THE GRAPHON AS A LIMIT FOR DENSE GRAPHS

VALERIE HAN

Abstract. This paper will introduce the graphon, which is the completion of the space of dense graphs. We will discuss homomorphism densities, an important property of graphs, and cut distance and sampling distance, two metrics used to compare graphs, in order to make sense of the graphon serving as a limit object for dense graphs.

Contents

1. Introduction

When working with large graphs, it is difficult to and, in fact, not particularly illuminating to know exactly which nodes are connected to one other. One issue we face, then, is how to determine the structure of such a large graph in an informative and not (unnecessarily) overwhelming way.

One approach is using graph homomorphisms, which can tell us a lot about the structure of a graph. However, finding homomorphisms becomes unwieldy as the size of the graph grows. To make this less unwieldy, in Section 3, we explore different kinds of homomorphisms, which are more useful or convenient in certain cases, as well as defining and relating homomorphism densities.

Another way to study large graphs is through sampling, examining small random induced subgraphs of the large graph. Using this approach, we can compare graphs using sampling distance, defined in Section 5. In that section, we will also define cut distance, a metric that is more directly informative structurally.

Another approach, the main focus of our paper, is to pass from sequences of larger and larger graphs to ideal limiting objects, which share important properties with their sequences. An advantage of this approach is having a more explicit representation of the limit object, which allows for greater analytic flexibility.

In fact, we shall see that these approaches are not separate at all. In Section 6, we will define convergence using graph homomorphism densities, and we will see

that the limit of homomorphism densities of a convergent sequence of dense graphs will be identical to the homomorphism densities of its limit object, the graphon. Moreover, the cut distance between the graphs in a sequence approaches 0 if and only if the graphs themselves converge, and the sampling distance converges to 0 if and only if the cut distance converges to 0.

This paper will assume familiarity with the very basics of graph theory and probability. We will be closely following Lovász's book [1] on the subject.

2. Preliminaries

First, we define terms and notation that will be used throughout the paper.

Remark 2.1. In the paper, we will use the terms *nodes* and *vertices* interchangeably.

Notation 2.2. Let G be a graph. Define $\mathrm{v}(G)$ to be the number of vertices of G and $\mathrm{V}(G)$ to be the set of vertices of G. Similarly, define $\mathrm{e}(G)$ to be the number of edges of G and $\mathrm{E}(G)$ to be the set of edges of G.

The following definitions of dense and sparse, terms which are not well-defined in graph theory, are not universally used but are convenient for our purposes.

Definition 2.3. Let G be a graph with n nodes. G is *dense* if the number of nodes that each node is adjacent to is linearly related to n. A graph is *sparse* if each node has a bounded number of neighbors.

Note that these may not seem mutually exclusive. However, they are mutually exclusive when considering a sequence of graphs where the number of vertices is increasing, important for investigating graph limits. While the fraction of edges of a sparse graph (normalized with respect to the number of nodes) will tend towards 0, the fraction of edges of a dense graph will tend towards some positive number. This will become important in the context of graph homomorphisms in Section 3.

Although sparse graphs are more important from a practical point of view, this paper will focus on dense graphs because they have a more natural limit object definition. We will explain why the graphon and the construction of homomorphism numbers in this paper unfortunately do not work for sparse graphs in Sections 6 and 3, respectively.

Next, we define the two types of graphs we will work with.

Definition 2.4. A *simple graph* is an undirected graph without weights, multiple edges, or loops.

Definition 2.5. A *weighted graph* is an undirected graph that has a positive real number, known as a weight, assigned to each edge and/or each node.

These two types of graphs will be sufficient for our purposes. We will mainly use simple graphs, but we will use weighted graphs to show how the same ideas illustrated with simple graphs can easily be extended. (For those readers familiar with other types of graphs, note that the weighted graph can easily represent other types of graphs such as multigraphs.)

Finally, we define some terms useful for examples and explanations.

Definition 2.6. A simple graph is a *complete graph* if every node is connected to every other node. We denote the complete graph with n vertices K_n.

Definition 2.7. Consider a graph G on the vertices $1, \ldots, n$. G is a *cycle* if each node is connected only to the ones before and after it. We denote the cycle with n vertices C_n.

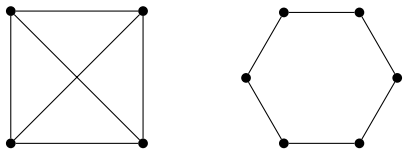


FIGURE 1. K_4 and C_6

Definition 2.8. A graph F is a *subgraph* of graph G if there exists a map $\varphi : \mathrm{V}(F) \to \mathrm{V}(G)$ such that φ preserves adjacency.

Definition 2.9. A graph F is an *induced subgraph* of graph G if there exists a map $\varphi : \mathrm{V}(F) \to \mathrm{V}(G)$ such that φ preserves adjacency and non-adjacency.

Note that F being an induced subgraph of G means there exists an exact copy of F in G, while F being a subgraph of G means there exists a copy but one that may have extra edges.

3. Graph Homomorphisms

It can be difficult to describe and characterize graphs, especially as the graphs become larger. One way we can characterize graphs is by how they're related to other graphs. Thus, we consider graph homomorphisms.

Definition 3.1. Let F, G be graphs. A map from $\mathrm{V}(F)$ to $\mathrm{V}(G)$ is a *graph homomorphism*, denoted by $F \to G$, if it preserves adjacency.

Note, however, that a homomorphism does not necessarily preserve non-adjacency.

Example 3.2. Consider the graphs in Figure 2. There exists the obvious homomorphism f which takes $F = C_4$ directly to $G = C_4$; $f(1) = A$, $f(2) = B$, $f(3) = C$, and $f(4) = D$. However, there also exists the less obvious homomorphism g that "squishes" two opposite corners together since they are not connected by an edge; $g(1) = g(3) = A$, $g(2) = B$, and $g(4) = D$.

Homomorphisms convey more information than may be immediately evident.

Example 3.3. A homomorphism from the complete graph K_n to G exists if and only if G contains a clique with n nodes. (A clique is a subset of nodes such that every pair of distinct nodes is connected. The term originates from a clique of people—people who all know each other.) The reverse homomorphism is also interesting; if there exists a homomorphism $G \to K_n$, the nodes in G can be split into n groups such that no node in each group is adjacent to another node in the same group. Thus, the existence of a homomorphism $G \to K_n$ means that G is n-colorable.

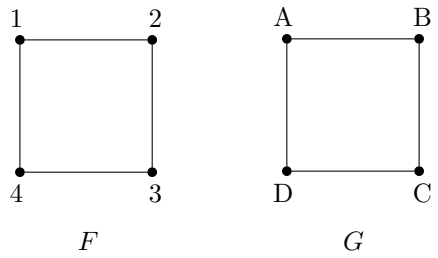


FIGURE 2. Two labeled C_4 graphs F and G

In the context of graph limits, we look at homomorphisms *into* a given graph G rather than from a given graph G. This can be understood in terms of sampling from the given graph. Testing for homomorphism(s) from a smaller graph into a large graph G can give us information on what is "present" in G. This will be made even clearer when we discuss injective and induced homomorphisms.

However, as these graphs become larger, it becomes difficult to determine or even record all the different homomorphisms from say K_3 to G. Instead, then, we consider homomorphism numbers and densities, allowing us to condense and quantify the information.

Definition 3.4. Define the *homomorphism number* $\hom(F, G)$ to be the number of homomorphisms from F to G.

Notation 3.5. Throughout the rest of this section, we will use $n = \mathrm{v}(G)$ and $k = \mathrm{v}(F)$ to make equations simpler.

In order to compare homomorphism numbers of two different graphs, we normalize them to get the probability that a random map from $V(F)$ to $V(G)$ is a homomorphism. The homomorphism number, which is the number of homomorphisms from F to G, is divided by the number of maps from F to G. Since maps are not constrained by preserving adjacency, any node in F can be mapped to any node in G, giving us n^k total maps from F to G.

Definition 3.6. Define the *homomorphism density*

$$t(F, G) = \frac{\hom(F, G)}{n^k}. \tag{3.7}$$

Note that it is possible for this value to be 1. If F were an empty graph, which gives the most homomorphisms, each vertex of F could be mapped to any vertex of G to create a homomorphism, leaving us with n^k total homomorphisms.

Although this definition may now seem natural and indeed works well for dense graphs, note that homomorphism density is essentially meaningless for sparse graphs as the number of vertices increases because the homomorphism density will always approach 0. Consider, for example, the homomorphism density of K_2 into a large sparse graph G with n vertices. Suppose the bound on the number of neighbors of each node is b. Then, the density $t(K_2, G) = \frac{\hom(K_2,G)}{n^2} \leq \frac{bn}{n^2} = \frac{b}{n}$, which will converge to 0 as $n \to \infty$.

We now extend these concepts to injective and induced homomorphisms.

Injective homomorphisms are exactly what they sound like, but the definition of induced homomorphisms may not be immediately apparent.

Definition 3.8. An *induced homomorphism* is an injective homomorphism that also preserves non-adjacency.

The induced homomorphism from F to G is so named because it is an embedding of F into G as an induced subgraph.

Using different types of homomorphisms may be useful for three reasons. First, one is sometimes more convenient to use in a calculation than another, producing a cleaner formula when working probabilistically. Second, they are useful in different contexts as we will see in Example 3.9 when compared to Example 3.3. Third, as we will see in Section 6, they can provide different intuition than the usual homomorphism.

Example 3.9. As we saw in Example 3.3, a homomorphism from F to G does not imply that there exists a subgraph F of G. However, there exists an injective homomorphism from F to G if and only if there exists such a subgraph in G. We can thus use the injective homomorphism as a tool to detect subgraphs such as cycles.

Having an induced homomorphism from one graph to another is an even stronger condition. An induced homomorphism from F to G is equivalent to a copy of F existing in G, with the same adjacencies and non-adjacencies. We can use the induced homomorphism to detect how many cycles are present as exact copies in G. Note that while K_4 will have an injective homomorphism from C_4, it will not have an induced one.

Notation 3.10. Let $\text{inj}(F, G)$ be the number of injective homomorphisms from F to G, and let $\text{ind}(F, G)$ be the number of induced homomorphisms from F to G.

Now we define injective and induced homomorphism densities, which are similar to the usual homomorphism density.

Instead of describing the probability of a random map from F to G being a homomorphism, the injective homomorphism density corresponds to the probability of a random injection from $\text{V}(F)$ to $\text{V}(G)$ being a homomorphism, and the induced homomorphism density corresponds to the probability of a random injection from $\text{V}(F)$ to $\text{V}(G)$ preserving both adjacency and non-adjacency. Instead of dividing by the total number of maps, we divide by the number of injective maps. Because we cannot map two vertices of F to the same vertex of G, we obtain $n(n-1)\cdots(n-k+1)$ instead of n^k.

Definition 3.11. Define the *injective homomorphism density*

$$t_{\text{inj}}(F, G) = \frac{\text{inj}(F, G)}{n(n-1)\cdots(n-k+1)}. \tag{3.12}$$

Similarly, define the *induced homomorphism density*

$$t_{\text{ind}}(F, G) = \frac{\text{ind}(F, G)}{n(n-1)\cdots(n-k+1)}. \tag{3.13}$$

Through combinatorial manipulation, we can relate the different homomorphism numbers to each other.

Induced and injective homomorphism numbers are closely related to one another since induced homomorphisms are injective homomorphisms that preserve non-adjacency. In order to relate them, we sum all induced homomorphisms from graphs with the same number of nodes that contain F as a subgraph. Then, with F' ranging over all simple graphs obtained from F by adding edges,

$$\text{inj}(F, G) = \sum_{F' \supseteq F} \text{ind}(F', G). \tag{3.14}$$

Using inclusion-exclusion, we also have

$$\text{ind}(F, G) = \sum_{\substack{F' \supseteq F \\ \text{V}(F')=\text{V}(F)}} (-1)^{e(F')-e(F)} \text{inj}(F', G). \tag{3.15}$$

It is also possible to relate homomorphism numbers and injective homomorphism numbers.

To do this, we use the partition and quotient. For those readers who are unfamiliar, a partition P groups vertices together. The quotient graph F/P is formed by collapsing the vertices in each of the groups in the partition into a single vertex; two vertices in F/P are adjacent if and only if any of the original vertices in the two groups are connected. Any parts of the partition that have internal edges collapse to a single vertex with a loop.

With P ranging over all partitions of $\text{V}(F)$ and F/P as the quotient graph,

$$\text{hom}(F, G) = \sum_{P} \text{inj}(F/P, G). \tag{3.16}$$

Using (3.16), we can also express inj in terms of hom by considering the values $\text{inj}(F, G)$ unknowns and solving the system. Deriving the explicit formula is a more involved computation and is not relevant for our discussion; thus, we direct the reader to (5.17)-(5.18) in [1] if the reader is interested.

Now, we turn to relating homomorphism densities rather than numbers.

Converting the relationships in (3.14) and (3.15) to relationships between injective and induced homomorphism numbers is easy. We simply divide by $n(n-1)\cdots(n-k+1)$ to obtain

$$t_{\text{inj}}(F, G) = \sum_{F' \supseteq F} t_{\text{ind}}(F', G), \text{ and} \tag{3.17}$$

$$t_{\text{ind}}(F, G) = \sum_{\substack{F' \supseteq F \\ \text{V}(F')=\text{V}(F)}} (-1)^{e(F')-e(F)} t_{\text{inj}}(F', G). \tag{3.18}$$

However, because t and t_{inj} have different normalizations, we can only obtain an inequality. Fortunately, this is sufficient for our purposes.

Recall that t_{inj} is the probability that an injective map is a homomorphism while t is the probability that a random map is a homomorphism.

Let φ be a random map from $\mathrm{V}(F)$ to $\mathrm{V}(G)$. If $t(F,G) \geq t_{\mathrm{inj}}(F,G)$,

$$\begin{aligned}
|t(F,G) - t_{\mathrm{inj}}(F,G)| &= t(F,G) - t_{\mathrm{inj}}(F,G) \\
&= \frac{\hom(F,G)}{n^k} - \frac{\mathrm{inj}(F,G)}{n(n-1)\cdots(n-k+1)} \\
&\leq \frac{\hom(F,G)}{n^k} - \frac{\mathrm{inj}(F,G)}{n^k} \\
&= \mathrm{P}(\varphi \text{ is a homomorphism) - P}(\varphi \text{ is injective}) \\
&= \mathrm{P}(\varphi \text{ is a non-injective homomorphism}) \\
&\leq \mathrm{P}(\varphi \text{ is non-injective}) \\
|t(F,G) - t_{\mathrm{inj}}(F,G)| &\leq \frac{1}{n}\binom{k}{2}, \qquad (3.19)
\end{aligned}$$

where $\frac{1}{n}$ is the probability of two given vertices of F being sent to the same vertex on G while $\binom{k}{2}$ is the number of ways to choose two vertices of F.

Meanwhile, if $t_{\mathrm{inj}}(F,G) > t(F,G)$,

$$\begin{aligned}
|t(F,G) - t_{\mathrm{inj}}(F,G)| &= t_{\mathrm{inj}}(F,G) - t(F,G) \\
&= \frac{\mathrm{inj}(F,G)}{n(n-1)\cdots(n-k+1)} - \frac{\hom(F,G)}{n^k} \\
&\leq \frac{\hom(F,G)}{n(n-1)\cdots(n-k+1)} - \frac{\hom(F,G)}{n^k} \\
&= \hom(F,G)\left(\frac{1}{n(n-1)\cdots(n-k+1)} - \frac{1}{n^k}\right) \\
&\leq n^k\left(\frac{1}{n(n-1)\cdots(n-k+1)} - \frac{1}{n^k}\right) \\
|t(F,G) - t_{\mathrm{inj}}(F,G)| &\leq \frac{n^k}{n(n-1)\cdots(n-k+1)} - 1, \qquad (3.20)
\end{aligned}$$

Note that these inequalities are sufficient for our purposes because we are concerned with graphs G with a large number of vertices, and the difference in the homomorphism densities approaches 0 as $\mathrm{v}(G) \to \infty$.

Finally, we extend homomorphism numbers to weighted graphs. Counting weighted homomorphisms is more difficult because we must "weight" the homomorphisms.

Definition 3.21. Let F be a simple graph, and let G be a weighted graph. Call the nodeweights $\alpha_v(G)$ and edgeweights $\beta_{uv}(G)$. For each map, $\varphi : \mathrm{V}(F) \to \mathrm{V}(G)$, define their weights

$$\alpha_\varphi = \prod_{u \in \mathrm{V}(F)} \alpha_{\varphi(u)}(G), \qquad (3.22)$$

and

$$\hom_\varphi(F,G) = \prod_{uv \in \mathrm{E}(F)} \beta_{\varphi(u)\varphi(v)}(G). \qquad (3.23)$$

Define the *weighted homomorphism number*

$$\hom(F,G) = \sum_{\varphi:\mathrm{V}(F)\to\mathrm{V}(G)} \alpha_\varphi \hom_\varphi(F,G), \tag{3.24}$$

and

$$\mathrm{inj}(F,G) = \sum_{\substack{\varphi:\mathrm{V}(F)\to\mathrm{V}(G)\\ \varphi \text{ injective}}} \alpha_\varphi \hom_\varphi(F,G), \tag{3.25}$$

Note that α_φ in (3.22) weights the nodes so that maps φ sending vertices of F to vertices of G with higher nodeweights are weighted more. If there does not exist a corresponding edge of $\mathrm{E}(F)$ in $\mathrm{E}(G)$, where we treat non-edges as edges with weight 0, $\hom_\varphi(F,G)$ in (3.23) is 0, so this map will not contribute to the sum, ensuring that only adjacency-preserving maps are counted.

Note also that this definition is truly an extension of the simple graph version. If we let the nodeweights and edgeweights be 1, (3.24) and (3.25) reduce to the simple graph definitions.

4. Graphons

The relationship between the graphon and the graph may not be immediately evident, but the reader may be reassured that this will be made clearer after the definition.

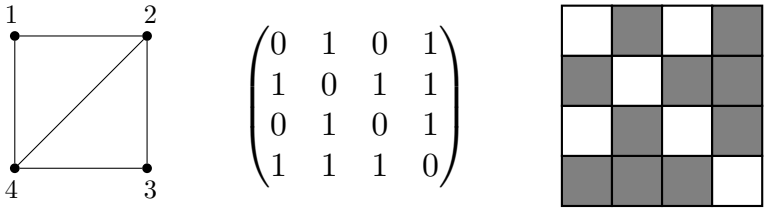
Definition 4.1. A *graphon* is a bounded symmetric measurable function from $[0,1]\times[0,1]$ to $\mathbb{R}$.

To show its relation to a graph, we explain how we can construct a graphon from a graph. The construction of the graphon from a graph is similar to the construction of a graph's adjacency matrix.

Let G be a graph with n vertices. Label the vertices $1,\dots,n$.

In the case of the simple graph, partition the interval $[0,1]$ into n intervals of length $1/n$. For each pair of vertices j and k , consider $x \in \left[\frac{j-1}{n}, \frac{j}{n}\right]$ and $y \in \left[\frac{k-1}{n}, \frac{k}{n}\right]$. We denote the graphon W_G as the graphon formed from the graph G, defined by $W(x,y) = 1$ if vertices j and k are adjacent and 0 if not.

The process can easily be extended to weighted graphs. Replace 1 by the edge weight, and partition the interval $[0,1]$ proportionally to the node weights.



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

Figure 3. A graph, its adjacency matrix, and its corresponding graphon. The white areas of the grid represent 0 on the graphon while the shaded ones represent 1. Note that the origin is in the upper lefthand corner.

Now that the relation between the graph and the graphon is clearer, the reader may wonder if this is all there is to it. Can we simply use the usual real analysis definition of convergence using a norm like the L_1 norm? Unfortunately, the answer is no.

The reason for this is that graphons can be "close" without being close in the L_1 distance. For example, recall our construction of a graphon from a graph. The labeling of vertices is arbitrary. The same graph can be represented by two graphons that look very different. We would like to be able to view these two graphons as the "same" in some sense. This is especially important when we consider the important property we defined earlier and that is the backbone of our notion of convergence—homomorphism densities—which do not change when confronted by relabeling of vertices.

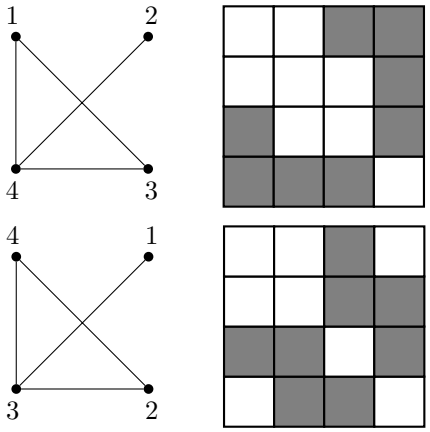


FIGURE 4. A graph (labeled two different ways) and two different graphon representations due to relabeling

Definition 4.2. Graphons U, W are *weakly isomorphic* if there exist measure preserving functions $\varphi, \psi : [0,1] \to [0,1]$ such that $U(\varphi(x), \varphi(y)) = W(\psi(x), \psi(y))$ almost everywhere. Denote $U(\varphi(x), \varphi(y))$ by U^{φ}.

Example 4.3. Consider a graph G. Label its vertices two different ways, and call the resulting labeled graphs G_1 and G_2. The graphons created from G_1 and G_2 are weakly isomorphic because relabeling only moves around the subintervals and does not resize them. For example, the two graphons in Figure 4 are weakly isomorphic.

While defining weak isomorphisms is motivated by the issue of relabeling, it is more versatile than that one use.

Example 4.4. Define $\varphi_k : x \mapsto kx \pmod 1$. The map φ_2 is a measure preserving map that make four smaller "copies" of a given graphon. Given graphon W, the graphons W and W^{φ_2} are weakly isomorphic.

There exists an equivalent definition of weak isomorphism, defining two graphons U, W to be weakly isomorphic if $t(F, U) = t(F, W)$ for every simple graph F.

Proposition 4.5. *Graphons U and W are weakly isomorphic if and only if $t(F, U) = t(F, W)$ for every simple graph F.*

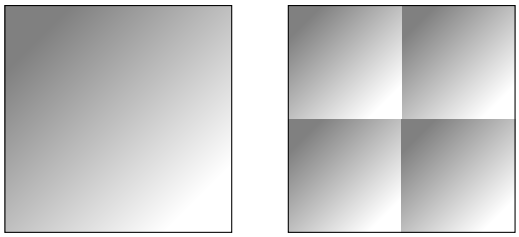


FIGURE 5. A graphon W and W^{φ_2}

For the proof, see Corollary 10.35(a) in [1].

We now connect graphons to our definition of convergence by defining homomorphism densities for graphons.

Call the vertices of F $1, \dots, \mathrm{v}(F)$. Choose $x_i \in [0,1]$ for each vertex i. For each edge ij of F, take the product of $W(x_i, x_j)$ to represent the weight of the homomorphism. By integrating over all choices of x_i, we obtain the homomorphism density. This process parallels the process of sampling $\mathrm{v}(F)$ vertices of a graph and summing the resulting homomorphism weights to obtain the weighted homomorphism density for the graph. In fact, $t(F, G) = t(F, W_G)$ for all weighted graphs G.

Definition 4.6. Let F be a graph, and let W be a graphon. Define the *homomorphism density*

$$t(F, W) = \int_{[0,1]^{\mathrm{v}(F)}} \prod_{ij \in \mathrm{E}(F)} W(x_i, x_j) \prod_{i \in \mathrm{V}(F)} dx_i. \tag{4.7}$$

Note that in this context, the injective homomorphism density is insignificant because when randomly assigning vertices i and j to x_i and x_j in the interval $[0,1]$, $x_i \neq x_j$ with probability 1. Thus, the injective homomorphism density and homomorphism density are essentially equivalent in this context.

However, it is useful to define the induced subgraph density.

Definition 4.8. Let F be a graph, and let W be a graphon. Denote $\binom{V}{2} \backslash \mathrm{E}(F)$ to be the set of all pairs of vertices with no edge between them. Define the *induced homomorphism density*

$$t_{\mathrm{ind}}(F, W) = \int_{[0,1]^{\mathrm{v}(F)}} \prod_{ij \in \mathrm{E}(F)} W(x_i, x_j) \prod_{ij \in \binom{V}{2} \backslash \mathrm{E}(F)} (1 - W(x_i, x_j)) \prod_{i \in \mathrm{V}(F)} dx_i. \tag{4.9}$$

These graphon versions of homomorphism numbers are very similar to their graph versions in that they can be related in similar ways.

Similarly to (3.18), we have

$$t_{\mathrm{ind}}(F, W) = \sum_{\substack{F' \supseteq F \\ \mathrm{V}(F') = \mathrm{V}(F)}} (-1)^{e(F') - e(F)} t(F', W) \tag{4.10}$$

by expanding the product in parentheses.

In addition, similarly to (3.17), we have

$$t_{\mathrm{inj}}(F, W) = \sum_{F' \supseteq F} t_{\mathrm{ind}}(F', W). \tag{4.11}$$

5. Cut Distance and Sampling Distance

Weakly isomorphic graphons are equivalent in every way that matters for us. Thus, we want to define a metric that will make the distance between two weakly isomorphic graphons 0. We thus turn to two different distances that both work for convergence, as we will see in Section 6. First, though, to provide motivation for the other two more involved distances, we define an intuitive distance and explain why it cannot be used.

Especially when considering the distance between graphs, one distance that may seem intuitive is the edit distance, the number of edges that need to be toggled to move from one graph to another. We may wish to normalize this because our graphs are very large. Thus, we define edit distance for our purposes as follows.

Definition 5.1. Let G, G' be dense graphs with the same set of nodes. The *edit distance* d_1 is defined by

$$d_1(G, G') = \frac{|\mathrm{E}(G) \triangle \mathrm{E}(G')|}{n^2}, \tag{5.2}$$

where $\triangle$ denotes the symmetric difference.

However, there are two main problems with using this distance.

First, as discussed in the Introduction, structural similarity is important when comparing large graphs, but the edit distance does not measure structural similarity well. For example, consider two random graphs on $\{1, \dots, n\}$ with edge density $1/2$, which are structurally very similar, including approaching the same homomorphism densities as $n \to \infty$ with probability 1. (For more structural similarities, see QR1-QR5 in [1].) Yet, the graphs' edit distance will be large; it will be close to $1/2$ with high probability.

The second problem is that edit distance is only defined for graphs with the same number of nodes. When analyzing large graphs, we generally do not care overmuch about exactly how many nodes the two graphs have. In fact, we may not even know exactly how many nodes a graph has. In a sense, we wish to determine whether the object is cheese and not be bogged down by how large a piece we are given.

We thus arrive at the idea of sampling a graph. Instead of looking at the graph as a whole, we can consider whether small samples of the two graphs are similar.

First, we define a preliminary distance to compare two probability distributions.

Definition 5.3. Let α, β be two probability distributions on the same set X. Define the *variation distance* d_{var} between α and β by

$$d_{\mathrm{var}}(\alpha, \beta) = \sup_{X'} |\alpha(X') - \beta(X')|, \tag{5.4}$$

where X' varies over all measurable subsets of X.

Now, we define probability distributions obtained by sampling to use variation distance to compare graphs.

Definition 5.5. Define $\sigma_{G,k}$ to be the probability distribution on graphs on $\{1, \dots, k\}$ obtained by selecting k random nodes of G, which will generally have more than k nodes, and taking the corresponding induced subgraph. If $k > \mathrm{v}(G)$, we return the empty k-node graph.

In order to measure the distance between two graphs with a single number, rather than a different variation distance for every k, we combine the variation distances using a sum. To ensure convergence, we divide by 2^k.

Definition 5.6. Let G, G' be dense graphs. Define the *sampling distance* δ_{samp} between G and G' by

$$\delta_{\text{samp}}(G, G') = \sum_{k=1}^{\infty} \frac{1}{2^k} d_{\text{var}}(\sigma_{G,k}, \sigma_{G',k}). \tag{5.7}$$

The $1/2^k$ is useful for another reason as well. It ensures that the sum is most influenced by "small" induced subgraphs even though sampling distance includes larger ones in the definition. However, the flip side of this quality is that sampling distance does not directly reflect global structural similarity, although we will see in Section 6 that it reflects structural similarity well enough, albeit indirectly.

This definition can be extended to graphons simply by defining a distribution to use in the graphon version.

Definition 5.8. Let W be a graphon such that $0 \leq W \leq 1$. Given $S = \{x_1, \ldots, x_n\}$ with $x_i \in [0, 1]$, define $\mathbb{G}_{S,W}$ to be the probability distribution on graphs obtained from W in the following way.

Form weighted graph $\mathbb{H}(S, W)$ by assigning weight $W(x_i, x_j)$ to edge ij, giving weight 0 to loops.

Given such a weighted graph H, form the random simple graph $\mathbb{G}(H)$ on $\mathrm{V}(H)$ by connecting nodes i and j with probability $\beta_{ij}(H)$, deciding independently for distinct pairs of nodes.

Notation 5.9. For easier notation in defining sampling distance for graphons, given $S = \{x_1, \ldots, x_n\}$, we denote $\mathbb{G}_{S,W}$ by $\mathbb{G}_{n,W}$.

The graphon version of sampling distance then immediately follows.

Definition 5.10. Let U, W be graphons. Define the *sampling distance* δ_{samp} between U and W by

$$\delta_{\text{samp}}(U, W) = \sum_{k=1}^{\infty} \frac{1}{2^k} d_{\text{var}}(\mathbb{G}_{k,U}, \mathbb{G}_{k,W}). \tag{5.11}$$

The sampling distance can also be written in terms of induced homomorphism numbers. This will be useful in Section 6 when we relate sampling distance and convergence of graphs.

Note that $\mathbb{G}_{k,U}(F) = t_{\text{ind}}(F, U)$ for any graphon U and simple k-node graph F.

$$\begin{aligned} d_{\text{var}}(\mathbb{G}_{k,U}, \mathbb{G}_{k,W}) &= \sup_X |\mathbb{G}_{k,U}(X) - \mathbb{G}_{k,W}(X)| \\ &= \sup_X \left| \sum_{F \in X} (\mathbb{G}_{k,U}(F) - \mathbb{G}_{k,W}(F)) \right| \\ &= \sup_X \left| \sum_{F \in X} (t_{\text{ind}}(F, U) - t_{\text{ind}}(F, W)) \right|. \end{aligned} \tag{5.12}$$

The quantity inside the supremum is maximized when $\mathbb{G}_{k,U}(F) > \mathbb{G}_{k,W}(F)$ for all graphs $F \in X$ or when $\mathbb{G}_{k,U}(F) < \mathbb{G}_{k,W}(F)$ for all graphs $F \in X$. The difference between $\mathbb{G}_{k,U}(F)$ and $\mathbb{G}_{k,W}(F)$ contributes positively towards the sum

$\sum_F |\mathbb{G}_{k,U}(F) - \mathbb{G}_{k,W}(F)|$ both when $\mathbb{G}_{k,U}(F)$ is larger and when $\mathbb{G}_{k,W}(F)$ is larger. The total differences in both cases must be equal for the overall probabilities to sum to 1. Thus, we divide by 2 to obtain the quantity in (5.12), giving us

$$d_{\text{var}}(\mathbb{G}_{k,U}, \mathbb{G}_{k,W}) = \frac{1}{2}\sum_F |t_{\text{ind}}(F,U) - t_{\text{ind}}(F,W)|, \tag{5.13}$$

where F varies over all simple graphs with k vertices.

Then, we have

$$\delta_{\text{samp}}(U,W) = \sum_F 2^{-\text{v}(F)-1}|t_{\text{ind}}(F,U) - t_{\text{ind}}(F,W)|, \tag{5.14}$$

where F varies over all finite graphs.

Now, we move to our discussion of cut distance. Unlike sampling distance, cut distance does directly reflect structural similarity, which makes it more useful to use in some cases.

Although cut distance can be defined on graphs, and, indeed, gives some intuition that is not as obvious in the graphon version, we give only the graphon definition because the graph version is quite involved. Moreover, we can use the graphon version for graphs as well, by transforming any graph into a graphon. The interested reader is encouraged to read Sections 8.1 and 8.2 in [1] for more information on the graph version.

Definition 5.15. Let U, W be graphons. Define cut distance $\delta_\square$ between U and W by

$$\delta_\square(U,W) = \inf_{\varphi,\psi}\sup_{S,T}\left|\int_{S\times T}(U(\varphi(x),\varphi(y)) - W(\psi(x),\psi(y)))dxdy\right|, \tag{5.16}$$

where φ, ψ range over all measure preserving maps on $[0,1]$ and S, T range over all measurable subsets of $[0,1]$.

The cut distance takes measurable subsets S, T of $[0,1]$, forming "boxes" $S \times T$ (hence the $\square$), which are similar to sets of vertices, then maximizes the difference between the two graphons integrating over the box $S \times T$, in a sense, counting the difference in the number of edges between the sets of vertices. Then, we take the infimum over all measure preserving maps to ensure that weakly isomorphic graphons have cut distance 0.

In Section 4, we have seen that a graph can easily be converted into a graphon. Thus, we define the cut distance for graphs by simply converting the graph into a graphon and using the graphon cut distance.

Definition 5.17. Let F, G be graphs. Define cut distance $\delta_\square$ between F and G by

$$\delta_\square(F,G) = \delta_\square(W_F, W_G). \tag{5.18}$$

6. Convergence of Dense Graphs

We have seen throughout the paper that graphs do not have to be identical or even a few edits away to be close structurally. Instead, we have looked at sampling and homomorphism densities. We thus define the convergence of a sequence of dense graphs (G_n) with $v(G_n) \to \infty$ by combining these approaches. We use induced homomorphism densities, in a sense "sampling" induced subgraphs.

Definition 6.1. A sequence (G_n) with $\mathrm{v}(G_n) \to \infty$ *converges* if its induced subgraph densities $t_{\mathrm{ind}}(F, G_n)$ converge for all finite graphs F.

It may sometimes be more convenient to use $t(F, G_n)$ or $t_{\mathrm{inj}}(F, G_n)$ instead. Because of the formulas relating the different densities, we have the following theorem.

Theorem 6.2. *Let (G_n) be a sequence of graphs with $v(G_n) \to \infty$. The following are equivalent:*

(i) $t_{ind}(F, G_n)$ converges for all finite graphs F.
(ii) $t_{inj}(F, G_n)$ converges for all finite graphs F.
(iii) $t(F, G_n)$ converges for all finite graphs F.

Proof. We first show that (ii) is equivalent to (i). By (3.17) and (3.18), we can express the induced subgraph densities as a linear combination of the subgraph densities and vice versa. Thus, $t_{\mathrm{ind}}(F, G_n)$ converges for all finite graphs F if and only if $t_{\mathrm{inj}}(F, G_n)$ does.

Now we show that (iii) is equivalent to (ii). By the inequalities relating $t_{\mathrm{inj}}(F, G_n)$ and $t(F, G_n)$ in (3.19) and (3.20), we have that as $n \to \infty$, $t(F, G_n)$ converges if and only if $t_{\mathrm{inj}}(F, G_n)$ converges. □

The following theorems follow a long chain of results and are difficult to prove. For these reasons, we omit the proofs and direct the interested reader to the relevant sections in the textbook.

We have defined convergence of a graph sequence by the convergence of the sequence's homomorphism densities to some number. In the following theorem, we give the significance of such numbers in terms of our limit object, the graphon.

Theorem 6.3. *Let (G_n) be a sequence of simple graphs. If (G_n) converges, there exists a graphon W such that $t(F, G_n) \to t(F, W)$ for all simple graphs F.*

For the proof, see either Theorem 11.21 in [1] or the original proof by Lovász and Szegedy in [5].

Given that the limits of each of a graph sequence's homomorphism densities, which form the basis of the definition of convergence, coincide with a graphon's homomorphism densities, we now define the graphon as the limit object.

Definition 6.4. A graphon W is the *limit* of a convergent sequence (G_n) of simple graphs if $t(F, G_n) \to t(F, W)$ for all simple graphs F.

However, we note here that the limit is not unique. Indeed, any weakly isomorphic graphons will result in the same homomorphism densities. Fortunately, this is the only weakness in the uniqueness; the limit is unique up to weak isomorphism, following from Proposition 4.5. This is sufficient for our purposes because graphons which are weakly isomorphic have the same probability distributions on graphs obtained from the graphons, unsurprising given that the graphon homomorphism densities are identical.

Although we have four different ways to prove convergence, the first three are not truly different ways of proving convergence so much as they are applying the same method to slightly different quantities, and the fourth requires an explicit

limit, which is often hard to determine. The below theorem allows us to instead use the cut distance to prove convergence.

Theorem 6.5. *Let (G_n) be a sequence of simple graphs with $v(G_n) \to \infty$. Then, G_n converges if and only if it is a Cauchy sequence in the metric $\delta_\square$.*

For a proof of this, see either Theorem 11.3 in [1] or the original proof by Borgs, Chayes, Lovász, Sós and Vesztergombi in [3] and [4].

We have a similar theorem for a sequence of graphons. The graphon thus is truly the completion of the space of graphs; a convergent sequence of graphons has a graphon as its limit and does not require the construction of a different limit object.

Theorem 6.6. *Let (W_n) be a sequence of graphons such that every W_n takes values only in $[0, 1]$, and likewise, let W be a graphon that takes values only in $[0, 1]$. Then $t(F, W_n)$ converges for all finite simple graphs F if and only if W_n is a Cauchy sequence in the cut distance. Furthermore, $t(F, W_n) \to t(F, W)$ for all finite simple graphs F if and only if $\delta_\square(W_n, W) \to 0$.*

See Theorem 11.5 in [1] for the proof.

The following theorem provides yet another way to prove convergence: showing the cut distance between the limit graphon and the graphs in the sequence approaches 0 as $n \to \infty$.

Theorem 6.7. *A graphon W is the limit of graph sequence (G_n) if and only if $\delta_\square(W_{G_n}, W) \to 0$.*

See Theorem 11.22 in [1] for the proof.

In fact, there is a similar result for sampling distance. Note that one direction is easy to prove. If a sequence of graphs converge, the induced homomorphism densities converge, forming a Cauchy sequence. Using (5.14), then, the sampling distance approaches 0 as $n \to \infty$. The other direction also holds but is less straightforward to prove because it does not follow from (5.14) that the sampling distance being 0 implies that the difference in induced homomorphism densities converges to 0.

As mentioned in the Introduction, the graphon only serves as a limit object for dense graphs. The graphon is not a meaningful limit object for sparse graphs. Aside from the definition of convergence relying on homomorphism densities, which we have already explained are not significant for sparse graphs, the graphon itself is not helpful for gleaning information about a sequence of sparse graphs.

Indeed, consider a sequence of sparse graphs (G_n) as $\mathrm{v}(G_n) \to \infty$. Then, the number of neighbors of each node is bounded above by some fixed b. As $\mathrm{v}(G_n) \to \infty$, the graphon constructed from such graphs will become the graphon $W(x, y) = 0$ for any sparse graph sequence, making it useless as a limit object for such a sequence.

7. Examples of Convergence to Graphons

Finding the graphon limit to a sequence of graphs can be difficult. However, here are some intuitive examples of sequences of graphs that converge to a graphon.

Example 7.1 (Random p graphs)**.** Let $(G(n,p))$ be a sequence of graphs formed by forming an edge between any two nodes with probability p. Because edges occur with probability p in $(G(n,p))$, we may guess the limit graphon to be $W(x,y)=p$.

Note that this is one of the few examples where it is easy to calculate the homomorphism density. Consider any graph F. Then, the graphon homomorphism density

$$\begin{aligned} t(F,W) &= \int_{[0,1]^{\mathrm{v}(F)}} \prod_{ij\in \mathrm{E}(F)} p \prod_{i\in \mathrm{V}(F)} dx_i \text{ by Definition 4.6} \\ &= \int_{[0,1]^{\mathrm{v}(F)}} p^{\mathrm{e}(F)} \prod_{i\in \mathrm{V}(F)} dx_i \\ &= p^{\mathrm{e}(F)} \end{aligned}$$

With probability 1, the limit of the graph homomorphism densities $t(F,G(n,p))$ will also be $p^{\mathrm{e}(F)}$ because the probability of a map being a homomorphism as $n\to\infty$ is determined by the probability that an edge in F is also an edge in G_n. Since edges occur with probability p, we have that the density is $p^{\mathrm{e}(F)}$.

Example 7.2 (Complete bipartite graphs)**.** A *complete bipartite graph* $K_{n,n}$ is a bipartite graph (a graph where $\mathrm{V}(K_{n,n})$ is decomposed into two disjoint sets of size n such that no pair of vertices within the same set are adjacent) such that every pair of vertices in different sets are adjacent. Let $(K_{n,n})$ be a sequence of complete bipartite graphs.

Consider the two disjoint node sets of size n. Label the nodes in one set $1,\dots,n$ and the nodes in the other set $n+1,\dots,2n$. Then, every node in the first set is adjacent to every node in the second, and no two nodes in the same set are adjacent. Then, every graphon representation $W_{K_{n,n}}$ is

$$\mathbb{1}(0\le x\le \frac{1}{2}\le y\le 1)+\mathbb{1}(0\le y\le \frac{1}{2}\le x\le 1),$$

where $\mathbb{1}$ denotes the indicator function. Then, $(K_{n,n})$ converges to the graphon

$$W(x,y)=\mathbb{1}(0\le x\le \frac{1}{2}\le y\le 1)+\mathbb{1}(0\le y\le \frac{1}{2}\le x\le 1).$$

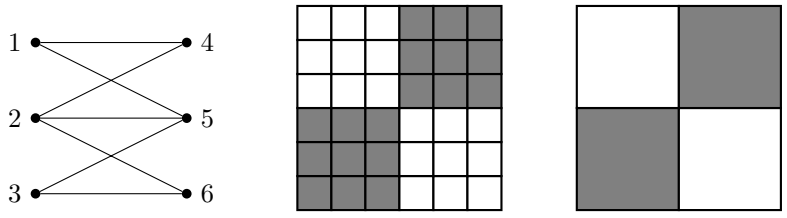


FIGURE 6. Graph $K_{6,6}$, its graphon $W_{K_{6,6}}$, and the limiting graphon, which is the same graphon as $W_{K_{6,6}}$.

Example 7.3 (Simple threshold graphs)**.** A *simple threshold graph* is defined on the set $\{1,\dots,n\}$ by connecting i and j if and only if $i+j\le n$. Consider a sequence of simple threshold graphs (G_n). Note that when we convert G_n to a graphon, we

obtain $W_{G_n} = \mathbb{1}(x + y \leq 1 + \frac{1}{n})$. Thus, as $n \to \infty$, (G_n) converges to the graphon $W(x, y) = \mathbb{1}(x + y \leq 1)$.

Acknowledgments

I'd like to thank my mentor Dylan Quintana for his patience and insight in answering my questions and reviewing my paper as well as for his help in finding me a topic that interested me. I would also like to thank Professor Peter May for organizing the program.

References

[1] László Lovász. Large networks and graph limits. American Mathematical Society. 2012.
[2] Daniel Glasscock. WHAT IS ... a Graphon? https://arxiv.org/pdf/1611.00718.pdf
[3] Christian Borgs, Jennifer Chayes, Lászl Lovász, Vera Sós, Balázs Szegedy and Katalin Vesztergombi. Graph Limits and Parameter Testing. STOC38. 2006. 261-270.
[4] Christian Borgs, Jennifer Chayes, Lászl Lovász, Vera Sós, Balázs Szegedy and Katalin Vesztergombi. Convergent Graph Sequences I: Subgraph frequencies, metric properties, and testing. Advances in Math. 2008. 1801-1851.
[5] Lászl Lovásza and Balázs Szegedy. Limits of dense graph sequences. Theory B 96. 2006. 933-957.