

# M2 – NSD (Practical 3)

## Graph models

Maximilien Danisch, Marwan Ghanem, Lionel Tabourier

This practical aims at generating graphs and comparing their properties to the ones of real-world graphs.

### Preamble

Depending on the efficiency of your codes, you will be able to manage graphs of different sizes. Here are a few datasets of different sizes available on <http://konect.uni-koblenz.de/> and <http://snap.stanford.edu/data/> which can be used for comparison purposes :

- small datasets :
  - Infrastructure : <http://konect.uni-koblenz.de/networks/contiguous-usa>  
 $n = 49$  and  $m = 107$
  - Bible : [http://konect.uni-koblenz.de/networks/moreno\\_names](http://konect.uni-koblenz.de/networks/moreno_names)  
 $n = 1773$  and  $m = 16401$
- medium datasets :
  - Coauthoring : <http://konect.uni-koblenz.de/networks/ca-AstroPh>  
 $n = 18771$  and  $m = 198050$
  - Internet-AS : <http://konect.uni-koblenz.de/networks/as-caida20071105>  
 $n = 26475$  and  $m = 53381$
- large datasets :
  - Amazon : <http://snap.stanford.edu/data/com-Amazon.html>  
 $n = 334863$  and  $m = 925872$
  - LiveJournal : <http://snap.stanford.edu/data/com-LiveJournal.html>  
 $n = 3997962$  and  $m = 34681189$

### Exercise 1 — *Erdős-Rényi (ER)*

**Question 1.** Make a program which, given two integers  $n$  and  $m$  generates an ER graph with  $n$  nodes and  $m$  edges (without multiple edges and self-loops) and writes the list of edges in a file.

**Question 2.** Generate ER graphs with  $n$  and  $m$  corresponding to the datasets listed above.

**Question 3.** Measure the following properties.

- the number of connected components, the size of the largest connected components, an approximation of the average distance and a (good) lower-bound to the diameter
- the number of triangles, the average clustering coefficient and the transitivity ratio
- plot the degree distribution of this graph using (for example) GNUPLOT

### Exercise 2 — *Barabási-Albert (BA)*

**Question 1.** Make a program which, given two integers  $n$  and  $m$  generates a BA graph with  $n$  nodes and approximately  $m$  edges (without multiple edges and self-loops) and writes the list of edges in a file.

**Question 2 and 3.** Same as the ones of exercise 1.

### Exercise 3 — *Configuration Model*

**Question 1.** Make a program which, given a degree sequence, creates a random graph with the same degree sequence using the configuration model seen during the course (we allow self-loops and multiple edges).

**Question 2.** Modify the program in question 1 so that it removes self-loops and multiple edges (for instance by randomly switching edge-ends).

**Question 3.** Extract the degrees of the two following graphs and make the two corresponding random graphs using your program made in question 2.

**Question 4.** Same as the question 3 of exercise 1.

**Exercise 4 — *Switches method***

**Question 1.** Create a program which, given a graph, achieves  $P$  random switches of edge-ends. Make sure that switches do not create any self-loop or multiple edge.

**Question 2.** Apply your algorithm on the two real-world graphs given in the previous exercise. Which value of  $P$  seems reasonable?

**Question 3.** Same as the one of Exercise 1.

**Exercise 5 — *Conclusion*** Compare the properties obtained for the models to the ones of the corresponding real-world graph and conclude.