Projet RMT4GRAPH

Evaluation à mi-parcours

Romain COUILLET
(en collaboration avec Hafiz Tiomoko Ali)

CentraleSupélec (Paris, France)

9 septembre 2016

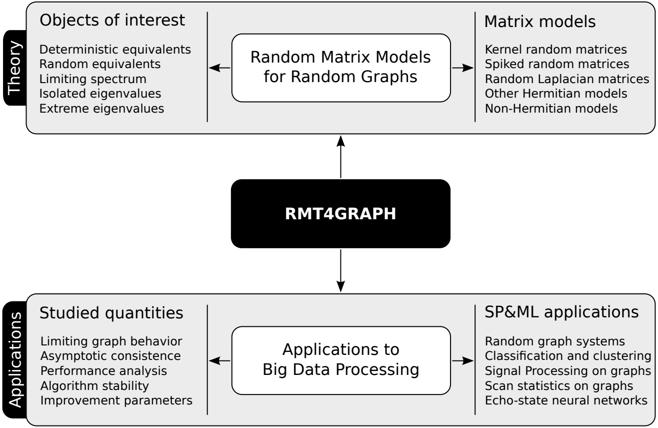
CentraleSupélec

Outline

Outline

Project definition and objectives

"develop a framework of big data processing analysis (notably graph-based methods) relying on random matrix tools"



Work Packages and Timeline

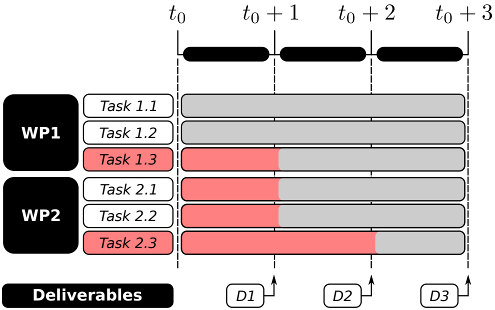
- **WP1. Random Matrix Models for Random Graphs.**
  - Task 1.1. Kernel random matrix models.
  - Task 1.2. Hermitian models and spikes.
  - Task 1.3. Random matrices with non-linear or recursive entries
    (formerly "Task 1.3. Non-hermitian random matrix models)

Work Packages and Timeline

- **WP1. Random Matrix Models for Random Graphs.**
  - Task 1.1. Kernel random matrix models.
  - Task 1.2. Hermitian models and spikes.
  - Task 1.3. Random matrices with non-linear or recursive entries
    (formerly "Task 1.3. Non-hermitian random matrix models)

- **WP2. Applications to Big Data Processing.**
  - Task 2.1. Applications to machine learning.
  - Task 2.2. Signal processing on graphs
  - Task 2.3. Neural networks
    (formerly restricted to "Task 2.3. Echo-state neural networks)

- **WP1. Random Matrix Models for Random Graphs.**
  - Task 1.1. Kernel random matrix models.
  - Task 1.2. Hermitian models and spikes.
  - Task 1.3. Random matrices with non-linear or recursive entries
    (formerly "Task 1.3. Non-hermitian random matrix models)
- **WP2. Applications to Big Data Processing.**
  - Task 2.1. Applications to machine learning.
  - Task 2.2. Signal processing on graphs
  - Task 2.3. Neural networks
    (formerly restricted to "Task 2.3. Echo-state neural networks)



Taskforce and Actions

Taskforce

- **Principal Investigator**. Romain Couillet.

Taskforce and Actions

Taskforce

- **Principal Investigator**. Romain Couillet.
- **Students**.
  - Hafiz Tiomoko Ali (PhD student, RMT4GRAPH grant, 2015-2018): community detection, neural networks.
  - Xiaoyi Mai (PhD student, DIGICOSME grant, 2016-2019): semi-supervised learning.
  - Zhenyu Liao (intern, ERC-MORE grant, 2016): support vector machines.
  - Cosme Louart (intern, ERC-MORE grant, 2016): neural networks (extreme learning machines).
  - Evgeny Kusmenko (PhD student, ERC-MORE grant, jan. 2015-dec. 2015): spectral clustering.
  - Harry Sevi (intern, ERC-MORE grant, 2015): echo-state neural networks.

Taskforce and Actions

Taskforce

- **Principal Investigator**. Romain Couillet.
- **Students**.
  - Hafiz Tiomoko Ali (PhD student, RMT4GRAPH grant, 2015-2018): community detection, neural networks.
  - Xiaoyi Mai (PhD student, DIGICOSME grant, 2016-2019): semi-supervised learning.
  - Zhenyu Liao (intern, ERC-MORE grant, 2016): support vector machines.
  - Cosme Louart (intern, ERC-MORE grant, 2016): neural networks (extreme learning machines).
  - Evgeny Kusmenko (PhD student, ERC-MORE grant, jan. 2015-dec. 2015): spectral clustering.
  - Harry Sevi (intern, ERC-MORE grant, 2015): echo-state neural networks.
- **Collaborators**.
  - Florent Benaych-Georges (professor at Universit Paris Descartes): kernel random matrices.
  - Gilles Wainrib (assistant professor at ENS Paris): neural networks.
  - Abla Kammoun (research scientist at KAUST): subspace clustering.

Taskforce and Actions

Taskforce

- **Principal Investigator**. Romain Couillet.
- **Students**.
  - Hafiz Tiomoko Ali (PhD student, RMT4GRAPH grant, 2015-2018): community detection, neural networks.
  - Xiaoyi Mai (PhD student, DIGICOSME grant, 2016-2019): semi-supervised learning.
  - Zhenyu Liao (intern, ERC-MORE grant, 2016): support vector machines.
  - Cosme Louart (intern, ERC-MORE grant, 2016): neural networks (extreme learning machines).
  - Evgeny Kusmenko (PhD student, ERC-MORE grant, jan. 2015-dec. 2015): spectral clustering.
  - Harry Sevi (intern, ERC-MORE grant, 2015): echo-state neural networks.
- **Collaborators**.
  - Florent Benaych-Georges (professor at Universit Paris Descartes): kernel random matrices.
  - Gilles Wainrib (assistant professor at ENS Paris): neural networks.
  - Abla Kammoun (research scientist at KAUST): subspace clustering.

Actions and Publications.

- **Publications**: 3 journal articles, 7 conference articles
- **Dissemination**:
  - Organization of the summer school "Large Random Matrices and High Dimensional Statistical Signal Processing", Telecom ParisTech, June 7-8, 2016.
  - SSP'16 Special Session "Random matrices in signal processing and machine learning"
  - Distinguished keynote speaker at EUSIPCO 2016
  - Special Issue on Random Matrices in "Revue du Traitement du Signal"
  - Several invited talks and contributions to local events

Outline

System Setting

Assume n-node, m-edge graph G, with

- "intrinsic" average connectivity $q_1, \ldots, q_n \sim \mu$ i.i.d.

System Setting

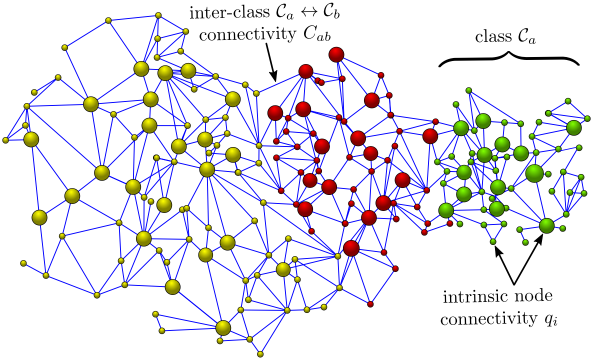
Assume n-node, m-edge graph G, with

- "intrinsic" average connectivity $q_1, \ldots, q_n \sim \mu$ i.i.d.
- k classes $\mathcal{C}_1, \ldots, \mathcal{C}_k$ independent of $\{q_i\}$ of (large) sizes $n_1, \ldots, n_k$, with preferential attachment C_{ab} between $\mathcal{C}_a$ and $\mathcal{C}_b$

Assume n-node, m-edge graph G, with

- "intrinsic" average connectivity $q_1, \ldots, q_n \sim \mu$ i.i.d.
- k classes $\mathcal{C}_1, \ldots, \mathcal{C}_k$ independent of $\{q_i\}$ of (large) sizes $n_1, \ldots, n_k$, with preferential attachment C_{ab} between $\mathcal{C}_a$ and $\mathcal{C}_b$
- induces edge probability for node $i \in \mathcal{C}_a$, $j \in \mathcal{C}_b$,

$$P(i \sim j) = q_i q_j C_{ab}.$$



System Setting

Objective:
Understand and improve performance of spectral community detection methods:

- based on adjacency A or modularity $A - \frac{dd^{\mathsf{T}}}{2m}$ matrices (adapted to **dense nets**)

System Setting

Objective:

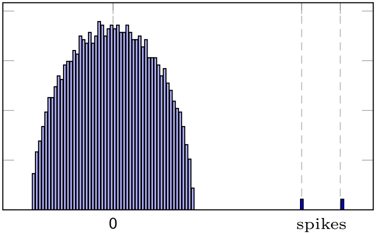
Understand and improve performance of spectral community detection methods:

- based on adjacency A or modularity $A - \frac{dd^{\mathsf{T}}}{2m}$ matrices (adapted to **dense nets**)
- based on Bethe Hessian $(r^2 - 1)I_n - rA + D$ (adapted to **sparse nets!**).

System Setting

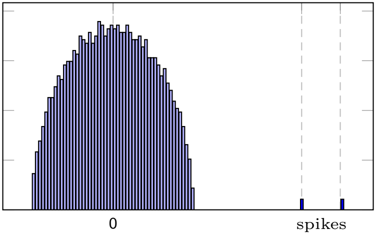
Objective:
Understand and improve performance of spectral community detection methods:

- based on adjacency A or modularity $A - \frac{dd^{\mathsf{T}}}{2m}$ matrices (adapted to **dense nets**)
- based on Bethe Hessian $(r^2 - 1)I_n - rA + D$ (adapted to **sparse nets!**).

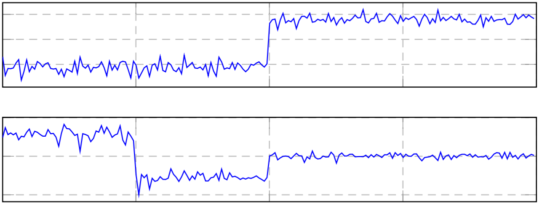


Objective:
Understand and improve performance of spectral community detection methods:

- based on adjacency A or modularity $A - \frac{dd^{\mathsf{T}}}{2m}$ matrices (adapted to **dense nets**)
- based on Bethe Hessian $(r^2 - 1)I_n - rA + D$ (adapted to **sparse nets!**).



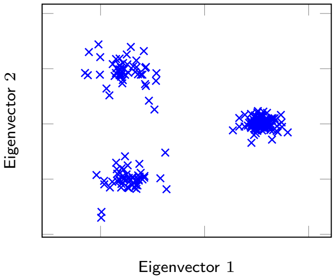
⇓ **Eigenvectors** ⇓



System Setting

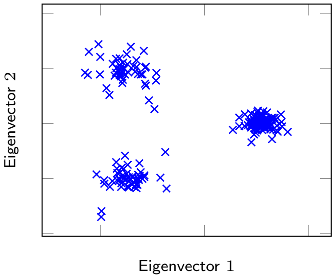
System Setting

⇓ p-**dimensional representation** ⇓



System Setting

⇓ p-dimensional representation ⇓



⇓

EM or k-means clustering.

Limitations of Adjacency/Modularity Approach

Scenario: 3 classes with μ bi-modal (e.g., $\mu = \frac{3}{4}\delta_{0.1} + \frac{1}{4}\delta_{0.5}$)

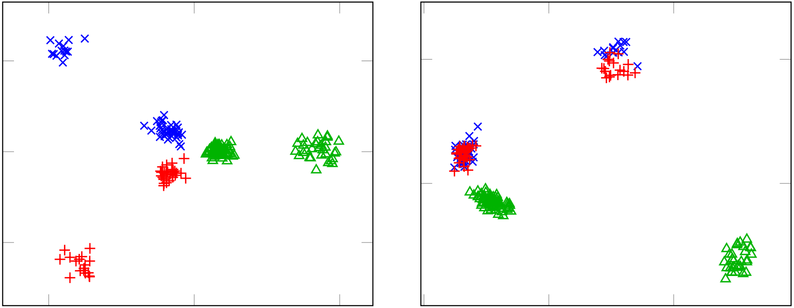
$\rightarrow$ Leading eigenvectors of A (or modularity $A - \frac{dd^{\mathsf{T}}}{2m}$) biased by q_i distribution.
$\rightarrow$ Similar behavior for Bethe Hessian.

Limitations of Adjacency/Modularity Approach

Scenario: 3 classes with μ bi-modal (e.g., $\mu = \frac{3}{4}\delta_{0.1} + \frac{1}{4}\delta_{0.5}$)

$\rightarrow$ Leading eigenvectors of A (or modularity $A - \frac{dd^{\mathsf{T}}}{2m}$) biased by q_i distribution.

$\rightarrow$ Similar behavior for Bethe Hessian.



(Modularity) (Bethe Hessian)

Connectivity Model: $P(i \sim j) = q_i q_j C_{ab}$ for $i \in \mathcal{C}_a$, $j \in \mathcal{C}_b$.

Dense Regime Assumptions: Non trivial regime when, as $n \to \infty$,

$$C_{ab} = 1 + \frac{M_{ab}}{\sqrt{n}}$$

with $M_{ab} = O(1)$ (fixed).

Regularized Modularity Approach

Connectivity Model: $P(i \sim j) = q_i q_j C_{ab}$ for $i \in \mathcal{C}_a$, $j \in \mathcal{C}_b$.

Dense Regime Assumptions: Non trivial regime when, as $n \to \infty$,

$$C_{ab} = 1 + \frac{M_{ab}}{\sqrt{n}}$$

with $M_{ab} = O(1)$ (fixed).

Considered Matrix:
For $\alpha \in [0, 1]$, (and with $D = \mathrm{diag}(A1_n) = \mathrm{diag}(d)$ the degree matrix)

$$L_\alpha = (2m)^\alpha \frac{1}{\sqrt{n}} D^{-\alpha} \left[A - \frac{dd^\mathsf{T}}{2m} \right] D^{-\alpha}.$$

Asymptotic Equivalence

Theorem (Limiting Random Matrix Equivalent)

For each $\alpha \in [0,1]$, *as* $n \to \infty$, $\|L_\alpha - \tilde{L}_\alpha\| \to 0$ *almost surely, where*

$$L_\alpha = (2m)^\alpha \frac{1}{\sqrt{n}} D^{-\alpha} \left[A - \frac{dd^\mathsf{T}}{2m} \right] D^{-\alpha}$$

$$\tilde{L}_\alpha = \frac{1}{\sqrt{n}} D_q^{-\alpha} X D_q^{-\alpha} + U \Lambda U^\mathsf{T}$$

with $D_q = \operatorname{diag}(\{q_i\})$, X *zero-mean random matrix,*

$$U = \left[D_q^{1-\alpha} \frac{J}{\sqrt{n}} \quad \frac{1}{1_n^\mathsf{T} D_q 1_n} D_q^{-\alpha} X 1_n \right], \quad \textit{rank } k+1$$

$$\Lambda = \begin{bmatrix} (I_k - 1_k c^\mathsf{T}) M (I_k - c 1_k^\mathsf{T}) & -1_k \\ 1_k^\mathsf{T} & 0 \end{bmatrix}$$

and $J = [j_1, \ldots, j_k]$, $j_a = [0, \ldots, 0, 1_{n_a}^\mathsf{T}, 0, \ldots, 0]^\mathsf{T} \in \mathbb{R}^n$ *canonical vector of class* $\mathcal{C}_a$.

Theorem (Limiting Random Matrix Equivalent)

For each $\alpha \in [0,1]$, *as* $n \to \infty$, $\|L_\alpha - \tilde{L}_\alpha\| \to 0$ *almost surely, where*

$$L_\alpha = (2m)^\alpha \frac{1}{\sqrt{n}} D^{-\alpha} \left[A - \frac{dd^\mathsf{T}}{2m} \right] D^{-\alpha}$$

$$\tilde{L}_\alpha = \frac{1}{\sqrt{n}} D_q^{-\alpha} X D_q^{-\alpha} + U \Lambda U^\mathsf{T}$$

with $D_q = \operatorname{diag}(\{q_i\})$, X *zero-mean random matrix,*

$$U = \begin{bmatrix} D_q^{1-\alpha} \frac{J}{\sqrt{n}} & \frac{1}{1_n^\mathsf{T} D_q 1_n} D_q^{-\alpha} X 1_n \end{bmatrix}, \quad \textit{rank } k+1$$

$$\Lambda = \begin{bmatrix} (I_k - 1_k c^\mathsf{T}) M (I_k - c 1_k^\mathsf{T}) & -1_k \\ 1_k^\mathsf{T} & 0 \end{bmatrix}$$

and $J = [j_1, \ldots, j_k]$, $j_a = [0, \ldots, 0, 1_{n_a}^\mathsf{T}, 0, \ldots, 0]^\mathsf{T} \in \mathbb{R}^n$ *canonical vector of class* $\mathcal{C}_a$.

Consequences:

- $\tilde{L}_\alpha$ is a well-known spiked random matrix

Theorem (Limiting Random Matrix Equivalent)

For each $\alpha \in [0,1]$, *as* $n \to \infty$, $\|L_\alpha - \tilde{L}_\alpha\| \to 0$ *almost surely, where*

$$L_\alpha = (2m)^\alpha \frac{1}{\sqrt{n}} D^{-\alpha} \left[A - \frac{dd^\mathsf{T}}{2m} \right] D^{-\alpha}$$

$$\tilde{L}_\alpha = \frac{1}{\sqrt{n}} D_q^{-\alpha} X D_q^{-\alpha} + U \Lambda U^\mathsf{T}$$

with $D_q = \operatorname{diag}(\{q_i\})$, X *zero-mean random matrix,*

$$U = \begin{bmatrix} D_q^{1-\alpha} \frac{J}{\sqrt{n}} & \frac{1}{1_n^\mathsf{T} D_q 1_n} D_q^{-\alpha} X 1_n \end{bmatrix}, \quad \textit{rank } k+1$$

$$\Lambda = \begin{bmatrix} (I_k - 1_k c^\mathsf{T}) M (I_k - c 1_k^\mathsf{T}) & -1_k \\ 1_k^\mathsf{T} & 0 \end{bmatrix}$$

and $J = [j_1, \ldots, j_k]$, $j_a = [0, \ldots, 0, 1_{n_a}^\mathsf{T}, 0, \ldots, 0]^\mathsf{T} \in \mathbb{R}^n$ *canonical vector of class* $\mathcal{C}_a$.

Consequences:

- $\tilde{L}_\alpha$ is a well-known spiked random matrix
- it is "easy" to study and leads to a full analysis of the spectral clustering performance!

Theorem (Limiting Random Matrix Equivalent)

For each $\alpha \in [0,1]$, *as* $n \to \infty$, $\|L_\alpha - \tilde{L}_\alpha\| \to 0$ *almost surely, where*

$$L_\alpha = (2m)^\alpha \frac{1}{\sqrt{n}} D^{-\alpha} \left[A - \frac{dd^\mathsf{T}}{2m} \right] D^{-\alpha}$$
$$\tilde{L}_\alpha = \frac{1}{\sqrt{n}} D_q^{-\alpha} X D_q^{-\alpha} + U \Lambda U^\mathsf{T}$$

with $D_q = \operatorname{diag}(\{q_i\})$, X *zero-mean random matrix,*

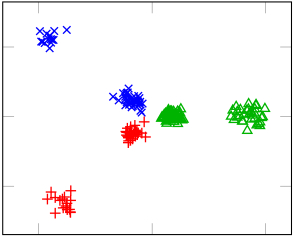
$$U = \begin{bmatrix} D_q^{1-\alpha} \frac{J}{\sqrt{n}} & \frac{1}{1_n^\mathsf{T} D_q 1_n} D_q^{-\alpha} X 1_n \end{bmatrix}, \quad \text{rank } k+1$$
$$\Lambda = \begin{bmatrix} (I_k - 1_k c^\mathsf{T}) M (I_k - c 1_k^\mathsf{T}) & -1_k \\ 1_k^\mathsf{T} & 0 \end{bmatrix}$$

and $J = [j_1, \ldots, j_k]$, $j_a = [0, \ldots, 0, 1_{n_a}^\mathsf{T}, 0, \ldots, 0]^\mathsf{T} \in \mathbb{R}^n$ *canonical vector of class* $\mathcal{C}_a$.

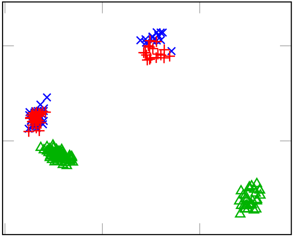
Consequences:

- $\tilde{L}_\alpha$ is a well-known spiked random matrix
- it is "easy" to study and leads to a full analysis of the spectral clustering performance!
- it helps us correct and optimize classical spectral clustering into a powerful **new algorithm**.

Performance Results (2 masses of q_i)

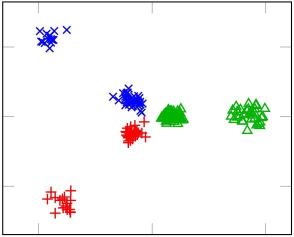


(Modularity)

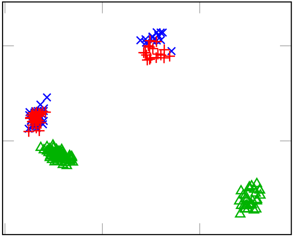


(Bethe Hessian)

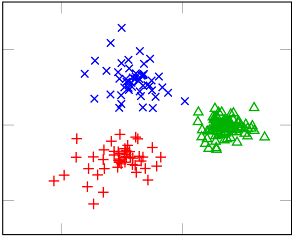
Performance Results (2 masses of q_i)



(Modularity)



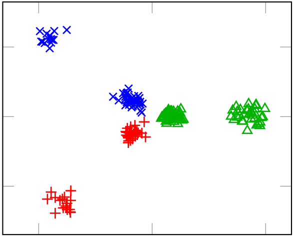
(Bethe Hessian)



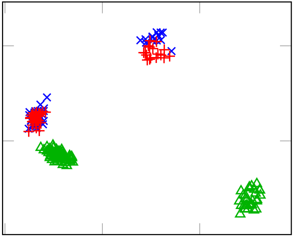
(Algo with $\alpha = 1$)

Figure: Two dominant eigenvectors (x-y axes) for $n = 2000$, $K = 3$, $\mu = \frac{3}{4}\delta_{q_1} + \frac{1}{4}\delta_{q_2}$, $q_1 = 0.1$, $q_2 = 0.5$, $c_1 = c_2 = \frac{1}{4}$, $c_3 = \frac{1}{2}$, $M = 100 I_3$.

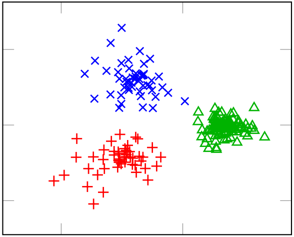
Performance Results (2 masses of q_i)



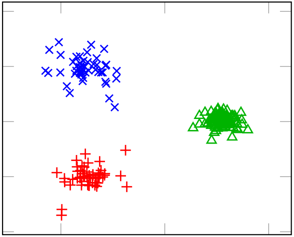
(Modularity)



(Bethe Hessian)



(Algo with $\alpha = 1$)



(Algo with α_{opt})

Figure: Two dominant eigenvectors (x-y axes) for $n = 2000$, $K = 3$, $\mu = \frac{3}{4}\delta_{q_1} + \frac{1}{4}\delta_{q_2}$, $q_1 = 0.1$, $q_2 = 0.5$, $c_1 = c_2 = \frac{1}{4}$, $c_3 = \frac{1}{2}$, $M = 100 I_3$.

Performance Results (2 masses for q_i)

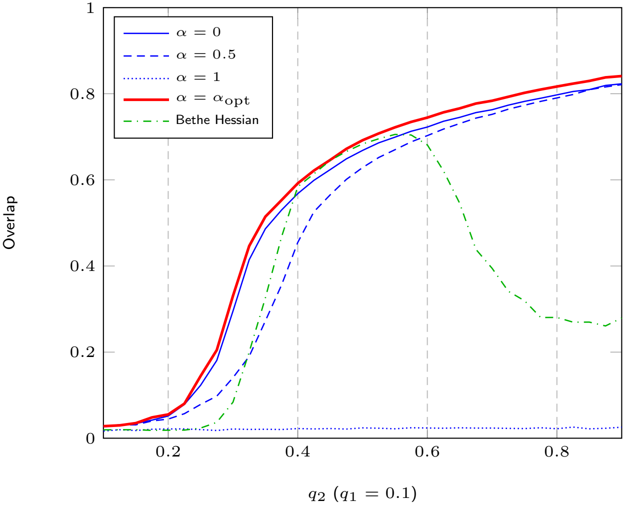


Figure: Overlap performance for $n = 3000$, $K = 3$, $\mu = \frac{3}{4}\delta_{q_1} + \frac{1}{4}\delta_{q_2}$ with $q_1 = 0.1$ and $q_2 \in [0.1, 0.9]$, $M = 10(2I_3 - 1_3 1_3^{\mathsf{T}})$, $c_i = \frac{1}{3}$.

Outline

Problem Statement

- Dataset $x_1, \ldots, x_n \in \mathbb{R}^p$
- Objective: "cluster" data in k similarity classes $\mathcal{S}_1, \ldots, \mathcal{S}_k$.

Problem Statement

- Dataset $x_1, \ldots, x_n \in \mathbb{R}^p$
- Objective: "cluster" data in k similarity classes $\mathcal{S}_1, \ldots, \mathcal{S}_k$.

- Typical metric to optimize:

$$\text{(RatioCut)}\ \ \operatorname{argmin}_{\mathcal{S}_1 \cup \ldots \cup \mathcal{S}_k = \{1,\ldots,n\}} \sum_{i=1}^{k} \sum_{\substack{j \in \mathcal{S}_i \\ \bar{j} \notin \mathcal{S}_i}} \frac{\kappa(x_j, x_{\bar{j}})}{|\mathcal{S}_i|}$$

for some similarity kernel $\kappa(x, y) \geq 0$ (large if x similar to y).

Kernel Spectral Clustering

Problem Statement

- Dataset $x_1, \ldots, x_n \in \mathbb{R}^p$
- Objective: "cluster" data in k similarity classes $\mathcal{S}_1, \ldots, \mathcal{S}_k$.

- Typical metric to optimize:

$$(\text{RatioCut}) \ \ \operatorname{argmin}_{\mathcal{S}_1 \cup \ldots \cup \mathcal{S}_k = \{1,\ldots,n\}} \sum_{i=1}^{k} \sum_{\substack{j \in \mathcal{S}_i \\ \bar{j} \notin \mathcal{S}_i}} \frac{\kappa(x_j, x_{\bar{j}})}{|\mathcal{S}_i|}$$

for some similarity kernel $\kappa(x,y) \geq 0$ (large if x similar to y).

- Can be shown equivalent to

$$(\text{RatioCut}) \ \ \operatorname{argmin}_{M \in \mathcal{M}} \operatorname{tr} M^{\mathsf{T}}(D-K)M$$

where $\mathcal{M} \subset \mathbb{R}^{n \times k} \cap \left\{M; \ M_{ij} \in \{0, |\mathcal{S}_j|^{-\frac{1}{2}}\}\right\}$ (in particular, $M^{\mathsf{T}}M = I_k$) and

$$K = \{\kappa(x_i, x_j)\}_{i,j=1}^{n}, \ D_{ii} = \sum_{j=1}^{n} K_{ij}.$$

Kernel Spectral Clustering

Problem Statement

- Dataset $x_1, \ldots, x_n \in \mathbb{R}^p$
- Objective: "cluster" data in k similarity classes $\mathcal{S}_1, \ldots, \mathcal{S}_k$.

- Typical metric to optimize:

$$(\text{RatioCut}) \;\; \text{argmin}_{\mathcal{S}_1 \cup \ldots \cup \mathcal{S}_k = \{1, \ldots, n\}} \sum_{i=1}^{k} \sum_{\substack{j \in \mathcal{S}_i \\ \bar{j} \notin \mathcal{S}_i}} \frac{\kappa(x_j, x_{\bar{j}})}{|\mathcal{S}_i|}$$

for some similarity kernel $\kappa(x, y) \geq 0$ (large if x similar to y).

- Can be shown equivalent to

$$(\text{RatioCut}) \;\; \text{argmin}_{M \in \mathcal{M}} \operatorname{tr} M^{\mathsf{T}}(D - K)M$$

where $\mathcal{M} \subset \mathbb{R}^{n \times k} \cap \left\{ M; \; M_{ij} \in \{0, |\mathcal{S}_j|^{-\frac{1}{2}}\} \right\}$ (in particular, $M^{\mathsf{T}} M = I_k$) and

$$K = \{\kappa(x_i, x_j)\}_{i,j=1}^{n}, \;\; D_{ii} = \sum_{j=1}^{n} K_{ij}.$$

- But integer problem! Usually NP-complete.

Towards kernel spectral clustering

- Kernel spectral clustering: discrete-to-continuous relaxations of such metrics

$$(\text{RatioCut}) \ \operatorname{argmin}_{M,\ M^\mathsf{T} M = I_K} \operatorname{tr} M^\mathsf{T}(D-K)M$$

i.e., eigenvector problem:

1. find eigenvectors of smallest eigenvalues
2. retrieve classes from eigenvector components

Towards kernel spectral clustering

- Kernel spectral clustering: discrete-to-continuous relaxations of such metrics

$$(\text{RatioCut})\ \ \operatorname{argmin}_{M,\ M^\mathsf{T} M=I_K} \operatorname{tr} M^\mathsf{T}(D-K)M$$

  i.e., eigenvector problem:

  1. find eigenvectors of smallest eigenvalues
  2. retrieve classes from eigenvector components

- Refinements:
  - working on K, $D-K$, $I_n - D^{-1}K$, $I_n - D^{-\frac{1}{2}}KD^{-\frac{1}{2}}$, etc.
  - several steps algorithms: Ng–Jordan–Weiss, Shi–Malik, etc.

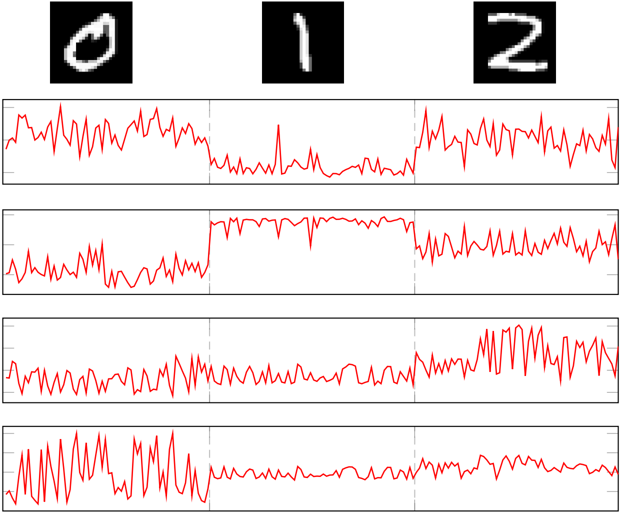


Figure: Leading four eigenvectors of $D^{-\frac{1}{2}} K D^{-\frac{1}{2}}$ for MNIST data.

Objectives and Roadmap:

- Develop mathematical analysis framework for BigData kernel spectral clustering ($p, n \to \infty$)

Objectives and Roadmap:

- Develop mathematical analysis framework for BigData kernel spectral clustering $(p, n \to \infty)$
- Understand:
  1. Phase transition effects (i.e., when is clustering possible?)
  2. Content of each eigenvector
  3. Influence of kernel function
  4. Performance comparison of clustering algorithms

Methodology and objectives

Objectives and Roadmap:

- Develop mathematical analysis framework for BigData kernel spectral clustering ($p, n \to \infty$)
- Understand:
  1. Phase transition effects (i.e., when is clustering possible?)
  2. Content of each eigenvector
  3. Influence of kernel function
  4. Performance comparison of clustering algorithms

Methodology:

- Use statistical assumptions (Gaussian mixture)
- Benefit from doubly-infinite independence and random matrix tools

Model and Assumptions

Gaussian mixture model:

- $x_1, \ldots, x_n \in \mathbb{R}^p$,
- k classes $\mathcal{C}_1, \ldots, \mathcal{C}_k$,
- $x \in \mathcal{C}_a \Leftrightarrow x \sim \mathcal{N}(\mu_a, C_a)$.

Model and Assumptions

Gaussian mixture model:

- $x_1, \ldots, x_n \in \mathbb{R}^p$,
- k classes $\mathcal{C}_1, \ldots, \mathcal{C}_k$,
- $x \in \mathcal{C}_a \Leftrightarrow x \sim \mathcal{N}(\mu_a, C_a)$.

Kernel Matrix:

- Kernel matrix of interest:

$$K = \left\{ f\left(\frac{1}{p} \|x_i - x_j\|^2 \right) \right\}_{i,j=1}^{n}$$

for some sufficiently smooth nonnegative f.

Model and Assumptions

Gaussian mixture model:

- $x_1, \ldots, x_n \in \mathbb{R}^p$,
- k classes $\mathcal{C}_1, \ldots, \mathcal{C}_k$,
- $x \in \mathcal{C}_a \Leftrightarrow x \sim \mathcal{N}(\mu_a, C_a)$.

Kernel Matrix:

- Kernel matrix of interest:

$$K = \left\{ f\left(\frac{1}{p}\|x_i - x_j\|^2\right) \right\}_{i,j=1}^{n}$$

  for some sufficiently smooth nonnegative f.

- We study the normalized Laplacian:

$$L = nD^{-\frac{1}{2}} K D^{-\frac{1}{2}}$$

  with $D = \mathrm{diag}(K 1_n)$.

Model and Assumptions

Difficulty: L is a very intractable random matrix

- non-linear f
- non-trivial dependence between entries of L

Model and Assumptions

Difficulty: L is a very intractable random matrix

- non-linear f
- non-trivial dependence between entries of L

Strategy:

1. Find random equivalent $\hat{L}$ (i.e., $\|L - \hat{L}\| \xrightarrow{\text{a.s.}} 0$ as $n, p \to \infty$) based on:
   - concentration: $K_{ij} \to \tau$, constant, as $n, p \to \infty$ (for all $i \neq j$)
   - Taylor expansion around limit point

Difficulty: L is a very intractable random matrix

- non-linear f
- non-trivial dependence between entries of L

Strategy:

1. Find random equivalent $\hat{L}$ (i.e., $\|L - \hat{L}\| \xrightarrow{\text{a.s.}} 0$ as $n, p \to \infty$) based on:
   - concentration: $K_{ij} \to \tau$, constant, as $n, p \to \infty$ (for all $i \neq j$)
   - Taylor expansion around limit point
2. Apply spiked random matrix approach to study:
   - existence of isolated eigenvalues in $\hat{L}$: phase transition

Model and Assumptions

Difficulty: L is a very intractable random matrix

- non-linear f
- non-trivial dependence between entries of L

Strategy:

1. Find random equivalent $\hat{L}$ (i.e., $\|L - \hat{L}\| \xrightarrow{\text{a.s.}} 0$ as $n, p \to \infty$) based on:
   - concentration: $K_{ij} \to \tau$, constant, as $n, p \to \infty$ (for all $i \neq j$)
   - Taylor expansion around limit point
2. Apply spiked random matrix approach to study:
   - existence of isolated eigenvalues in $\hat{L}$: phase transition
   - eigenvector projections on canonical class-basis

Theoretical Findings versus MNIST

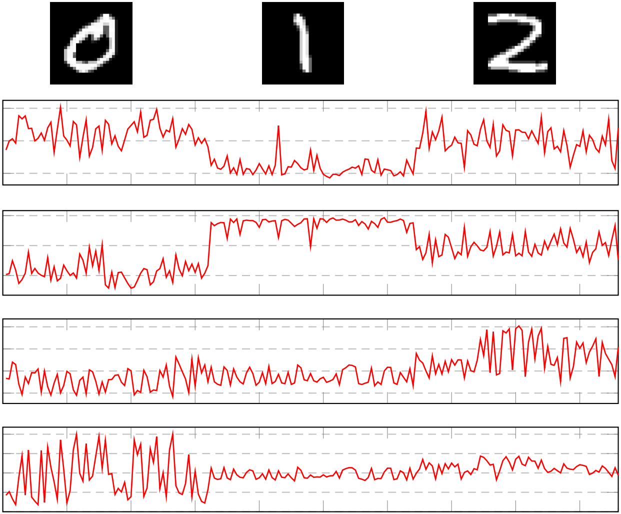


Figure: Leading four eigenvectors of $D^{-\frac{1}{2}}KD^{-\frac{1}{2}}$ for MNIST data (**red**), versus Gaussian equivalent model (**black**), and theoretical findings (**blue**).

Theoretical Findings versus MNIST

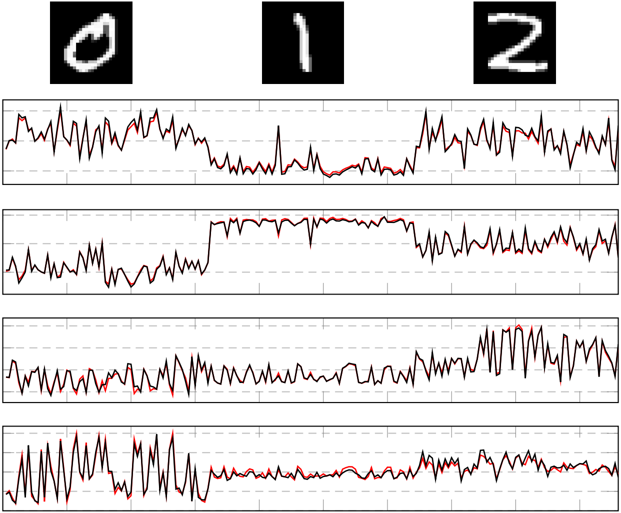


Figure: Leading four eigenvectors of $D^{-\frac{1}{2}}KD^{-\frac{1}{2}}$ for MNIST data (**red**), versus Gaussian equivalent model (**black**), and theoretical findings (**blue**).

Theoretical Findings versus MNIST

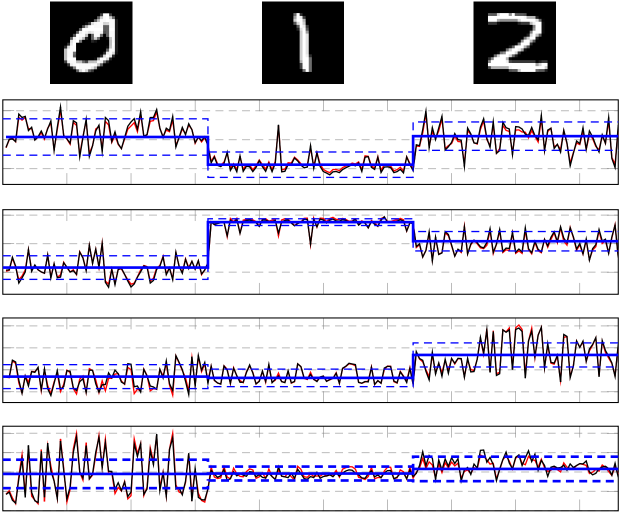


Figure: Leading four eigenvectors of $D^{-\frac{1}{2}}KD^{-\frac{1}{2}}$ for MNIST data (**red**), versus Gaussian equivalent model (**black**), and theoretical findings (**blue**).

Theoretical Findings versus MNIST

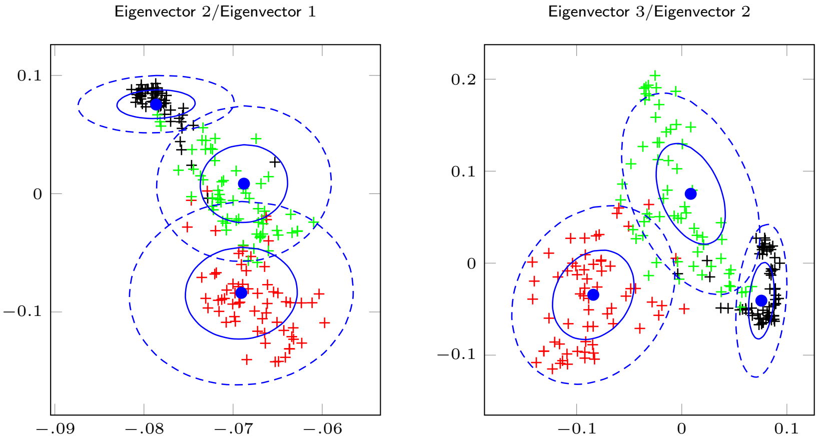


Figure: 2D representation of eigenvectors of L, for the MNIST dataset. Theoretical means and 1- and 2-standard deviations in **blue**. Class 1 in **red**, Class 2 in **black**, Class 3 in **green**.

Outline

Other Results and Perspectives

Objectives:

- **Kernel methods.**
  - ✔ Subspace spectral clustering (dramatically different case of $f'(\tau) = 0$)
  - ✎ Spectral clustering with outer product kernel $f(x^{\mathsf{T}} y)$
  - ✎ Semi-supervised learning, kernel approaches.
  - ✎ Support vector machines (SVM).

Other Results and Perspectives

Objectives:

- **Kernel methods**.
  - ✔ Subspace spectral clustering (dramatically different case of $f'(\tau) = 0$)
  - ✎ Spectral clustering with outer product kernel $f(x^{\mathsf{T}} y)$
  - ✎ Semi-supervised learning, kernel approaches.
  - ✎ Support vector machines (SVM).
- **Community detection**.
  - ✔ Complete study of eigenvector contents in adjacency/modularity methods.
  - Study of Bethe Hessian approach.
  - Analysis of non-necessarily spectral approaches (wavelet approaches).

Other Results and Perspectives

Objectives:

- **Kernel methods.**
  - ✔ Subspace spectral clustering (dramatically different case of $f'(\tau) = 0$)
  - ✎ Spectral clustering with outer product kernel $f(x^{\mathsf{T}} y)$
  - ✎ Semi-supervised learning, kernel approaches.
  - ✎ Support vector machines (SVM).
- **Community detection.**
  - ✔ Complete study of eigenvector contents in adjacency/modularity methods.
  - 💡 Study of Bethe Hessian approach.
  - 💡 Analysis of non-necessarily spectral approaches (wavelet approaches).
- **Neural Networks.**
  - ✎ Analysis of non-linear extreme learning machines
  - 💡 non-linear echo-state

Other Results and Perspectives

Objectives:

- **Kernel methods**.
  - ✔ Subspace spectral clustering (dramatically different case of $f'(\tau) = 0$)
  - ✎ Spectral clustering with outer product kernel $f(x^{\mathsf{T}}y)$
  - ✎ Semi-supervised learning, kernel approaches.
  - ✎ Support vector machines (SVM).
- **Community detection**.
  - ✔ Complete study of eigenvector contents in adjacency/modularity methods.
  - 💡 Study of Bethe Hessian approach.
  - 💡 Analysis of non-necessarily spectral approaches (wavelet approaches).
- **Neural Networks**.
  - ✎ Analysis of non-linear extreme learning machines
  - 💡 non-linear echo-state
- **Signal processing on graphs, further graph inference, etc.**
  - 💡 Making graph methods random.

Thank you.