Complex Networks - Homework 4

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

8th October 2024

1 Homework 4.1

A graph is called highly clustered when:

$$lim_{n\to\infty}\frac{E(\Delta_{G_n})}{E(W_{G_n})} = C \tag{1}$$

where $\mathbb{E}[\Delta_{G_n}]$ refers to the expected numbers of triangles in a graph, while $\mathbb{E}[W_{G_n}]$ refers to the expected numbers of wedges. The more C is equal to 0, the less clustered it is, resembling more a tree structure. We're going to analyse both terms of the ratio at equation (1) separately.

1.1 $\mathbb{E}[\Delta_{G_n}]$

In the Configuration Model, each vertex of the graph gets a degree randomly assigned, and the edges are formed depending on these degrees. The degree sequence is the sequence of these degrees, which are in this case independent and identically distributed.

If a vertex i has degree d_i, it will have d_i stubs that can be used to connect it to other vertices (or itself): the probability of forming an edge between two vertices i and j is then:

$$\mathbb{P}(i \sim j) = \frac{d_i d_j}{2m-1} \tag{2}$$

where $2m$ is the total number of stubs. In case of large amount of vertices the $2m-1$ can be replaced by $2m$

$$2m = \sum_{i=1}^{n} d_i \tag{3}$$

It follows that if each triangle can be formed between three vertex i, j and k, which degrees are, respectively, d_i, d_j and d_k, the probability of a triangle forming between these vertex then depends on the presence of three edges between them at the same time.

$$\mathbb{P}(i \sim j \cap j \sim k \cap i \sim k) \tag{4}$$

We know that the degrees are i.i.d., and after forming one edge the number of stubs available for a certain vertex decreases by one. So equation (4) can be simplified as:

$$\mathbb{P}(i \sim j \cap j \sim k \cap i \sim k) = \frac{d_i d_j}{2m} \cdot \frac{(d_j-1)d_k}{2m} \cdot \frac{(d_i-1)(d_k-1)}{(2m)} = \frac{(d_i d_j d_k)(d_i-1)(d_j-1)(d_k-1)}{(2m)^3} = \mathbb{P}(I_{ijk}=1) \tag{5}$$

where I_{ijk} is equal to 1 if a triangle between those vertices is formed, and 0 if not. The total number of triangles is then:

$$\Delta_{G_n} = \binom{n}{3}\mathbb{P}[I_{ijk}] \tag{6}$$

where n is the total amount of nodes in the graph and $\binom{n}{3}$ is the total possible trios of vertices. We can condition this probability over having a certain degree sequence $\vec{d}$.

$$\mathbb{E}[\Delta_{G_n}|\vec{d}] = \binom{n}{3}\mathbb{E}\left[I_{ijk}|\vec{d}\right] = \binom{n}{3}\frac{(d_i d_j d_k)(d_i-1)(d_j-1)(d_k-1)}{(2m)^3} \tag{7}$$

Now we rewrite this for every possible degree sequence following a degree distribution $f(d)$:

$$\mathbb{E}[\Delta_{G_n}] = \sum_d \mathbb{E}[\Delta_{G_n}|\vec{(d)}]f(d) = \binom{n}{3}\frac{1}{(2m)^3}\sum_d (d_i d_j d_k)(d_i-1)(d_j-1)(d_k-1)f(d) \tag{8}$$

Which can be rewritten as:

$$\mathbb{E}[\Delta_{G_n}] = \sum_d \mathbb{E}[\Delta_{G_n}|\vec{(d)}]f(d) = \binom{n}{3}\frac{1}{(2m)^3}\sum_d (d_i^2-d_i)(d_j^2-d_j)(d_k^2-d_k)f(d) \tag{9}$$

By knowing that the first and the second moment of a function are denoted respectively as:

$$\mathbb{E}[d] = \sum_i d_i f(d) \tag{10}$$

$$\mathbb{E}[d^2] = \sum_i d_i^2 f(d) \tag{11}$$

Then equation (8) becomes:

$$\mathbb{E}[\Delta_{G_n}] = \binom{n}{3}\frac{1}{(2m)^3}\left[(\mathbb{E}[d_i^2]-\mathbb{E}[d_i])(\mathbb{E}[d_j^2]-\mathbb{E}[d_j])(\mathbb{E}[d_k^2]-\mathbb{E}[d_k])\right] \tag{12}$$

We assumed that the degrees d_i, d_j and d_k were i.i.d. and extracted from the same distribution f(d), so:

$$\mathbb{E}[d_i] = \mathbb{E}[d_j] = \mathbb{E}[d_k] = \mathbb{E}[d] \tag{13}$$

And equation (12) becomes:

$$\mathbb{E}[\Delta_{G_n}] = \binom{n}{3}\frac{\left(\mathbb{E}[d^2]-\mathbb{E}[d]\right)^3}{(2m)^3} \tag{14}$$

1.2 $\mathbb{E}[W_{G_n}]$

Each trio of vertex i, j and k forms a wedge when there is a connection between i and j, and between j and k. So, similarly as before, we can estimate that:

$$\mathbb{E}[W_{G_n}] = \sum_d \mathbb{E}[W_{G_n}|\vec{d}]f(d) = 3\binom{n}{3}\frac{(\mathbb{E}[d^2]-\mathbb{E}[d])^2}{(2m)^2} \tag{15}$$

1.3 Together

For $n \to \infty$ the previous equation for the clustering coefficient (C) becomes:

$$\lim_{n\to\infty}\frac{\mathbb{E}[\Delta_{G_n}]}{\mathbb{E}[W_{G_n}]} = \lim_{n\to\infty}\binom{n}{3}\frac{(\mathbb{E}[d^2]-\mathbb{E}[d])^3}{(2m)^3}\cdot\frac{1}{3\binom{n}{3}}\frac{(2m)^2}{(\mathbb{E}[d^2]-\mathbb{E}[d])^2} = \lim_{n\to\infty}\frac{(\mathbb{E}[d^2]-\mathbb{E}[d])}{6m} \tag{16}$$

We can rewrite m as a function of the average degree:

$$2m = \sum_{i=1}^{n} d_i = n\left\langle d\right\rangle \tag{17}$$

And so

$$\lim_{n\to\infty}\frac{(\mathbb{E}[d^2]-\mathbb{E}[d])}{6m} = \lim_{n\to\infty}\frac{(\mathbb{E}[d^2]-\mathbb{E}[d])}{3n\left\langle d\right\rangle} \to 0 \tag{18}$$

which means that as n increases the clustering coefficient C will tend to 0, so the configuration model with i.i.d. degrees isn't highly clustered and has a tree-like shape.

2 Homework 4.2

Given $i \in \mathbb{N}$ we define a sequence of i.i.d. {0,1}-valued variables $(I_j)_{j=i}^{\infty}$, where $I_j = 1$ means that the new vertex i is attached to an old vertex j in the graph.

If the probability that $I_j = 1$ is:

$$\mathbb{P}(I_j = 1) = \frac{1+\delta}{j(2+\delta)+(1+\delta)} \tag{19}$$

then that means I_j is the event of the vertex i forming a self loop (j = i) when it's attached to the graph. $\delta > 0$ is a parameter that represents a shift of the proportionality in the preferential attachment.

The sum $\sum_{j=i}^{n} I_j$ represents the total number of self loops formed by adding vertices, so it's stochastically dominated by $D_i(n)$, which is the degree of the vertices in the graph at iteration n. Since $D_i(n)$ gets updated at every iteration, it will grow either faster or at the same rate as the sum of I_j, since for any vertex we have at most one self loop, while the degree depends on any type of edge.

The Borel-Cantelli lemma states that the sum of the events $\{E_n\}$ is finite then the probability that infinitely many of them occur is 0:

$$\mathbb{P}(limsup_{n\to\infty} E_n) = 0 \tag{20}$$

On the contrary, if the sum is infinite, then the probability that infinitely many of them occur is 1.

We want to apply this lemma to the sum:

$$\sum_j \mathbb{P}(I_j = 1) \tag{21}$$

For big values of j (big number of nodes already in the graph), we can rewrite (19) as:

$$\mathbb{P}(I_j = 1) \sim \frac{1+\delta}{j(2+\delta)} \tag{22}$$

This is an harmonic series $\sum_{j=1}^{\infty} \frac{1}{j}$ with $\frac{1+\delta}{2+\delta}$ as a multiplying factor that depends solely on δ We know that every harmonic series diverges, so we can conclude that the number of links formed by vertex i grows to infinity with probability 1.

Since $D_i(n)$ is stochastically bigger than a sum that grows to infinity, then we can conclude that:

$$\mathbb{P}(\lim_{n\to\infty} D_i(n) = \infty) = 1$$

which means that the degree of a vertex i tends to infinity with probability 1.