Random graphs

Spring 2017/2018

Ágnes Backhausz

Tartalomjegyzék

1. Introduction

Goals:

- overview of the theory of random graphs;
- Erdős–Rényi random graphs, preferential attachment graphs, graphons and random graphs, bounded degree random graph models;
- presenting some of the mathematical tools of the analysis of random graphs (e.g. branching processes, martingales)

2. Erdős–Rényi random graph and connectivity

Based on [12], Chapter 4.

1. Definition (Erdős–Rényi random graph). *Let $n \geq 2$ be an integer and $0 \leq p \leq 1$. The Erdős–Rényi random graph $\mathrm{ER}_n(p)$ is a graph on vertices $[n] = \{1, \ldots, n\}$. As for the edges, for every pair $1 \leq s < t \leq n$ we connect vertices s and t with probability p, independently.*

Notation: $\mathbb{P}_{n,p}(A)$ is the probability of A in an $\mathrm{ER}_n(p)$. Similarly for expectation and variance.

2.1. The size of the largest component in the subcritical case

We denote by $\mathscr{C}(v)$ the connectivity component of a vertex v (i.e. the set of vertices from which there is a path to v), while $\mathscr{C}_{\max}$ is the largest connectivity component (or one of them, is it is not unique).

We will compare a so-called exploration algorithm on $\mathscr{C}(v)$ to a similar algorithm on a branching process, which is the following.

2. Definition (Branching process). *The sequence of random variables $(Z_n)_{n\geq 0}$ is a branching process, if*

$$Z_n = \sum_{i=1}^{Z_{n-1}} X_{n,i}, \qquad Z_0 = 1,$$

where $(X_{n,i})_{n,i\geq 0}$ is a family of independent identically distributed non-negative integer valued random variables. The common distribution of $X_{n,i}$ is called offspring distribution.

The following theorem is related to the behaviour of the largest component in the Erdős–Rényi graph.

3. Proposition (Extinction for branching processes). *Let $(Z_n)_{n\geq 0}$ be a branching process, and η be the probability of extinction:*

$$\eta = \mathbb{P}(\exists n : \; Z_n = 0).$$

Then

(i) $\eta = 1$ *if* $\mathbb{E}(X_{1,1}) < 1$;
(ii) $\eta = 1$ *if* $\mathbb{E}(X_{1,1}) = 1$ *and* $\mathbb{P}(X = 1) < 1$;
(iii) $\eta < 1$ *if* $\mathbb{E}(X_{1,1}) > 1$.

First we deal with the subcritical case, which corresponds to $\mathbb{E}(X_{1,1}) < 1$.

We will need the following function:

$$I_\lambda = \lambda - 1 - \log \lambda.$$

1. Theorem ([12], Theorem 4.4, lower bound). *Assume that for n and p we have $np = \lambda < 1$. Then for every $a > 1/I_\lambda$ there exists $\delta = \delta(a, \lambda) > 0$ such that*

$$\mathbb{P}_{n,p}(|\mathscr{C}_{\max}| \geq a \log n) = O(n^{-\delta}).$$

2. Theorem ([12], Theorem 4.5, upper bound). *Assumed that for n and p we have $np = \lambda < 1$. Then for every $a < 1/I_\lambda$ there exists $\delta = \delta(a, \lambda) > 0$ such that*

$$\mathbb{P}_{n,p}(|\mathscr{C}_{\max}| \leq a \log n) = O(n^{-\delta}).$$

Proof of Theorem 1

Exploration of the connectivity component. In an graph $\mathrm{ER}_n(p)$ we run the following process. Vertices might be active, inactive or neutral. At the beginning, one vertex (denoted by v) is active, every other one is neutral. At each step, we choose an arbitrary active vertex. All its neutral neighbors become active, while the vertex itself becomes inactive. We repeat this, until there are no active vertices left.

Let S_t be the number of acvtive vertices after t steps. We have $S_0 = 1$ (as only v is active in the beginning). Note that the set of inactive vertices at the end is exactly $\mathscr{C}(v)$. Since exactly one vertex becomes inactive at each step, we obtain that

$$\mathscr{C}(v) = \min\{t : S_t = 0\}. \tag{1}$$

Let X_t be the number of vertices becoming active at step t. Then we have

$$S_t = S_{t-1} + X_t - 1. \tag{2}$$

Exploration of the branching process. We consider a branching process (Definition 2) in which the offspring distribution is binomial, more precisely, each $X_{n,i}$ has binomial distribution with parameters n and p. Let

$$\overline{T} = \sum_{n=0}^{\infty} Z_n$$

be the total number of individuals in the process.

We use the same exploration process as in the Erdős–Rényi graph. At the beginning the first individual is active, the others are neutral. At each step, we choose an arbitrary active individual, all its offsprings become active, while the individual itself becomes inactive. We repeat this until the are no active vertices left.

Let $\overline{S}_t$ be the number of active individuals after t steps in the exploration algorithm of the branching process. Since in this case we find all individuals, and exactly one individual becomes inactive at each step, we have

$$\overline{T} = \min\{t : \overline{S}_t = 0\}. \tag{3}$$

On the other hand, let $\overline{X}_t$ be the number of individuals becoming active at step t, that is,

$$\overline{S}_t = \overline{S}_{t-1} + \overline{X}_t - 1. \tag{4}$$

Since $\overline{X}_t$ is the total number of offsprings of an individual, the random variables $\overline{X}_t$ have binomial distribution with parameters n and p.

Comparison of the graph and the branching process. In the exploration algorithm of the Erdős–Rényi graph, X_t, which is the number of vertices becoming active at step t, is the number of neutral neighbors of the actual active vertex. Since each pair is connected independently with probability p, conditionally on the past, X_t has binomial distribution, where the first parameter is the actual number of neutral vertices, the second one is p. The number of neutral vertices is at most n, hence we can couple X_t and $\overline{X}_t$ such that $X_t \le \overline{X}_t$ holds. Furthermore, X_ts are conditionally independent of each other, and $\overline{X}_t$s are independent. Hence we can choose the sequences $(X_t)_{t\ge 1}$ and $(\overline{X}_t)_{t\ge 1}$ such that $X_t \le \overline{X}_t$ holds for all t. By equations (2) and (4), we also have $S_t \le \overline{S}_t$ for all t. This implies

$$\mathbb{P}_{n,p}(S_k > 0\ \forall k \le t-1) \le \mathbb{P}_{n,p}(\overline{S}_k > 0\ \forall k \le t-1).$$

(Here $\mathbb{P}_{n,p}$ refers to the probability in a branching process with $\mathrm{Bin}(n,p)$ offspring distribution.) By (1), the left-hand side means that the exploration process does not stop during the first $t-1$ steps, that is, there are at least t vertices becoming inactive and hence belonging to the connectivity component of v. Similarly, by equation (3), the right-hand side is the probability that there are at least t individuals in the branching process. We conclude that

$$\mathbb{P}_{n,p}(|\mathscr{C}(v)| \geq t) \leq \mathbb{P}_{n,p}(\overline{T} \geq t),$$

which is Theorem 4.2. in [12].

Binomial distribution and exponential inequalities. From now on we examine the branching process with binomial offspring distribution. If the total number of individuals is more than t, then there is at least one active vertex after t steps in the exploration algorithm, that is, $S_t > 0$. Equation (4) implies

$$\overline{S}_t = \overline{X}_1 + \ldots + \overline{X}_t - (t-1),$$

therefore

$$\mathbb{P}_{n,p}(\overline{T} > t) \leq \mathbb{P}_{n,p}(\overline{S}_t > 0) = \mathbb{P}(\overline{X}_1 + \ldots + \overline{X}_t \geq t).$$

The random variables $\overline{X}_1, \ldots, \overline{X}_t$ are independent, and have binomial distribution with parameters n and p. Hence $Y = \overline{X}_1 + \ldots + \overline{X}_t$ also has binomial distribution, with parameters $m = tn$ and p. We give an upper bound on the tail probability by the usual exponential inequality (which is used Azuma–Hoeffding argument):

$$\mathbb{P}(Y \geq t) = \mathbb{P}(e^{uY} \geq e^{ut}) \leq \frac{\mathbb{E}(e^{uY})}{e^{ut}} = \frac{(e^u \cdot p + 1 - p)^m}{e^{ut}}.$$

To make the calculations easier, we use $1 + x \leq e^x$ to get

$$\mathbb{P}(Y \geq t) \leq (1+p(e^u-1))^m e^{-ut} \leq \exp\big(p(e^u-1)\cdot nt - ut\big) = \exp(-t(u-\lambda(e^u-1))).$$

The positive number $u > 0$ is chosen such that the upper bound becomes as small as possible, that is, we minimize $u - \lambda(e^u - 1)$. It is easy to check that we get $e^u = 1/\lambda$. Note that we use $\lambda < 1$ at this point, otherwise this would give $u \leq 0$, for which the inequality does not hold (in this case, $t \leq \mathbb{E}(Y)$).

Putting this together, we have

$$\mathbb{P}_{n,p}(|\mathscr{C}(v)| > t) \leq \mathbb{P}_{n,p}(\overline{T} > t) \leq \exp(-t(\lambda - \log\lambda - 1)) = e^{-tI_\lambda}.$$

Finally, we use the union bound for the largest component, and we substitute $t = a\log n$ to get

$$\begin{aligned}\mathbb{P}_{n,p}(|\mathscr{C}_{\max}| \geq a\log n) &\leq n\mathbb{P}_{n,p}(|\mathscr{C}(v)| > a\log n) = n \cdot e^{-a\log n \cdot I_\lambda} \\ &= n^{1-aI_\lambda} = O(n^{-\delta})\end{aligned}$$

for arbitrary $\delta > 0$, because of the condition $aI_\lambda < 1$. □

Remark. The proof of Theorem 2 can be done by the application of the second moment method on the number of vertices belonging to a component larger than t (and not the number of large components), with appropriate t. For this random variable Z by Chebyshev's inequality we have

$$\mathbb{P}(Z=0) \leq \frac{\mathrm{Var}(Z)}{\mathbb{E}(Z)^2}.$$

2.2. Phase transition of the Erdős–Rényi graph

Supercritical case.

Now we consider the Erdős–Rényi graph model with $\lambda = np > 1$.

Let (Z_n) be a branching process such that all $X_{n,i}$ has Poisson distribution λ. We denote by ζ_λ the survival probability of this process, that is, the probability of the event that $Z_n > 0$ holds for all n. Proposition 3 implies that $\zeta_\lambda = 1 - \eta$ is positive in this case. Hence the following theorem is a statement on the existence of a large component in the supercritical case.

3. Theorem (Theorem 4.8, [12]). *Let $\lambda = np > 1$. Then for every $a \in \left(\frac{1}{2}, 1\right)$ there exists $\delta = \delta(a, \lambda)$ such that*

$$\mathbb{P}_{n,p}\left(\big||\mathscr{C}_{\max}| - \zeta_\lambda \cdot n\big| \geq n^a\right) = O(n^{-\delta}).$$

Key steps of the proof. Let Z be the number of vertices belonging to a component larger than $K \log n$ (with an appropriate constant K). Calculate the expectation of Z, and give an upper bound for its variance (better than in the subcritical case). Calculate the expected number of vertices in connected components of size between $K \log n$ and αn with $\alpha < \zeta_\lambda$, and show that the probability that there exists such a component tends to 0 as goes to infinity. It follows that with high probability the size of the largest component is equal to Z, whose asymptotic behavior has already been described.

4. Theorem (Theorem 4.16, [12]). *Let $\lambda = np > 1$. Then*

$$\frac{|\mathscr{C}_{\max}| - \zeta_\lambda \cdot n}{\sqrt{n}} \xrightarrow{d} X,$$

where X is a normal random variable with mean 0 and variance

$$\sigma_\lambda^2 = \frac{\zeta_\lambda(1-\zeta_\lambda)}{(1-\lambda+\lambda\zeta_\lambda)^2}.$$

Critical case.

5. Theorem (Theorem 5.1, [12]). *Let $\lambda = np = 1 + \vartheta n^{-1/3}$, where $\vartheta \in \mathbb{R}$. There exists a constant $b = b(\vartheta) > 0$ such that for all $\omega > 1$ we have*

$$\mathbb{P}_{n,p}\left(\frac{n^{2/3}}{\omega} \leq |\mathscr{C}_{\max}| \leq \omega n^{2/3}\right) \geq 1 - \frac{b}{\omega}.$$

Transition connectivity.

6. Theorem (Theorem 5.8, [12]). *Suppose that for $\lambda_n = p_n \cdot n$ we have $\lambda_n - \log n \to \infty$. Then*

$$\lim_{n\to\infty} \mathbb{P}_{n,p_n}(\textit{the graph is connected}) = 1.$$

On the other hand, if $\lambda_n - \log n \to -\infty$, then

$$\lim_{n\to\infty} \mathbb{P}_{n,p_n}(\textit{the graph is connected}) = 0.$$

Main idea of the proof. Let Y be the number of isolated vertices. Show that $\mathbb{E}_{n,p}(Y) = ne^{-\lambda}(1 + O(\lambda^2/n))$, give an upper bound on the variance of Y, and use the proposition below.

4. Proposition (Proposition 5.10, [12]). *Let Y be the number of isolated vertices in an Erdős–Rényi graph: $Y = \sum_{i=1}^n \mathbb{I}(|\mathscr{C}(i)| = 1)$. For all $\lambda = np$ and $n \geq 2$, we have*

$$\mathbb{P}_{n,p}(\mathrm{ER}_n(p) \textit{ is connected}) \leq \mathbb{P}_{n,p}(Y = 0).$$

Moreover, if there exists an $a > 1/2$ such that $\lambda \geq a \log n$, then, for $n \to \infty$,

$$\mathbb{P}_{n,p}(\mathrm{ER}_n(p) \textit{ is connected}) = \mathbb{P}_{n,p}(Y = 0) + o(1).$$

3. Stochastic block model

Based on: [1]

The stochastic block model is an inhomogeneous generalization of the Erdős–Rényi random graph model.

5. Definition (Stochastic block model). *Let n, k be positive integers, $p = (p_1, \ldots, p_k)$ be a probability distribution on $[k] = \{1, \ldots, k\}$, and W be a $k \times k$ symmetric matrix with entries in $[0, 1]$.*

The random graph $\mathrm{SBM}(n, p, W)$ is defined as follows. Let $X_1, \ldots, X_n$ be independent random variables with distribution p. Then vertices i and j are connected with probability W_{X_i,X_j} independently for each pairs of vertices.

6. Definition (Symmetric stochastic block model). *The random graph* SSBM *with parameters n, k, a, b is a stochastic block model where p is the uniform distribution on $[k]$, and W takes value a on the diagonal and b otherwise.*

Properties:

- for $a, b > 0$, the random graph $\mathrm{SSBM}(n, k, a\log n/n, b\log n/n)$ is connected with high probability if and only if $(a + (k-1)b)/k > 1$ (that is, in this case, the probability that the graph is connected tends to 1 as n goes to infinity).
- the random graph $\mathrm{SSBM}(n, k, a/n, b/n)$ has a large component if and only if $d = (a + (k-1)b)/k > 1$.
- For $\delta < 1/2$, the neighborhood at depth $r = \delta \log_d n$ of a vertex v in $\mathrm{SSBM}(n, k, a/n, b/n)$ tends in total variation to a branching process of offspring distribution Poisson(d), where $d = (a + (k-1)b)/k$.

3.1. Weak recovery

7. Definition (Agreement). *Let $x, y \in [k]^n$ be two vectors. Their agreement is obtained by maximizing the common components between x and any relabelling of y, that is,*

$$A(x, y) = \max_{\pi \in S_k} \frac{1}{n} \sum_{i=1}^{n} \mathbb{I}(x_i = \pi(y_i)).$$

8. Definition (Weak recovery). *Weak recovery or detection is solved in* $\mathrm{SSBM}(n, k, a, b)$ *if there exists $\varepsilon > 0$ and an algorithm that takes G as an input an outputs $\hat{X}$ such that*

$$\mathbb{P}(A(X, \hat{X}) \geq 1/k + \varepsilon) = 1 - o(1)$$

as $n \to \infty$.

The following is Theorem 8 from [1]. The bound is the so-called Kesten–Stigum bound.

7. Theorem (Mossel, Neeman, Sly, 2015). *We consider the* $\mathrm{SSBM}(n, k, a/n, b/n)$ *symmetric block model. We define the signal-to-noise ratio as follows:*

$$\mathrm{SNR} = \frac{(a-b)^2}{k(a + (k-1)b)}.$$

For $k=2$, weak recovery is not solvable if $\mathrm{SNR} \le 1$*, that is, if* $(a-b)^2 \le 2(a+b)$*.*

Method of the proof: message passing algorithms, in particular, reconstruction on trees. Consider the infinte rooted tree where each vertex has exactly d descendants. We assign a random bit to the root, which is 0 or 1 with probability 1/2. Each vertex passes along to its descendants the bit its gets, but the bit changes on each edge with probability ε independently. Let $X^{(0)}$ be the bit of the root, while $X^{(t)}$ the set of bits at the vertices in the tth generation. Detection (reconstruction) is solvable if $\lim_{t\to\infty} \mathbb{E}(|\mathbb{E}(X^{(0)}|X^{(t)}) - 1/2|) > 0$. This is equivalent to $\lim_{t\to\infty} I(X^{(0)}, X^{(t)}) > 0$, where $I(X,Y) = H(X) + H(Y) - H(X,Y)$ is the mutual information of two (vector-valued) random variables. In this model, detection is solvable if and only if $d(1-2\varepsilon)^2 \ge 1$. (Kesten and Stigum proved in 1966 that this condition is sufficient, and Bleher, Ruiz, Zagrebnov in 1995 and Evans, Kenyon, Peres and Schulmann in 2000 proved that it is necessary.) The stochastic block model locally looks like a Poisson Galton–Watson tree, for which similar results hold. It is then enough to show that if weak recovery is possible, then recovery is possible from the information of the communities of the leaves.

The following are Theorem 10 and Theorem 11 in [1], proving the existence of algorithms for recovery.

8. Theorem (Massoulié, Mossel–Neeman–Sly, 2014). *For $k=2$, weak recovery is efficiently solvable if* $\mathrm{SNR} > 1$*.*

9. Theorem (Abbe–Sandon, 2015). *For $k \ge 4$, weak recovery is information-theoretically solvable for some* SNR *strictly less than* 1*.*

For general $\mathrm{SBM}(n, p, Q/n)$ models, the signal-to-noise ratio is defined as λ_2^2/λ_1, where λ_i is the ith largest eigenvalue of PQ, with P being a diagonal matrix with the same diagonal as Q.

For a signal-to-noise ratio larger than 1, weak recovery can be solved in $O(n \log n)$ steps (Theorem 12 in [1]).

4. Limits of dense random graphs

4.1. Random graphs from a graphon

For this section, see also [17].

The following can be considered as a generalization of the stochastic block model, in the case when W is a step function in the sense that it is a finite linear combination of indicator functions of (disjoint) squares.

9. Definition (W-random graph, [18], 2006). *The W-random graph model $G(n, W)$ produces a graph on vertices $v_1, \ldots, v_n$ by using a symmetric measurable function $W : [0,1]^2 \to [0,1]$. Let $X_1, \ldots, X_n$ be independent random variables with uniform distribution on $[0,1]$. Then connect vertices v_i and v_j independently with probability $W(X_i, X_j)$ for every $1 \leq i < j \leq n$.*

A symmetric measurable function $W : [0,1]^2 \to [0,1]$ is called a graphon.

We will be interested in the limit of a sequence of W-random graphs. Clearly, the limit object will be W, the question is: what does it mean for a graph sequence to be convergent.

10. Definition (Homomorphism density). *Let F and G be finite simple graphs. We say that $\varphi : V(F) \to V(G)$ is a* homomorphism *if it is adjacency-preserving; that is, for every $(v, v') \in E(F)$ we have $(\varphi(v), \varphi(v')) \in E(G)$. We denote by $\hom(F, G)$ the number of homomorphisms from F to G. The* homomorphism density *of F in G is defined as follows:*

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

In other words, $t(F, G)$ is the probability of the following event when $u_1, \ldots, u_m$ are randomly chosen vertices from G (uniformly with replacement): for every edge (v_i, v_j) of F the vertices u_i and u_j are connected in G (the vertex set of F is $v_1, \ldots, v_m$). That is, $t(F, G)$ is the probability that if we map the vertices of F to the vertices of G, edges go to edges, and we get a homomorphism.

11. Definition (Graph convergence, [7], 2008). *A sequence of finite simple graphs (G_n) is convergent if the sequence $\big(t(F, G_n)\big)$ converges for all finite simple graphs F.*

Now we examine whether the sequence of random graphs $\tilde{G}_n = G(n, W)$ can be convergent when W is a step function. Namely, suppose that $W(x, y) = w_{i,j}$ if $x \in I_i$ and $y \in I_j$, where $[0,1] = I_1 \cup I_2 \cup \ldots \cup I_k$ with disjoint intervals $I_1, I_2, \ldots, I_k$.

First we consider the case when F is a triangle. The question is: what is the probability that three randomly chosen vertices of $\tilde{G}_n$ are all connected to each other; this will be the expectation of the homomorphism density $t(F, \tilde{G}_n)$. Instead of randomly chosen vertices, we can take v_1, v_2, v_3 for sake of simplicity. Then by the law of total expectation, we have

$$\begin{aligned}\mathbb{E}(t(F, \tilde{G}_n)) &= \mathbb{P}(v_1, v_2, v_3 \text{ form a triangle}) \\ &= \sum_{p,q,r} w_{p,q} w_{q,r} w_{p,r} \mathbb{P}(X_1 \in I_p, X_2 \in I_q, X_3 \in I_r) \\ &= \sum_{p,q,r} w_{p,q} w_{q,r} w_{p,r} \cdot |I_p| \cdot |I_q| \cdot |I_r| \\ &\int_{[0,1]^3} W(x_1, x_2) W(x_2, x_3) W(x_3, x_1)\, dx_1\, dx_2\, dx_3.\end{aligned}$$

If F is not necessarily a triangle, then the product goes for all edges of F (recall that in Definition 10, the condition is only for the edges of F). This can motivate the following definition of homomorphism density of finite simple graphs in graphons.

12. Definition. *The homomorphism density of a finite simple graph $F = ([m], E(F))$ in a graphon $W : [0,1]^2 \to [0,1]$ is given by*

$$t(F, W) = \int_{[0,1]^m} \prod_{ij \in E(F)} W(x_i, x_j) dx_1 \ldots dx_m.$$

A sequence of finite simple graphs (G_n) converges to a graphon W if for every finite simple graph F the sequence of homomorphism densities $t(F, G_n)$ converges to $t(F, W)$ as $n \to \infty$.

Now we can state one of the fundamental theorems of the theory of dense graph limits, which can be proved by using the martingale convergence theorem.

10. Theorem (Lovász–Szegedy, 2006, [18]). *Let (G_n) be a sequence of finite simple graphs that is convergent. Then there exists a graphon W such that G_n converges to W.*

As for W-random graphs, more is true than the calculation of the expectation shows. To prove this, we will need Azuma's inequality for martingales.

11. Theorem (Azuma inequality for martingales). *Let a $X_0, X_1, X_2, \ldots$ martingale with respect to $\mathcal{F}_0, \mathcal{F}_1, \mathcal{F}_2, \ldots$, i.e. each X_m has finite expectation, X_m is measurable with respect to $\mathcal{F}_m$ and $\mathbb{E}(X_m | \mathcal{F}_{m-1}) = X_{m-1}$ holds* *for every $m \geq 1$. In addition, suppose that $|X_m - X_{m-1}| \leq c_m$ holds for some $c_m > 0$ for every $m \geq 1$. Then for every $t > 0$ and $n \geq 1$ we have*

$$\mathbb{P}(|X_n - X_0| > t) \leq 2\exp\left(-\frac{t^2}{2\sum_{m=1}^{n} c_m^2}\right).$$

12. Theorem (Lovász–Szegedy, 2006, [18]). *For every symmetric measurable function $W : [0,1]^2 \to [0,1]$, the sequence $G(n,W)$ converges to W with probability 1. That is,*

$$\lim_{n\to\infty} t(F, G(n,W)) = t(F,W) = \int_{[0,1]^m} \prod_{ij\in E(F)} W(x_i,x_j)dx_1\dots dx_m$$

holds for every finite simple graph F as $n \to \infty$ almost surely.

Proof for a slightly modified version [18]. In the definition of homomorphism densities, one can consider only the maps that are injective. Since the map is from the vertex set of F to the vertex set of G_n, and the first one is fixed, while the latter tends to infinity, most of the maps are injective anyway. In the sequel we present the proof of the statement for injective homomorphisms. The proof can be completed by using Lemma 2.1 from [18].

Let F be a finite simple graph on vertices $\{1,\dots,k\}$, and fix n (until the very end of the proof). We denote by G_n a $G(n,W)$-random graph. Let H be the set of injective maps from $\{1,\dots,k\}$ to the vertex set of G_n. There are $n(n-1)\dots(n-k+1)$ elements in H. For $\varphi \in H$, let A_φ be the event that φ is a homomorphism from F to G_n. This is the same for all φ (by the symmetry of the random graph model). In addition, as we discussed after Definition 10, the homomorphism density is the probability that a randomly chosen map is a homomorphism. By symmetry, $\mathbb{P}(A_\varphi)$ is the same for all φ, and a calculation similar to the argument after Definition 11 shows that it is equal to the integral $t(F,W)$. Therefore we have

$$t(F,W) = \frac{1}{|H|}\sum_{\varphi\in H} \mathbb{P}(A_\varphi).$$

Given n, for $1 \leq m \leq n$, let G_m^* be the induced subgraph of G_n corresponding to the vertices $v_1,\dots,v_m$. The following sequence becomes a martingale for $1 \leq m \leq n$ with respect to $\mathcal{F}_m = \sigma(G_m^*)$:

$$B_m = \frac{1}{|H|}\sum_{\varphi\in H} \mathbb{P}(A_\varphi|G_m^*).$$

Indeed, we have

$$\mathbb{E}(B_m|\mathcal{F}_{m-1}) = \frac{1}{|H|}\sum_{\varphi\in H} \mathbb{E}\big(\mathbb{P}(A_\varphi|G_m^*)\big|\mathcal{F}_{m-1}\big) = B_{m-1},$$

as the σ-algebra generated by G^*_{m-1} is contained by the σ-algebra generated by G^*_m.

Since $\mathcal{F}_0$ is the trivial σ-algebra, $B_0 = t(F, W)$. On the other hand, the event A_φ is measurable with respect to $\sigma(G_n)$, hence

$$B_n = \frac{1}{|H|} \sum_{\varphi \in H} \mathbb{P}(A_\varphi | G_n) = \frac{1}{|H|} \sum_{\varphi \in H} \mathbb{I}(A_\varphi) = t(F, G_n),$$

by the definition of homomorphism density.

Notice that

$$\begin{aligned}|B_m - B_{m-1}| &= \frac{1}{|H|} \Big| \sum_{\varphi \in H} (\mathbb{P}(A_\varphi | G_m) - \mathbb{P}(A_\varphi | G_{m-1})) \Big| \\ &\le \frac{1}{|H|} \sum_{\varphi \in H} \big| \mathbb{P}(A_\varphi | G_m) - \mathbb{P}(A_\varphi | G_{m-1}) \big|.\end{aligned}$$

The graphs G_{m-1} and G_m differ only in v_m and the edges connected to it. The event A_φ (whether φ is a homomorphism or not) depends on the subgraph corresponding to $v_{\varphi(1)}, \dots, v_{\varphi(k)}$. Hence if m is not in the range of φ, then the conditional expectation of A_φ is the same with respect to G_{m-1} and G_m. The number of maps φ whose range contains k is $k(n-1)(n-2)\dots(n-k+1)$. Therefore

$$|B_m - B_{m-1}| \le \frac{k(n-1)\dots(n-k+1)}{|H|} = \frac{k(n-1)\dots(n-k+1)}{n(n-1)\dots(n-k+1)} = \frac{k}{n}.$$

By the Azuma inequality on martingales with bounded differences (with $c_m = k/m$ for every m, see Theorem 11) we obtain

$$\mathbb{P}(|B_n - B_0| \ge \varepsilon) \le 2 \exp\Big(- \frac{\varepsilon^2}{2k^2} n \Big).$$

This means

$$\mathbb{P}(|t(F, G_n) - t(F, W)| \ge \varepsilon) \le 2 \exp\Big(- \frac{\varepsilon^2}{2k^2} n \Big).$$

For fixed k (which is the number of vertices of F), the right-hand side is finite when we sum it up for n. Hence by the Borel–Cantelli lemma, with probability 1, the event on the left-hand side occurs only finitely many times. Since this holds for every fixed F and ε, we get that $t(F, G_n)$ converges with probability 1 to $t(F, W)$ for every fixed F. $\square$

4.2. Speed of convergence

Given a general convergence theorem, it is natural to ask how fast the convergence is in some appropriate distance. As we will see, for some particular models, more can be proved, but first we summarize some general results on this question.

First we need a notion for distance of graphs, which induces the convergence notion of Definition 11. As in [8], we define this for graphons (based on the cut-norm defined in [13]), and then for finite graphs.

13. Definition (Cut distance, [8], 2008). *The* cut norm *of a graphon W is defined by*

$$\|W\|_\Box = \sup_{S,T\subset[0,1]} \left| \int_{S\times T} W(x,y)dxdy \right|,$$

where the supremum goes over measurable subsets of $[0,1]$.

The cut distance *of two graphons U, W is given by*

$$\delta_\Box(U,W) = \inf_{\phi:[0,1]\to[0,1]} \|U - W^\phi\|_\Box,$$

where the infimum goes over all invertible maps $\phi : [0,1] \to [0,1]$ such that both ϕ and its inverse are measure preserving, and the graphon W^ϕ is defined by $W^\phi(x,y) = W(\phi(x), \phi(y))$.

We assign a graphon W_G to every finite simple graph $G = ([n], E(G))$. For $1 \le k < n$, let J_k be the interval $[(k-1)/n, k/n)$, and let $J_n = [(n-1)/n, 1]$.) Then

$$W_G(x,y) = \begin{cases} 1, & \text{if } x \in J_i, y \in J_j \text{ and } ij \in E(G); \\ 0, & \text{otherwise.} \end{cases}$$

Finally, the cut distance of two finite simple graphs is the following:

$$\delta_\Box(G,G') = \delta_\Box(W_G, W_{G'}).$$

In the sequel, we will often write G instead of W_G when we consider distances of graphs, or a graph and a graphon.

In general, the following is known for the connection between cut distance and convergence.

13. Theorem ([7], Borgs–Chayes–Lovász–T. Sós–Vesztergombi, 2008). *A sequence of finite simple graphs (G_n) is convergent if and only if it is a*

Cauchy sequence in the cut distance, i.e. for every $\varepsilon > 0$ there exists n_0 such that for every $n, m \geq 0$ we have $\delta_\square(G_n, G_m) \leq \varepsilon$.

A sequence of finite simple graphs (G_n) converges to W if and only if $\delta_\square(W_{G_n}, W)$ tends to 0 as $n \to \infty$. Moreover, the vertices of the graphs can be labeled such that $\|W_{G_n} - W\|_\square \to 0$.

14. Corollary (Uniqueness of the limit). *For a sequence of finite simple graphs (G_n), if (G_n) converges to W and also to W^*, then $\delta_\square(W, W^*) = 0$, that is,*

$$\inf_{\phi:[0,1]\to[0,1]} \|W^\phi - W^*\|_\square = 0.$$

We are interested in the cut-distance of a W-random graph on n vertices and W as a random variable. There are some general results on this. In general, without any assumption on the structure of the graphon, the following can be proved.

15. Proposition ([17], 2012, Lemma 10.16). *Let W be a graphon, and $G(n, W)$ is a W-random graph on n vertices. Then the following holds:*

$$\mathbb{P}\left(\delta_\square\big(G(n, W), W\big) \leq \frac{22}{\sqrt{\log n}}\right) \geq 1 - e^{-\frac{n}{2\log n}}.$$

However, one expects that for graphons with simpler structures, the W-random graph on n vertices is closer to the original graphon. Recently, Klopp and Verzelen [15] have proved a result in this direction, by understanding the speed of convergence of graphons with finitely many values. The next theorem shows that faster convergence can be proved if the range of the graphon has much smaller size than the number of vertices. It also implies that the uniform result of Proposition 15 is the best possible.

14. Theorem ([15], 2017+). *Let W be a graphon that has k possible values ($2 \leq k \leq n$). Then for the W-random graph on n vertices we have*

$$\mathbb{E}\big(\delta_\square(G(n, W), W)\big) \leq C\sqrt{\frac{k}{n \log k}},$$

where C is a universal constant (depending neither on n, nor or k).

This is formulated for dense graphs, but in [15], a more general version is proved for sparse random graphs generated from graphons. A related result on L^p-graphons (symmetric L^p-functions on $[0,1]^2$) can be found in [5] and [6].

4.3. Special models

Now we summarize some random graph models defined by Borgs, Chayes, Lovász, T. Sós and Vesztergombi [8, 2011], who also investigated the limits of these sequences.

16. Definition (Growing uniform attachment model). *We start with a single vertex $\{v_0\}$. At step $n \geq 1$, we add a new vertex v_n to the graph, and independently for each $1 \leq i < j \leq n$ we connect v_i and v_j with probability $1/n$ (if they are not connected yet).*

Notice that for fixed i and j the probability that v_i and v_j are not connected to each other after n steps is the following:

$$\frac{j}{j+1} \cdot \frac{j+1}{j+2} \cdot \ldots \cdot \frac{n-1}{n} = \frac{j}{n},$$

which tends to 0 as $n \to \infty$. That is, each pair of vertices gets connected at some point with probability 1.

15. Theorem ([8]). *The sequence of growing uniform attachment converges to the graphon $W(x, y) = 1 - \max(x, y)$ with probability 1 as the number of vertices tends to infinity.*

17. Definition (Growing ranked attachment graph). *We start with a single vertex $\{v_1\}$. At step $n \geq 2$, we add a new vertex v_n to the graph, and connect it to v_i with probability $1 - i/n$ for every $1 \leq i \leq n - 1$ independently. Then every pair of nonadjacent vertices (among $v_1, \ldots, v_n$) is connected independently with probability $2/n$.*

Again, for fixed i and j the probability that v_i and v_j are not connected tends to 0. Now the limit object is different.

16. Theorem ([8], 2011). *The sequence of growing uniform attachment converges to the graphon $W(x, y) = 1 - xy$ with probability 1 as the number of vertices tends to infinity.*

18. Definition (Simplified dense preferential attachment graph). *Given n, the vertices will be $\{v_1, \ldots, v_n\}$. Fix $c > 0$. For $j = 1, 2, \ldots, 2\lceil cn^2 \rceil$, choose one of the vertices according to an n-color Pólya urn process: at the beginning, each vertex has a single ball. Then at each step we choose one of the vertices with probabilities proportional to the number of balls, and* *assign a new ball to the urn of this chosen vertex. We repeat this* $2\lceil nc^2 \rceil$ *times. Finally, we connect the pairs of vertices chosen in steps* $2j-1$ *and* $2j$ *for* $j = 1, \ldots, \lceil cn^2 \rceil$ *if they are different, but only once if a pair appears several times.*

We get a graph which has n vertices and $\lceil cn^2 \rceil$ edges. This means that the edge density is

$$\frac{\lceil cn^2 \rceil}{\binom{n}{2}} \to 2c > 0.$$

This graph model is called "dense" because the edge density converges to a positive number. Without the restriction on loops (i.e. not connecting a vertex to itself) and multiple edges, the limit can also be defined but not in the sense of Definition 11. In that case, instead of cut distance, the so-called jumble distance can be used [16]. As for the current simplified version, the following holds.

17. Theorem ([20], 2012). *The sequence of simplified dense preferential attachment graphs converges to the graphon* $W(x, y) = 1 - \exp(-c \ln x \ln y)$ *with probability* 1 *as the number of vertices tends to infinity.*

The version of the preferential attachment graph where loops and multiple edges are allowed can be reformulated as follows. The vertices are $\{v_1, \ldots, v_n\}$, and there are no edges in the beginning. We add edges one by one for steps $k = 1, \ldots, \lceil nc^2 \rceil$. We denote by $\deg_k(v)$ the degree of vertex v after k steps. Then, at step $k+1$, the probability that vertices u and v are connected with an edge is the following:

$$\frac{2(\deg_k(u) + 1)(\deg_k(v) + 1)}{(n + 2k)(n + 2k + 1)},$$

if $u \neq v$, and the probability that we add a loop to vertex v is given by

$$\frac{(\deg_k(v) + 1)(\deg_k(v) + 2)}{(n + 2k)(n + 2k + 1)}.$$

This shows that vertices of larger degree have more chance to get new edges. This is the preferential attachment property, which will be a common property of the random graph models in the next chapter.

5. Preferential attachment gráfok

A preferential attachment attachment véletlen gráfokat [12] alapján először $m = 1$-re definiáljuk. Ekkor minden lépésben egy új csúcsot és egy új élt veszünk a gráfhoz. Az él egyik végpontja az új csúcs, a régit pedig az éppen aktuális fokszámok egy lineáris függvényével arányosan választjuk. Ennek a modellnek egy speciális esete szerepelt [2]-ben, bár nem pontos definícióval – ezért Barabási–Albert-modellnek is szokták nevezni. Egy lehetséges pontos definíció [4]-ban jelent meg.

19. Definíció (PA-modell.). *Legyen $\delta \geq -1$. Kiindulunk egy v_1 csúcsból és egy hurokélből, és minden lépésben egy új csúcsot és $m = 1$ új élt adunk a gráfhoz, a következő módon. Minden $n \geq 1$-re*

$$\mathbb{P}(v_{n+1} \to v_i | G_n) = \begin{cases} \frac{1+\delta}{n(2+\delta)+(1+\delta)}; & i = n+1; \\ \frac{\deg_i(n)+\delta}{n(2+\delta)+(1+\delta)}; & i = 1, 2, \ldots, n, \end{cases}$$

ahol $\to$ azt jelenti, hogy v_{n+1}-ből v_i-be megy egy él, G_n a gráf n lépés után, $\deg_i(n)$ pedig a v_i szomszédainak száma a G_n gráfban, vagyis n lépés után. Így kapjuk az n csúcsú preferential attachment véletlen gráfot $m = 1$-gyel és $\delta \geq -1$ paraméterrel.

Legyen most $m \geq 1$ egész és $\delta \geq -1$. Ekkor először elkészítünk egy mn csúcsú, δ/m paraméterű preferential attachment véletlen gráfot az előző definíció alapján, minden lépésben egy új csúcsot és egy új élt véve a gráfhoz. Majd összevonjuk a $v_1, \ldots, v_m$ csúcsokat egyetlen csúccsá: a köztük menő élekből hurokélek lesznek, az összevont csúcs pedig ahhoz a többihez csatlakozik éllel, amelyikhez valamelyikük csatlakozott. Párhuzamos élek is kialakulhatnak. Ezt folytatjuk: a $v_{m+1}, \ldots, v_{2m}$ csúcsokat is hasonlóképpen összevonjuk egyetlen csúccsá, és általában a $v_{mj+1}, \ldots, v_{m(j+1)}$ csúcsokat egyetlen csúccsá ($j = 1, 2, \ldots, n$ esetén). Így kapjuk az n csúcsú preferential attachment véletlen gráfot m-mel és $\delta \geq -m$ paraméterrel.

Világos, hogy az élek száma n csúcsú gráf esetén $mn+1$, ami azt jelenti, hogy az élszám a csúcsszám lineáris függvénye, az élsűrűség pedig $(m+1/n)/2n \to 0$. Vagyis a sűrű gráfok limeszelmélete értelmében ezeknek a gráfoknak az azonosan 0 grafon a limesze.

5.1. Martingálok

Legyen $G_1, G_2, G_3, \ldots$ véletlen gráfok sorozata. A valószínűségi mező $(\Omega, \mathcal{A}, \mathbb{P})$. Legyen továbbá $\mathcal{F}_n = \sigma(G_1, \ldots, G_n)$ az első n gráf által generált σ-algebra.

Azt mondjuk, hogy az $X : \Omega \to \mathbb{R}$ valószínűségi változó *mérhető* az $\mathcal{F}_n$ szigma-algebrára nézve, ha $X = f(G_1, \ldots, G_n)$ valamely f függvényre (vegyük észre, hogy a $G_1, \ldots, G_n$ sorozat véges sokféle lehet).

Ha pedig g egy gráfokon értelmezett függvény, akkor a $g(G_{n+1})$-nek az $\mathcal{F}_n$-re vonatkozó feltételes várható értéke, azaz $\mathbb{E}(g(G_{n+1})|\mathcal{F}_n)$ egy olyan valószínűségi változó, mely mérhető $\mathcal{F}_n$-re, azaz $f(G_1, \ldots, G_n)$ alakú, és minden $\omega \in \Omega$-hoz a

$$\mathbb{E}(g(G_{n+1})|G_1 = G_1(\omega), G_2 = G_2(\omega), \ldots, G_n = G_n(\omega))$$

feltételes várható értéket rendeli.

18. Tétel (Teljes várható érték tétele). *Ha X véges várható értékű valószínűségi változó, akkor $\mathbb{E}(\mathbb{E}(X|\mathcal{F}_n)) = \mathbb{E}(X)$. Továbbá, ha $m \leq n$, akkor $\mathbb{E}(\mathbb{E}(X|\mathcal{F}_n)|\mathcal{F}_m) = \mathbb{E}(\mathbb{E}(X|\mathcal{F}_m)|\mathcal{F}_n) = \mathbb{E}(X|\mathcal{F}_m)$.*

Például: legyen $G_1, \ldots, G_n$ preferential attachment véletlen gráfok sorozata $m = 1$-gyel és $\delta = 0$-val. Ekkor ha a $g(G)$ a G-ben található 1 fokú (pontosan egy szomszéddal rendelkező) csúcsok száma, akkor

$$\mathbb{E}(g(G_{n+1})|\mathcal{F}_n) = g(G_n) + \left(1 - \frac{1}{2n+1}\right) - g(G_n) \cdot \frac{1}{2n+1}.$$

Ugyanis: létrejövő elsőfokú csúcs csak az új csúcs lehet, és ez akkor elsőfokú, ha nem saját magához kötjük hozzá, ebből adódik a középső tag. A meglévő elsőfokú csúcsok mindegyike $\frac{1}{2n+1}$ valószínűséggel hozzákötődik az új csúcshoz, ekkor már két szomszédja lesz. Ezért az új csúcshoz csatlakozó csúcsok számának feltételes várható értéke (indikátorokra bontással) $g(G_n) \cdot \frac{1}{2n+1}$, ezt le kell vonni.

20. Definíció (Martingál). *Legyenek $\mathcal{F}_1 \subseteq \mathcal{F}_2 \subseteq F_3 \ldots$ szigma-algebrák. Legyen továbbá $X_1, X_2, \ldots$ valószínűségi változók sorozata úgy, hogy*

- *minden n-re $\mathbb{E}(X_n)$ véges;*
- *minden n-re X_n mérhető az $\mathcal{F}_n$ szigma-algebrára;*
- *minden n-re $\mathbb{E}(X_{n+1}|\mathcal{F}_n) = X_n$.*

Ekkor az $(X_n, \mathcal{F}_n)_{n\geq 1}$ sorozatot martingálnak *nevezzük.*

Az $(X_n, \mathcal{F}_n)_{n\geq 1}$ sorozat szubmartingál, *ha a martingál definíciójában az első két tulajdonság teljesül, továbbá minden n-re $\mathbb{E}(X_{n+1}|\mathcal{F}_n) \geq X_n$.*

Az $(X_n, \mathcal{F}_n)_{n\geq 1}$ sorozat szupermartingál, *ha a martingál definíciójában az első két tulajdonság teljesül, továbbá minden n-re $\mathbb{E}(X_{n+1}|\mathcal{F}_n) \leq X_n$.*

Minden martingál szubmartingál és szupermartingál is.

19. Tétel (Martingál-konvergenciatétel). *Tegyük fel, hogy $(X_n, \mathcal{F}_n)$ szubmartingál, melyre $\mathbb{E}(\sup_n X_n) < \infty$ teljesül (például $X_n \leq a$ minden n-re valamely a valós számra). Ekkor az X_n sorozat 1 valószínűséggel konvergens.*

Tegyük fel, hogy $(X_n, \mathcal{F}_n)$ szupermartingál, melyre $\mathbb{E}(\inf_n X_n) < \infty$ teljesül (például $X_n \geq a$ minden n-re valamely a valós számra). Ekkor az X_n sorozat 1 valószínűséggel konvergens.

5.2. Rögzített csúcs foka

Használni fogjuk a következő jelölést:

$$\Gamma(t) = \int_0^t x^{t-1} e^{-x} dx \qquad (t > 0).$$

Parciális integrálással könnyen látható, hogy $\Gamma(t+1) = t\Gamma(t)$, és $\Gamma(n) = (n-1)!$, ha n pozitív egész.

20. Tétel ([12], Theorem 8.2). *Legyen $G_1, G_2, \ldots$ preferential attachment véletlen gráfok sorozata $m = 1$-gyel és δ-val. Legyen i rögzített, és $\deg_i(n)$ a v_i csúcs (vagyis az i. csúcs) szomszédainak száma (foka) G_n-ben. Ekkor a*

$$\frac{\deg_i(n)}{n^{1/(2+\delta)}}$$

sorozat 1 valószínűséggel konvergál egy ξ_i valószínűségi változóhoz $n \to \infty$ esetén, továbbá

$$\mathbb{E}(\deg_i(n)) = (1+\delta)\frac{\Gamma(n+1)\Gamma(i - \frac{1}{2+\delta})}{\Gamma(n + \frac{1+\delta}{2+\delta})\Gamma(i)} - \delta.$$

Bizonyítás. Legyen most is $\mathcal{F}_n$ a $G_1, \ldots, G_n$ által generált σ-algebra. Számítsuk ki $\deg_i(n+1) + \delta$-nak $\mathcal{F}_n$-re vonatkozó feltételes várható értékét.

$$\begin{aligned}
\mathbb{E}(\deg_i(n+1) + \delta | \mathcal{F}_n) &= \deg_i(n) + \delta + \frac{\deg_i(n) + \delta}{(2+\delta)n + 1 + \delta} = \\
&= (\deg_i(n) + \delta)\frac{(2+\delta)n + 1 + \delta + 1}{(2+\delta)n + 1 + \delta} = \\
&= (\deg_i(n) + \delta)\frac{(2+\delta)(n+1)}{(2+\delta)n + 1 + \delta}.
\end{aligned}$$

Ebből következik, hogy ha az alábbi sorozatot tekintjük:

$$X_n = \frac{\deg_i(n) + \delta}{1+\delta} \prod_{s=i-1}^{n-1} \frac{(2+\delta)s + 1 + \delta}{(2+\delta)(s+1)},$$

akkor $\mathbb{E}(X_{n+1}|\mathcal{F}_n) = X_n$. Az is világos, hogy X_n várható értéke véges (a csúcsok foka n lépés után legfeljebb $mn+1$, illetve, hogy X_n mérhető $\mathcal{F}_n$-re, hiszen a G_n gráfból ki lehet számolni. Vagyis az $(X_n, \mathcal{F}_n)_{n\geq 1}$ sorozat martingál.

Mivel ez egy martingál, és $X_n \geq 0$ minden n-re, a 19. tétel szerint az X_n sorozat 1 valószínűséggel konvergens. Könnyen látható, hogy

$$\prod_{s=i-1}^{n-1} \frac{(2+\delta)s+1+\delta}{(2+\delta)(s+1)} = \frac{\Gamma(n+1)\Gamma(i-\frac{1}{2+\delta})}{\Gamma(n+\frac{1+\delta}{2+\delta})\Gamma(i)} = \frac{n^{1-\frac{1+\delta}{2+\delta}}\left(1+O(\frac{1}{n})\right)\Gamma(i-\frac{1}{2+\delta})}{\Gamma(i)},$$

mivel $\Gamma(n+a)/\Gamma(n) = n^a(1+O(\frac{1}{n}))$ teljesül $n\to\infty$ esetén a Stirling-formula alapján. Vagyis $\deg_i(n)n^{1-\frac{1+\delta}{2+\delta}}$ is 1 valószínűséggel konvergens, és ebből az állítás első része következik.

Az állítás második részét úgy kapjuk, ha észrevesszük, hogy martingálok várható értéke állandó ($\mathbb{E}(X_n) = \mathbb{E}(\mathbb{E}(X_{n+1}|\mathcal{F}_n)) = \mathbb{E}(X_n)$ a teljes várható érték tétele alapján), így $\mathbb{E}(X_n) = \mathbb{E}(X_i)$, továbbá (mivel v_i az i. lépésben jön létre, és fokszáma vagy 1, vagy 2, ha saját magával kötjük össze)

$$\mathbb{E}(\deg_i(i)+\delta) = 1+\delta+\frac{1+\delta}{(2+\delta)(i-1)+1+\delta} = \frac{(1+\delta)(2+\delta)i}{(2+\delta)(i-1)+1+\delta}.$$

□

Ha nagyobb m-et tekintünk, akkor a két modell közötti kapcsolat alapján az előző tételből könnyen következik, hogy

$$\frac{\deg_i(n)}{n^{1/(2+\delta/m)}}$$

lesz 1 valószínűséggel konvergens, amint $n\to\infty$.

5.3. Fokszámeloszlás a preferential attachment modellben

Az alábbi tételből kiderül, hogy a preferential attachment gráfokból álló véletlengráf-sorozatban minden rögzített k-ra a k fokszámú csúcsok aránya sztochasztikusan tart egy rögzített, m-ből és δ-ból a gamma függvény segítségével meghatározható számhoz. Pontosabban, az eltérések maximumának nagyságrendje is meghatározható.

21. Tétel ([12], Theorem 8.3). *Legyen $m \geq 1$ és $\delta > -m$ rögzített, és tekintsük az m, δ paraméterekkel definiált preferential attachment gráfot $n = 1, 2, \ldots$ csúcsszámmal. Ekkor létezik olyan pozitív $C = C(m,\delta)$, hogy*

$$\lim_{n\to\infty} \mathbb{P}\left(\max_k |X_k(n) - x_k| \geq C\sqrt{\frac{\log n}{n}}\right) = 0,$$

ahol $X_k(n)$ a k fokszámú csúcsok aránya az n csúcsú gráfban, és

$$x_k = (2+\delta/m)\frac{\Gamma(k+\delta)\cdot\Gamma(m+2+\delta+\delta/m)}{\Gamma(k+3+\delta+\delta/m)\cdot\Gamma(m+\delta)}.$$

Továbbá, $\sum_{k=1}^{\infty} x_k = 1$.

Ebből következik, hogy $X_k(n) \to x_k$ sztochasztikusan $n \to \infty$ esetén minden $k = m, m+1, \ldots$ esetén. A $\delta = 0$ speciális esetben a határérték az alábbi alakban írható:

$$x_k = \frac{2\Gamma(k)\Gamma(m+2)}{\Gamma(k+3)\Gamma(m)} = \frac{2m(m+1)}{k(k+1)(k+2)} \qquad (k = m, m+1, \ldots).$$

Ez azt jelenti, hogy

$$x_k \cdot k^3 \to 2m(m+1), \text{ azaz } x_k \sim 2m(m+1)\cdot k^{-3} \quad (k \to \infty).$$

(Jelölés: $x_k \sim y_k$, ha $\lim_{k\to\infty} x_k/y_k = 1$.)

Az általános m-re visszatérve:

$$x_k = (2+\delta/m)\frac{\Gamma(k+\delta)\cdot\Gamma(m+2+\delta+\delta/m)}{\Gamma(k+3+\delta+\delta/m)\cdot\Gamma(m+\delta)} = C_{m,\delta}\frac{\Gamma(k+\delta)}{\Gamma(k+3+\delta+\delta/m)}$$
$$\sim C_{m,\delta} k^{-(3+\delta/m)},$$

ahol $C_{m,\delta}$ pozitív, és k-tól nem függ. Vagyis a fokszámeloszlás határértéke polinomiálisan cseng le, aszimptotikusan megegyezik k-nak egy hatványfüggvényével. Ez azt jelenti, hogy preferential attachment modell skálafüggetlen az alábbi értelemben.

21. Definíció (Skálafüggetlenség sztochasztikus értelemben). *Legyen $G_1, G_2, \ldots$ véletlen gráfok sorozata, és $X_k(n)$ a k fokszámú csúcsok aránya G_n-ben. Tegyük fel, hogy valamely $(x_k)_{k=0}^{\infty}$ sorozatra minden $k = 0, 1, 2, \ldots$ esetén $X_k(n) \to x_k$ teljesül $n \to \infty$ esetén sztochasztikusan, továbbá*

- $\sum_{k=0}^{\infty} x_k = 1$*;*
- $x_k \sim Ck^{-\tau}$ *teljesül valamely C, τ pozitív számokra. Ekkor azt mondjuk, hogy a (G_n) sorozat* skálafüggetlen *τ karakterisztikus kitevővel.*

A példában tehát $\tau = 3 + \delta/m > 2$.

Összehasonlításként: a G_n gráfnak n csúcsa és $mn+1$ éle van, vagyis egy csúcs átlagos fokszáma $m + 1/n$. Ha egy Erdős–Rényi-gráfot tekintünk szintén n csúccsal úgy, hogy az átlagos fokszám várható értéke m legyen (de a számolás $m + 1/n$-re is hasonló eredményt adna), azaz $p = m/n$-nel, akkor hasonló tétel bizonyítható ([12], 5.12. tétel) $y_k = e^{-m} \cdot \frac{m^k}{k!}$ értékekkel ($k \geq 0$). Ebben az esetben $y_k \cdot k^\tau$ mindig nullához tart, nem lehet pozitív a limesz. Vagyis a fokszámeloszlás várható értéke nem polinomiálisan, hanem exponenciálisnál is gyorsabban cseng le, amint k-val végtelenhez tartunk.

Valós hálózatok. A [2] cikk egyik fő állítása, hogy az internet gráfjának fokszámeloszlása hatványrendben cseng le, vagyis a gráf skálafüggetlen tulajdonságú. Az azóta eltelt időben ezzel ellentétes megfigyelések is születtek, például [10] összehasonlító vizsgálatokban az exponenciális és lognormális eloszlásokat jobban illeszkedőnek találja az esetek legalább harmadában, illetve felhívja a figyelmet arra, hogy a hatványrendben lecsengő és lognormális eloszlásokat nehéz véges mintán megkülönböztetni. A skálafüggetlenséget állító mérések esetén a τ kitevő általában 1 és 4 közé esik.

5.4. A preferential attachment modell további változatai

5.4.1. Fitness

Az eredeti preferential attachment modellben a csúcsokat csak az különböztette meg egymástól, hogy mely időpontban születtek. Dereich és Mörters [11] a következő modellt vizsgálták. A gráf csúcsai $1, 2, \ldots$. A csúcsokhoz független azonos eloszlású valószínűségi változókat sorsolunk: $Y_1, Y_2, \ldots$. Az $n+1$. csúcs létrejöttekor ezt az új csúcsot minden régihez a többitől függetlenül kötünk hozzá (párhuzamos, illetve hurokélek most nincsenek megengedve), az éleket mindig az újabbak felől a régiek felé irányítva (az első lépésben megszületik az 1 csúcs, de nincsenek még élek). Ha az $n + 1$. lépésben az i csúcs (mely az i. lépésben jött létre) pontosan ${\deg^{\text{in}}}_i(n)$ darab bemenő éllel rendelkezik (azaz d darab nála később létrejövő csúcs csatlakozik hozzá), akkor annak valószínűsége, hogy az új csúcsból indul felé egy él, a következő képlettel adható meg:

$$\frac{Y_i(\deg_i^{\text{in}}(n) + 1)}{\sum_{j=1}^n Y_j(\deg_j^{\text{in}}(n) + 1)},$$

ahol $\deg_j^{\text{in}}(n)$ a j csúcsba bemenő élek száma n lépés után. Vagyis most már nem a fokszám egy lineáris függvényétől, hanem az előre kisorsolt fittnestől is függ, hogy egy csúcs mennyi valószínűséggel kap új éleket.

Megfelelő feltételek mellett (például az Y_i-k szórása nem lehet nulla, és további feltételek is vannak), azt látják be a szerzők, hogy a pontosan $k - 1$ befelé menő éllel rendelkező csúcsok aránya 1 valószínűséggel konvergens, és a határértéke

$$\mathbb{E}\left(\frac{1}{k+\frac{\vartheta^*}{Y_1}}\cdot\frac{\vartheta^*}{Y_1}\cdot\prod_{i=1}^{k-1}\frac{i}{i+\frac{\vartheta^*}{Y_1}}\right),$$

ahol ϑ^* a modellből meghatározható konstans szám. Vagyis ebben az esetben 1 valószínűséggel konvergens az adott fokszámmal rendelkező csúcsok számának aránya.

5.4.2. Fitness és idősödés

A valós hálózatokban az is megtörténhet, hogy azonos számú szomszéddal rendelkező csúcsok közül a később létrejövők nagyobb valószínűséggel kapnak új éleket, vagyis egy adott csúcs súlya az idővel csökken. Garavaglia, van der Hofstad és Woeginger [14] modellje a preferential attachment modell $\delta = 0$ és $m = 1$ esetének egy folytonos idejű változata (az új csúcsok nem diszkrét időpontokban, hanem véletlenszerű időpontokban jönnek létre). Ebben az esetben annak valószínűsége, hogy az új csúcs egy adott régihez csatlakozik, függ a régi csúcs életkorától (ezt t-vel jelöljük), a fokszámától (ezt $\mathrm{deg_i}$-vel jelöljük) és a fitnessétől (ezt Y_i-vel jelöljük), melyet most is minden csúcshoz előre, független azonos eloszlás szerint sorsolunk ki. Ezek ismeretében egy adott régi csúcshoz az

$$Y_i \cdot g(t) \cdot f(\mathrm{deg_i})$$

mennyiséggel arányos valószínűséggel csatlakozik az új csúcs, ahol g megfelelő tulajdonságokkal rendelkező monoton csökkenő és integrálható függvény, f pedig lineáris, monoton növekvő függvény. A szerzők azt látják be, hogy ha az Y_i eloszlása korlátos (azaz $0 \le Y_i \le c$ valamely c konstansra), akkor nem teljesül a skálafüggetlenség: a k fokszámú csúcsok aránya 1 valószínűséggel konvergál egy x_k számhoz, de $x_k \sim Cb^{-k}$ teljesül valamely C, b-vel, vagyis ez a sorozat exponenciális sebességgel cseng le. Ha az Y_i eloszlása nem korlátos, de $\mathbb{E}(e^{tY}) < \infty$ valamely $t > 0$-ra, akkor viszont a skálafüggetlenség igazolható. A bizonyítások a folytonos idejű elágazó folyamatok elméletét használják.

5.4.3. Törlés

Tekintsük az alábbi modellt [21] alapján. Minden lépésben négyféle lehetőség valamelyikét hajtjuk végre, véletlenszerűen választva, hogy a négy közül melyiket.

(1) π_1 valószínűséggel hozzáadunk a gráfhoz egy új csúcsot egy új éllel. Az új él másik végének végpontját véletlenszerűen választjuk ki, a fokszámokkal

arányos valószínűséggel (most hurokél nincs a modellben, csak a régi csúcsok közül választunk).

(2) π_2 valószínűséggel két véletlenszerűen választott régi csúcs között húzunk be új élt. Az élnek mindkét végpontját a fokszámokkal arányos valószínűséggel választjuk.

(3) π_3 valószínűséggel egy véletlenszerűen egyenletesen választott régi élt törlünk.

(4) Különben, azaz $\pi_4 = 1 - \pi_1 - \pi_2 - \pi_3$ valószínűséggel egy egyenletesen választott régi csúcsot törlünk.

A fokszámeloszlás viselkedése a paraméterektől függően különféle lehet.

Legyen p_k a k fokú csúcsok arányának várható értékének limesze, amint a csúcsok száma végtelenhez tart. Legyen

$$\eta = \frac{(\pi_1 + \pi_2 - \pi_3)(\pi_1 - \pi_4)}{\pi_1 + \pi_4}, \quad \alpha_0 = \left(\frac{\pi_1}{2} + \pi_2\right)\frac{1}{\eta}, \quad \alpha_2 = \frac{\pi_3}{\eta} + \frac{\pi_4}{\nu}.$$

Ekkor

$$\frac{\alpha_0}{\alpha_2} = 1 - \frac{\pi_1\left(\frac{\pi_1}{2} + \pi_2 - \pi_3\right)}{\pi_1\pi_4 + \pi_1\pi_3 + \pi_2\pi_4}.$$

Könnyen igazolható, hogy

$$\frac{\alpha_0}{\alpha_1} < 1 \qquad \Leftrightarrow \qquad \pi_1 + 2\pi_2 - 2\pi_3 - \pi_4 < 0.$$

Ez alapján pedig az alábbi három lehetőség alakulhat ki a fokszámeloszlás viselkedése szempontjából, a paraméterek választása szerint.

- Skálafüggetlen viselkedés: ha $\frac{\alpha_0}{\alpha_2} > 1$, azaz $\pi_1 + 2\pi_2 - 2\pi_3 - \pi_4 > 0$, akkor a (p_k) sorozat hatványrendben cseng le

$$\tau = 2 + \frac{\pi_1 + \pi_4}{\pi_1 + 2\pi_2 - 2\pi_3 - \pi_4}$$

  kitevővel, azaz

$$p_k \sim Ck^{-\tau} \qquad (C > 0,\ k \to \infty).$$

- Exponenciális lecsengés: ha $\frac{\alpha_0}{\alpha_2} < 1$, azaz $\pi_1 + 2\pi_2 - 2\pi_3 - \pi_4 < 0$, akkor a (p_k) sorozat exponenciálisan cseng le $\frac{\alpha_0}{\alpha_2}$ rátával, azaz

$$p_k \leq C\left(\frac{\alpha_0}{\alpha_2}\right)^k \qquad (C > 0).$$

- Kritikus viselkedés: ha $\frac{\alpha_0}{\alpha_2} = 1$, azaz $\pi_1 + 2\pi_2 - 2\pi_3 - \pi_4 = 0$, akkor a (p_k) a polinomiális lecsengésű sorozatoknál kisebb, az exponenciális lecsengésűeknél nagyobb.

5.4.4. Másolás

Olyan modelleket is tekinthetünk, ahol minden lépésben egy már létező csúcsot lemásolunk, azaz létrehozunk egy új csúcsot, aminek a szomszédai épp a régi csúcs szomszédai. A lemásolt régi csúcsot választhatjuk véletlenszerűen egyenletesen, vagy fokszámokkal arányosan (például [9]). A létrejövő új éleket bizonyos valószínűséggel rögtön törölhetjük.

A modell paramétereitől függően itt is előállhat skálafüggetlen viselkedés, exponenciális lecsengés illetve a kettő közötti viselkedés.

5.5. Klaszteresedési együttható

22. Definíció. *Legyen G véges egyszerű gráf. Ennek klaszteresedési együtthatója a háromszögek számának háromszorosa elosztva a szomszédos élpárok (cseresznyék) számával.*

Preferential attachment modellekben a klaszteresedési együtthatót is érdemes lehet vizsgáni, ennek a limeszét meghatározni vagy előírni (például [19]).

6. Véletlen reguláris gráfok

A továbbiakban olyan véletlen gráfokkal foglalkozunk, ahol a csúcsok szomszédainak száma adott (vagy valamilyen módon kisorsolt), és ennek ismeretében az élek halmaza az, amit véletlenszerűen választunk a megfelelő lehetőségek közül.

6.1. Konfigurációs modell

Tegyük fel, hogy adott n, illetve $d_1, \dots, d_n$ pozitív egész számok, ez utóbbiak összege páros. Szeretnénk egy olyan véletlen gráfot előállítani, melynek csúcsai $v_1, \dots, v_n$, és a v_j csúcsnak pontosan d_j a fokszáma. Ezt egyrészt megtehetjük úgy, hogy egyenletesen választunk az összes lehetőség közül: véges sok ilyen gráf van, választhatunk egyet úgy, hogy mindet azonos valószínűséggel választjuk. Az is egy lehetőség, hogy az összes, adott fokszámsorozattal rendelkező egyszerű (párhuzamos élek és hurokélek nélküli) gráf közül választunk egyet véletlenszerűen, mindegyiket azonos valószínűséggel választva.

Azonban ezekkel a modellekkel egyrészt nehéz számolni, másrészt nehéz ezt számítógépes szimulációval megvalósítani (ehhez az összes lehetőséget fel kellene sorolni és tárolni). Ezért lehet hasznos az alábbi véletlengráf-modell.

23. Definíció. *Adott n, illetve $d_1, \ldots, d_n$ pozitív egész számok, ez utóbbiak összege páros. Tekintsük a $v_1, v_2, \ldots, v_n$ csúcsokat, és rendeljünk hozzá mindegyikhez annyi darab fél-élt, amennyi a hozzá tartozó d_j (vagyis v_j-hez d_j darabot). Ezután az így kapott $D = \sum_{j=1}^{n} d_j$ fél-élnek válasszuk ki egy véletlenszerűen, egyenletesen választott teljes párosítását. A teljes párosítással összekapcsolt fél-élekből készítsünk éleket. Így egy véletlen gráfot kapunk, ez a konfigurációs modell adott $d_1, \ldots, d_n$ mellett.*

Ez azért könnyebben megvalósítható, mert a véletlen teljes párosítást sorban is elkészíthetjük: kiválaszthatjuk bármelyik fél-élt, ennek véletlenszerűen sorsolunk egy párt, mind a $D-1$ darabot azonos valószínűséggel választva. Majd a megmaradó $D-2$ közül választunk egyet tetszőlegesen, és ennek sorsolunk egy párt, mind a $D-3$ darabot azonos valószínűséggel választva. Ezt ismételjük, amíg el nem fogynak a fél-élek. Könnyen látható, hogy ezzel valóban minden párosítás azonos valószínűséggel állt elő.

A konfigurációs modell szerint előállított gráfban lehetnek párhuzamos élek és hurokélek. Ha azonban a konfigurációs modellt amellett a feltétel mellett tekintjük, hogy egyszerű gráfot kapunk (ez egy pozitív valószínűségű esemény megfelelő fokszámok mellett), akkor a feltételes eloszlás egyenletes lesz, azaz minden egyszerű gráfot azonos valószínűséggel kapunk meg. Ezt továbbfejlesztve lehet szimulációs algoritmust készíteni egyenletesen választott egyszerű gráf előállítására.

6.2. Reguláris gráfok

24. Definíció (Véletlen reguláris gráf). *Egy gráf d-reguláris, ha minden csúcs fokszáma d. Egy n csúcsú véletlen d-reguláris gráf a konfigurációs modellből kapott gráf, ha nd páros, és $d_j = d$ minden $j = 1, 2, \ldots, n$-re.*

Ahogy fent említettük, a konfigurációs modellben, és így a véletlen reguláris gráfokban is előfordulhatnak hurokélek. Az alábbi állítás szerint azonban a hurokéllel rendelkező csúcsok száma sztochasztikusan 0-hoz tart, ha a csúcsok számával végtelenhez tartunk.

22. Állítás. *Legyen d rögzített, S_n pedig a hurokélek száma egy n csúcsú véletlen d-reguláris gráfban. Ekkor*

$$\lim_{n\to\infty} \frac{\mathbb{E}(S_n)}{n} = 0 \quad \Rightarrow \quad \frac{S_n}{n} \to 0 \text{ sztochasztikusan.}$$

Bizonyítás. [12] Ha $1 \le j \le n$ és $1 \le s < t \le d$, akkor legyen $\mathbb{I}_{st,j}$ annak indikátora, hogy a j csúcs s és t fél-éleit összekötöttük (azaz értéke 1, ha ez igaz, 0 különben). Ekkor

$$\mathbb{E}(S_n) = \mathbb{E}\Bigg(\sum_{j=1}^{n} \sum_{1\le s<t\le d} \mathbb{I}_{st,j}\Bigg) = \sum_{j=1}^{n} \sum_{1\le s<t\le d} \mathbb{E}(\mathbb{I}_{st,j}).$$

Az $\mathbb{E}(\mathbb{I}_{st,j})$ várható érték annak valószínűsége, hogy a j csúcs s és t fél-éleit összekötöttük. Mivel az s fél-élt ugyanannyi valószínűséggel kötjük bármelyik másik fél-élhez a teljes párosítás választásánál, és ilyenből $nd-1$ darab van, ez a várható érték $\frac{1}{nd-1}$. A megfelelő s,t párok száma pedig $d(d-1)/2$, a j-től függetlenül. Tehát

$$\mathbb{E}(S_n) = \sum_{j=1}^{n} \sum_{1\le s<t\le d} \mathbb{E}(\mathbb{I}_{st,j}) = n \cdot \frac{d(d-1)}{2} \cdot \frac{1}{nd-1} \le \frac{d}{2}.$$

Ebből következik az állítás első része. Ha pedig tetszőleges $\varepsilon > 0$-ra a Markov-egyenlőtlenséget alkalmazzuk az $S_n/n \ge 0$ valószínűségi változóra:

$$\mathbb{P}\left(\left|\frac{S_n}{n}\right| > \varepsilon\right) = \mathbb{P}\left(\frac{S_n}{n} > \varepsilon\right) \le \frac{\mathbb{E}(S_n)}{n\varepsilon} \le \frac{d}{2n\varepsilon} \to 0 \qquad (n \to \infty). \qquad \square$$

Hasonlóképpen igazolható, hogy a $k = 2$ hosszú körök számának várható értéke legfeljebb egy konstans, és így azon csúcsok száma, melyeken átmegy kettő hosszú kör, sztochasztikusan nullához tart. Sőt hasonló érvelés működik tetszőleges k-ra a k hosszú körök számára, amiből kapjuk az alábbi állítást ($k = 2r$ alapján).

23. Állítás. *Legyen d és r rögzített. Legyen $V_n(r)$ pedig az olyan csúcsok száma egy n csúcsú véletlen d-reguláris gráfban, amelynek r sugarú környezete fa, azaz nincs benne kör (hurokél sem). Ekkor*

$$\lim_{n\to\infty} \frac{\mathbb{E}(V_n(r))}{n} = 1 \quad \Rightarrow \quad \frac{V_n(r)}{n} \to 1 \text{ sztochasztikusan } n \to \infty \text{ esetén.}$$

Átfogalmazva: legyen G_n egy n csúcsú véletlen d-reguláris gráf, és v ennek egy véletlenszerűen, egyenletesen választott csúcsa. Ekkor minden r-re annak valószínűsége, hogy v-nek az r sugarú környezete egy fa, nullához tart, ha a csúcsok számával végtelenhez tartunk.

25. Definíció ([3]). *Egy véges egyszerű gráfban egy v csúcs r sugarú környezete azon csúcsok halmaza, melyek legfeljebb r távolságra vannak v-től.*

Jelölje $B_d(r)$ az olyan véges, egyszerű, összefüggő gráfok halmazát, melyekben minden csúcs foka legfeljebb d, tartalmaznak egy o kijelölt csúcsot *(gyökeret), és a gráf csúcsainak o-tól vett távolságainak maximuma éppen r.*

Az F_1 és F_2 véges, egyszerű, gyökeres (egy kijelölt csúcsot is tartalmazó) gráfok izomorfak, ha van olyan $\varphi : V(F_1) \to V(F_2)$ kölcsönösen egyértelmű leképezés, melyre (v_1, v_2) pontosan akkor vannak összekötve F_1-ben, ha $\varphi(v_1)$ és $\varphi(v_2)$ össze vannak kötve F_2-ben.

26. Definíció. *Legyen (G_n) véges egyszerű gráfok sorozata úgy, hogy minden gráfban minden csúcs fokszáma legfeljebb d (ami rögzített pozitív egész). Azt mondjuk, hogy (G_n) konvergens Benjamini–Schramm-értelemben (lokálisan), ha minden $r \geq 1$-re és minden $F \in B_d(r)$-re annak valószínűsége, hogy G_n-ből véletlenszerűen, egyenletesen választva egy csúcsot, azt gyökérnek tekintve az r sugarú környezete éppen az F gyökeres gráffal izomorf, konvergens $n \to \infty$ esetén.*

Példa. Legyen P_n egy n hosszú út, illetve C_n az n hosszúságú kör. Mindkét gráfsorozat konvergens, limeszük: $\mathbb{Z}$, ahol a szomszédos egész számokat kötjük össze.

Legyen G_n az n oldalhosszú kocka a $\mathbb{Z}^n$ gráfban. Ekkor a (G_n) gráfsorozat konvergens, és limesze $\mathbb{Z}^n$.

24. Állítás. *Legyen $d \geq 3$ rögzített, és G_n egy n csúcsú véletlen d-reguláris gráf minden $n \geq 1$-re. Ekkor a (G_n) gráfsorozat egy valószínűséggel konvergens Benjamini–Schramm-értelemben, limesze a végtelen d-reguláris fa (azaz egy olyan végtelen fa, ahol minden csúcsnak d szomszédja van).*

6.3. Véletlen d-reguláris gráfok sajátértékei

Legyen $G = (V, E)$ egy véges egyszerű gráf, csúcsainak száma n. Tekinthetjük ennek az adjacenciamátrixát: ez egy $n \times n$-es mátrix, melynek az i. sorának j. eleme 1, ha $ij \in E$ (azaz az i. és j. csúcsok össze vannak kötve), és 0 különben.

Az A mátrix sajátértékeit a G gráf sajátértékeinek, sajátvektorait pedig a G gráf sajátvektorainak nevezzük. Azaz ha λ komplex szám, $v \in \mathbb{R}^n \setminus \{0\}$, és $Av = \lambda v$ teljesül, akkor v egy sajátvektor λ sajátértékkel. Mivel az A mátrix elemei csak 0-k vagy 1-ek lehetnek, az Av vektor j. koordinátája a v bizonyos koordinátáinak összege: pontosan azoké, amik a j. csúcs szomszédaihoz tartoznak. Tehát az alábbi definícióhoz jutunk. (Mivel az adjacenciamátrix szimmetrikus, minden sajátértéke valós.)

27. Definíció (Gráfok sajátértékei). *Legyen $G = (V, E)$ véges egyszerű gráf, ahol $V = \{1, 2, \ldots, n\}$. A $v \in \mathbb{R}^n \setminus \{0\}$ vektor a G gráf sajátvektora $\lambda \in \mathbb{R}$ sajátértékkel, ha minden $j = 1, 2, \ldots, n$-re teljesül, hogy*

$$\sum_{i:ij \in E} v_i = \lambda v_j.$$

Vagyis a gráf sajátvektora λ sajátértékkel egy olyan függvény a csúcsokon, melynél minden csúcsra igaz, hogy a szomszédjain összegezve a függvény értékét az eredeti csúcsban felvett érték λ-szorosát kapjuk.

Mivel az adjacenciamátrix valós szimmetrikus mátrix, egy n csúcsú gráfnak mindig legalább n olyan sajátvektora van, melyek nem egymás konstansszorosai. Egy d-reguláris gráfnak $\lambda = d$-vel a csupa 1-esekből álló vektor az egyik sajátvektora. Továbbá, egy d-reguláris gráf esetén minden λ sajátértékre $|\lambda| \leq d$, és $\lambda = -d$ pontosan akkor sajátérték, ha a gráf páros. Vagyis a sajátértékek a gráf szerkezetéről árulnak el információkat. A sajátvektorok a klaszterek megtalálásában lehetnek hasznosak (ahogy ezt a sztochasztikus blokkmodellnél láttuk), illetve a gráf vizualizációjában.

25. Állítás. *Legyen d rögzített, és minden $n \geq 1$-re G_n egy n csúcsú véletlen d-reguláris gráf. Legyen $T_n(x)$ a G_n gráf x-nél kisebb sajátértékeinek aránya. Ekkor minden $x \in [-2\sqrt{d-1}, 2\sqrt{d-1}]$-re 1 valószínűséggel*

$$\lim_{n\to\infty} T_n(x) = \int_{-2\sqrt{d-1}}^{x} \frac{d}{2\pi} \cdot \frac{\sqrt{4(d-1) - y^2}}{d^2 - y^2} dy.$$

Azt az eloszlást, aminek a fenti függvény a sűrűségfüggvénye, Kesten–McKay eloszlásnak nevezik.

Hivatkozások

[1] E. Abbe, Community detection and stochastic block models: recent developments. arXiv:1703.10146 math.PR

[2] A.-L. Barabási and R. Albert, Emergence of scaling in random networks, Science **286** (1999), no. 5439, 509–512.

[3] I. Benjamini and O. Schramm, Recurrence of distributional limits of finite planar graphs, Electron. J. Probab. **6** (2001), no. 23, 13 pp.

[4] B. Bollobás, O. Riordan, J. Spencer, G. Tusnády, The degree sequence of a scale-free random graph process, Random Structures Algorithms **18** (2001), no. 3, 279–290.

[5] C. Borgs, J. T. Chayes, H. Cohn and S. Ganguly, Consistent nonparametric estimation for heavy-tailed sparse graphs. Preprint. ArXiv: 1508.06675

[6] C. Borgs, J. T. Chayes, H. Cohn and Y. Zhao, An L^p-theory of sparse graph convergence I: limits, sparse random graph models, and power law distributions. Preprint. ArXiv: 1401.2906.

[7] C. Borgs, J. T. Chayes, L. Lovász, V. T. Sós and K. Vesztergombi, Convergent sequences of dense graphs. I. Subgraph frequencies, metric properties and testing, Adv. Math. **219** (2008), no. 6, 1801–1851.

[8] C. Borgs, J. Chayes, L. Lovász, V. Sós, K. Vesztergombi, Limits of randomly grown graph sequences. *Eur. J. Combin.* **32(7)** (2011), pp. 985–999.

[9] N. Cohen, J. Jordan and M. Voliotis, Preferential duplication graphs, J. Appl. Probab. **47** (2010), no. 2, 572–585.

[10] A. D. Broido, A. Clauset, Scale-free networks are rare, arXiv preprint, 1801.03400.

[11] S. Dereich and M. Ortgiese, Robust analysis of preferential attachment models with fitness, Combin. Probab. Comput. **23** (2014), no. 3, 386–411.

[12] R. van der Hofstad, *Random graphs and complex networks. Vol. 1*, Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press, Cambridge, 2017.

[13] A. Frieze and R. Kannan, Quick approximation to matrices and applications, Combinatorica **19** (1999), no. 2, 175–220.

[14] A. Garavaglia, R. van der Hofstad and G. Woeginger, The dynamics of power laws: fitness and aging in preferential attachment trees, J. Stat. Phys. **168** (2017), no. 6, 1137–1179.

[15] O. Klopp, N. Verzelen, Optimal graphon estimation in cut distance. Preprint. ArXiv: 1703.05101.

[16] D. Kunszenti-Kovács, L. Lovász, B. Szegedy, Multigraph limits, unbounded kernels, and Banach space decorated graphs Preprint. ArXiv:1406.7846

[17] L. Lovász, *Large networks and graph limits*, American Mathematical Society Colloquium Publications, 60, Amer. Math. Soc., Providence, RI, 2012.

[18] L. Lovász and B. Szegedy, Limits of dense graph sequences, J. Combin. Theory Ser. B **96** (2006), no. 6, 933–957.

[19] L. Ostroumova, A. Ryabchenko and E. Samosvat, Generalized preferential attachment: tunable power-law degree distribution and clustering coefficient, in *Algorithms and models for the web graph*, 185–202, Lecture Notes in Comput. Sci., 8305, Springer, Cham.

[20] B. Ráth, L. Szakács, Multigraph limit of the dense configuration model and the preferential attachment graph, *Acta Math. Hung.* **136** (2012), pp. 196–221.

[21] T. Vallier, Transition of the degree sequence in the random graph model of Cooper, Frieze, and Vera, Stoch. Models **29** (2013), no. 3, 341–352.