Clique Topology of the Stochastic Block Model

Derrick Lee, Robert Ghrist, Danielle Bassett

1 University of Pennsylvania, Department of Mathematics
2 University of Pennsylvania, Department of Electrical & Systems Engineering
3 University of Pennsylvania, Department of Bioengineering

Introduction

The stochastic block model has recently been employed as a generative graph model for real-world networks in a variety of fields including neuroscience [2]. In addition, persistent homology is gaining traction as a tool to study the higher order topology of networks, by providing topological signatures such as Betti curves [6]. Recent results in random topology provide the expected topological behaviour of the clique complex generated by common graph models such as the Erdős-Renyi graph [5]. We aim to study asymptotic homological properties of the clique complex generated by the stochastic block model via spectral methods first introduced by Hoffman, Kahle and Paquette [4].

Stochastic Block Model

The Stochastic Block Model (SBM) is a generative random graph model in which the vertices are partitioned into blocks (or communities) that share a similar connectivity profile. It is defined by the following parameters:

- n ∈ N: the number of vertices
- k ∈ N: the number of blocks
- \( \{b_i \}_{i=1}^k \): a partition of the vertices into blocks
- \( P \in [0,1]^{k \times k} \): a symmetric matrix of edge probabilities

Normalized Laplacian and Spectral Gap

Graph-theoretic properties such as connectedness can often be described in terms of the spectra of matrices derived from the graph. Common matrices to study are the adjacency matrix \( A \) and Laplacian matrix \( L \). Here we study the normalized Laplacian \( \mathcal{L} \), defined as follows. Let \( G \) be a graph and let \( d_i \) denote the degree of vertex \( v_i \). Then

\[
\mathcal{L}(u,v) = \begin{cases} 
1 & \text{if } u = v \text{ and } d_i \neq 0 \\
0 & \text{if } u \neq v \text{ and } v \text{ are adjacent} \\
\frac{-1}{n} & \text{otherwise}
\end{cases}
\]

Alternatively, we can define \( \mathcal{L} \) in terms of other matrices as \( \mathcal{L} = I - T^{-1/2} AT^{-1/2} \) where \( T \) is the diagonal matrix of degrees. The eigenvalues of \( \mathcal{L} \) are denoted in ascending order as \( 0 = \lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_n \).

Then, the spectral gap is defined to be \( \lambda_2 \).

Numerical Results

**Model A:** \( P \propto \begin{pmatrix} 1 & \gamma \\ \gamma & 1 \end{pmatrix} \)

- For \( \gamma \leq 1 \), \( \lambda_2 = \frac{2 \gamma}{1+\gamma} \)
- For \( \gamma \geq 1 \), \( \lambda_2 \) is undefined

**Betti 1**

- \( \gamma = 0.06 \)
- \( \gamma = 0.50 \)
- \( \gamma = 1.00 \)

**Betti 2**

- \( \gamma = 0.06 \)
- \( \gamma = 0.50 \)
- \( \gamma = 1.00 \)

**Model B:** \( P \propto \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \)

- For \( \gamma \leq 1 \), \( \lambda_2 = \frac{2 \gamma}{1+\gamma} \)
- For \( \gamma \geq 1 \), \( \lambda_2 \) is undefined

**Betti 1**

- \( \gamma = 0.06 \)
- \( \gamma = 0.50 \)
- \( \gamma = 1.00 \)

**Betti 2**

- \( \gamma = 0.06 \)
- \( \gamma = 0.50 \)
- \( \gamma = 1.00 \)

Questions

- Can we adapt the spectral methods of Hoffman, Kahle and Paquette to derive threshold functions for vanishing homology for the SBM?
- For certain cases, the clique complex has nonvanishing homology at its saturation point (when one part of the graph becomes fully connected). For which values of \( \gamma \) is there non-vanishing homology with high probability at its saturation point?
- There exist more than two transitions (see the Betti 2 curves) in some cases; how can we detect these intermediate regions with vanishing homology?

References


Acknowledgements

Work supported in part by the Basic Research Office of the ASD&R&E through ONR N00014-16-1-2010 and the Fulbright Canada program.