
Lab3: Significance of Network Metrics

Simon Van den Eynde
Martí Renedo Mirambell

October 26, 2016

1 Introduction

2 Results

3 Discussion

4 Methods

4.1 Analytical estimation of the p -value

4.1.1 Erdős-Rényi Graph

Given our original graph with N vertices and E edges, the Erdős-Rényi graph has the same size and order but with its edges randomized. Of the two possible Erdős-Rényi 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 whether 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:

$$\begin{aligned} \mathbb{E}[X_j] &= 0 \cdot (1 - p) + 1 \cdot p = p \\ \text{Var}(X_j) &= \mathbb{E}[X_j^2] - (\mathbb{E}[X_j])^2 = p - p^2 \end{aligned}$$

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 $N(E(C_v^{ER}), \frac{\text{Var}(X_j)}{n}) \approx N(C_v^S, \frac{p-p^2}{p \binom{N-1}{2}}) = N(C_v^S, \frac{1-p}{p \binom{N-1}{2}})$.

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

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

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

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