Complex Networks exam questions

Mariya Karlashchuk

October 2020

What is friendship paradox and how to explain it? How will friendship paradox manifest itself in a scale-free network with a power law degree distribution?

The friendship paradox was first introduced by Scott L. Feld in 1991. The phenomenon that Feld described was that your friends tend to have more friends than you do, in general. This paradox is based on the fact that people tend to compare their popularity (expressed in their number of friends) to that of their friends. We will use the provided example from the course to illustrate the nature of this paradox.

INSERT FIGURE HERE

First of all, we assume a finite graph $G = ([n], E)$. We can define the average degree of a node i in graph G as

$$E[D_n] = \frac{1}{n} \sum_{i=1}^{n} d_i \tag{1}$$

and the average degree of the neighbour j of a node i as

$$E[D_n^*] = \frac{1}{n} \sum_{i=1}^{n} \left(\frac{1}{d_i} \sum_{(i,j) \in E^*} d_j \right) \tag{2}$$

We will use these notations to rewrite 2 and to show the result of the friendship paradox:

$$E[D^*] = \frac{\sum_{i=1}^{n} \sum_{(i,j) \in E^*} d_j}{n \sum_{i=1}^{n} d_i}$$
$$=$$

What are the definition and interpretation of convergence in distribution and convergence in probability? Find an example related to one of the random graph models that we studied, where: convergence in probability occurs; convergence in distribution occurs but convergence in probability does not.

Definition 1) Convergence in distribution. A sequence X_n of random variables converges in distribution to a limiting random variable X when

$$\lim_{n\to\infty} P(X_n \leq x) = P(X \leq x) = F(x)$$

for all points of continuity of $F(x)$, with $x \in \mathbf{R}$.

Definition 2) Convergence in probability. A sequence X_n of random variables converges in probability to a limiting random variable X when

$$\lim_{n\to\infty} P(|X_n - X| > \epsilon) = 0$$

for all $\epsilon > 0$.

We can also conclude that convergence in probability implies convergence in distribution since

$$P(X_n \leq x) \leq P(X \leq x + \epsilon) + P(|X_n - X| > \epsilon) \tag{3}$$

and

$$P(X \leq x - \epsilon) \leq P(X_n \leq x) + P(|X_n - X| > \epsilon)$$

Combining both equation we achieve

$$P(X \leq x - \epsilon) - P(|X_n - X| > \epsilon) \leq P(X_n \leq x) \leq P(X \leq x + \epsilon) + P(|X_n - X| > \epsilon) \tag{4}$$

Taking the limit $n \to \infty$ for all component of the inequality then gives us

$$P(X \leq x - \epsilon) \leq \lim_{n\to\infty} P(X_n \leq x) \leq P(X \leq x + \epsilon) \tag{5}$$

due to convergence in probability and so we conclude

$$\lim_{n\to\infty} P(X_n \leq x) = P(X \leq x) \tag{6}$$

In this course we have encountered a couple of random graph models that show convergence in probability. In

Explain how to solve Exercise 3.3. Which theoretical results do you use? What implication of this results have we seen in one of the random graphs models?

Exercise 3.3. We assume a probability distribution $(p_k)_{k\geq 0}$ that is given by

$$\begin{cases} p_k = b(1-p)^{k-1} & \text{for } k = 1, 2, \dots \\ p_0 = 1 - b/p & \text{for } k = 0 \end{cases} \tag{7}$$

so that, for $b = p$ the offspring distribution has a geometric distribution with success probability p. Show that the extinction probability η is given by $\eta = 1$ if $\mu = E[X] = b/p^2 \leq 1$, while, with the abbreviation $q = 1 - p$ and for $b/p^2 > 1$, $\eta = (1 - \mu p)/q$.

More generally, let 1 and s_0 be the two solutions to $G_X(s) = s$. Show that $s_0 = (1 - \mu p)/q$.

Solution. First of all, we need to determine $G_X(s)$. That will be

$$\begin{aligned} G_X(s) &= E[s^X] \\ &= \sum_{i=0}^{\infty} p_i s^i \\ &= 1 - \frac{b}{p} + \sum_{i=1}^{\infty} b(1-p)^{i-1} s^i \\ &= 1 - \frac{b}{p} + \frac{bs}{1-qs} \end{aligned}$$

so that

$$G'_X = \frac{b(1-qs) + bqs}{(1-qs)^2}\Big|_{s=1} = \frac{b}{p^2}$$

[chain rule]

Before we finish our proof, we will need Theorem 3.1 to support our claim. This theorem says the following.

Theorem 3.1 For a branching process with iid offspring X, $\eta = 1$ when $E[X] < 1$, while $\eta < 1$ when $E[X] > 1$. Further, $\eta = 1$ when $E[X] = 1$ and $P(X = 1) < 1$. The extinction probability η is the smallest solution in $[0, 1]$ of $\eta = G_X(\eta)$ with $s \mapsto G_X(s)$ the probability generating function of the offspring distribution X, i.e.

$$G_X(s) = E[s^X]$$

Hence, the result that follows is $\eta = 1$ if $\mu = b/p^2 \leq 1$. Now when $b/p^2 > 1$, then $\eta < 1$ is the solution of $\eta = G_X(\eta)$, which will then be

$$\eta = 1 - \frac{b}{p} + \frac{b\eta}{1-q\eta}$$

Explain the random walk perspective to the total progeny of a branching process. Can you give examples of results that have been obtained in the course, using this perspective?

First the definition of the total progeny of a branching process and random walk will be defined.

Definition 3) Total Progeny. The total progeny can be interpreted as the number of total individuals in a branching process and is defined as

$$T = \sum_{n=0}^{\infty} Z_n \tag{8}$$

where Z_n is the number of individuals in the n-th generation.

Definition 4) Random walk. Assume $X_1, X_2, \ldots$ be independent and identically distributed random variables with the same distribution as $X_{1,1}$ (offspring distribution). We can define the sun S_i by the recursion

$$\begin{cases} S_0 = 1 \\ S_i = S_{i-1} + X_i - 1 = X_1 + \ldots + X_i + (i-1) \end{cases} \tag{9}$$

PLAATJE?!?!?!

The idea of considering the random walk in a branching process is that we can apply breadth-first-search in order to determine the number of active nodes S_t at time step t, which is equivalent to counting all the nodes. The algorithm is performed as follows:

1. $t = 0$; Start with one active node
2. $t = 1$; Make the offspring of this node active and the first node inactive
3. $t > 1$; Pick one active node, make its offspring active and make chosen node inactive
4. $t = t_{end}$; All nodes are inactive

We now can apply this principle on the random walk. This will give us the following results for S_i:

$$\begin{cases} S_0 = 1 \\ S_1 = S_0 + X_1 - 1 \\ S_t = S_{t-1} + X_t - 1 \end{cases} \tag{10}$$

and hence BFS is a random walk with increments $X_t - 1$.

We know that the process stops when there are no active nodes, that is when $S_t = 0$. This is equal to the total progeny for $T = t$, since the branching process is extinct. This is due to the fact that all nodes have been made active and after that inactive. That way one has encountered t nodes.

We have seen quite some results that have been acquired due to this principle, namely the following.

Theorem 3.7 (Duality principle for branching processes.)
Let $(p_k)_{k\geq 0}$ and $(p'_k)_{k\geq 0}$ be a conjugate pair of offspring distributions. The branching process with distribution $(p_k)_{k\geq 0}$ conditioned on extinction, has the same distribution as the branching process with offspring distribution $(p'_k)_{k\geq 0}$.
Proof. The proof is based on showing that for the finite history $H = (x_1, \ldots, x_t)$ the probability $P(H = bla) = \Pi_{i=1}^{t} p_{x_i}$ is the same for both branching processes with the before mentioned probability distributions, where the first one is conditioned on extinction.

Another convenient feature of the random-walk perspective for branching processes is that it allows one to study what the probability of extinction is when the family tree has at least a given size.

Theorem 3.8 (Extinction probability with large total progeny.)
For a branching process with iid offspring X having mean $\mu = E[X] > 1$

$$P(k \leq T < \infty) \leq \frac{e^{-Ik}}{1 - e^{-I}}$$

where the exponential rate I is given by

$$I = \sup_{t \leq 0} t - \log\big(E[e^{tX}]\big) > 0$$

This theorem can be reformulated by saying that when the total progeny is large, than then the branching process will survive in high probability. There must be noted that it is not assumed that $\mu < \infty$.

MORE?

What is the expected number and the limiting distribution of the number of k-cycles in the $ER_n\left(\frac{\lambda}{n}\right)$ random graph? Compute the expectation. Provide a conjecture for the limiting distribution and explain the intuition behind it. What methods and theorems will you use to prove this conjecture?

A k-cycle in an ER-graph are k vertices connected in a closed chain. An example of a k-cycle is a triangle (3-cycle) or a square (4-cycle) in a graph. The expected number of k-cycles in an $ER_n\left(\frac{\lambda}{n}\right)$ random graph are

$$E[\#\text{ k-cycles}] = \binom{n}{k}\frac{(k-1)!}{2}\left(\frac{\lambda}{n}\right)^k \tag{11}$$

Proof. Assume that X_k is the random variable denoting the number of k-cycles in $ER_n\left(\frac{\lambda}{n}\right)$. Then

$$DING$$

How large is the size of the largest connected component in the $ER_n\left(\frac{\lambda}{n}\right)$ random graph in the subcritical regime? How is this related to branching processes? Explain the phase transition phenomenon.

First of all, we will provide equivalence relations that we will use for the definition of the component of a graph. Say $i, j \in E_n$, then:

- $i \longleftrightarrow i$
- $i \longleftrightarrow j$ implies $j \longleftrightarrow i$
- $i \longleftrightarrow j$ and $j \longleftrightarrow k$ imply $i \longleftrightarrow k$

Now we can provide the definition of a connected component in a graph.

Definition X) Connected component. The connected component containing node v or a cluster of v is given by

$$\mathcal{C}(v) = \{x \in [n] \mid v \longleftrightarrow x\} \tag{12}$$

The size of the connected component is denoted by $|\mathcal{C}(v)|$, which denotes the number of vertices connected to v by an occupied path. Now we provide a definition for the size of the largest component.

Definition X) Size of largest component. The largest connected component $\mathcal{C}_{max}$ is equal to any cluster $\mathcal{C}(v)$ for which $|\mathcal{C}(v)|$ is maximal, so that

$$|\mathcal{C}_{max}| = \max_{v \in [n]} |\mathcal{C}(v)|$$

We consider a subcritical regime, meaning that $\mu < 1$. Since it is rather difficult to derive the exact size of the largest component in an $ER_n\left(\frac{\lambda}{n}\right)$ graph, we can first derive the upper and lower bound for the largest component size. For this we will make use of Theorem 4.4 and 4.5.

Theorem 4.4 (Upper bound on largest subcritical component) Fix $\lambda < 1$. Then, for every $a > \frac{1}{I_\lambda}$, there exists a $\delta = \delta(a, \lambda) > 0$ such that

$$P_\lambda(|\mathcal{C}_{max}| \geq a \log n) = O(n^{-delta})$$

Theorem 4.5 (Upper bound on largest subcritical component) Fix $\lambda < 1$. Then, for every $a < \frac{1}{I_\lambda}$, there exists a $\delta = \delta(a, \lambda) > 0$ such that

$$P_\lambda(|\mathcal{C}_{max}| \leq a \log n) = O(n^{-delta})$$

When combining these two theorems we can draw the conclusion that can derive a convergence in probability:

$$\frac{|\mathcal{C}_{max}|}{\log n} \xrightarrow{P} \frac{1}{I_\lambda}$$

as n grows very large.

In general we can view an $ER_n\left(\frac{\lambda}{n}\right)$ graph as a branching process with Poiss(λ), where the average number of offspring is denoted as λ. In that case, we encounter two situations for the value of λ:

- $\lambda < 1$; in this case the branching process will become extinct, due to the fact that we will have many small sized components in the ER-graph
- $\lambda > 1$; the branching process will survive forever with a positive probability, we will see that the ER-graph will have one giant component

This phenomenon can be defined as the phase transition phenomenon and it refers to the fact that there is a sharp transition in the largest component.

How large is the size of the largest connected component in the $ER_n\left(\frac{\lambda}{n}\right)$ random graph in a supercritical regime? How is this related to branching processes? Explain the phase transition phenomenon.

In a supercritical regime we have that $\mu > 1$. The main result that can be derived for this case is the law of large numbers for the size of the maximal component in an $ER-$graph (stated in Theorem 4.8). Write $\zeta_\lambda = 1 - \eta_\lambda$ as the survival probability of a Poisson branching process with mean offspring λ. Note that $\zeta_\lambda > 0$, since we have that $\lambda > 1$. Theorem 4.8 now states that there exists one giant component.

Theorem 4.8 (Law of large numbers for giant component) Fix $\lambda > 1$. Then, for every $\nu \in (\frac{1}{2}, 1)$ there exists a $\delta = \delta(\nu, \lambda) > 0$ such that

$$P_\lambda\Big(\Big||\mathcal{C}_{max}| - \zeta_\lambda n\Big| \geq n^\nu\Big) = O(n^{-\delta})$$

We can interpret the theorem as follows. A vertex has a large component with probability ζ_λ. Therefore, at least in expectation, there are roughly $\zeta_\lambda n$ vertices with large connected components. Theorem 4.8 implies that almost all these vertices are in fact in the same component, which is called the giant component (since the order of the probability is very very small).

USE PICTURE WITH EXAMPLES

Consider a network of size n that has a power law degree distribution with exponent τ . What can you say about mean degree, variance of the degree, and the maximal degree? What consequences it has for the random graph models?

Power-law: relationship between two variables when one is proportional to a power of the other. Power laws are related to 80/20 rules (wealth among people, 20 percent of people earn 80 percent of total income).

Denote the random variable X to be the degree of a node in a real life network. We can derive the power law as:

$$P(X = k) = p_k = c \cdot k^{-\tau}$$

where p_k is the fraction of nodes of degree k, c a constant and $\tau \in (2, 4)$. This power law implies that rare event will indeed happen (tries to capture non-normal behaviour, ex. Project X Haren). The power law distribution is also known as the Pareto distribution.

Sometimes it is possible to use a slowly varying function $L(x)$ instead of a constant to describe a certain phenomenon:

Definition X) A function $L(x)$ is slowly varying if for all $t > 0$,

$$\lim_{x\to\infty} \frac{L(tx)}{L(x)} = 1$$

For the mean and variance we need to consider the fact that $\sum_{k=k_0}^{\infty} P(X = k) = 1$, hence $\sum_{k=k_0}^{\infty} p_k = \sum_{k=k_0}^{\infty} c \cdot k^{-\tau} = 1$. Now we will use this in order to derive the mean and the variance. We must use that $\tau > 1$, otherwise convergence will not hold in this case.

Assume X has a Pareto distribution, then

$$E[X] = \sum_{k=k_0}^{\infty} k \cdot p_k = \sum_{k=k_0}^{\infty} k \cdot c \cdot k^{-\tau} = \sum_{k=k_0}^{\infty} c \cdot k^{-(\tau-1)} \tag{13}$$

$$E[X^2] = \sum_{k=k_0}^{\infty} k^2 \cdot p_k = \sum_{k=k_0}^{\infty} k^2 \cdot c \cdot k^{-\tau} = \sum_{k=k_0}^{\infty} c \cdot k^{-(\tau-2)} \tag{14}$$

$$\tag{15}$$

So we see that there also must hold that $\tau > 2$, hence we have derived a tighter lower bound for the power τ. Now we can compute the variance, which will be

$$var(X) = E[X^2] - (E[X])^2 = \sum_{k=k_0}^{\infty} c \cdot k^{-(\tau-2)} - \Big(\sum_{k=k_0}^{\infty} c \cdot k^{-(\tau-1)} \Big)^2$$

In order for $E[X^2] < \infty$ we need to have that $\tau > 3$.

We cannot derive the maximal degree easily in a real life network. Hence we will approach the derivation of the maximal degree heuristically. For the maximal degree we need to consider ordered statistics first.

Definition X) Ordered statistics. Let $X_1, \ldots, X_n$ be independent copies of X. Then

$$X_{(1)} \leq X_{(2)} \leq \ldots \leq X_{(n)}$$

are ordered statistics where $X_{(1)}$ is the smallest and $X_{(n)}$ is the largest.

Assume now as the heuristic argument that $P(X > x_{max}) \approx \frac{1}{n}$. Applying the Pareto distribution (which says that

$$1 - F(X) = P(X > x) = \sum_{k=x+1}^{\infty} c \cdot k^{-\tau} \approx \frac{bla}{bla}$$

holds)

Why is ER random graph not a suitable model for social networks? Give as many reasons as possible, formulated in mathematical terms.

First of all, we need a definition for the Erdös-Renyi graph, in order to determine why it is not suited to model social networks.

For an ER-graph holds:

- We have a known finite set of nodes $V = \{1, \dots, n\} = [n]$
- The graph $G = (V, E)$ is undirected
- The edges are placed with probability $P(ij \in E) = p = \frac{\lambda}{n}$, $\forall i, j \in [n]$ and $i = neqj$
- An edge is present or not independently from other edges (the number of edges has a Binomial distribution)

Now we can explain properly why the ER-graph is not suited to model social networks.

*** The edge probabilities are all equal.**
Since all edges have the same edge probability p, the ER-graph is not suited as a model for a social network; this is e.g. due to preferential attachment. In a social network you are likely to know friends of your friends or popular people instead of a random person who you have not any connection with. Hence, we would like to introduce some type of probability for a random graph that is dependent on the weight of a node (a.k.a. the 'popularity' of a person). Therefore, we would like to consider a generalized random graph (GRG), where

$$p = p_{ij} = \frac{w_i w_j}{l_n + w_i w_j} \tag{16}$$

with $i, j \in [n]$, w_i the weight of a graph and $l_n = \sum_{i=1}^{n} w_i$

*** Degree sequence is close to a Poisson distribution.**
At first sight, this might not seem as a big problem. However, a Poisson distribution has thin tails as we have seen in Theorem 4.4 and 4.5. Also, its moment generating function is finite. When we consider the degree sequence for $n = 10.000$ and $n = 1.000.000$ we will see that applying the log-log scale on these networks will result in a very small edge probability when the degree starts to increase (for $\lambda = 1, 10$). We can conclude that when power-law degree sequences are observed, we cannot use the ER-graph.

*** Clusters in a graph.** We define the clustering coefficient as

$$CC_G = \frac{E[\Delta_G]}{E[W_G]} \tag{17}$$

where

$$\Delta_G = 6 \cdot \text{number of triangles in G} = \sum_{i,j,k \in [n]} \mathbf{1}\{ij, ik, jk \text{ occupied}\} \tag{18}$$

$$\Delta_G = 2 \cdot \text{number of wedges in G} = \sum_{i,j,k \in [n]} \mathbf{1}\{ij, jk \text{ occupied}\} \tag{19}$$

WORK THIS OUT

How is Generalized Random Graph model (GRG) defined and what is the motivation behind this model? What can you say about the average degree of each vertex? How this relates to the original motivation of the GRG?

A generalized random graph model can be written as $GRG(\bar{w})$, where $\bar{w} = (w_1, \ldots, w_n)$ a weight sequence. This model has an edge probability of

$$p = p_{ij} = \frac{w_i w_j}{l_n + w_i w_j} \tag{20}$$

with $i, j \in [n]$, w_i the weight of a graph and $l_n = \sum_{i=1}^{n} w_i$. This means that we can assign weights to all vertices. By doing so, you can have a large variability in the degrees of the vertex, which makes it well-suited for real-life networks. Given that $\bar{w}$ edges are independent, if the weights of nodes are unknown, then the edges are dependent.

First, we define the degree of a node i in the $GRG(\bar{w})$ model:

$$D_i = \sum_{j=1, j\neq i}^{n} \mathbf{1}\{ij \in E_n\} \tag{21}$$

The degree here is Poisson distributed with parameter w_i.

Now we can derive the average degree of a node in the $GRG(\bar{w})$ model:

$$\begin{aligned} E[D_i] &= \sum_{j=1, j\neq i}^{n} 1 \cdot p_{ij} && (22)\\ &= \sum_{j=1, j\neq i}^{n} \frac{w_i w_j}{l_n + w_i w_j} && (23)\\ &= w_i \sum_{j=1, j\neq i}^{n} \frac{w_j}{l_n + w_i w_j} && (24)\\ &= w_i \sum_{j=1, j\neq i}^{n} \frac{w_j}{l_n} \frac{1}{1 + \frac{w_i w_j}{l_n}} && (25)\\ \leq w_i &\sum_{j=1, j\neq i}^{n} \frac{w_j}{l_n} && (26) \end{aligned}$$

As $n \to \infty$ we will see that the average degree will be approximately w_i.

The total number of edges is given in Theorem 6.6 of the book.

Theorem 6.6 (Total number of edges in $GRG(\bar{w})$)
Assume that the regularity conditions (a) and (b) of Theorem 6.4 hold. Then

$$\frac{1}{n} E\left(GRG_n(\mathbf{w})\right) \xrightarrow{P} \frac{1}{2} E(W)$$

where $|E(GRG(\bar{w}))|$ is the expected total number of edges and W the random variable of the vertex weight.

Since the distribution of degrees in $GRG(\bar{w})$ is dependent on the vertex weight, we can say that in probability the degrees and weights are approximately distributed the same. The intuition behind this, is that the number of edges can be expressed using the degrees:

$$2|E(GRG(\bar{w}))| = \sum_{i=1}^{n} D_i = nE(W) + o(n) \tag{27}$$

How is the mixed Poisson distribution defined? What does this mean that the limiting degree distribution in the GRG is mixed Poisson? How this relates to the original motivation of behind the GRG?

What is the motivation behind the regularity conditions (a)-(c) for the GRG and the Configuration Model (CM)? What is the meaning of these conditions? What can you say about these conditions when the weights in GRG or the degrees in CM are independent identically distributed random variables drawn from a power law distribution?

First we will denote the regularity conditions.

Condition 6.4 (Regularity conditions for vertex weights). There exists a distribution function F such that, as $n \to \infty$, the following conditions hold

a. Weak convergence of vertex weight; As $n \to \infty$

$$W_n \xrightarrow{d} W$$

where W_n and W have distribution functions F_n and F, respectively. Equivalently, for any x for which $x \mapsto F(x)$ is continuous,

$$\lim_{n\to\infty} F_n(x) = F(x)$$

b. Convergence of vertex weight;

$$\lim_{n\to\infty} \mathbb{E}\left[W_n\right] = \mathbb{E}[W]$$

where W_n and W have distribution functions F_n and F, respectively. Further, we assume that $E[W] > 0$.

c. Convergence of second moment vertex weight.

$$\lim_{n\to\infty} \mathbb{E}\left[W_n^2\right] = \mathbb{E}\left[W^2\right]$$

For the $GRG(\bar{w})$ we want to identify the asymptotic degree distribution, i.e. we aim to prove that the proportion of vertices of degree k approaches a limit when we let the size of the network n tend to infinity. The degree distribution can only converge when the vertex weights are sufficiently regular (therefore we are talking about regularity conditions), which imply convergence of the degree distribution in the $GRG(\bar{w})$. So we get the following:

a. Guarantees that the weight of a typical vertex is close to a random variable that is independent of n.

b. Implies that the average weight of the vertices in $GRG(\bar{w})$ converges to the expectation of the limiting weight variable. In turn, this implies that the average degree in $GRG(\bar{w})$ converges to the expectation of the limit random variable of the vertex weights.

c. Ensures the convergence of the second moment of the limiting weight variable.

When the weights in the weight sequence $\bar{w}$ are realisations of iid random variables themselves, the function F_n is also a random distribution function. This is then called the empirical distribution function of the random weights. Moreover, $E(W_n)$ is then to be interpreted as $\frac{1}{n}\sum_{i\in[n]} w_i$, which itself is random. From this we may conclude that

$$P_n(W_n \leq x) \xrightarrow{P} P(W \leq x) = F(x) \tag{28}$$

holds for each point of continuity x and P_n denotes the conditional probability given the (random) degrees $(w_i)_{i\in[n]}$. Similarly we can say that conditions (b) and (c) can also be replaced by the same conclusion regarding convergence in probability.

We can deduce the same reasoning for the Configuration Model. We get the following regularity conditions:

a. Weak convergence of vertex weight; There exists a distribution function F such that

$$D_n \xrightarrow{d} D$$

where D_n and D have distribution functions F_n and F, respectively. Equivalently, for any x

$$\lim_{n\to\infty} F_n(x) = F(x)$$

Further we assume that $F(0) = 0$, i.e. $P(D \geq 1) = 1$

b. Convergence of vertex degrees;

$$\lim_{n\to\infty} \mathbb{E}\left[D_n\right] = \mathbb{E}[D]$$

where D_n and D have distribution functions F_n and F, respectively, from (a).

c. Convergence of second moment vertex degree.

$$\lim_{n\to\infty} \mathbb{E}\left[D_n^2\right] = \mathbb{E}\left[D^2\right]$$

where again D_n and D have distribution function F_n and F from part (a), respectively.

We note that, since the degrees d_i only take values in integers, so does D_n and therefore so must the limiting random variable D. As a result, the limiting distribution function F is constant between integers and makes a jump $P(D = x)$ at $x \in \mathbf{N}$.

When a graph has random degrees, we can change the conditions in the same fashion as for $GRG(\bar{w})$ with a convergence in probability. Moreover, we could also define the $CM(\bar{d})$ in terms of the number of vertices with fixed degrees, namely

$$n_k = \sum_{i\in[n]} \mathbf{1}\{d_i = k\} \tag{29}$$

which denotes the number of vertices with degree k. Then clearly, apart from the vertex labels, the degree sequence is uniquely determined by the sequence $(n_k)_{k\geq 0}$.

NOG WAT TOE TE VOEGEN?

Why a Generalized Random Graph, conditioned on the degrees, is sampled uniformly at random from all simple graphs with these degrees? Compare to a similar result for the Configuration Model.

[Build up to theorem 6.15, p199] First we need to consider when $GRG(\bar{w})$ is conditioned on its degrees. We then get that

$$X = (X_{ij})_{1 \geq i < \geq n} \tag{30}$$
$$X_{ij} = \mathbf{1}\{ij \in E\} \tag{31}$$

The first one has length $n(n-1)$.

Define now that

$$P(X_{ij} = 1) = 1 - P(X_{ij} = 0) = p_{ij} < 1 \tag{32}$$

Using this information, consider the probability $P(X = x)$, then
(just rewrite using the odds ratio and def of pij and qij)

$$\begin{aligned}
P(X = x) &= \prod_{1<i<i<n} p_{ij}^{x_{ij}} q_{ij}^{1-x_{ij}} \\
&= \prod_{1 \leq i < j \leq n} \frac{r_{ij}^{x_{ij}}}{(1+r_{ij})^{x_{ij}}} \cdot \frac{1}{(1+r_{ij})^{1-x_{ij}}} \\
&= \prod_{1<i<j \geq n} \frac{r_{ij}^{r_{ij}}}{1+r_{ij}} \\
&= \prod_{1 \leq i < j \leq n} \frac{1}{1+r_{ij}} \prod_{1 \leq i < j \leq n} r_{ij}^{x_{ij}} \\
&= \prod_{1 \leq i < j \leq n} \frac{1}{1+u_i u_j} \prod_{1 \leq i < j \leq n} (u_i u_j)^{x_{ij}}
\end{aligned}$$

We now see that the left multiplication does not depend on x and is equal to $G^{-1}(u)$

BLABLA

By using Proposition 7.15 and Theorem 6.15, we see that $GRG(\bar{w})$ conditionally on its degrees,and $CM_n(\bar{d})$ with the same degrees conditioned on producing a simple graph, are equal in distribution. This partially explains the popularity of the configuration model: Some results for $GRG(\bar{w})$, and sometimes even the Erdös-Renyi random graph, are more easily proved by proving the result for the configuration model and conditioning on the degree sequence, and using the fact that the degree distribution of the Erdös-Renyi random graph is very close to a sequence of independent Poisson random variables. We will see that a relation between GRG and CM is defined in Theorem 7.18 and is stated as follows

Theorem 7.18 (Relation between $GRG(\bar{w})$ and $CM_n(\bar{d})$) Let $D_i = \sum_{j \in [n]} X_{ij}$ be the degree of vertex i in $GRG(\bar{w})$ and $CM_n(\bar{d})$ and let $\mathbf{D} = (D_i)_{i \in [n]}$. Then

$$P\left(GRG_n(\mathbf{w}) = G \mid \mathbf{D} = \mathbf{d}\right) = P\left(CM_n(\mathbf{d}) = G \mid CM_n(\mathbf{d}) \text{ simple }\right) \tag{33}$$

Why $CM(d)$ for a given degree sequence d, conditioned on the graph being simple, is sampled uniformly at random from all simple graphs with the given degrees?

Under conditions (a)-(c), what is the probability that Configuration Model produces a simple graph? How can you explain this result? For which degree distributions is condition (c) violated, and what are the consequences of this for the probability that the resulting graph is simple?

Assume you have a dataset that contains a list of all edges in a real-life network. The real-life network is a simple graph. Consider a degree sequence of this network. You would like to sample a graph uniformly at random from a set of all graphs with this degree sequence. Discuss advantages and disadvantages of Configuration Model and Generalized Random Graph.

We will consider both models in order.

GRG
In a GRG we extend the ER-graph model to include a higher amount of inhomogeneity. This means that the ER and GRG model are similar, but we have added weights to the nodes of the latter. In a GRG model the edge probability of the edge between vertices i and j for $i \neq j$ is defined as

$$p_{ij} = \frac{w_i w_j}{l_n + w_i w_j} \tag{34}$$

AND CONDITIONS

When we sample a graph uniformly at random from a set of all graphs with this degree sequence, we know that with the GRG model all graphs with the same degree sequence are equally likely. However, it is not possible to set the degree beforehand, since the weights do not correspond one-to-one with the degrees. Hence, there will always be a variability in the degree sequence.

CM
In the CM model the degrees of the vertices are fixed beforehand, unlike for the GRG model, and when leaving the constraint that we have a simple graph those degrees remain the same. Therefore you can sample an exact degree sequence. However, with the erased configuration model, the degrees will change after erasing some eges. This makes the degree sequence slightly different, but we can proof that it is not a lot of difference.

What is the motivation behind the Preferential Attachment Model? Define the model, solve Exercises 8.2-8.4, and explain intuitively why this model will result in a power law degree sequence.

In a PA-model a new node is added to a network and connects to another node with a probability proportional to the degree of the old node (we take the power-law into consideration here). We denote a PA-model as $(PA_t^{m,\delta})_{t\geq 1}$, where we have t vertices and each vertex has m outgoing edges. Moreover, we can define the probability that the new vertex $v_{t+1}^{(1)}$ is connected to an old vertex $v_t^{(1)}$ (for $m = 1$): SEE DEF 8.2.1

Exercise 8.2
Fix $m = 1$. Verify that $D_i(t) \geq 1$ for all i and t with $t \geq i$, so that $D_i(t) + \delta =\geq 0$ for all $\delta \geq -1$.

Since at every time step t a node gets connected to another node or to itself, the degree of node i at time t for $t \geq i$ is always $D_i(t) \geq 1$. Therefore $D_i(t) + \delta =\geq 0$ for all $\delta \geq -1$. This is also necessary for the probabilities: otherwise we won't have any edges between nodes and only self-loops. When we only have self loops, we can see that the degree of every node is 1.

Exercise 8.3
Verify that the probabilities in 8.2.1. sum up to one.

At time $t + 1$ there are $t + 1$ possibilities for a node to connect to either itself or other t nodes. Hence

$$\sum_{i=1}^{t+1} P\left(v_{t+1} \to v_i\right) = \frac{1+\delta}{l(2+\delta)+(1+\delta)} + \sum_{i=1}^{t} \frac{D_i(t)+\delta}{t(2+\delta)+(1+\delta)} \tag{35}$$

Since this is the case for $m = 1$, we can say that the total sum of all degrees is $2t$. Therefore

$$\frac{1+\delta}{t(2+\delta)+(1+\delta)} + \sum_{i=1}^{t} \frac{D_i(t)+\delta}{t(2+\delta)+(1+\delta)} = \frac{1+\delta}{t(2+\delta)+(1+\delta)} + \frac{t(2+\delta)}{t(2+\delta)+(1+\delta)} \tag{36}$$

$$= \frac{t(2+\delta)+(1+\delta)}{t(2+\delta)+(1+\delta)} = 1 \tag{37}$$

Exercise 8.4
Prove that the total degree of $(PA_t^{m,\delta})_{t\geq 1}$ equals to $2mt$.

Since for a PA-model all nodes are connected to themselves or to another node with m edges, we know that the total degree of all nodes together is $2mt$, since every edge is counted twice.

What is the limiting distribution of the degrees in the Preferential Attachment Model? Provide a heuristic mean-field argument, based on the recursive equations for $p_{k,t}$ - the fraction of vertices with degree k at time t.

In the PA-model a new node arrives to a network and connects to another node with a probability that is 'roughly' proportional to the degree of the old node (rich-get-richer) idea.

Define m to be the number of connections of a newly arriving node (out degree). Furthermore we define

$$p_{k,t} = \frac{1}{t} \sum_{i=1}^{t} \mathbf{1} \{D_i(t) = k\} \tag{38}$$

to be the fraction of nodes with degree k at time t (t is also the number of vertices at time t). To use the heuristic argument we will use model B of the PA-model defined in the book. We then get that

$$\sum_{k=m}^{\infty} k p_{k,t} = \frac{1}{t} \cdot \text{total degree} = \frac{2mt}{t} = 2m \tag{39}$$

We can then apply the 'Master equation', which denotes the change in vertices of degree k

SEE NOTES