Lecture 5: November 9

Lecturer: Cosma Shalizi *Scribes: Alden Green*

5.1 Review

We begin by quickly re-defining some concepts from previous lectures. Specify a motif f on k nodes, and a graph g on n nodes.

Definition 5.1 (Homomorphism Density) *The **homomorphism density** of f into g is*

$$t(f,g) = \frac{Hom(f,g)}{n^k} \tag{5.1}$$

where $Hom(f,g)$ counts the number of homomorphisms from f into g.

Recall that for any graph g with corresponding adjacency matrix $\mathbf{a}$, we can define a function $w_g : [0,1] \times [0,1] \to \{0,1\}$ as follows:

$$w_g(u,v) = a_{\lceil nu \rceil \lceil nv \rceil} \tag{5.2}$$

We can think of w_g as taking the adjacency matrix and squashing it into the unit square, and sampling a graph from w_g is the same as picking nodes from g, then connecting them if there is an edge between them in g. This leads to a second (and equivalent) definition for $t(f,g)$,

$$t(f,g) = \int_{[0,1]^k} \prod_{(i,j) \in E(f)} w_g(u_i, u_j) du_1 \dots du_k \tag{5.3}$$

Recall the injective homomorphism density $t_{inj}(f,g) = \mathbb{P}(f \precsim G[k])$ where $G[k]$ is the induced subgraph we get by randomly sampling k nodes from g. The following proposition bounds the difference between t and t_{inj}.

Proposition 5.2

$$|t(f,g) - t_{inj}(f,g)| \leq \frac{k^2}{2n} \tag{5.4}$$

Notice that we can define the homomorphism density for a function $w : [0,1] \times [0,1] \to [0,1]$ in a similar manner as we did for a graph.

$$t(f,w) = \int_{[0,1]^k} \prod_{(i,j)\in E(f)} w(u_i,u_j)du_1...du_k \tag{5.5}$$

and so $t(f,g) = t(f,w_g)$. This allows us to talk about the convergence of graph sequences to a graphon function.

Definition 5.3 (Convergence of Graph Sequences) *A sequence $g_1, g_2, ..., g_m, ...$ converges when $\forall f$, $t(f,g_m)$ converges. The sequence converges to w when $\forall f$,*

$$t(f,g_m) \to t(f,w) \tag{5.6}$$

Now we are ready to define a graphon.

Definition 5.4 (Graphon) *Two functions w_1 and w_2 are **equivalent** iff $\forall f, t(f,w_1) = t(f,w_2)$. An equivalence class of ws is called a **graphon**.*

*For any graphon functon, the w-**random graph** on n nodes, $G_n(w)$, is the Conditionally Independent Dyad model with w as the edge probability function.*

Finally, we come to the main question for today: does $G_n(w) \to w$ in any useful sense? Recall that we might be interested in any of the four main types of convergence for random quantities: convergence in probability, almost sure convergence, convergence in distribution, and convergence in squared mean. We will focus on the first two today.

5.2 Convergence of w-random graphs

Our first theorem will cover convergence in probability of $G_n(w)$ to w.

Theorem 5.5 *For any motif f and any $\epsilon \in (0,1)$,*

$$Pr(|t(f,G_n(w)) - t(f,w)| > \epsilon) \leq 2e^{\frac{-\epsilon^2 n}{4k^2}} \tag{5.7}$$

Proof: To begin with, recall that one way to generate $G_n(w)$ is to say that (i,j) appears in $G_n(w)$ if $\xi_{ij} > w(U_i,U_j)$ where the ξ_{ij} and U_i are independent draws from $U(0,1)$. Let $Z_i = \{U_i \text{ and } \xi_{1,i}, ..., \xi_{i-1i}\}$. Then, changing Z_i changes the value of $t(f,G_n(w))$ by at most $\frac{k}{n}$. We can therefore use the Bounded Difference Inequality and say

$$\mathbb{P}\left(|t(f,G_n(w)) - \mathbb{E}[t(f,G_n(w))]| > \epsilon\right) \leq 2\exp\{-\frac{\epsilon^2 n}{k^2}\} \tag{5.8}$$

We then need to bound $\mathbb{E}[t(f, G_n(w))] - t(f, w)$. To begin with, we use the previously stated fact that for any graph g on n nodes,

$$|t(f, g) - t_{inj}(f, g)| \leq \frac{k^2}{2n} \tag{5.9}$$

Also,

$$\mathbb{E}[t_{inj}(f, G_n(w))] = t(f, w) \tag{5.10}$$

Putting these together yields

$$|\mathbb{E}[t(f, G_n(w))] - t(f, w)| \leq \frac{k^2}{2n} \tag{5.11}$$

Moreover, we have that if $2\exp\{-\frac{\epsilon^2 n}{4k^2}\} \leq 1$ and $\epsilon \in (0, 1)$, then

$$\frac{k^2}{2n} \leq \frac{\epsilon^2}{4\log 2} \leq \frac{\epsilon}{2} \tag{5.12}$$

Thus, we have that $|\mathbb{E}[t(f, G_n(w))] - t(f, w)| \leq \frac{\epsilon}{2}$, and by extension,

$$\mathbb{P}\left(|t(f, G_n(w)) - t(f, w)| > \epsilon\right) \leq \mathbb{P}\left(|t(f, G_n(w)) - \mathbb{E}[t(f, G_n(w))]| > \frac{\epsilon}{2}\right) \tag{5.13}$$

$$\leq 2\exp\{-\frac{\epsilon^2 n}{4k^2}\} \tag{5.14}$$

■

A simple consequence of this theorem is that $\forall f, t(f, G_n(w)) \xrightarrow{p} t(f, w)$. However, we can show something even stronger.

Corollary 5.6 *For each $f, t(f, G_n(w)) \xrightarrow{a.s} t(f, w)$.*

Proof: Since the previous deviation inequality decreases exponentially in n, we have that $\forall \epsilon > 0$

$$\sum_{n=1}^{\infty} \mathbb{P}\left(|t(f, G_n(w)) - t(f, w)| > \epsilon\right) < \infty \tag{5.15}$$

so by the Borel-Cantelli Lemma, we have almost sure convergence. In particular, the Borel-Cantelli Lemma gives us

$$\mathbb{P}\left(\cap_{m=0}^{\infty}\left\{|t(f,G_n(w))-t(f,w)|>2^{-m} f.o.\right\}\right)=1 \tag{5.16}$$

which is equivalent to $\mathbb{P}\left(t(f,G_n(w))\to t(f,w)\right)=1$. ■

This leads us to a strong Law of Large Numbers for graphs.

Theorem 5.7 (LLN for graphs) $G_n(w)\overset{a.s.}{\to} w$

Proof: We want to show that $\mathbb{P}\left(\forall f, t(f,G_n(w))\to t(f,w)\right)=1$. Let $B_k=\{f : |V(f)|=k, t(f,G_n(w))\not\to t(f,w)\}$. Then, $\forall k>0, \mathbb{P}(B_k)=0$, and there are countably many B_k, so $\mathbb{P}(\cup_{k=1}^{\infty}B_k)=0$. ■