Complex Networks - Homework 5

Group 1: Iwan Pasveer, Flavia Leotta, Noah Frinking, Dylan Gavron

15th October 2024

1 Homework 5.1

Given that every possible simple graph with L_u links and n vertices is generated with the same probability, we can just compute the amount x of simple graphs with n vertices and L_u links, since then:

$$\mathbb{P}(G) = \frac{1}{x} \tag{1}$$

A full graph has $\binom{n}{2}$ links, while we need to consider only L_u of these. So:

$$x = \binom{\binom{n}{2}}{L_u} = \frac{\left(\frac{n!}{2!(n-2)!}\right)!}{L_u!\left(\frac{n!}{2!(n-2)!} - L_u\right)!} = \frac{\left(\frac{n(n-1)}{2!}\right)!}{L_u!\left(\frac{n(n-1)}{2!} - L_u\right)!} \tag{2}$$

Which gives:

$$\mathbb{P}(G) = \frac{L_u!\left(\frac{n(n-1)}{2!} - L_u\right)!}{\left(\frac{n(n-1)}{2!}\right)!} \tag{3}$$

2 Homework 5.2

Equation 5.8 states that:

$$\mathbb{E}(k_i^{nn}) \approx \frac{\mathbb{E}\left(\sum_{j\neq i} g_{ij}k_j\right)}{\mathbb{E}(k_i)} \approx \frac{\sum_j p_{ij}k_j^*}{k_i^*} \tag{4}$$

We can rewrite the numerator of the first approximation as:

$$\mathbb{E}\left(\sum_{j\neq i} g_{ij}k_j\right) = \sum_{j\neq i} \mathbb{E}(g_{ij}k_j) \tag{5}$$

If g_{ij} and k_j were independent, then:

$$\mathbb{E}(g_{ij}k_j) = \mathbb{E}(g_{ij})\mathbb{E}(k_j) \tag{6}$$

but this is not always the case, as the degree k_j of a vertex j is affected by the value of the entries in the adjacency matrix (g_{ij}). Vice versa, in a real world network, hubs are more likely to form connections with other nodes, especially other hubs: the higher the degree k_j, the higher is $\mathbb{P}(g_{ij}=1)$, that's why we cannot assume they're independent.

Instead, we rewrite the expectation from equation (5) as conditional expectation:

$$\mathbb{E}(g_{ij}k_j) = \mathbb{E}(k_j|g_{ij}=1)\cdot\mathbb{P}(g_{ij}=1) + \mathbb{E}(k_j|g_{ij}=0)\cdot\mathbb{P}(g_{ij}=0) \tag{7}$$

and substitute it in equation (4), to obtain:

$$\mathbb{E}(k_i^{nn}) \approx \frac{\mathbb{E}\left(\sum_{j\neq i} g_{ij}k_j\right)}{\mathbb{E}(k_i)} = \frac{\sum_{j\neq i}\mathbb{E}(g_{ij}|k_j)k_j}{\mathbb{E}(k_i} \tag{8}$$

The expected value for the clustering coefficient becomes:

$$\mathbb{E}(C) = \frac{2\sum_{j,k} g_{ij}g_{ik}g_{jk}}{\sum_{j,k} g_{ij}g_{ik}} = \frac{2\sum_{j,k}\mathbb{E}(g_{ij}g_{ik}g_{jk})}{\sum_{j,k}\mathbb{E}(g_{ij}g_{ik})} \tag{9}$$

$$= \frac{2\sum_{j,k}\sum_{k_j,k_k}\mathbb{P}(g_{ij}=1|k_j)\cdot\mathbb{P}(g_{ik}=1|k_k)\cdot\mathbb{P}(g_{jk}=1|k_j,k_k)\cdot\mathbb{P}(k_j)\mathbb{P}(k_k)}{\sum_{j,k}\sum_{k_j,k_k}\mathbb{P}(g_{ij}=1|k_j)\cdot\mathbb{P}(g_{ik}=1|k_k)\cdot\mathbb{P}(k_j)\cdot\mathbb{P}(k_k)} \tag{10}$$

$$= 2\sum_{j,k}\sum_{k_j,k_k}\mathbb{P}(g_{jk}=1|k_j,k_k) \tag{11}$$

3 Homework 5.3

We assume that G^* is a regular network, where $k_i^* = z$ for all the vertices i in it. Since it's a regular graph, then $max\{k_i^*\} = z$. We want to examine condition 5.10, which is $max\{k_i^*\} \leq \sqrt{2L_u^*}$.

In this graph the total number of links is:

$$L_u^* = \frac{1}{2}\sum_{j=1}^{n} = \frac{n\cdot z}{2} \tag{12}$$

So when we assume that $z \neq 0$, the condition is met when:

$$z^2 \leq n\cdot z \longrightarrow z \leq n \tag{13}$$

This holds for any regular network since the degree z of each vertex must always be less than or equal to the total number of vertices n.

Now we use $p_{ij} = \frac{k_i^*k_j^*}{2L_u^*}$ and $k_i = z$ to generate the graph assemble. So:

$$p_{ij} = \frac{z^2}{2\cdot\frac{n\cdot z}{2}} = \frac{z}{n} \tag{14}$$

The likelihood of an edge existing between two vertices i and j in the graph ensemble, depends only on the degree z of the nodes and the total number of nodes n in the network. This means that edges are distributed uniformly across the graph so it behaves the same as an ER graph with $p = \frac{z}{n}$.

We write down the expression for the difference $L_u^* - \langle L_u\rangle$:

$$L_u^* - \langle L_u\rangle = L_u^* - \frac{1}{2}\sum_{i}^{n} = 1\mathbb{E}(k_i) = L_u^* - \frac{1}{2}\sum_{i}^{n} = 1\sum_{j\neq i}\mathbb{P}(g_{ij}=1) \tag{15}$$

Now we consider G a star graph with n vertices, where a central vertex is connected to all the other vertices (it's an hub) and all the other vertices are not directly connected to each other. This means that:

$$k_{hub}^* = n-1 = L_u^* = max\{k_i^*\} \tag{16}$$

The condition at 5.10 imposes that $max\{k_i^*\} \leq \sqrt{2L_u^*}$, so:

$$n-1 \leq \sqrt{2(n-1)} \longrightarrow (n-1)^2 \leq 2(n-1) \longrightarrow (n-3)(n-1) \leq 0 \tag{17}$$

And this holds only for those star graphs that have $1 \leq n \leq 3$ vertices.

If we want to generate the graph ensemble in this case, then we compute the probability of forming a link between vertices as:

$$p_{ij} = \begin{cases} \frac{(n-1)\cdot 1}{1(n-1)} = \frac{1}{2} & \text{if j = central vertex} \\ \frac{1}{2(n-2)} & \text{if j = leaf vertex} \end{cases}$$

This probability won't generate a star graph because the probability of two leaves vertices connecting isn't equal to 0.

4 Homework 5.4

In the Park-Newmann model, the probability of connecting two nodes is given by equation 5.14 :

$$p_{ij} = \frac{x_i x_j}{1 + x_i x_j} \tag{18}$$

where x_i and x_j are parameters characterizing some property of vertex i and j.

The probability of generating a certain graph G is expressed by equation 5.11:

$$\mathbb{P}(G) = \prod_{i<j} p_{ij}^{g_{ij}} (1 - p_{ij})^{1-g_{ij}} \tag{19}$$

where g_{ij} is the adjacency matrix entry that is equal to 1 if the link between i and j is realized.

We know that the degree of vertex i is given by:

$$k_i(G) = \sum_{j \neq i} g_i \tag{20}$$

It is evident that the probability of the realisation of graph G depends only on the link probability p_{ij} between each couple of vertices that, in turn, is linked to parameters x_i and x_j of the vertices themselves. In the Park-Newmann model, by construction, those parameters are optimized to reproduce the degree sequence $k_i(G)$: this means that there's no bias to which nodes are connected to each other, as long as, on average, each node obtains its desired degree $\langle k_i \rangle$.

It follows that $\mathbb{P}(G)$ depends solely on the degree sequence: once that is fixed and respected, it's not important how the links are distributed in the graph. This implies that graphs with the same degree sequence are equiprobable.

In the Chung-Lu model, however, $\mathbb{P}(g_{ij} = 1)$, is given by:

$$p_{ij} = \frac{k_i k_j}{2L_u} \tag{21}$$

where k_i and k_j are the observed degrees of i and j (which are fixed), and L_u is the total number of links in the real graph.

We can use equation 5.11 to express the probability of the realization of graph G but using the probability defined by the Chung-Lu model. It's evident that now $\mathbb{P}(G)$ depends directly on the product between the degrees k_i and k_j, for each pair of vertices i and j in the graph, and so two vertices with high degrees are going to have a higher probability of being connected. Because the arrangement of edges affects the probability, two graphs with the same degree sequences but different organizations of edges may have different probabilites of generation. When looking at two randomly generated graph G and G', the one with higher degrees nodes connected to each other (lets say G), will have a bigger probability of being generated than the other one (G'). There's no equiprobability, meaning that even if $k_i(G) = k_i(G')$, we have $\mathbb{P}(G) > \mathbb{P}(G')$.

5 Homework 5.5

In the sparse model, we assume that $x_i x_j \ll 1$, so equation 5.14 becomes:

$$p_{ij} \approx x_i x_j \tag{22}$$

and equation 5.16:

$$k_i \approx \sum_{j \neq i} x_i x_j = x_i \sum_{j \neq i} x_j \tag{23}$$

To solve for x_i we do:

$$x_i = \frac{k_i}{\sum_{j \neq i} x_j} \tag{24}$$

We know that we can approximate L_u, the total number of links in the graph, as:

$$L_u \approx \frac{1}{2} \sum_{i,j \neq i} p_{ij} \tag{25}$$

And by substituting p_{ij} we obtain:

$$L_u \approx \frac{1}{2} \sum_{i,j \neq i} x_i x_j = \frac{1}{2} \sum_i \sum_{j \neq i} x_i x_j \tag{26}$$

$\sum_{i,j \neq i} x_i x_j$ can be rewritten as:

$$\sum_{i,j \neq i} x_i x_j = \sum_i \sum_j x_i x_j - \sum_{j=i} x_i x_j = \sum_i \sum_j x_i x_j - \sum_i x_i^2 = \left(\sum_i x_i \right) \left(\sum_j x_j \right) - \sum_i x_i^2 = \left(\sum_i x_i \right)^2 - \sum_i x_i^2 \tag{27}$$

Since we assumed a sparse graph, and so that $x_i x_j \ll 1$, this implies that $x_i \ll 1$ for every i. It follows that the $\sum_i x_i$ grows proportionally to the number of nodes in the graph, while x_i^2 remains fixed and very small: even when summing all x_i^2, it will be smaller than $(\sum_i x_i)^2$. We can then remove the second sum from the equation.

We're then left with:

$$L_u \approx \frac{1}{2} \left(\sum_i x_i \right)^2 \tag{28}$$

that can be rewritten as:

$$\sum_i x_i = \sqrt{2L_u} \tag{29}$$

$\sum_{j \neq i} x_j$ is the sum of x_j for all nodes excluding node i, while $\sum_i x_i$ sums x_i for all nodes including node i. Since we assumed a sparse graph in which $x_i \ll 1$ then we can consider x_i negligible in comparison to the sum over all nodes: consequentially, excluding just x_i won't change the sum by a lot, and we can approximate $\sum_{j \neq i} x_j \approx \sum_i x_i$. We substitute it in the equation (24) and obtain:

$$x_i = \frac{k_i}{\sqrt{2L_u}} \tag{30}$$

6 Homework 5.6

From equation 5.22 we have:

$$\mathbb{E}(x_\alpha) \equiv \sum_{G \in \mathcal{G}} x_\alpha(G) \mathbb{P}(G) = x_\alpha \tag{31}$$

From statistical physics can express $\mathbb{P}(G)$ in terms of the partition function $Z(\vec{\theta})$, using the Hamiltonian of graph G (equation 5.23):

$$\mathbb{P}(G) = \frac{e^{-H(G,\vec{\theta})}}{Z(\vec{\theta})} \tag{32}$$

where the partition function, expressed as

$$Z(\vec{\theta}) \equiv \sum_{G \in \mathcal{G}} e^{-H(G,\vec{\theta})} \tag{33}$$

normalizes this probability over all the graphs of the ensemble $\mathcal{G}$. Knowing this, we can rewrite the equation for $\mathbb{E}(x_\alpha)$ as:

$$\mathbb{E}(x_\alpha) \equiv \sum_{G \in \mathcal{G}} x_\alpha(G) \frac{e^{-H(G,\vec{\theta})}}{Z(\vec{\theta})} \tag{34}$$

We can differentiate the partition function over $\vec{\theta}$, to obtain:

$$\frac{\partial Z(\vec{\theta})}{\partial \vec{\theta}_\alpha} = \frac{\partial}{\partial \vec{\theta}_\alpha} \sum_{G \in \mathcal{G}} e^{-H(G,\vec{\theta})} = \sum_{G \in \mathcal{G}} - \left(\frac{\partial H(G,\vec{\theta})}{\partial \vec{\theta}_\alpha} \right) e^{-H(G,\vec{\theta})} \tag{35}$$

From equation 5.24 we know:

$$H(G,\vec{\theta}) \equiv \sum_{\alpha=1}^{m} \vec{\theta}_\alpha x_\alpha(G) \tag{36}$$

and its derivative over $\vec{\theta}$ is thus calculated as:

$$\frac{\partial H(G,\vec{\theta})}{\partial \vec{\theta}_\alpha} = x_\alpha(G) \tag{37}$$

We can substitute it in the derivative of the partition function:

$$\frac{\partial Z(\vec{\theta})}{\partial \vec{\theta}_\alpha} = - \sum_{G \in \mathcal{G}} x_\alpha(G) e^{-H(G,\vec{\theta})} \tag{38}$$

and rewrite the expectation as:

$$\mathbb{E}(x_\alpha) = - \frac{1}{Z(\vec{\theta})} \cdot \frac{\partial Z(\vec{\theta})}{\partial \vec{\theta}_\alpha} \tag{39}$$

Now, we introduce the free energy $\Omega(\vec{\theta}) = -lnZ(\vec{\theta})$ and we differentiate it:

$$\frac{\partial \Omega(\vec{\theta})}{\partial \vec{\theta}_\alpha} = - \frac{1}{Z(\vec{\theta})} \cdot \frac{\partial Z(\vec{\theta})}{\partial \vec{\theta}_\alpha} \tag{40}$$

With it, the expected value for x_α is:

$$\mathbb{E}(x_\alpha) = \frac{1}{Z(\vec{\theta})} \cdot \frac{\partial Z(\vec{\theta})}{\partial \vec{\theta}_\alpha} = \frac{\partial \Omega(\vec{\theta})}{\partial \vec{\theta}_\alpha} \tag{41}$$