A model of networks with structured nodes

Pierluigi Frisco
School of Math. and Comp. Sciences, Heriot-Watt University,
EH14 4AS Edinburgh, UK, P.Frisco@hw.ac.uk

We defined a network model with structured nodes and a definition of distance between these nodes. This model creates networks with different combinations of power law degree distribution exponent, average path length and average clustering coefficient. The obtained results, still in an initial phase, are definitely very promising a they are close to the ones present in empirical networks.

The model. Each node in the network has a *structure* (also called *name*, *description*, etc.): a string of a fixed length over an alphabet. Initially there is only one node. New nodes are added to the network one by one and their structure is a modification of the structure of any of the already present nodes. New nodes have edges with previously existing nodes only if the distance between their structure is smaller than a certain value. The distance uses two parameters: `Unit_distance` and `Max_distance` and now we describe it through an example.

Let us assume to have two nodes with structures: `ABBCCA` and `BABCCC`.

If `Unit_distance = 1`, then the distance of these two structures is 3, this is because, when compared symbol by symbol, there are three symbols that are different. The three different symbols are: the first, the second and the sixth. In this case, as `Unit_distance = 1`, we matched one symbol with another symbol.

If `Unit_distance = 2`, then we have to compare pairs of symbols in one structure with pairs of symbols in another structure. Two pairs are considered equal if they have the same symbols, independently from their order. The first pair or symbols in the first structure is `AB`, while in the second structure is `BA`. These two pairs are not different, as each of them contain one `A` and one `B`. The second pair of symbols in the first and second structure is `BC`. These two pairs are not different. The third pair of symbols in the first structure is `CA`, while in the second structure is `CC`. These two pairs are different, as the first pair contains one `C` and one `A`, while the second pair contains two `C`'s. In this case, as `Unit_distance = 2`, the distance is 1 (due to the third pair).

In general, `Unit_distance` denotes how the symbols in a structure are grouped.

Having said this, two nodes have a common edge if their distance is smaller/equal than the value of `Max_distance`. So, in the first case (`Unit_distance = 1`) the two nodes will have a common edge only if the value of `Max_distance` is smaller/equal 3; in the second case (`Unit_distance = 2`) the two nodes have a common edge only if the value of `Max_distance` is 1.

At the end of each run all the nodes with less than 6 edges and all the edges to them are removed.

Results. The following table lists some obtained combinations of power law degree distribution exponent (γ), average path length (l_G) and average clustering coefficient ($\bar{C}$):

γ	l_G	$\bar{C}$
-2	3.9	0.78
-2.7	4.7	0.84
-3.3	3.8	0.86

The following figure displays a typical outcome of a network with 3000 nodes: