

Asymptotic Behavior of the Random 3-Regular Bipartite Graph

Tim Novikoff^{*}
Courant Institute
New York University, 10012

December 17, 2002

Abstract

In 2001, two numerical experiments were performed to observe whether or not the second largest eigenvalue of the adjacency matrix for the random cubic bipartite graph approaches $2\sqrt{2}$ as the size of the graph increases. In the first experiment, by Kevin Chang, the graphs were chosen using an algorithm that constructed entirely new graphs at each step using three random permutations, in contrast to the second experiment, by Peter Richter, which used a random walk in the space of simple cubic connected bipartite graphs. Although the walk in Richter's experiment was random, in that two randomly chosen edges were swapped, the eigenvalues of the graphs from two consecutive steps of the walk are shown here to be correlated. A walk in which the eigenvalues are uncorrelated is used here in a similar experiment. In addition, an experiment similar to Kevin Chang's experiment is performed in which graphs are constructed using an algorithm that is proven to choose uniformly at random from the space of simple cubic connected bipartite graphs. The distributions of the eigenvalues, after being normalized to have mean zero and standard deviation one, appear to be stable but not symmetric, similar to the Tracy-Widom distributions. The mean and standard deviation appear to approach $2\sqrt{2}$ and zero, respectively, according to power laws with the mean approaching quicker than the standard deviation.

^{*}E-mail: tn8@nyu.edu

From this, a conjecture is made that the probability that a random connected cubic bipartite graph is Ramanujan approaches .52 as the size of the graph increases to infinity.

1 Introduction

This report is the culmination of a new course at NYU called Undergraduate Math Laboratory, in which we learned about some unsolved problems in pure math, and performed numerical experiments to investigate the problems. The results of this paper pertain to a particular problem in graph theory, and so we start with an introduction to the theory needed to understand the results

1.1 Graph Theory

A graph G is described by a set of vertices V and a set of edges E , where elements of E are pairs of vertices $(v_i, v_j) = (v_j, v_i)$. Any two vertices, v_i and v_j , are said to be *adjacent* if and only if the pair (v_i, v_j) is an element of E . A graph is considered to be *simple* if all the elements of E are distinct (no multiple edges), and is considered to be *connected* if the vertices cannot be partitioned into two sets, A and B , such that no vertex from A is adjacent to a vertex from B . A graph is said to be *bipartite*, or *bicolourable*, if one could associate one of two colors with every vertex, and yet have no two adjacent vertices be associated with the same color. Finally, a graph is said to be k -regular if every vertex is adjacent to exactly k vertices.

Associated with a graph G is its *adjacency matrix*, $A(G)$. The adjacency matrix is constructed very simply: label the vertices $v_1 \dots v_n$, and in the entries (i, j) and (j, i) of A , put the multiplicity of the edge $(v_i, v_j) = (v_j, v_i)$. Thus, for a simple graph, the adjacency matrix contains only zeros and ones. Furthermore, the sum of any row or column of the adjacency matrix for a simple k -regular graph is k . (Note that $A(G)$ is not unique, for it depends on the labeling of the vertices. The eigenvalues of $A(G)$ remain the same regardless of the labeling, however, and that is all that we will be interested in here.)

Also associated with a graph G is the *expanding constant* $h(G)$, a measure of the connectivity of G . Graphs with large expanding constants are desirable in many applications of graph theory, including information networks, where h is a measure of how quickly information propagates on a network.

It can be shown that if a k -regular graph G is connected and bipartite, then the eigenvalues of $A(G)$ will form a set that is symmetric about the origin, with k as the unique largest element. Furthermore, $h(G)$ can be estimated using the second largest eigenvalue of A , $\lambda_1(G)$, with the Cheeger-Buser inequalities:

$$\frac{k-1}{2} \leq h(G) \leq 2\sqrt{2k(k-\lambda_1(G))} \quad (1)$$

In fact, G need not be bipartite for the above inequality to hold.

It can also be shown that for any sequence, or *family*, G_m of connected k -regular graphs such that the size (number of vertices) of the graphs goes to infinity,

$$\liminf_{m \rightarrow \infty} \lambda_1(G_m) \geq 2\sqrt{k-1} \quad (2)$$

Thus a family of k -regular graphs G_m which satisfies the condition $\lambda_1(G_m) \leq 2\sqrt{k-1}$ for all m is the ideal family in terms the bounds placed on $h(g)$ by (1). Such a family is called a *Ramanujan Family*, and in fact any graph G with the property $\lambda_1(G) \leq 2\sqrt{k-1}$ is said to be a *Ramanujan Graph*.

1.2 The Tracy-Widom Distributions

The Tracy-Widom distributions were discovered in 1995 by Craig Tracy and Harold Widom. In their paper, On Orthogonal and Symplectic Matrix Ensembles, they derive explicitly the probability density functions (see Figure 1) for the largest eigenvalue of a random member of three classes of Gaussian ensembles (orthogonal, unitary, and symplectic) in terms of a *Painlevé* function. Although these probability density functions apply only in the limit as the size of the matrices goes to infinity, computer simulations indicate that they are good approximations of the probability density functions for finite matrices of size $N \geq 200$.

2 Statement of Problem

Given the importance of Ramanujan graphs, a natural question arises: *What is the probability that the random k -regular graph is Ramanujan?* Here, two experiments are performed in order to formulate a conjecture regarding the answer to this problem for the set of 3-regular graphs that are simple, connected, and bipartite.

The first experiment is a refinement of an experiment described in *Richter 2001*. In Richter’s experiment, graphs were chosen by constructing an initial simple, bipartite, connected 3-regular graph, calculating the second largest eigenvalue of the adjacency matrix, swapping two randomly chosen edges, calculating the second largest eigenvalue of the new adjacency matrix, swapping two randomly chosen edges again, and so on. Figure 1 shows the autocorrelation of a sequence of eigenvalues obtained using this method. Clearly one needs to do many more than one swap to get uncorrelated eigenvalues. The first experiment here is performed using a similar random walk, but with a number of edge-swaps at each step such that the eigenvalues obtained are relatively uncorrelated (see *Methods*).

The second experiment is similar to the experiment in *Chang 2001* except that the algorithm used here is *proven* to choose uniformly at random from the set of simple, bipartite, connected, 3-regular graphs. Also, much larger graphs are used here (up to 20,000; Chang only went up to 1,000) and a more careful statistical analysis of the data is presented here.

3 Methods

3.1 Experiment 1: The Random Walk

The sizes of the graphs, which ranged from 200 to 2,000, were logarithmically spaced. To determine how many swaps were to be used to get uncorellated eigenvalues for a given graph size, the first value at which the autocorrelation function achieves zero was taken. Then, to be on the safe side, the number was increased until it was larger than the number of swaps to be used for all smaller graphs. The original experiment is then repeated, with the appropriate number of edge-swaps performed at each step. The eigenvalues were calculated to machine precision using the *eigs* function in MATLAB, which implements the Arnoldi Method through ARPACK.

3.2 Experiment 2: Choosing Graphs Uniformly at Random

The sizes of the graphs, which ranged from 200 to 20,000, were logarithmically spaced. For the graph sizes up to 2000, 10,000 graphs were constructed, and for sizes above 2000, only 1000 graphs were constructed, because the time needed to compute more eigenvalues of the large adjacency matrices would have exceeded the time constraint of this project. The eigenvalues were calculated to machine

precision using the the *eigs* function in MATLAB, which implements the Arnoldi Method through ARPACK.

3.3 The Configuration Model Algorithm

The algorithm that was used to randomly sample the space of simple 3-regular bipartite graphs is based on Bollobas' "configuration model" for simple k -regular graphs (not necessarily bipartite). The configuration model algorithm was altered so that bipartite graphs were produced. The altered algorithm is as follows:

Let N be the size of the graph, and k be the degree of regularity.

1. Create two vectors x_1 and x_2 , each random permutations of the integers from 1 to $Nk/2$.
2. Reassign all the entries in the vectors with their values mod $N/2$, thus making each vector have exactly three instances of each integer from 1 to $N/2$.
3. Add $N/2$ to all the entries in x_2 , so that x_2 has exactly three instances of each of the integers from $(N/2 + 1)$ to N .
4. Construct the graph by defining vertices $v_1, v_2 \dots v_n$, and defining edges from the vertices labelled by corresponding entries in x_1 and x_2 . (Thus if the first entries of x_1 and x_2 are a and b , then an edge is defined connecting v_a to v_b .)
5. If the graph is not simple, repeat steps 1-4 until a simple graph is created.

Since any simple, k -regular bipartite graph can be constructed from exactly $k!^{N/2}$ choices of x_1 and x_2 , this algorithm will choose uniformly at random (u.a.r.) from the space of simple, k -regular, bipartite graphs so long as the permutations in Step 1 are chosen u.a.r. It should be noted that without Step 5, the algorithm would NOT choose u.a.r. from the space of 3-regular bipartite graphs. This is because graphs with multiple edges can be constructed from less possible choices of x_1 and x_2 than simple graphs. (In fact, the number of choices for a simple graph exceeds the number of choices for a non-simple graph with m k -tuples by a factor of $k!m$.)

4 Results

Based on the results of the experiments (see Figures), several conjectures can be made:

1. For future experiments, edge-swapping can be used to obtain a sampling of eigenvalues similar to the one obtained when choosing uniformly at random, provided that the number of edge-swaps between graph samples yields uncorrelated eigenvalues. Support for this is the similarity of the results obtained in the two experiments: the difference in the standard deviations and the means obtained from the two experiments were negligible. Comparing Figures 2 and 3 with Figures 5 and 6 is convincing of this as well.
2. The distribution of the second largest eigenvalue of the adjacency matrix for a random simple, connected, bipartite, 3-regular graph is stable (once normalized for mean and variance) and can be approximated by one of the similarly normalized Tracy-Widom distributions. This is supported by Figures 5 and 6 (and if you believe the first conjecture, by Figures 2 and 3). It is natural to also conjecture that the approximation becomes exact in the limit as the size of the graphs goes to infinity, although it is not clear which of the Tracy-Widom distributions it approaches, since they are so similar. One could speculate, however, that it is the distribution for the GOE, given the symmetry of the adjacency matrix.
3. The probability that a random simple, connected, bipartite, 3-regular graph is Ramanujan approaches a fixed probability, approximately .52, as the size of the graph goes to infinity. This is supported by the data in Figure 7. The reasoning is this: It appears that the standard deviation σ and the mean μ are well approximated by the equations

$$\sigma(N) = c_\sigma N^\alpha \quad (3)$$

and

$$\mu(N) = \mu_\infty + c_\mu N^\gamma \quad (4)$$

where $\alpha \approx -.697$, $\gamma \approx -.778$ and $\mu_\infty = 2\sqrt{2}$. If this is true then

$$\lim_{N \rightarrow \infty} \frac{\mu(N) - \mu_\infty}{\sigma(N)} = \lim_{N \rightarrow \infty} \frac{c_\mu}{c_\sigma} N^{\gamma - \alpha} = 0 \quad (5)$$

and so the distribution tends towards a distribution centered around $\mu_\infty = 2\sqrt{2}$. (In fact (5) holds so long as $\gamma < \alpha$.) So (3) and (4) are true approximations as N goes to infinity, and if the previous conjecture is true, then the probability of a random simple, connected, bipartite 3-regular graph being Ramanujan is the fraction of the area under the Tracy-Widom probability

density function which lies to the left of the mean. This was calculated numerically to be approximately .52.

It is important to note that the third conjecture is the least certain (though probably the most important, if true.) Linear fits were made for the first m data points, with m going from 3 to 20, and there was a general tendency for the difference $|\alpha - \gamma|$ to decrease as more data points were included, though it increased with the last couple of fits. This simply highlights how reliant the fits are on each data point, and thus the uncertainty of the third conjecture.

Given this uncertainty, it is also important to note what the probability of the random simple, connected, bipartite 3-regular graph being Ramanujan would be if the true values of γ and α turn out to be equal, or if γ turns out to be *greater* than α . In the case where $\gamma = \alpha$ (that is to say, as N goes to infinity, the γ and α needed to satisfy (3) and (4) approach the same limit), the probability would be some number between .52 and 1. This is because the limit in (5) would tend to $\frac{c_\mu}{c_\sigma}$, and so μ_∞ would tend to fall $\frac{c_\mu}{c_\sigma}$ standard deviations away from $\mu(N)$ in the limit as N approaches infinity. Thus the probability of the graph being Ramanujan would tend to the fraction of the area under the Tracy-Widom probability density function that lies to the left of $\frac{c_\mu}{c_\sigma}$ standard deviations to the right of the mean. The general rule, of course, is that the limit in (5) determines how many standard deviations to the right of the mean we must go before we reach the portion of the distribution that is not Ramanujan, as N goes to infinity. Thus, if $\gamma > \alpha$, the limit in (5) would be infinity, and the probability of the graph being Ramanujan would tend to 1.

5 Future Work

Much more work can be done to investigate the nature of these random graphs. The experiment which uses the altered Bollobas algorithm should be repeated with the unaltered algorithm in order to see if/how the results change if the graphs are not necessarily bipartite. Investigations into graphs with greater degrees of regularity should be conducted as well. As we increase the degree of regularity, however, the Bollobas algorithm will take much longer to produce simple graphs, so edge-swapping may be considered as an alternative method.

6 Acknowledgements

Thanks to Peter Sarnak for introducing me to this exciting experience. Thanks to Steven Miller for enlightening conversations. Thanks to Craig Tracy for sending me code to produce the Tracy-Widom distributions. Supreme thanks to Alex Barnett for insight and guidance at every step of the process.

References

- [Ch] Kevin Chang, *An Experimental Approach to Studying Ramanujan Graphs*, Unpublished Undergraduate Thesis, see Princeton Website. 2001
- [DSV] Giuliana Davidoff, Peter Sarnak, Alain Valette, *Elementary Number Theory, Group Theory, and Ramanujan Graphs*, Unpublished Notes.
- [JLR] Svante Janson, Tomasz Luczak, Andrzej Rucinski, *Random Graphs*, John Wiley & Sons, Inc., New York 2000.
- [Ri] Peter Richter, *An Investigation of Expanders and Ramanujan Graphs along Random walks of Cubic Bipartite Graphs*, Student Thesis, see Princeton Website. 2001.
- [Tra] Craig Tracy, Harold Widom, *On Orthogonal and Symplectic Matrix Ensembles*, Communications in Mathematical Physics, 1996

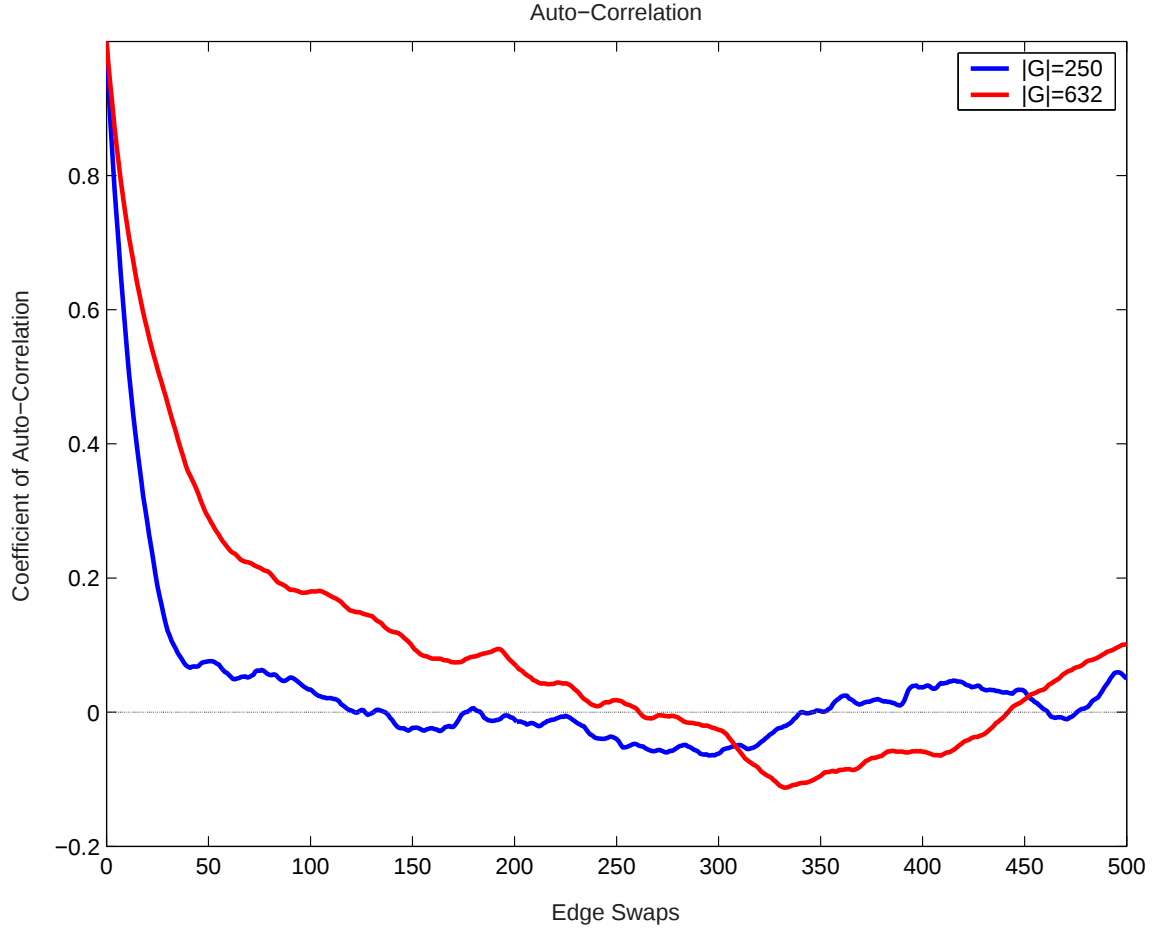


Figure 1: Shows the coefficient of autocorrelation for the sequence of $\lambda_1(G)$ found using Richter's experiment with single edge-swaps at every step (for two values of N). One can clearly see that consecutive eigenvalues are highly correlated, and the lag required for statistically uncorrelated eigenvalues is significant (on the order of 10^2 for these values of N).

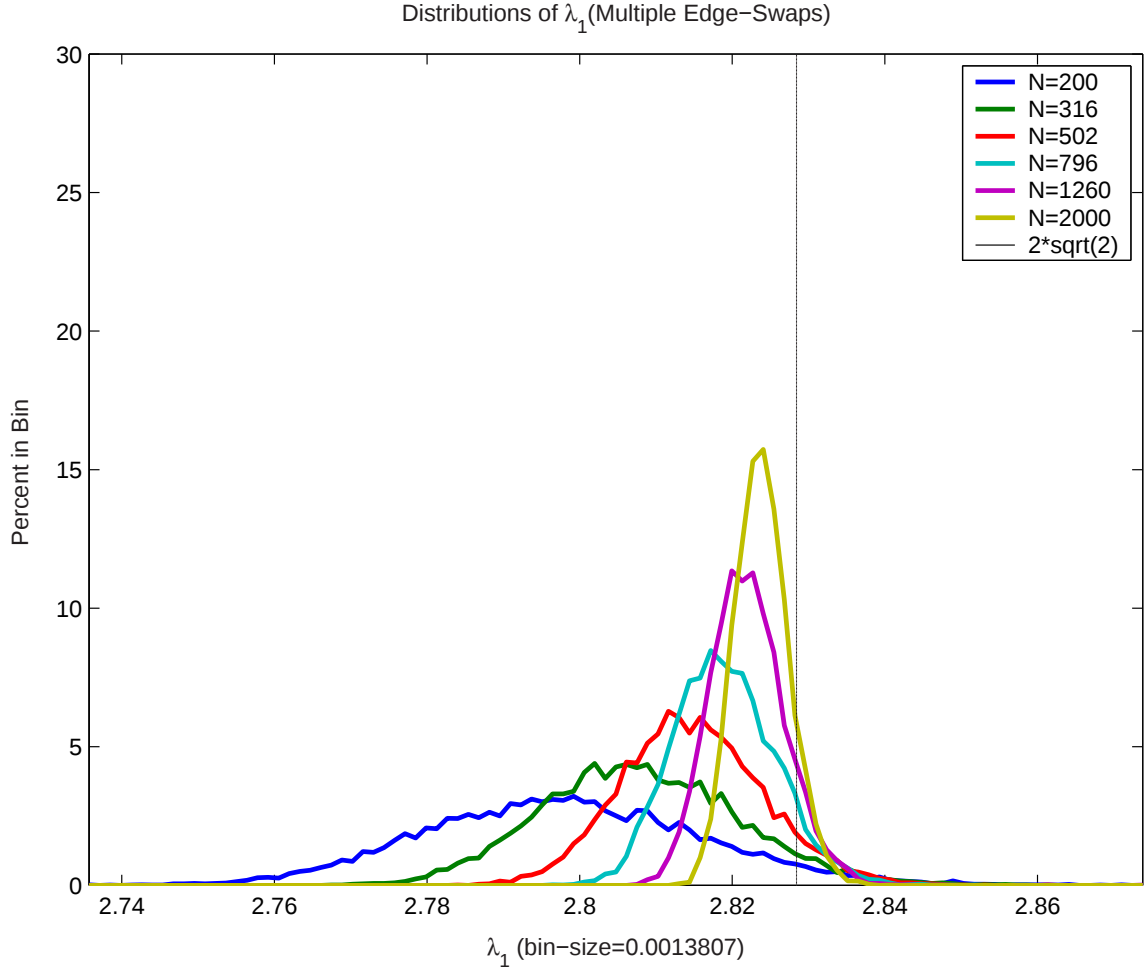


Figure 2: Shows the distributions of $\lambda_1(G)$ for a few different graphs sizes, collected using the appropriate number of edge-swaps at each step. Note that the distribution gets thinner as the graph size increases, and also that the distribution is slowly shifting to the right.

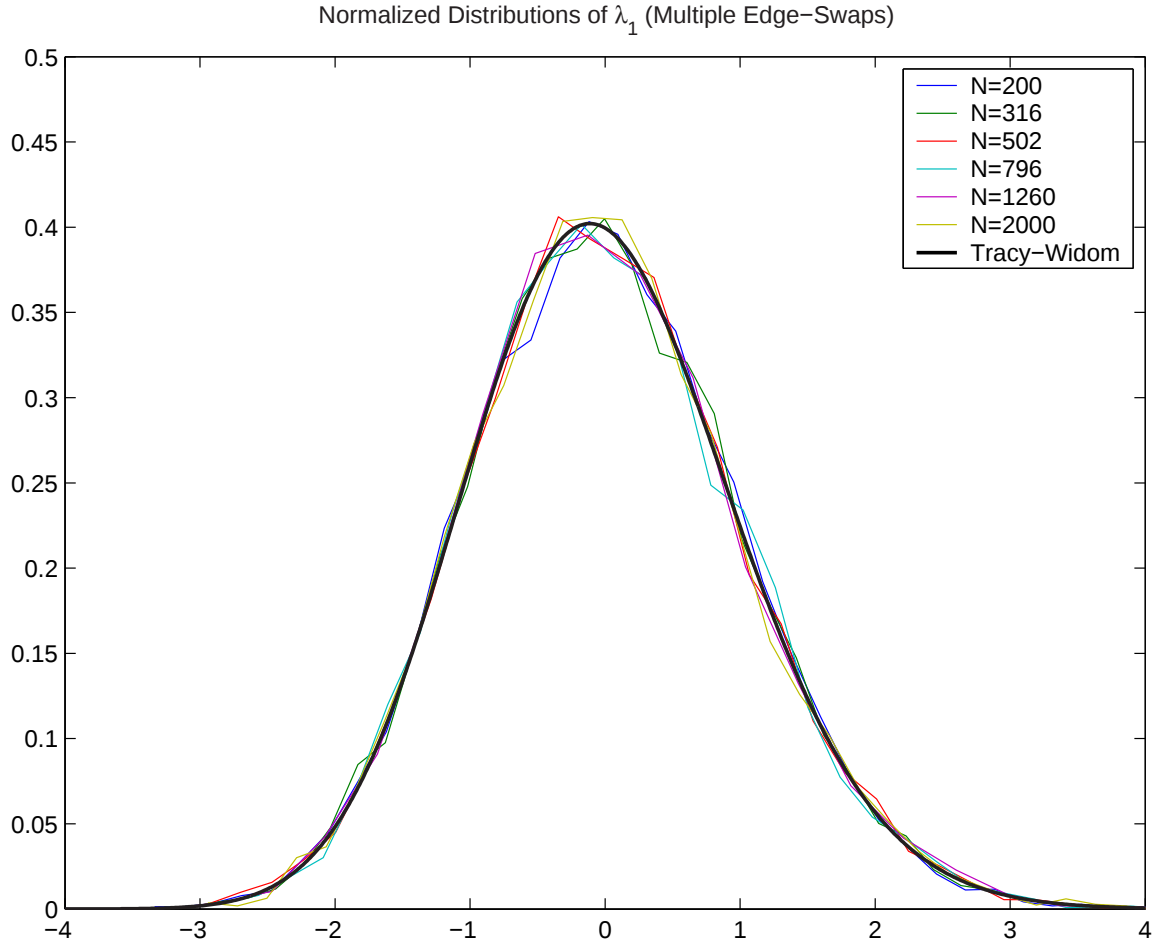


Figure 3: Shows the distributions of Figure 3, normalized to have mean=0, variance=1, and area=1. Plotted as well is the Tracy-Widom distribution for beta=1, but the distribution for beta=2 may as well have been plotted, for as one can see in Figure 4, the two are practically indistinguishable once normalized. The distributions of λ_1 are clearly not symmetric about their mean, but rather possess the same sort of asymmetry as the Tracy-Widom distributions.

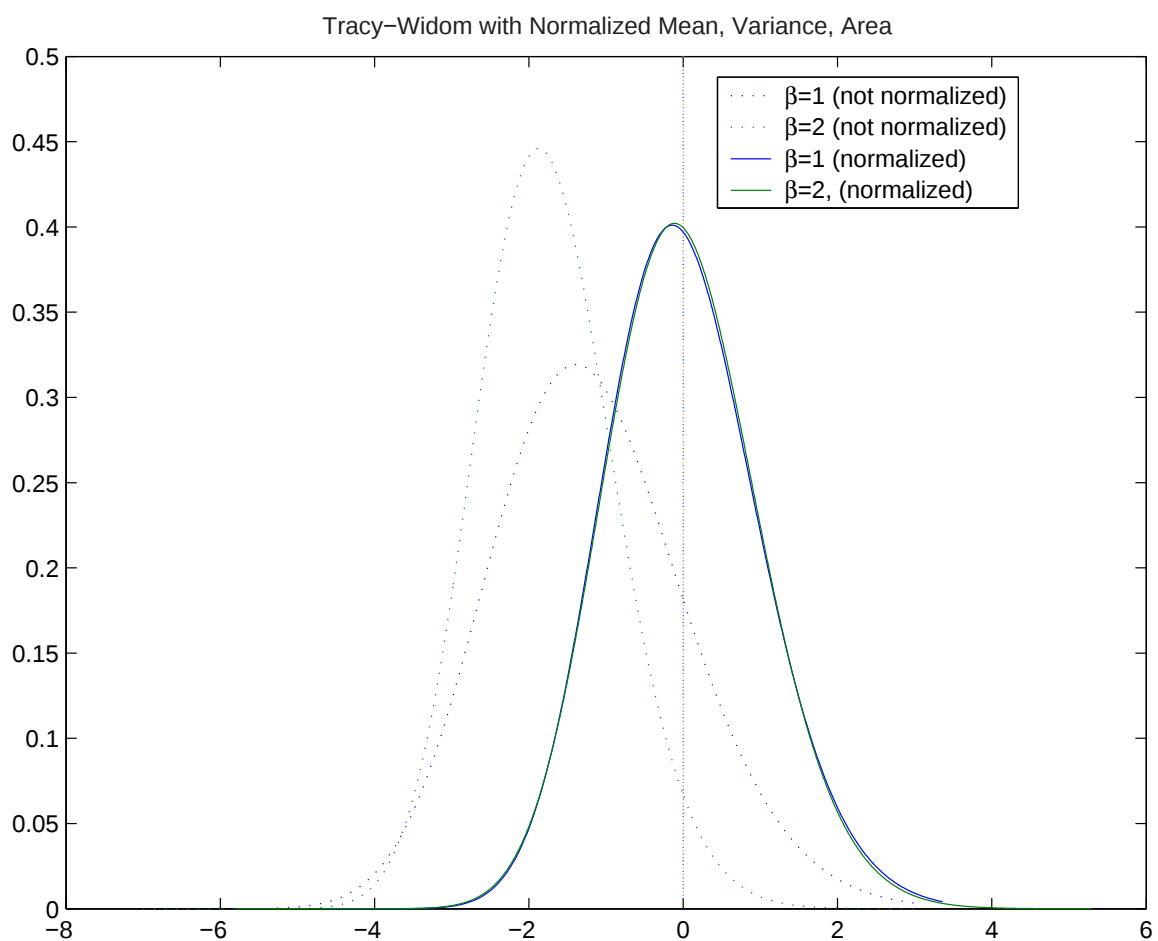


Figure 4: The dotted curves are the Tracy-Widom distributions for “ $\beta = 1$ ” (GOE) and “ $\beta = 2$ ” (GUE), as they were presented in the original paper by Tracy et al, and the solid curves are the same curves normalized to have mean=0, variance=1, and area=1.

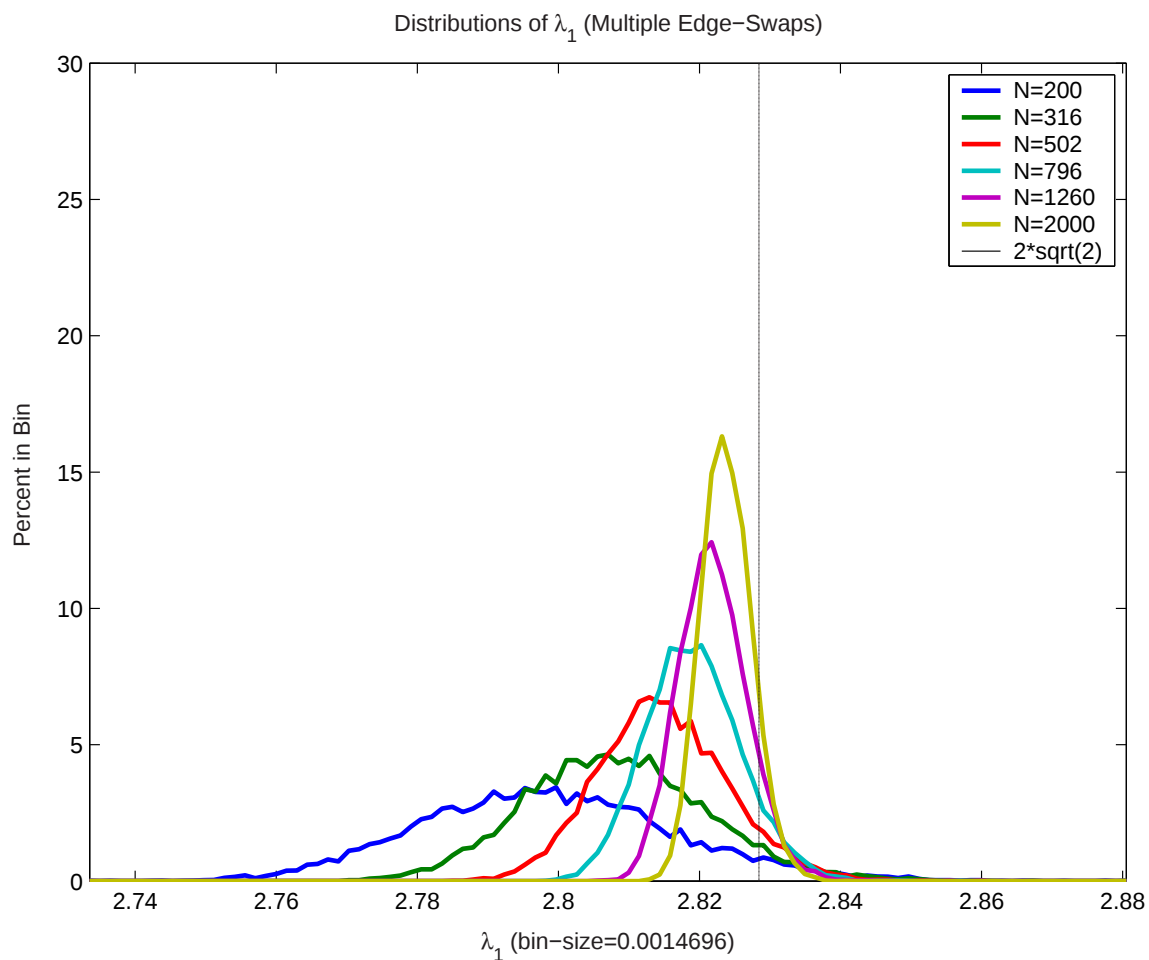


Figure 5: This figure is analogous to Figure 2, except that the data here was collected using the modified Bollobas algorithm, as opposed to the random walks used to collect the data in Figure 2.

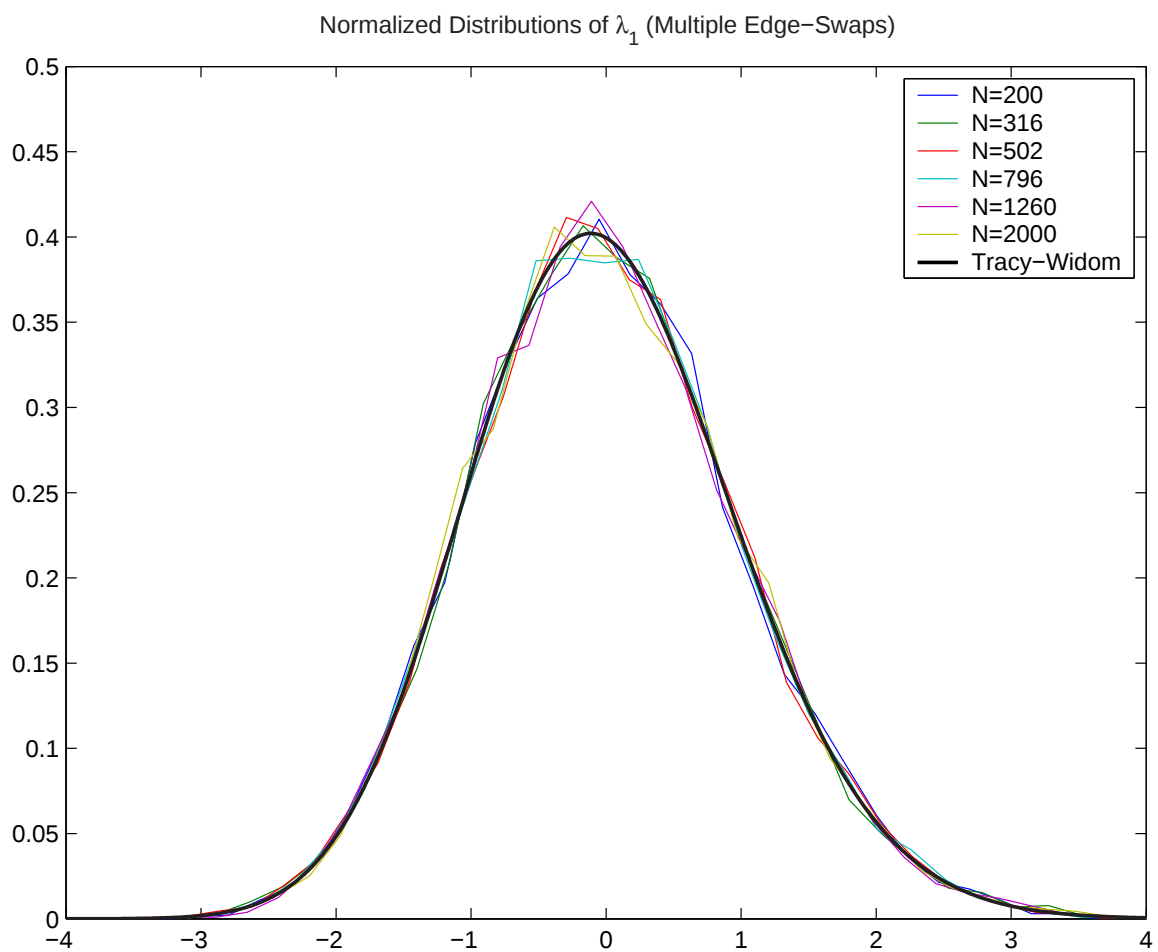


Figure 6: This figure is analogous to Figure 3, except that the data here was collected using the modified Bollobas algorithm, as opposed to the random walks used to collect the data in Figure 3.

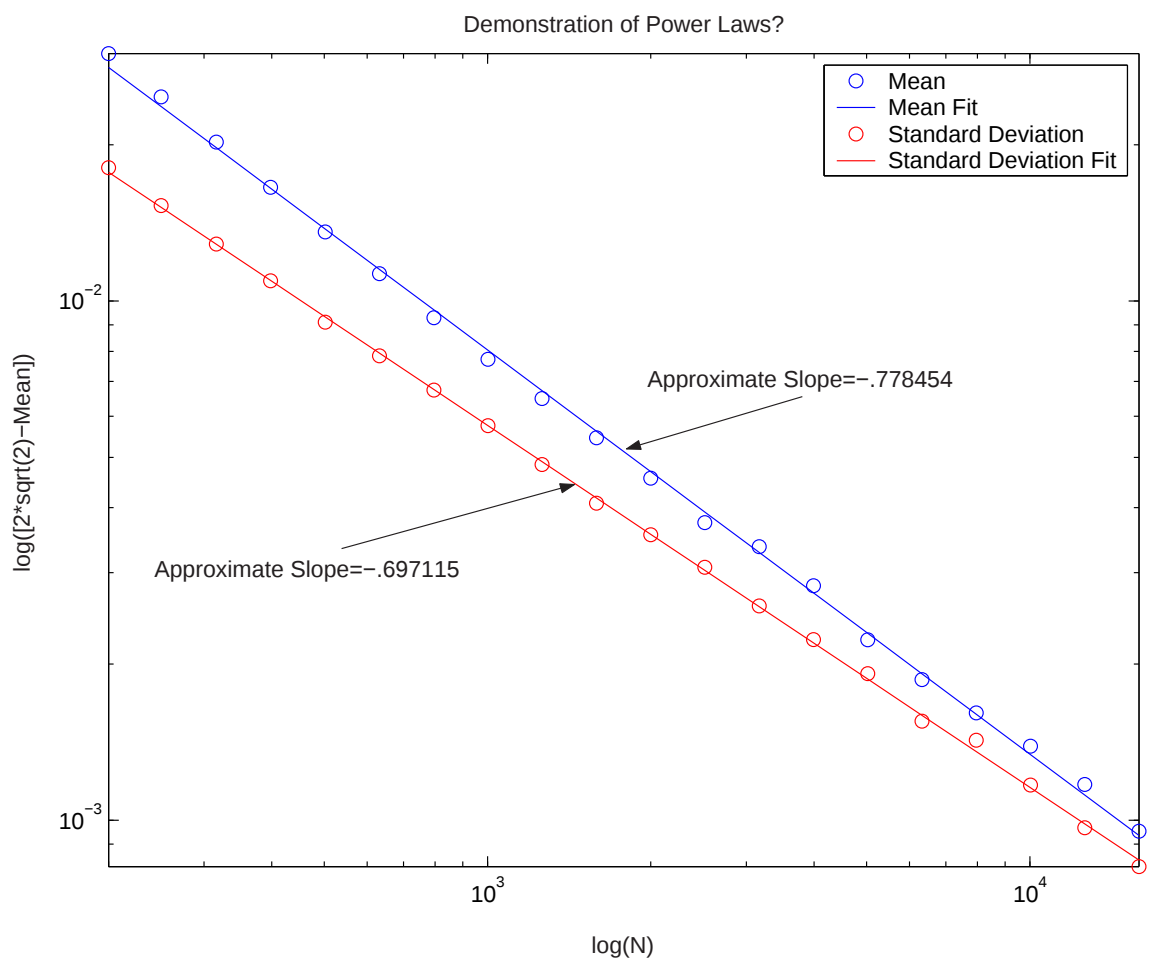


Figure 7: Shows the rates at which the mean and standard deviation (from Experiment 2) approach $2\sqrt{2}$ and zero respectively.