Clique Topology of the Stochastic Block Model

Darrick Lee[1], **Robert Ghrist**[1,2], **Danielle Bassett**[2,3]

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

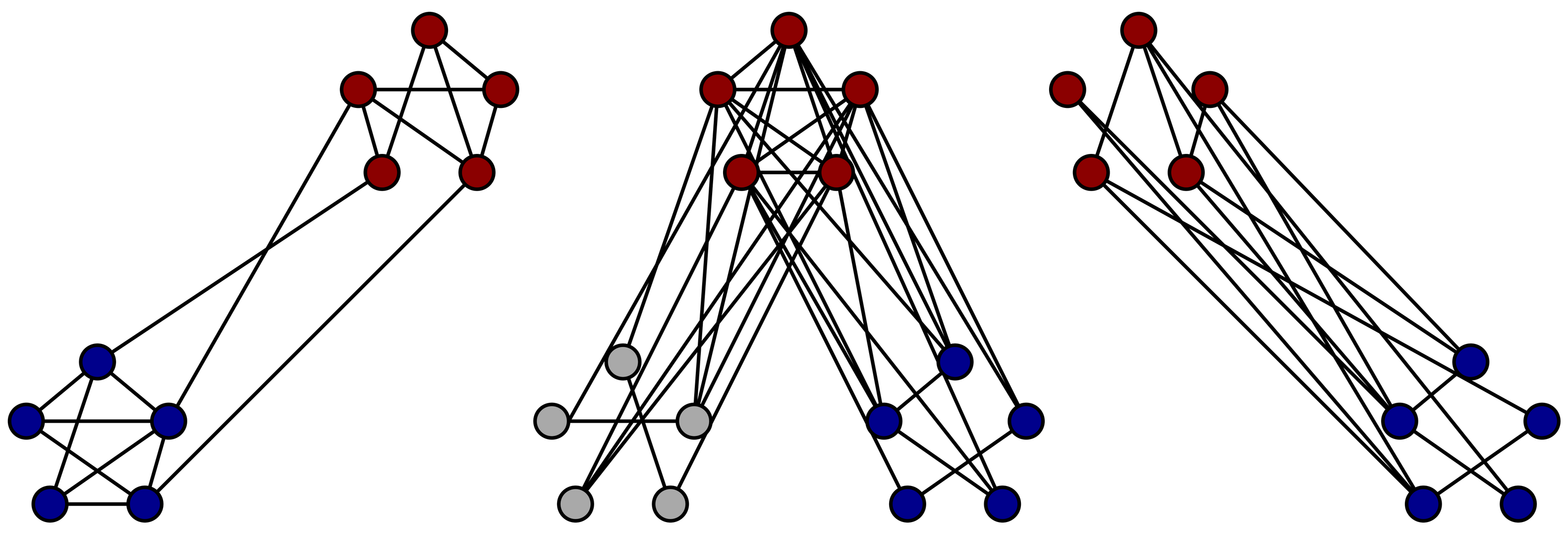
Introduction

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 Erdös-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:

- $n \in \mathbb{N}$: the number of vertices
- $k \in \mathbb{N}$: the number of blocks
- $\{q_i\}_{i=1}^n, q_i \in \{1, 2, \ldots, k\}$: 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_v denote the degree of vertex v. Then

$$\mathcal{L}(u,v) = \begin{cases} 1 & \text{if } u = v \text{ and } d_v \neq 0 \\ \frac{-1}{\sqrt{d_u d_v}} & \text{if } u \text{ and } v \text{ are adjacent} \\ 0 & \text{otherwise.} \end{cases}$$

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 λ_2.

Previous Work and Approach

Sharp threshold functions for vanishing and non-vanishing homology for the Erdös-Renyi graph have been shown by Kahle [5], and is summarized by the following theorem.

Theorem (Clique Complex Topology of Erdös-Renyi Graphs) Let $k \geq 1$ and $\epsilon > 0$ be fixed, and let $X(n,p)$ be the clique complex generated by the Erdös-Renyi graph $G(n,p)$.

1. If
$$p \geq \left(\left(\frac{k}{2} + 1 + \epsilon\right)\frac{\log n}{n}\right)^{1/(k+1)}$$
then w.h.p. $H^k(X, \mathbb{Q}) = 0$.
2. If
$$\left(\frac{k+1+\epsilon}{n}\right)^{1/k} \leq p \leq \left(\left(\frac{k}{2} + 1 - \epsilon\right)\frac{\log n}{n}\right)^{1/(k+1)}$$
then w.h.p. $H^k(X, \mathbb{Q}) \neq 0$.

The following theorem connects the cohomology of a simplicial complex with the spectral gap of the underlying graph [1].

Theorem (Cohomology Vanishing Theorem) Let Δ be a pure D-dimensional finite simplicial complex such that for every $(D-2)$-dimensional face σ, the link $\text{lk}_\Delta(\sigma)$ and has spectral gap

$$\lambda_2[\text{lk}_\Delta(\sigma)] > 1 - \frac{1}{D}.$$

Then $H^{D-1}(\Delta, \mathbb{Q}) = 0$.

With this theorem, the first step is to understand and bound the spectral gap of the SBM. Several authors have studied the spectral gap of the Erdös-Renyi graph. We note that $T^{1/2}\mathbf{1}$ is an eigenvector of $\mathcal{L}$ with eigenvalue 0. Suppose $\mathcal{S} = \{\mathbf{v} \in \mathbb{R}^n : \mathbf{v} \perp T^{1/2}\mathbf{1} : \|\mathbf{v}\| \leq 1\}$, then $\lambda_2(\mathcal{L}) = \min_{\mathbf{v} \in \mathcal{S}} \langle \mathbf{v}, \mathcal{L}\mathbf{v}\rangle$.

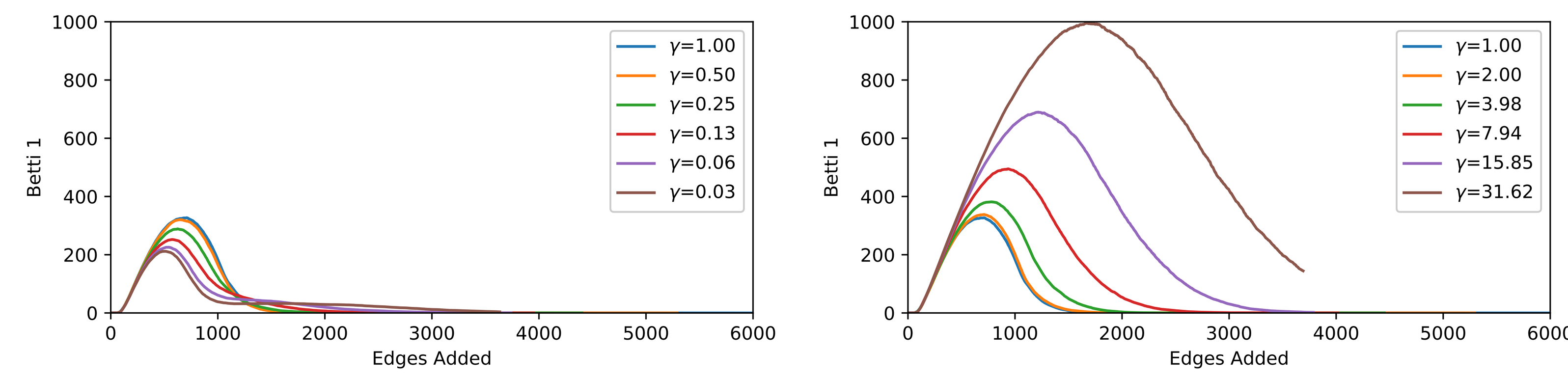
To bound this, one approach is to first bound the contribution from the adjacency matrix A. The result for the Erdös-Renyi graphs is $\max_{(\mathbf{v},\mathbf{w}) \in \mathcal{S}\times\mathcal{S}} |\langle \mathbf{v}, A\mathbf{w}\rangle| \leq c\sqrt{np}$, with probability at least $1 - O(n^{-\alpha})$, where c is a constant. A common technique is due to Kahn and Szemeredi [3] in which the summands of $\langle \mathbf{v}, A\mathbf{w}\rangle = \sum_{i,j} v_i A_{i,j} w_j$ are considered in two cases.

- **Light couples** ($|v_i w_j| \leq \sqrt{p/n}$) whose contribution can be bounded using a concentration inequality
- **Heavy couples** ($|v_i w_j| > \sqrt{p/n}$) whose contribution is bounded using properties of the random graph such as bounded degree and discrepancy

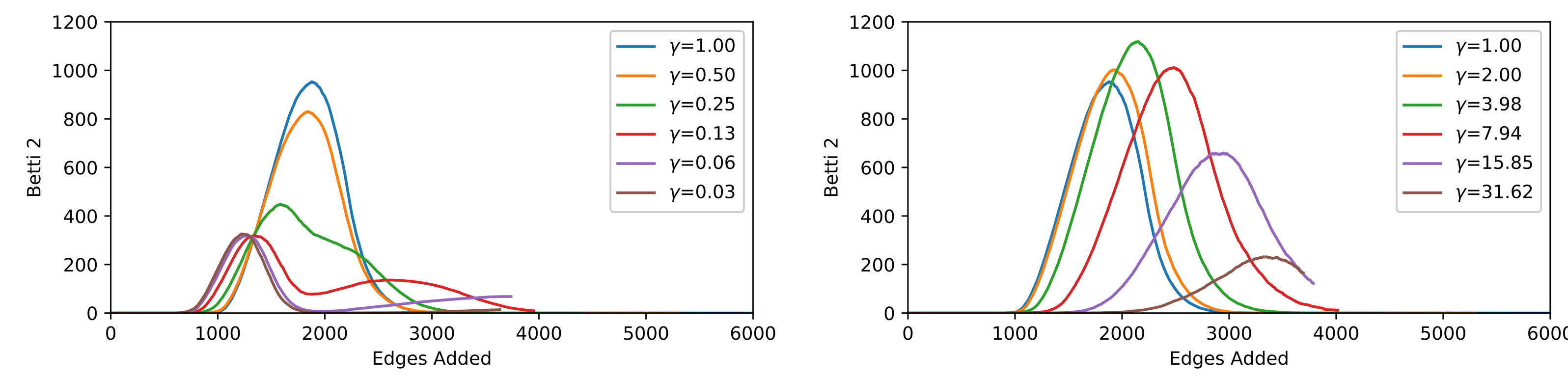
Numerical Results

Model A: $P \propto \begin{pmatrix} 1 & \gamma \\ \gamma & 1 \end{pmatrix}$

Betti 1

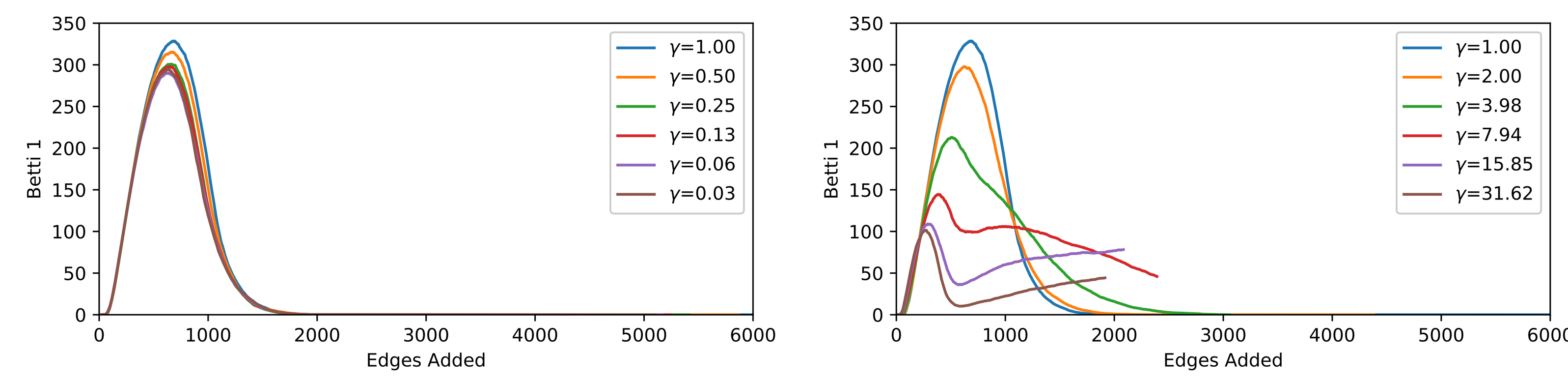


Betti 2

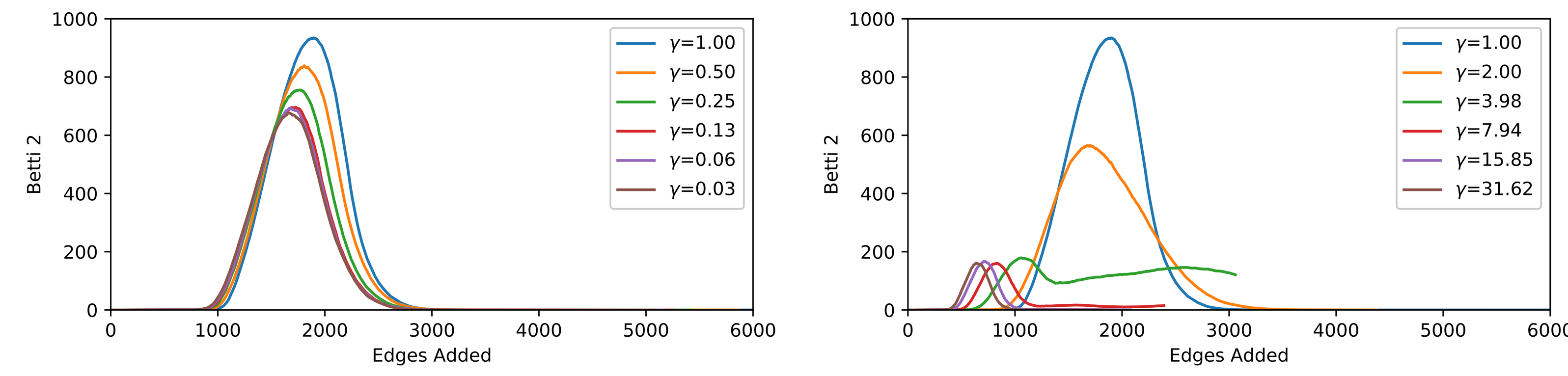


Model B: $P \propto \begin{pmatrix} 1 & 1 \\ 1 & \gamma \end{pmatrix}$

Betti 1



Betti 2



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 γ 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

[1] W. Ballmann and J. Światkowski. On L^2-cohomology and Property (T) for Automorphism Groups of Polyhedral Cell Complexes. *Geometric and Functional Analysis*, 7(4):615–645, August 1997.

[2] Richard F. Betzel, John D. Medaglia, and Danielle S. Bassett. Diversity of meso-scale architecture in human and non-human connectomes. *arXiv:1702.02807 [q-bio]*, February 2017.

[3] J. Friedman, J. Kahn, and E. Szemerdi. On the Second Eigenvalue of Random Regular Graphs. In *Proceedings of the Twenty-first Annual ACM Symposium on Theory of Computing*, STOC '89, pages 587–598, New York, NY, USA, 1989. ACM.

[4] Christopher Hoffman, Matthew Kahle, and Elliot Paquette. Spectral gaps of random graphs and applications. *arXiv:1201.0425 [math]*, January 2012. arXiv: 1201.0425.

[5] Matthew Kahle. Sharp vanishing thresholds for cohomology of random flag complexes. *Annals of Mathematics*, 179(3):1085–1107, 2014.

[6] Bernadette J. Stolz, Heather A. Harrington, and Mason A. Porter. Persistent homology of time-dependent functional networks constructed from coupled time series. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 27(4):047410, April 2017.

Acknowledgements

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