Community Detection in Sparse Networks by the Symmetrized Laplacian Inverse Matrix (SLIM)

Xianshi Yu

the Hong Kong University of Science and Technology

Joint work with Bing-Yi Jing, Ting Li & Ningchen Ying

19 Aug 2017

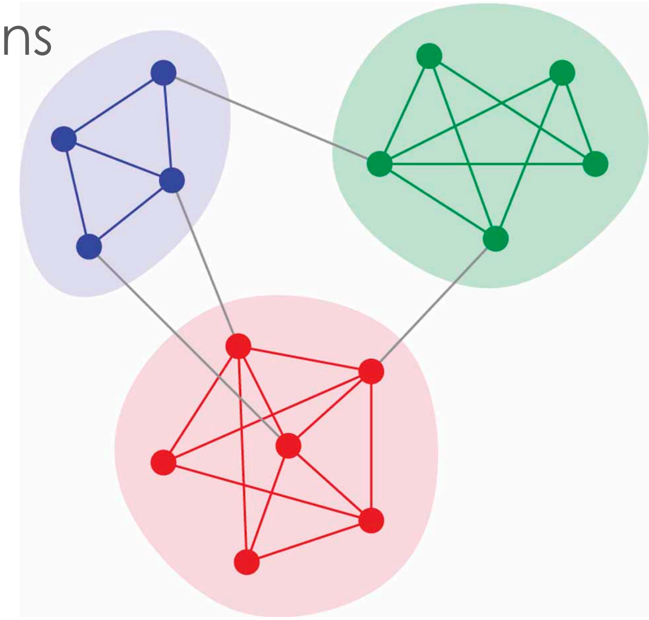
Outline

- Network community detection
- Community detection in **sparse** networks
- SLIM method
- Empirical performance of the SLIM method
- Comparison between the SLIM method & other spectral methods
- Theoretical result

Network Community Detection

Motivations

- Parallel Computation
- Detect functional modules of proteins
- Find friend circles



Network Community Detection

Methods

- Algorithm-based methods
- Maximize Modularity

  e.g. Newman and Girvan, 2004

- Search and combine cliques

  e.g. Palla et al., 2005

Methods

- Model-based methods
- MLE

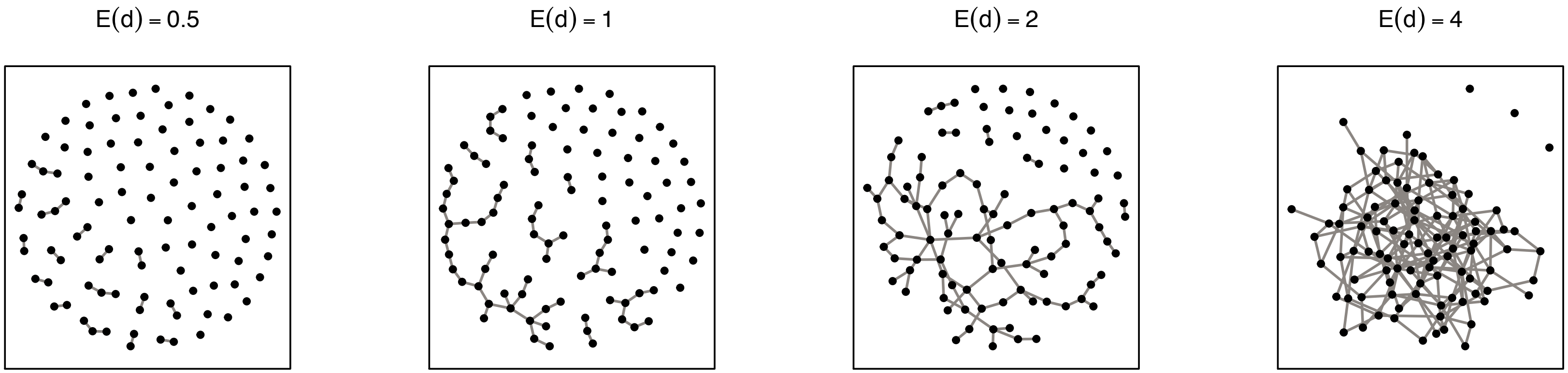
  e.g. Clauset et al., 2008; Airoldi, 2009; Karrer1 and Newman, 2011

- Spectral methods for Stochastic Block Model

  Rohe et al., 2011; Jin, 2015; Zhang et al., 2015

Community Detection in **Sparse** Networks

Networks with Different Edge Densities



ER Random Graphs, n=100

Three Scenarios

① $E(\text{degree}) = \Omega(\log n)$

② $E(\text{degree}) = \infty$

③ $E(\text{degree}) = O(1)$

2. and 3. are commonly seen but understudied

Community Detection in Sparse Networks

Pioneer Works

- Krzakala et al. 2013; Newman 2013

  Consider the network of edges, non-backtracking

- Amini et al. 2013; Joseph and Yu 2013; Gao et al. 2015

  Normalized Spectral Clustering with Regularization

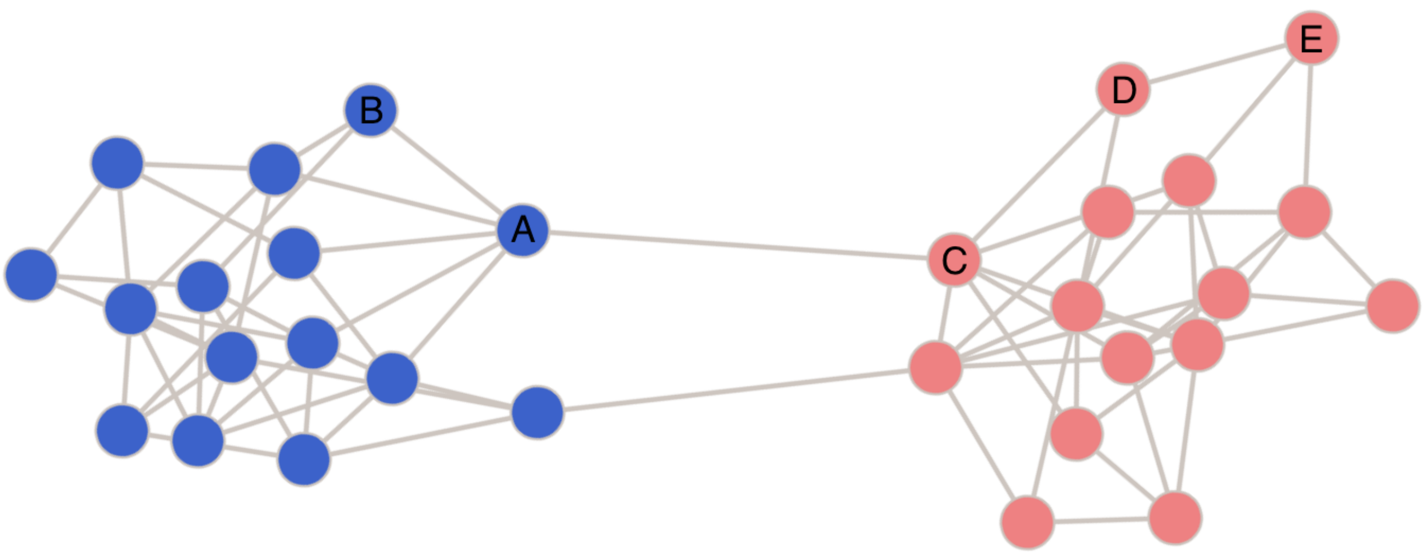
- Bhattacharyya and Bickel 2014; Bhattacharyya and Bickel 2015

  Graph distance based (GDB), multi-dimensional scaling

SLIM (Symmetrized Laplacian Inverse Matrix) Method

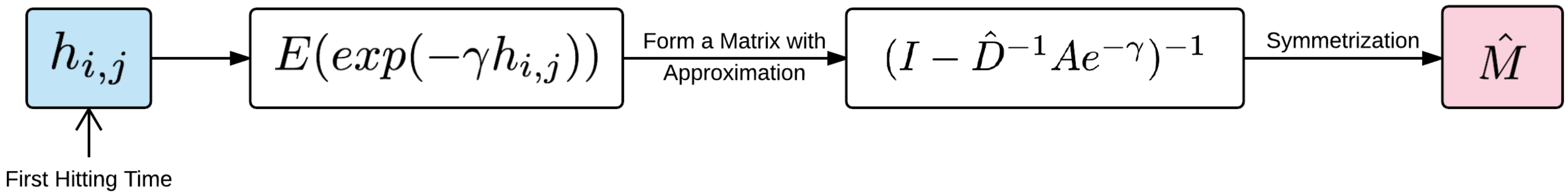
SLIM Method

Consider First Hitting Time



SLIM Method

Symmetrized Laplacian Inverse Matrix



- Laplacian Inverse Matrix

$$W = \sum_{k=0}^{\infty} exp(-\gamma k)(D^{-1}A)^k = (I - D^{-1}Ae^{-\gamma})^{-1}$$

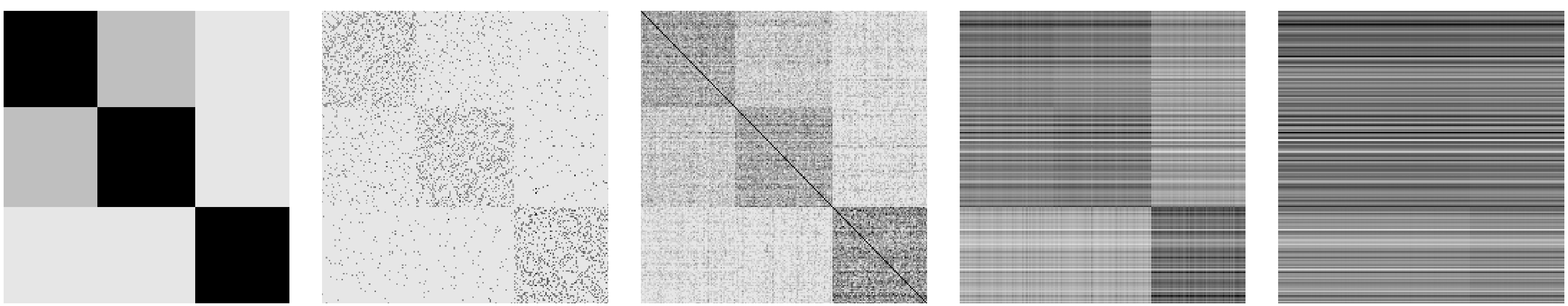
- Symmetrized Laplacian Inverse Matrix (SLIM)

$$M = \frac{W + W^T}{2}$$

SLIM Method

Alternative Explanation of SLIM

- $W = \sum_{k=0}^{\infty} exp(-\gamma k)(D^{-1}A)^k$ is a weighted summation of different scales of representations of network structure
- $W_m = \sum_{k=0}^{m} exp(-\gamma k)(D^{-1}A)^k$ is reasonable approximation of W



$E(A)$ $(D^{-1}A)^1$ $(D^{-1}A)^2$ $(D^{-1}A)^8$ $(D^{-1}A)^{20}$

Apply Classical Spectral Method to M

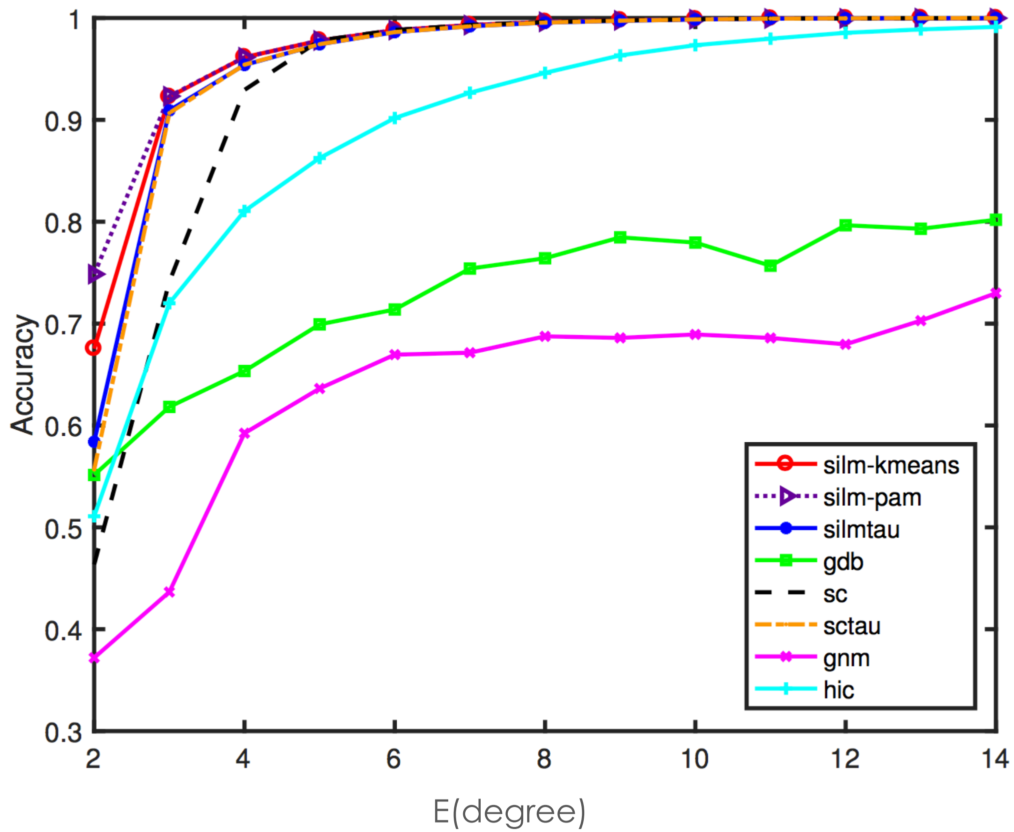
Algorithm

- Compute M (M_m)
- Perform Spectral decomposition to M
- View first k eigenvectors as locations of nodes
- Clustering(e.g. k-means)

Empirical Performance of the SLIM Method

Empirical Performance of the SLIM Method

Simulation under SBM



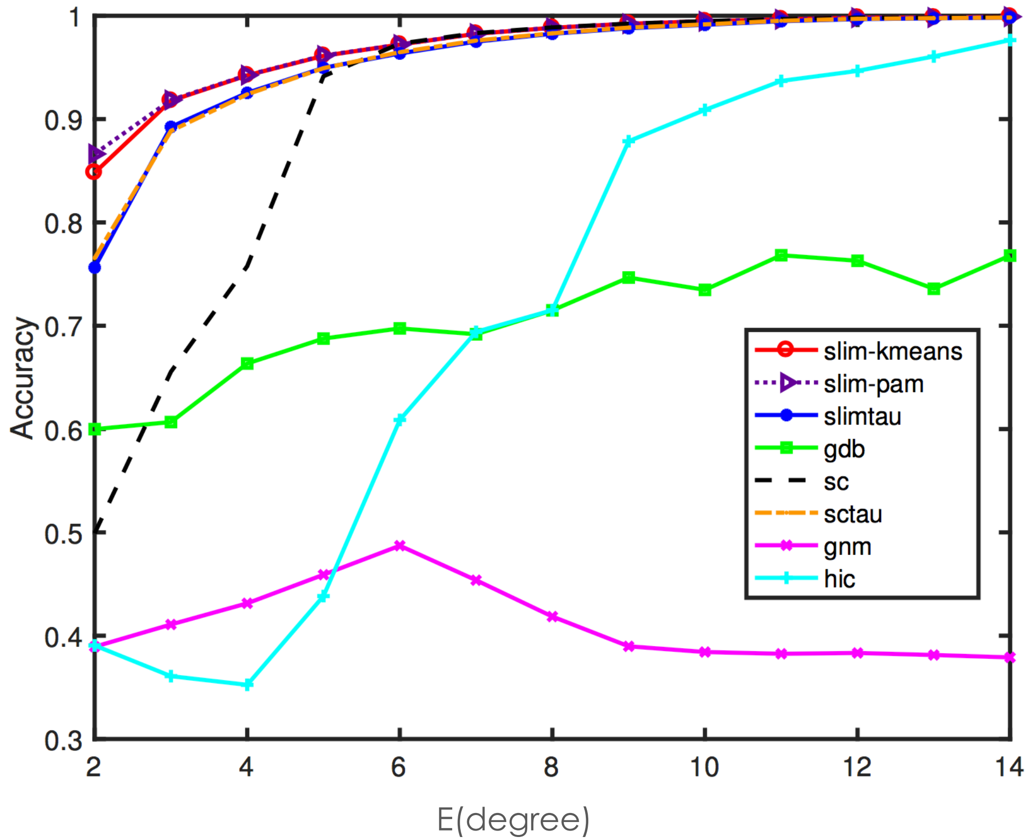
N=1200

3 communities

Π =(0.33,0.33,0.33)

Empirical Performance of the SLIM Method

Simulation under DCSBM



N=1200

3 communities

Π=(0.33,0.33,0.33)

10% nodes are hubs

Empirical Performance of the SLIM Method

Simulation under DCSBM

Replicate Experiment 2 in Jin, J. (2015) (a paper addressing Degree Corrected SBM specifically which proposed SCORE)

Method	SC	NSC	NGM	SCORE	SLIM
Mean (SD)	0.378 (0.041)	0.165 (0.084)	0.355 (0.01)	0.070 (0.004)	0.0619 (0.009)

Mean & SD of **error rate**

Real Data Analysis

Politic Blogs with Hyperlinks

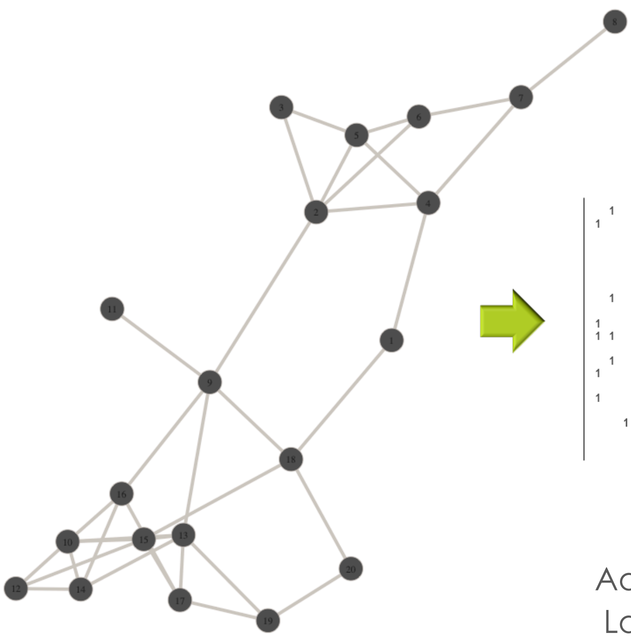
- Each has been manually labeled to be Conservative or Liberal
- Sparse and inhomogeneous in degree

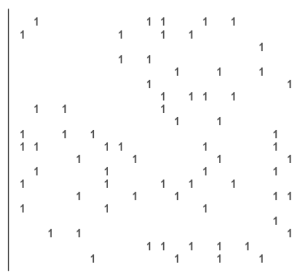
error rate of the SLIM method is **50/1222**

The best performance know is 55/1222 (SCORE)

Comparison Between the SLIM Method & Other Spectral Methods

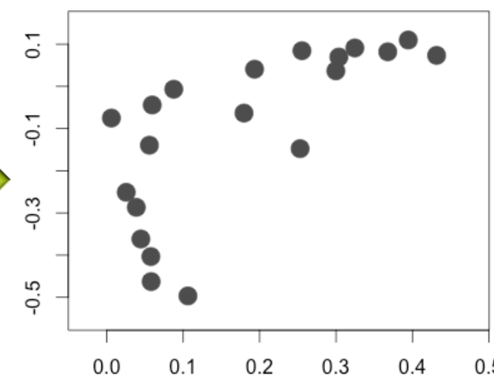
Illustration of Spectral Method





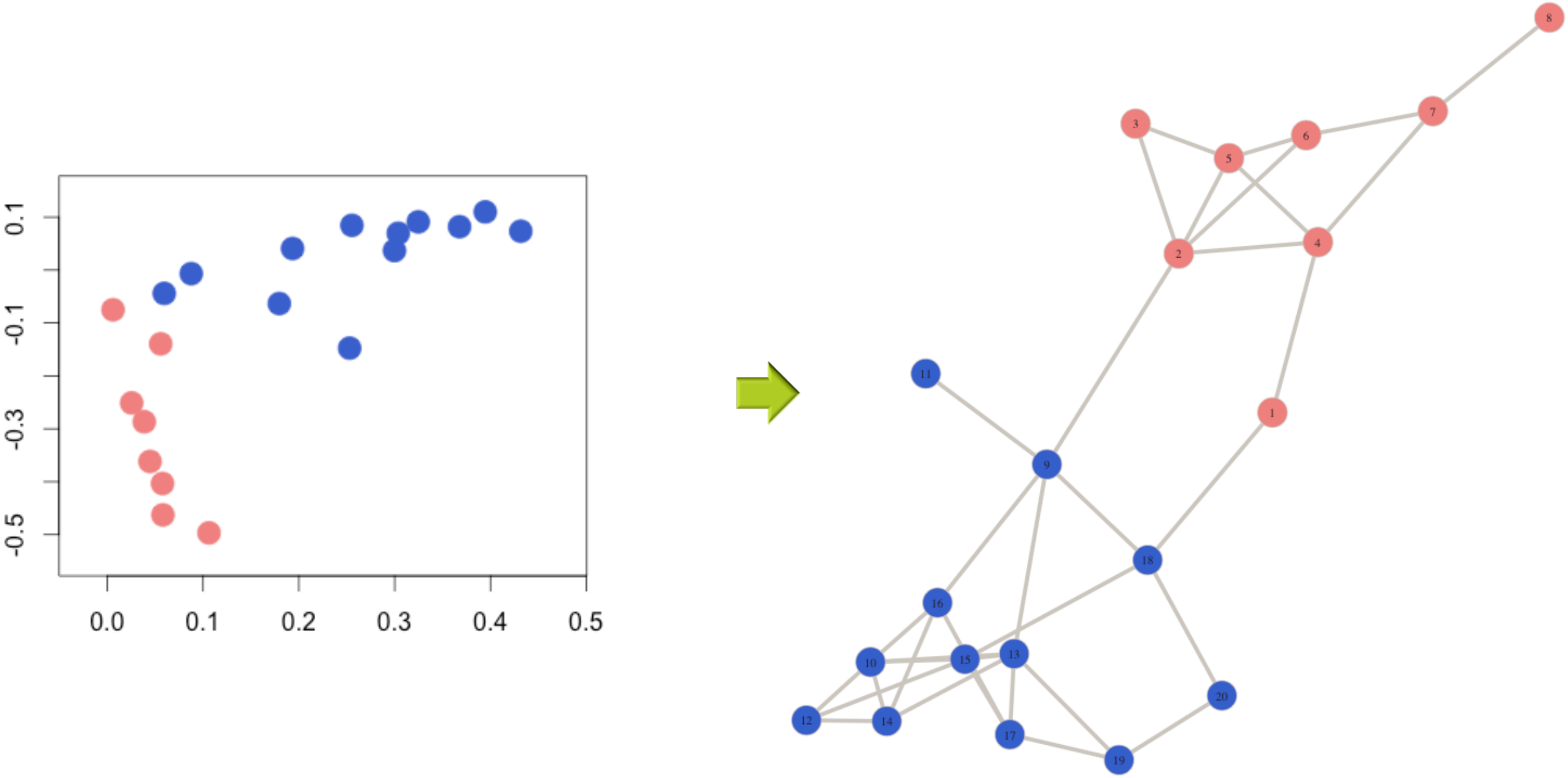
Adjacency Matrix
Laplacian Matrix
Distance Matrix
…

0.37	0.08
0.3	0.07
0.01	-0.08
0.09	-0.01
0.04	-0.36
0.06	-0.14
0.32	0.09
0.19	0.04
0.04	-0.29
0.18	-0.06
0.43	0.07
0.06	-0.46
0.3	0.04
0.39	0.11
0.11	-0.5
0.26	0.08
0.06	-0.04
0.03	-0.25
0.25	-0.15
0.06	-0.4



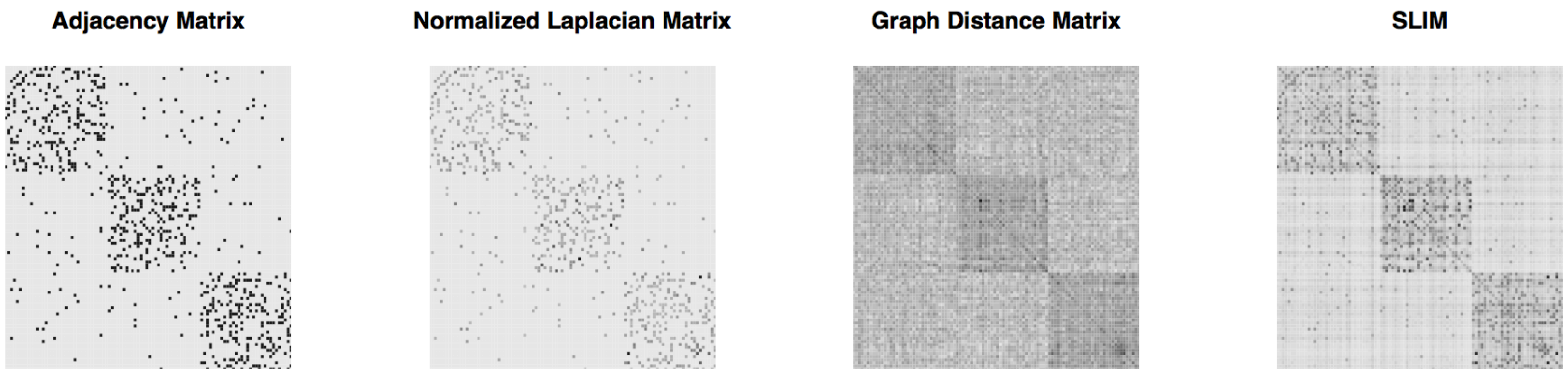
Comparison Between the SLIM Method & Other Spectral Methods

Illustration of Spectral Method



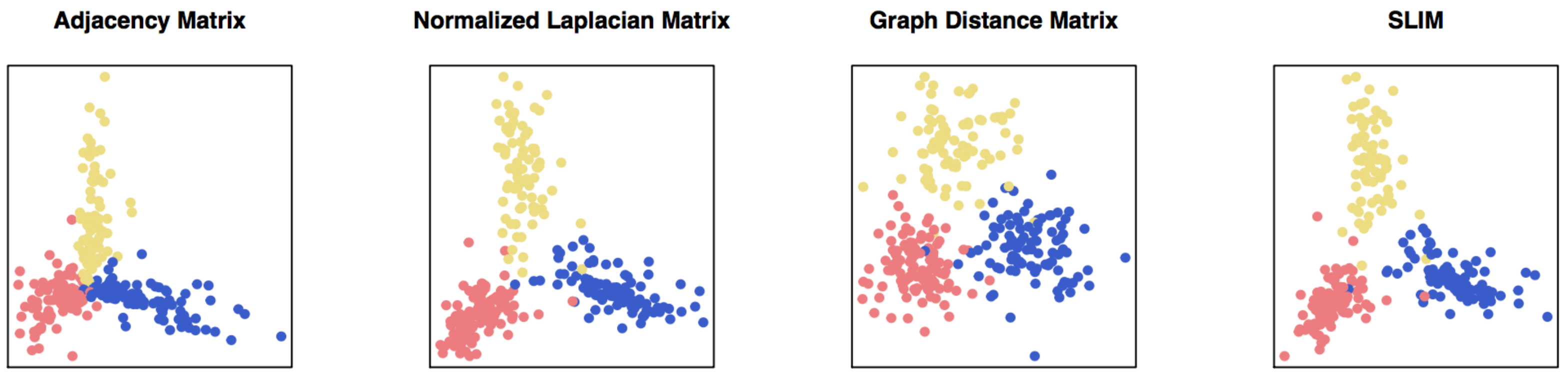
Compare Matrices

- Under Stochastic Block Model (SBM), in expextation, the matrices to be decomposed are BLOCK MATRICES



Comparison Between the SLIM Method & Other Spectral Methods

Compare Eigenvectors



Theoretical Result

Theoretical Result

Consistency in Sparse Scenario

Accuracy $\xrightarrow{P}$ 1 if d → ∞ as n → ∞, under certain conditions

- Consistency of the SLIM method is proven for dense networks
- Consistency of “SLIM with regularization” is proven for sparse networks
- Here the “regularization” is to deal with nodes with with very large or small degrees

Summary

Motivation

- First hitting time describes the closeness between nodes
- The exponential transformation emphasizes local similarities
- SLIM is never sparse

Performance

- Spectral method is more computationally scalable than MLE methods
- Empirical study show that the SLIM method works well especially for <u>sparse</u> networks and networks with <u>inhomogeneous degrees</u>

Reference

- Airoldi, E. M., Blei, D. M., Fienberg, S. E. and Xing, E. P. (2009). Mixed membership stochas- tic blockmodels. *In Advances in Neural Information Processing Systems* (pp. 33-40).
- Bhattacharyya, S. and Bickel, P. J. (2014). Community detection in networks using graph dis- tance. arXiv preprint arXiv:1401.3915.
- Bhattacharyya, Sharmodeep, and Peter J. Bickel. "Spectral Clustering and Block Models: A Review And A New Algorithm." arXiv preprint arXiv:1508.01819 (2015).
- Clauset, A., Moore, C. and Newman, M. E. (2008). Hierarchical structure and the prediction of missing links in networks. *Nature*, 453(7191), 98-101.
- Jin, J. (2015). Fast community detection by SCORE. *The Annals of Statistics*, 43(1), 57-89.
- Karrer, B. and Newman, M. E. (2011). Stochastic blockmodels and community structure in networks. *Physical Review E*, 83(1), 016107.

Reference

- Krzakala, F., Cristopher, M., Elchanan, M., Joe N., Allan S., Lenka Z., and Pan Z. (2013) ”Spectral redemption in clustering sparse networks.” Proceedings of the National Academy of Sciences 110, no. 52: 20935-20940.

- Newman, M. E. and Girvan, M. (2004). Finding and evaluating community structure in net- works. *Physical Review E*, 69(2), 026113.

- Newman, M. E. J. (2013) Spectral community detection in sparse networks. arXiv preprint arXiv:1308.6494 .

- Palla, G., Derenyi, I., Farkas, I. and Vicsek, T. (2005). Uncovering the overlapping community structure of complex networks in nature and society. *Nature*, 435(7043), 814-818.

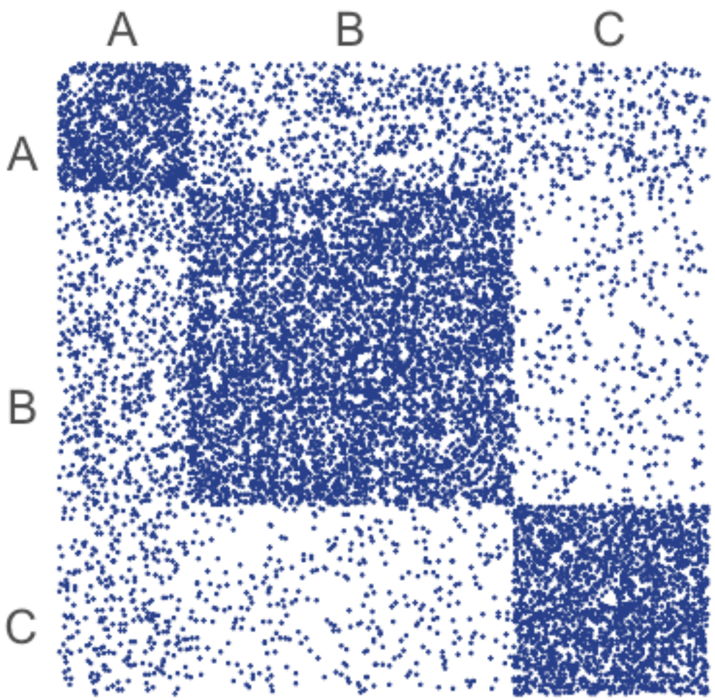
- Rohe, K., Chatterjee, S. and Yu, B. (2011). Spectral clustering and the high-dimensional stochastic blockmodel. *The Annals of Statistics*, 1878-1915.

- Zhang, Y., Levina, E. and Zhu, J. (2014). Detecting overlapping communities in networks with spectral methods. arXiv preprint arXiv:1412.3432.

Appendix

Appendix

Stochastic Block Model (SBM)



- n nodes and K groups
- g(i) ∊ {1,…,K} indicates the group assignment
- Node i and j connects with probability P(g(i),g(j)), where P is a K×K matrix

Appendix

Stochastic Block Model (SBM)

Example: 5 nodes and 2 groups

$$P = \begin{pmatrix} 0.8 & 0.8 & 0.8 & 0.2 & 0.2 \\ 0.8 & 0.8 & 0.8 & 0.2 & 0.2 \\ 0.8 & 0.8 & 0.8 & 0.2 & 0.2 \\ 0.2 & 0.2 & 0.2 & 0.6 & 0.6 \\ 0.2 & 0.2 & 0.2 & 0.6 & 0.6 \end{pmatrix} \Rightarrow A = \begin{pmatrix} 1 & 1 & 1 & & \\ 1 & 1 & & & 1 \\ 1 & & 1 & & \\ & & & 1 & \\ & 1 & & & 1 \end{pmatrix}$$