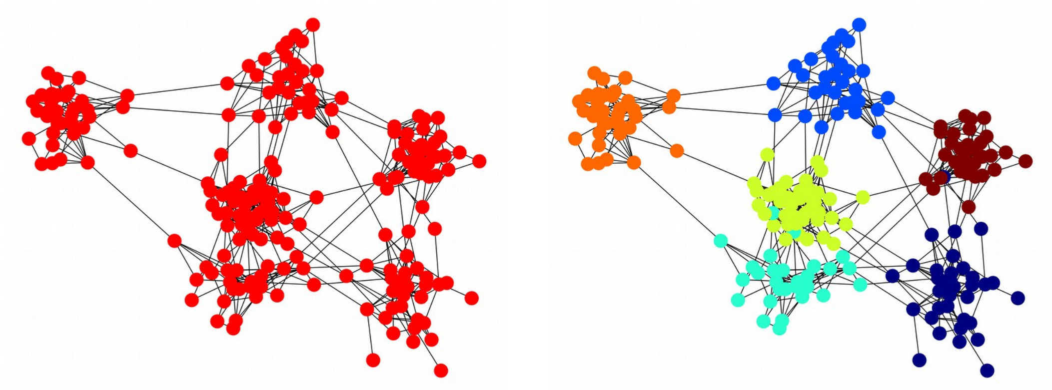
Sparse and Smooth:
Spectral Clustering in the dynamic SBM

Nicolas Keriven[1], Samuel Vaiter[2]

[1]CNRS, GIPSA-lab [2]CNRS, IMB

nicolas.keriven@gipsa-lab.grenoble-inp.fr

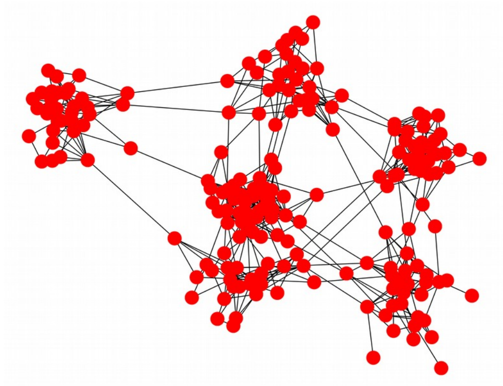
Spectral Clustering

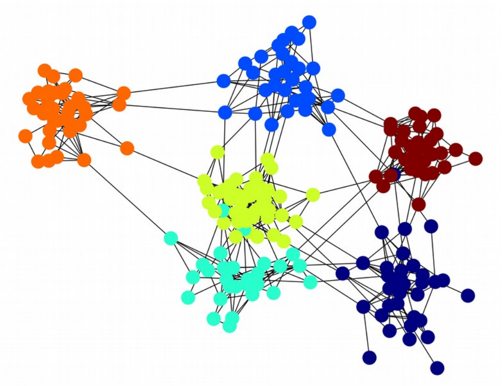


Cluster the nodes of a graph using its structure.

Application in:
- Social network analysis
- Point cloud segmentation
- Etc, etc

Spectral Clustering





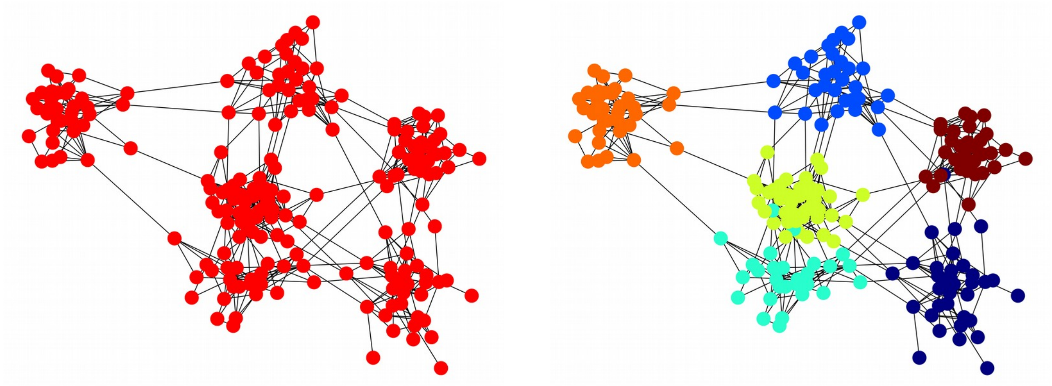
Cluster the nodes of a graph using its structure.

Application in:

- Social network analysis
- Point cloud segmentation
- Etc, etc

Classical algorithm:

Spectral Clustering



Cluster the nodes of a graph using its structure.

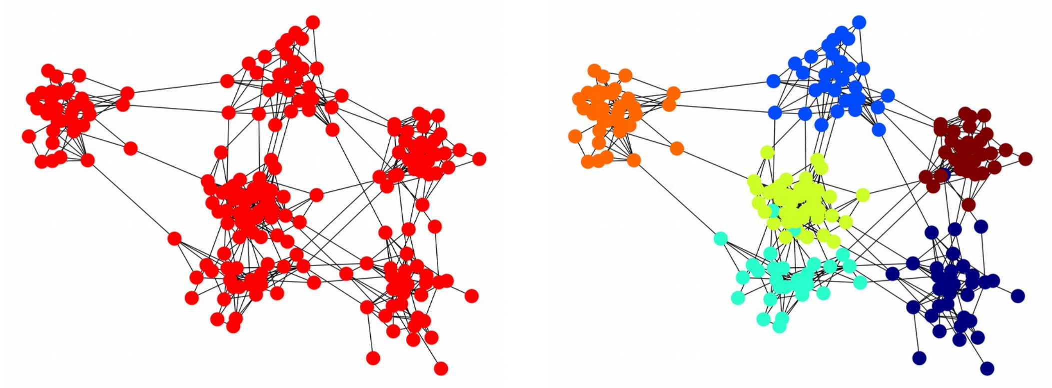
Application in:

- Social network analysis
- Point cloud segmentation
- Etc, etc

Classical algorithm:

- Take $W = \begin{cases} A \\ D - A \\ Id - D^{-1/2} A D^{-1/2} \end{cases}$

Spectral Clustering



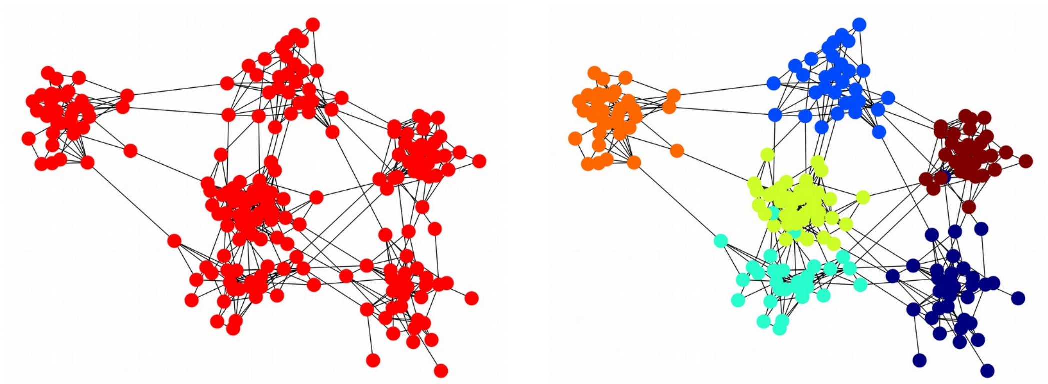
Cluster the nodes of a graph using its structure.

Application in:
- Social network analysis
- Point cloud segmentation
- Etc, etc

Classical algorithm:

- Take $W = \begin{cases} A \\ D - A \\ Id - D^{-1/2} A D^{-1/2} \end{cases}$
- Compute its k-SVD $W = U \Delta U^\top, \quad \tilde{U} = U_{:,1:k}$

Spectral Clustering



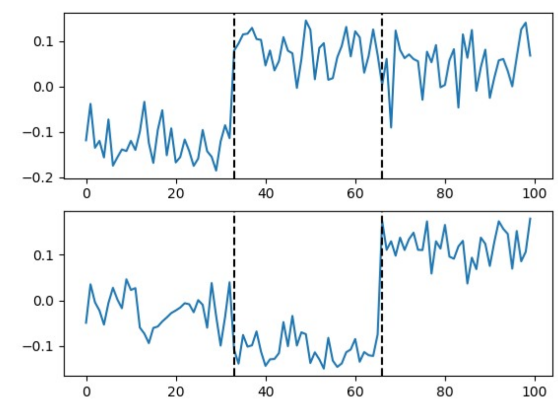
Cluster the nodes of a graph using its structure.

Application in:
- Social network analysis
- Point cloud segmentation
- Etc, etc

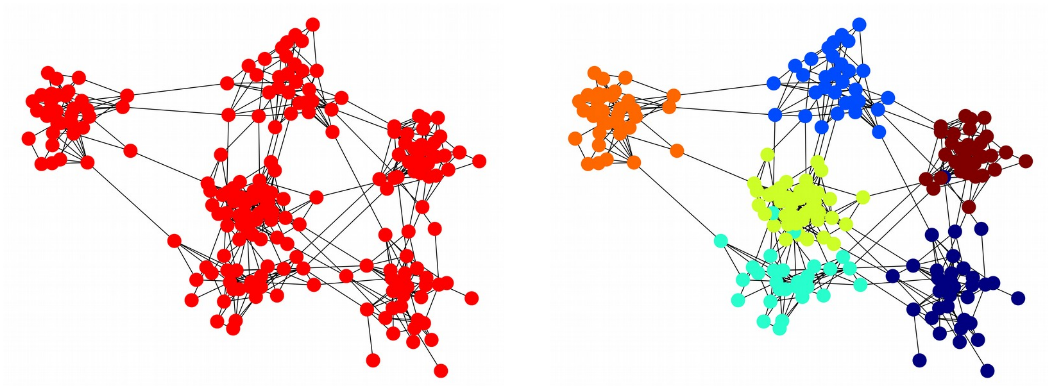
Classical algorithm:

- Take $W = \begin{cases} A \\ D - A \\ Id - D^{-1/2} A D^{-1/2} \end{cases}$
- Compute its k-SVD $W = U \Delta U^{\top}, \quad \tilde{U} = U_{:,1:k}$
- Cluster the rows of $\tilde{U}$ with k-means

Eigenvectors :



Spectral Clustering



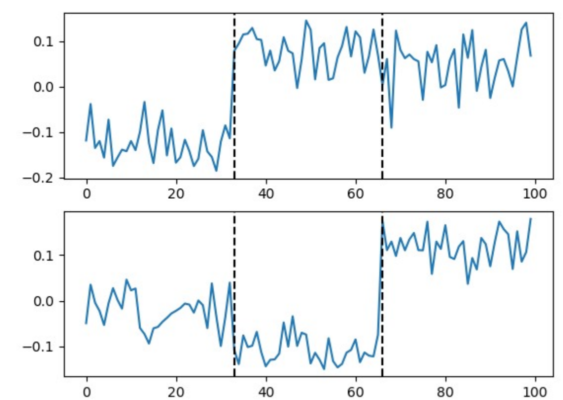
Cluster the nodes of a graph using its structure.

Application in:
- Social network analysis
- Point cloud segmentation
- Etc, etc

Classical algorithm:

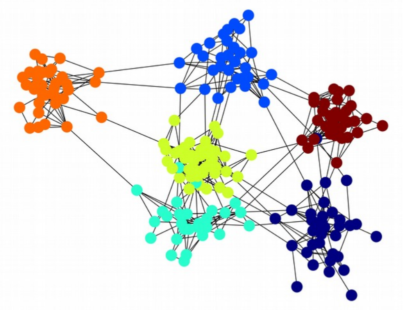
- Take $W = \begin{cases} A \\ D - A \\ Id - D^{-1/2}AD^{-1/2} \end{cases}$
- Compute its k-SVD $W = U\Delta U^\top, \quad \tilde{U} = U_{:,1:k}$
- Cluster the rows of $\tilde{U}$ with k-means
- Many many (fast) variants...

Eigenvectors :



SC and SBM

Stochastic Block Model (SBM)



$\Theta \in \{0,1\}^{n \times K}$ $\Theta_{ik} = 1$ *if node* i *belongs to comm.* k

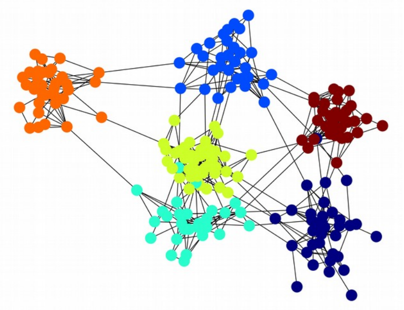
$B \in [0,1]^{K \times K}$ *Symmetric probability matrix*

$$\begin{cases} a_{ij} \sim \mathrm{Ber}(B_{k\ell}) \\ \text{when } \Theta_{ik} = \Theta_{j\ell} = 1 \end{cases}$$

Independent edges

SC and SBM

Stochastic Block Model (SBM)



$\Theta \in \{0,1\}^{n \times K}$ $\quad \Theta_{ik} = 1$ *if node* i *belongs to comm.* k

$B \in [0,1]^{K \times K}$ $\quad$ *Symmetric probability matrix*

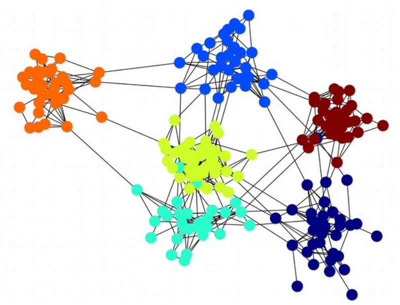
$$\begin{cases} a_{ij} \sim \mathrm{Ber}(B_{k\ell}) \\ \text{when } \Theta_{ik} = \Theta_{j\ell} = 1 \end{cases}$$

Independent edges

Sparsity = *density of edges*

SC and SBM

Stochastic Block Model (SBM)



$\Theta \in \{0,1\}^{n\times K}$ $\Theta_{ik} = 1$ *if node* i *belongs to comm.* k

$B \in [0,1]^{K\times K}$ *Symmetric probability matrix*

$$\begin{cases} a_{ij} \sim \mathrm{Ber}(B_{k\ell}) \\ \text{when } \Theta_{ik} = \Theta_{j\ell} = 1 \end{cases}$$

Independent edges

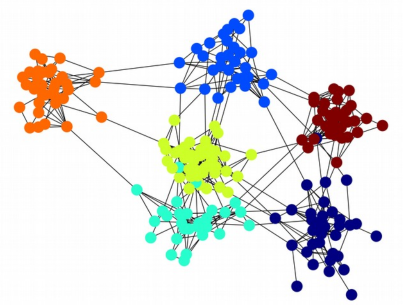
Sparsity = *density of edges*

$$B = \alpha_n B^0$$

Often $B^0 = (1-\tau) Id + \tau 1_{n\times n}$

SC and SBM

Stochastic Block Model (SBM)



$\Theta \in \{0,1\}^{n\times K}$ $\Theta_{ik} = 1$ *if node* i *belongs to comm.* k

$B \in [0,1]^{K\times K}$ *Symmetric probability matrix*

$$\begin{cases} a_{ij} \sim \mathrm{Ber}(B_{k\ell}) \\ \text{when } \Theta_{ik} = \Theta_{j\ell} = 1 \end{cases}$$

Independent edges

Sparsity = *density of edges*

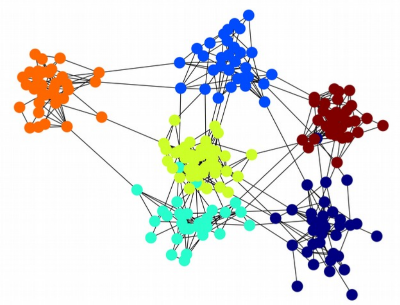
$$B = \alpha_n B^0$$

Often $B^0 = (1-\tau)Id + \tau 1_{n\times n}$

- **Dense** $\alpha_n \sim 1$
  - Easy
- **Sparse** $\alpha_n \sim 1/n$
  - Hard *(some **asymptotic** results from statistical physics)* *[Krzakala, Mossel, Massoulié, Abbe...]*
- ***Relatively* sparse** $\alpha_n \sim \log n / n$
  - easier...

SC and SBM

Stochastic Block Model (SBM)



$\Theta \in \{0,1\}^{n\times K}$ $\quad\Theta_{ik}=1$ *if node* i *belongs to comm.* k

$B \in [0,1]^{K\times K}$ $\quad$ *Symmetric probability matrix*

$$\begin{cases} a_{ij} \sim \mathrm{Ber}(B_{k\ell}) \\ \text{when } \Theta_{ik}=\Theta_{j\ell}=1 \end{cases}$$

Independent edges

Sparsity = *density of edges*

$$B = \alpha_n B^0$$

Often $B^0 = (1-\tau)Id + \tau 1_{n\times n}$

- **Dense** $\quad \alpha_n \sim 1$
  - Easy
- **Sparse** $\quad \alpha_n \sim 1/n$
  - Hard *(some **asymptotic** results from statistical physics)* *[Krzakala, Mossel, Massoulié, Abbe...]*
- ***Relatively* sparse** $\quad \alpha_n \sim \log n / n$
  - easier...

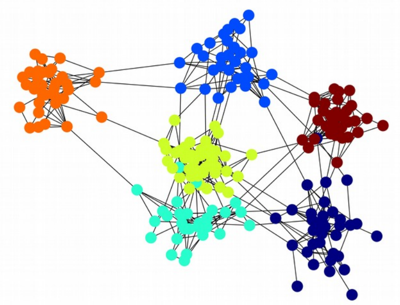
Non-asymptotic analysis

[Lei 2015]

- SC with $W = A$
- Relatively sparse: $\alpha_n \gtrsim \dfrac{\log n}{n}$

SC and SBM

Stochastic Block Model (SBM)



$\Theta \in \{0,1\}^{n\times K}$ $\Theta_{ik}=1$ *if node* i *belongs to comm.* k

$B \in [0,1]^{K\times K}$ *Symmetric probability matrix*

$$\begin{cases} a_{ij} \sim \mathrm{Ber}(B_{k\ell}) & \textit{Independent edges} \\ \text{when } \Theta_{ik}=\Theta_{j\ell}=1 \end{cases}$$

Sparsity = *density of edges*

$$B = \alpha_n B^0$$

Often $B^0 = (1-\tau)Id + \tau 1_{n\times n}$

- **Dense** $\alpha_n \sim 1$
  - Easy
- **Sparse** $\alpha_n \sim 1/n$
  - Hard *(some **asymptotic** results from statistical physics)* *[Krzakala, Mossel, Massoulié, Abbe...]*
- ***Relatively* sparse** $\alpha_n \sim \log n/n$
  - easier...

Non-asymptotic analysis

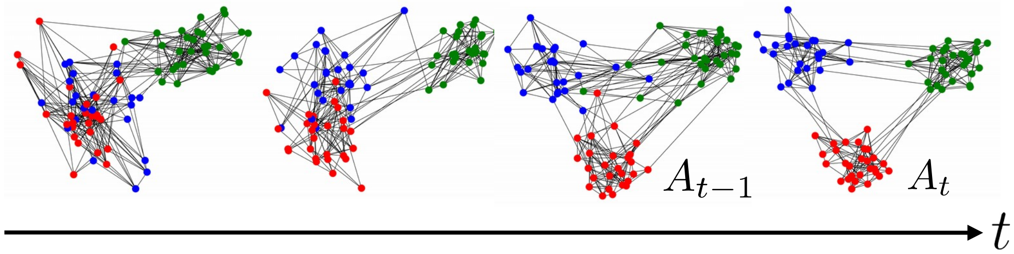
[Lei 2015]

- SC with $W = A$
- Relatively sparse: $\alpha_n \gtrsim \frac{\log n}{n}$
- With proba $1-n^{-r}$:

$$L(\hat\Theta,\Theta) \leq C_r \frac{K^2}{n\alpha_n(1-\tau)^2}$$

$$L(\hat\Theta,\Theta) = n^{-1}\min_\sigma \|\hat\Theta\sigma - \Theta\|_0$$

SC and DSBM

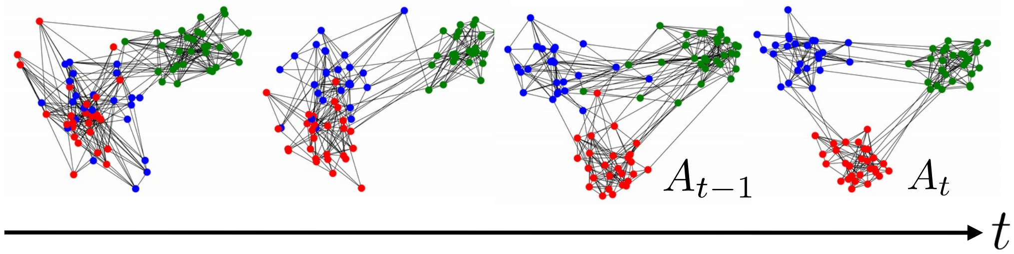


Goal: Exploit past data to

- Track communities
- Enforce smoothness/consistency
- **Improve result at time t**

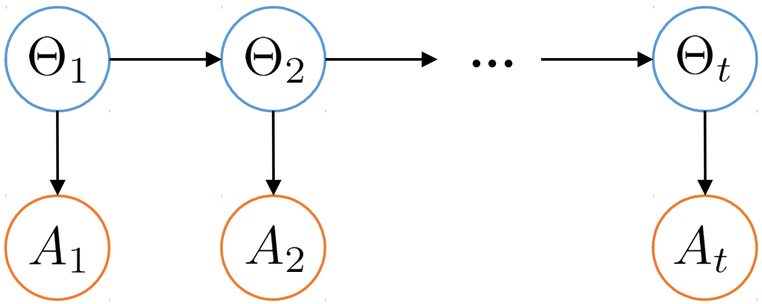
Many approaches (Bayesian, variational…)

SC and DSBM



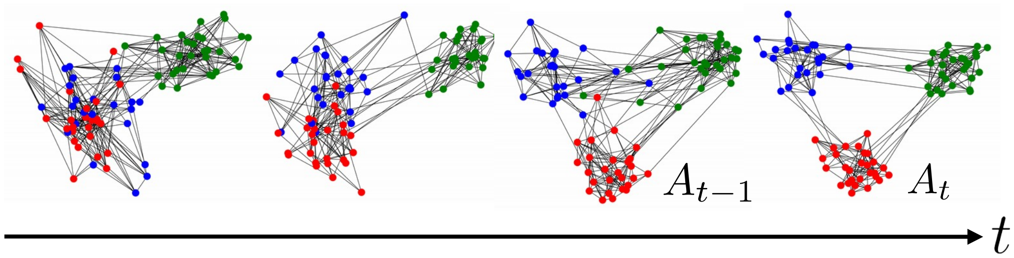
Goal: Exploit past data to
- Track communities
- Enforce smoothness/consistency
- **Improve result at time t**

Many approaches (Bayesian, variational…)



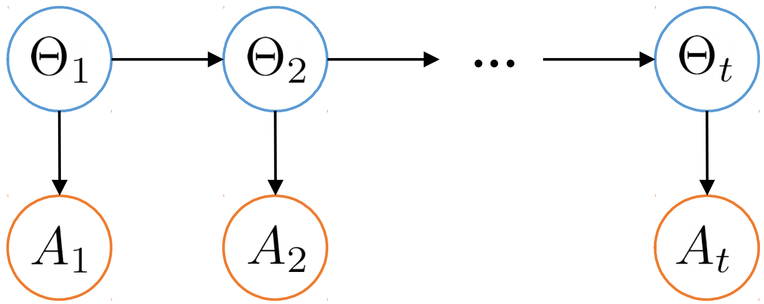
HMM
- At each time step, each node change community with proba ε_n **(smoothness)**
- To simplify, B does not change

SC and DSBM



Goal: Exploit past data to
- Track communities
- Enforce smoothness/consistency
- **Improve result at time t**

Many approaches (Bayesian, variational...)

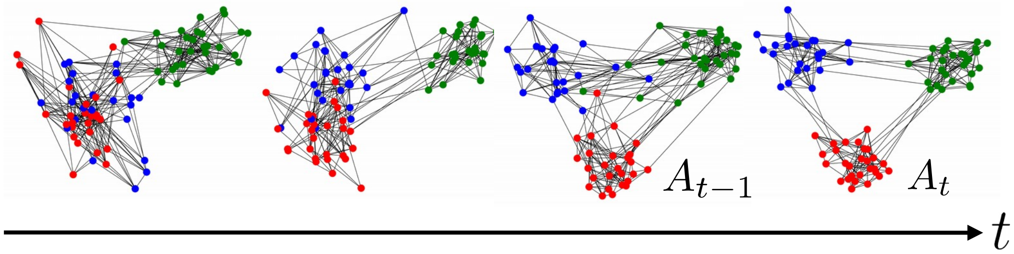


HMM
- At each time step, each node change community with proba ε_n **(smoothness)**
- To simplify, B does not change

Smoothed SC
[Pensky 2019]

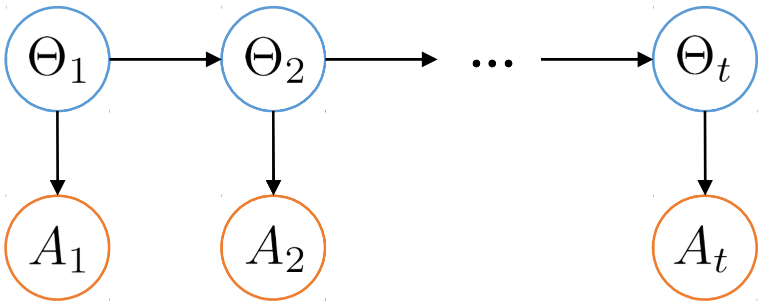
- SC with $W = \frac{1}{r}\sum_{k=0}^{r-1} A_{t-k}$
- Relatively sparse: $\alpha_n \gtrsim \frac{\log n}{n}$

SC and DSBM



Goal: Exploit past data to
- Track communities
- Enforce smoothness/consistency
- **Improve result at time t**

Many approaches (Bayesian, variational…)



HMM

- At each time step, each node change community with proba ε_n **(smoothness)**
- To simplify, B does not change

Smoothed SC

[Pensky 2019]

- SC with $W = \frac{1}{r}\sum_{k=0}^{r-1} A_{t-k}$
- Relatively sparse: $\alpha_n \gtrsim \frac{\log n}{n}$

- Choose $r \sim \rho_n^{-1}$ with

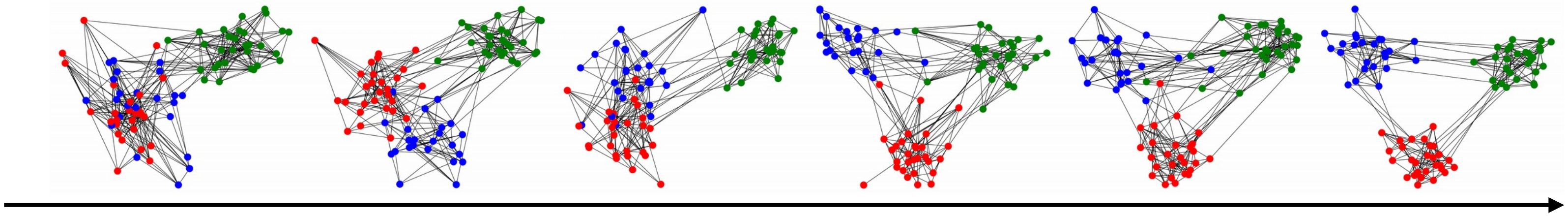
$$\rho_n = \min(1, \sqrt{\alpha_n n \varepsilon_n})$$

- With proba $1 - n^{-r}$:

$$L(\hat{\Theta}, \Theta) \lesssim C_r \frac{K^2}{n\alpha_n(1-\tau)^2}\rho_n$$

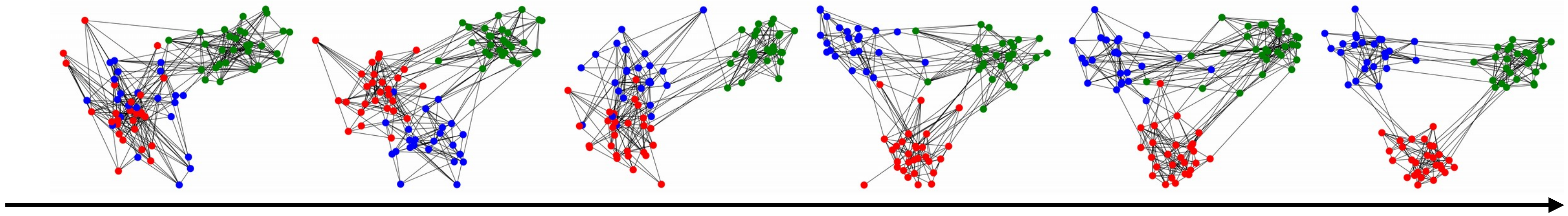
Better if ***smooth enough*** $\varepsilon_n = o(1/(\alpha_n n))$

Sparse and smooth...?



Sparse and smooth: intuitively, the smoother the data, the sparser it could be...

Sparse and smooth...?

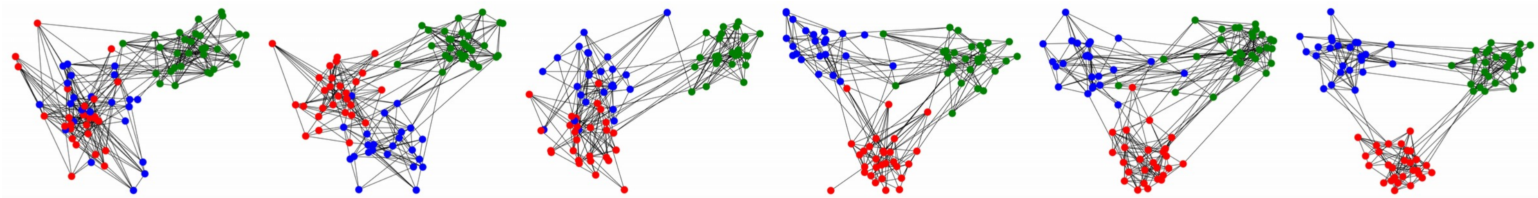


Sparse and smooth: intuitively, the smoother the data, the sparser it could be...

Thm *(Keriven, Vaiter)*

- SC with $W = \sum_k \beta_k A_{t-k}$

Sparse and smooth...?



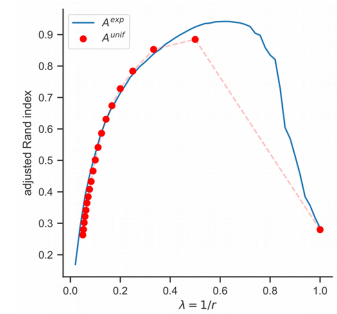
Sparse and smooth: intuitively, the smoother the data, the sparser it could be...

Thm *(Keriven, Vaiter)*

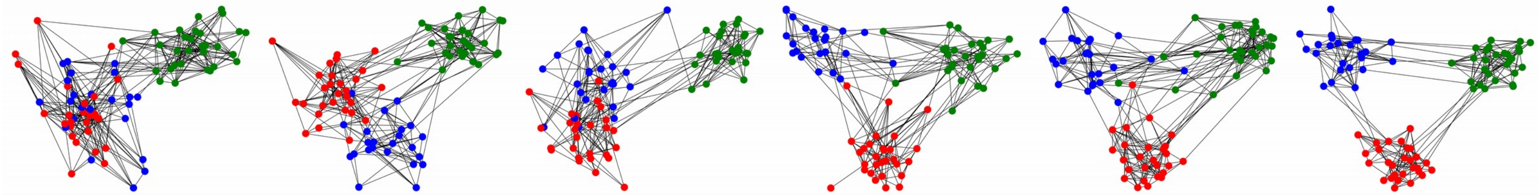
- SC with $W = \sum_k \beta_k A_{t-k}$

- Uniform *[Pensky]* $r^{-1} \sum_{k=0}^{r-1} A_{t-k}$
- Exponential *[Chi, Xu]* $W_t = (1-\lambda) W_{t-1} + \lambda A_t$

Exponential performs better in practice, and λ *is easier to choose than* r



Sparse and smooth...?



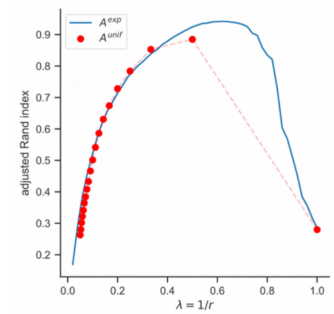
Sparse and smooth: intuitively, the smoother the data, the sparser it could be...

Thm *(Keriven, Vaiter)*

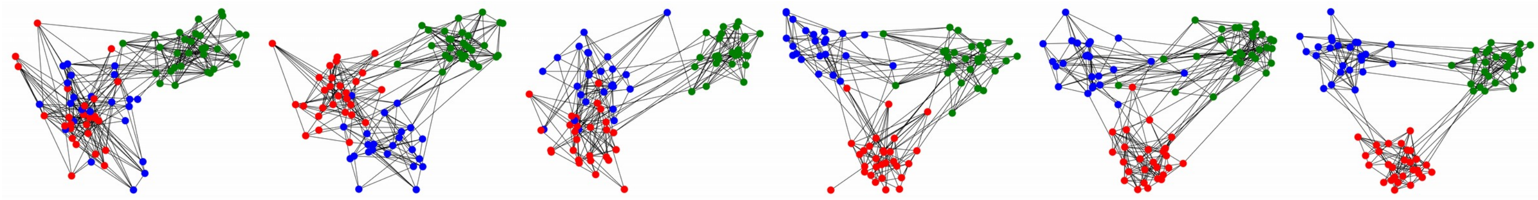
- SC with $W = \sum_k \beta_k A_{t-k}$
  - Uniform *[Pensky]* $r^{-1} \sum_{k=0}^{r-1} A_{t-k}$
  - Exponential *[Chi, Xu]* $W_t = (1-\lambda) W_{t-1} + \lambda A_t$

- **Improved** sparsity: $\frac{\alpha_n}{\rho_n} \gtrsim \frac{\log n}{n}$

Exponential performs better in practice, and λ is easier to choose than r



Sparse and smooth...?

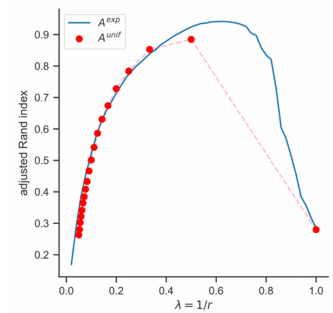


Sparse and smooth: intuitively, the smoother the data, the sparser it could be...

Thm *(Keriven, Vaiter)*

- SC with $W = \sum_k \beta_k A_{t-k}$
  - Uniform *[Pensky]* $r^{-1} \sum_{k=0}^{r-1} A_{t-k}$
  - Exponential *[Chi, Xu]* $W_t = (1-\lambda) W_{t-1} + \lambda A_t$

Exponential performs better in practice, and λ is easier to choose than r

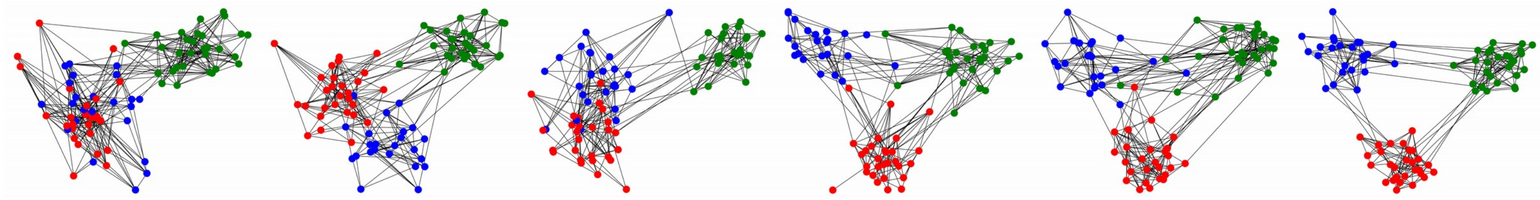


- **Improved** sparsity: $\dfrac{\alpha_n}{\rho_n} \gtrsim \dfrac{\log n}{n}$

Can handle the **sparse case** $\alpha_n \sim \frac{1}{n}$ when $\varepsilon_n \sim o\left(\frac{1}{\log^2 n}\right)$!!

Remember we already had $\varepsilon_n \sim o\left(\frac{1}{n\alpha_n}\right)$

Sparse and smooth...?



Sparse and smooth: intuitively, the smoother the data, the sparser it could be...

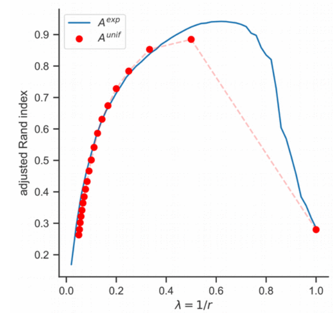
Thm *(Keriven, Vaiter)*

- SC with $W = \sum_k \beta_k A_{t-k}$
- **Improved** sparsity: $\frac{\alpha_n}{\rho_n} \gtrsim \frac{\log n}{n}$
- With proba $1 - n^{-r}$:

$$L(\hat{\Theta}, \Theta) \lesssim C_r \frac{K^2}{n\alpha_n(1-\tau)^2} \rho_n$$

- Uniform *[Pensky]* $r^{-1} \sum_{k=0}^{r-1} A_{t-k}$
- Exponential *[Chi, Xu]* $W_t = (1-\lambda)W_{t-1} + \lambda A_t$

Exponential performs better in practice, and λ is easier to choose than r



Can handle the **sparse case** $\alpha_n \sim \frac{1}{n}$ when $\varepsilon_n \sim o\left(\frac{1}{\log^2 n}\right)$!!

Remember we already had $\varepsilon_n \sim o\left(\frac{1}{n\alpha_n}\right)$

Normalized Laplacian

In practice, the *normalized Laplacian* works better. $L(A) = D(A)^{-1/2} A D(A)^{-1/2}$

- No real explanation (some hints…)

- Here we show that **it's not worse**, but require (very slightly) stronger hypothesis

    - *The proof has interesting by-products...*

Normalized Laplacian

In practice, the *normalized Laplacian* works better. $L(A) = D(A)^{-1/2} A D(A)^{-1/2}$

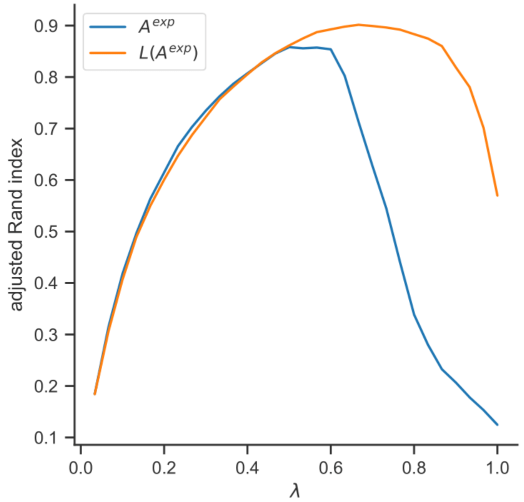
- No real explanation (some hints…)
- Here we show that **it's not worse**, but require (very slightly) stronger hypothesis
  - *The proof has interesting by-products...*

Thm *(Keriven, Vaiter)*

- SC with $W = L\left(\sum_k \beta_k A_{t-k}\right)$
- Sparsity: $\frac{\alpha_n}{\rho_n} \geq C'_r \frac{\log n}{n}$
- With proba $1 - n^{-r}$:

$$L(\hat\Theta, \Theta) \lesssim C_r \frac{K^2}{n\alpha_n(1-\tau)^2} \rho_n$$

The multiplicative constant is lower-bounded !

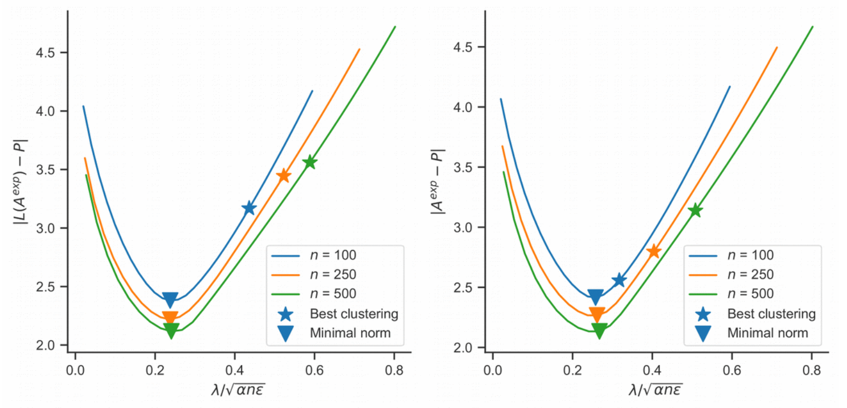


Sketch of sketch of proof

Proof is based on **spectral norm concentration**

$$\|A - \mathbb{E}A\| \qquad \|L(A) - L(\mathbb{E}A)\|$$

+ *classical "Davis-Kahan"-based perturbation analysis*
+ *almost-optimal k-means*



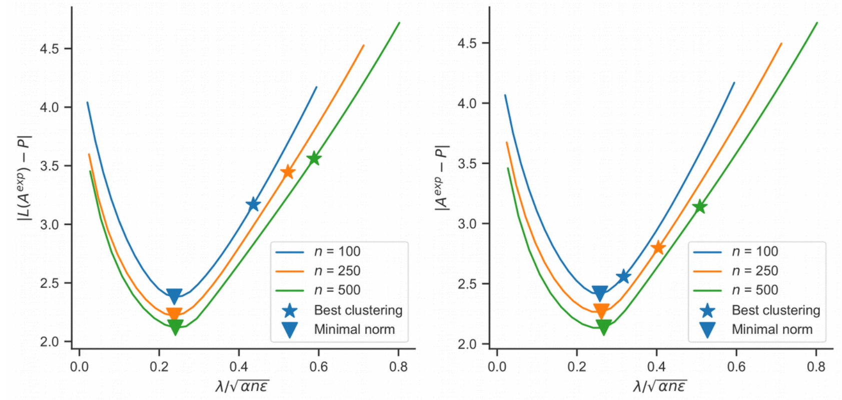
May not be the best criterion...

Sketch of sketch of proof

Proof is based on **spectral norm concentration**

$$\|A - \mathbb{E}A\| \qquad \|L(A) - L(\mathbb{E}A)\|$$

+ classical "Davis-Kahan"-based perturbation analysis
+ almost-optimal k-means



May not be the best criterion...

*Summary (valid for **any matrix with Bernoulli entries**)*

	Static		Dynamic		
	Adjacency	Laplacian	Adjacency	Laplacian	Sparsity
Oliveira 2009	$\log n$	1			$\alpha_n \gtrsim n^{-1} \log n$
Bandeira 2016	$\sqrt{\log n}$				$\alpha_n \gtrsim n^{-1}$
Lei 2015	$\sqrt{\alpha_n n}$				$\alpha_n \gtrsim n^{-1} \log n$
Pensky 2019	$\sqrt{\alpha_n n}$		$\sqrt{\alpha_n n \rho_n}$		$\alpha_n \gtrsim n^{-1} \log n$
Us	$\sqrt{\alpha_n n}$	$\sqrt{1/(\alpha_n n)}$	$\sqrt{\alpha_n n \rho_n}$	$\sqrt{\rho_n/(\alpha_n n)}$	$\alpha_n/\rho_n \gtrsim n^{-1} \log n$

Useful in many other contexts !

Conclusion and outlooks

- We showed a theoretical link between smoothness and sparsity in dynamic SC
- As a by-product, obtained best spectral concentration for normalized Laplacian

Conclusion and outlooks

- We showed a theoretical link between smoothness and sparsity in dynamic SC
- As a by-product, obtained best spectral concentration for normalized Laplacian

Outlooks

Conclusion and outlooks

- We showed a theoretical link between smoothness and sparsity in dynamic SC
- As a by-product, obtained best spectral concentration for normalized Laplacian

Outlooks

- Choice of forgetting factor in practice?
  - Theoretical bounds-based methods do not work in practice…

Conclusion and outlooks

- We showed a theoretical link between smoothness and sparsity in dynamic SC
- As a by-product, obtained best spectral concentration for normalized Laplacian

Outlooks

- Choice of forgetting factor in practice?
  - Theoretical bounds-based methods do not work in practice…
- Sparse, constant smoothness analysis
  - Some conjectures from statistical physics

Conclusion and outlooks

- We showed a theoretical link between smoothness and sparsity in dynamic SC
- As a by-product, obtained best spectral concentration for normalized Laplacian

Outlooks

- Choice of forgetting factor in practice?
  - Theoretical bounds-based methods do not work in practice…
- Sparse, constant smoothness analysis
  - Some conjectures from statistical physics

Keriven, Vaiter. **Sparse and Smooth: improved guarantees for Spectral Clustering in the Dynamic Stochastic Block Model.** *arXiv:2002.02892*