

1) The file `facebook_sample_anon.txt` is a data table containing the list of edges of an anonymized sample of the Facebook friendship network. Download it on your computer, upload it to R as a dataframe, and define an undirected graph with this list of edges.

- (a) What are its order and size?
- (b) Is it connected?
- (c) Compute its density. Do you consider it sparse?
- (d) Plot its degrees distribution in linear and in log-log scale. Which is more helpful to understand this distribution? Does this distribution seem to follow a power-law?
- (e) What are the range, the mode, the mean, the median and the standard deviation of its degree distribution? What do they tell you?
- (f) Compute its diameter, its average distance, and its 90% effective diameter (the smallest distance d such that at least a 90% of the pairs of different nodes are at distance at most d).
- (g) Plot a histogram of its clustering coefficients.
- (h) Compute its average clustering coefficient and its transitivity coefficient. What do they tell you about the clustering coefficient of the nodes with large degree? Confirm it by computing the clustering coefficients of the 10 nodes with largest degree.
- (i) This graph was obtained by gathering all friends of 10 students enrolled in a course and establishing the friendship relations among the union of their friends (in the old times when Facebook APIs allowed it easily). Can you make an educated guess about which are these 10 students?

2) The file `WikiVote.txt` is a data table containing the votations on promotion to administratorship of Wikipedia (till January 2008). Each arc (x, y) in it means that user x voted on y becoming Wikipedia administrator. More details at <http://snap.stanford.edu/data/wiki-Vote.html>.

Download it on your computer and upload it to R as a dataframe. Define a directed graph DW with this list of arcs, and the underlying unweighted undirected graph UW .

- (a) Does DW contain self-loops? How many? What do self-loops mean, in the context of this graph? If any, remove them (as well as from UW).
- (b) What are the order and size of DW ? What is the size of UW ? What does the difference in sizes correspond to?
- (c) Does DW have a giant weakly connected component? And a giant strongly connected component?
- (d) Plot its out-degree distribution on a log-log scale (omitting the nodes with out-degree 0).
- (e) Do the same with its in-degree distribution.
- (f) For both the in-degrees and the out-degrees, compute their range, mode, mean, median, and standard deviation. What do they tell you?
- (g) What kind (or kinds) of transitivity do you consider worth studying in this network? Choose at least one, compute it, and comment the results.

You must deliver an R Markdown document with the code, the results and your comments, and the html file obtained by compiling it.