gipsa-lab

Grenoble | images | parole | signal | automatique | laboratoire

Optimal Laplacian regularization for sparse spectral community detection

ICASSP 2020

Lorenzo Dall'Amico

Romain Couillet, Nicolas Tremblay

Laboratoire Gipsa-lab, UMR 5216, CNRS, UGA
11 rue des mathématiques 38420 Grenoble, France
lorenzo.dall-amico@gipsa-lab.fr

April 15, 2020

cnrs Grenoble INP UNIVERSITÉ Grenoble Alpes UMR 5216

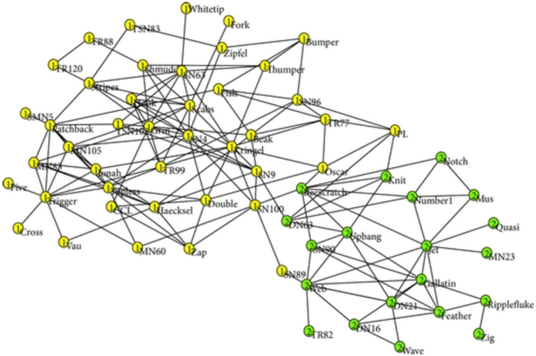


Figure: A representation of the *dolphin* network (Lusseau 2003)

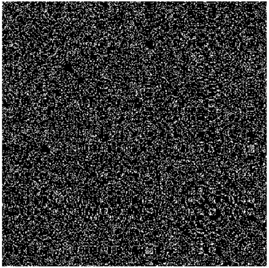
More formally

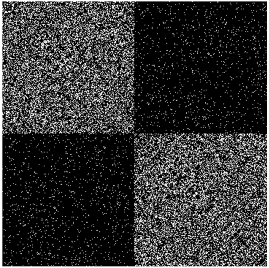
Given a graph $\mathcal{G}(\mathcal{V}, \mathcal{E})$ with $|\mathcal{V}| = n$ nodes and k communities, assign to each node the correct class label.

More formally

Given a graph $\mathcal{G}(\mathcal{V}, \mathcal{E})$ with $|\mathcal{V}| = n$ nodes and k communities, assign to each node the correct class label.

The problem





The solution

A representation of the adjacency matrix $A_{ij} = 0$ (black) if i, j are not connected and $A_{ij} = 1$ (white) if they are connected

Spectral clustering

Node embedding to low dimensional space

Spectral clustering

Node embedding to low dimensional space $\rightarrow$ *k-means*

Spectral clustering

Node embedding to low dimensional space $\rightarrow$ *k-means*

$D = \mathrm{diag}(A\mathbf{1})$

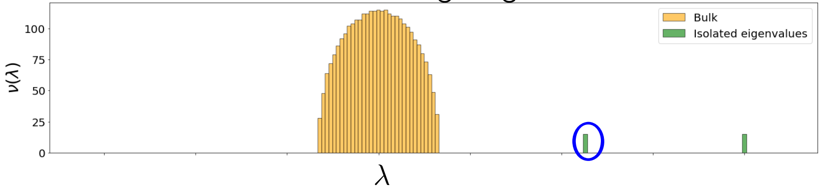
Spectral clustering

Node embedding to low dimensional space $\rightarrow$ *k-means*

$D = \mathrm{diag}(A\mathbf{1})$

Spectrum of $D^{-1}A$

Dense case: average degree $\sim n$

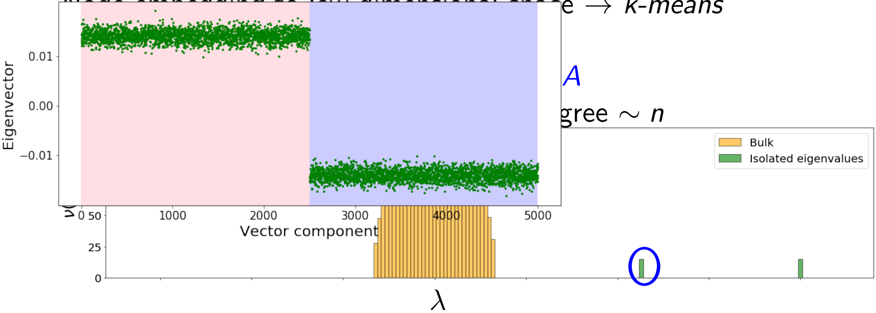


Spectral clustering

Node embedding to low dimensional space → *k-means*

A

gree $\sim n$



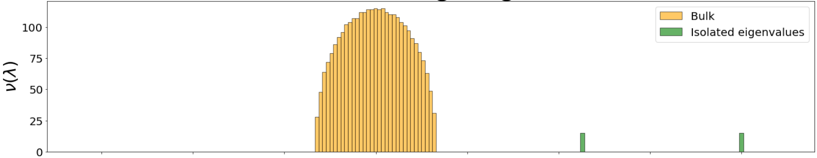
Spectral clustering

Node embedding to low dimensional space $\rightarrow$ *k-means*

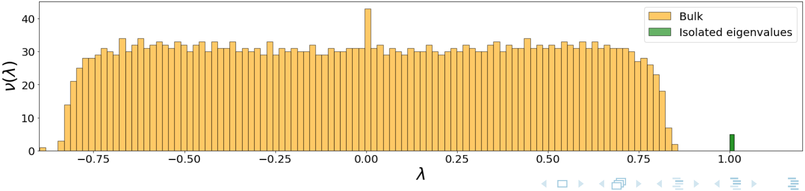
$D = \mathrm{diag}(A\mathbf{1})$

Spectrum of $D^{-1}A$

Dense case: average degree $\sim n$



Sparse case: average degree $= O_n(1)$



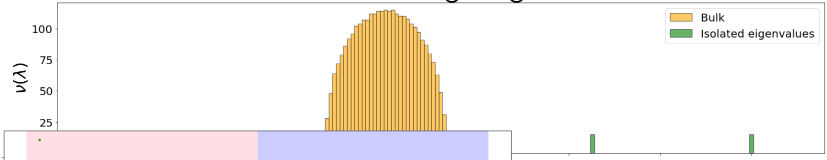
Spectral clustering

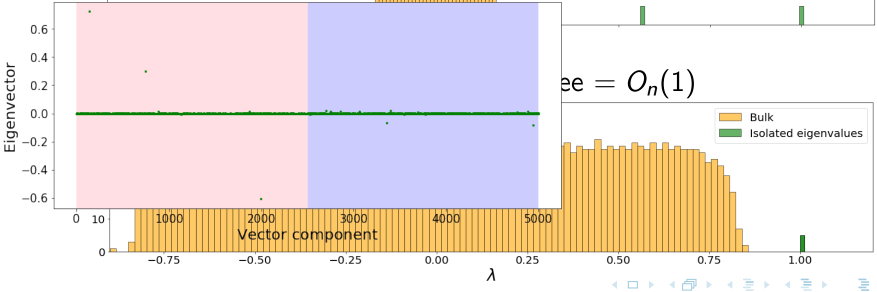
Node embedding to low dimensional space $\rightarrow$ *k-means*

$D = \mathrm{diag}(A\mathbf{1})$

Spectrum of $D^{-1}A$

Dense case: average degree $\sim n$





ee $= O_n(1)$

The degree corrected stochastic block model (DC-SBM)

Dealing with sparsity and heterogeneous degree distributions

The degree corrected stochastic block model (DC-SBM)

Dealing with sparsity and heterogeneous degree distributions

- n: number of nodes

The degree corrected stochastic block model (DC-SBM)

Dealing with sparsity and heterogeneous degree distributions

- n: number of nodes
- $k = 2$: number of communities

The degree corrected stochastic block model (DC-SBM)

Dealing with sparsity and heterogeneous degree distributions

- n: number of nodes
- $k = 2$: number of communities
- $\sigma_i \in \{-1, 1\}$: label of node i

The degree corrected stochastic block model (DC-SBM)

Dealing with sparsity and heterogeneous degree distributions

- n: number of nodes
- $k = 2$: number of communities
- $\sigma_i \in \{-1, 1\}$: label of node i
- $C = \begin{pmatrix} c_{\rm in} & c_{\rm out} \\ c_{\rm out} & c_{\rm in} \end{pmatrix}$: class affinity matrix

The degree corrected stochastic block model (DC-SBM)

Dealing with sparsity and heterogeneous degree distributions

- n: number of nodes
- $k = 2$: number of communities
- $\sigma_i \in \{-1, 1\}$: label of node i
- $C = \begin{pmatrix} c_{\text{in}} & c_{\text{out}} \\ c_{\text{out}} & c_{\text{in}} \end{pmatrix}$: class affinity matrix

Degree-corrected stochastic block model

$$\mathbb{P}(A_{ij} = 1 | \theta_i, \theta_j, \sigma_i, \sigma_j) = \theta_i \theta_j \frac{C_{\sigma_i, \sigma_j}}{n}$$

Theoretical bounds

Define

- $c = \frac{c_{\mathrm{in}} + c_{\mathrm{out}}}{2}$, expected average degree
- $\Phi = \sum_i \theta_i^2$

Theoretical bounds

Define

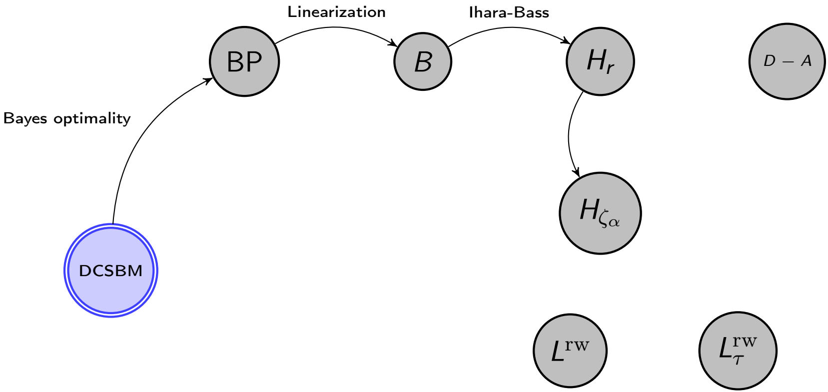
- $c = \frac{c_{\rm in}+c_{\rm out}}{2}$, expected average degree
- $\Phi = \sum_i \theta_i^2$

Detectability threshold

Non-trivial reconstruction iff[1] $\alpha = \frac{c_{\rm in}-c_{\rm out}}{\sqrt{c}} > \frac{2}{\sqrt{\Phi}}$.

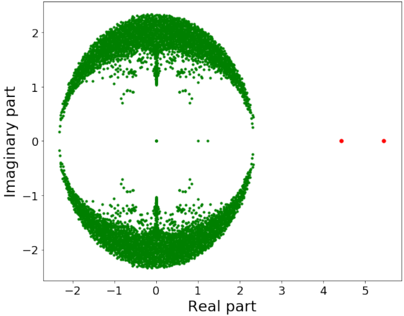
[1] Gulikers *et.al.*, An impossibility result for reconstruction in the degree-corrected stochastic block model

State of the art



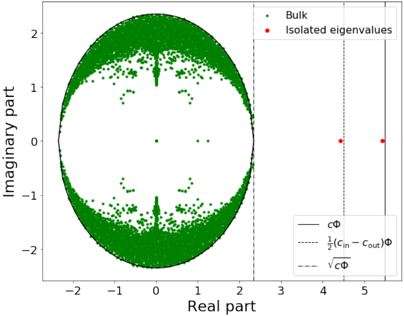
The non-backtracking matrix

$$B_{(ij),(kl)} = \delta_{jk}(1 - \delta_{il}), \quad \forall\ (ij),(kl) \in \mathcal{E}^d$$



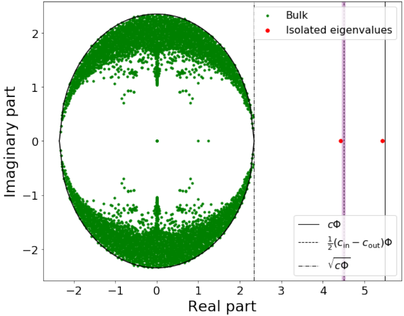
The non-backtracking matrix

$$B_{(ij),(kl)} = \delta_{jk}(1 - \delta_{il}), \quad \forall\ (ij),(kl) \in \mathcal{E}^d$$



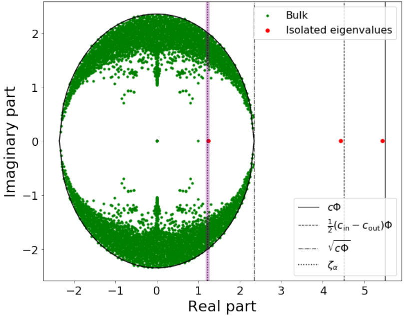
The non-backtracking matrix

$$B_{(ij),(kl)} = \delta_{jk}(1 - \delta_{il}), \quad \forall\ (ij), (kl) \in \mathcal{E}^d$$



The non-backtracking matrix

$$B_{(ij),(kl)} = \delta_{jk}(1-\delta_{il}), \quad \forall\ (ij),(kl) \in \mathcal{E}^d$$



Linearization of BP

$$B\boldsymbol{\delta} = \zeta_\alpha \boldsymbol{\delta} \tag{1}$$

$$\zeta_\alpha = \frac{c_{\rm in} + c_{\rm out}}{c_{\rm in} - c_{\rm out}} = \frac{2\sqrt{c}}{\alpha} \tag{2}$$

To recap

✓ Detects communities down to the threshold

To recap

✓ Detects communities down to the threshold
✓ An informative eigenvalue *inside* the bulk of B

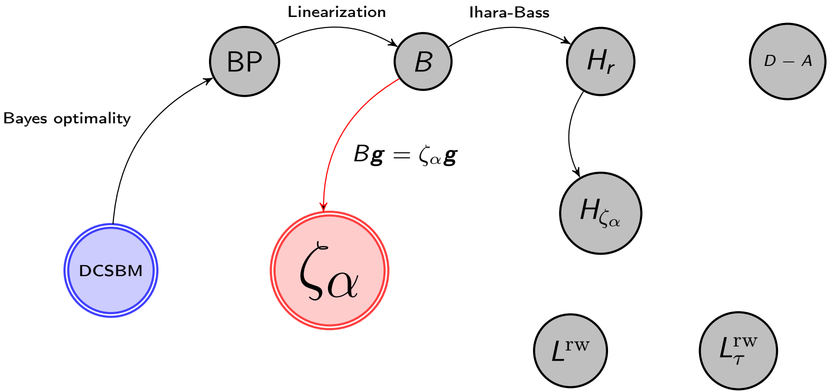
To recap

✓ Detects communities down to the threshold

✓ An informative eigenvalue *inside* the bulk of B

Introduces the parameter ζ_α

A unified framework



The Bethe-Hessian matrix

Ihara-Bass formula

$$B\boldsymbol{g} = \zeta_\alpha \boldsymbol{g}$$
$$[(\zeta_\alpha^2 - 1)I_n + D - \zeta_\alpha A]\boldsymbol{x} = 0$$

The Bethe-Hessian matrix

Ihara-Bass formula

$$Bg = \zeta_\alpha g$$

$$\underbrace{[(\zeta_\alpha^2 - 1)I_n + D - \zeta_\alpha A]}_{\text{Bethe-Hessian } H_{\zeta_\alpha}} x = 0$$

The Bethe-Hessian matrix

Ihara-Bass formula

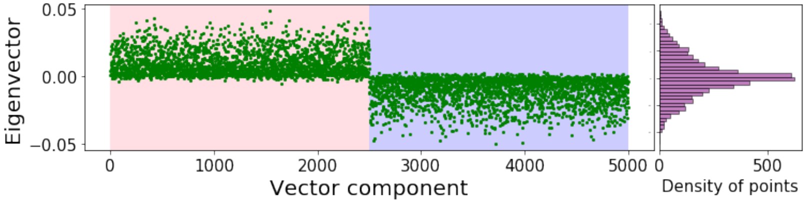
$$B\boldsymbol{g} = \zeta_\alpha \boldsymbol{g}$$

$$\underbrace{[(\zeta_\alpha^2 - 1)I_n + D - \zeta_\alpha A]}_{\text{Bethe-Hessian } H_{\zeta_\alpha}}\boldsymbol{x} = 0$$

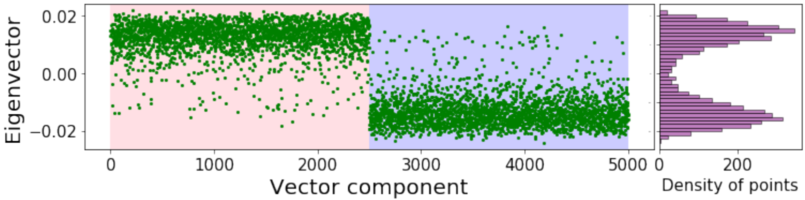
We showed, for all D

$$\mathbb{E}[\boldsymbol{x}] = \boldsymbol{\sigma}$$

Initially proposed value[6] $r = \sqrt{c\Phi}$



Optimal value $r = \zeta_\alpha = \frac{c_{\rm in}+c_{\rm out}}{c_{\rm in}-c_{\rm out}}$



[6] Saade (2014) *Spectral clustering of graphs with the Bethe Hessian*

To recap

H_{ζ_α}

✓ The second smallest eigenvalue is zero and is informative

To recap

H_{ζ_α}

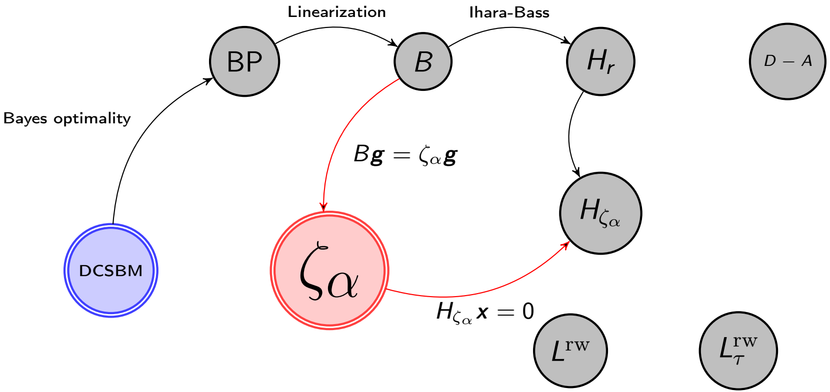
✓ The second smallest eigenvalue is zero and is informative
✓ Detects communities down to the threshold

To recap

H_{ζ_α}

✓ The second smallest eigenvalue is zero and is informative
✓ Detects communities down to the threshold
✓ The eigenvector is resilient to the degree distribution

A unified framework



Regularized Laplacian matrix

$$L_\tau = D_\tau^{-1/2} A D_\tau^{-1/2}$$
$$L_\tau^{\mathrm{rw}} = D_\tau^{-1} A$$

Where $D_\tau = D + \tau I_n$.

[1] Qin (2013) *Regularized spectral clustering under the degree-corrected stochastic blockmodel*

Regularized Laplacian matrix

$$L_\tau = D_\tau^{-1/2} A D_\tau^{-1/2}$$
$$L_\tau^{\mathrm{rw}} = D_\tau^{-1} A$$

Where $D_\tau = D + \tau I_n$. [1]Proposed (heuristic) regularization : $\tau = c$.

[1] Qin (2013) *Regularized spectral clustering under the degree-corrected stochastic blockmodel*

Regularized Laplacian matrix

$$L_\tau = D_\tau^{-1/2} A D_\tau^{-1/2}$$
$$L_\tau^{\mathrm{rw}} = D_\tau^{-1} A$$

Where $D_\tau = D + \tau I_n$. [1]Proposed (heuristic) regularization : $\tau = c$.

From H_{ζ_α} to L_τ

$$H_{\zeta_\alpha}\boldsymbol{x} = [(\zeta_\alpha^2 - 1)I_n + D - \zeta_\alpha A]\boldsymbol{x} = 0$$
$$[D + (\zeta_\alpha^2 - 1)I_n]^{-1} A\boldsymbol{x} = \frac{1}{\zeta_\alpha}\boldsymbol{x}$$

[1] Qin (2013) *Regularized spectral clustering under the degree-corrected stochastic blockmodel*

Regularized Laplacian matrix

$$L_\tau = D_\tau^{-1/2} A D_\tau^{-1/2}$$
$$L_\tau^{\mathrm{rw}} = D_\tau^{-1} A$$

Where $D_\tau = D + \tau I_n$. [1]Proposed (heuristic) regularization : $\tau = c$.

From H_{ζ_α} to L_τ

$$H_{\zeta_\alpha}\boldsymbol{x} = [(\zeta_\alpha^2 - 1)I_n + D - \zeta_\alpha A]\boldsymbol{x} = 0$$
$$[D + (\zeta_\alpha^2 - 1)I_n]^{-1}A\boldsymbol{x} = \frac{1}{\zeta_\alpha}\boldsymbol{x}$$

So

$$\tau = \zeta_\alpha^2 - 1 \leq c\Phi - 1 \approx c$$

[1] Qin (2013) *Regularized spectral clustering under the degree-corrected stochastic blockmodel*

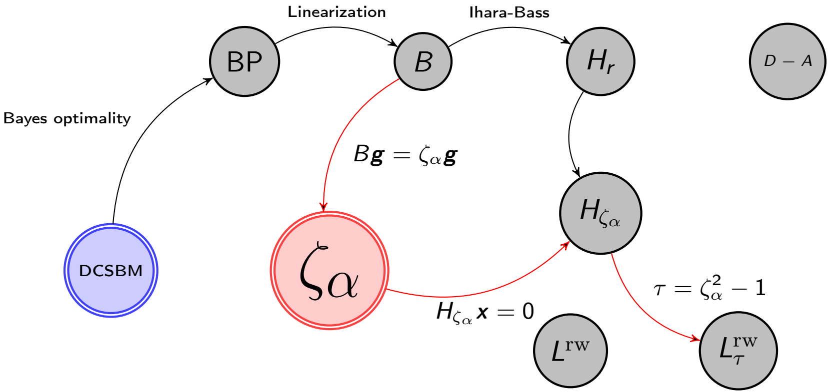
$$L^{\mathrm{rw}}_{\zeta_\alpha^2-1}$$

✓ Explains why $\tau = c$ is a good choice, in practice

$$L^{\mathrm{rw}}_{\zeta_\alpha^2-1}$$

✓ Explains why $\tau = c$ is a good choice, in practice
✓ $\tau = \zeta_\alpha^2 - 1$: minimal regularization for detection down to the threshold

A unified framework



The classical Laplacians

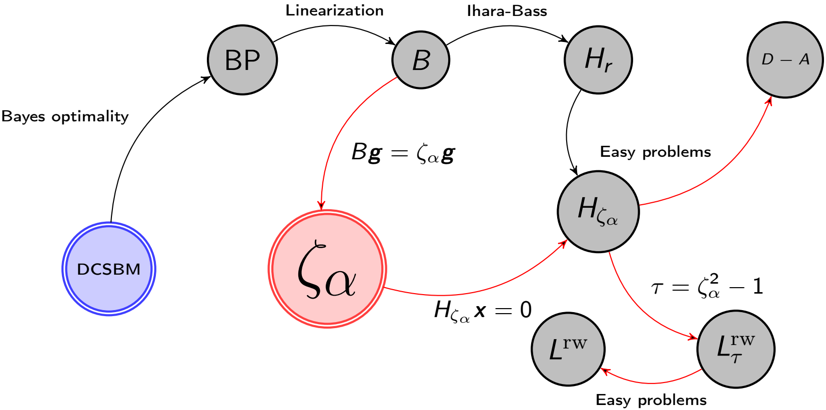
For easy detection problems: $\zeta_\alpha \to 1$

The classical Laplacians

For easy detection problems: $\zeta_\alpha \to 1$

$$[(\zeta_\alpha^2 - 1)I_n + D - \zeta_\alpha A] \to D - A$$
$$[D + (\zeta_\alpha^2 - 1)I_n]^{-1}A \to D^{-1}A$$

A unified framework



Performance on real networks

Dataset	n	c	Φ	k	Alg	$H_{\sqrt{c\Phi}}$	B	L^{rw}	L_{τ}^{sym}
Karate	34	4.6	1.7	2	**0.37**	**0.37**	**0.37**	**0.37**	**0.37**
Dolphins	62	5	1.3	2	**0.38**	0.34	0.22	**0.38**	**0.38**
Polbooks	105	8.4	1.4	3	**0.50**	**0.50**	0.45	**0.50**	**0.50**
Football	115	10.7	1	12	**0.60**	**0.60**	**0.60**	**0.60**	**0.60**
Mail	1133	9.6	1.9	21	**0.50**	0.40	0.37	0.48	**0.50**
Polblogs	1222	27,4	3	2	**0.43**	0.27	0.23	0.00	**0.43**
Tv	3892	8.9	3	41	**0.85**	0.56	0.55	0.55	0.78
Facebook	4039	43.7	2.4	55	**0.79**	0.49	0.48	0.70	0.58
GrQc	4158	6.5	2.8	29	**0.80**	0.51	0.51	0.33	**0.79**
Power grid	4941	2.7	1.5	25	**0.92**	0.33	0.31	**0.92**	0.85
Politicians	5908	14.1	3	62	**0.85**	0.54	0.51	0.74	0.74
GNutella P2P	6299	6.6	2.7	4	**0.40**	0.14	0.14	0.00	0.35
Wikipedia	7066	28.3	5.1	22	0.27	0.18	0.16	**0.34**	0.27
HepPh	11204	21.0	6.2	60	**0.57**	0.42	0.42	0.27	0.52
Vip	11565	11.6	4.4	53	**0.65**	0.32	0.32	0.16	0.54

Conclusion

Contributions

✓ A unified framework for spectral clustering in sparse graphs

Conclusion

Contributions

✓ A unified framework for spectral clustering in sparse graphs
✓ Sparsity and heterogeneity are properly taken into account

Conclusion

Contributions

✓ A unified framework for spectral clustering in sparse graphs
✓ Sparsity and heterogeneity are properly taken into account
✓ Best performing algorithm

Conclusion

Contributions

✓ A unified framework for spectral clustering in sparse graphs
✓ Sparsity and heterogeneity are properly taken into account
✓ Best performing algorithm

Future perspectives

✓ More structured graphs (time-evolving, multi-modal...)

Conclusion

Contributions

✓ A unified framework for spectral clustering in sparse graphs
✓ Sparsity and heterogeneity are properly taken into account
✓ Best performing algorithm

Future perspectives

✓ More structured graphs (time-evolving, multi-modal...)
✓ Is hardness-dependent regularization more general? (SSL kernel methods, weighted graphs...)

Main references (**Dall'Amico**, Couillet, Tremblay)

- *Optimal Laplacian regularization for sparse spectral community detection*, ICASSP 2020
- *A unified framework for spectral clustering in sparse graphs*, arXiv:2003.09198
- *Revisiting the Bethe-Hessian: improved community detection in sparse heterogeneous graphs*, NeurIPS 2019.

Main references (**Dall'Amico**, Couillet, Tremblay)

- *Optimal Laplacian regularization for sparse spectral community detection*, ICASSP 2020
- *A unified framework for spectral clustering in sparse graphs*, arXiv:2003.09198
- *Revisiting the Bethe-Hessian: improved community detection in sparse heterogeneous graphs*, NeurIPS 2019.

Thank you!