UPC

Lab3: Significance of Network Metrics

Simon Van den Eynde
Martí Renedo Mirambell

October 27, 2016

1 Introduction

2 Results

3 Discussion

4 Methods

4.1 Analytical estimation of the p-value

4.1.1 Erdös-Rényi Graph

Given our original graph with N vertices and E edges, the Erdös-Rényi graph has the same size and order but with its edges randomized. Of the two possible Erdös-Renyi models, we will use $G(N, p)$ where p will be such that the expected number of edges is E (which is $p = \frac{E}{\binom{N}{2}}$). This has the advantage that X_j the random variables that for every possible vertex indicate wether it exists (X_j=1) or not (X_j=0) are independent, which will be very useful later on. When working with large values of N and E (such as those we study in this lab session), the models $G(N, E)$ and $G(N, p)$ should give very similar random graphs. For X_j where j is the index of any vertex, we calculate its expectation and variance:

$$\mathrm{E}[X_j] = 0 \cdot (1 - p) + 1 \cdot p = p$$

$$\mathrm{Var}(X_j) = \mathrm{E}[X^2] - (\mathrm{E}[X])^2 = p - p^2$$

Table 1: Summary of the properties of the degree sequences. N is the number of vertices of the network, E is the number of edges, $\langle k \rangle = 2E/N$ is the mean degree and $\delta = 2E/(N(N-1))$ is the network density of edges.

Language	N	E	$\langle k \rangle$	δ
Arabic	21532	68767	6,4	3,0e-04
Basque	12207	25558	4,2	3,4e-04
Catalan	36865	197318	10,7	2,9e-04
Chinese	40298	181081	9,0	2,2e-04
Czech	69303	257295	7,4	1,0e-03
English	29634	193186	13,0	4,4e-04
Greek	13283	43974	6,6	5,0e-04
Hungarian	36126	106716	5,9	1,6e-04
Italian	14726	56042	7,6	5,2e-04
Turkish	20409	45642	4,5	2,2e-04

Table 2: Values of the mean local clustering coefficient and its p-values with respect to the binomial (Erdös-Renyi) and switching models.

Language	C_{WS}	p-value (binomial)	p-value (switching)

Given a vertex v of the graph, it can have $N-1$ neighbours (each with probability p), which results in $\binom{N-1}{2}$ possible pairs of neighbours. The expected number of pairs. This gives an expected number of pairs of neighbours $n_v = p^2\binom{N-1}{2}$. Then, we can estimate the local clustering of v by

$$C_v^{ER} \approx \frac{\sum_{j=1}^{\lfloor n_v \rfloor} X_j}{n}.$$

Since X_j are independent and equally distributed, we can apply the central limit theorem, which states that for a large enough sample size, the distribution of C_v^{ER} is the normal $N(\mathrm{E}(C_v^{ER}), \frac{\mathrm{Var}(X_j)}{n}) \approx N(C_v^S, \frac{p-p^2}{p\binom{n-1}{2}}) = N(C_v^S, \frac{1-p}{p\binom{N-1}{2}})$.

We can apply the central limit theorem again to $C_{WS}^{ER} = \frac{1}{N}\sum_{i=1}^{N} C_i$, wich gives us the distribution of C_{WS}^{ER}: the normal distribution

$$N(C_v^S, \frac{1-p}{pN\binom{N-1}{2}}) = N(C_v^S, \frac{1-p}{E(N-2)}).$$

Note that in the last step we used the equality $pN\binom{N-1}{2} = \frac{EN\binom{N-1}{2}}{\binom{N}{2}} = E(N-2)$.

Now the p-value of C_{WS}, which is given by $p(C_{WS}^{ER} \geq C_{WS})$, can be calculated