The Distribution of the Largest Nontrivial Eigenvalues in Families of Random Regular Graphs

Steven J. Miller, Tim Novikoff, and Anthony Sabelli

CONTENTS

Recently, Friedman proved Alon's conjecture for many families of d-regular graphs, namely that given any $\epsilon > 0$, "most" graphs have their largest nontrivial eigenvalue at most $2\sqrt{d-1}+\epsilon$ in absolute value; if the absolute value of the largest nontrivial eigenvalue is at most $2\sqrt{d-1}$, then the graph is said to be Ramanujan. These graphs have important applications in communication network theory, allowing the construction of superconcentrators and nonblocking networks, as well as in coding theory and cryptography. Since many of these applications depend on the size of the largest nontrivial positive and negative eigenvalues, it is natural to investigate their distributions. We show that these are well modeled by the $\beta = 1$ Tracy–Widom distribution for several families. If the observed growth rates of the mean and standard deviation as a function of the number of vertices hold in the limit, then in the limit, approximately 52% of d-regular graphs from bipartite families should be Ramanujan, and about 27% from nonbipartite families (assuming that the largest positive and negative eigenvalues are independent).

1. INTRODUCTION

1.1 Families of Graphs

In this paper we investigate the distribution of the largest nontrivial eigenvalues associated to d-regular undirected graphs.[1] A graph G is bipartite if the vertex set of G can be split into two disjoint sets A and B such that every edge connects a vertex in A with one in B, and G is d-regular if every vertex is connected to exactly d vertices. To any graph G we may associate a real symmetric matrix, called its adjacency matrix, by setting a_{ij} to be the number of edges connecting vertices i and j. Let us write the eigenvalues of G as $\lambda_1(G) \geq \cdots \geq \lambda_N(G)$, where G

[1] An undirected graph G is a collection of vertices V and edges E connecting pairs of vertices. A graph G is simple if there are no multiple edges between vertices; G has a self-loop if a vertex is connected to itself, and G is connected if given any two vertices u and w there is a sequence of vertices $v_1, \dots, v_n$ such that there is an edge from v_i to v_{i+1} for $i \in \{0, \dots, n+1\}$ (where $v_0 = u$ and $v_{n+1} = w$).

2000 AMS Subject Classification: Primary 05C80; Secondary 05C50, 15A52.

Keywords: Ramanujan graphs, random graphs, largest nontrivial eigenvalues, Tracy–Widom distribution

has N vertices. We call any eigenvalue equal to $\pm d$ a *trivial* eigenvalue (there is an eigenvalue of $-d$ if and only if the graph is bipartite), and all other eigenvalues are called *nontrivial*.

The eigenvalues of the adjacency matrix provide much information about the graph. We give two such properties to motivate investigations of the eigenvalues; see [Davidoff et al. 03, Sarnak 90] for more details.

First, if G is d-regular, then $\lambda_1(G) = d$ (the corresponding eigenvector is all 1's); further, $\lambda_2(G) < d$ if and only if G is connected. Thus if we think of our graph as a network, $\lambda_2(G)$ tells us whether all nodes can communicate with each other. For network purposes, it is natural to restrict to connected graphs without self-loops.

Second, a fundamental problem is to construct a well-connected network such that each node can communicate with any other node "quickly" (i.e., there is a short path of edges connecting any two vertices). While a simple solution is to take the complete graph as our network, these graphs are expensive: there are N vertices and $\binom{N}{2} = N(N-1)/2$ edges. We want a well-connected network in which the number of edges grows linearly with N. Let V be the set of vertices for a graph G, and E its set of edges. The *boundary* ∂U of a vertex set $U \subseteq V$ is the set of edges connecting U to $V \setminus U$. The *expanding constant* $h(G)$ is

$$h(G) := \inf \left\{ \frac{|\partial U|}{\min(|U|, |V \setminus U|)} : U \subset V,\ |U| > 0 \right\}.$$

It measures the connectivity of G. If $\{G_m\}$ is a family of connected d-regular graphs, then we call $\{G_m\}$ a family of *expanders* if $\lim_{m\to\infty} |G_m| = \infty$ and there exists an $\epsilon > 0$ such that for all m, $h(G_m) \geq \epsilon$. Expanders have two very important properties: they are sparse ($|E|$ grows at most linearly with $|V|$), and they are highly connected (the expanding constants have a positive lower bound). These graphs have important applications in communication network theory, allowing the construction of superconcentrators and nonblocking networks [Bien 89, Pippenger 77], as well as applications to coding theory [Sipser and Spielman 96] and cryptography [Goldreich et al. 90]; see [Sarnak 04] for a brief introduction to expanders. The Cheeger–Buser inequalities[2] [Alon and Milman 85, Dodziuk 84] give upper and lower bounds for the expanding constant of a finite d-regular connected graph in terms of the spectral gap (the separation between the largest and second-largest eigenvalues) $d - \lambda_2(G)$:

$$\frac{d - \lambda_2(G)}{2} \leq h(G) \leq 2\sqrt{2d(d - \lambda_2(G))}.$$

Thus we have a family of expanders if and only if there exists an $\epsilon > 0$ such that for all m, $d - \lambda_2(G_m) \geq \epsilon$. Finding graphs with small $\lambda_2(G)$ leads to large spectral gaps and thus sparse, highly connected graphs.

For many problems, the behavior is controlled by the largest absolute value of a nontrivial eigenvalue. We write $\lambda_+(G)$ (respectively $\lambda_-(G)$) for the largest nontrivial positive eigenvalue (respectively the most negative nontrivial eigenvalue) of G, and set $\lambda(G) = \max(|\lambda_+(G)|, |\lambda_-(G)|)$. Alon–Boppana, Burger, and Serre proved that for any family $\{G_m\}$ of finite connected d-regular graphs with $\lim_{m\to\infty} |G_m| = \infty$, we have $\liminf_{m\to\infty} \lambda(G_m) \geq 2\sqrt{d-1}$; in fact, Friedman [Friedman 93] proved that if G is a d-regular ($d \geq 3$) graph with n vertices, then

$$\lambda(G) \geq 2\sqrt{d-1} \left(1 - \frac{2\pi^2}{(\log_{d-1} n)^2} + O\left(\frac{1}{(\log_{d-1} n)^4}\right)\right). \tag{1–1}$$

Thus we are led to search for graphs with $\lambda(G) \leq 2\sqrt{d-1}$; such graphs are called Ramanujan graphs.[3] See [Murty 03] for a nice survey. Explicit constructions are known when d is 3 [Chiu 92] or $q+1$, where q is either an odd prime [Lubotzky et al. 88, Margulis 88] or a prime power [Morgenstern 94].

Alon [Alon 86] conjectured that as $N \to \infty$, for $d \geq 3$ and any $\epsilon > 0$, "most" d-regular graphs on N vertices have $\lambda(G) \leq 2\sqrt{d-1} + \epsilon$; it is known that the bound $2\sqrt{d-1}$ cannot be improved. Upper bounds on $\lambda(G)$ of this form give a good spectral gap. Recently, Friedman [Friedman 03] proved Alon's conjecture for many models of d-regular graphs.

Our goal in this work is to investigate the distribution of $\lambda_\pm(G)$ and $\lambda(G)$ numerically for these and other families of d-regular graphs. By identifying the limiting distribution of these eigenvalues, we are led to the conjecture that for many families of d-regular graphs, in the limit as the number of vertices tends to infinity, the probability that a graph in the family has $\lambda(G) \leq 2\sqrt{d-1}$ tends to approximately 52% if the family is bipartite, and about 27% otherwise.

[2]The name is from an analogy with the isoperimetric constant of a compact Riemann manifold.

[3]Lubotzky, Phillips, and Sarnak [Lubotzky et al. 88] construct an infinite family of $(p+1)$-regular Ramanujan graphs for primes $p \equiv 1 \bmod 4$. Their proof uses the Ramanujan conjecture for bounds on Fourier coefficients of cusp forms, which led to the name Ramanujan graphs.

Specifically, consider a family $\mathcal{F}_{N,d}$ of d-regular graphs on N vertices. For each $G \in \mathcal{F}_{N,d}$, we study

$$\widetilde{\lambda_\pm}(G) = \frac{|\lambda_\pm(G)| - 2\sqrt{d-1} + c_{\mu,N,d,\pm} N^{m_\pm(\mathcal{F}_{N,d})}}{c_{\sigma,N,d,\pm} N^{s_\pm(\mathcal{F}_{N,d})}}; \tag{1–2}$$

we use m for the first exponent, since it arises from studying the means, and s for the second, since it arises from studying the standard deviations. Our objective is to see whether as G varies in a family $\mathcal{F}_{N,d}$, $\widetilde{\lambda_\pm}(G)$ converges to a universal distribution as $N \to \infty$. We therefore subtract off the sample mean and divide by the standard deviation to obtain a mean-0, variance-1 data set, which will facilitate comparisons to candidate distributions. We write the subtracted mean as a sum of two terms. The first is $2\sqrt{d-1}$, the expected mean as $N \to \infty$. The second is the remaining effect, which has been observed to be negative (see the concluding remarks in [Friedman 03] and [Hoory et al. 06]), and was found to be negative in all our experiments. We shall assume in our discussions below that $c_{\mu,N,d,\pm} < 0$. Of particular interest is whether $m_\pm(\mathcal{F}_{N,d}) - s_\pm(\mathcal{F}_{N,d}) < 0$. If this is negative (for both $\lambda_\pm(G)$), if $\widetilde{\lambda_\pm}(G)$ converges to a universal distribution, and if $\lambda_+(G)$ and $\lambda_-(G)$ are independent for the nonbipartite families, then in the limit, a positive percentage of graphs in $\mathcal{F}_{N,d}$ are *not* Ramanujan. This follows from the fact that for $|\lambda_\pm(G)|$, in the limit a negligible fraction of the standard deviation suffices to move beyond $2\sqrt{d-1}$; if $m_\pm(\mathcal{F}_{N,d}) - s_\pm(\mathcal{F}_{N,d}) > 0$, then we may move many multiples of the standard deviation and still be below $2\sqrt{d-1}$ (see Remark 2.1 for a more detailed explanation).

Remark 1.1. (Families of d-regular graphs.) We describe the families we investigate. For convenience in our studies we always take N to be even. Friedman [Friedman 03] showed that for fixed ϵ, for the families $\mathcal{G}_{N,d}$, $\mathcal{H}_{N,d}$, and $\mathcal{I}_{N,d}$ defined below, as $N \to \infty$ "most" graphs[4] have $\lambda(G) \leq 2\sqrt{d-1} + \epsilon$:

- $\mathcal{B}_{N,d}$. We let $\mathcal{B}_{N,d}$ denote the set of d-regular bipartite graphs on N vertices. We may model these by letting π_1 denote the identity permutation and choosing $d-1$ independent permutations of $\{1, \ldots, N/2\}$. For each choice we consider the graph with edge set
$$E : \big\{(i, \pi_j(i) + N/2) : i \in \{1, \ldots, N/2\},\ j \in \{1, \ldots, d\}\big\}.$$
- $\mathcal{G}_{N,d}$. For d even, let $\pi_1, \ldots, \pi_{d/2}$ be chosen independently from the $N!$ permutations of $\{1, \ldots, N\}$. For each choice of $\pi_1, \ldots, \pi_{d/2}$ form the graph with edge set
$$E : \big\{(i, \pi_j(i)),\ (i, \pi_j^{-1}(i)) : i \in \{1, \ldots, N\},\ j \in \{1, \ldots, d/2\}\big\}.$$
Note that $\mathcal{G}_{N,d}$ can have multiple edges and self-loops, and a self-loop at vertex i contributes 2 to a_{ii}.
- $\mathcal{H}_{N,d}$. These are constructed in the same manner as $\mathcal{G}_{N,d}$, with the additional constraint that the permutations are chosen independently from the $(N-1)!$ permutations whose cyclic decomposition is one cycle of length N.
- $\mathcal{I}_{N,d}$. These are constructed similarly, except that instead of choosing $d/2$ permutations, we choose d perfect matchings; the d matchings are independently chosen from the $(N-1)!!$ perfect matchings.[5]
- Connected and simple graphs. If $\mathcal{F}_{N,d}$ is any of the families above ($\mathcal{B}_{N,d}$, $\mathcal{G}_{N,d}$, $\mathcal{H}_{N,d}$, $\mathcal{I}_{N,d}$), let $\mathcal{CF}_{N,d}$ denote the subset of graphs that are connected, and $\mathcal{SCF}_{N,d}$ the subset of graphs that are simple and connected.

Remark 1.2. The eigenvalues of bipartite graphs are symmetric about zero. We sketch the proof. Let G be a bipartite graph with $2N$ vertices. Its adjacency matrix is of the form $A(G) = \begin{pmatrix} Z & B \\ B & Z \end{pmatrix}$, where Z is the $N \times N$ zero matrix and B is an $N \times N$ matrix. Let $J = \begin{pmatrix} Z & I \\ -I & Z \end{pmatrix}$, where I is the $N \times N$ identity matrix. Simple calculations show that $J^{-1} = -J$ and $J^{-1}A(G)J = -A(G)$. Noting that similar matrices have the same eigenvalues, we see that the eigenvalues of $A(G)$ must be symmetric about zero.

[4]Friedman shows that given $\epsilon > 0$, with probability at least $1 - c_{\mathcal{F}_d} N^{-\tau(\mathcal{F}_d)}$ we have $\lambda(G) \leq 2\sqrt{d-1} + \epsilon$ for $G \in \mathcal{F}_{N,d}$, and with probability at least $\widetilde{c}_{\mathcal{F}_d} N^{-\widetilde{\tau}(\mathcal{F}_d)}$ we have $\lambda(G) > 2\sqrt{d-1}$; see [Friedman 03] for the values of the exponents.

[5]For example, if $d = 3$ and $N = 8$, our three permutations might be (43876152), (31248675), and (87641325). Each permutation generates $8/2 = 4$ edges. Thus the first permutation gives edges between vertices 4 and 3, between 8 and 7, between 6 and 1, and between 5 and 2. A permutation whose cyclic decomposition is one cycle of length N can be written in N different ways (depending on which element is listed first). This permutation generates two different perfect matchings, depending on where we start. Note that there are no self-loops.

1.2 Tracy–Widom Distributions

We investigate in detail the distribution of $\lambda_\pm(G)$ for d-regular graphs related to two of the families above: the perfect-matching family $\mathcal{I}_{N,d}$ and the bipartite family $\mathcal{B}_{N,d}$ (by Remark 1.2 we need to study $\lambda_+(G)$ for only the bipartite family). Explicitly, for N even, we study $\mathcal{CI}_{N,d}$, $\mathcal{SCI}_{N,d}$, $\mathcal{CB}_{N,d}$, and $\mathcal{SCB}_{N,d}$; we restrict to connected graphs, since d is a multiple eigenvalue for disconnected graphs. As d and N increase, so too does the time required to choose uniformly a simple connected graph from our families; we concentrate on $d \in \{3, 4, 7, 10\}$ and $N \leq 20000$.

Since there are known constructions of Ramanujan graphs for d equal to 3 or $q + 1$ (where q is either an odd prime or a prime power), $d = 7$ is the first instance for which there is no known explicit construction to produce Ramanujan graphs. In the interest of space, we report in detail on the $d = 3$ computations for $\lambda_+(G)$. We remark briefly on the other computations and results, which are similar and are available from the authors on request; much of the data and programs used are available online (http://www.math.princeton.edu/mathlab/ramanujan/).

We conjecture that the distributions of $\lambda_\pm(G)$ are independent in nonbipartite families and that each converges to the $\beta = 1$ Tracy–Widom distribution (see Conjecture 1.3 for exact statements). We summarize our numerical investigations supporting this conjecture in Section 1.3, and content ourselves here with describing why it is natural to expect the $\beta = 1$ Tracy–Widom distribution to be the answer. The Tracy–Widom distributions model the limiting distribution of the normalized largest eigenvalues for many ensembles of matrices. There are three distributions $f_\beta(s)$: (i) $\beta = 1$, corresponding to orthogonal symmetry (GOE); (ii) $\beta = 2$, corresponding to unitary symmetry (GUE); (iii) $\beta = 4$, corresponding to symplectic symmetry (GSE). These distributions can be expressed in terms of a particular Painlevé II function, and are plotted in Figure 1.

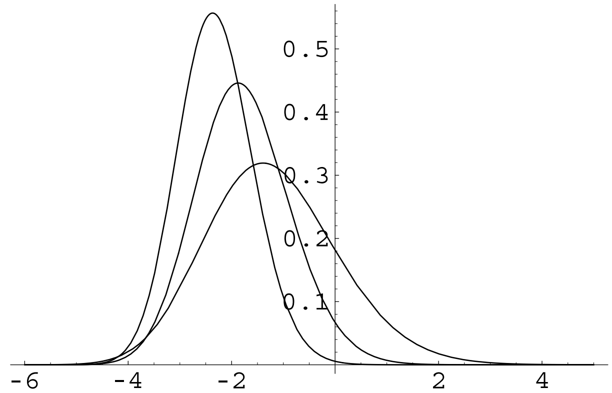

FIGURE 1. Plots of the three Tracy–Widom distributions: $f_1(s)$ has the smallest maximum amplitude, then $f_2(s)$, and then $f_4(s)$.

We describe some of the problems in which the Tracy–Widom distributions arise, and why the $\beta = 1$ distribution should describe the distributions of $\lambda_\pm(G)$. The first is in the distribution of the largest eigenvalue (as $N \to \infty$) in the $N \times N$ Gaussian orthogonal, unitary, and symplectic ensembles [Tracy and Widom 96]. For example, consider the $N \times N$ Gaussian orthogonal ensemble. From the scaling in Wigner's semicircle law [Mehta 91, Wigner 57], we expect the eigenvalues to be of order $\sqrt{N}$. With $\lambda_{\max}(A)$ denoting the largest eigenvalue of A, the normalized largest eigenvalue $\widetilde{\lambda}_{\max}(A)$ satisfies

$$\lambda_{\max}(A) \;=\; 2\sigma\sqrt{N} + \frac{\widetilde{\lambda}_{\max}(A)}{N^{1/6}}; \tag{1–3}$$

here σ is the standard deviation of the Gaussian distribution of the off-diagonal entries, and is often taken to be 1 or $1/\sqrt{2}$. As $N \to \infty$, the distribution of $\widetilde{\lambda}_{\max}(A)$ converges to $f_1(s)$.

The Tracy–Widom distributions also arise in combinatorics in the analysis of the length of the largest increasing subsequence of a random permutation and the number of boxes in rows of random standard Young tableaux [Baik et al. 99, Borodin et al. 00, Baik and Rains 01a, Baik and Rains 01b, Johansson 01], in growth problems [Baik and Rains 00, Gravner et al. 01, Johansson 02b, Prähofer and Spohn 00a, Prähofer and Spohn 00b], random tilings [Johansson 02a], the largest principal component of covariances matrices [Soshnikov 08], queuing theory [Baryshnikov 01, Gravner et al. 01], and superconductors [Vavilov et al. 01]; see [Tracy and Widom 02] for more details and references.

It is reasonable to conjecture that appropriately normalized, the limiting distributions of $\lambda_\pm(G)$ in the families of d-regular graphs considered by Friedman converges to the $\beta = 1$ Tracy–Widom distribution (the largest eigenvalue is always d). One reason for this is that to any graph G we may associate its adjacency matrix $A(G)$, where a_{ij} is the number of edges connecting vertices i and j. Thus a family of d-regular graphs on N vertices gives us a subfamily of $N \times N$ real symmetric matrices, and real symmetric matrices typically have $\beta = 1$ symmetries. While [McKay 81] showed that for fixed d, the density of normalized eigenvalues is different from the semicircle found for the GOE (though as $d \to \infty$ the limiting distribution does converge to the semicircle), [Jakobson et al. 99] experimentally found that the spacings between adjacent normalized eigenvalues agree with the GOE.

Since the spacings in the bulk agree in the limit, it is plausible to conjecture that the spacings at the edge agree in the limit as well, in particular, that the density of the normalized second largest eigenvalue converges to $f_1(s)$.

1.3 Summary of Experiments, Results, and Conjectures

We numerically investigated the eigenvalues for the families $\mathcal{CI}_{N,d}$, $\mathcal{SCI}_{N,d}$, $\mathcal{CB}_{N,d}$, and $\mathcal{SCB}_{N,d}$. Most of the simulations were performed on a 1.6-GHz Centrino processor running version 7 of MATLAB over several months; the data indicate that the rate of convergence is probably controlled by the logarithm of the number of vertices, and thus there would not be significant gains in understanding the limiting behavior by switching to more powerful systems.[6] The data are available online (http://www.math.princeton.edu/mathlab/ramanujan/).

We varied N from 26 up to 50,000. For each N we randomly chose 1000 graphs G from the various ensembles, and calculated $\lambda_\pm(G)$. Letting $\mu^{\text{sample}}_{\mathcal{F}_{N,d,\pm}}$ and $\sigma^{\text{sample}}_{\mathcal{F}_{N,d,\pm}}$ denote the mean and standard deviation of the sample data (these are functions of N and $\lim_{N\to\infty} \mu^{\text{sample}}_{\mathcal{F}_{N,d,\pm}} = 2\sqrt{d-1}$), we studied the distribution of

$$\left(\lambda_\pm(G) - \mu^{\text{sample}}_{\mathcal{F}_{N,d,\pm}}\right) \Big/ \sigma^{\text{sample}}_{\mathcal{F}_{N,d,\pm}}. \tag{1–4}$$

This normalizes our data to have mean 0 and variance 1, which we compared to the $\beta = 1$ Tracy–Widom distribution; as an additional test, we also compared our data to $\beta = 2$ and 4 Tracy–Widom distributions, as well as the standard normal.

Before stating our results, we comment on some of the difficulties of these numerical investigations.[7] If $g(s)$ is a probability distribution with mean μ and variance σ^2, then $\sigma g(\sigma x + \mu)$ has mean 0 and variance 1. Since we do not know the normalization constants in (1–2) for the second-largest eigenvalue, it is natural to study (1–4) and compare our sample distributions to the normalized $\beta = 1$ Tracy–Widom distribution.[8] In fact, even if we did know the constants it would still be worth normalizing our data in order to determine whether other distributions, appropriately scaled, provide good fits as well. As remarked in Section 1.2, there are natural reasons to suspect that the $\beta = 1$ Tracy–Widom is the limiting distribution; however, as Figure 2 shows, if we normalize the three Tracy–Widom distributions to have mean 0 and variance 1, then they are all extremely close to the standard normal.

The fact that several different distributions can provide good fits to the data is common in random matrix theory. For example, Wigner's surmise[9] for the spacings between adjacent normalized eigenvalues in the bulk of the spectrum is extremely close to the actual answer (and in fact, Wigner's surmise is often used for comparison purposes, since it is easier to plot than the actual answer).[10] While the two distributions are quite close (see [Gaudin 61, Mehta 60, Mehta 91]) and both often provide good fits to data, they are unequal, and it is the Fredholm determinant that is correct.[11] We see a similar

[6]In fact, many quantities and results related to these families of graphs are controlled by the logarithm of the number of vertices. For example, a family of graphs is said to have large girth if the girths are greater than a constant times the logarithm of the number of vertices [Davidoff et al. 03, p. 10]. For another example, see (1–1).

[7]Another difficulty is that the MATLAB code was originally written to investigate bipartite graphs. The symmetry of the eigenvalues allowed us to look at just the second-largest eigenvalue; when we ran the code for nonbipartite graphs, we originally did not realize that this had been hardwired. Thus we were implicitly assuming $\lambda(G) = \lambda_+(G)$, which is frequently false for nonbipartite graphs. This error led us initially to conjecture that 52% of these graphs are Ramanujan in the limit, instead of the 27% that we discuss later.

[8]The Tracy–Widom distributions [Tracy and H. Widom 94] could have been defined in an alternative way as mean-zero distributions if lower-order terms had been subtracted off; since these terms were kept, the resulting distributions have nonzero means. These correction factors vanish in the limit, but for finite N, they result in an N-dependent correction (we divide by a quantity with the same N-dependence, so the resulting answer is a nonzero mean). This is similar to other situations in number theory and random matrix theory. For example, originally "high" critical zeros of $\zeta(s)$ were shown to be well modeled by the $N \to \infty$ scaling limits of the $N \times N$ GUE ensemble [Odlyzko 87, Odlyzko 01]; however, for zeros with imaginary part about T, a better fit is obtained using finite N (in particular, $N \sim \log T$; see [Keating and Snaith 00]).

[9]Wigner conjectured that as $N \to \infty$, the spacing between adjacent normalized eigenvalues in the bulk of the spectrum of the $N \times N$ GOE ensemble tends to $p_W(s) = (\pi s/2)\exp\left(-\pi s^2/4\right)$. He was led to this by assuming that (1) given an eigenvalue at x, the probability that another one lies s units to its right is proportional to s; (2) given an eigenvalue at x and $I_1, I_2, I_3, \ldots$ any disjoint intervals to the right of x, then the events of observing an eigenvalue in I_j are independent for all j; (3) the mean spacing between consecutive eigenvalues is 1.

[10]The distribution is $(\pi^2/4)d^2\Psi/dt^2$, where $\Psi(t)$ is (up to constants) the Fredholm determinant of the operator $f \mapsto \int_{-t}^{t} K * f$ with kernel $K = \frac{1}{2\pi}\left(\frac{\sin(\xi-\eta)}{\xi-\eta} + \frac{\sin(\xi+\eta)}{\xi+\eta}\right)$.

[11]While this is true for number-theoretic systems with large numbers of data points, there is often not enough data for physical systems to make a similar claim. The number of energy levels from heavy nuclei in nuclear physics is typically between 100 and 2000, which can be insufficient to distinguish between GOE and GUE behavior (while we expect GOE from physical symmetries, there is a maximum of about a 2% difference in their cumulative distribution functions). Current research in quantum dots (see [Alhassid 00]) shows promise for obtaining sufficiently large data sets to detect such subtle differences.

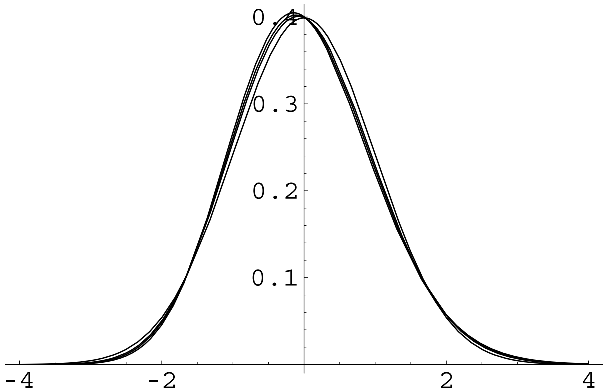

FIGURE 2. Plots of the three Tracy–Widom distributions, normalized to have mean 0 and variance 1, and the standard normal.

phenomenon, since for many of our data sets we obtain good fits from the three normalized Tracy–Widom distributions and the standard normal. It is therefore essential that we find a statistic sensitive to the subtle differences among the four normalized distributions.

We record the mean, standard deviation, and the percentage of the mass to the left of the mean for the three Tracy–Widom distributions (and the standard normal) in Table 1. The fact that the four distributions have different percentages of their mass to the left of the mean gives us a statistical test to determine which of the four distributions best models the observed data.

Thus, in addition to comparing the distribution of the normalized eigenvalues in (1–4) to the normalized Tracy–Widom distributions, we also computed the percentage of time that $\lambda_{\pm}(G)$ was less than the sample mean. We compared this percentage to the three different values for the Tracy–Widom distribution and the value for the standard normal (which is just 0.5). Since the four percentages are different, this comparison provides evidence that of the four distributions, the second-largest eigenvalues are modeled *only* by a $\beta = 1$ Tracy–Widom distribution.

	Mean μ	Std. Dev. σ	$F_\beta(\mu_\beta)$
TW($\beta = 1$)	-1.2065	1.26798	0.519652
TW($\beta = 2$)	-1.7711	0.90177	0.515016
TW($\beta = 4$)	-2.3069	0.71953	0.511072
Standard Normal	0.0000	1.00000	0.500000

TABLE 1. Parameters for the Tracy–Widom distributions (before being normalized to have mean 0 and variance 1). Here F_β is the cumulative distribution function for f_β, and $F_\beta(\mu_\beta)$ is the mass of f_β to the left of its mean.

We now briefly summarize our results and the conjecture what they suggest. We concentrate on the families (see Remark 1.1 for definitions) $\mathcal{CI}_{N,d}$, $\mathcal{SCI}_{N,d}$, $\mathcal{CB}_{N,d}$, and $\mathcal{SCB}_{N,d}$ with $d \in \{3,4\}$, as well as $\mathcal{CI}_{N,7}$ and $\mathcal{CI}_{N,10}$. For each

$$N \in \{26, 32, 40, 50, 64, 80, 100, 126, 158, 200, 252, 316, 400, 502, 632, 796, 1002, 1262, 1588, 2000, 2516, 3168, 3990, 5022, 6324, 7962, 10022, 12618, 15886, 20000\},$$

we randomly chose 1000 graphs from each family. We analyze the data for the 3-regular graphs in Section 2. Since the results are similar, the data and analysis for the other families are available online (http://www.math.princeton.edu/mathlab/ramanujan/), where we include our data for $d = 3$ as well.

1.3.1 Chi-squared tests for goodness of fit.

Chi-squared tests show that the distribution of the normalized eigenvalues $\lambda_{\pm}(G)$ are well modeled by a $\beta = 1$ Tracy–Widom distribution, although the other two Tracy–Widom distributions and the standard normal also provide good fits; see Tables 2 and 3.

The χ^2 values are somewhat large for small $N \le 100$, but once $N \ge 200$, they are small for all families except for the connected bipartite graphs, indicating good fits. For the connected bipartite graphs, the χ^2 values are small for N large. This indicates that perhaps the rate of convergence is slower for connected bipartite graphs; we shall see additional differences in behavior for these graphs below.

Further, on average, the χ^2 values are lowest for the $\beta = 1$ case. While this suggests that the correct model

N	$\mathrm{TW}_1^{\mathrm{norm}}$	$\mathrm{TW}_2^{\mathrm{norm}}$	$\mathrm{TW}_4^{\mathrm{norm}}$	N(0, 1)
mean (all N)	27.0	24.5	24.0	29.4
median (all N)	21.2	19.1	20.0	26.5
mean (last 10)	21.7	22.2	23.7	35.0
median (last 10)	21.2	20.9	22.4	35.4

TABLE 2. Summary of χ^2 values: each set is 1000 random 3-regular graphs from $\mathcal{CI}_{N,3}$ with $N \in \{26, 32, 40, 50, 64, 80, 100, 126, 158, 200, 252, 316, 400, 502, 632, 796, 1002, 1262, 1588, 2000, 2516, 3168, 3990, 5022, 6324, 7962, 10022, 12618, 15886, 20000\}$. The sample distribution in each set is normalized to have mean 0 and variance 1, and is then compared to normalized Tracy–Widom distributions $\mathrm{TW}_\beta^{\mathrm{norm}}$ ($\beta \in \{1, 2, 4\}$, normalized to have mean 0 and variance 1) and the standard normal $\mathrm{N}(0, 1)$. There are 19 degrees of freedom, and the critical values are 30.1435 (for $\alpha = 0.05$) and 36.1908 (for $\alpha = 0.01$).

N	$\mathcal{CI}_{N,3}$	$\mathcal{SCI}_{N,3}$	$\mathcal{CB}_{N,3}$	$\mathcal{SCB}_{N,3}$
mean (all N)	27	19	78	19
s.d. (all N)	21	8	180	7
mean (last 10)	22	18	44	17
s.d. (last 10)	11	6	37	8
mean (last 5)	23	18	32	14
s.d. (last 5)	13	8	23	1

TABLE 3. Summary of χ^2 values: each set is 1000 random 3-regular graphs with N vertices from our families, with $N \in \{26, 32, 40, 50, 64, 80, 100, 126, 158, 200, 252, 316, 400, 502, 632, 796, 1002, 1262, 1588, 2000, 2516, 3168, 3990, 5022, 6324, 7962, 10022, 12618, 15886, 20000\}$. The sample distribution in each set is normalized to have mean 0 and variance 1, and is then compared to the normalized $\beta = 1$ Tracy–Widom distributions. There are 19 degrees of freedom, and the critical values are 30.1435 (for $\alpha = 0.05$) and 36.1908 (for $\alpha = 0.01$).

is a $\beta = 1$ Tracy–Widom distribution, the data are not conclusive.

1.3.2 Percentage of eigenvalues to the left of the mean.

As remarked, the four distributions, while close, differ in the percentage of their mass to the left of the mean. By studying the percentage of normalized eigenvalues in a sample less than the sample mean, we see that the $\beta = 1$ distribution provides a better fit to the observed results; however, with sample sizes of 1000, all four distributions provide good fits (see Table 4).

We therefore increased the number of graphs in the samples from 1000 to 100,000 for $N \in \{1002, 2000, 5002\}$ for the four families; increasing the sample size by a factor of 100 gives us an additional decimal digit of accuracy in measuring the percentages. See Table 5 for the results; this is *the* most important experiment in the paper, showing that for the families $\mathcal{CI}_{N,d}$, $\mathcal{SCI}_{N,d}$, and $\mathcal{SCB}_{N,d}$, the $\beta = 1$ Tracy–Widom distribution provides a significant fit, but the other three distributions do not.

Thus we have found a statistic that is sensitive to very fine differences among the four normalized distributions. However, *none* of the four candidate distributions provides a good fit for the family $\mathcal{CB}_{N,d}$ for these values of N. For this family the best fit is still with $\beta = 1$, but the z statistics are high (between 3 and 4), which suggests either that the distribution of eigenvalues for d-regular connected bipartite graphs might not be given by a $\beta = 1$ Tracy–Widom distribution or that the rate of convergence is slower; note that our χ^2 tests suggest that the rate of convergence is indeed slower for the connected bipartite family. In fact, upon increasing N to 10,022, we obtain a good fit for connected bipartite graphs; the z statistic is about 2 for $\beta = 1$, and almost 5 or larger

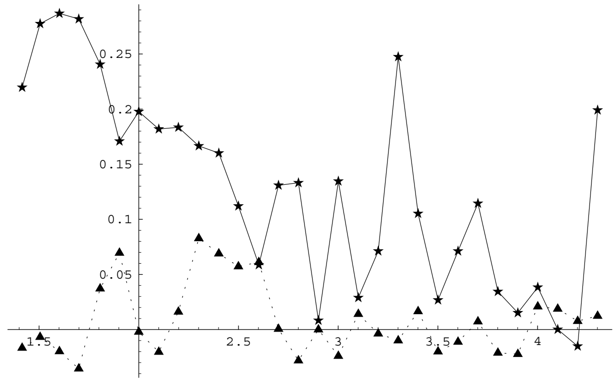

FIGURE 3. Sample correlation coefficients of $\lambda_{\pm}(G)$: each set is 1000 random 3-regular graphs with N vertices, chosen according to the specified construction. We plot the sample correlation coefficient versus the logarithm of the number of vertices. The $\mathcal{CI}_{N,3}$ are stars and $\mathcal{SCI}_{N,3}$ are triangles.

for the other three distributions. We shall see below that there are other statistics for which this family behaves differently from the other three, strongly suggesting that its rate of convergence is slower.

1.3.3 Independence of $\lambda_{\pm}(G)$.

A graph is Ramanujan if $|\lambda_{\pm}(G)| \leq 1$. For bipartite graphs it suffices to study $\lambda_{+}(G)$, since $\lambda_{-}(G) = -\lambda_{+}(G)$. For the nonbipartite families, however, we must investigate both. For our nonbipartite families we computed the sample correlation coefficient[12] for $\lambda_{+}(G)$ and $\lambda_{-}(G)$ as G varied through our random sample of 1000 graphs with N vertices. For the $\mathcal{SCI}_{N,d}$ families we found the correlation coefficients to be quite small; when $d = 3$ they were in $[-.0355, 0.0827]$. For the $\mathcal{CI}_{N,d}$ the values were larger, but still small. When $d = 3$, the correlation coefficients were in $[-0.0151, 0.2868]$, and all but two families with at least 5000 vertices had a correlation coefficient less than 0.1 in absolute value (and the values were generally decreasing with increasing N); see Figure 3 for the values. Thus the data suggest that the $\lambda_{\pm}(G)$ are independent (for nonbipartite families).

1.3.4 Percentage of graphs that are Ramanujan.

Except occasionally for the connected bipartite families, almost always $s_{\pm}(\mathcal{F}_{N,d}) > m_{\pm}(\mathcal{F}_{N,d})$. Recall our normalization of the eigenvalues from (1–2):

$$\widetilde{\lambda_{\pm}}(G) \;=\; \frac{\lambda_{\pm}(G) - 2\sqrt{d-1} + c_{\mu,N,d,\pm} N^{m_{\pm}(\mathcal{F}_{N,d})}}{c_{\sigma,N,d,\pm} N^{s_{\pm}(\mathcal{F}_{N,d})}};$$

[12]The sample correlation coefficient r_{xy} is $S_{xy}/\sqrt{S_{xx}S_{yy}}$, where $S_{uv} = \sum_{i=1}^{n}(u_i - \overline{u})(v_i - \overline{v})$ (with $\overline{u}$ the mean of the u_i's). By Cauchy–Schwarz, $|r_{xy}| \leq 1$. If the x_i and y_i are independent, then $r_{xy} = 0$, though the converse need not hold.

N	Observed mass	$z_{\text{TW},1}$	$z_{\text{TW},2}$	$z_{\text{TW},4}$	z_{StdNorm}
26	0.477	−2.700	−2.405	−2.155	−1.455
100	0.522	0.149	0.442	0.691	1.391
400	0.522	0.149	0.442	0.691	1.391
1588	0.526	0.402	0.695	0.944	1.644
6324	0.524	0.275	0.568	0.818	1.518
20000	0.551	1.984	2.277	2.526	3.226
mean (last 10)	0.519	0.861	0.873	0.960	1.341
median (last 10)	0.519	0.696	0.758	0.854	1.170
mean (last 5)	0.514	1.186	1.126	1.076	1.138
median (last 5)	0.508	1.434	1.140	0.890	0.506

TABLE 4. The mass to the left of the sample mean for $\lambda_+(G)$ from each set of 1000 3-regular graphs from $\mathcal{CI}_{N,3}$ and the corresponding z statistics comparing that to the mass to the left of the mean of the three Tracy–Widom distributions (0.519652 for $\beta = 1$, 0.515016 for $\beta = 2$, 0.511072 for $\beta = 4$) and the standard normal (0.500). We use the absolute value of the z statistics for the means and medians. For a two-sided z test, the critical thresholds are 1.96 (for $\alpha = 0.05$) and 2.575 (for $\alpha = 0.01$). For brevity we report only some of the values for $N \in \{26, 32, 40, 50, 64, 80, 100, 126, 158, 200, 252, 316, 400, 502, 632, 796, 1002, 1262, 1588, 2000, 2516, 3168, 3990, 5022, 6324, 7962, 10022, 12618, 15886, 20000\}$, but list the mean and medians for the last five and last ten values of N.

$\mathcal{CI}_{N,3}$	$z_{\text{TW},1}$	$z_{\text{TW},2}$	$z_{\text{TW},4}$	z_{StdNorm}	Dis.
1002	1.2773	4.2103	6.7044	13.7053	0
2000	0.9671	3.9002	6.3944	13.3954	0
5022	0.3152	3.2485	5.7428	12.744	0
$\mathcal{SCI}_{N,3}$	$z_{\text{TW},1}$	$z_{\text{TW},2}$	$z_{\text{TW},4}$	z_{StdNorm}	Dis.
1002	−0.7481	2.1855	4.6801	11.6815	0
2000	−0.5899	2.3437	4.8382	11.8396	0
5022	−1.0456	1.8881	4.3827	11.3842	0
$\mathcal{CB}_{N,3}$	$z_{\text{TW},1}$	$z_{\text{TW},2}$	$z_{\text{TW},4}$	z_{StdNorm}	Dis.
1002	3.151	6.083	8.577	15.577	0
2000	3.787	6.719	9.213	16.213	1
5022	3.563	6.495	8.989	15.989	4
10022	2.049	4.982	7.476	14.477	0
$\mathcal{SCB}_{N,3}$	$z_{\text{TW},1}$	$z_{\text{TW},2}$	$z_{\text{TW},4}$	z_{StdNorm}	Dis.
1002	−1.963	0.971	3.465	10.467	0
2000	−0.767	2.167	4.661	11.663	2
5022	−0.064	2.869	5.364	12.365	4

TABLE 5. The mass to the left of the sample mean of $\lambda_+(G)$ for each set of 100,000 3-regular graphs from our four families ($\mathcal{CI}_{N,3}$, $\mathcal{SCI}_{N,3}$, $\mathcal{CB}_{N,3}$ and $\mathcal{SCB}_{N,3}$), and the corresponding z statistics comparing that to the mass to the left of the mean of the three Tracy–Widom distributions (0.519652 for $\beta = 1$, 0.515016 for $\beta = 2$, 0.511072 for $\beta = 4$) and the standard normal (0.500). The abbreviation Dis., for "discarded," refers to the number of graphs for which MATLAB's algorithm to determine the second largest eigenvalue did not converge; this was never greater than 4 for any data set. For a two-sided z test, the critical thresholds are 1.96 (for $\alpha = 0.05$) and 2.575 (for $\alpha = 0.01$).

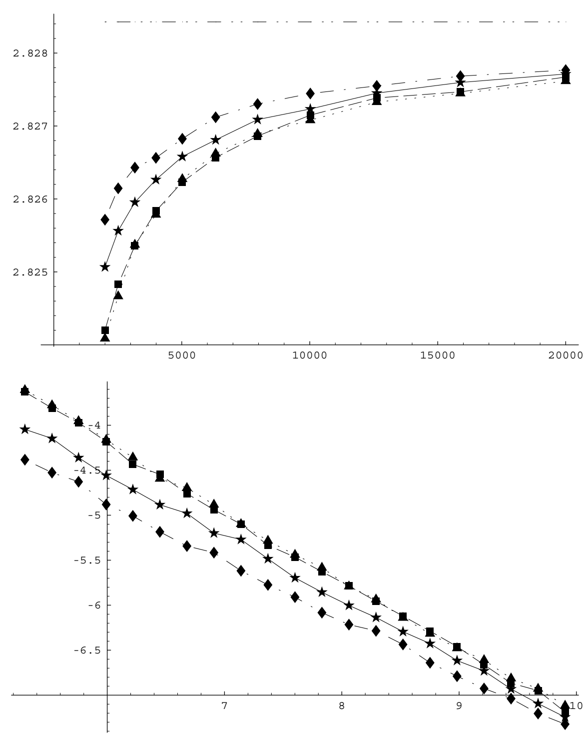

FIGURE 4. Sample means of $\lambda_+(G)$: each set is 1000 random 3-regular graphs with N vertices, chosen according to the specified construction. The first plot is the mean versus the number of vertices; the second plot is a log-log plot of the mean and the number of vertices; $\mathcal{CI}_{N,3}$ are stars, $\mathcal{SCI}_{N,3}$ are triangles, $\mathcal{CB}_{N,3}$ are diamonds, $\mathcal{SCB}_{N,3}$ are boxes; the dashed line is $2\sqrt{2} \approx 2.8284$.

Log-log plots of the differences between the sample means and the predicted values together with standard deviations yield behavior that is approximately linear as a function of $\log N$, supporting the claimed normalization. Further, the exponents appear to be almost constant in N, depending mostly only on d (see Figure 4).

If this behavior holds as $N \to \infty$, then in the limit, approximately 52% of the time we have $\lambda_+(G) \leq$

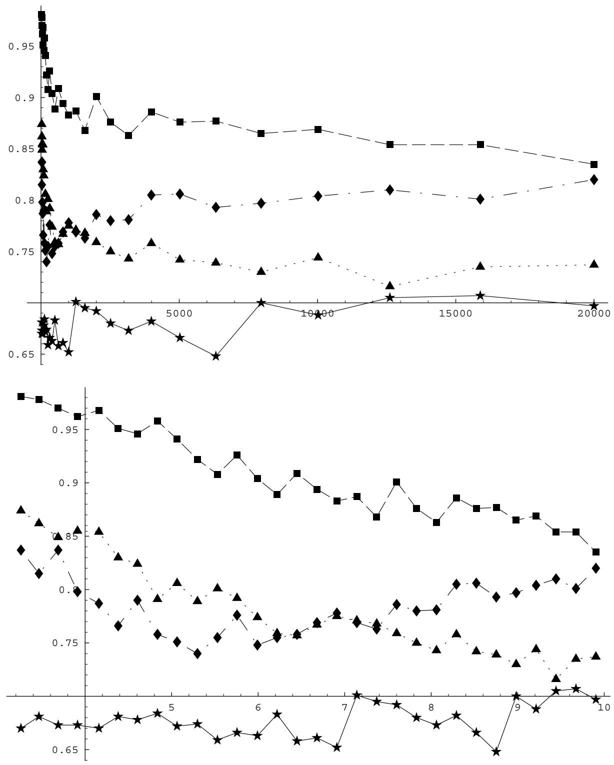

FIGURE 5. Percentage Ramanujan: each set is 1000 random 3-regular graphs with N vertices, chosen according to the specified construction. The first plot is the percentage versus the number of vertices; the second plot is the percentage versus the logarithm of the number of vertices. Here $\mathcal{CI}_{N,3}$ are stars, $\mathcal{SCI}_{N,3}$ are diamonds, $\mathcal{CB}_{N,3}$ are triangles, $\mathcal{SCB}_{N,3}$ are boxes.

$2\sqrt{d-1}$ (and similarly, about 52% of the time, $|\lambda_-(G)| \leq 2\sqrt{d-1}$). Since $\lambda_-(G) = -\lambda_+(G)$ for bipartite graphs, this implies that about 52% of the time, bipartite graphs will be Ramanujan. Nonbipartite families behave differently. Assuming that $\lambda_+(G)$ and $\lambda_-(G)$ are independent, the probability that both are at most $2\sqrt{d-1}$ in absolute value is about 27% (52% $\cdot$ 52%). See Figure 5 for plots of the percentages and Conjecture 1.3 for exact statements of these probabilities. Unfortunately, the rate of convergence is too slow for us to see the conjectured limiting behavior.

1.3.5 Conjecture. Based on our results, we are led to the following conjecture.

Conjecture 1.3. *Let $\mathcal{F}_{N,d}$ be one of the following families of d-regular graphs: $\mathcal{CI}_{N,d}$, $\mathcal{SCI}_{N,d}$, $\mathcal{SCB}_{N,d}$ (see Remark 1.1 for definitions). The distribution of $\lambda_\pm(G)$, appropriately normalized as in* (1–2)*, converges as $N \to \infty$ to the $\beta = 1$ Tracy–Widom distribution (and not to a normalized $\beta = 2$ or $\beta = 4$ Tracy–Widom distribution or the standard normal distribution). For nonbipartite graphs, $\lambda_+(G)$ and $\lambda_-(G)$ are statistically independent. The normalization constants have $c_{\mu,N,d,\pm} < 0$ and $s_\pm(\mathcal{F}_{N,d}) > m_\pm(\mathcal{F}_{N,d})$, implying that in the limit as $N \to \infty$, approximately* 52% *of the graphs in the bipartite families and* 27% *otherwise are Ramanujan (i.e., $\lambda(G) \leq 2\sqrt{d-1}$); the actual percentage for the bipartite graphs is the percentage of mass in a $\beta = 1$ Tracy–Widom distribution to the left of the mean (to six digits, it is* 51.9652%*), and the square of this otherwise.*

Remark 1.4. The evidence for the above conjecture is very strong for three families. While the conjecture is likely to be true for the connected bipartite graphs as well, different behavior is observed for smaller N, though this may simply indicate a slower rate of convergence. For example, when we studied the percentage of eigenvalues to the left of the sample mean, this was the only family for which we did not obtain good fits to the normalized $\beta = 1$ Tracy–Widom distribution for $N \leq 5002$, though we did obtain good fits at $N = 10022$ (see Table 5).

2. RESULTS FOR 3-REGULAR GRAPHS

For

$$N \in \{26, 32, 40, 50, 64, 80, 100, 126, 158, 200, 252, 316, 400, 502, 632, 796, 1002, 1262, 1588, 2000, 2516, 3168, 3990, 5022, 6324, 7962, 10022, 12618, 15886, 20000\},$$

we randomly chose 1000 3-regular graphs from the families $\mathcal{CI}_{N,3}$, $\mathcal{SCI}_{N,3}$, $\mathcal{CB}_{N,3}$, and $\mathcal{SCB}_{N,3}$. We analyzed the distributions of $\lambda_\pm(G)$ for each sample using MATLAB's `eigs` function and investigated whether it is well modeled by the $\beta = 1$ Tracy–Widom distribution. Further, we calculated what percentage of graphs were Ramanujan as well as what percentage of graphs had $|\lambda_\pm(G)|$ less than the sample mean; these statistics help elucidate the behavior as the number of vertices tends to infinity.

2.1 Distribution of $\lambda_\pm(G)$

In Figure 6 we plot the histogram distribution of $\lambda_+(G)$ for $\mathcal{CI}_{N,3}$; the plots for the other families and for $\lambda_-(G)$ are similar. This is a plot of the actual eigenvalues. To determine whether the $\beta = 1$ Tracy–Widom distribution (or another value of β or even a normal distribution) gives a good fit to the data, we rescale the samples to

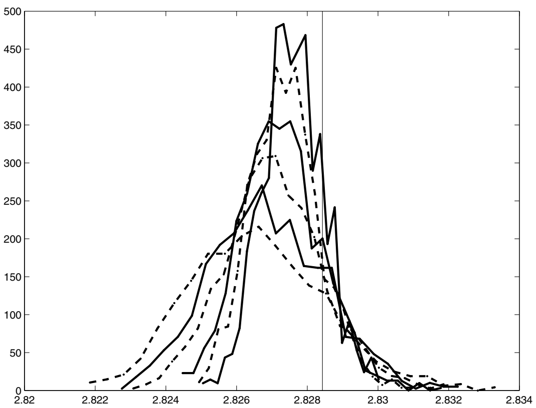

FIGURE 6. Distribution of $\lambda_+(G)$ for 1000 graphs randomly chosen from the ensemble $\mathcal{CI}_{N,3}$ for various N. The vertical line is $2\sqrt{2}$ and $N \in \{3990, 5022, 6324, 7962, 10022, 12618\}$. The curve with the lowest maximum value corresponds to $N = 3990$, and as N increases, the maximum value increases (so $N = 12618$ corresponds to the curve with greatest maximum value).

have mean 0 and variance 1, and then compare the results to scaled Tracy–Widom distributions (and the standard normal). In Table 2 we study the χ^2 values for the fits from the three Tracy–Widom distributions and the normal distribution.

As Table 2 shows, the three normalized Tracy–Widom distributions all give good fits, and even the standard normal gives a reasonable fit.[13] We divided the data into 20 bins and calculated the χ^2 values; with 19 degrees of freedom. The $\alpha = 0.05$ threshold is 30.1435, and the $\alpha = 0.01$ threshold is 36.1908.[14] We investigate below another statistic that is better able to distinguish the four candidate distributions. We note that the normalized $\beta = 1$ distribution gives good fits as $N \to \infty$ for all the families, as indicated by Table 3. The fits are good for modest N for all families but the connected bipartite graphs; there the fit is poor until N is large. This indicates that the connected bipartite graphs may have slower convergence properties than the other families.

In Table 1 we have listed the mass to the left of the mean for the Tracy–Widom distributions; it is 0.519652 for $\beta = 1$, 0.515016 for $\beta = 2$, and 0.511072 for $\beta = 4$ (note that it is 0.5 for the standard normal). Thus looking at the mass to the left of the sample mean provides a way to distinguish the four candidate distributions; we present the results of these computations for each set of 1000 graphs from $\mathcal{CI}_{N,3}$ in Table 4 (the other families behave similarly). If θ_{obs} is the observed percentage of the sample data (of size 1000) below the sample mean, then the z statistic

$$z \;=\; (\theta_{\text{obs}} - \theta_{\text{pred}}) \Big/ \sqrt{\theta_{\text{pred}} \cdot (1 - \theta_{\text{pred}})/1000}$$

measures whether the data support that θ_{pred} is the percentage below the mean.

While the data in Table 4 suggest that the $\beta = 1$ Tracy–Widom is the best fit, the other three distributions provide good fits as well. Since we expect the fit to improve as N increases, the last few rows of the table are the most important. In five of the last ten rows, the smallest z statistic is with the $\beta = 1$ Tracy–Widom distribution. Further, the average of the absolute values of the z-values for the last ten rows are 0.861 ($\beta = 1$), 0.873 ($\beta = 2$), 0.960 ($\beta = 4$), and 1.341 (for the standard normal), again supporting the claim that the best fit is from the $\beta = 1$ Tracy–Widom distribution.

In order to obtain more-conclusive evidence as to which distribution best models the second-largest normalized eigenvalue, we considered larger sample sizes (100,000 instead of 1000) for all four families; see Table 5 for the analysis. While there is a sizable increase in run time (it took on the order of a few days to run the simulations for the three different values of N for the four families), we gain a decimal digit of precision in estimating the percentages. This will allow us to statistically distinguish the four candidate distributions.

This is the most important test in the paper. The results are striking, and strongly support that only the $\beta = 1$ Tracy–Widom distribution models $\lambda_\pm(G)$ (the results for $\lambda_-(G)$ were similar to those for $\lambda_+(G)$). Except for $\mathcal{SCB}_{1002,3}$, for each of the families and each N, the z statistic increases in absolute value as we move from $\beta = 1$ to $\beta = 2$ to $\beta = 4$ to the standard normal. Further, the z-values indicate excellent fits with the $\beta = 1$ distribution for all N and all families *except* the 3-regular connected bipartite graphs; no other value of β or the standard normal gives as good a fit. In fact, the other fits are often terrible. The $\beta = 4$ and standard normal typically have z-values greater than 4; $\beta = 2$ gives a better fit, but it is significantly worse than $\beta = 1$.

Thus, except for 3-regular connected bipartite graphs, the data are consistent only with a $\beta = 1$ Tracy–Widom distribution. In the next subsections we shall study the sample means, standard deviations, and percentage of

[13] While the data displayed above are for $\lambda_+(G)$, the χ^2 values for $\lambda_-(G)$ are comparable.

[14] We could use the (pessimistic) Bonferroni adjustments for multiple comparisons (for ten comparisons these numbers become 38.5822 and 43.8201); we do not do this, because the fits are already quite good.

graphs in a family that are Ramanujan. We shall see that the 3-regular connected bipartite graphs consistently behave differently from the other three families (see in particular Figure 5).

2.2 Means and Standard Deviations

In Figure 4 we have plotted the sample means of sets of 1000 3-regular graphs chosen randomly from $\mathcal{CI}_{N,3}$ (connected perfect matchings), $\mathcal{SCI}_{N,3}$ (simple connected perfect matchings), $\mathcal{CB}_{N,3}$ (connected bipartite), and $\mathcal{SCB}_{N,3}$ (simple connected bipartite) against the numbers of vertices.

Because of analogies with similar systems whose largest eigenvalue satisfies a Tracy–Widom distribution, we expect the normalization factor for the second-largest eigenvalue to be similar to that in (1–3). Since we do not expect that the factors will still be $N^{1/2}$ and $N^{1/6}$, we consider the general normalization given in (1–2); for a 3-regular graph in one of our families, we study

$$\widetilde{\lambda_\pm}(G) \;=\; \frac{|\lambda_\pm(G)| - 2\sqrt{2} + c_{\mu,N,3,\pm} N^{m_\pm(\mathcal{F}_{N,3})}}{c_{\sigma,N,3,\pm} N^{s_\pm(\mathcal{F}_{N,3})}}.$$

Remark 2.1. The most important parameters are the exponents $m_\pm(\mathcal{F}_{N,3})$ and $s_\pm(\mathcal{F}_{N,3})$; previous work [Friedman 03] (and our investigations) suggests that $c_{\mu,N,3,\pm} < 0$. Let us assume that in the limit as the number of vertices tends to infinity, the distributions of $|\lambda_\pm(G)|$ converge to the $\beta = 1$ Tracy–Widom distribution and that $c_{\mu,N,3,\pm} < 0$. If $s_\pm(\mathcal{F}_{N,3}) > m_\pm(\mathcal{F}_{N,3})$, then in the limit we expect about 52% of the graphs to have $\lambda_+(G) \le 2\sqrt{2}$ (and similarly for $|\lambda_-(G)|$), since this is the mass of the $\beta = 1$ Tracy–Widom distribution to the left of the mean. To see why this is true, note that if $\mu_{\mathcal{F}_{N,3,+}}$ and $\sigma_{\mathcal{F}_{N,3,+}}$ are the mean and standard deviation of the data set of $\lambda_+(G)$ for all $G \in \mathcal{F}_{N,3}$, then $\mu_{\mathcal{F}_{N,3,+}} \approx 2\sqrt{2} - c_{\mu,N,d,+} N^{m_+(\mathcal{F}_{N,3})}$ and $\sigma_{\mathcal{F}_{N,3,+}} \approx c_{\sigma,N,3,+} N^{s_+(\mathcal{F}_{N,3})}$, so

$$2\sqrt{2} \;\approx\; \mu_{\mathcal{F}_{N,3,+}} + \frac{c_{\mu,N,3,+}}{c_{\sigma,N,3,+}} \cdot N^{m_+(\mathcal{F}_{N,3}) - s_+(\mathcal{F}_{N,3})} \cdot \sigma_{\mathcal{F}_{N,3,+}}.$$

Thus the Ramanujan threshold, $2\sqrt{2}$, will fall approximately $\frac{c_{\mu,N,3,+}}{c_{\sigma,N,3,+}} N^{m_+(\mathcal{F}_{N,3}) - s_+(\mathcal{F}_{N,3})}$ standard deviations away from the mean. In the limit as N goes to infinity, we see that the threshold falls to zero to the right of the mean if $m_+(\mathcal{F}_{N,3}) < s_+(\mathcal{F}_{N,3})$, but infinitely many standard deviations if $m_+(\mathcal{F}_{N,3}) > s_+(\mathcal{F}_{N,3})$.

We record (some of) the best-fit exponents in Table 6; the remaining values are similar. To simplify the calculations, we changed variables and did a log-log plot.

Several trends can be seen from the best-fit exponents in Table 6. Most of the time, $s_\pm(\mathcal{F}_{N,3}) > m_\pm(\mathcal{F}_{N,3})$, which indicates that it is more likely in the limit that 52% (and not all) of the bipartite graphs are Ramanujan (and about 27% of the nonbipartite). Except for $\mathcal{CB}_{N,3}$ (connected bipartite graphs), only once do we have $s_+(\mathcal{F}_{N,3}) < m_+(\mathcal{F}_{N,3})$; for $\mathcal{CB}_{N,3}$, we have $s_+(\mathcal{F}_{N,3}) < m_+(\mathcal{F}_{N,3})$ approximately half of the time. Further, the best-fit exponents $s_+(\mathcal{F}_{N,3})$ and $m_+(\mathcal{F}_{N,3})$ are mostly monotonically increasing with increasing N (recall that all exponents are negative), and $c_{\mu,N,3,+}$ and $c_{\sigma,N,3,+}$ do not seem to get too large or small (these are the least important of the parameters, and are dwarfed by the exponents).

This suggests that either the relationship is more complicated than we have modeled, or N is not large enough for us to see the limiting behavior. While our largest N is 20,000, log(20000) is only about 10. Thus we may not have gone far enough to see the true behavior. If the correct parameter is $\log N$, it is unlikely that larger simulations will help.

In Figure 7 we have plotted the N-dependence of the logarithm of the difference of the mean from $2\sqrt{2}$ versus the logarithm of $-c_{\mu,N,3,+} N^{m_+(\mathcal{CI}_{N,3})}$, as well as the best-fit lines obtained using all the data and just the last ten data points. As the plot shows, the slope of the best-fit line (the key parameter for our investigations) noticeably changes in the region we investigate, suggesting either that we have not gone high enough to see the limiting, asymptotic behavior or that it is not precisely linear.

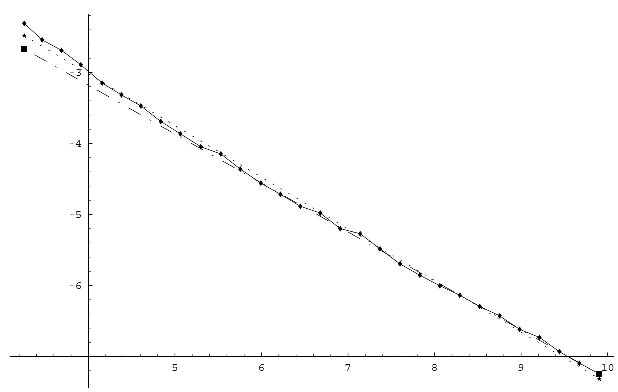

FIGURE 7. Dependence on N of the logarithm of the mean of $\lambda_+(G)$ versus $\log\left(-c_{\mu,N,3,+} N^{m_+(\mathcal{CI}_{N,3})}\right)$, showing the best-fit lines using all 30 values of N as well as just the last 10 values.

N	$\mathcal{CI}_{N,3}$	$\mathcal{SCI}_{N,3}$	$\mathcal{CB}_{N,3}$	$\mathcal{SCB}_{N,3}$	$\mathcal{CI}_{N,3}$	$\mathcal{SCI}_{N,3}$	$\mathcal{CB}_{N,3}$	$\mathcal{SCB}_{N,3}$
$\{26,\ldots,20000\}$	−0.792	−0.830	−0.723	−0.833	−0.718	−0.722	−0.709	−0.729
$\{80,\ldots,20000\}$	−0.756	−0.790	−0.671	−0.789	−0.701	−0.700	**−0.697**	−0.706
$\{252,\ldots,20000\}$	−0.727	−0.761	−0.638	−0.761	−0.695	−0.688	**−0.688**	−0.696
$\{26,\ldots,64\}$	−1.045	−1.097	−1.065	−1.151	−0.863	−0.906	−0.794	−0.957
$\{80,\ldots,200\}$	−0.887	−0.982	−0.982	−0.968	−0.769	−0.717	−0.719	−0.750
$\{232,\ldots,632\}$	−0.801	−0.885	−0.737	−0.842	−0.688	−0.713	−0.714	−0.734
$\{796,\ldots,2000\}$	−0.771	−0.819	−0.649	−0.785	−0.606	−0.719	**−0.705**	−0.763
$\{2516,\ldots,6324\}$	−0.745	−0.788	−0.579	−0.718	−0.714	−0.671	**−0.770**	−0.688
$\{7962,\ldots,20000\}$	−0.719	−0.692	−0.584	−0.757	−0.592	**−0.707**	**−0.671**	−0.648

TABLE 6. The graph sizes are chosen from {26, 32, 40, 50, 64, 80, 100, 126, 158, 200, 252, 316, 400, 502, 632, 796, 1002, 1262, 1588, 2000, 2516, 3168, 3990, 5022, 6324, 7962, 10022, 12618, 15886, 20000}. The first four columns are the best-fit values of $m(\mathcal{F}_{N,3})$; the last four columns are the best-fit values of $s(\mathcal{F}_{N,3})$. Bold entries are those for which $s(\mathcal{F}_{N,3}) < m(\mathcal{F}_{N,3})$; all other entries are for $s(\mathcal{F}_{N,3}) > m(\mathcal{F}_{N,3})$.

2.3 Independence of $\lambda_\pm(G)$ in Nonbipartite Families

In determining what percentage of graphs in a nonbipartite family is Ramanujan, it is important to know whether $\lambda_+(G)$ and $\lambda_-(G)$ are statistically independent as G varies in a family. For example, if they are perfectly correlated, the percentage could be 100%, while if they are perfectly anticorrelated, it could be 0%.

In Figure 3 we have plotted the sample correlation coefficient for $\lambda_\pm(G)$ for the nonbipartite families. For $\mathcal{CI}_{N,3}$ the values are generally positive and decreasing with increasing N; for $\mathcal{SCI}_{N,3}$ the data appear uncorrelated, with very small coefficients oscillating about zero. As another test we compared the product of the observed probabilities that $\lambda_+(G) < 2\sqrt{2}$ and $|\lambda_-(G)| < 2\sqrt{2}$ to the observed probability that $\lambda(G) < 2\sqrt{2}$; these values were virtually identical, which is what we would expect if $\lambda_\pm(G)$ are statistically independent.

2.4 Percentage of Graphs That Are Ramanujan

In Figure 5 we have plotted the percentage of graphs in each sample of 1000 from the four families that are Ramanujan (the first plot is the percentage against the number of vertices; the second is the percentage against the logarithm of the number of vertices). The most interesting observation is that for the most part, the probability that a random graph from the bipartite families is Ramanujan decreases as N increases, while the probability that a random graph from the nonbipartite families is Ramanujan oscillates in the range.

ACKNOWLEDGMENTS

We thank Alex Barnett, Jon Bober, Peter Sarnak, Brad Weir, and the participants of the AMS Special Session on Expanders and Ramanujan Graphs: Construction and Applications at the 2008 Joint Meetings (organized by Michael T. Krebs, Anthony M. Shaheen, and Audrey A. Terras) for many enlightening discussions, Craig Tracy for sharing Mathematica code to compute the Tracy–Widom distributions, and the information technology managers at the mathematics departments at Princeton, the Courant Institute, and Brown University for help in getting all the programs to run compatibly. The first author was partly supported by NSF grant DMS0600848.

REFERENCES

[Alhassid 00] Y. Alhassid. "The Statistical Theory of Quantum Dots." *Rev. Mod. Phys.* 72:4 (2000), 895–968.

[Alon 86] N. Alon. "Eigenvalues and Expanders." *Combinatorica* 6:2 (1986), 83–96.

[Alon and Milman 85] N. Alon and V. Milman. "λ_1, Isoperimetric Inequalities for Graphs, and Superconcentrators." *J. Combin. Theory Ser. B* 38:1 (1985), 73–88.

[Baik and Rains 00] J. Baik and E. M. Rains. *Limiting distributions for a polynuclear growth model*, J. Stat. Phys. 100 (2000), 523–541.

[Baik and Rains 01a] J. Baik and E. M. Rains. "The Asymptotics of Monotone Subsequences of Involutions." *Duke Math. J.* 109 (2001), 205–281.

[Baik and Rains 01b] J. Baik and E. M. Rains. "Symmetrized Random Permutations." In *Random Matrix Models and Their Applications*, edited by P. Bleher and A. Its, pp. 1–19, Math. Sci. Res. Inst. Publications 40. Cambridge: Cambridge University Press, 2001.

[Baik et al. 99] J. Baik, P. Deift, and K. Johansson. "On the Distribution of the Length of the Longest Increasing Subsequence of Random Permutations." *J. Amer. Math. Soc.* 12 (1999), 1119–1178.

[Baryshnikov 01] Yu. Baryshnikov. "GUEs and Queues." *Probab. Th. Rel. Fields* 119 (2001), 256–274.

[Bien 89] F. Bien. "Constructions of Telephone Networks by Group Representations." *Notices of the AMS* 36:1 (1989), 5–22.

[Bollobás 01] B. Bollobás. *Random Graphs.* Cambridge: Cambridge University Press, 2001.

[Borodin et al. 00] A. Borodin, A. Okounkov, and G. Olshanski. "Asymptotics of Plancherel Measures for Symmetric Groups." *J. Amer. Math. Soc.* 13 (2000), 481–515.

[Chiu 92] P. Chiu. "Cubic Ramanujan Graphs. *Combinatorica* 12:3 (1992), 275–285.

[Davidoff et al. 03] G. Davidoff, P. Sarnak, and A. Valette. *Elementary Number Theory, Group Theory, and Ramanujan Graphs*, London Mathematical Society, Student Texts 55. Cambridge: Cambridge University Press, 2003.

[Dodziuk 84] J. Dodziuk. "Difference Equations, Isoperimetric Inequality and Transience of Certain Random Walks." *Trans. Amer. Math. Soc.* 284:2 (1984), 787–794.

[Friedman 93] J. Friedman. "Some Geometric Aspects of Graphs and Their Eigenfunctions." *Duke Math. J.* 69:3 (1993), 487–525.

[Friedman 03] J. Friedman. "A Proof of Alon's Second Eigenvalue Conjecture." In *Proceedings of the Thirty-Fifth Annual ACM Symposium on Theory of Computing*, pp. 720–724 (electronic). New York: ACM, 2003.

[Gaudin 61] M. Gaudin. "Sur la loi limite de l'espacement des valeurs propres d'une matrice aléatoire." *Nucl. Phys.* 25 (1961), 447–458.

[Goldreich et al. 90] O. Goldreich, R. Impagliazzo, L. Levin, R. Venkatesan, and D. Zuckerman. "Security Preserving Amplification of Hardness." In *31st Annual Symposium on Foundations of Computer Science (St. Louis, MO, 1990)*, Vols. I, II, pp. 318–326. Los Alamitos, CA: IEEE Comput. Soc. Press, 1990.

[Gravner et al. 01] J. Gravner, C. A. Tracy, and H. Widom, "Limit Theorems for Height Fluctuations in a Class of Discrete Space and Time Growth Models." *J. Stat. Phys.* 102 (2001), 1085–1132.

[Hoory et al. 06] S. Hoory, N. Linial, and A. Wigderson, "Expander Graphs and Their Applications." *Bull. Amer. Math. Soc.* 43 (2006), 439–561.

[Jakobson et al. 99] D. Jakobson, S. D. Miller, I. Rivin, and Z. Rudnick. "Eigenvalue Spacings for Regular Graphs." In *Emerging Applications of Number Theory (Minneapolis, 1996)*, pp. 317–327, IMA Volumes in Mathematics and Its Applications 109. New York: Springer, 1999.

[Johansson 01] K. Johansson. "Discrete Orthogonal Polynomial Ensembles and the Plancherel Measure." *Ann. Math.* 153 (2001), 259–296.

[Johansson 02a] K. Johansson. "Non-intersecting Paths, Random Tilings and Random Matrices." *Probab. Th. Rel. Fields* 123 (2002), 225–280.

[Johansson 02b] K. Johansson. *Toeplitz determinants, random growth and determinantal processes*, ICM Vol. III (2002), 53–62.

[Keating and Snaith 00] J. P. Keating and N. C. Snaith. "Random Matrix Theory and $\zeta(1/2+it)$." *Comm. Math. Phys.* 214:1 (2000), 57–89.

[Lubotzky et al. 88] A. Lubotzky, R. Phillips, and P. Sarnak. "Ramanujan Graphs." *Combinatorica* 8:3 (1988), 261–277.

[Margulis 88] G. A. Margulis. "Explicit Group-Theoretic Constructions of Combinatorial Schemes and Their Applications in the Construction of Expanders and Concentrators" (Russian). *Problemy Peredachi Informatsii* 24:1 (1988), 51–60; translation in *Problems Inform. Transmission* 24:1 (1988), 39–46.

[McKay 81] B. McKay. "The Expected Eigenvalue Distribution of a Large Regular Graph." *Linear Algebra Appl.* 40 (1981), 203–216.

[Mehta 60] M. Mehta. "On the Statistical Properties of Level Spacings in Nuclear Spectra." *Nucl. Phys.* 18 (1960), 395–419.

[Mehta 91] M. Mehta. *Random Matrices*, 2nd edition. Boston: Academic Press, 1991.

[Morgenstern 94] M. Morgenstern. "Existence and Explicit Constructions of $q+1$ Regular Ramanujan Graphs for Every Prime Power q." *J. Combin. Theory Ser. B* 62:1 (1994), 44–62.

[Murty 03] M. Ram Murty, "Ramanujan Graphs." *J. Ramanujan Math. Soc.* 18:1 (2003), 33–52.

[Odlyzko 87] A. Odlyzko. "On the Distribution of Spacings between Zeros of the Zeta Function." *Math. Comp.* 48:177 (1987), 273–308.

[Odlyzko 01] A. Odlyzko. "The 10^{22}-nd Zero of the Riemann Zeta Function." In *Proc. Conference on Dynamical, Spectral and Arithmetic Zeta-Functions*, edited by M. van Frankenhuysen and M. L. Lapidus, Amer. Math. Soc., Contemporary Math. Series. Available online (http://www.research.att.com/~amo/doc/zeta.html), 2001.

[Pippenger 77] N. Pippenger. "Super Concentrators." *SIAM Journal Comp.* 6 (1977), 298–304.

[Prähofer and Spohn 00a] M. Prähofer and H. Spohn. "Statistical Self-Similarity of One-Dimensional Growth Processes." *Physica A* 279 (2000), 342–352.

[Prähofer and Spohn 00b] M. Prähofer and H. Spohn. "Universal Distributions for Growth Processes in $1+1$ Dimensions and Random Matrices." *Phys. Rev. Lett.* 84 (2000), 4882–4885.

[Sarnak 90] P. Sarnak. *Some Applications of Modular Forms*, Cambridge Trusts in Mathematics, 99, Cambridge: Cambridge University Press, 1990.

[Sarnak 04] P. Sarnak. "What Is an Expander?" *Notices of the AMS* 51:7 (2004), 762–763.

[Soshnikov 08] A. Soshnikov. "A Note on Universality of the Distribution of the Largest Eigenvalue in Certain Classes of Sample Covariance Matrices. Preprint, arXiv: math.PR/0104113, 2008.

[Sipser and Spielman 96] M. Sipser and Spielman, "Expander Codes." *IEEE Trans. Inform. Theory* 42:6 (1996), 1710–1722.

[Tracy and H. Widom 94] C. A. Tracy and H. Widom. "Level-Spacing Distributions and the Airy Kernel." *Comm. Math. Phys.* 159 (1994), 151–174.

[Tracy and Widom 96] C. Tracy and H. Widom. "On Orthogonal and Sympletic Matrix Ensembles. *Communications in Mathematical Physics* 177 (1996), 727–754.

[Tracy and Widom 02] C. Tracy and H. Widom. "Distribution Functions for Largest Eigenvalues and Their Applications." *ICM Vol. I* (2002), 587–596.

[Vavilov et al. 01] M. G. Vavilov, P. W. Brouwer, V. Ambegaokar, and C. W. J. Beenakker. "Universal Gap Fluctuations in the Superconductor Proximity Effect." *Phys. Rev. Lett.* 86 (2001), 874–877.

[Wigner 57] E. Wigner. "Statistical Properties of Real Symmetric Matrices." In *Canadian Mathematical Congress Proceedings*, pp. 174–184. Toronto: University of Toronto Press, 1957.

Steven J. Miller, Department of Mathematics, Williams College, Williamstown, MA 01267 (Steven.J.Miller@williams.edu).

Tim Novikoff, Center for Applied Mathematics, 657 Frank Rhodes Hall, Cornell University, Ithaca, NY 14853 (tnovikoff@gmail.com)

Anthony Sabelli, Center for Applied Mathematics, 657 Frank Rhodes Hall, Cornell University, Ithaca, NY 14853 (Anthony_Sabelli@brown.edu)

Received November 22, 2006; accepted September 28, 2007.