WHAT IS . . .

a Graphon?

Daniel Glasscock

Large graphs are ubiquitous in mathematics, and describing their structure is an important goal of modern combinatorics. The most famous description is given by Szemerédi's Regularity Lemma (1978), which asserts that the vertices of any large enough finite graph may be partitioned into sets of roughly equal size in such a way that the edges between most of the sets appear to be nearly random.

One way to study large, finite objects is to pass from sequences of larger and larger such objects to ideal limiting objects. When done properly, properties of the limiting objects reflect properties of the finite objects which approximate them, and vice versa.

Graphons, short for graph functions, are the limiting objects for sequences of large, finite graphs with respect to the so-called cut metric. They were introduced and developed by C. Borgs, J. T. Chayes, L. Lovász, V. T. Sós, B. Szegedy, and K. Vesztergombi in [1] and [3]. Graphons arise naturally wherever sequences of large graphs appear: extremal graph theory, property testing of large graphs, quasi-random graphs, random networks, the thermodynamic limit of statistical physics systems, et cetera.

Let's begin with a few definitions and a motivating example. A *graph* G is a set of vertices $V(G)$ and a set of edges $E(G)$ whose elements are pairs of distinct vertices. A *graph homomorphism* from H to G is a map from $V(H)$ to $V(G)$ that preserves edge adjacency; that is, for every edge $\{v, w\}$ in $E(H)$, the edge $\{\varphi(v), \varphi(w)\}$ is in $E(G)$. Denote by $\hom(H, G)$ the number of homomorphisms from H to G. For example, $\hom(\bullet, G) = |V(G)|$, $\hom(\bullet\!-\!\bullet, G) = 2|E(G)|$, and $\hom(\triangle, G)$ is 6 times the number of triangles in G. Normalizing by the total number of possible maps, we get the *homomorphism density* of H into G,

$$t(H, G) = \frac{\hom(H, G)}{|V(G)|^{|V(H)|}},$$

the probability that a randomly chosen map from $V(H)$ to $V(G)$ preserves edge adjacency. This number also represents the density of H as a subgraph in G asymptotically as $n = |V(G)| \to \infty$. For example, $t(\bullet\!-\!\bullet, G) = 2|E(G)|/n^2$, while the density of edges in G is $2|E(G)|/n(n-1)$; these two expressions are nearly the same when n is large.

Consider the following problem from extremal graph theory:

How many 4-cycles must there be in a graph with edge density at least $1/2$*?*

It is easy to see that there are at most on the order of n^4 4-cycles in any graph with n vertices. A theorem of Erdős says that graphs with at least half the number of possible edges have *at least* on the order of n^4 4-cycles. More precisely, for any graph G,

$$t(\square, G) \geq t(\bullet\!-\!\bullet, G)^4;$$

in particular, if $t(\bullet\!-\!\bullet, G) \geq 1/2$, then $t(\square, G) \geq 1/16$. In light of this, the problem can be reformulated as a minimization one: *Minimize* $t(\square, G)$ *over finite graphs* G *satisfying* $t(\bullet\!-\!\bullet, G) \geq 1/2$. With some work, it can be shown that no finite graph G with $t(\bullet\!-\!\bullet, G) \geq 1/2$ achieves the minimum $t(\square, G) = 1/16$.

It's useful at this point to draw an analogy with a problem from elementary analysis: *Minimize* $x^3 - 6x$ *over rational numbers* x *satisfying* $x \geq 0$. This polynomial has a unique minimum on $[0, \infty)$ at $x = \sqrt{2}$, so the best we can do over the rationals is show that the polynomial achieves values approaching this minimum along a sequence of rationals approaching $\sqrt{2}$. We know well to avoid this complication by completing the rational numbers to the reals and realizing the limit of such a sequence as $\sqrt{2}$.

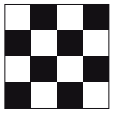
Daniel Glasscock is a graduate student in mathematics at The Ohio State University. His email address is glasscock.4@osu.edu.

DOI: http://dx.doi.org/10.1090/noti1204

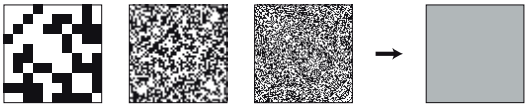
There is a sequence of finite graphs with edge density at least $1/2$ and 4-cycle density approaching $1/16$. Let R_n be an instance of a random graph on n vertices where each edge is decided independently with probability $1/2$. Throwing away those R_n's for which $t(\bullet\!\!-\!\!\bullet, R_n) < 1/2$, the 4-cycle density in the remaining graph sequence converges almost surely to $1/16$. Following the $\sqrt{2}$ analogy, we should look to realize the limit of this sequence of finite graphs and understand how it solves the minimization problem at hand.

What might the limit of this sequence of random graphs $(R_n)_n$ be? From the adjacency matrix of a labeled graph, construct the graph's *pixel picture* by turning the 1's into black squares, erasing the 0's, and scaling to the unit square $[0,1]^2$:

$$\begin{pmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \end{pmatrix} \longrightarrow$$

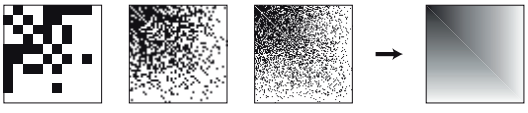


Pixel pictures can be seen to "converge" graphically; those of larger and larger random graphs with edge probability $1/2$, regardless of how they are labeled, seem to converge to a gray square, the constant $1/2$ function on $[0,1]^2$.

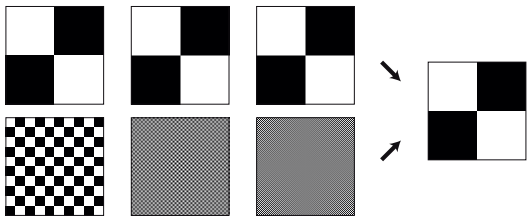


The constant $1/2$ function on $[0,1]^2$ is an example of a labeled graphon. A *labeled graphon* is a symmetric, Lebesgue-measurable function from $[0,1]^2$ to $[0,1]$ (modulo the usual identification almost everywhere); such functions can be thought of as edge-weighted graphs on the vertex set $[0,1]$. An *unlabeled graphon* is a graphon up to relabeling, where a *relabeling* is the result of applying an invertible, measure-preserving transformation to the $[0,1]$ interval. Note that any pixel picture is a labeled graphon, meaning that (labeled) graphs are (labeled) graphons.

As another example of this convergence, consider the *growing uniform attachment* graph sequence $(G_n)_n$ defined inductively as follows. Let $G_1 = \bullet$. For $n \geq 2$, construct G_n from G_{n-1} by adding one new vertex, then including each edge not already in $E(G_{n-1})$ independently with probability $1/n$. It turns out that this graph sequence converges almost surely to the graphon $1 - \max(x, y)$. (Since matrices are indexed with $(0,0)$ in the top left corner, so too are graphons.)



There are two natural ways to label a complete bipartite graph, and each suggests a different limit graphon for the complete bipartite graph sequence. Both sequences of labeled graphons in fact have the same limit, as indicated in the diagram; the reader is encouraged to return to this example after we define this convergence more precisely.



Homomorphism densities extend naturally to graphons. For a finite graph G, the density $t(\bullet\!\!-\!\!\bullet, G)$ can be computed by giving each vertex of G a mass of $1/n$ and integrating the edge indicator function over all pairs of vertices. In exactly the same way, the edge density $t(\bullet\!\!-\!\!\bullet, W)$ of a labeled graphon W is

$$\int_{[0,1]^2} W(x,y)\, dxdy,$$

and the 4-cycle density $t(\square, W)$ is

$$\int_{[0,1]^4} W(x_1,x_2)W(x_2,x_3)\, W(x_3,x_4)W(x_4,x_1)\, dx_1dx_2dx_3dx_4.$$

It is straightforward from here to write the expression for the homomorphism density $t(H, W)$ of a finite graph H into a graphon W. This allows us to see how the constant graphon $W \equiv 1/2$ solves the minimization problem: $t(\bullet\!\!-\!\!\bullet, W) = 1/2$ while $t(\square, W) = 1/16$.

To see the space of graphons as the completion of the space of finite graphs and make graphon convergence precise, define the *cut distance* $\delta_\square(W, U)$ between two labeled graphons W and U by

$$\inf_{\varphi,\psi} \sup_{S,T} \left| \int_{S\times T} W(\varphi(x),\varphi(y)) - U(\psi(x),\psi(y))\, dxdy \right|,$$

where the infimum is taken over all relabelings φ of W and ψ of U, and the supremum is taken over all measurable subsets S and T of $[0,1]$. The cut distance first measures the maximum discrepancy between the integrals of two labeled graphons over measurable boxes (hence the $\square$) of $[0,1]^2$, then minimizes that maximum discrepancy over all possible relabelings. (It is possible to define the cut distance between two finite graphs combinatorially, without any analysis, but the definition is quite involved.)

The infimum in the definition of the cut distance makes $\delta_\square$ well defined on the space of unlabeled graphons, but it is not yet a metric. Graphons W

and U for which $t(H,W) = t(H,U)$ for all finite graphs H are called *weakly isomorphic*; it turns out that W and U are weakly isomorphic if and only if $\delta_\square(W,U) = 0$. The cut distance becomes a genuine metric on the space $\mathcal{G}$ of unlabeled graphons up to weak isomorphism. The examples of pixel picture convergence above provide examples of convergent sequences and their limits in $\mathcal{G}$ (up to weak isomorphism).

We conclude by highlighting a few foundational results about graphons. The accompanying page references are from Lovász's book [2], to which the interested reader is encouraged to refer for more details.

Theorem 1 (Prop. 11.32, p. 185)**.** *Every graphon is the $\delta_\square$-limit of a sequence of finite graphs.*

To approximate a labeled graphon W by a finite labeled graph, let S be a set of n randomly chosen points from $[0,1]$, then construct a graph on S in which the edge $\{s_i, s_j\}$ is included with probability $W(s_i, s_j)$. With high probability (as $|S| \to \infty$), this labeled graph approximates W well in cut distance.

Theorem 2 (Thm. 9.23, p. 149)**.** *The space $(\mathcal{G}, \delta_\square)$ is compact.*

This implies that $(\mathcal{G}, \delta_\square)$ is a complete metric space. Combining this fact with Theorem 1, we see that the space of graphons is the completion of the space of finite graphs with the cut metric! This theorem also demonstrates the way in which graphons provide a bridge between different forms of Szemerédi's Regularity Lemma: Theorem 2 may be deduced from a weak form of the lemma, while a stronger regularity lemma follows from the compactness of $\mathcal{G}$.

Theorem 3 (Lem. 10.23, p. 167)**.** *For every finite graph H, the map $t(H,\cdot) : \mathcal{G} \to [0,1]$ is Lipschitz continuous.*

Theorems 2 and 3 combine with elementary analysis to show that minimization problems in extremal graph theory (such as the one considered above) are guaranteed to have solutions in the space of graphons. These graphon solutions provide "templates," via Theorem 1, for approximate solutions in the space of finite graphs.

References

1. C. Borgs, J. T. Chayes, L. Lovász, V. T. Sós, and K. Vesztergombi, Convergent sequences of dense graphs. I. Subgraph frequencies, metric properties and testing, *Adv. Math.* **219** (2008), no. 6, 1801–1851. MR 2455626 (2009m:05161)
2. L. Lovász, *Large Networks and Graph Limits*, American Mathematical Society Colloquium Publications, vol. 60, American Mathematical Society, Providence, RI, 2012. MR 3012035
3. L. Lovász and B. Szegedy, Limits of dense graph sequences, *J. Combin. Theory Ser. B* **96** (2006), no. 6, 933–957. MR 2274085 (2007m:05132)