Lecture 7 (summary)

In this lecture, we introduce the notions of convergent sequences of graphs and graphons, analytic objects associating with convergent sequences of graphs. Throughout this lecture, we use $v(G)$ to denote the number of vertices of G.

The *density* of a graph H in G is the probability that $v(H)$ randomly chosen vertices of G induce a copy of H and is denoted by $d(H,G)$; if $v(H) > v(G)$, we set $d(H,G) = 0$. The *homomorphism density* of H in G, denoted by $t(H,G)$, is the probability that a random mapping from $V(H)$ to $V(G)$ is a homomorphism. We will also work with an injective version of homomorphism densities, denoted by $t_i(H,G)$, which is the probability that a random *injective* mapping from $V(H)$ to $V(G)$ is a homomorphism; again, if $v(H) > v(G)$, we set $t_i(H,G) = 0$.

We next observe relations between the quantities that we have just defined. Consider graphs H and G and an integer ℓ such that $v(H) \le \ell \le v(G)$, and observe that the following identity holds:

$$d(H,G) = \sum_{G', v(G')=\ell} d(H,G')d(G',G).$$

Likewise, the following identity holds:

$$t_i(H,G) = \sum_{G', v(G')=v(H)} t_i(H,G')d(G',G).$$

Fix $\ell \in \mathbb{N}$ and let $H_1, \dots, H_L$ be all graphs with ℓ vertices, listed in the non-decreasing number of their edges. The above identity implies that the values of $d(H_i,G)$, $i \in [L]$, determine the values of $t_i(H_i,G)$, $i \in [L]$. Consider a matrix $A \in [0,1]^{L \times L}$ where $A_{ij} = t(H_i,H_j)$. Observe that $A_{ii} \neq 0$ for every $i \in [L]$ and $A_{ij} = 0$ for all $L \ge i > j \ge 1$, i.e. A is an upper diagonal matrix with a non-zero diagonal and so A^{-1} exists. Note that $A\vec{d} = \vec{t_i}$ where $\vec{d} = (d(H_1,G), \dots, d(H_L,G))$ and $\vec{t_i} = (t_i(H_1,G), \dots, t_i(H_L,G))$, and so $A^{-1}\vec{t_i} = \vec{d}$, i.e. the values of $t_i(H_i,G)$, $i \in [L]$, determine the values of $d(H_i,G)$, $i \in [L]$.

A sequence of graphs $(G_n)_{n\in\mathbb{N}}$ is *convergent* if $v(G_n)$ tends to infinity and the sequence $d(H,G_n)_{n\in\mathbb{N}}$ converges for every graph H. The latter is equivalent to that the sequence $t_i(H,G_n)$ converges for every graph H, which is equivalent to that the sequence $t(H,G_n)$ converges for every graph H (here, we use that $v(G_n) \to \infty$). Convergent sequences of graphs can be represented by the following analytic object: a *graphon* W is symmetric measurable $[0,1]^2 \to [0,1]$, where *symmetric* stands for the property that $W(x,y) = W(y,x)$ for all $(x,y) \in [0,1]^2$. If W is a graphon, a *W-random graph with n vertices* is a random graph obtained in the following way: choose $x_1, \dots, x_n$ uniformly at random and join the i-th vertex and j-th vertex with probability $W(x_i,x_j)$ independently of the other pairs of vertices. Note that if W is a constant graphon equal to $p \in [0,1]$, then the W-random graph with n vertices is just the Erdős-Rényi random graph $G_{n,p}$.

The *density* of a graph H in a graphon W, denoted by $d(H,W)$ is the probability that a W-random graph with $v(H)$ vertices is H. Observe that

$$d(H,W) = \frac{v(H)!}{|\mathrm{Aut}(H)|}\int_{[0,1]^{V(H)}} \prod_{vw\in E(H)} W(x_v,x_w) \prod_{vw\notin E(H)} 1 - W(x_v,x_w)\,\mathrm{d}x_{V(H)}.$$

We say that a graphon W is a *limit* of a convergent sequence $(G_n)_{n\in\mathbb{N}}$ of graphs if

$$d(H,W) = \lim_{n\to\infty} d(H,G_n)$$

for every graph H. In the analogy to homomorphism densities, we define the *homomorphism density* of a graph H in a graphon W, denoted by $t(H,W)$, as

$$t(H,W) = \int_{[0,1]^{V(H)}} \prod_{vw\in E(H)} W(x_v,x_w)\,\mathrm{d}x_{V(H)}.$$

If a graphon W is a limit of a convergent sequence $(G_n)_{n\in\mathbb{N}}$ of graphs, then it holds that

$$t(H,W) = \lim_{n\to\infty} t(H,G_n) = \lim_{n\to\infty} t_i(H,G_n)$$

for every graph H. A limit graphon of a convergent sequence $(G_n)_{n\in\mathbb{N}}$ of graphs is not unique. If W is a graphon and φ is a measure preserving map $[0,1]\to[0,1]$, then we define W^φ to be the graphon defined as $W^\varphi(x,y) = W(\varphi(x),\varphi(y))$. If W is a limit of a convergent sequence $(G_n)_{n\in\mathbb{N}}$, then W^φ is also a limit of $(G_n)_{n\in\mathbb{N}}$ for any measure preserving map $\varphi:[0,1]\to[0,1]$. It can be shown that if graphons W_1 and W_2 satisfy that $d(H,W_1) = d(H,W_2)$ for every graph H, then there exist measure preserving maps φ_1 and φ_2 such that $W_1^{\varphi_1}(x,y) = W_2^{\varphi_2}(x,y)$ for almost all $(x,y)\in[0,1]^2$.

In one of the future lectures, we show that every convergent sequence of graphs has a limit graphon. We finish this lecture by showing that for every graphon W there exists a convergent sequence of graphs such that W is its limit. Specifically, we show that if G_n is a W-random graph with n vertices, then the sequence $(G_n)_{n\in\mathbb{N}}$ is convergent and its limit is W with probability one. To do so, it is enough to show that for every H, the sequence $d(H,G_n)_{n\in\mathbb{N}}$ converges to $d(H,W)$ with probability one. We will define a sequence of X_k, $k=0,\dots,n$, of random variables on the space $[0,1]^n\times[0,1]^{\binom{n}{2}}$, which we understand to determine the result of the process of generating the W-random graph with n vertices; indeed, if $(x_1,\dots,x_n,y_{11},\dots,y_{1n},\dots,y_{n-1,n})\in[0,1]^n\times[0,1]^{\binom{n}{2}}$, we can think that the resulting graph contains an edge between the i-th vertex and the j-th vertex iff $y_{ij}\le W(x_i,x_j)$. Let X_k for $k=0,\dots,n$ be the expected number of induced copies of H condition on the values of $x_1,\dots,x_k$ and $y_{11},\dots,y_{k-1,k}$. Note that $X_0 = d(H,W)\binom{n}{v(H)}$ and X_k is the random variable equal to the actual number of copies of H in the graph obtained by the above process, which has the same distribution as the number of copies of H in the W-random graph with n vertices. Note that the expected value of X_k, $k\in[n]$, conditioned on $X_{k-1}=z$ is equal to z, i.e. the sequence of random variables $X_0,\dots,X_k$

forms a martingale. Next observe that $|X_k - X_{k-1}| \leq \binom{n}{v(H)-1} \leq n^{v(H)-1}$, which implies using the Azuma-Hoeffding Inequality that the probability that $|X_k - X_0| \geq \varepsilon\binom{n}{v(H)}$, which is equal to the probability that $d(H, G_n)$ and $d(H, W)$ differ by at least ε, is at most $2e^{-\frac{\varepsilon^2\binom{n}{v(H)}^2}{n \cdot n^{2(v(H)-1)}}} = 2e^{-\Theta(n)}$. The Borel-Cantelli Lemma now implies that with probability one, there exists n_0 such that $|d(H, G_n) - d(H, W)| \leq \varepsilon$ for all $n \geq n_0$. It follows that $d(H, G_n)_{n\in\mathbb{N}}$ converges to $d(H, W)$ with probability one.

We leave as an exercise to show the analogous statement in the realm of permutations that for every permuton μ there exists a convergent sequence of permutations such that μ is its limit.

Exercise. *Show that for every permuton μ the sequence of μ-random permutations with increasing sizes converges to μ with probability one.*