What Is... a Ramanujan Graph?

Vilas Winstein

July 26, 2022

1 The Heat Equation

1.1 In $\mathbb{R}^n$ (and other nice spaces)

Let X be a nice space (eg a subset of $\mathbb{R}^n$), and let $u(x,t)$ denote the temperature at the point $x \in X$ at time $t \in [0, \infty)$. What happens if we set $u(x,0) = f(x)$ to be some initial temperature distribution? How does the temperature evolve over time? For example:

- $X =$ a disk. (heat spreads evenly to the whole disk)
- $X =$ two disjoint disks. (heat spreads to only one disk, not the other)
- $X =$ two disjoint disks connected by a tiny strip. (heat spreads to the other disk, but slowly)
- $X =$ a hyperbolic or euclidean disk of the same *area*. (heat spreads faster in the hyperbolic disk)

So understanding how heat flows in a space can tell us about the geometry and topology of the space in a more quantitative way. Rather than just being "connected" or "disconnected," we have a way to measure *how* connected a space is.

The equation which $u(x,t)$ satisfies is the *heat equation*

$$\partial_t u = \Delta u, \qquad \text{where} \quad \Delta = \partial_{x_1}^2 + \cdots + \partial_{x_n}^2 \quad \text{is the } \textit{Laplacian} \text{ (in } \mathbb{R}^n).$$

This is a reasonable equation for heat flow, because

Theorem (Pizzetti; 1909)**.** *let $\overline{f_R}(x)$ denote the average value of f on the sphere of radius R around x. Then*

$$\lim_{R \to 0} \frac{\overline{f_R}(x) - f(x)}{R^2} = \frac{(\Delta f)(x)}{2n}.$$

So $\partial_t u = \Delta u$ says that the rate of change of temperature at a point is proportional to the difference between the temperature *near* that point and the temperature *at* that point. If you are near something hot, your temperature will increase. If you are near something cold, your temperature will decrease.

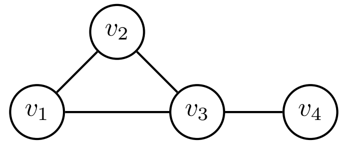
1.2 In Graphs

If you are a data scientist, you might want to understand the connectivity of a network G of data in a quantitative way. Can we do that using the heat equation? If $u(v,t)$ represents the temperature at a vertex $v \in V(G)$ at time t, what can we say about the evolution of u over time?

We can't just write $\partial_t u = \Delta u$, since a graph isn't a smooth space: it's just a collection of vertices and edges. But the theorem of Pizzetti motivates the *Normalized Discrete Laplacian*:

$$(\tilde{L} f)(v) := \overline{f}(v) - f(v),$$

where $\overline{f}(v)$ is the average value of f on all the neighbors of v. For example, with this graph:



if we represent a function $f : V(G) \to \mathbb{R}$ as a column vector

$$f = \begin{pmatrix} f(v_1) \\ f(v_2) \\ f(v_3) \\ f(v_4) \end{pmatrix} \in \mathbb{R}^{V(G)}, \quad \text{then} \quad \tilde{L}f = \begin{pmatrix} -1 & 1/2 & 1/2 & 0 \\ 1/2 & -1 & 1/2 & 0 \\ 1/3 & 1/3 & -1 & 1/3 \\ 0 & 0 & 1 & -1 \end{pmatrix} \begin{pmatrix} f(v_1) \\ f(v_2) \\ f(v_3) \\ f(v_4) \end{pmatrix}.$$

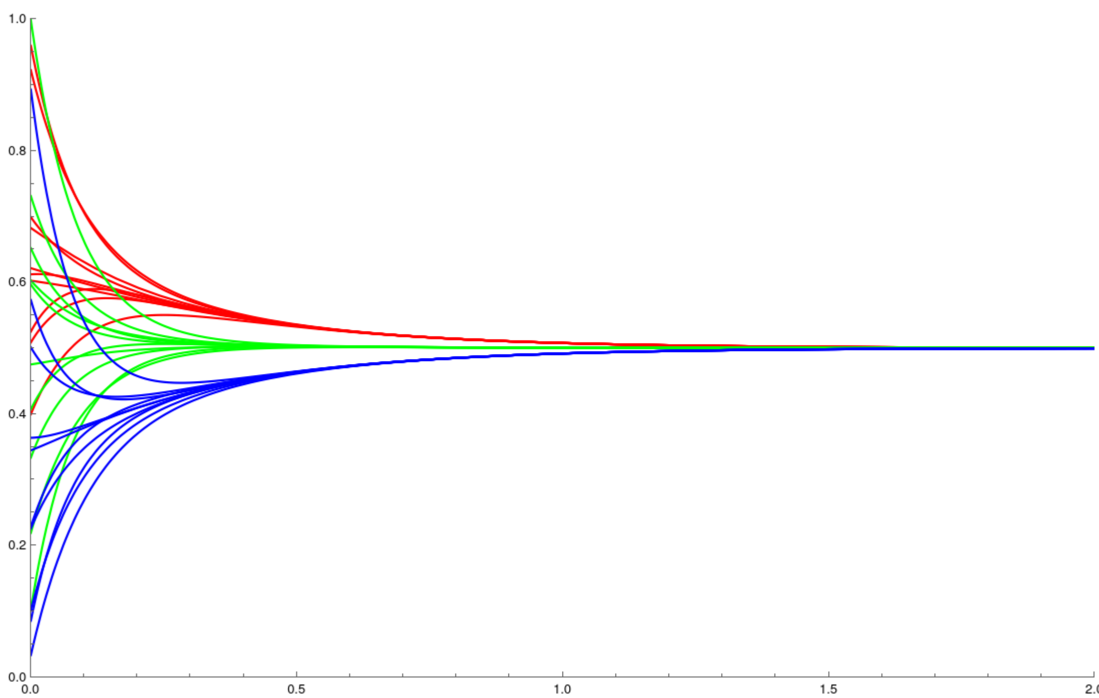
We will actually clear the denominators of the matrix $\tilde{L}$ to obtain the unnormalized *Discrete Laplacian*:

$$L = \begin{pmatrix} -2 & 1 & 1 & 0 \\ 1 & -2 & 1 & 0 \\ 1 & 1 & -3 & 1 \\ 0 & 0 & 1 & -1 \end{pmatrix},$$

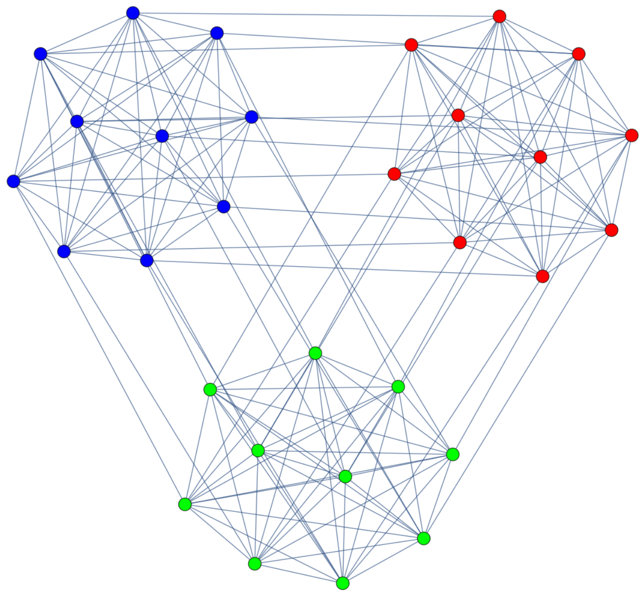
which (in general) has 1 in the (i, j) position if vertex i is connected to vertex j, and has the negatives of the degrees of the vertices along the diagonal. This will have advantages that we will see later, but for now let's consider some examples of the evolution of the *discrete heat equation* $\partial_t u = Lu$:

- $G =$ a complete graph. (heat spreads evenly very quickly)
- $G =$ a path graph. (heat spreads evenly, but more slowly)
- $G =$ a disjoint union of two complete graphs. (heat does not spread to the other component)
- $G =$ a graph with two different highly connected clusters, and a few edges between them. (heat spreads within each cluster very quickly, but only very slowly does it equalize between clusters)

The last example inspires a simple algorithm for detecting clusters in a graph of data (which may just be provided to you as a list of vertice and edges, without any nice picture drawn). First, randomize the temperatures at all vertices. Then let the heat equation evolve. You will see a picture like this:



After a while the heat equalizes across the whole network, but if you pause the evolution after a short time (for example, at time $t = 0.5$ in this plot), you might see a few distinct clusters of heat values; these correspond to the clusters of the graph. The above plot was generated using the following graph, where the color of the vertex corresponds to the color of the curve:



Of course, if we were doing this in a real application, we would not have these colors, but we could still recover the clusters by, for example, running the k-means algorithm on the set of numbers represented by the plot at time 0.5, say.

1.3 The Spectrum of the Discrete Laplacian

One advantage of clearing the denominators is that now L is a symmetric matrix, so we can easily apply the Spectral Theorem to obtain *real* eigenvalues

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n, \qquad \text{(now } n = |V(G)|\text{)}$$

and their corresponding orthonormal basis of eigenvectors $\phi_1, \ldots, \phi_n \in \mathbb{R}^{V(G)}$ satisfying

$$L\phi_i = \lambda_i \phi_i.$$

This will make studying the *discrete heat initial value problem* of finding u such that

$$\partial_t u = Lu \qquad \text{and} \qquad u(v, 0) = f(v)$$

much easier. If we write f and u in terms of ϕ-coordinates as

$$f = \sum_{i=1}^{n} f_i \cdot \phi_i, \qquad \text{and} \qquad u(t) = \sum_{i=1}^{n} u_i(t) \cdot \phi_i$$

then the discrete heat equation becomes

$$\sum_{i=1}^{n} \partial_t u_i(t) \cdot \phi_i = \sum_{i=1}^{n} u_i(t) \cdot L\phi_i = \sum_{i=1}^{n} u_i(t) \cdot \lambda_i \phi_i,$$

and so we get a bunch of ordinary differential equations

$$\partial_t u_i = \lambda_i u_i \qquad \text{with initial condition} \qquad u_i(0) = f_i.$$

This is easily solved by

$$u_i(t) = f_i e^{\lambda_i t}.$$

So, each ϕ-coefficient of u evolves separately and behaves like an exponential; in total, we have

$$u(t) = \sum_{i=1}^{n} f_i e^{\lambda_i t} \cdot \phi_i.$$

Notice that at least one of the eigenvalues is zero, since any constant function is in the kernel of the discrete Laplacian. In fact, the multiplicity of the eigenvalue 0 is exactly the same as the number of connected components of the graph. The eigenvectors corresponding to these 0 eigenvalues are all constant on connected components. Intuitively, the heat distribution should converge to something which is constant on connected components, and this means that all of the other eigenvalues should be negative, so that the coefficients of the non-constant-on-components coordinates all go to zero. Thus

$$0 = \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n.$$

1.4 Spectral Gap

Suppose now that the graph G is connected, so that λ_2 is strictly less than 0. The value of λ_2 gives a bound for how fast a heat distribution converges to the constant distribution:

$$\begin{aligned}
\|u(t) - f_1\phi_1\| &= \sqrt{\sum_{i=2}^{n} |f_i e^{\lambda_i t}|^2} \\
&\leq \sqrt{\left(\sum_{i=2}^{n} |f_i|^2\right)\left(\sum_{i=2}^{n} |e^{\lambda_i t}|^2\right)} \\
&\leq \sqrt{\|f\|^2 \cdot n \cdot e^{2\lambda_2 t}} \\
&= \sqrt{n} \cdot \|f\| \cdot e^{\lambda_2 t}.
\end{aligned}$$

So, the more negative λ_2 is, the faster the convergence to constant heat. It's reasonable to say that the speed of heat equalizing in the graph is a good quantitative measure of the "connectivity" of the graph. Thus, the absolute value of λ_2 is a good quantitative measure of the connectivity of a graph, and $|\lambda_2|$ is called the "spectral gap" of the graph, denoted by $\mathrm{sg}(G)$. Let's see some examples:

- A cycle graph C_n: $\mathrm{sg}(C_n) = 2 - 2\cos\left(\frac{2\pi}{n}\right)$, so $\mathrm{sg}(C_n) \to 0$ as $n \to \infty$.
- A complete graph K_n: $\mathrm{sg}(K_n) = n$, so $\mathrm{sg}(K_n) \to \infty$ as $n \to \infty$.

Notice that the spectral gap equals zero if and only if the graph is disconnected, so this number carries more information than just knowing whether the graph is connected or not. This is just true for finite graphs, by the way. For infinite graphs, a similar notion can be defined, and the spectral gap of an infinite graph is zero if and only if that graph is *amenable*. For example, the spectral gap of the Cayley graph of the free group on two generators is

$$4 - 2\sqrt{4-1} > 0.$$

You may have heard of amenability only for groups, but the Følner set definition of amenability works for infinite graphs too, and a group is amenable if and only if every *Cayley graph* of that group is amenable. The spectral gap we defined above can also be viewed as a quantitative analog of nonamenability for finite graphs. For more about the relationship between spectral gap and amenability, look up *Cheeger's Inequality*.

2 Regular Graphs

A graph is called *regular* if all of its vertices have the same degree (i.e. the number of adjacent edges). If that degree is d, the graph is *d-regular*.

2.1 Alon-Boppana Bound

A d-regular graph can have at most $\frac{dn}{2}$ edges. Compared with a complete graph, which has $\binom{n}{2}$ edges, this is quite sparse. So, one might expect that a d-regular graph cannot be *too* well connected. Indeed,

Theorem (Alon-Boppana; 1986)**.** *If (G_n) is a sequence of d-regular graphs with $|V(G_n)| \to \infty$, then*

$$\limsup_{n\to\infty} \operatorname{sg}(G_n) \leq d - 2\sqrt{d-1}.$$

A *Ramanujan graph* is a graph which is exceptional, with regards to the Alon-Boppana bound. Specifically, a Ramanujan graph is a d-regular graph G such that $\operatorname{sg}(G) \geq d - 2\sqrt{d-1}$. For example, the complete graphs are all Ramanujan graphs: they are $(n-1)$-regular and they have spectral gap

$$n, \qquad \text{which is greater than} \qquad (n-1) - 2\sqrt{n-2}.$$

However, more interesting and useful would be a *family* of growing Ramanujan graphs, all with the *same* degree. Finding such a family turns out to be much trickier.

2.2 Adjacency Matrix

There is another useful graph matrix to consider, which might be more familiar to people with computer science experience. That matrix is the *adjacency matrix* A, which has a 1 in the (i, j) entry if vertex i and j are connected, and which has a 0 in that entry otherwise. For d-regular graphs,

$$A = L + dI,$$

which means that the eigenvalues $\alpha_1 \geq \cdots \geq \alpha_n$ of A are exactly the same as the eigenvalues of L, but just shifted by d. So the highest eigenvalue of A is $\alpha_1 = d$, and the second-highest is $\alpha_2 = \lambda_2 + d$. The spectral gap is still the difference between the top two eigenvalues, but often people just refer to the second-highest eigenvalue of the adjacency matrix α_2 as the "top of the spectrum" of the adjacency matrix (or something like that).

This number α_2 is still a good measure of the connectivity of G, but now as α_2 increases, the connectivity decreases. A d-regular graph is disconnected if and only if $\alpha_2 = d$. By the way, d is also the *operator norm* of A.

One reason we want to switch the notation at this point is that the adjacency matrix carries a lot of useful combinatorial information. For example, the (i, j) entry of A^k is exactly the number of length-k paths from vertex i to vertex j in G. Also, A d-regular graph is *bipartite* if and only if the lowest eigenvalue α_n of A, is equal to $-d$, since in this case the function which is positive on one part and negative on the other is an eigenvector for this eigenvalue.

2.3 Proof of the Alon-Boppana Bound

We will actually prove a slightly weaker statement, but this proof allows us to see where the $2\sqrt{d-1}$ comes from more easily. By the way, this proof is almost directly copied from the page for "Alon-Boppana bound" on Wikipedia.

Theorem. *Let $\alpha(G) = \max(|\alpha_2(G)|, |\alpha_n(G)|)$. If (G_n) is a sequence of d-regular graphs, then*

$$\liminf_{n\to\infty} \alpha(G_n) \geq 2\sqrt{d-1}.$$

Proof. Pick $k \in \mathbb{N}$. The number of closed walks of length $2k$ in G_n is

$$\operatorname{tr}(A(G_n)^{2k}) = \sum_{i=1}^{n} \alpha_i(G_n)^{2k} \leq d^{2k} + n \cdot \alpha(G_n)^{2k}.$$

However, the number of closed walks of length $2k$ starting from a fixed vertex in in G_n is at most the corresponding number for the d-regular tree, because the d-regular tree covers G_n. The number of returning walks of length $2k$ in the d-regular tree is at least

$$C_k \cdot (d-1)^k, \qquad \text{where} \qquad C_k = \frac{1}{k+1}\binom{2k}{k}$$

is the kth Catalan number. To see this, notice that a returning path in the d-regular tree can be "projected" down to a returning path in $\mathbb{N}$, by considering the path's distance from it's starting point. The number of returning paths of length $2k$ in $\mathbb{N}$ is exactly C_k. Given a returning path in $\mathbb{N}$, to create a returning path in the d-regular tree, you just need to pick which directions to go when you are moving *away* from the starting point. There are exactly k steps where you move away, and at each step you have at least $d-1$ choices (sometimes you will have d choices, if you are standing at the starting point). Anyway, it follows that

$$\operatorname{tr}(A(G_n)^{2k}) \geq n\frac{1}{k+1}\binom{2k}{k}(d-1)^k,$$

and so

$$\alpha(G_n)^{2k} \geq \frac{1}{k+1}\binom{2k}{k}(d-1)^k - \frac{d^{2k}}{n}.$$

Now, as you can prove with Stirling's formula, as $k \to \infty$ we have

$$\binom{2k}{k} \sim \frac{2^{2k}}{\sqrt{n\pi}}.$$

If we let both n and k go to infinity but ensure that $k \ll \log(n)$, then we find that

$$\alpha(G_n) \geq 2\sqrt{d-1} - o(1),$$

and this is equivalent to what we want to prove. □

3 Ramanujan and Almost-Ramanujan Graphs

Again, a *Ramanujan graph* is a regular graph whose spectral gap is almost as large as possible. In our new notation, a Ramanujan graph is a d-regular graph G such that $\alpha_2(G) \le 2\sqrt{d-1}$.

3.1 Ramanujan Graphs: Past and Future

The first examples of Ramanujan graphs come from a number-theoretic construction, and used the *Ramanujan conjecture*, which is what led to the name.

Theorem (Lubotzky, Phillips, Sarnak; Margulis; 1988)**.** *There exists an infinite family of* $(p+1)$*-regular Ramanujan graphs, whenever* p *is prime and* $p \equiv 1 \mod 4$*.*

Construction Sketch. Let $q \equiv 1 \mod 4$ be a prime not equal to p. By Jacobi's four-square theorem, there are $p+1$ solutions to the equation

$$p = a_0^2 + a_1^2 + a_2^2 + a_3^2,$$

where $a_0 > 0$ is odd and a_1, a_2, a_3 are even. To each such solution associate the $\mathrm{PGL}(2, \mathbb{Z}/q\mathbb{Z})$ matrix

$$\begin{pmatrix} a_0 + ia_1 & a_2 + ia_3 \\ -a_2 + ia_3 & a_0 - ia_1 \end{pmatrix}, \qquad \text{where } i \text{ is a fixed solution to } i^2 \equiv -1 \mod q.$$

If p is a quadratic residue modulo q, let $X^{p,q}$ be the Cayley graph of $\mathrm{PGL}(2, \mathbb{Z}/q\mathbb{Z})$ with these $p+1$ generators, and otherwise let $X^{p,q}$ be the cayley graph of $\mathrm{PSL}(2, \mathbb{Z}/q\mathbb{Z})$ with the same generators. Then $X^{p,q}$ is a $(p+1)$-regular Ramanujan graph on $n = q(q^2-1)$ or $q(q^2-1)/2$ vertices (depending on whether or not p is a quadratic residue modulo q). □

Later, this construction was extended to prime powers.

Theorem (Morgenstern; 1994)**.** *There exists an infinite family of* $(q+1)$*-regular Ramanujan graphs, whenever* q *is a prime power and* $q \equiv 1 \mod 4$*.*

More recently, a separate construction via the probabilistic method was used to show the existence of *bipartite* Ramanujan graphs of all degrees and all sizes.

Theorem (Marcus, Spielman, Srivastava; 2013)**.** *For any* $d \ge 3$ *and any* $n \ge d+1$*, there exists a* d*-regular bipartite Ramanujan graph with* n *vertices.*

However, the bipartiteness means these graphs are Ramanujan in a weaker sense—they are not exceptional from the point of view of the weaker version of the Alon-Boppana bound which we proved in section 2.3 above, since all of their lowest eigenvalues are equal to $-d$. The question of whether there are Ramanujan graphs of every degree which are *not* bipartite is still open.

3.2 Random Regular Graphs

What are some examples of d-regular graphs? Well, one (cheap) way to get d-regular graphs is to generate them randomly. For instance, one might ask for a d-regular graph chosen *uniformly* among all d-regular graphs on n nodes. But how can one actually generate such a random graph on a computer?

The configuration model is a random algorithm for constructing a d-regular graph, and it gives an approximation to the uniformly random d-regular graph on n nodes. First, start with n nodes, and attach d half-edges to each one. Then choose a random pairing of all of the half-edges, and join paired half-edges to form a full edge. One way to choose a pairing is just to go one-by-one through the unjoined half-edges, choose a random unjoined half-edge, and join the two.

If you do this, you might notice that it's possible to get self-loops and multiple edges, which we would prefer to avoid. but, the good news is that the probability of getting too many self-loops or multiple edges is very low, even as $n \to \infty$. In fact, for all n, there is a uniformly positive probability that there will be *no* self-loops or multiple edges in the graph resulting from the configuration model.

If we condition on this positive-probability event, then we actually recover the *uniform* random d-regular graph on n nodes. Because of this, any statement that holds with asymptotic probability 1 for the configuration model also holds with asymptotic probability 1 for the uniform random d-regular graph, and vice versa.

By the way, the proof that there are few self-loops and multiple edges works in greater generality to show that there are also very few short cycles in a random regular graph. Informally, a random d-regular graph looks locally like a tree. More formally (but perhaps unhelpfully), growing random d-regular graphs converge to the infinite d-regular tree (in the sense of Benjamini and Schramm). More precisely, for any integer k, the proportion of nodes in a random d-regular graph which are part of a k-cycle goes to zero (almost surely) in any sequence of growing random d-regular graphs.

Perhaps that makes the following theorem easier to believe:

Theorem (Friedman; 2003)**.** *If (G_n) is a sequence of uniformly random d-regular graphs on n vertices (or $2n$ vertices if d is odd), then with probability 1, we have*

$$\lim_{n\to\infty} \operatorname{sg}(G_n) = d - 2\sqrt{d-1}, \quad \textit{or equivalently}, \quad \lim_{n\to\infty} \alpha_2(G_n) = 2\sqrt{d-1}.$$

It was originally conjectured by Alon in 1986. The proof is over 100 pages and is quite complicated. However, with the intuition that random regular graphs are locally tree-like, combined with knowledge of the proof of the Alon-Boppana bound which compares a regular graph to the regular tree in a local way, it is at least believable.

References

Wikipedia: Alon-Boppana Bound
https://en.wikipedia.org/wiki/Alon-Boppana_bound

Wikipedia: Cheeger Constant (Graph Theory)
https://en.wikipedia.org/wiki/Cheeger_constant_(graph_theory)

Wikipedia: Expander Graph
https://en.wikipedia.org/wiki/Expander_graph

Wikipedia: Laplace Operator
https://en.wikipedia.org/wiki/Laplace_operator

Wikipedia: Configuration Model
https://en.wikipedia.org/wiki/Configuration_model

Wikipedia: Ramanujan Graph
https://en.wikipedia.org/wiki/Ramanujan_graph