1 Introduction

We have been discussing the Gibbs measure on graphs $\nu(G) = e^{\beta\cdot\Delta(G)}\mu(G)$, where $\mu = \mathcal{G}(n, 1/2)$ and $\Delta(G)$ is the number of triangles in G (or $3\times$ that quantity). So we'd like to understand triangle counts in general graphs to see where this measure lives.

More generally, we'd like to understand the count of any fixed subgraph. In these notes, we'll introduce a class of models where the subgraph counts are easy to understand, and then prove that *every* graph looks like a graph from one of these models, in the sense that the subgraph counts are similar. The main reference for these notes is Lovász's book "Large Networks and Graph Limits," mostly chapters 7-10.

1.1 Subgraph densities

To set some notation, for any fixed finite graph F, let $V(F)$ and $E(F)$ denote the vertex set and edge set respectively. If G is another graph, we define the *subgraph density of F in G* as

$$t(F, G) = \mathbb{P}[|V(F)| \text{ random vertices in } G \text{ form a copy of } F]. \tag{1}$$

Here, "$|V(F)|$ random vertices" means, for each $v \in F$, sampling $X_v \in V(G)$ uniformly, independently for different v. These vertices "form a copy of F" if, for each $uv \in E(F)$, $X_uX_v \in E(G)$. Note that this means the subgraph F *does not need to be induced* by the vertices.

A word of warning: for some graphs F, it is not easy to see how $t(F, G)$ relates to the actual subgraph *count* of F in G. For instance, $t(\square, -) = \frac{1}{8}$, but a single edge has no squares as subgraphs. However, when $|V(G)| \gg |V(F)|^2$ so that there is little chance of sampled vertices being the same, we have

$$t(F, G) \approx \frac{\text{number of } \textit{labeled} \text{ copies of } F \text{ in } G}{|V(G)|^{|V(F)|}} = |V(F)|! \frac{\text{number of } \textit{unlabeled} \text{ copies of } F \text{ in } G}{|V(G)|^{|V(F)|}}.$$

1.2 Homogenizing with respect to a partition

Let G be a graph and $\mathcal{P} = \{V_1, \ldots, V_k\}$ be a partition of its vertices $V(G)$. We define a weighted graph $G_{\mathcal{P}}$ on the same edge set by "averaging the edges between pieces" of the partition. More precisely, for every $u, v \in V$, we include the edge uv in $E(G_{\mathcal{P}})$ with weight $d_G(V_i, V_j)$ if $u \in V_i$ and $v \in V_j$, where

$$d_G(X, Y) = \frac{\text{number of edges between } X \text{ and } Y}{|X||Y|}$$

is the density of edges between X and Y.

We can also define subgraph densities $t(F, H)$ for weighted graphs H with edge-weights in $[0, 1]$; in (1), we just interpret this probability as also sampling the edges between two chosen vertices with probability equal to the edge weight. Now for a graph G and a partition $\mathcal{P} = \{V_1, \ldots, V_k\}$ of its vertices, $t(F, G_{\mathcal{P}})$ is often easier to compute, and can be written in terms of the densities between the pieces.

$$t(F, G_{\mathcal{P}}) = \sum_{(i_v)\in\{1,\ldots,k\}^{V(F)}} \prod_{v\in V(F)} \frac{|V_{i_v}|}{n} \prod_{uv\in E(F)} d_G(V_{i_u}, V_{i_v}).$$

For instance, if $F = \triangle$ is a triangle, then

$$t(\triangle, G_{\mathcal{P}}) = \frac{1}{n^3} \sum_{i,j,\ell}^{k} |V_i||V_j||V_\ell| d_G(V_i, V_j) d_G(V_j, V_\ell) d_G(V_\ell, V_i).$$

1.3 The (weak) regularity lemma

The regularity lemma states that every graph G looks like $G_{\mathcal{P}}$ for some partition, in the sense that we have $t(F,G) \approx t(F,G_{\mathcal{P}})$ for all small enough finite graphs. Of course, this would be trivial if we allowed the number k of pieces in the partition to depend on the graph; just make a different part for each vertex of G.

The strength of the regularity lemma is the fact that the error in the approximation $t(F,G) \approx t(F,G_{\mathcal{P}})$ *only depends on the number of parts k in the partition, and not on the graph G.* Here's the statement.

Theorem 1 (weak regularity lemma I)**.** *For any $k \geq 1$ and any graph G, there is a partition $\mathcal{P}$ of $V(G)$ with k parts such that*

$$|t(F,G) - t(F,G_{\mathcal{P}})| < \frac{2|E(F)|}{\sqrt{\log k}}$$

for every finite graph F.

A brief word on why this is called the "weak" regularity lemma is in order. The original regularity lemma, due to Szemerédi in 1975, states that for every $\varepsilon > 0$, there is some number $S(\varepsilon) \in \mathbb{N}$ such that every graph $G = (V,E)$ has an *equitable partition* $V = V_1 \sqcup \cdots \sqcup V_k$ (meaning the pieces differ in size by at most 1) with $k \leq S(\varepsilon)$, and such that the bipartite graph between V_i and V_j is ε-homogeneous, for all but εk^2 pairs of indices i,j. Here ε-homogeneity means that every subgraph has close to "the right number of edges", and ε measures the error tolerance.

This is stronger than Theorem 1 for a few reasons: first, in Theorem 1 there is no guarantee of an equitable partition (although this can be remedied). More importantly, ε-homogeneity is a slightly stronger condition and it implies that the subgraph densities are close to what they should be. However, the original regularity lemma is much more difficult to prove. Moreover the number $S(\varepsilon)$ is humongous: it is a power tower $2^{2^{2^{\cdots}}}$ of height about $1/\varepsilon^2$.

The weaker regularity lemma, Theorem 1, has a much more reasonable partition size of about $\exp(2/\varepsilon^2)$ (of course, the ε means something different here). And, more relevant for us, it is easier to prove. Additionally, the formulation deals more directly with subgraph densities, the quantity of interest to us, although we will need to make a detour through some more analytic formulations in order to prove it. Theorem 1 is due to Frieze and Kannan in 1999.

Finally, as a special case which we will use later in the course, here is what the theorem says when we take $F = \triangle$, and rewrite the number of partitions in terms of the desired error bound:

Corollary 2. *Let $\varepsilon > 0$ and let G be a graph with n vertices. Then there exists a partition $\mathcal{P}$ of the vertices into at most $4^{1/\varepsilon^2}$ parts $V_1, \ldots, V_m$ such that if $d_{ij} = d_G(V_i, V_j)$, then*

$$\left| t(\triangle, G) - \frac{1}{n^3} \sum_{i,j,k=1}^{m} |V_i||V_j||V_k| d_{ij} d_{jk} d_{ki} \right| \leq 3\varepsilon.$$

2 Graphons

It is much easier to understand and prove the weak regularity lemma if we use the language of graphons, which are typically thought of as a form of *graph limit*, although this idea will not be too relevant for us.

2.1 Definitions

Simply put, a graphon is a symmetric measurable function $W : [0,1]^2 \to [0,1]$. Here "symmetric" means that $W(x,y) = W(y,x)$. Every graph can be turned into a graphon in a canonical (not bijective) way, and from every graphon one can sample a finite graph, in a way which generalizes the stochastic block model.

To turn a graph G into a graphon W_G, first split up the interval $[0,1]$ into pieces Q_v for $v \in V(G)$, with equal measure, and then define

$$W_G = \sum_{uv \in E(G)} \mathbf{1}_{Q_u \times Q_v}.$$

In words, this is a matrix plot of the adjacency matrix of G. Note that the map $G \mapsto W_G$ is not bijective, since subdividing each Q_v into two, for instance, gives $W_{G^{\bowtie}} = W_G$. Additionally, since we consider graphs to be isomorphic when the vertices are relabeled, we consider graphons to be isomorphic when the points of $[0,1]$ are relabeled by a *measure-preserving transformation* (which does not need to be invertible (!)). This is why the ambiguity in the definition of Q_v above is okay, but in general we won't worry too much about this notion of isomorphism in these notes.

To sample a graph $G(n, W)$ from a graphon W, sample n points $x_1, \dots, x_n \in [0,1]$ uniformly and independently, and add each edge uv to $E[G(n,W)]$ with probability $W(x_u, x_v)$ independently of one another. This is completely analogous to the stochastic block model, and indeed a stochastic block model with k parts can be thought of as a special kind of graphon which is constant on k^2 sets $Q_i \times Q_j$, for some partition $Q = (Q_1, \dots, Q_k)$ of $[0,1]$. From here on out, we call this type of graphon a *stepfunction with k steps*.

We can generalize the subgraph density to a graphon as follows:

$$t(F, W) = \mathbb{P}[G(|V(F)|, W) \text{ contains } F] = \int_{[0,1]^{V(F)}} \prod_{uv \in E(F)} W(x_u, x_v) \prod_{v \in V(F)} dx_v.$$

As before, we have $t(F, G(n, W)) \to t(F, W)$ almost surely. Also note that for any finite graph G, we have $t(F, G) = t(F, W_G)$. This hints that we can generalize Theorem 1 to graphons.

Theorem 3 (weak regularity lemma II)**.** *For any $k \geq 1$ and any graphon W, there is a stepfunction S with k steps such that*

$$|t(F, W) - t(F, S)| < \frac{2|E(F)|}{\sqrt{\log k}}$$

for every finite graph F.

It is not too hard to see (perhaps easier from a later version of the statement) that if we start with W_G for some finite graph, then the steps of the stepfunction S produced by the theorem can be chosen to align with the partition Q_v corresponding to the vertices of G. Then $t(F, S) = t(F, G_{\mathcal{P}})$ for the corresponding partition $\mathcal{P}$ of the vertices, and we recover Theorem 1.

Without this extra massaging, we still get an interesting theorem, which states that every graph G is close to some stochastic block model S with k pieces, in the sense that $t(F, G) \approx t(F, S)$, where the error bound only depends on k and F.

2.2 Cut metric

Now we have put the two objects in the statement of the weak regularity lemma on equal footing (they are both graphons). But to actually prove this, we will need to introduce a metric which induces the same topology on graphons as the one where $W_n \to W$ if $t(F, W_n) \to t(F, W)$ for all F.

First, for any symmetric function $W : [0,1]^2 \to \mathbb{R}$, we define

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

This is called the *cut norm*, since it measures the maximal "edge weight" between two sets of "vertices."

One way to measure the distance between two graphons W and U would be to calculate $\|W - U\|_\square$. However, this does not account for the fact that graphons should be considered isomorphic under relabeling $[0,1]$ by a measure-preserving transformation. The fix is simple enough: define the *cut metric*

$$\delta_\square(W, U) = \inf_{\substack{\varphi : [0,1] \to [0,1] \\ \text{measure-preserving}}} \|W - U^\varphi\|_\square,$$

where U^φ is the graphon U under relabeling by φ, i.e. $U^\varphi(x, y) = U(\varphi(x), \varphi(y))$. The "min-max" feature in the cut metric can make it a bit tough to reason about. Luckily, in these notes, as we will see, we don't really need to worry about $\delta_\square$; we will only need to work with the cut norm $\|\cdot\|_\square$.

2.2.1 Example: Erdős-Rényi graph converges to "one-half" graphon

To help make sense of this, let's see (at least heuristically) why $\delta_\square(W_{G(n,1/2)}, \frac{1}{2}) \to 0$, which means that in the topology of the cut metric, the Erdős-Renyi graph converges to the graphon which is the constant $\frac{1}{2}$ on the unit square. It suffices to show that $\|W_{G(n,1/2)} - \frac{1}{2}\|_\square \to 0$ for any realization of $W_{G(n,1/2)}$; let's just fix an ordering on the vertices and say that vertex i corresponds to the interval $Q_i = [\frac{i-1}{n}, \frac{i}{n}]$.

For any $S, T \subseteq [0,1]$ of positive measure, $W_{G(n,1/2)} - \frac{1}{2}$ consists of a bunch of tiny squares with value $\pm\frac{1}{2}$, with the sign being decided independently for each square (ignoring the symmetry across the diagonal, which can be easily accounted for). So, by some uniform law of large numbers,

$$\int_{S\times T} \left(W_{G(n,1/2)}(x,y) - \frac{1}{2} \right) dx\, dy = o(1)$$

uniformly among all $S, T \subseteq [0,1]$.

This explains why it is very important that the absolute value bars are on the *outside* of the integral. If they were on the inside we would just recover the L^1 norm, and $\|W_{G(n,1/2)} - \frac{1}{2}\|_1 = \frac{1}{2}$ for every n. Of course, by the Triangle inequality and Jensen's inequality, we have

$$\|W\|_\square \le \|W\|_1 \le \|W\|_2 \le \|W\|_\infty.$$

These comparisons are not so important for us in these notes, but we will make use of the L^2 norm of a graphon in order to prove Theorem 3.

2.2.2 Functional equivalence, and maximum is obtained

Finally, we present a technical result about the cut norm which we will need for later proofs.

Lemma 4. *For any symmetric function $W : [0,1]^2 \to \mathbb{R}$,*

$$\|W\|_\square = \sup_{S,T\subseteq[0,1]} \left| \int_{S\times T} W(x,y)\, dx\, dy \right| = \sup_{f,g:[0,1]\to[0,1]} \left| \int_{[0,1]^2} f(x)g(y)W(x,y)\, dx\, dy \right|,$$

and both suprema are attained (i.e. they are maxima).

Proof sketch. The supremum over functions is attained because the set of functions $[0,1] \to [0,1]$ is compact, and the given integral is a continuous function of f and g, both statements holding for the weak* topology. Additionally, the maximizing functions can be chosen to be $\{0,1\}$-valued, e.g. by replacing the maximizing f by $\mathbf{1}_{\{f>0\}}$ and similarly for g. ■

2.3 Counting Lemma

As alluded to earlier, $\delta_\square(W_n, W) \to 0$ if and only if $t(F, W_n) \to t(F, W)$ for every finite graph F. The counting lemma is one half of this equivalence

Lemma 5 (counting lemma)**.** *Let F be a simple graph and let U, W be graphons. Then*

$$|t(F,W) - t(F,U)| \le |E(F)| \cdot \|W - U\|_\square.$$

Note that we can get away with $\|W - U\|_\square$ in the right-hand side, instead of $\delta_\square(W,U)$, since the left-hand side does not depend on the "coupling" between W and U, i.e. it remains fixed if we permute only W (and not U) by some measurable map. So we can just take the minimum over measurable maps permuting W to obtain

$$|t(F,W) - t(F,U)| \le |E(F)| \cdot \delta_\square(W,U).$$

Proof of counting lemma. By the definition of $t(F,W)$, we have

$$|t(F,W)-t(F,U)| = \left| \int_{[0,1]^{V(F)}} \left(\prod_{uv\in E(F)} W(x_u,x_v) - \prod_{uv\in E(F)} U(x_u,x_v) \right) \prod_{v\in V(F)} dx_v \right|.$$

Fix some ordering on $E(F)$, and let u_iv_i denote the ith edge. Then the difference in the interior of the integral can be written as a telescoping sum:

$$\begin{aligned}\prod_i W(x_{u_i},x_{v_i}) - \prod_i U(x_{u_i},x_{v_i}) &= \sum_{j=1}^{|E(F)|} \left(\prod_{i\le j} W(x_{u_i},x_{v_i}) \prod_{i>j} U(x_{u_i},x_{v_i}) - \prod_{i<j} W(x_{u_i},x_{v_i}) \prod_{i\ge j} U(x_{u_i},x_{v_i}) \right) \\ &= \sum_{j=1}^{|E(F)|} \prod_{i<j} W(x_{u_i},x_{v_i}) \prod_{i>j} U(x_u,x_v) \left(W(x_{u_j},x_{v_j}) - U(x_{u_j},x_{v_j}) \right).\end{aligned}$$

So, using the triangle inequality and Fubini's theorem, we can write

$$|t(F,W)-t(F,U)| \le \sum_{j=1}^{|E(F)|} \left| \int_{[0,1]^2} f_j(x_{u_j}) g_j(x_{v_j}) \left(W(x_{u_j},x_{v_j}) - U(x_{u_j},x_{v_j}) \right) dx_{u_j}\, dx_{v_j} \right|, \tag{2}$$

where

$$f_j(x_{u_j}) = \int_{[0,1]^{V(F)\setminus\{u_j,v_j\}}} \prod_{\substack{v\sim u_j\\ v\neq v_j}} W(x_{u_j},x_v) \prod_{v\in V(F)\setminus\{u_j,v_j\}} dx_v$$

and

$$g_j(x_{v_j}) = \int_{[0,1]^{V(F)\setminus\{u_j,v_j\}}} \prod_{\substack{u\neq u_j\\ v\sim u}} W(x_u,x_v) \prod_{v\in V(F)\setminus\{u_j,v_j\}} dx_v.$$

Since these are averages of products of functions which take values between 0 and 1, we have $0 \le f_j, g_j \le 1$ for each j. So, by Lemma 4, each term of (2) is at most $\|W-U\|_\square$. This finishes the proof. ■

3 Proof of the weak regularity lemma

By Lemma 5 (the counting lemma), Theorem 3 is implied by the following reformulation, which we will prove in this final section.

Theorem 6 (weak regularity lemma III)**.** *For any $k \ge 1$ and any graphon W, there is a stepfunction S with k steps such that*

$$\|W-S\|_\square < \frac{2}{\sqrt{\log k}}.$$

3.1 Reduction to polynomial-size description

The bound given in Theorem 6 implies that we can obtain error δ with $k = \exp(2/\delta^2)$ steps. This is exponentially large, but luckily there is a description of the stepfunction which has polynomial size in terms of the error.

Theorem 7 (weak regularity lemma IV)**.** *For any $m \ge 1$ and any graphon W, there are m pairs of subsets $S_i, T_i \subseteq [0,1]$ and m real numbers $a_i \in [0,1]$ such that*

$$\left\| W - \sum_{i=1}^m a_i \mathbf{1}_{S_i\times T_i} \right\|_\square < \frac{1}{\sqrt{m}}.$$

By symmetrizing the sum by averaging with $\sum_{i=1}^{m} a_i \mathbf{1}_{T_i \times S_i}$ (which does not change the cut norm), we get a stepfunction with at most 2^{2m} pieces. We can thus obtain Theorem 6 from Theorem 7 when $k = 2^{2m}$. Since we need m to be an integer, for k not a power of 4, we can take $m = \lfloor \log_4(k) \rfloor$ in general, and obtain the bound

$$\frac{1}{\sqrt{m}} = \frac{1}{\sqrt{\lfloor \log_4(k) \rfloor}} \leq \frac{2}{\sqrt{\log(k)}}.$$

the last bound holds whenever $k \geq 4$, but actually if $k = 1, 2, 3$ then $\frac{2}{\sqrt{\log(k)}} > 1$ whereas $t(F, W) \in [0, 1]$ for any W, and so for these k the statement (in all versions) is trivial. So it just remains to prove Theorem 7.

3.2 Finishing the proof

We can prove Theorem 7 by successively removing the best approximation $a\mathbf{1}_{S \times T}$ from W. The following lemma tells us how good these approximations can be, in terms of the L^2 norm.

Lemma 8 (L^2 increments)**.** *For every symmetric function $W : [0,1]^2 \to \mathbb{R}$, there are two sets $S, T \subseteq [0,1]$ and a real number $a \in [0, \|W\|_\infty]$ such that*

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

Proof. Let $S, T \subseteq [0,1]$ such that

$$\|W\|_\square = \left| \int_{S \times T} W(x,y)\, dx\, dy \right|,$$

as guaranteed by Lemma 4. Let $a = \frac{1}{\lambda(S \times T)} \int_{S \times T} W(x,y)\, dx\, dy$, where λ is the Lebesgue measure. Then

$$\begin{aligned}
\|W - a\mathbf{1}_{S \times T}\|_2^2 &= \|W\|_2^2 - \int_{S \times T} aW(x,y)\, dx\, dy + a^2 \lambda(S \times T) \\
&= \|W\|_2^2 - \frac{1}{\lambda(S \times T)} \left(\int_{S \times T} W(x,y)\, dx\, dy \right)^2 \\
&\leq \|W\|_2^2 - \|W\|_\square^2,
\end{aligned}$$

since $\lambda(S \times T) \leq 1$, and the integral in the second line is $\pm \|W\|_\square$. ■

With Lemma 8 in hand, we can finish the proof of the weak regularity lemma.

Proof of Theorem 7. Apply Lemma 8 repeatedly to get pairs of sets S_i, T_i, and real numbers a_i, such that

$$\begin{aligned}
\|W - a_1\mathbf{1}_{S_1 \times T_1}\|_2^2 &\leq \|W\|_2^2 - \|W\|_\square^2, \\
\|W - a_1\mathbf{1}_{S_1 \times T_1} - a_2\mathbf{1}_{S_2 \times T_2}\|_2^2 &\leq \|W - a_1\mathbf{1}_{S_1 \times T_1}\|_2^2 - \|W - a_1\mathbf{1}_{S_1 \times T_1}\|_\square^2 \\
&\leq \|W\|_2^2 - \|W\|_\square^2 - \|W - a_1\mathbf{1}_{S_1 \times T_1}\|_\square^2 \\
&\vdots \\
\left\| W - \sum_{i=1}^{k} a_i \mathbf{1}_{S_i \times T_i} \right\|_2^2 &\leq \|W\|_2^2 - \sum_{j=0}^{k-1} \left\| W - \sum_{i=1}^{j} a_i \mathbf{1}_{S_i \times T_i} \right\|_\square^2.
\end{aligned}$$

Now the left-hand side is ≥ 0, and $\|W\|_2^2 \leq 1$, so there is at least one $j \in \{0, \ldots, k-1\}$ for which

$$\left\| W - \sum_{i=1}^{j} a_i \mathbf{1}_{S_i \times T_i} \right\|_\square^2 \leq \frac{1}{k}.$$

This finishes the proof, since we can just set $a_{j+1}, \ldots, a_k = 0$. ■