Graph Limit theory: Graphons, Cut Distance, and the Regularity Lemma

Youyou, Wu

April 6, 2026

Outline

Section 1

Motivation

Graph limit theory is the study of very large graphs

- People are interested in studying the behavior of **very large graphs.** Questions like:
- What is the average degree of the graph?

Graph limit theory is the study of very large graphs

- People are interested in studying the behavior of **very large graphs.** Questions like:
- What is the average degree of the graph?
- How well is the graph connected? (well it's almost certainly not connected, but we care about large components)

Graph limit theory is the study of very large graphs

- People are interested in studying the behavior of **very large graphs.** Questions like:
- What is the average degree of the graph?
- How well is the graph connected? (well it's almost certainly not connected, but we care about large components)
- What is the largest cut in the graph? (even we have the answer, in what form do we expect the answer? you can't just return a list of vertices because of time/space complexity too high!)

Graph limit theory is the study of very large graphs

- People are interested in studying the behavior of **very large graphs.** Questions like:
- What is the average degree of the graph?
- How well is the graph connected? (well it's almost certainly not connected, but we care about large components)
- What is the largest cut in the graph? (even we have the answer, in what form do we expect the answer? you can't just return a list of vertices because of time/space complexity too high!)
- And more. (Network design, graph learning, etc.)

But how do we study them

- How to obtain information from them (look at local structures: **graph homomorphism**, sampling methods)

But how do we study them

- How to obtain information from them (look at local structures: **graph homomorphism**, sampling methods)
- How to model them (random/quasirandom graph modeling)

But how do we study them

- How to obtain information from them (look at local structures: **graph homomorphism**, sampling methods)
- How to model them (random/quasirandom graph modeling)
- How to approximate them (global approximation: **Cut distance**, **Regularity Lemma**)

But how do we study them

- How to obtain information from them (look at local structures: **graph homomorphism**, sampling methods)
- How to model them (random/quasirandom graph modeling)
- How to approximate them (global approximation: **Cut distance**, **Regularity Lemma**)
- **Goal today:** We will focus on the first and third question today, and see how they are actually related. We will first look at a bunch of definitions to build intuitions (bear with me), and then we will hopefully look at one or two applications to see why graphon is useful.

Section 2

Preliminaries

Subsection 1

graph homomorphism

homomorphism number

Definition

- Let G and H be two simple graphs. An **adjacency-preserving map** from G to H is called a **homomorphism**.

homomorphism number

Definition

- Let G and H be two simple graphs. An **adjacency-preserving map** from G to H is called a **homomorphism**.
- We denote the set of homomorphisms as **Hom(F, G)**, and let **hom(F, G)** to be their number.

homomorphism number

Definition

- Let G and H be two simple graphs. An **adjacency-preserving map** from G to H is called a **homomorphism**.
- We denote the set of homomorphisms as **Hom(F, G)**, and let **hom(F, G)** to be their number.
- We define **inj(F, G)** to be the number of **injective homomorphism**, and **ind(F, G)** to be the number of **induced homomorphism**. Induced homorphism preserves both adjacency and non-adjacency.

homomorphism number

Definition

- Let G and H be two simple graphs. An **adjacency-preserving map** from G to H is called a **homomorphism**.
- We denote the set of homomorphisms as **Hom(F, G)**, and let **hom(F, G)** to be their number.
- We define **inj(F, G)** to be the number of **injective homomorphism**, and **ind(F, G)** to be the number of **induced homomorphism**. Induced homorphism preserves both adjacency and non-adjacency.
- For **multigraphs**, we usually require the specification of the mapping on the edges in addition to the vertices, so it's really a composition of two mappings.

Homomorphism density

Definition (homomorphism density)

For a fixed finite graph F and a large graph G, define

$$t(F, G) := \frac{\hom(F, G)}{|V(G)|^{|V(F)|}}.$$

Homomorphism density

Definition (homomorphism density)

For a fixed finite graph F and a large graph G, define

$$t(F,G) := \frac{\hom(F,G)}{|V(G)|^{|V(F)|}}.$$

- $t(F,G)$ is the probability that a uniformly random map

  $$\phi : V(F) \to V(G)$$

  is adjacency-preserving.

Homomorphism density

Definition (homomorphism density)

For a fixed finite graph F and a large graph G, define

$$t(F,G) := \frac{\hom(F,G)}{|V(G)|^{|V(F)|}}.$$

- $t(F,G)$ is the probability that a uniformly random map

  $$\phi : V(F) \to V(G)$$

  is adjacency-preserving.
- Similarly define $t_{\mathrm{ind}}(F,G)$ and $t_{\mathrm{inj}}(F,G)$ for induced and injective densities.

Homomorphism density

Definition (homomorphism density)

For a fixed finite graph F and a large graph G, define

$$t(F,G) := \frac{\hom(F,G)}{|V(G)|^{|V(F)|}}.$$

- $t(F,G)$ is the probability that a uniformly random map

  $$\phi : V(F) \to V(G)$$

  is adjacency-preserving.
- Similarly define $t_{\text{ind}}(F,G)$ and $t_{\text{inj}}(F,G)$ for induced and injective densities.
- These are asymptotically equivalent for our purposes.

Homomorphism density

- In particular,

$$|t_{\mathrm{inj}}(F,G) - t(F,G)| \le \frac{1}{|V(G)|}\binom{|V(F)|}{2}.$$

Homomorphism density

- In particular,

$$|t_{\text{inj}}(F,G) - t(F,G)| \leq \frac{1}{|V(G)|}\binom{|V(F)|}{2}.$$

- Homomorphism densities are extremely useful in understanding large graphs.

Homomorphism density

- In particular,

$$|t_{\mathrm{inj}}(F,G) - t(F,G)| \leq \frac{1}{|V(G)|}\binom{|V(F)|}{2}.$$

- Homomorphism densities are extremely useful in understanding large graphs.
- For example, $t_{\mathrm{ind}}(F,G)$ is the probability that $|V(F)|$ sampled vertices induce a labeled copy of F.

Homomorphism density

- In particular,

$$|t_{\mathrm{inj}}(F,G) - t(F,G)| \leq \frac{1}{|V(G)|}\binom{|V(F)|}{2}.$$

- Homomorphism densities are extremely useful in understanding large graphs.
- For example, $t_{\mathrm{ind}}(F,G)$ is the probability that $|V(F)|$ sampled vertices induce a labeled copy of F.
- This will be important when we discuss convergence of graph sequences. In particular,

Definition (Subgraph sampling method)

Recall the definition of $t_{\mathrm{ind}}(F,G)$: we sample uniformly at random a k-element subset of $V(G)$ and return the subgraph induced by it. The probability that the sampled graph is F is $t_{\mathrm{ind}}(F,G)$.

Subsection 2

Graphons and Kernels

Graphons and Kernels

Let $\mathcal{W}$ denote the space of all bounded symmetric measurable functions

$$W : [0,1]^2 \to \mathbb{R}.$$

Definition (Kernel)

The elements of $\mathcal{W}$ are called **kernels**.

Graphons and Kernels

Let $\mathcal{W}$ denote the space of all bounded symmetric measurable functions

$$W : [0,1]^2 \to \mathbb{R}.$$

Definition (Kernel)

The elements of $\mathcal{W}$ are called **kernels**.

Let $\mathcal{W}_0$ denote the set of all kernels such that $0 \leq W \leq 1$.

Definition (Graphon)

The elements of $\mathcal{W}_0$ are called **graphons**.

Graphons and Kernels

- A graphon which only takes values 0 and 1 can be considered as a **graph on node set** $[0, 1]$. Think of it as an continuous adjacency matrix.

Graphons and Kernels

- A graphon which only takes values 0 and 1 can be considered as a **graph on node set** $[0, 1]$. Think of it as an continuous adjacency matrix.
- Kernels generalize **weighted graphs** in the following sense: partition the interval $[0, 1]$ into finitely many measurable sets $S_1 \cup \cdots \cup S_k$, such that W is constant on each product $S_i \times S_j$. We call this object the **Stepfunction**.

Graphons and Kernels

- A graphon which only takes values 0 and 1 can be considered as a **graph on node set** $[0,1]$. Think of it as an continuous adjacency matrix.
- Kernels generalize **weighted graphs** in the following sense: partition the interval $[0,1]$ into finitely many measurable sets $S_1 \cup \cdots \cup S_k$, such that W is constant on each product $S_i \times S_j$. We call this object the **Stepfunction**.
- We say two kernels W, U are **weakly isomorphic** if

$$t(F,W) = t(F,U)$$

for every graph F.

Graphons and Kernels

- A graphon which only takes values 0 and 1 can be considered as a **graph on node set** $[0,1]$. Think of it as an continuous adjacency matrix.
- Kernels generalize **weighted graphs** in the following sense: partition the interval $[0,1]$ into finitely many measurable sets $S_1 \cup \cdots \cup S_k$, such that W is constant on each product $S_i \times S_j$. We call this object the **Stepfunction**.
- We say two kernels W, U are **weakly isomorphic** if

$$t(F, W) = t(F, U)$$

  for every graph F.
- Graphon will serve as the **limit object** when we talk convergence of graph sequence. We'll see why this is true later.

Subsection 3

Cut distance

Cut distance

Definition (Cut norm)

(Frieze, Kannan, 1999) Let matrix $A \in \mathbb{R}^{n\times n}$. Its **cut norm** is

$$\|A\|_\square = \frac{1}{n^2} \max_{S,T\subseteq[n]} \left| \sum_{i\in S,\, j\in T} A_{ij} \right|.$$

We derive **cut distance** from this definition.

Cut distance

Definition (Cut norm)

(Frieze, Kannan, 1999) Let matrix $A \in \mathbb{R}^{n\times n}$. Its **cut norm** is

$$\|A\|_\square = \frac{1}{n^2} \max_{S,T\subseteq[n]} \left| \sum_{i\in S,\, j\in T} A_{ij} \right|.$$

We derive **cut distance** from this definition.

Definition (Cut distance of graph with same vertex set)

For two graphs on the same vertex set, we define their **cut distance** to be

$$d_\square(G,G') = \frac{1}{n^2} \max_{S,T\subseteq V} \left| e_G(S,T) - e_{G'}(S,T) \right| = \|A_G - A_{G'}\|_\square.$$

Cut distance

Definition (Cut distance of graph with same vertex set)

For two graphs on the same vertex set, we define their **cut distance** to be

$$d_\square(G, G') = \frac{1}{n^2} \max_{S,T \subseteq V} \left| e_G(S,T) - e_{G'}(S,T) \right| = \|A_G - A_{G'}\|_\square.$$

Definition (Cut distance of graph with the same number of vertices)

$$d_\square(G, G') = \min_{\hat{G}, \hat{G}'} d_\square(G, G').$$

where $\hat{G}, \hat{G}'$are all relabelings of G and G'.

The notion can be extend to arbitrary weighted graphs, with a bit more machinery.

Why don't we just use L_1 norm?

If $G, G' \sim G(n, 1/2)$ (random graph with edge probability $\frac{1}{2}$)are independent, then typically

$$d_1(G, G') \approx \tfrac{1}{2}, \qquad d_\square(G, G') = O\left(\frac{1}{\sqrt{n}}\right).$$

Why don't we just use L_1 norm?

If $G, G' \sim G(n, 1/2)$ (random graph with edge probability $\frac{1}{2}$)are independent, then typically

$$d_1(G, G') \approx \frac{1}{2}, \qquad d_\square(G, G') = O\left(\frac{1}{\sqrt{n}}\right).$$

- Edge-wise, the graphs disagree a lot.

Why don't we just use L_1 norm?

If $G, G' \sim G(n, 1/2)$ (random graph with edge probability $\frac{1}{2}$)are independent, then typically

$$d_1(G, G') \approx \tfrac{1}{2}, \qquad d_\square(G, G') = O\left(\frac{1}{\sqrt{n}}\right).$$

- Edge-wise, the graphs disagree a lot.
- But structurally, they look the same: both are just "random of density 1/2."

Why don't we just use L_1 norm?

If $G, G' \sim G(n, 1/2)$ (random graph with edge probability $\frac{1}{2}$)are independent, then typically

$$d_1(G, G') \approx \tfrac{1}{2}, \qquad d_\square(G, G') = O\left(\frac{1}{\sqrt{n}}\right).$$

- Edge-wise, the graphs disagree a lot.
- But structurally, they look the same: both are just "random of density 1/2."
- Therefore, cut distance measure the global structure similarity between graphs instead of looking at local structure.

Cut distance for graphons

Now it's intuitive to extend this notion to graphons. Compare the two definitions.

Definition (Cut norm for matrix)

(Frieze, Kannan, 1999) Let matrix $A \in \mathbb{R}^{n\times n}$. Its **cut norm** is

$$\|A\|_\square = \frac{1}{n^2} \max_{S,T\subseteq[n]} \left| \sum_{i\in S,\, j\in T} A_{ij} \right|.$$

Cut distance for graphons

Now it's intuitive to extend this notion to graphons. Compare the two definitions.

Definition (Cut norm for matrix)

(Frieze, Kannan, 1999) Let matrix $A \in \mathbb{R}^{n\times n}$. Its **cut norm** is

$$\|A\|_{\square} = \frac{1}{n^2} \max_{S,T\subseteq[n]} \left| \sum_{i\in S,\, j\in T} A_{ij} \right| .$$

Definition (Cut distance for graphons)

For graphons, the cut norm becomes

$$\|W\|_{\square} = \sup_{S,T\subseteq[0,1]} \left| \int_{S\times T} W(x,y)\, dx\, dy \right| .$$

Cut distance for graphons

Definition (Cut distance for graphons)

For graphons, the cut norm becomes

$$\|W\|_\square = \sup_{S,T \subseteq [0,1]} \left| \int_{S \times T} W(x,y)\,dx\,dy \right|.$$

Hence, define the **cut distance**

$$d_\square(U,W) = \|U - W\|_\square.$$

Section 3

Regularity Lemma

Dense graphs look like block graphs

Suppose $\mathcal{P} = \{V_1, \ldots, V_k\}$ is a partition of $V(G)$. Define the averaged graph $G_{\mathcal{P}}$ by replacing each block $V_i \times V_j$ by its edge density.

Dense graphs look like block graphs

Suppose $\mathcal{P} = \{V_1, \ldots, V_k\}$ is a partition of $V(G)$. Define the averaged graph $G_{\mathcal{P}}$ by replacing each block $V_i \times V_j$ by its edge density.

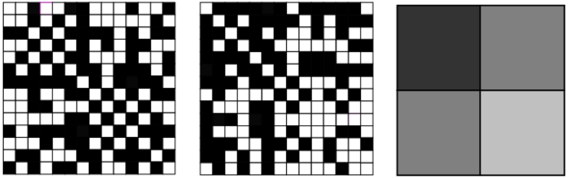

FIGURE 1.6. A random-looking pixel picture, an informative re-arrangement, and its regularity partition

Dense graphs look like block graphs

Suppose $\mathcal{P} = \{V_1, \ldots, V_k\}$ is a partition of $V(G)$. Define the averaged graph $G_{\mathcal{P}}$ by replacing each block $V_i \times V_j$ by its edge density.

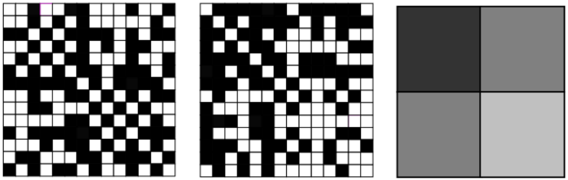

FIGURE 1.6. A random-looking pixel picture, an informative re-arrangement, and its regularity partition

- The regularity lemma says, roughly speaking, the vertex set of every graph G has an partition P into a "small" number of classes, such that G_P is " close " to G. Depending on what we mean by " close ", various forms of the lemma can be stated. Today we focus on **the weak form**.

Dense graphs look like block graphs

Suppose $\mathcal{P} = \{V_1, \ldots, V_k\}$ is a partition of $V(G)$. Define the averaged graph $G_{\mathcal{P}}$ by replacing each block $V_i \times V_j$ by its edge density.

Theorem (Weak regularity lemma)

(Frieze, Kannan, 1999)For every integer $k > 1$ and every graph G, there exists a partition $\mathcal{P}$ into k classes such that

$$d_{\square}(G, G_{\mathcal{P}}) \leq \frac{2}{\sqrt{\log k}}.$$

- More classes $\Rightarrow$ better approximation in cut distance.

Dense graphs look like block graphs

Suppose $\mathcal{P} = \{V_1, \ldots, V_k\}$ is a partition of $V(G)$. Define the averaged graph $G_{\mathcal{P}}$ by replacing each block $V_i \times V_j$ by its edge density.

Theorem (Weak regularity lemma)

(Frieze, Kannan, 1999)For every integer $k > 1$ and every graph G, there exists a partition $\mathcal{P}$ into k classes such that

$$d_{\square}(G, G_{\mathcal{P}}) \leq \frac{2}{\sqrt{\log k}}.$$

- More classes $\Rightarrow$ better approximation in cut distance.
- For people who knows the original form of Regularity Lemma, the weak regularity lemma here is weak in the sense it has a worse partition, but we have finer control on the number of classes. This often has more practical uses.

Regularity Lemma for Kernels

Theorem (Weak regularity lemma for kernels)

For every function $W \in \mathcal{W}$ and $k \geq 1$, there is a ***stepfunction*** *U such that*

$$\|W - U\|_{\square} < \frac{2}{\sqrt{log k}} \|W\|_2$$

Regularity Lemma for kernels

In fact, something stronger is true: the approximating stepfunction admits a polynomial-size description.

Regularity Lemma for kernels

In fact, something stronger is true: the approximating stepfunction admits a polynomial-size description.

Theorem (Frieze–Kannan)

For every kernel U and every $k \geq 1$, there exist measurable sets $S_i, T_i \subseteq [0,1]$ and real numbers a_i such that

$$\left\| U - \sum_{i=1}^{k} a_i \mathbb{1}_{S_i \times T_i} \right\|_{\square} \leq \frac{\|U\|_2}{\sqrt{k}}.$$

In particular, if U is a graphon, then the right-hand side is at most $\frac{1}{\sqrt{k}}$.

Regularity Lemma for kernels

In fact, something stronger is true: the approximating stepfunction admits a polynomial-size description.

Theorem (Frieze–Kannan)

For every kernel U and every $k \geq 1$, there exist measurable sets $S_i, T_i \subseteq [0,1]$ and real numbers a_i such that

$$\left\| U - \sum_{i=1}^{k} a_i \mathbb{1}_{S_i \times T_i} \right\|_{\square} \leq \frac{\|U\|_2}{\sqrt{k}}.$$

In particular, if U is a graphon, then the right-hand side is at most $\frac{1}{\sqrt{k}}$.

To see the proof of this we first look at a lemma.

Regularity Lemma for kernels: A lemma

Lemma (A good Lemma)

For every kernel $U \in W$, there exist measurable sets $S, T \subseteq [0,1]$ and a real number a such that

$$\|U - a\mathbb{1}_{S\times T}\|_2^2 \leq \|U\|_2^2 - \|U\|_\square^2.$$

Proof.

Choose S, T so that

$$\left| \int_{S\times T} U \right| = \|U\|_\square,$$

and set

$$a := \frac{\int_{S\times T} U}{\lambda(S)\lambda(T)}.$$

$\square$

Regularity Lemma for kernels: A lemma

Lemma (A good Lemma)

For every kernel $U \in W$, there exist measurable sets $S, T \subseteq [0,1]$ and a real number a such that

$$\|U - a\mathbb{1}_{S\times T}\|_2^2 \leq \|U\|_2^2 - \|U\|_\square^2.$$

Proof.

Then

$$\|U - a\mathbb{1}_{S\times T}\|_2^2 = \|U\|_2^2 - 2a\int_{S\times T} U + a^2\|\mathbb{1}_{S\times T}\|_2^2.$$

Since $\|\mathbb{1}_{S\times T}\|_2^2 = \lambda(S)\lambda(T)$, this becomes

$$\|U - a\mathbb{1}_{S\times T}\|_2^2 = \|U\|_2^2 - \frac{\left(\int_{S\times T} U\right)^2}{\lambda(S)\lambda(T)}.$$

Regularity Lemma for kernels: A lemma

Lemma (A good Lemma)

For every kernel $U \in W$, there exist measurable sets $S, T \subseteq [0,1]$ and a real number a such that

$$\|U - a\mathbb{1}_{S\times T}\|_2^2 \leq \|U\|_2^2 - \|U\|_\square^2.$$

Proof.

Finally, since $\lambda(S)\lambda(T) \leq 1$,

$$\|U - a\mathbb{1}_{S\times T}\|_2^2 \leq \|U\|_2^2 - \left|\int_{S\times T} U\right|^2 = \|U\|_2^2 - \|U\|_\square^2.$$

$\square$

Regularity Lemma for kernels: greedy step

Apply the lemma greedily.

Set
$$U_0 := U.$$
For $j = 0, 1, \ldots, k-1$, choose S_{j+1}, T_{j+1} and a_{j+1} from the lemma applied to U_j, and define
$$U_{j+1} := U_j - a_{j+1} \mathbb{1}_{S_{j+1} \times T_{j+1}}.$$

Regularity Lemma for kernels: greedy step

Apply the lemma greedily.

Set

$$U_0 := U.$$

For $j = 0, 1, \ldots, k-1$, choose S_{j+1}, T_{j+1} and a_{j+1} from the lemma applied to U_j, and define

$$U_{j+1} := U_j - a_{j+1} \mathbb{1}_{S_{j+1} \times T_{j+1}}.$$

Then for each j,

$$\|U_{j+1}\|_2^2 \leq \|U_j\|_2^2 - \|U_j\|_\square^2.$$

Regularity Lemma for kernels: greedy step

Apply the lemma greedily.

Set

$$U_0 := U.$$

For $j = 0, 1, \ldots, k-1$, choose S_{j+1}, T_{j+1} and a_{j+1} from the lemma applied to U_j, and define

$$U_{j+1} := U_j - a_{j+1} \mathbb{1}_{S_{j+1} \times T_{j+1}}.$$

Then for each j,

$$\|U_{j+1}\|_2^2 \leq \|U_j\|_2^2 - \|U_j\|_\square^2.$$

Summing over $j = 0, \ldots, k-1$ gives

$$\sum_{j=0}^{k-1} \|U_j\|_\square^2 \leq \|U\|_2^2.$$

Regularity Lemma for kernels: conclusion

Since

$$\sum_{j=0}^{k-1} \|U_j\|_\square^2 \leq \|U\|_2^2,$$

there exists some $j < k$ such that

$$\|U_j\|_\square \leq \frac{\|U\|_2}{\sqrt{k}}.$$

Regularity Lemma for kernels: conclusion

Since

$$\sum_{j=0}^{k-1} \|U_j\|_\square^2 \le \|U\|_2^2,$$

there exists some $j < k$ such that

$$\|U_j\|_\square \le \frac{\|U\|_2}{\sqrt{k}}.$$

But by construction,

$$U_j = U - \sum_{i=1}^{j} a_i \mathbb{1}_{S_i \times T_i}.$$

Regularity Lemma for kernels: conclusion

Setting the remaining coefficients equal to 0, we obtain

$$\left\| U - \sum_{i=1}^{k} a_i \mathbb{1}_{S_i \times T_i} \right\|_{\square} \leq \frac{\|U\|_2}{\sqrt{k}}.$$

Regularity Lemma for kernels: conclusion

Setting the remaining coefficients equal to 0, we obtain

$$\left\| U - \sum_{i=1}^{k} a_i \mathbb{1}_{S_i \times T_i} \right\|_\square \leq \frac{\|U\|_2}{\sqrt{k}}.$$

If U is a graphon, then $\|U\|_2 \leq 1$, so

$$\left\| U - \sum_{i=1}^{k} a_i \mathbb{1}_{S_i \times T_i} \right\|_\square \leq \frac{1}{\sqrt{k}}.$$

Remarks

- The approximating stepfunction obtained in this lemma is not necessarily the original stepping of W. However, one can fix that by a constant factor of 2. (Laszlo Lovasz, large network and graph limits, lemma 9.12)

Remarks

- The approximating stepfunction obtained in this lemma is not necessarily the original stepping of W. However, one can fix that by a constant factor of 2. (Laszlo Lovasz, large network and graph limits, lemma 9.12)
- The regularity lemma is equivalent to the statement that **The graphon space equipped with the cut distance is compact.** (Lovasz, Szegedy, 2007) This is in fact the strongest notion of regularity we can obtain. We will use this theorem in our discussion of graph sequence convergence.

Section 4

Convergence of graph sequence

Convergence of graph sequence: preliminaries

We now state two theorems that allow us to relate the two main quantities we have seen: homomorphism densities and the cut distance.

Convergence of graph sequence: preliminaries

We now state two theorems that allow us to relate the two main quantities we have seen: homomorphism densities and the cut distance.

Theorem (Counting Lemma)

(Lovasz, and Szegedy, 2006) Let F be a simple graph, and let W, W' be graphons. Then,

$$|t(F, W) - t(F, W')| \leq e(F)\delta_{\square}(W, W').$$

Convergence of graph sequence: preliminaries

We now state two theorems that allow us to relate the two main quantities we have seen: homomorphism densities and the cut distance.

Theorem (Counting Lemma)

(Lovasz, and Szegedy, 2006) Let F be a simple graph, and let W, W' be graphons. Then,

$$|t(F, W) - t(F, W')| \leq e(F)\delta_\square(W, W').$$

- Counting Lemma implies the function $W \to t(F, W)$ is Lipschitz-continuous on the graphon space $(W, \delta_\square)$. Use this in the proof of convergence theorem.
- local similarity $\Rightarrow$ global similarity

Convergence of graph sequence: preliminaries

In fact the converse is true: if two large graphs are locally similar w.r.t **homomorphism densities**, they are also globally close w.r.t **cut distance**.

Theorem (Inverse Counting Lemma)

(Borgs, Chayes, Lovasz, Sos, and Vesztergombi, 2008) Let k be a positive integer, let $U, W \in W_0$, and assume that for every simple graph F on k nodes, we have

$$|t(F,U) - t(F,W)| \leq 2^{-k^2}.$$

Then,

$$\delta_{\square}(U,W) \leq \frac{50}{\sqrt{logk}}$$

Convergence of graph sequence

We are now ready to state the definition of convergent graph sequence. Recall we already have Compactness and sampling with induced homomorphism density.

Convergence of graph sequence

We are now ready to state the definition of convergent graph sequence. Recall we already have Compactness and sampling with induced homomorphism density.

Theorem (Compactness)

The space $(\mathcal{W}, \delta_\square)$ is compact. Here $\delta_\square$ denotes the cut distance.

Convergence of graph sequence

We are now ready to state the definition of convergent graph sequence. Recall we already have Compactness and sampling with induced homomorphism density.

Theorem (Compactness)

The space $(W, \delta_\square)$ *is compact. Here* $\delta_\square$ *denotes the cut distance.*

Definition (Subgraph sampling method)

Recall the definition of $t_{\text{ind}}(F, G)$: we sample uniformly at random a k-element subset of $V(G)$ and return the subgraph induced by it. The probability that the sampled graph is F is $t_{\text{ind}}(F, G)$.

Convergence of graph sequence

We are now ready to state the definition of convergent graph sequence. Recall we already have Compactness and sampling with induced homomorphism density.

Theorem (Compactness)

The space $(W, \delta_\square)$ is compact. Here $\delta_\square$ denotes the cut distance.

Definition (Subgraph sampling method)

Recall the definition of $t_{\mathrm{ind}}(F, G)$: we sample uniformly at random a k-element subset of $V(G)$ and return the subgraph induced by it. The probability that the sampled graph is F is $t_{\mathrm{ind}}(F, G)$.

Definition (**Convergence of graph sequence**)

A sequence of graphs (G_n) with $|V(G_n)| \to \infty$ is **convergent** if the induced subgraph density $t_{\mathrm{ind}}(F, G_n)$ converges for every finite graph F.

Convergence of graph sequence

Definition (**Convergence of graph sequence**)

A sequence of graphs (G_n) with $|V(G_n)| \to \infty$ is **convergent** if the induced subgraph density $t_{\text{ind}}(F, G_n)$ converges for every finite graph F.

Convergence of graph sequence

Definition (Convergence of graph sequence)

A sequence of graphs (G_n) with $|V(G_n)| \to \infty$ is **convergent** if the induced subgraph density $t_{\text{ind}}(F, G_n)$ converges for every finite graph F.

- Exchanging $t(F, G)$ and $t_{\text{ind}}(F, G)$ in the definition is allowed here because they are asymptotically equivalent for large graphs in our setting.

Convergence of graph sequence

Definition (**Convergence of graph sequence**)

A sequence of graphs (G_n) with $|V(G_n)| \to \infty$ is **convergent** if the induced subgraph density $t_{\mathrm{ind}}(F, G_n)$ converges for every finite graph F.

- Exchanging $t(F, G)$ and $t_{\mathrm{ind}}(F, G)$ in the definition is allowed here because they are asymptotically equivalent for large graphs in our setting.
- Example: the sequence of complete graphs K_n is convergent, since a random induced subgraph is still complete.

Convergence of graph sequence

Definition (**Convergence of graph sequence**)

A sequence of graphs (G_n) with $|V(G_n)| \to \infty$ is **convergent** if the induced subgraph density $t_{\text{ind}}(F, G_n)$ converges for every finite graph F.

- Exchanging $t(F, G)$ and $t_{\text{ind}}(F, G)$ in the definition is allowed here because they are asymptotically equivalent for large graphs in our setting.
- Example: the sequence of complete graphs K_n is convergent, since a random induced subgraph is still complete.
- Example: The random graph sequence $G_n = \mathbb{G}(n, p)$ is convergent with probability 1. Indeed, since an induced k-subgraph $G_n[S]$ of G_n will be very close in distribution to $\mathbb{G}(k, p)$.

Cauchy sequence in the cut metric is convergent

Theorem (Cauchy graph sequence is convergent)

(Borgs, Chayes, Lovasz, Sos, and Vesztergombi, 2006, 2008) A sequence of simple graphs (G_n) *with* $V(G_n) \to \infty$ *is convergent if and only if it is a Cauchy sequence in the cut metric* $\delta_\square$*.*

Cauchy sequence in the cut metric is convergent

Theorem (Cauchy graph sequence is convergent)

(Borgs, Chayes, Lovasz, Sos, and Vesztergombi, 2006, 2008) A sequence of simple graphs (G_n) *with* $V(G_n) \to \infty$ *is convergent if and only if it is a Cauchy sequence in the cut metric* $\delta_\square$.

Proof.

The counting lemma implies that every Cauchy sequence is convergent. Apply the inverse counting lemma to W_{G_n}, we get the converse direction. $\square$

Cauchy sequence in the cut metric is convergent

Theorem (Cauchy graph sequence is convergent)

(Borgs, Chayes, Lovasz, Sos, and Vesztergombi, 2006, 2008) A sequence of simple graphs (G_n) with $V(G_n) \to \infty$ is convergent if and only if it is a Cauchy sequence in the cut metric $\delta_\square$.

Proof.

The counting lemma implies that every Cauchy sequence is convergent. Apply the inverse counting lemma to W_{G_n}, we get the converse direction. $\square$

We can generalize this to graphons. Proof is similar.

Cauchy sequence in the cut metric is convergent

Theorem (Cauchy graph sequence is convergent)

(Borgs, Chayes, Lovasz, Sos, and Vesztergombi, 2006, 2008) A sequence of simple graphs (G_n) *with* $V(G_n) \to \infty$ *is convergent if and only if it is a Cauchy sequence in the cut metric* $\delta_\square$.

Proof.

The counting lemma implies that every Cauchy sequence is convergent. Apply the inverse counting lemma to W_{G_n}, we get the converse direction. □

We can generalize this to graphons. Proof is similar.

Theorem (Cauchy graphon sequence is convergent)

Let (W_n) *be a sequence of graphons, and let* W *be a kernel. Then* $t(F, W_n)$ *converges for all finite simple graph* F *if and only if* W_n *is a cauchy sequence in the cut metric.*

But...umm is the right construction?

What is the limit object the sequence converges to?

- Now we have defined convergence of a graph sequence, we are of course interested in **what object it converges to**. (One will hope it's the graphon, otherwise why are we doing all this :))

But...umm is the right construction?

What is the limit object the sequence converges to?

- Now we have defined convergence of a graph sequence, we are of course interested in **what object it converges to**. (One will hope it's the graphon, otherwise why are we doing all this :))
- If you use computer to draw out the adjacency matrix of all the graphs as grayscale images, a **proof by picture** will tell that they converge to **graphons**. But how to describe it more formally?

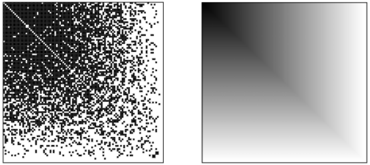

FIGURE 1.8. A randomly grown uniform attachment graph with 100 nodes, and a (continuous) function approximating it

But...umm is the right construction?

What is the limit object the sequence converges to?

- Now we have defined convergence of a graph sequence, we are of course interested in **what object it converges to**. (One will hope it's the graphon, otherwise why are we doing all this :))
- If you use computer to draw out the adjacency matrix of all the graphs as grayscale images, a **proof by picture** will tell that they converge to **graphons**. But how to describe it more formally?
- **Intuition**: View each graphon in the sequence as a **probabiltiy distribution**, then the limit should tend to some distribution σ_k on k node labeled graph. If such a distribution satisfy some certian properties, we can **uniquely reconstruct a graphon** from it!

Random Graph Models

- More formally, given a simple graph and $k \in V(G)$, the random sample $G(k, G)$ is a random graph on k labeled nodes, and we denote it by $\sigma_{G,k}$.

Random Graph Models

- More formally, given a simple graph and $k \in V(G)$, the random sample $G(k, G)$ is a random graph on k labeled nodes, and we denote it by $\sigma_{G,k}$.
- By this construction, we have $\sigma_{G,k}(F) = t_{ind}(F, G)$.

Random Graph Models

- More formally, given a simple graph and $k \in V(G)$, the random sample $G(k, G)$ is a random graph on k labeled nodes, and we denote it by $\sigma_{G,k}$.
- By this construction, we have $\sigma_{G,k}(F) = t_{ind}(F, G)$.
- If G_n is a convergent sequence, then according to the convergence theorem, the sequence of distributions $\{\sigma_{G,k}\}$ tends to some distribution σ_k.

Random Graph Models

- More formally, given a simple graph and $k \in V(G)$, the random sample $G(k, G)$ is a random graph on k labeled nodes, and we denote it by $\sigma_{G,k}$.
- By this construction, we have $\sigma_{G,k}(F) = t_{ind}(F, G)$.
- If G_n is a convergent sequence, then according to the convergence theorem, the sequence of distributions $\{\sigma_{G,k}\}$ tends to some distribution σ_k.
- Conversely, if the distribution sequence tends to some limit for every k, then the graph sequence is convergent.

Random Graph Models

- More formally, given a simple graph and $k \in V(G)$, the random sample $G(k, G)$ is a random graph on k labeled nodes, and we denote it by $\sigma_{G,k}$.
- By this construction, we have $\sigma_{G,k}(F) = t_{ind}(F, G)$.
- If G_n is a convergent sequence, then according to the convergence theorem, the sequence of distributions $\{\sigma_{G,k}\}$ tends to some distribution σ_k.
- Conversely, if the distribution sequence tends to some limit for every k, then the graph sequence is convergent.
- We will call such limit object the **Random Graph Model (rgm)**. It turns out that if it satisfies some condition called **consistent and local**, we can uniquely recover a graphon from it!

Random Graph model is the right limit object

Theorem

If a graph sequence $\{G_n\}$ is convergent, then the distribution $\sigma_k = \lim_{n\to\infty} \sigma_{k,G_n}$ forms a ***consistent and local random graph model****. Conversely, every consistent and local random graph model arises this way.*

Random Graph model is the right limit object

Theorem

If a graph sequence $\{G_n\}$ is convergent, then the distribution $\sigma_k = \lim_{n\to\infty} \sigma_{k,G_n}$ forms a ***consistent and local random graph model****. Conversely, every consistent and local random graph model arises this way.*

Proof: a rough sketch.

- Forward direction follows from consistency and locality.
- For the reverse direction, consider a consistent and local random graph model $(\sigma_1, \sigma_2, ...)$, we can generate a graph G_n from every σ_n, independently for each n. This forms a graph sequence $(G_1, G_2, ...)$.

Random Graph model is the right limit object

Theorem

If a graph sequence $\{G_n\}$ is convergent, then the distribution $\sigma_k = \lim_{n\to\infty} \sigma_{k,G_n}$ forms a ***consistent and local random graph model****. Conversely, every consistent and local random graph model arises this way.*

Proof: a rough sketch.

- Forward direction follows from consistency and locality.
- For the reverse direction, consider a consistent and local random graph model $(\sigma_1, \sigma_2, ...)$, we can generate a graph G_n from every σ_n, independently for each n. This forms a graph sequence $(G_1, G_2, ...)$.
- Then the sequence should be convergent with probability 1 [Laszlo, Large network and graph limits, lemma 11.8]. But since the sequence is convergent, this is precisely the right model we are looking for.

$\square$

Section 5

An application: the Removal Lemma

Recap for what we've done so far

- Graph homomorphism $t_{ind}(F, G)$ gives us a way to sample subgraph from very large graphs.

Recap for what we've done so far

- Graph homomorphism $t_{ind}(F, G)$ gives us a way to sample subgraph from very large graphs.
- Graphon is a continuous model to describe graphs (A continuous adjacency matrix). It also gives a nice formulation of the Regularity Lemma.

Recap for what we've done so far

- Graph homomorphism $t_{ind}(F, G)$ gives us a way to sample subgraph from very large graphs.
- Graphon is a continuous model to describe graphs (A continuous adjacency matrix). It also gives a nice formulation of the Regularity Lemma.
- Cut distance $\delta_\square$ is the right metric to measure similarity between graphs. Graphs/Graphons are convergent in the cut metric.

Recap for what we've done so far

- Graph homomorphism $t_{ind}(F, G)$ gives us a way to sample subgraph from very large graphs.
- Graphon is a continuous model to describe graphs (A continuous adjacency matrix). It also gives a nice formulation of the Regularity Lemma.
- Cut distance $\delta_\square$ is the right metric to measure similarity between graphs. Graphs/Graphons are convergent in the cut metric.
- It turns out that the limit object of graphon sequence are random graph models, which are in fact still graphons. So this construction is exactly what we expected.

Recap for what we've done so far

- Graph homomorphism $t_{ind}(F, G)$ gives us a way to sample subgraph from very large graphs.
- Graphon is a continuous model to describe graphs (A continuous adjacency matrix). It also gives a nice formulation of the Regularity Lemma.
- Cut distance $\delta_\square$ is the right metric to measure similarity between graphs. Graphs/Graphons are convergent in the cut metric.
- It turns out that the limit object of graphon sequence are random graph models, which are in fact still graphons. So this construction is exactly what we expected.
- Now we look at a concrete example of using these, which is the triangle removal lemma.

the Removal Lemma

Theorem (Removal Lemma)

For every $\epsilon > 0$, there exists $\delta > 0$ such that if a simple graph G with n nodes has at most δn^3 triangles, then we can delete ϵn^2 edges from G so that the remaining graph has no triangles.

Proof.

Proof by hand-waving. □

Questions?

Bibliography I

Christian Borgs, Jennifer T. Chayes, László Lovász, Vera T. Sós, and Katalin Vesztergombi.

Convergent sequences of dense graphs i: Subgraph frequencies, metric properties and testing.

Advances in Mathematics, 219(6):1801–1851, 2008.

F. R. K. Chung, R. L. Graham, and R. M. Wilson.

Quasi-random graphs.

Combinatorica, 9(4):345–362, 1989.

Persi Diaconis and Svante Janson.

Graph limits and exchangeable random graphs.

Rendiconti di Matematica e delle sue Applicazioni, 28:33–61, 2008.

Alan Frieze and Ravi Kannan.

Quick approximation to matrices and applications.

Combinatorica, 19(2):175–220, 1999.

Bibliography II

László Lovász.

Large Networks and Graph Limits, volume 60 of *Colloquium Publications*.

American Mathematical Society, 2012.

László Lovász and Balázs Szegedy.

Limits of dense graph sequences.

Journal of Combinatorial Theory, Series B, 96(6):933–957, 2006.

László Lovász and Balázs Szegedy.

Szemerédi's lemma for the analyst.

Geometric and Functional Analysis, 17(1):252–270, 2007.