Spectral Clustering of Graphs with the Bethe Hessian

Alaa Saade[1] Florent Krzakala[1,2] Lenka Zdeborová[3]

[1]Laboratoire de Physique Statistique, CNRS UMR 8550,
Université P. et M. Curie Paris 6 et École Normale Supérieure, 24, rue Lhomond, 75005 Paris, France.

[2]ESPCI and CNRS UMR 7083 Gulliver, 10 rue Vauquelin,Paris 75005

[3]Institut de Physique Théorique, CEA Saclay and URA 2306, CNRS, 91191 Gif-sur-Yvette, France

Speaker: 李寰

Clustering by the Laplacian $L = D - A$[1]

Input: Adjacency matrix $A \in \mathbb{R}^{n \times n}$, number of clusters q

- $q = 2$
  - Cluster by the signs of the 2nd eigenvector of L
- $q > 2$
  - Let $u_1, \cdots, u_q$ be the first q eigenvectors of L
  - Let $U = (u_1, \cdots, u_q) \in \mathbb{R}^{n \times q}$
  - Let $\begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = U$
  - Cluster the points $y_1, \cdots, y_n \in \mathbb{R}^q$ with k-means

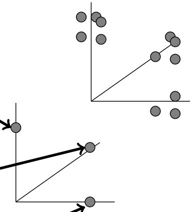
[1]Ulrike Luxburg (2007). A tutorial on spectral clustering. Statistics and Computing, 17(4):395-416.

Clustering by the Adjacency Matrix A[2]

$$\begin{pmatrix}
1&1&1&1&1&0&0&0&0&0&0&0\\
1&1&1&1&1&0&0&0&0&0&0&0\\
1&1&1&1&1&0&0&0&0&0&0&0\\
1&1&1&1&1&0&0&0&0&0&0&0\\
1&1&1&1&1&0&0&0&0&0&0&0\\
0&0&0&0&0&1&1&1&1&0&0&0\\
0&0&0&0&0&1&1&1&1&0&0&0\\
0&0&0&0&0&1&1&1&1&0&0&0\\
0&0&0&0&0&1&1&1&1&0&0&0\\
0&0&0&0&0&0&0&0&0&1&1&1\\
0&0&0&0&0&0&0&0&0&1&1&1\\
0&0&0&0&0&0&0&0&0&1&1&1
\end{pmatrix}
\quad
\begin{matrix}\begin{pmatrix}
1&0&0\\
1&0&0\\
1&0&0\\
1&0&0\\
1&0&0\\
0&1&0\\
0&1&0\\
0&1&0\\
0&1&0\\
0&0&1\\
0&0&1\\
0&0&1
\end{pmatrix}\end{matrix}$$

adjacency matrix

singular vectors



[2] Avrim Blum, John Hopcroft, and Ravindran Kannan (2016). Foundations of Data Science (pp. 275-281).

Spectral Clustering in Sparse Networks

- Clustering based on A fails to detect communities
- Locally tree-like structure
- Leading eigenvalues of A are dictated by the vertices of highest degree[3]
- Corresponding eigenvectors are localized around these vertices[3]

[3] Michael Krivelevich and Benny Sudakov (2003). The largest eigenvalue of sparse random graphs. Combinatorics, Probability and Computing, 12(01):61-72.

Clustering by the Non-Backtracking Matrix[4]

- The $2m \times 2m$ non-backtracking matrix B

$$B_{(u\to v),(x\to y)} = \begin{cases} 1 & \text{if } v = x \text{ and } u \neq y \\ 0 & \text{otherwise .} \end{cases}$$

- The spectrum of B is not sensitive to high-degree vertices
- Properties of B
  - A tree contributes zero eigenvalues to the spectrum
  - Unicyclic components yield eigenvalues either 1 or -1

[4]Florent Krzakala, Cristopher Moore, et al (2013). Spectral redemption in clustering sparse networks. Proceedings of the National Academy of Sciences, 110(52):20935–20940.

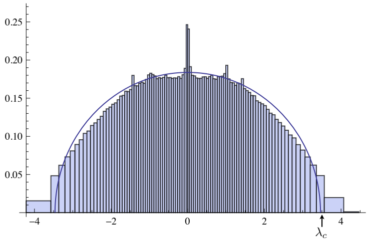
Stochastic Block Model

- Stochastic block model
  - q groups of vertices with size n/q
  - Each vertex v has a group label $g_v \in \{1, \cdots, q\}$
  - Adjacency matrix A is generated according to the distribution

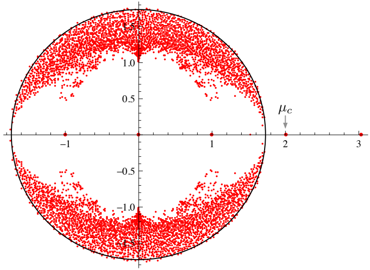
$$\Pr[A_{u,v} = 1] = \begin{cases} c_{\text{in}}/n & g_u = g_v \\ c_{\text{out}}/n & g_u \neq g_v \end{cases}, \text{ where } c_{\text{in}} > c_{\text{out}}$$

- Average degree $c = (c_{\text{in}} + c_{\text{out}})/2$ when $q = 2$
- G becomes sparse when n is large

Spectrum of A and B



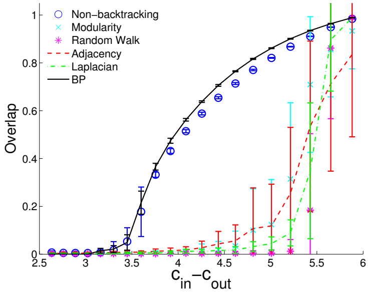
(a) The spectrum of A

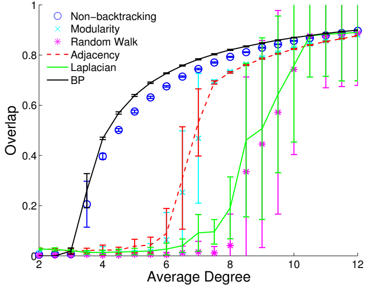


(b) The spectrum of B

- $n = 4000$, $c_{\text{in}} = 5$, $c_{\text{out}} = 1$, $c = (c_{\text{in}} + c_{\text{out}})/2 = 3$
- The radius of the bulk
  - For A, $2\sqrt{c} = 3.46$
  - For B, $\sqrt{c} = 1.73$

Experiments $n = 10^5$

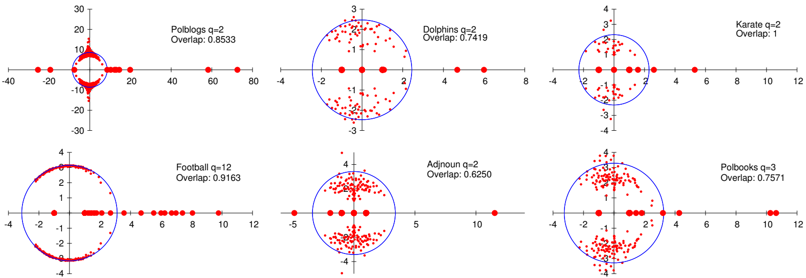




- Overlap $\triangleq \left(\frac{1}{n}\sum_u \delta_{g_u,\tilde{g}_u} - \frac{1}{q}\right) \Big/ \left(1 - \frac{1}{q}\right)$
- Theoretical threshold[5] ≈ 3.46

[5]Elchanan Mossel, Joe Neeman, and Allan Sly (2012). Stochastic block models and reconstruction. arXiv preprint, arXiv:1202.1499.

Detecting Number of Clusters



- Each circle's radius is $\sqrt{\rho(B)}$
- The number of real eigenvalues outside the circle indicates the number q of clusters

Reducing the Computational Complexity

- All eigenvalues λ of B not ± 1 are the roots of the equation[6]

$$\det\left[\lambda^2 \mathbb{1} - \lambda A + (D - \mathbb{1})\right] = 0 \tag{1}$$

- By the first companion linearization[7], roots of eq. (1) are eigenvalues of

$$B' = \begin{pmatrix} 0 & D - \mathbb{1} \\ -\mathbb{1} & A \end{pmatrix}$$

- Clustering by a $2n \times 2n$ matrix rather than a $2m \times 2m$ one

[6] Omer Angel, Joel Friedman, and Shlomo Hoory (2015). The non-backtracking spectrum of the universal cover of a graph. Trans. Amer. Math. Soc., 367(6):4287-4318.

[7] Francoise Tisseur and Karl Meerbergen (2001). The quadratic eigenvalue problem. SIAM Review, 43(2):235-286.

The Bethe Hessian Matrix

- All eigenvalues λ of B not ± 1 are the roots of the equation

$$\det\left[\lambda^2 \mathbb{1} - \lambda A + (D - \mathbb{1})\right] = 0 \quad (1)$$

- The Bethe Hessian matrix $H(r)$

$$H(r) := r^2 \mathbb{1} - rA + (D - \mathbb{1})$$

- $\forall$ real eigenvalue λ of B, $H(\lambda)$ has a eigenvalue 0
  - $H(\lambda\ )$'s null space $\longleftrightarrow$ B's eigenvector corresponding to λ
  - Let $r_c = \sqrt{\rho(B)}$
  - Eigenvectors of $H(r_c)$'s negative eigenvalues reveal clustering

The Bethe Hessian Matrix

- All eigenvalues λ of B not ± 1 are the roots of the equation

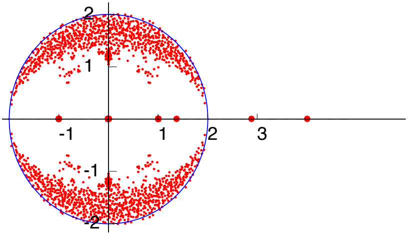
$$\det\left[\lambda^2 \mathbb{1} - \lambda A + (D - \mathbb{1})\right] = 0 \tag{1}$$

- The Bethe Hessian matrix $H(r)$

$$H(r) := r^2 \mathbb{1} - rA + (D - \mathbb{1})$$

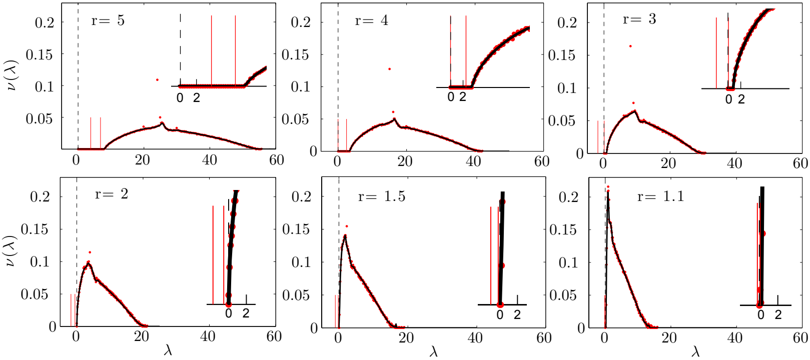
- $\forall$ real eigenvalue λ of B, $H(\lambda)$ has a eigenvalue 0
  - $H(\lambda_2)$'s null space $\longleftrightarrow$ B's eigenvector corresponding to λ_2
  - Let $r_c = \sqrt{\rho(B)}$
  - Eigenvectors of $H(r_c)$'s negative eigenvalues reveal clustering

Spectrum of B and $H(r)$

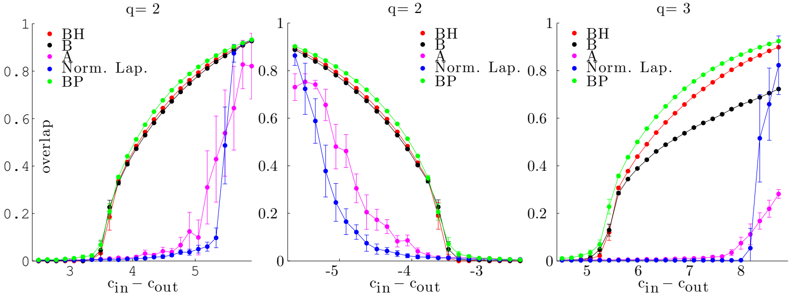


Stochastic block model

- $n = 10^4$
- $c = 4$
- $c_{\text{out}}/c_{\text{in}} = \frac{1}{7}$



Experiments



PART	Non-backtracking [9]	Bethe Hessian
Polbooks ($q = 3$) [1]	0.742857	0.757143
Polblogs ($q = 2$) [10]	0.864157	0.865794
Karate ($q = 2$) [24]	1	1
Football ($q = 12$) [6]	0.924111	0.924111
Dolphins ($q = 2$) [16]	0.741935	0.806452
Adjnoun ($q = 2$) [8]	0.625000	0.660714

Conclusions and Perspectives

- Comparing to non-backtracking clustering
  - $H(r_c)$ is $n \times n$ symmetric, while B' is $2n \times 2n$ non-symmetric
  - Need to compute $\rho(B)$ to calculate $r_c = \sqrt{\rho(B)}$
    - By solving quadratic eigenproblem (2) using a SLP algorithm[8]

$$\det H(\lambda) = \det\left[\lambda^2 \mathbb{1} - \lambda A + (D - \mathbb{1})\right] = 0 \tag{2}$$

- Detecting the communities in sparse network
  - All the way down to the threshold in SBM
- The number of negative eigenvalues indicating the number of clusters

[8] Axel Ruhe (1973). Algorithms for the nonlinear eigenvalue problem. SIAM Journal on Numerical Analysis, 10(4):674–689.

The End