

## 1 Introduction

My research interests are in the field of probabilistic combinatorics. My thesis research focuses on properties of random graphs, especially spectral properties. The study of these sorts of problems incorporates tools from probability theory, and in particular the study of random matrices, as well as combinatorics and graph theory.

In one-dimensional probabilistic combinatorics, theorems regarding concentration of a random variable on its expectation are key tools. For example, theorems such as the Chernoff bounds, Bernstein's Inequality, and many others, provide an upper bound on the probability that a sum of random variables deviates from its expectation. If this probability is sufficiently low, one can often replace a random variable by its expected value with a relatively small error term. Much of my current work focuses on deriving and applying similar concentration inequalities for matrices whose entries are random variables. In my thesis, I discuss the application of one such inequality to the spectrum of random graphs, providing concentration of the spectrum on that of the weighted expectation graph for random graphs with independent edges.

The motivation to study random matrices comes from many different fields of mathematics, including operator theory, analysis, and combinatorics. In my work, random matrices often represent matrices associated to random graphs. Given a graph  $G$  on  $n$  vertices, the adjacency matrix  $A$  of  $G$  is the  $n \times n$  matrix with a 1 in the  $ij$  position if  $v_i$  is adjacent to  $v_j$  and 0 otherwise. The spectrum of the  $A$  has been linked to various topological features of the graph, such as connectivity and the occurrence of certain subgraphs [3], [1]. If the edges of a graph appear randomly, the entries of the adjacency matrix are random variables, and we can use concentration inequalities as described to approximate the eigenvalues of  $A$ .

For a graph  $G$ , the degree matrix  $D$  is a diagonal matrix whose  $ii$  entry is the number of nodes to which  $v_i$  is adjacent, the degree of  $v_i$ . The (normalized) Laplacian matrix for  $G$  is given by  $L = I - D^{-1/2}AD^{-1/2}$ . The spectrum of the Laplacian is also tied to connectivity properties, as well as the study of random walks on graphs and approximation algorithms [1]. For a random graph, we use concentration inequalities to approximate the eigenvalues of  $L$ .

## 2 Current Work

Suppose  $X$  is an  $n \times n$  dimensional Hermitian matrix, whose entries are independent random variables (excepting those dependencies required for the matrix to be Hermitian, in particular, that  $X_{ij} = \overline{X_{ji}}$ ). Define the expectation of  $X$  to

be taken coordinate-wise, so that  $E(X)_{ij} = E(X_{ij})$ , and the variance of  $X$  to be  $\text{var}(X) = E[(X - E(X))^2]$ . Let  $\|\cdot\|$  denote the spectral norm of a matrix, so  $\|X\| = \max |\lambda|$ , where the max is taken over all eigenvalues of  $X$ . We note that in the space of Hermitian matrices,  $\|\cdot\|$  is a norm, although it is not a norm in the space of all matrices. The following is a generalization of Bernstein's Inequality to random matrices, found in [5].

**Theorem 2.1.** *Let  $X_1, X_2, \dots, X_m$  be independent random  $n \times n$  Hermitian matrices. Moreover, assume that  $\|X_i - E(X_i)\| \leq M$  for all  $i$ , and put  $v^2 = \|\sum \text{var}(X_i)\|$ . Let  $X = \sum X_i$ . Then for any  $a > 0$ ,*

$$\text{pr}(\|X - E(X)\| > a) \leq 2n \exp\left(-\frac{a^2}{2v^2 + 2Ma/3}\right).$$

Now, suppose we have a random graph  $G$  whose edges appear independently, where the probability of edge  $v_i v_j$  is  $p_{ij}$ . Take  $X_{ij}$  to be a random Hermitian matrix, with entries  $ij$  and  $ji$  equal to 1 with probability  $p_{ij}$  and 0 otherwise, and all other entries 0. In this way,  $X_{ij}$  represents edge  $ij$ , and the adjacency matrix for  $G$  is given by  $A = \sum_{i \leq j} X_{ij}$ . Applying the above theorem, we obtain the following result:

**Theorem 2.2.** [2] *Let  $G$  be a random graph, where  $\text{pr}(v_i \sim v_j) = p_{ij}$ , and each edge is independent of each other edge. Let  $A$  be the adjacency matrix of  $G$ , so  $A_{ij} = 1$  if  $v_i \sim v_j$  and 0 otherwise, and  $\bar{A} = E(A)$ , so  $\bar{A}_{ij} = p_{ij}$ . Let  $\Delta$  denote the maximum expected degree of  $G$ . Let  $\epsilon > 0$ , and suppose that for  $n$  sufficiently large,  $\Delta > \frac{4}{9} \ln(2n/\epsilon)$ . Then with probability at least  $1 - \epsilon$ , for  $n$  sufficiently large, the eigenvalues of  $A$  and  $\bar{A}$  satisfy*

$$|\lambda_i(A) - \lambda_i(\bar{A})| \leq \sqrt{4\Delta \ln(2n/\epsilon)}$$

for all  $1 \leq i \leq n$ .

We note that the bound holds simultaneously for all eigenvalues with probability  $1 - \epsilon$ . Moreover, the restriction on  $\Delta$  is minor, as most random graph families arising from real-world applications have a maximum expected degree of  $O(n)$ .

Moreover, let  $\bar{G}$  be a weighted complete graph where the weight of an edge  $v_i v_j$  is the expectation of that edge occurring in  $G$ . We can think of  $\bar{G}$  as the expectation of  $G$ . Take  $\bar{L}$  to be the Laplacian for  $\bar{G}$ , so  $\bar{L} = I - \bar{D}^{-1/2} \bar{A} \bar{D}^{-1/2}$ . We note that  $\bar{L} \neq E(L)$ , however, as the following theorem shows,  $\bar{L}$  does capture useful information about the spectrum of  $L$ .

**Theorem 2.3.** [2] *Let  $G$  be a random graph, where  $\text{pr}(v_i \sim v_j) = p_{ij}$ , and each edge is independent of each other edge. Let  $\delta$  be the minimum expected degree of  $G$ , and  $L = I - D^{-1/2}AD^{-1/2}$  the (normalized) Laplacian matrix for  $G$ . Choose  $\epsilon > 0$ . Then there exists a constant  $k = k(\epsilon)$  such that if  $\delta > k \ln n$ , then with probability at least  $1 - \epsilon$ , the eigenvalues of  $L$  and  $\bar{L}$  satisfy*

$$|\lambda_j(L) - \lambda_j(\bar{L})| \leq 3\sqrt{\frac{3 \ln(4n/\epsilon)}{\delta}}$$

for all  $1 \leq j \leq n$ , where  $\bar{L} = I - \bar{D}^{-1/2}\bar{A}\bar{D}^{-1/2}$ .

Again, as above, this bound holds simultaneously for all eigenvalues of  $L$  with probability  $1 - \epsilon$ . The restriction on the minimum expected degree can be seen as a density requirement, ensuring a sufficient number of edges in the graph.

These two theorems are particularly useful due to their generality. The degree requirements, as noted, are fairly minor, and thus these theorems can be applied to many different random graph models. Moreover, calculating the spectra of  $\bar{A}$  and  $\bar{L}$  is often straightforward, as these are fixed matrices, often with useful algebraic properties.

We note that for some models, these theorems are not optimal. For well studied graph models such as  $G_{n,p}$ , where every edge appears independently and with probability  $p$ , techniques that are specific to the structure of the model are better. However, these theorems have been used to approximate the spectra of graphs that have not been studied or obtained with other techniques. Indeed, we have used these two theorems to approximate the spectra of graphs with a given expected degree sequence [2], as well as Multiplicative Attribute Graphs and Stochastic Kronecker Graphs.

Multiplicative Attribute Graphs are random graphs in which the vertices are a random multi-subset of length  $n$  words in a finite alphabet  $\Gamma$ , and the probability of an edge between two vertices is determined by comparing the corresponding words coordinatewise. Analyzing such a graph using traditional techniques is difficult, as the vertex set and the edge set are both random. However, by using the concentration results stated, we have been able to produce an approximation of the spectra of the Laplacian for certain types of MAGs. Stochastic Kronecker Graphs are a related graph model in which the vertex set is the set of all length  $n$  words in  $\Gamma$ ; we have also approximated the spectra of these graphs.

### 3 Future Directions

There are many unanswered questions in the use of random matrices to describe random graphs. The results described above give information about the eigenvalues of a random graph, but do not reveal anything about the distribution of the eigenvalues. Wigner's Semicircular Law [6] says that the eigenvalues of a random matrix are distributed according to the semicircular law when the off-diagonal entries are independent and identically distributed, but this is often not the case for random graphs. Different graph models have often required individualized methods to determine an empirical distribution for eigenvalues. A general theorem similar to the concentration theorems above is yet to be established.

Another interesting path for future work is to continue to investigate Multiplicative Attribute Graphs and Stochastic Kronecker Graphs. While the results cited give information about their eigenvalues and connectivity, there are still many open questions. For Multiplicative Attribute Graphs, what conditions will ensure a giant component, that is, a component of size  $\Theta(n)$ ? For Stochastic Kronecker Graphs, we have established a necessary and sufficient condition [4], but there are still lingering questions. What is the behavior of the Stochastic Kronecker Graph near the critical threshold for a giant component? With  $G_{n,p}$ , there are various component structures that emerge near this threshold, and similar structure should be present in the Stochastic Kronecker Graph. Moreover, the component structure, other than the emergence of the giant component, is yet to be studied. How large is the second-largest component? How many components should the graph have?

### References

- [1] F. Chung, *Spectral Graph Theory*, AMS Publications, 1997.
- [2] F. Chung and M. Radcliffe. On the Spectra of General Random Graphs. *Electronic Journal of Combinatorics*. **18(1)** (2011), P215.
- [3] D. M. Cvetković, M. Doob, and H. Sachs, *Spectra of Graphs, Theory and Applications*, Academic Press, 1980.
- [4] P. Horn and M. Radcliffe. Giant components in Kronecker graphs. *Random Structures and Algorithms*. To appear.
- [5] J. Tropp, User-Friendly Tail Bounds for Sums of Random Matrices. Available at <http://arxiv.org/abs/1004.4389>

- [6] E. Wigner, On the Distribution of the Roots of Certain Symmetric Matrices. *Ann. of Math.* **67** (1958), pp. 325-328.