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 \textit{sparse} networks
- SLIM method
- Empirical performance of the SLIM method
- Comparison between the SLIM method & other spectral methods
- Theoretical result
Network Community Detection
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
Network Community Detection

Methods

- Model-based methods
  - MLE
    e.g. Clauset et al., 2008; Airoldi, 2009; Karrer and Newman, 2011
  - Spectral methods for Stochastic Block Model
    Rohe et al., 2011; Jin, 2015; Zhang et al., 2015
Community Detection in Sparse Networks
Community Detection in Sparse Networks

Networks with Different Edge Densities

$E(d) = 0.5$

$E(d) = 1$

$E(d) = 2$

$E(d) = 4$

ER Random Graphs, $n=100$
Community Detection in Sparse Networks

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

\( h_{i,j} \rightarrow E(\exp(-\gamma h_{i,j})) \rightarrow (I - \hat{D}^{-1}Ae^{-\gamma})^{-1} \rightarrow \hat{M} \)

- Laplacian Inverse Matrix

\[ W = \sum_{k=0}^{\infty} \exp(-\gamma k)(D^{-1} A)^k = (I - D^{-1} A e^{-\gamma})^{-1} \]

- 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 \( W \).
SLIM Method

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
\(\Pi = (0.33, 0.33, 0.33)\)
Empirical Performance of the SLIM Method
Simulation under DCSBM

N=1200
3 communities
\( \Pi = (0.33, 0.33, 0.33) \)
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)

<table>
<thead>
<tr>
<th>Method</th>
<th>SC</th>
<th>NSC</th>
<th>NGM</th>
<th>SCORE</th>
<th>SLIM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean (SD)</td>
<td>0.378 (0.041)</td>
<td>0.165 (0.084)</td>
<td>0.355 (0.01)</td>
<td>0.070 (0.004)</td>
<td>0.0619 (0.009)</td>
</tr>
</tbody>
</table>

Mean & SD of error rate
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 known is *55/1222* (SCORE)
Comparison Between the SLIM Method & Other Spectral Methods
Comparison Between the SLIM Method & Other Spectral Methods

Illustration of Spectral Method

- Adjacency Matrix
- Laplacian Matrix
- Distance Matrix

...
Comparison Between the SLIM Method & Other Spectral Methods

Illustration of Spectral Method
Comparison Between the SLIM Method & Other Spectral Methods

**Compare Matrices**

- Under Stochastic Block Model (SBM), in expectation, the matrices to be decomposed are BLOCK MATRICES
Comparison Between the SLIM Method & Other Spectral Methods

Compare Eigenvectors
Theoretical Result
Consistency in Sparse Scenario

Accuracy $\xrightarrow{P} 1$ if $d \rightarrow \infty$ as $n \rightarrow \infty$, 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 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 sparse networks and networks with inhomogeneous degrees
Reference


Stochastic Block Model (SBM)

- n nodes and K groups
- \( g(i) \in \{1,...,K\} \) 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

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