Random Sampling of Bandlimited Signals on Graphs

Pierre Vandergheynst

École Polytechnique Fédérale de Lausanne (EPFL)
School of Engineering & School of Computer and Communication Sciences

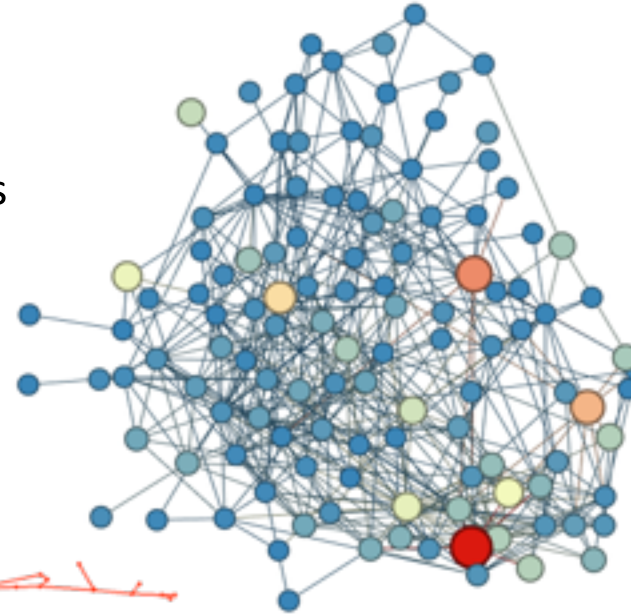
Joint work with Gilles Puy (INRIA), Nicolas Tremblay (INRIA)
and Rémi Gribonval (INRIA)

NIPS2015 Workshop Multiresolution Methods for Large Scale Learning

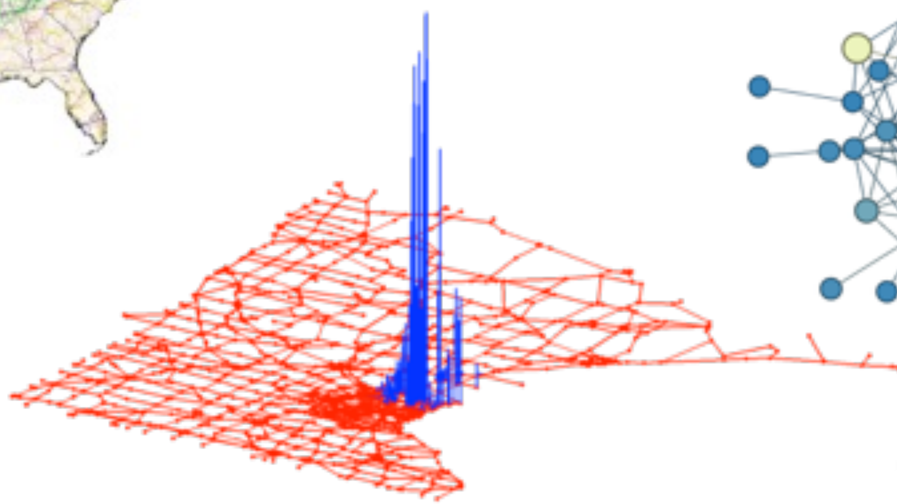
Motivation

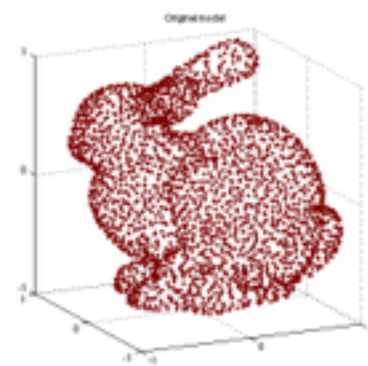
Energy Networks

Social Networks



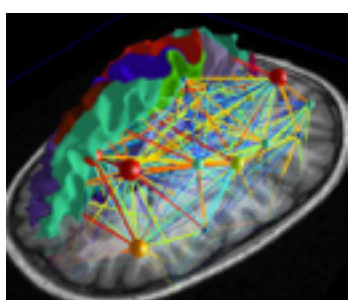
Transportation Networks





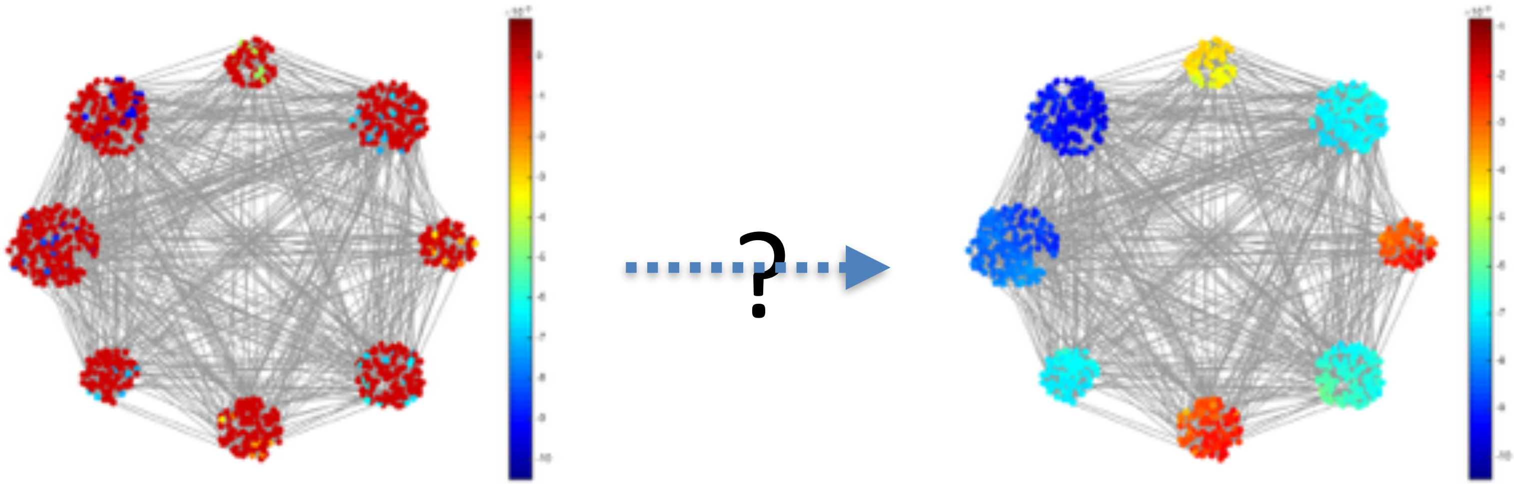
Point Clouds

Biological Networks



Goal

Given partially observed information at the nodes of a graph



Can we robustly and efficiently infer missing information ?

What signal model ?

How many observations ?

Influence of the structure of the graph ?

Notations

$\mathcal{G} = \{\mathcal{V}, \mathcal{E}, \mathsf{W}\}$ weighted, undirected

$\mathcal{V}$ is the set of n nodes

$\mathcal{E}$ is the set of edges

$\mathsf{W} \in \mathbb{R}^{n \times n}$ is the weighted adjacency matrix

$\mathsf{L} \in \mathbb{R}^{n \times n}$

combinatorial graph Laplacian $\mathsf{L} := \mathsf{D} - \mathsf{W}$

normalised Laplacian $\mathsf{L} := \mathsf{I} - \mathsf{D}^{-1/2}\mathsf{W}\mathsf{D}^{-1/2}$

diagonal degree matrix D has entries $d_i := \sum_{i \neq j} \mathsf{W}_{ij}$

Notations

$$\mathsf{L} \text{ is real, symmetric PSD}$$

orthonormal eigenvectors $\mathsf{U} \in \mathbb{R}^{n \times n}$ Graph Fourier Matrix

non-negative eigenvalues $\boldsymbol{\lambda}_1 \leqslant \boldsymbol{\lambda}_2 \leqslant \ldots, \boldsymbol{\lambda}_n$

$$\mathsf{L} = \mathsf{U}\Lambda\mathsf{U}^\mathsf{T}$$

Notations

L is real, symmetric PSD

orthonormal eigenvectors $\mathsf{U} \in \mathbb{R}^{n \times n}$ Graph Fourier Matrix

non-negative eigenvalues $\boldsymbol{\lambda}_1 \leqslant \boldsymbol{\lambda}_2 \leqslant \ldots, \boldsymbol{\lambda}_n$

$$\mathsf{L} = \mathsf{U} \Lambda \mathsf{U}^\intercal$$

k-bandlimited signals $\boldsymbol{x} \in \mathbb{R}^n$

Fourier coefficients $\hat{\boldsymbol{x}} = \mathsf{U}^\intercal \boldsymbol{x}$

$$\boldsymbol{x} = \mathsf{U}_k \hat{\boldsymbol{x}}^k \qquad \hat{\boldsymbol{x}}^k \in \mathbb{R}^k$$

$\mathsf{U}_k := (\boldsymbol{u}_1, \ldots, \boldsymbol{u}_k) \in \mathbb{R}^{n \times k}$ first k eigenvectors only

Sampling Model

$$\boldsymbol{p} \in \mathbb{R}^n \qquad \boldsymbol{p}_i > 0 \qquad \|\boldsymbol{p}\|_1 = \sum_{i=1}^{n} \boldsymbol{p}_i = 1$$

$$\mathsf{P} := \mathrm{diag}(\boldsymbol{p}) \in \mathbb{R}^{n \times n}$$

Sampling Model

$$\boldsymbol{p} \in \mathbb{R}^n \qquad \boldsymbol{p}_i > 0 \qquad \|\boldsymbol{p}\|_1 = \sum_{i=1}^{n} \boldsymbol{p}_i = 1$$

$$\mathsf{P} := \operatorname{diag}(\boldsymbol{p}) \in \mathbb{R}^{n \times n}$$

Draw independently m samples (random sampling)

$$\mathbb{P}(\omega_j = i) = \boldsymbol{p}_i, \quad \forall j \in \{1, \ldots, m\} \text{ and } \forall i \in \{1, \ldots, n\}$$

Sampling Model

$$\boldsymbol{p} \in \mathbb{R}^n \qquad \boldsymbol{p}_i > 0 \qquad \|\boldsymbol{p}\|_1 = \sum_{i=1}^{n} \boldsymbol{p}_i = 1$$

$$\mathsf{P} := \mathrm{diag}(\boldsymbol{p}) \in \mathbb{R}^{n \times n}$$

Draw independently m samples (random sampling)

$$\mathbb{P}(\omega_j = i) = \boldsymbol{p}_i, \quad \forall j \in \{1, \ldots, m\} \text{ and } \forall i \in \{1, \ldots, n\}$$

$$\boldsymbol{y}_j := \boldsymbol{x}_{\omega_j}, \quad \forall j \in \{1, \ldots, m\}$$

$$\boldsymbol{y} = \mathsf{M}\boldsymbol{x}$$

Sampling Model

$$\frac{\|\mathsf{U}_k^\intercal \boldsymbol{\delta}_i\|_2}{\|\mathsf{U}^\intercal \boldsymbol{\delta}_i\|_2} = \frac{\|\mathsf{U}_k^\intercal \boldsymbol{\delta}_i\|_2}{\|\boldsymbol{\delta}_i\|_2} = \|\mathsf{U}_k^\intercal \boldsymbol{\delta}_i\|_2$$

How much a perfect impulse can be concentrated on first k eigenvectors

Carries interesting information about the graph

Sampling Model

$$\frac{\|\mathsf{U}_k^\mathsf{T}\boldsymbol{\delta}_i\|_2}{\|\mathsf{U}^\mathsf{T}\boldsymbol{\delta}_i\|_2} = \frac{\|\mathsf{U}_k^\mathsf{T}\boldsymbol{\delta}_i\|_2}{\|\boldsymbol{\delta}_i\|_2} = \|\mathsf{U}_k^\mathsf{T}\boldsymbol{\delta}_i\|_2$$

How much a perfect impulse can be concentrated on first k eigenvectors

Carries interesting information about the graph

Ideally: $\boldsymbol{p}_i$ large wherever $\|\mathsf{U}_k^\mathsf{T}\boldsymbol{\delta}_i\|_2$ is large

Sampling Model

$$\frac{\|\mathsf{U}_k^{\intercal}\boldsymbol{\delta}_i\|_2}{\|\mathsf{U}^{\intercal}\boldsymbol{\delta}_i\|_2} = \frac{\|\mathsf{U}_k^{\intercal}\boldsymbol{\delta}_i\|_2}{\|\boldsymbol{\delta}_i\|_2} = \|\mathsf{U}_k^{\intercal}\boldsymbol{\delta}_i\|_2$$

How much a perfect impulse can be concentrated on first k eigenvectors

Carries interesting information about the graph

Ideally: $\boldsymbol{p}_i$ large wherever $\|\mathsf{U}_k^{\intercal}\boldsymbol{\delta}_i\|_2$ is large

Graph Coherence

$$\nu_{\boldsymbol{p}}^k := \max_{1\leqslant i\leqslant n}\left\{\boldsymbol{p}_i^{-1/2}\,\|\mathsf{U}_k^{\intercal}\boldsymbol{\delta}_i\|_2\right\}$$

Rem: $\nu_{\boldsymbol{p}}^k \geqslant \sqrt{k}$

Stable Embedding

Theorem 1 (Restricted isometry property)**.** *Let* M *be a random subsampling matrix with the sampling distribution* $\boldsymbol{p}$*. For any* $\delta, \epsilon \in (0,1)$*, with probability at least* $1-\epsilon$*,*

$$(1-\delta)\,\|\boldsymbol{x}\|_2^2 \leqslant \frac{1}{m}\left\|\mathsf{M}\mathsf{P}^{-1/2}\,\boldsymbol{x}\right\|_2^2 \leqslant (1+\delta)\,\|\boldsymbol{x}\|_2^2 \tag{1}$$

for all $\boldsymbol{x} \in \mathrm{span}(\mathsf{U}_k)$ *provided that*

$$m \geqslant \frac{3}{\delta^2}\,(\nu_{\boldsymbol{p}}^k)^2\,\log\left(\frac{2k}{\epsilon}\right). \tag{2}$$

Stable Embedding

Theorem 1 (Restricted isometry property)**.** *Let* M *be a random subsampling matrix with the sampling distribution* $\boldsymbol{p}$*. For any* $\delta, \epsilon \in (0,1)$*, with probability at least* $1-\epsilon$*,*

$$(1-\delta)\,\|\boldsymbol{x}\|_2^2 \leqslant \frac{1}{m}\left\|\mathsf{M}\mathsf{P}^{-1/2}\,\boldsymbol{x}\right\|_2^2 \leqslant (1+\delta)\,\|\boldsymbol{x}\|_2^2 \tag{1}$$

for all $\boldsymbol{x} \in \operatorname{span}(\mathsf{U}_k)$ *provided that*

$$m \geqslant \frac{3}{\delta^2}\,(\nu_{\boldsymbol{p}}^k)^2\,\log\left(\frac{2k}{\epsilon}\right). \tag{2}$$

$\mathsf{M}\mathsf{P}^{-1/2}\,\boldsymbol{x} = \mathsf{P}_{\Omega}^{-1/2}\mathsf{M}\boldsymbol{x}$ Only need M, re-weighting offline

Stable Embedding

Theorem 1 (Restricted isometry property)**.** *Let* M *be a random subsampling matrix with the sampling distribution* $\boldsymbol{p}$*. For any* $\delta, \epsilon \in (0, 1)$*, with probability at least* $1 - \epsilon$*,*

$$(1-\delta)\, \|\boldsymbol{x}\|_2^2 \leqslant \frac{1}{m} \left\| \mathsf{M} \mathsf{P}^{-1/2}\, \boldsymbol{x} \right\|_2^2 \leqslant (1+\delta)\, \|\boldsymbol{x}\|_2^2 \tag{1}$$

for all $\boldsymbol{x} \in \mathrm{span}(\mathsf{U}_k)$ *provided that*

$$m \geqslant \frac{3}{\delta^2}\, (\nu_{\boldsymbol{p}}^k)^2\, \log\left(\frac{2k}{\epsilon}\right). \tag{2}$$

$\mathsf{M}\mathsf{P}^{-1/2}\, \boldsymbol{x} = \mathsf{P}_\Omega^{-1/2} \mathsf{M} \boldsymbol{x}$ Only need M, re-weighting offline

$(\nu_{\boldsymbol{p}}^k)^2 \geqslant k$ Need to sample at least k nodes

Stable Embedding

Theorem 1 (Restricted isometry property). *Let* M *be a random subsampling matrix with the sampling distribution* $\boldsymbol{p}$. *For any* $\delta, \epsilon \in (0,1)$, *with probability at least* $1-\epsilon$,

$$(1-\delta)\,\|\boldsymbol{x}\|_2^2 \leqslant \frac{1}{m}\left\|\mathsf{M}\mathsf{P}^{-1/2}\,\boldsymbol{x}\right\|_2^2 \leqslant (1+\delta)\,\|\boldsymbol{x}\|_2^2 \tag{1}$$

for all $\boldsymbol{x} \in \mathrm{span}(\mathsf{U}_k)$ *provided that*

$$m \geqslant \frac{3}{\delta^2}\,(\nu_{\boldsymbol{p}}^k)^2\,\log\left(\frac{2k}{\epsilon}\right). \tag{2}$$

$\mathsf{M}\mathsf{P}^{-1/2}\,\boldsymbol{x} = \mathsf{P}_\Omega^{-1/2}\mathsf{M}\boldsymbol{x}$ — Only need M, re-weighting offline

$(\nu_{\boldsymbol{p}}^k)^2 \geqslant k$ — Need to sample at least k nodes

Proof similar to CS in bounded ONB but simpler since model is a subspace (not a union)

Stable Embedding

$(\nu_{\boldsymbol{p}}^k)^2 \geqslant k$ Need to sample at least k nodes

Stable Embedding

$(\nu_{\boldsymbol{p}}^k)^2 \geqslant k$ Need to sample at least k nodes

Can we reduce to optimal amount ?

Stable Embedding

$$(\nu_{\boldsymbol{p}}^k)^2 \geqslant k$$ Need to sample at least k nodes

Can we reduce to optimal amount ?

Variable Density Sampling $$\boldsymbol{p}_i^* := \frac{\|\mathsf{U}_k^\intercal \boldsymbol{\delta}_i\|_2^2}{k}, \quad i = 1, \ldots, n$$

is such that: $(\nu_{\boldsymbol{p}}^k)^2 = k$ and depends on structure of graph

Stable Embedding

$$(\nu_{\boldsymbol{p}}^{k})^2 \geqslant k \qquad \text{Need to sample at least } k \text{ nodes}$$

Can we reduce to optimal amount ?

Variable Density Sampling $\quad \boldsymbol{p}_i^* := \frac{\|\mathsf{U}_k^\intercal \boldsymbol{\delta}_i\|_2^2}{k}, \quad i = 1, \ldots, n$

is such that: $\quad (\nu_{\boldsymbol{p}}^{k})^2 = k \quad$ and depends on structure of graph

Corollary 1. *Let* M *be a random subsampling matrix constructed with the sampling distribution* $\boldsymbol{p}^*$. *For any* $\delta, \epsilon \in (0,1)$, *with probability at least* $1-\epsilon$,

$$(1-\delta)\,\|\boldsymbol{x}\|_2^2 \leqslant \frac{1}{m}\left\|\mathsf{M}\mathsf{P}^{-1/2}\,\boldsymbol{x}\right\|_2^2 \leqslant (1+\delta)\,\|\boldsymbol{x}\|_2^2$$

for all $\boldsymbol{x} \in \operatorname{span}(\mathsf{U}_k)$ *provided that*

$$m \geqslant \frac{3}{\delta^2}\, k\, \log\left(\frac{2k}{\epsilon}\right).$$

Recovery Procedures

$\boldsymbol{y} = \mathsf{M}\boldsymbol{x} + \boldsymbol{n}$ $\quad\quad$ $\boldsymbol{y} \in \mathbb{R}^m$

$\boldsymbol{x} \in \mathrm{span}(\mathsf{U}_k)$ $\quad$ stable embedding

Recovery Procedures

$$\boldsymbol{y} = \mathsf{M}\boldsymbol{x} + \boldsymbol{n} \qquad \boldsymbol{y} \in \mathbb{R}^m$$

$$\boldsymbol{x} \in \operatorname{span}(\mathsf{U}_k) \quad \text{stable embedding}$$

Standard Decoder

$$\min_{\boldsymbol{z} \in \operatorname{span}(\mathsf{U}_k)} \left\| \mathsf{P}_\Omega^{-1/2} (\mathsf{M}\boldsymbol{z} - \boldsymbol{y}) \right\|_2$$

Recovery Procedures

$$\boldsymbol{y} = \mathsf{M}\boldsymbol{x} + \boldsymbol{n} \qquad \boldsymbol{y} \in \mathbb{R}^m$$

$$\boldsymbol{x} \in \mathrm{span}(\mathsf{U}_k) \quad \text{stable embedding}$$

Standard Decoder

$$\min_{\boldsymbol{z} \in \mathrm{span}(\mathsf{U}_k)} \left\| \mathsf{P}_\Omega^{-1/2} (\mathsf{M}\boldsymbol{z} - \boldsymbol{y}) \right\|_2$$

need projector

Recovery Procedures

$$\boldsymbol{y} = \mathsf{M}\boldsymbol{x} + \boldsymbol{n} \qquad \boldsymbol{y} \in \mathbb{R}^m$$

$$\boldsymbol{x} \in \mathrm{span}(\mathsf{U}_k)$$ stable embedding

Standard Decoder

$$\min_{\boldsymbol{z} \in \mathrm{span}(\mathsf{U}_k)} \left\| \mathsf{P}_\Omega^{-1/2} (\mathsf{M}\boldsymbol{z} - \boldsymbol{y}) \right\|_2$$

need projector

re-weighting for RIP

Recovery Procedures

$$\boldsymbol{y} = \mathsf{M}\boldsymbol{x} + \boldsymbol{n} \qquad \boldsymbol{y} \in \mathbb{R}^m$$

$$\boldsymbol{x} \in \mathrm{span}(\mathsf{U}_k) \quad \text{stable embedding}$$

Recovery Procedures

$$\boldsymbol{y} = \mathsf{M}\boldsymbol{x} + \boldsymbol{n} \qquad \boldsymbol{y} \in \mathbb{R}^m$$

$$\boldsymbol{x} \in \operatorname{span}(\mathsf{U}_k) \quad \text{stable embedding}$$

Efficient Decoder:

$$\min_{\boldsymbol{z} \in \mathbb{R}^n} \left\| \mathsf{P}_\Omega^{-1/2} (\mathsf{M}\boldsymbol{z} - \boldsymbol{y}) \right\|_2^2 + \gamma \, \boldsymbol{z}^\intercal g(\mathsf{L}) \boldsymbol{z}$$

Recovery Procedures

$$\boldsymbol{y} = \mathsf{M}\boldsymbol{x} + \boldsymbol{n} \qquad \boldsymbol{y} \in \mathbb{R}^m$$

$\boldsymbol{x} \in \operatorname{span}(\mathsf{U}_k)$ stable embedding

Efficient Decoder:

$$\min_{\boldsymbol{z} \in \mathbb{R}^n} \left\| \mathsf{P}_{\Omega}^{-1/2} (\mathsf{M}\boldsymbol{z} - \boldsymbol{y}) \right\|_2^2 + \gamma\, \boldsymbol{z}^{\intercal} g(\mathsf{L}) \boldsymbol{z}$$

soft constrain on frequencies

efficient implementation

Analysis of Standard Decoder

Standard Decoder:

$$\min_{\boldsymbol{z} \in \mathrm{span}(\mathsf{U}_k)} \left\| \mathsf{P}_\Omega^{-1/2} \left(\mathsf{M}\boldsymbol{z} - \boldsymbol{y} \right) \right\|_2$$

Analysis of Standard Decoder

Standard Decoder:

$$\min_{\boldsymbol{z} \in \mathrm{span}(\mathsf{U}_k)} \left\| \mathsf{P}_\Omega^{-1/2} \left(\mathsf{M}\boldsymbol{z} - \boldsymbol{y} \right) \right\|_2$$

Theorem 1. *Let Ω be a set of m indices selected independently from $\{1, \ldots, n\}$ with sampling distribution $\boldsymbol{p} \in \mathbb{R}^n$, and M the associated sampling matrix. Let $\epsilon, \delta \in (0, 1)$ and $m \geqslant \frac{3}{\delta^2} \, (\nu_{\boldsymbol{p}}^k)^2 \, \log\left(\frac{2k}{\epsilon}\right)$. With probability at least $1 - \epsilon$, the following holds for all $\boldsymbol{x} \in \mathrm{span}(\mathsf{U}_k)$ and all $\boldsymbol{n} \in \mathbb{R}^m$.*

i) Let $\boldsymbol{x}^*$ be the solution of Standard Decoder with $\boldsymbol{y} = \mathsf{M}\boldsymbol{x} + \boldsymbol{n}$. Then,

$$\|\boldsymbol{x}^* - \boldsymbol{x}\|_2 \leqslant \frac{2}{\sqrt{m\,(1-\delta)}} \left\| \mathsf{P}_\Omega^{-1/2} \boldsymbol{n} \right\|_2 . \tag{1}$$

ii) There exist particular vectors $\boldsymbol{n}_0 \in \mathbb{R}^m$ such that the solution $\boldsymbol{x}^*$ of Standard Decoder with $\boldsymbol{y} = \mathsf{M}\boldsymbol{x} + \boldsymbol{n}_0$ satisfies

$$\|\boldsymbol{x}^* - \boldsymbol{x}\|_2 \geqslant \frac{1}{\sqrt{m\,(1+\delta)}} \left\| \mathsf{P}_\Omega^{-1/2} \boldsymbol{n}_0 \right\|_2 . \tag{2}$$

Analysis of Standard Decoder

Standard Decoder:

$$\min_{\boldsymbol{z} \in \mathrm{span}(\mathsf{U}_k)} \left\| \mathsf{P}_\Omega^{-1/2} \left(\mathsf{M}\boldsymbol{z} - \boldsymbol{y} \right) \right\|_2$$

Theorem 1. *Let Ω be a set of m indices selected independently from $\{1, \ldots, n\}$ with sampling distribution $\boldsymbol{p} \in \mathbb{R}^n$, and M the associated sampling matrix. Let $\epsilon, \delta \in (0,1)$ and $m \geqslant \frac{3}{\delta^2} \, (\nu_{\boldsymbol{p}}^k)^2 \, \log\left(\frac{2k}{\epsilon}\right)$. With probability at least $1-\epsilon$, the following holds for all $\boldsymbol{x} \in \mathrm{span}(\mathsf{U}_k)$ and all $\boldsymbol{n} \in \mathbb{R}^m$.*

i) Let $\boldsymbol{x}^*$ be the solution of Standard Decoder with $\boldsymbol{y} = \mathsf{M}\boldsymbol{x} + \boldsymbol{n}$. Then,

$$\|\boldsymbol{x}^* - \boldsymbol{x}\|_2 \leqslant \frac{2}{\sqrt{m\,(1-\delta)}} \left\| \mathsf{P}_\Omega^{-1/2} \boldsymbol{n} \right\|_2 . \quad (1)$$

Exact recovery when noiseless

ii) There exist particular vectors $\boldsymbol{n}_0 \in \mathbb{R}^m$ such that the solution $\boldsymbol{x}^*$ of Standard Decoder with $\boldsymbol{y} = \mathsf{M}\boldsymbol{x} + \boldsymbol{n}_0$ satisfies

$$\|\boldsymbol{x}^* - \boldsymbol{x}\|_2 \geqslant \frac{1}{\sqrt{m\,(1+\delta)}} \left\| \mathsf{P}_\Omega^{-1/2} \boldsymbol{n}_0 \right\|_2 . \quad (2)$$

Analysis of Standard Decoder

Standard Decoder:

$$\min_{\boldsymbol{z} \in \text{span}(\mathsf{U}_k)} \left\| \mathsf{P}_\Omega^{-1/2} \left(\mathsf{M}\boldsymbol{z} - \boldsymbol{y} \right) \right\|_2$$

Theorem 1. *Let Ω be a set of m indices selected independently from $\{1, \ldots, n\}$ with sampling distribution $\boldsymbol{p} \in \mathbb{R}^n$, and M the associated sampling matrix. Let $\epsilon, \delta \in (0,1)$ and $m \geqslant \frac{3}{\delta^2} \, (\nu_{\boldsymbol{p}}^k)^2 \, \log\left(\frac{2k}{\epsilon}\right)$. With probability at least $1-\epsilon$, the following holds for all $\boldsymbol{x} \in \text{span}(\mathsf{U}_k)$ and all $\boldsymbol{n} \in \mathbb{R}^m$.*

i) Let $\boldsymbol{x}^*$ be the solution of Standard Decoder with $\boldsymbol{y} = \mathsf{M}\boldsymbol{x} + \boldsymbol{n}$. Then,

$$\|\boldsymbol{x}^* - \boldsymbol{x}\|_2 \leqslant \frac{2}{\sqrt{m\,(1-\delta)}} \left\| \mathsf{P}_\Omega^{-1/2} \boldsymbol{n} \right\|_2 . \tag{1}$$

Exact recovery when noiseless

ii) There exist particular vectors $\boldsymbol{n}_0 \in \mathbb{R}^m$ such that the solution $\boldsymbol{x}^*$ of Standard Decoder with $\boldsymbol{y} = \mathsf{M}\boldsymbol{x} + \boldsymbol{n}_0$ satisfies

$$\|\boldsymbol{x}^* - \boldsymbol{x}\|_2 \geqslant \frac{1}{\sqrt{m\,(1+\delta)}} \left\| \mathsf{P}_\Omega^{-1/2} \boldsymbol{n}_0 \right\|_2 . \tag{2}$$

Analysis of Efficient Decoder

Efficient Decoder:

$$\min_{\boldsymbol{z} \in \mathbb{R}^n} \left\| \mathsf{P}_{\Omega}^{-1/2} \left(\mathsf{M}\boldsymbol{z} - \boldsymbol{y} \right) \right\|_2^2 + \gamma \; \underbrace{\boldsymbol{z}^{\intercal} g(\mathsf{L}) \boldsymbol{z}}_{\text{non-negative}}$$

Analysis of Efficient Decoder

Efficient Decoder:

$$\min_{\boldsymbol{z}\in\mathbb{R}^n} \left\| \mathsf{P}_\Omega^{-1/2} (\mathsf{M}\boldsymbol{z} - \boldsymbol{y}) \right\|_2^2 + \gamma\, \boldsymbol{z}^\intercal g(\mathsf{L}) \boldsymbol{z}$$

non-negative

Filter reshapes Fourier coefficients

$$h : \mathbb{R} \to \mathbb{R} \qquad \boldsymbol{x}_h := \mathsf{U}\,\mathrm{diag}(\hat{\boldsymbol{h}})\,\mathsf{U}^\intercal \boldsymbol{x} \in \mathbb{R}^n$$

$$\hat{\boldsymbol{h}} = (h(\boldsymbol{\lambda}_1), \ldots, h(\boldsymbol{\lambda}_n))^\intercal \in \mathbb{R}^n$$

Analysis of Efficient Decoder

Efficient Decoder:

$$\min_{\boldsymbol{z}\in\mathbb{R}^n} \left\| \mathsf{P}_\Omega^{-1/2} (\mathsf{M}\boldsymbol{z} - \boldsymbol{y}) \right\|_2^2 + \gamma\, \boldsymbol{z}^\intercal g(\mathsf{L})\boldsymbol{z}$$

non-negative

Filter reshapes Fourier coefficients

$$h : \mathbb{R} \to \mathbb{R} \qquad \boldsymbol{x}_h := \mathsf{U}\,\mathrm{diag}(\hat{\boldsymbol{h}})\,\mathsf{U}^\intercal \boldsymbol{x} \in \mathbb{R}^n$$

$$\hat{\boldsymbol{h}} = (h(\boldsymbol{\lambda}_1), \ldots, h(\boldsymbol{\lambda}_n))^\intercal \in \mathbb{R}^n$$

$$p(t) = \sum_{i=0}^{d} \alpha_i\, t^i \qquad \boldsymbol{x}_p = \mathsf{U}\,\mathrm{diag}(\hat{\boldsymbol{p}})\,\mathsf{U}^\intercal \boldsymbol{x} = \sum_{i=0}^{d} \alpha_i\, \mathsf{L}^i \boldsymbol{x}$$

Pick special polynomials and use e.g. recurrence relations for fast filtering (with sparse matrix-vector multiply only)

Analysis of Efficient Decoder

Efficient Decoder:

$$\min_{\boldsymbol{z}\in\mathbb{R}^n} \left\| \mathsf{P}_{\Omega}^{-1/2} \left(\mathsf{M}\boldsymbol{z} - \boldsymbol{y} \right) \right\|_2^2 + \gamma \, \boldsymbol{z}^{\intercal} g(\mathsf{L}) \boldsymbol{z}$$

non-negative

non-decreasing =
penalizes high-frequencies

Analysis of Efficient Decoder

Efficient Decoder:

$$\min_{z \in \mathbb{R}^n} \left\| \mathsf{P}_\Omega^{-1/2} (\mathsf{M}z - y) \right\|_2^2 + \gamma\, z^\intercal g(\mathsf{L}) z$$

non-negative

non-decreasing $\Longrightarrow$
penalizes high-frequencies

Favours reconstruction of approximately band-limited signals

Ideal filter yields Standard Decoder

$$i_{\lambda_k}(t) := \begin{cases} 0 & \text{if } t \in [0, \lambda_k], \\ +\infty & \text{otherwise,} \end{cases}$$

Analysis of Efficient Decoder

Theorem 1. *Let* Ω, M, P, m *as before and* $M_{\max} > 0$ *be a constant such that* $\left\|\mathsf{M}\mathsf{P}^{-1/2}\right\|_2 \leqslant M_{\max}$. *Let* $\epsilon, \delta \in (0,1)$. *With probability at least* $1-\epsilon$, *the following holds for all* $\boldsymbol{x} \in \operatorname{span}(\mathsf{U}_k)$, *all* $\boldsymbol{n} \in \mathbb{R}^n$, *all* $\gamma > 0$, *and all nonnegative and nondecreasing polynomial functions* g *such that* $g(\boldsymbol{\lambda}_{k+1}) > 0$.

Let $\boldsymbol{x}^*$ *be the solution of Efficient Decoder with* $\boldsymbol{y} = \mathsf{M}\boldsymbol{x} + \boldsymbol{n}$. *Then,*

$$\|\boldsymbol{\alpha}^* - \boldsymbol{x}\|_2 \leqslant \frac{1}{\sqrt{m(1-\delta)}} \left[\left(2 + \frac{M_{\max}}{\sqrt{\gamma g(\boldsymbol{\lambda}_{k+1})}}\right) \left\|\mathsf{P}_\Omega^{-1/2}\boldsymbol{n}\right\|_2 + \left(M_{\max}\sqrt{\frac{g(\boldsymbol{\lambda}_k)}{g(\boldsymbol{\lambda}_{k+1})}} + \sqrt{\gamma g(\boldsymbol{\lambda}_k)}\right)\|\boldsymbol{x}\|_2 \right], \tag{1}$$

and

$$\|\boldsymbol{\beta}^*\|_2 \leqslant \frac{1}{\sqrt{\gamma g(\boldsymbol{\lambda}_{k+1})}} \left\|\mathsf{P}_\Omega^{-1/2}\boldsymbol{n}\right\|_2 + \sqrt{\frac{g(\boldsymbol{\lambda}_k)}{g(\boldsymbol{\lambda}_{k+1})}}\,\|\boldsymbol{x}\|_2, \tag{2}$$

where $\boldsymbol{\alpha}^* := \mathsf{U}_k\mathsf{U}_k^\intercal\,\boldsymbol{x}^*$ *and* $\boldsymbol{\beta}^* := (\mathsf{I} - \mathsf{U}_k\mathsf{U}_k^\intercal)\,\boldsymbol{x}^*$.

Analysis of Efficient Decoder

Noiseless case:

$$\|\boldsymbol{x}^* - \boldsymbol{x}\|_2 \leqslant \frac{1}{\sqrt{m(1-\delta)}} \left(M_{\max} \sqrt{\frac{g(\boldsymbol{\lambda}_k)}{g(\boldsymbol{\lambda}_{k+1})}} + \sqrt{\gamma g(\boldsymbol{\lambda}_k)} \right) \|\boldsymbol{x}\|_2 + \sqrt{\frac{g(\boldsymbol{\lambda}_k)}{g(\boldsymbol{\lambda}_{k+1})}} \|\boldsymbol{x}\|_2$$

$g(\boldsymbol{\lambda}_k) = 0$ + non-decreasing implies perfect reconstruction

Analysis of Efficient Decoder

Noiseless case:

$$\|\boldsymbol{x}^* - \boldsymbol{x}\|_2 \leqslant \frac{1}{\sqrt{m(1-\delta)}} \left(M_{\max} \sqrt{\frac{g(\boldsymbol{\lambda}_k)}{g(\boldsymbol{\lambda}_{k+1})}} + \sqrt{\gamma g(\boldsymbol{\lambda}_k)} \right) \|\boldsymbol{x}\|_2 + \sqrt{\frac{g(\boldsymbol{\lambda}_k)}{g(\boldsymbol{\lambda}_{k+1})}} \|\boldsymbol{x}\|_2$$

$g(\boldsymbol{\lambda}_k) = 0$ + non-decreasing implies perfect reconstruction

Otherwise:

choose γ as close as possible to 0 and seek to minimise the ratio $g(\boldsymbol{\lambda}_k)/g(\boldsymbol{\lambda}_{k+1})$

Choose filter to increase spectral gap ?

Clusters are of course good

Noise: $\|\mathsf{P}_\Omega^{-1/2}\boldsymbol{n}\|_2 / \|\boldsymbol{x}\|_2$

Estimating the Optimal Distribution

Estimating the Optimal Distribution

Need to estimate $\|\mathsf{U}_k^\intercal \boldsymbol{\delta}_i\|_2^2$

Estimating the Optimal Distribution

Need to estimate $\|\mathsf{U}_k^{\intercal}\boldsymbol{\delta}_i\|_2^2$

Filter random signals with ideal low-pass filter:

$$\boldsymbol{r}_{b_{\boldsymbol{\lambda}_k}} = \mathsf{U}\,\mathrm{diag}(\boldsymbol{\lambda}_1, \ldots, \boldsymbol{\lambda}_k, 0, \ldots, 0)\,\mathsf{U}^{\intercal}\,\boldsymbol{r} = \mathsf{U}_k\mathsf{U}_k^{\intercal}\,\boldsymbol{r}$$

$$\mathbb{E}\left(\boldsymbol{r}_{b_{\boldsymbol{\lambda}_k}}\right)_i^2 = \boldsymbol{\delta}_i^{\intercal}\mathsf{U}_k\mathsf{U}_k^{\intercal}\,\mathbb{E}(\boldsymbol{r}\boldsymbol{r}^{\intercal})\,\mathsf{U}_k\mathsf{U}_k^{\intercal}\boldsymbol{\delta}_i = \|\mathsf{U}_k^{\intercal}\boldsymbol{\delta}_i\|_2^2$$

Estimating the Optimal Distribution

Need to estimate $\|\mathsf{U}_k^\intercal \boldsymbol{\delta}_i\|_2^2$

Filter random signals with ideal low-pass filter:

$$\boldsymbol{r}_{b_{\boldsymbol{\lambda}_k}} = \mathsf{U}\,\mathrm{diag}(\boldsymbol{\lambda}_1, \ldots, \boldsymbol{\lambda}_k, 0, \ldots, 0)\,\mathsf{U}^\intercal\,\boldsymbol{r} = \mathsf{U}_k \mathsf{U}_k^\intercal\,\boldsymbol{r}$$

$$\mathbb{E}\,(\boldsymbol{r}_{b_{\boldsymbol{\lambda}_k}})_i^2 = \boldsymbol{\delta}_i^\intercal \mathsf{U}_k \mathsf{U}_k^\intercal\,\mathbb{E}(\boldsymbol{r}\boldsymbol{r}^\intercal)\,\mathsf{U}_k \mathsf{U}_k^\intercal \boldsymbol{\delta}_i = \|\mathsf{U}_k^\intercal \boldsymbol{\delta}_i\|_2^2$$

In practice, one may use a polynomial approximation of the ideal filter and:

$$\tilde{\boldsymbol{p}}_i := \frac{\sum_{l=1}^{L} (\boldsymbol{r}^l_{c_{\lambda_k}})_i^2}{\sum_{i=1}^{n} \sum_{l=1}^{L} (\boldsymbol{r}^l_{c_{\lambda_k}})_i^2}$$

$$L \geqslant \frac{C}{\delta^2} \log\left(\frac{2n}{\epsilon}\right)$$

Estimating the Eigengap

Estimating the Eigengap

Again, low-pass filtering random signals:

$$(1-\delta)\sum_{i=1}^{n}\left\|\mathsf{U}_{j*}^{\mathsf{T}}\boldsymbol{\delta}_i\right\|_2^2 \leqslant \sum_{i=1}^{n}\sum_{l=1}^{L}(\boldsymbol{r}_{b_\lambda}^{l})_i^2 \leqslant (1+\delta)\sum_{i=1}^{n}\left\|\mathsf{U}_{j*}^{\mathsf{T}}\boldsymbol{\delta}_i\right\|_2^2$$

Estimating the Eigengap

Again, low-pass filtering random signals:

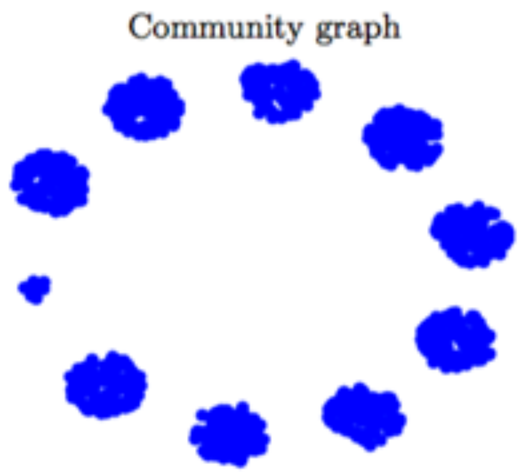
$$(1-\delta)\sum_{i=1}^{n}\left\|\mathsf{U}_{j^*}^{\mathsf{T}}\boldsymbol{\delta}_i\right\|_2^2 \leqslant \sum_{i=1}^{n}\sum_{l=1}^{L}(\boldsymbol{r}_{b_\lambda}^l)_i^2 \leqslant (1+\delta)\sum_{i=1}^{n}\left\|\mathsf{U}_{j^*}^{\mathsf{T}}\boldsymbol{\delta}_i\right\|_2^2$$

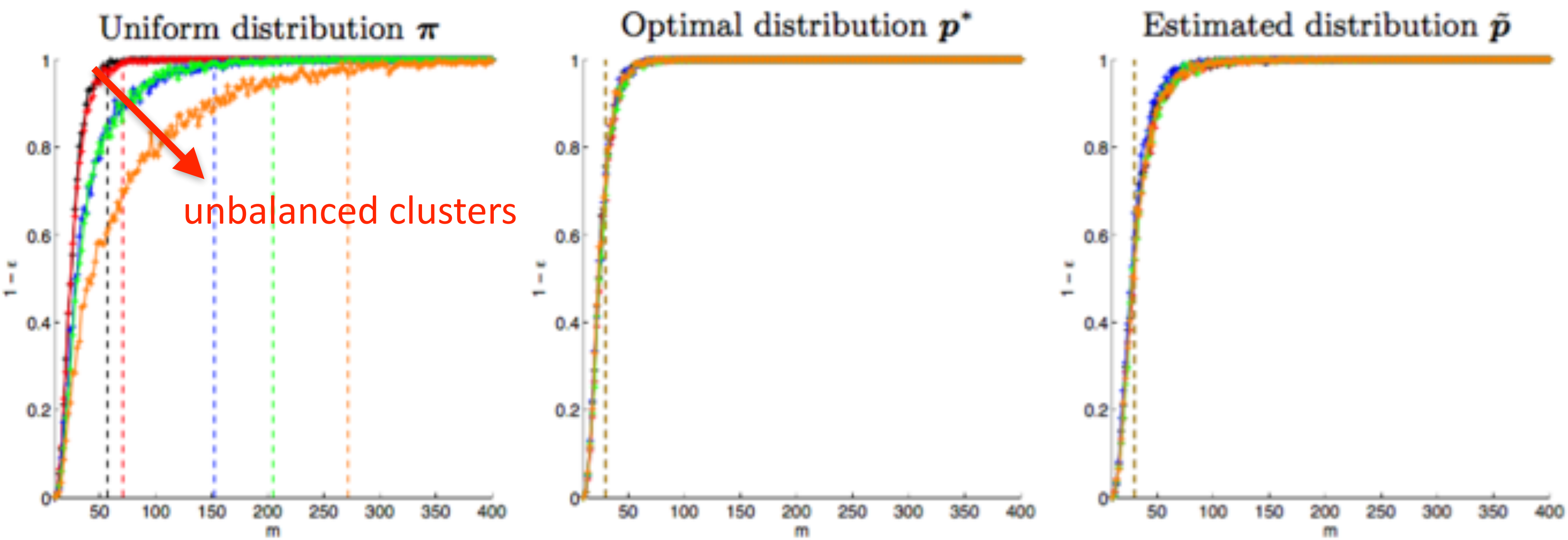
Since: $$\sum_{i=1}^{n}\left\|\mathsf{U}_{j^*}^{\mathsf{T}}\boldsymbol{\delta}_i\right\|_2^2 = \|\mathsf{U}_{j^*}\|_{\text{Frob}}^2 = j^*$$

We have: $$(1-\delta)\, j^* \leqslant \sum_{i=1}^{n}\sum_{l=1}^{L}(\boldsymbol{r}_{b_\lambda}^l)_i^2 \leqslant (1+\delta)\, j^*$$

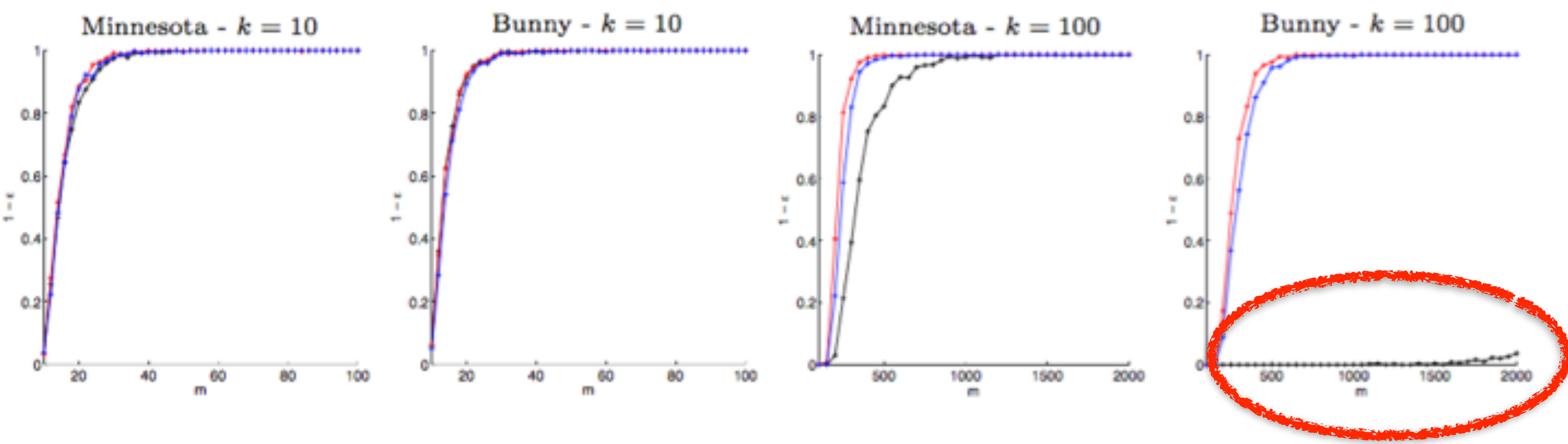
Dichotomy using the filter bandwidth

Experiments

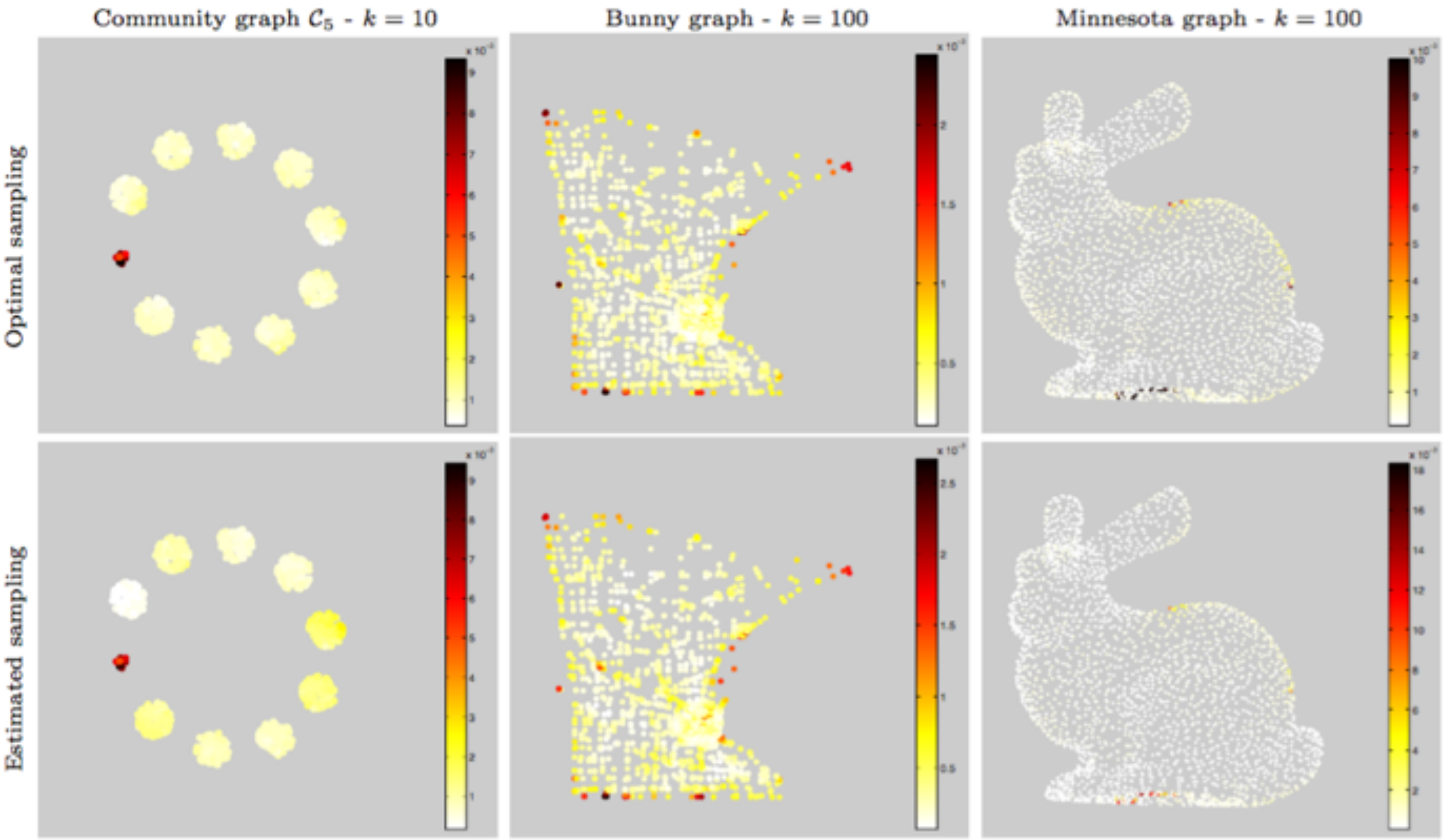




Experiments



Experiments



Experiments

(a) (b)

7%

Compressive Spectral Clustering

Clustering equivalent to recovery of cluster assignment functions

Well-defined clusters -> band-limited assignment functions!

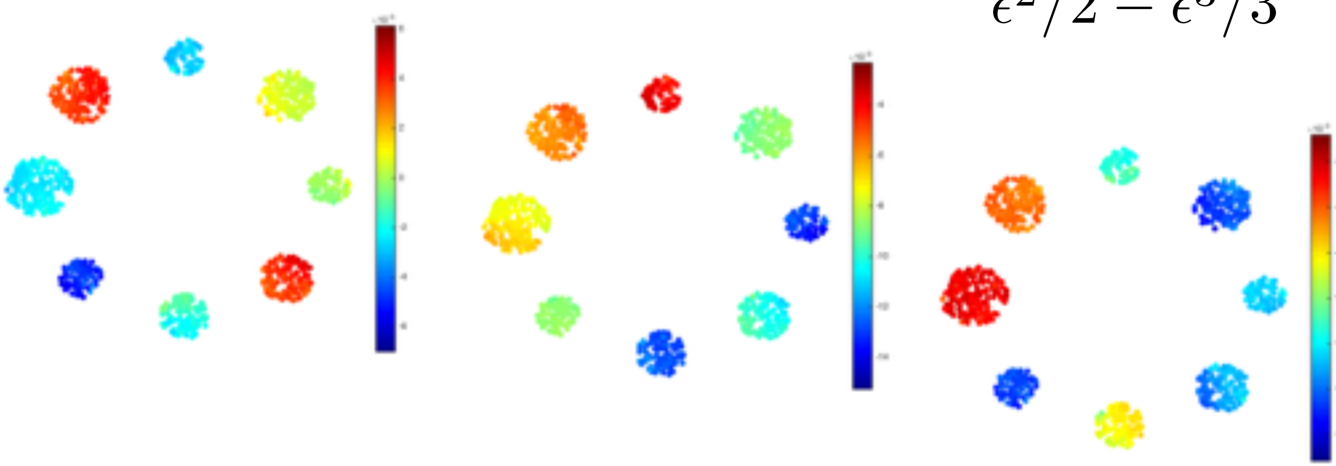
Compressive Spectral Clustering

Clustering equivalent to recovery of cluster assignment functions

Well-defined clusters -> band-limited assignment functions!

Generate features by filtering random signals

by Johnson-Lindenstrauss $\eta = \frac{4 + 2\beta}{\epsilon^2/2 - \epsilon^3/3} \log n$



Compressive Spectral Clustering

Clustering equivalent to recovery of cluster assignment functions

Well-defined clusters -> band-limited assignment functions!

Generate features by filtering random signals

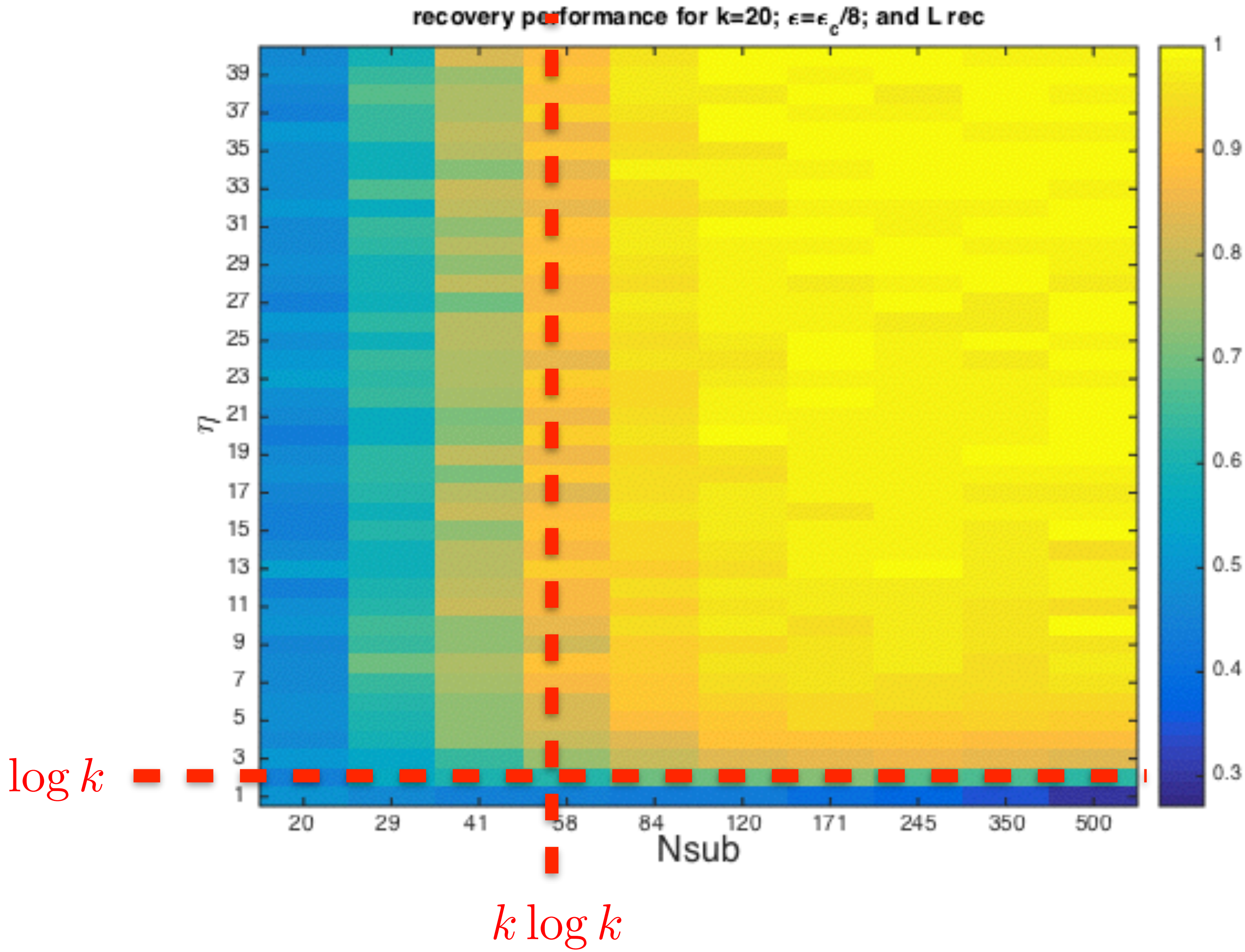
by Johnson-Lindenstrauss $$\eta = \frac{4 + 2\beta}{\epsilon^2/2 - \epsilon^3/3} \log n$$

Each feature map is smooth, therefore keep

$$m \geqslant \frac{6}{\delta^2} \, \nu_k^2 \, \log\left(\frac{k}{\epsilon'}\right)$$

Use k-means on compressed data and feed into Efficient Decoder

Compressive Spectral Clustering



Conclusion

Conclusion

- ***Stable, robust and universal random sampling*** of smoothly varying information on graphs.

Conclusion

- ***Stable, robust and universal random sampling*** of smoothly varying information on graphs.

- Tractable decoder with guarantees

Conclusion

- ***Stable, robust and universal random sampling*** of smoothly varying information on graphs.
- Tractable decoder with guarantees
- ***Optimal sampling distribution*** depends on graph structure

Conclusion

- ***Stable, robust and universal random sampling*** of smoothly varying information on graphs.
- Tractable decoder with guarantees
- ***Optimal sampling distribution*** depends on graph structure
- Can be used for inference, (SVD less) compressive clustering

Thank you !