Math C5.4, Networks, University of Oxford
Problem Sheet 2

Renaud Lambiotte, Mathematical Institute

1. *Erdos-Renyi.* The simplest and most famous type of random graph is an Erdős–Rényi (ER) graph (which was first studied by Solomonoff and Rapoport). In particular, consider the random-graph ensemble $\mathcal{G}(N, p)$, which is defined as follows: Suppose that there are N nodes. Between each pair of distinct nodes, a single edge exists with uniform and independent probability p. There are no self-edges. A single graph $G \in \mathcal{G}(N, p)$ is generated using this process, and it is interesting to study the properties of collections ("ensembles") of graphs that are generated in this way. The probability in which each simple graph G with m edges appears in a graph $G \in \mathcal{G}(N, p)$ is

$$P(G) = p^m (1-p)^{{}^nC_2 - m} . \tag{1}$$

   (a) Write down the total probability of drawing a graph G with exactly m edges from the ensemble $\mathcal{G}(N, p)$, and use it to find the mean value of edges $\langle m \rangle$.

   (b) Calculate the expected mean degree of an ER graph.

   (c) Show, under an appropriate assumption (which you should state), that the degree distribution for an ER graph (in expectation over the ensemble) satisfies

$$p_k \sim e^{-c} \frac{c^k}{k!} , \quad N \to \infty , \tag{2}$$

   where $c = (N-1)p$.

   (d) Reproduce numerically the results of Figure (14) in the lecture notes.

   (e) Calculate the global clustering coefficient C of an ER graph (in expectation over the ensemble), and verify your prediction numerically.

   (f) Show that the diameter of an ER graph is $B + \frac{\ln N}{\ln c}$ as $N \to \infty$, where B is a constant and $c = (N-1)p$.

2. *Configuration model.*

   (a) Ex.V.2 : Generate randomised version of different empirical networks and verify that the number of self-loops and multiple edges becomes negligible when the system is sufficiently large.

   (b) Take an undirected network and its randomised version according to the configuration model. Compare the correlations between node degree and Pagerank for each network. Describe and comment the results.

3. *Preferential attachment.* Consider a directed version of the BA model defined on page 25, modelling citation networks: nodes are articles citing the previous literature. When a new article comes in, it draws one single directed edge (for simplicity) to an existing article, with a probability proportional its a constant C + its in-degree. Start from an initial configuration with one single paper with one self-citation.

   (a) Derive the master equation for the in- and -out degree

   (b) Estimate the properties of the asymptotic in-degree distribution and verify your result by performing numerical simulations of the model.

   (c) Is this model a good model for citation networks? How could it be improved?