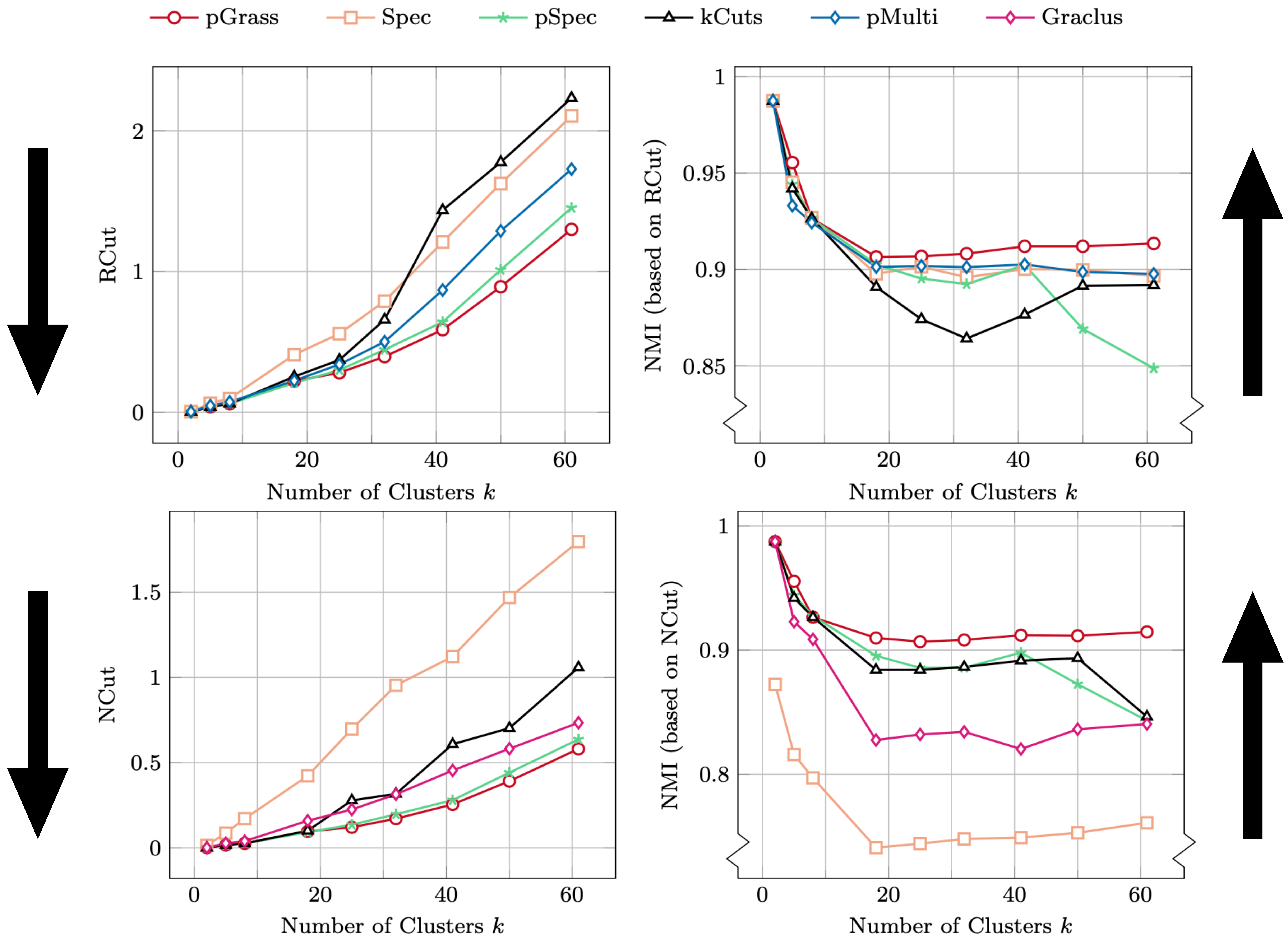
✓ Normalized mutual information

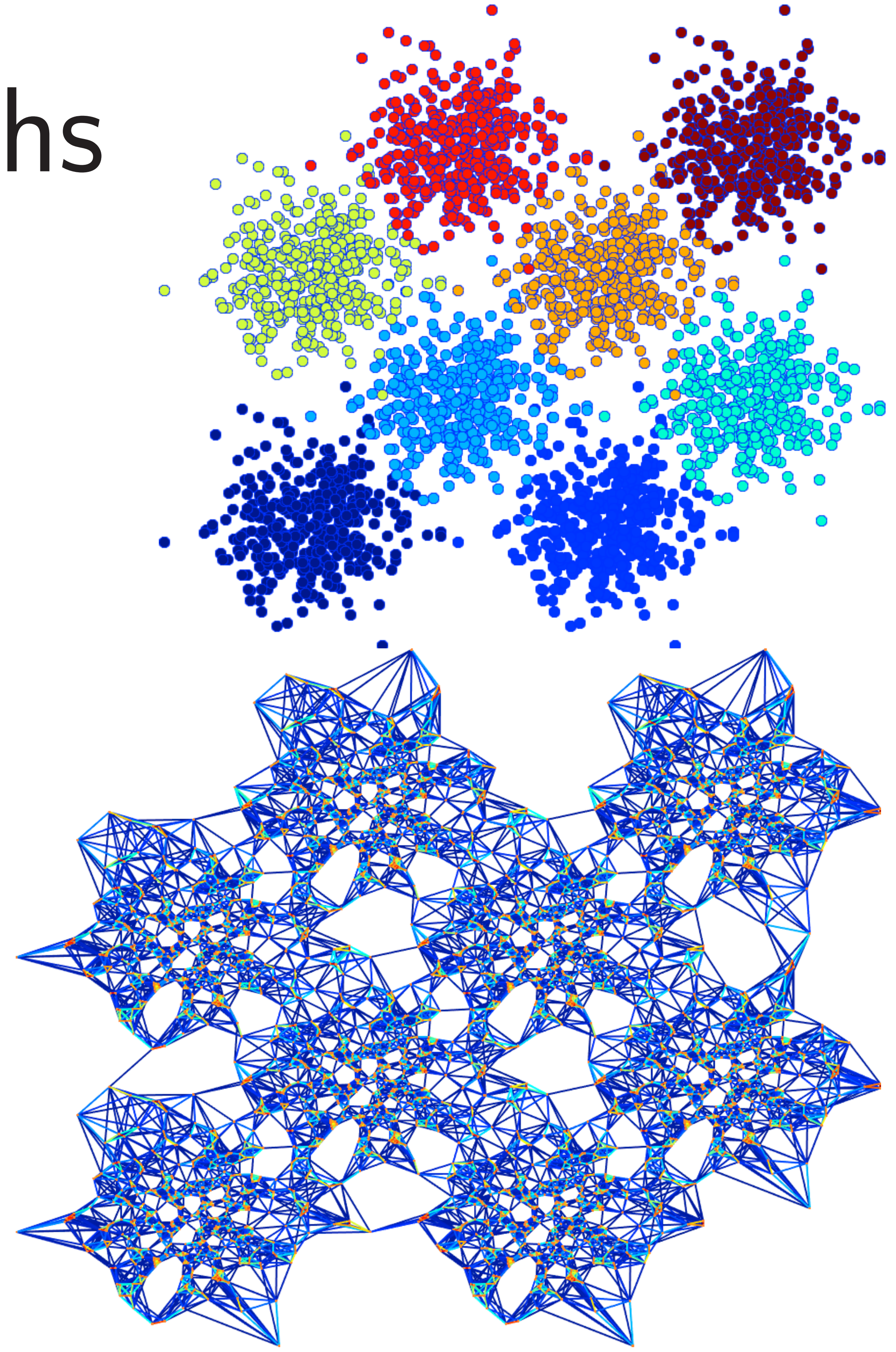
$$\mathrm{NMI} = \frac{I(l, c)}{\max\{H(l), H(c)\}} \in [0, 1],$$

$I(l, c)$: mutual information between l, c, $H(\cdot)$ their entropy.

Highlighted Results – Artificial Graphs

Increasing the number of clusters k





Gaussian datasets

$k \in [2,61]$

nodes $\in [800,25000]$

edges $\in [4900,145000]$

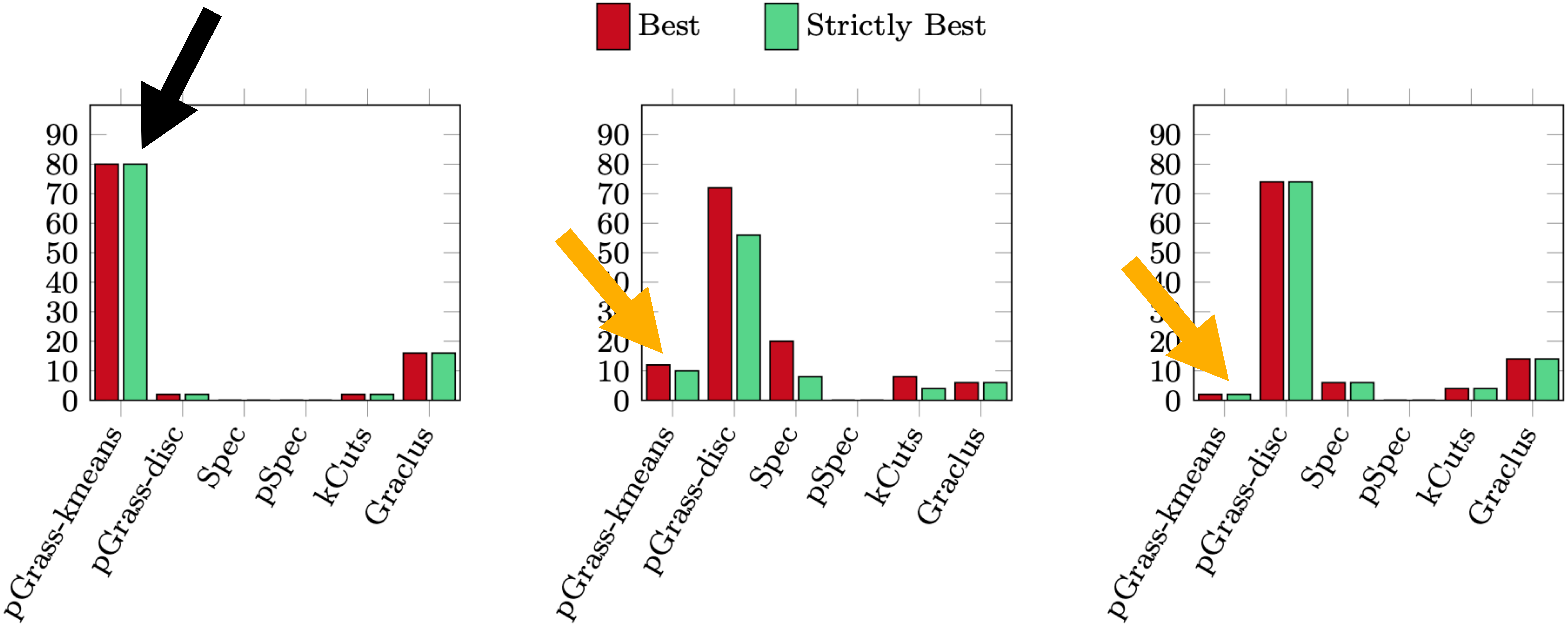
Highlighted Results – Real-World Graphs

Classification of Handwritten Characters

- Omniglot database: 1623 different handwritten characters from 50 alphabets.
  - https://github.com/brendenlake/omniglot

Highlighted Results – Real-World Graphs

Classification of Handwritten Characters

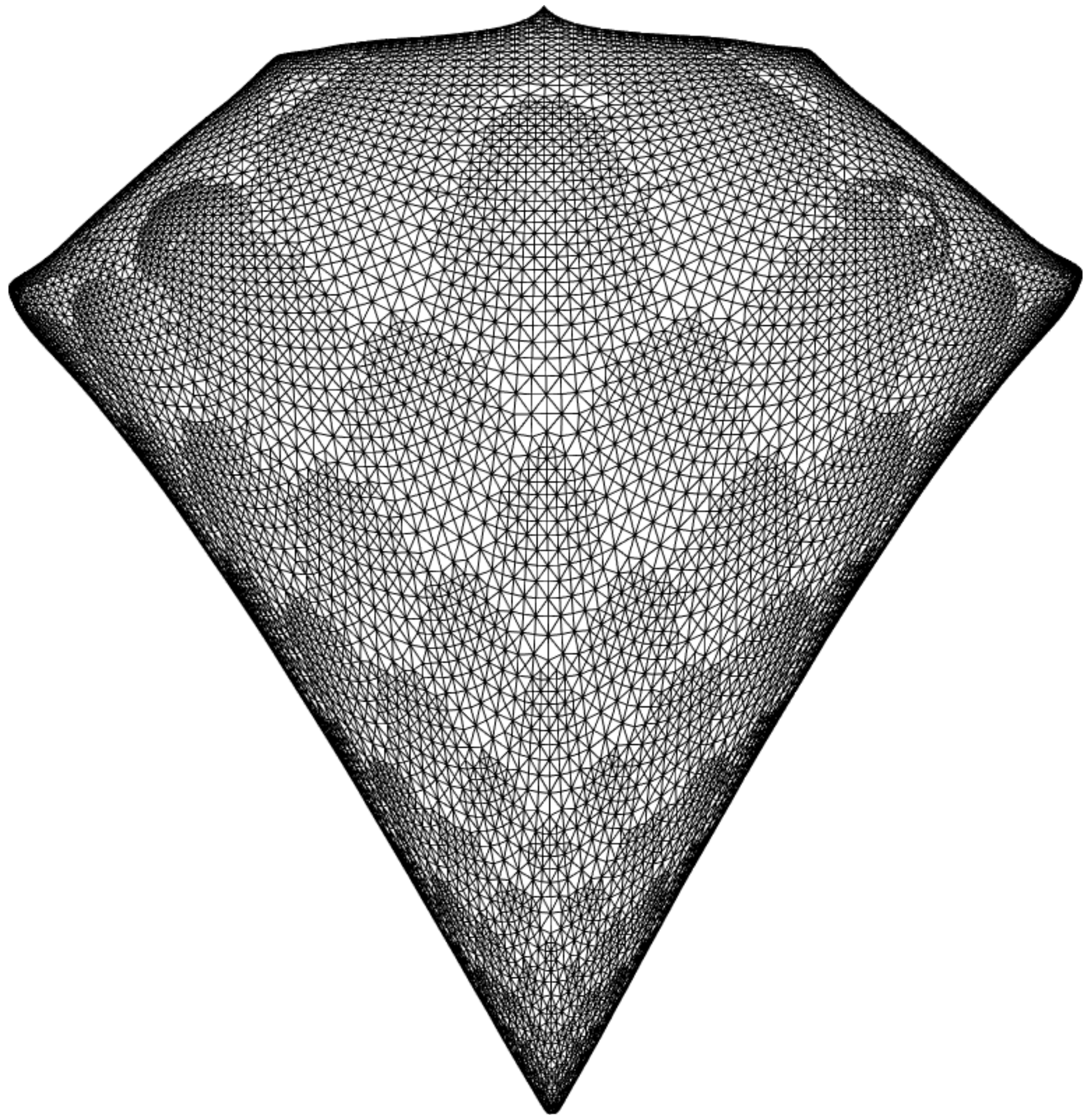


(a) NCut (b) NCut-based ACC (c) NCut-based NMI

Conclusions

- A direct multiway p-spectral graph clustering framework.
- Simple algorithm, utilizing packages of Riemannian optimization.
- pGrass embeddings lead to either superior graph cut values or labelling accuracy metrics.
- Consistent results over synthetic and real-world graphs.

FNSNF
FONDS NATIONAL SUISSE
SCHWEIZERISCHER NATIONALFONDS
FONDO NAZIONALE SVIZZERO
SWISS NATIONAL SCIENCE FOUNDATION



References

1. T. Bühler and M. Hein
Spectral clustering based on the graph p-Laplacian.
In Proceedings of the 26th Annual International Conference on Machine Learning, ICML '09, pages 81-88, New York, NY, USA, 2009, ACM.

2. W. Huang, P.-A. Absil, K.A. Gallivan, and P. Hand.
Roptlib: An object-oriented c++ library for optimization on riemannian manifolds
ACM Trans. Math. Softw., 44(4), July 2018.

3. D. Luo, H. Huang, C. Ding, and F. Nie.
On the eigenvectors of p-Laplacian
Machine Learning, 81(1):37-51, 2010.

4. D. Pasadakis, C. L. Alappat, O. Schenk, G. Wellein
K-way p-spectral clustering on Grassmann manifolds
Preprint: https://arxiv.org/abs/2008.13210

Additional Material

Highlighted Results – Real-World Graphs

Classification of Facial Images

1. Olivetti:400 images – 40 subjects. ‣ https://cam-orl.co.uk/facedatabase.html
2. Faces95:1440 images – 72 subjects. ‣ https://cmp.felk.cvut.cz/ spacelib/faces/
3. FACES: 2052 images – 171 subjects. ‣ https://faces.mpdl.mpg.de/imeji/

	Olivetti			Faces95			FACES		
Method	NCut	ACC	NMI	NCut	ACC	NMI	NCut	ACC	NMI
pGrass - kmeans	**3.984**	-4.15%	-2.28%	**2.658**	-5.77%	-4.24%	**29.42**	-3.58%	-2.41%
pGrass - disc	-4.50%	**0.716**	**0.831**	-4.50%	**0.609**	**0.758**	-6.08%	**0.802**	**0.91**
Spec	-24.84%	-9.19%	-5.27%	-24.84%	-4.23%	-0.90%	-15.05%	-2.50%	-1.23%
pSpec	-8.04%	-7.41%	-3.06%	-8.04%	-6.86%	-6.02%	-4.34%	-6.73%	-2.71%
kCuts	-1.41%	-6.78%	-3.20%	-1.41%	-10.37%	-7.70%	-7.67%	-13.0%	-6.99%
Graclus	-23.10%	-6.36%	-2.25%	-23.11%	-9.25%	-2.38%	-9.98%	-3.70%	-2.56%

Highlighted Results – Real-World Graphs

Classification of Facial Images

1. Olivetti:400 images – 40 subjects. ▸ https://cam-orl.co.uk/facedatabase.html
2. Faces95:1440 images – 72 subjects. ▸ https://cmp.felk.cvut.cz/ spacelib/faces/
3. FACES: 2052 images – 171 subjects. ▸ https://faces.mpdl.mpg.de/imeji/

	Olivetti			Faces95			FACES		
Method	NCut	ACC	NMI	NCut	ACC	NMI	NCut	ACC	NMI
pGrass - kmeans	**3.984**	-4.15%	-2.28%	**2.658**	-5.77%	-4.24%	**29.42**	-3.58%	-2.41%
pGrass - disc	-4.50%	**0.716**	**0.831**	-4.50%	**0.609**	**0.758**	-6.08%	**0.802**	**0.91**
Spec	-24.84%	-9.19%	-5.27%	-24.84%	-4.23%	-0.90%	-15.05%	-2.50%	-1.23%
pSpec	-8.04%	-7.41%	-3.06%	-8.04%	-6.86%	-6.02%	-4.34%	-6.73%	-2.71%
kCuts	-1.41%	-6.78%	-3.20%	-1.41%	-10.37%	-7.70%	-7.67%	-13.0%	-6.99%
Graclus	-23.10%	-6.36%	-2.25%	-23.11%	-9.25%	-2.38%	-9.98%	-3.70%	-2.56%

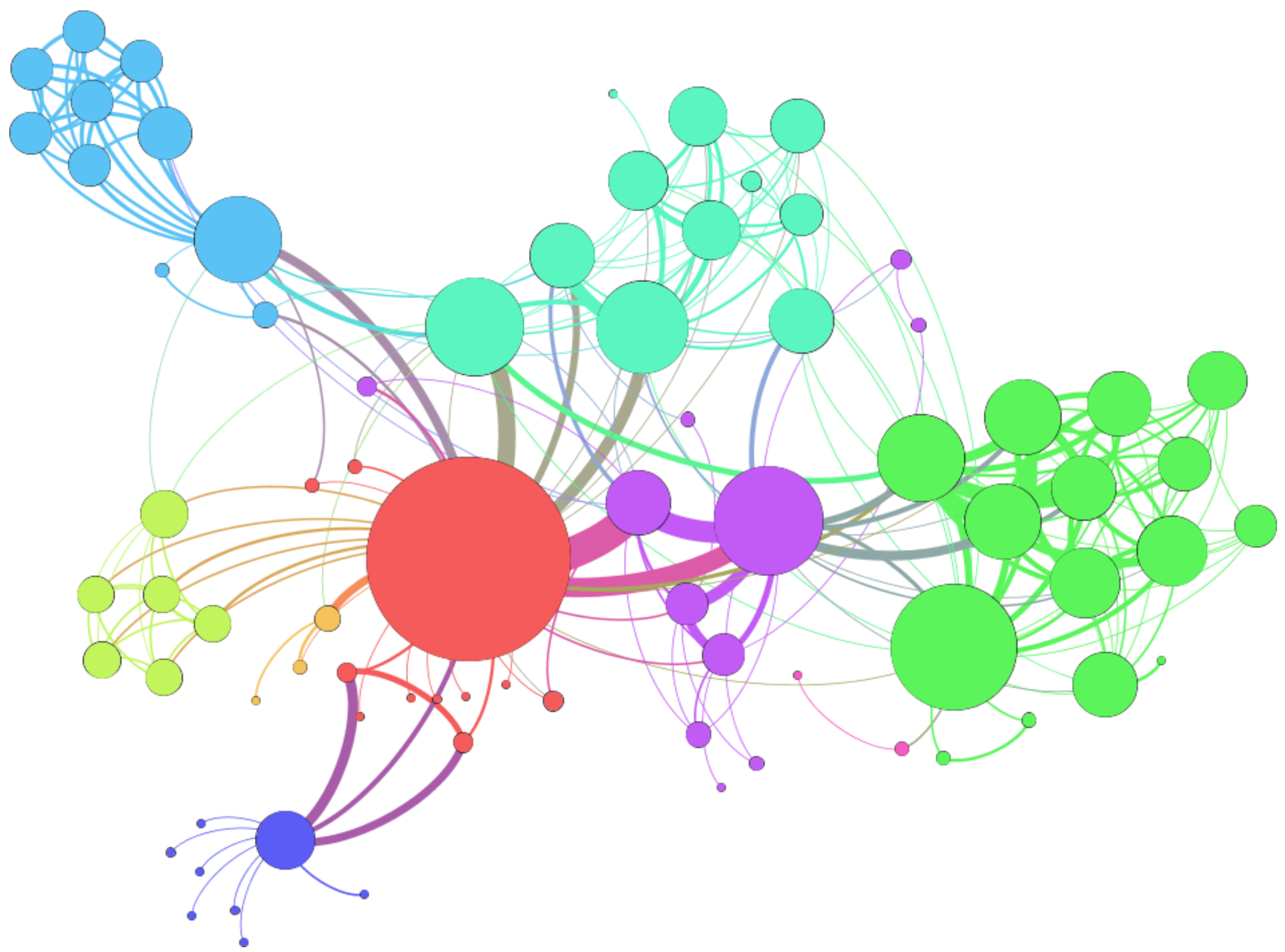
Highlighted Results – Real-World Graphs

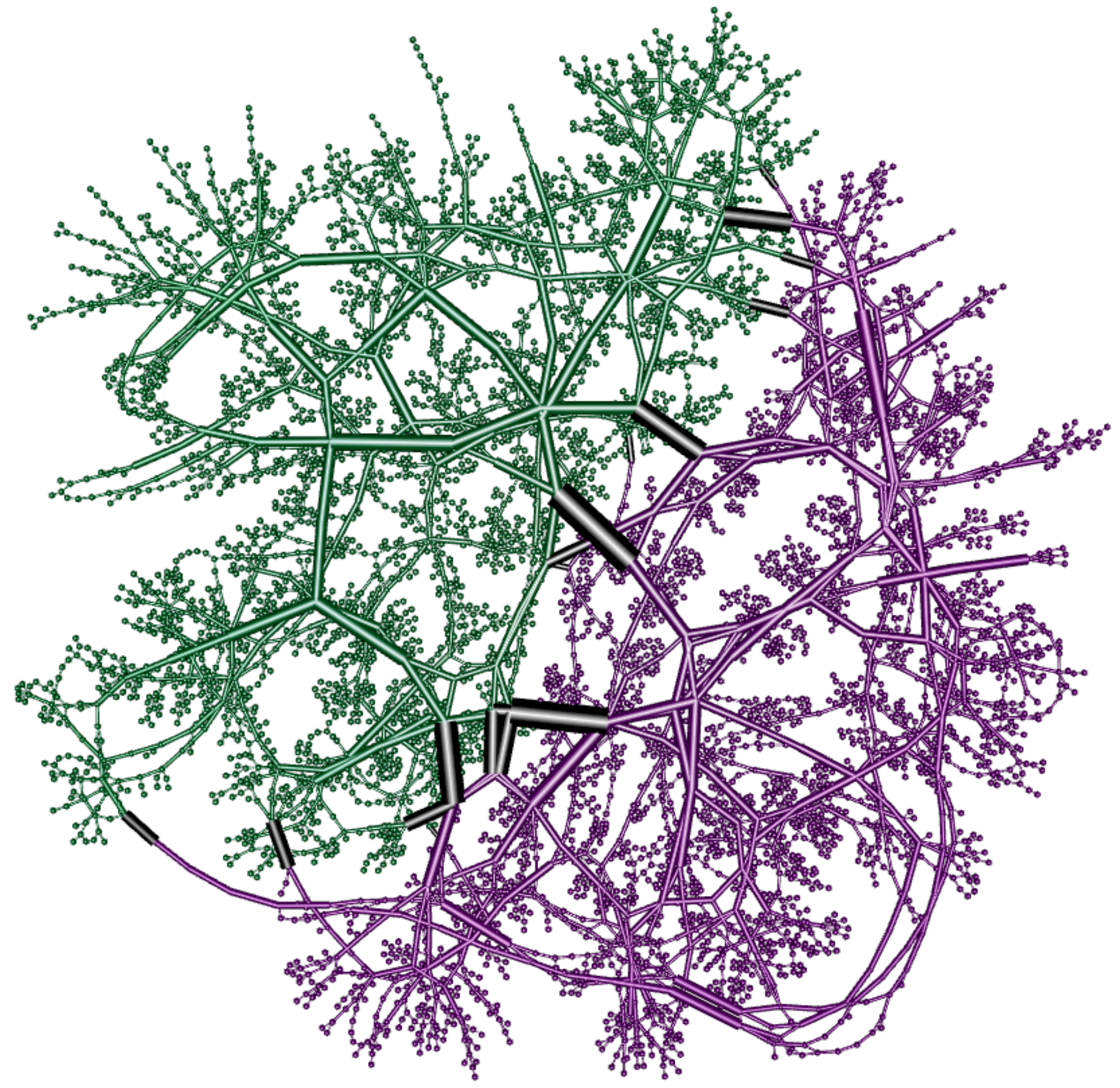
Classification of Facial Images

1. Olivetti:400 images – 40 subjects. ‣ https://cam-orl.co.uk/facedatabase.html
2. Faces95:1440 images – 72 subjects. ‣ https://cmp.felk.cvut.cz/ spacelib/faces/
3. FACES: 2052 images – 171 subjects. ‣ https://faces.mpdl.mpg.de/imeji/

	Olivetti			Faces95			FACES		
Method	NCut	ACC	NMI	NCut	ACC	NMI	NCut	ACC	NMI
pGrass - kmeans	**3.984**	-4.15%	-2.28%	**2.658**	-5.77%	-4.24%	**29.42**	-3.58%	-2.41%
pGrass - disc	-4.50%	**0.716**	**0.831**	-4.50%	**0.609**	**0.758**	-6.08%	**0.802**	**0.91**
Spec	-24.84%	-9.19%	-5.27%	-24.84%	-4.23%	-0.90%	-15.05%	-2.50%	-1.23%
pSpec	-8.04%	-7.41%	-3.06%	-8.04%	-6.86%	-6.02%	-4.34%	-6.73%	-2.71%
kCuts	-1.41%	-6.78%	-3.20%	-1.41%	-10.37%	-7.70%	-7.67%	-13.0%	-6.99%
Graclus	-23.10%	-6.36%	-2.25%	-23.11%	-9.25%	-2.38%	-9.98%	-3.70%	-2.56%

Spectral Clustering Applications





Spectral Bi-Partitioning

2-Laplacian

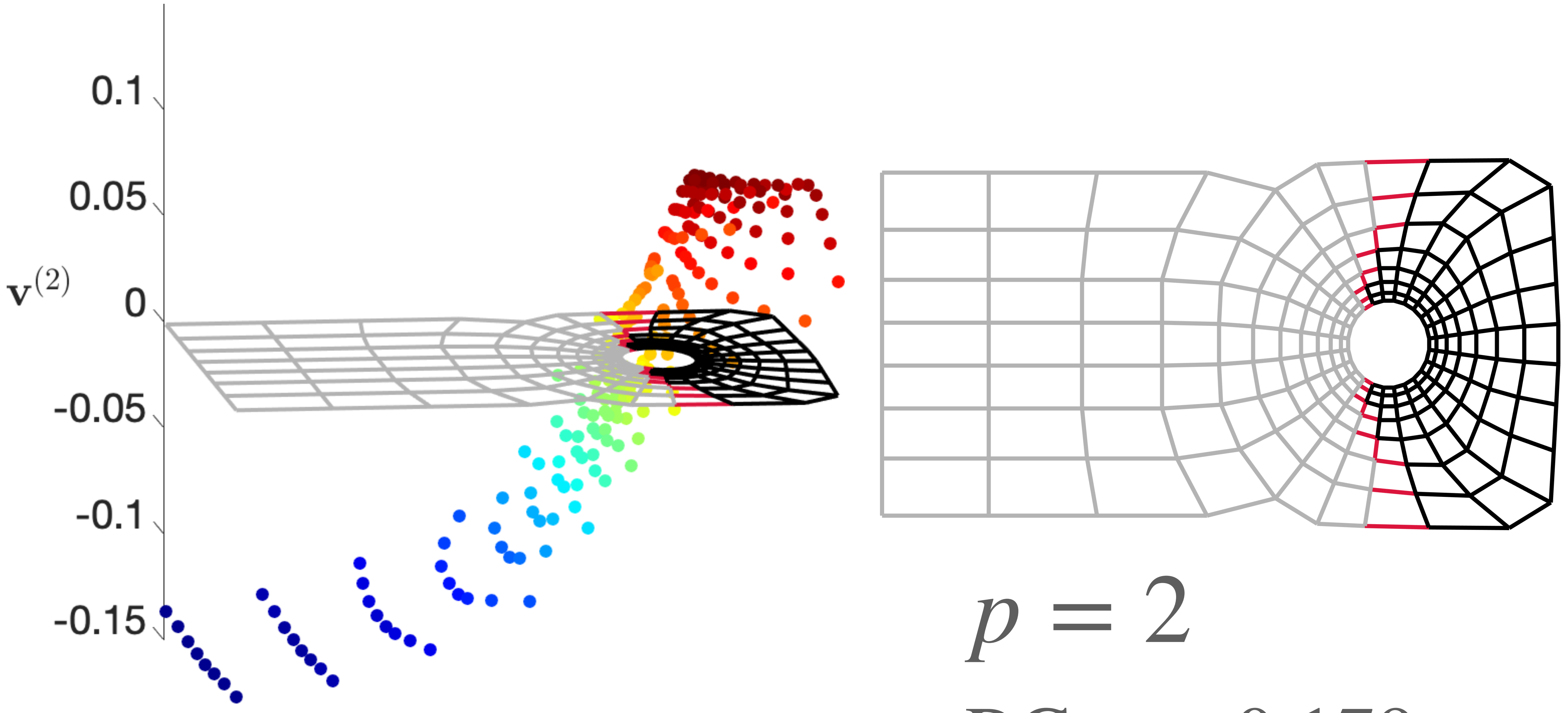
$$\min_{\boldsymbol{u}\in\mathbb{R}^n} \frac{\langle \boldsymbol{u}, \Delta_2 \boldsymbol{u}\rangle}{\|\boldsymbol{u}\|_2^2} = \min_{\boldsymbol{u}\in\mathbb{R}^n} \frac{1}{2}\frac{\sum_{i,j=1}^{n} w_{ij}\left(u_i - u_j\right)^2}{\|\boldsymbol{u}\|_2^2},$$

$$\text{s.t.} \quad \boldsymbol{u}^T \cdot \boldsymbol{e} = 0.$$

p-Laplacian, $p \in (1, 2]$

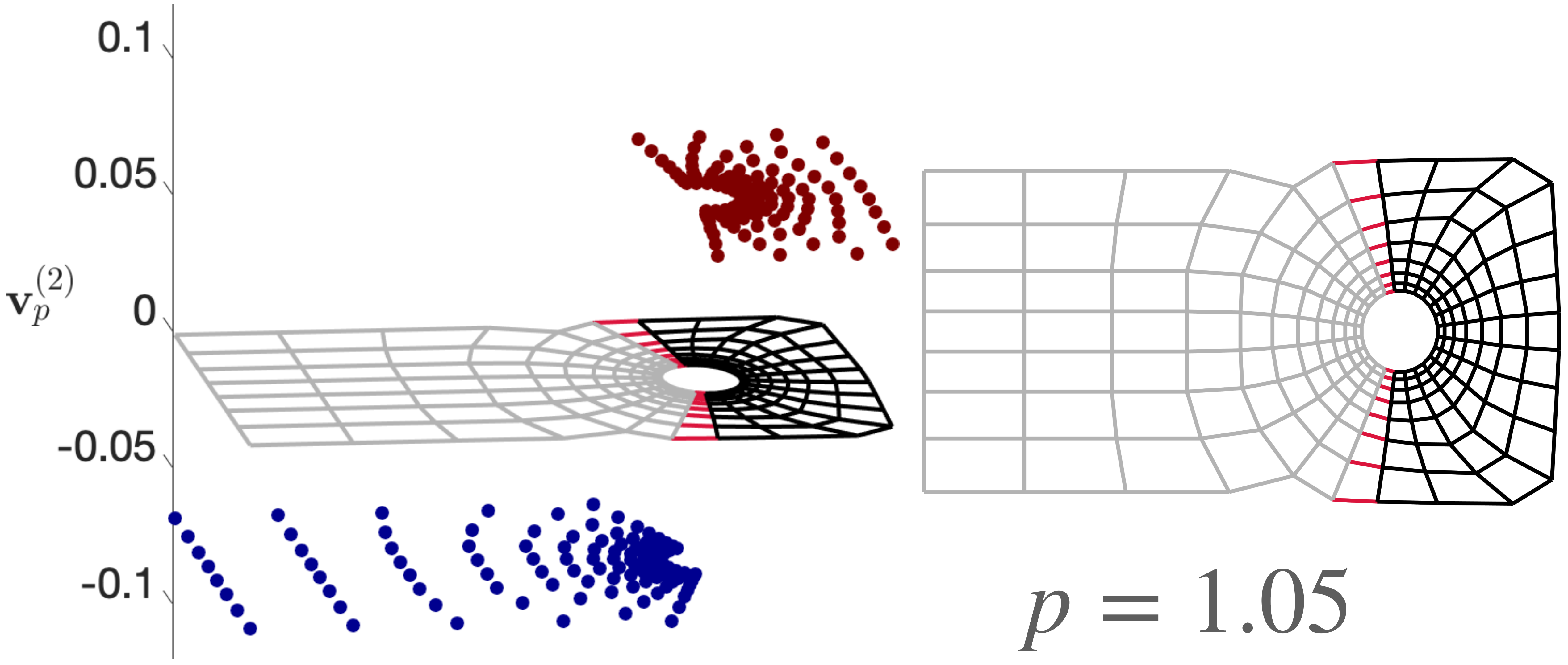
$$\min_{\boldsymbol{u}\in\mathbb{R}^n} \frac{\langle \boldsymbol{u}, \Delta_p \boldsymbol{u}\rangle}{\|\boldsymbol{u}\|_p^p} = \min_{\boldsymbol{u}\in\mathbb{R}^n} \frac{1}{2}\frac{\sum_{i,j=1}^{n} w_{ij}\left|u_i - u_j\right|^p}{\|\boldsymbol{u}\|_p^p},$$

$$\text{s.t.} \quad \boldsymbol{e}^\top \phi_p(\boldsymbol{u}) = 0$$



$p = 2$

RCut $= 0.179$



$p = 1.05$

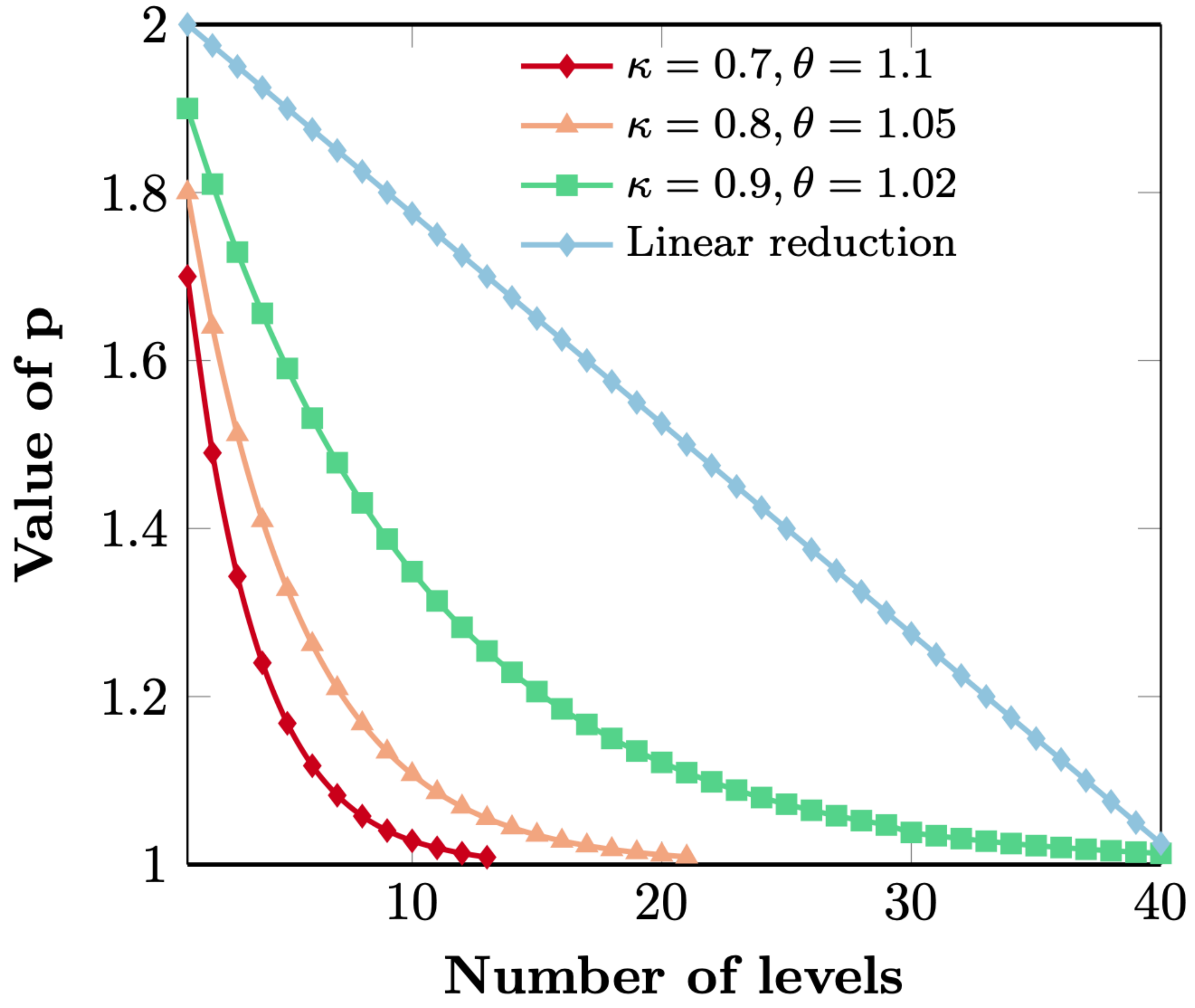
RCut $= 0.143$

Key Algorithmic Components

Pseudocontinuous minimization

$$p = 1 + \max\left(\text{tol}, \min\left(\kappa \cdot (p-1), (p-1)^{\theta}\right)\right),$$

with $\kappa \in (0,1), \theta \in (1,2)$, and $\text{tol} = 10^{-1}$.



ALGORITHM: main pGrass loop

Initialize: $\boldsymbol{c}, \mathrm{r}_{\text{new,old,best}} = \mathrm{Cut}(\boldsymbol{c})$ ▷ $p = 2$

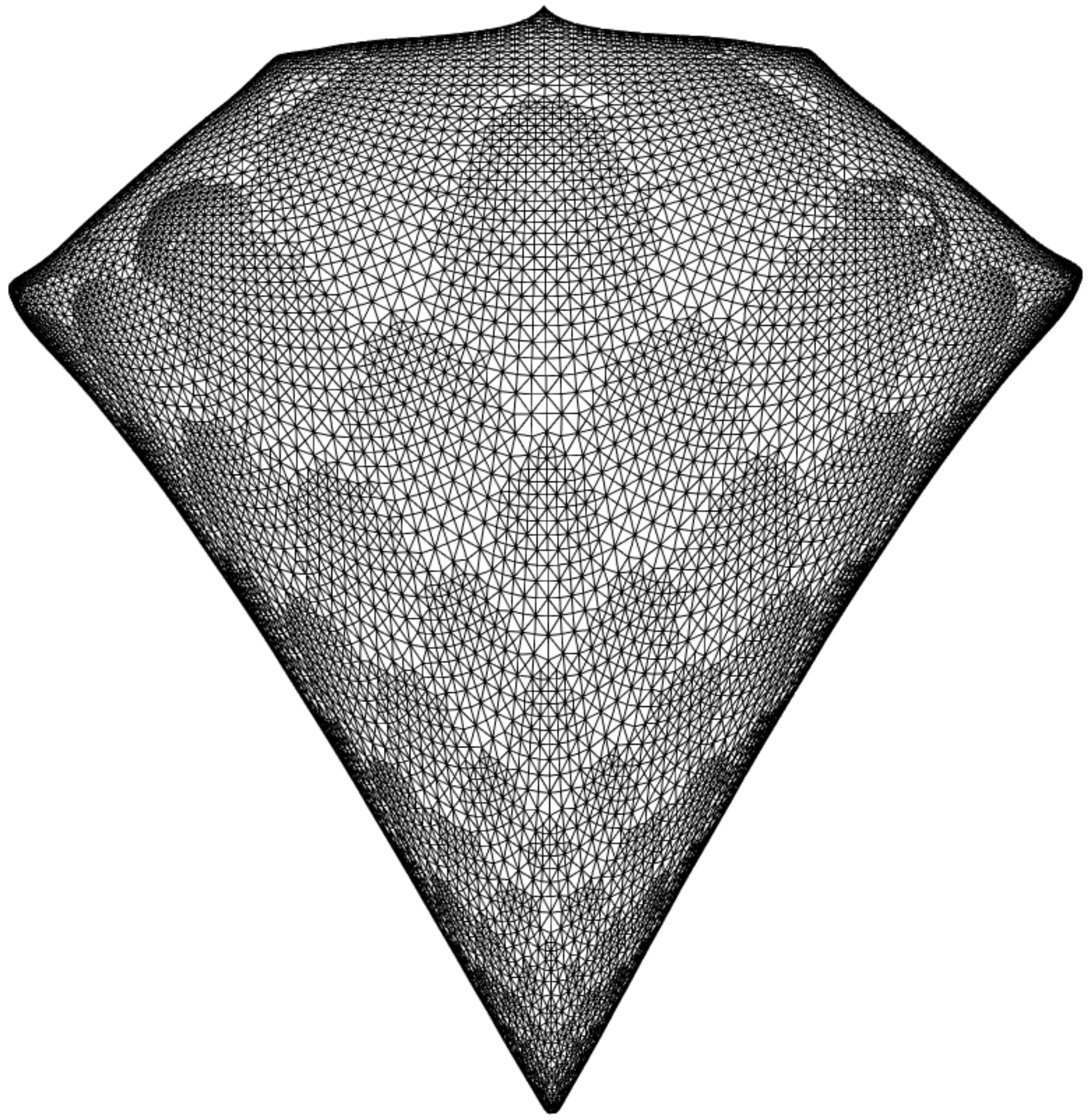
1. **while** $p \geq p_w$ && $\mathrm{r}_{\text{new}} \leq 1.05 \cdot \mathrm{r}_{\text{old}}$ **do**
2. Reduce p
3. Find $\boldsymbol{U}$: $\underset{\boldsymbol{U} \in \mathcal{Gr}(k,n)}{\text{minimize}}\ F_p(\boldsymbol{U})$ using $\boldsymbol{W}$
4. $\boldsymbol{c} = \text{discretize}(\boldsymbol{U})$
5. $\mathrm{r}_{\text{old}} = \mathrm{r}_{\text{new}}$
6. $\mathrm{r}_{\text{new}} = \mathrm{Cut}(\boldsymbol{c})$
7. **if** $\mathrm{r}_{\text{new}} < \mathrm{r}_{\text{best}}$ **then**
8. $\mathrm{r}_{\text{best}} = \mathrm{r}_{\text{new}}$
9. $\boldsymbol{c}_{\text{best}} = \boldsymbol{c}$
10. **end if**
11. **end while**

Conclusions

- A direct multiway p-spectral graph clustering framework.
- Simple algorithm, utilizing packages of Riemannian optimization.
- Consistent results over synthetic and real-world graphs.

Future Perspectives

- Embody the pGrass algorithm in a multilevel hierarchy based framework.
- Estimate optimal value of p and number of clusters k.
- High performance implementation $\rightarrow$ block eigenvalue computations.



FNSNF
Fonds national suisse
Schweizerischer Nationalfonds
Fondo nazionale svizzero
Swiss National Science Foundation