CHALMERS UNIVERSITY OF TECHNOLOGY

Simulation of Complex Systems FFR120 2015

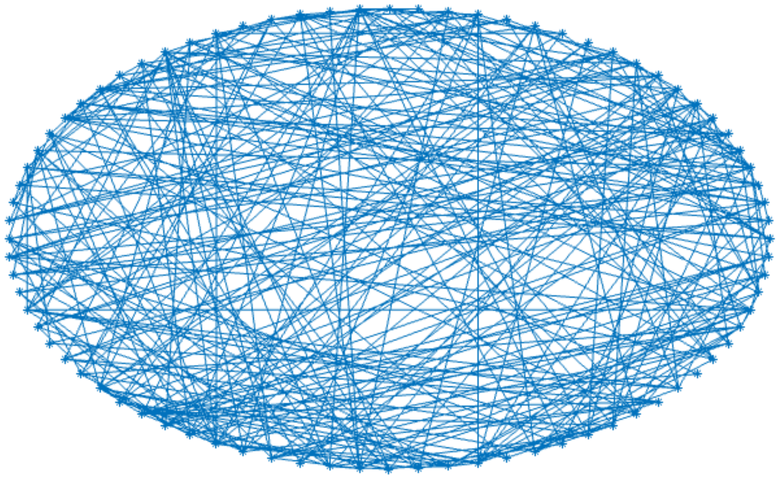
Homework 4 : Network models

Author:
August von Hacht
930513-1592
hacht@student.chalmers.se

Examiner:
Kolbjörn Tunström

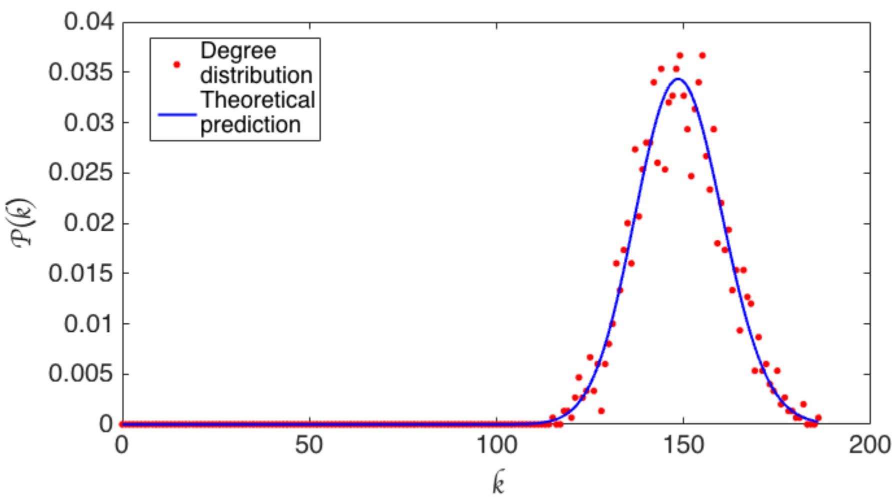
Erdõs-Rényi random graph :

I generate node positions on the unit-circle by dividing them equally spaced $2\pi/n$ where n is the number of nodes. The resulting networks follows below.



Figur 1: *Erdõs-Rényi random graph using $n = 80$ and $p = 0.1$.*

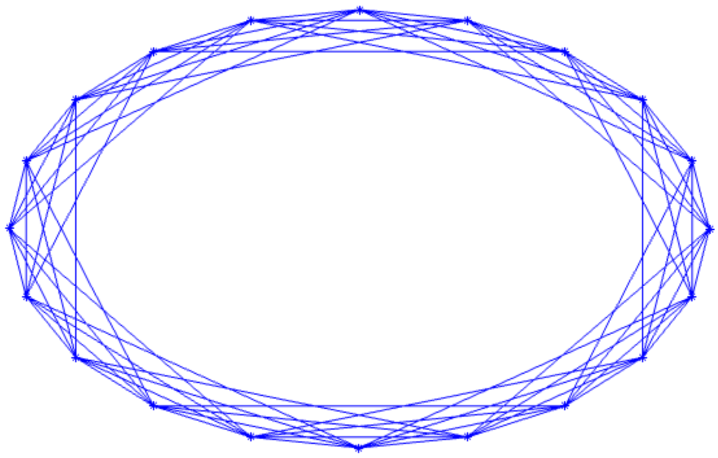
The degree distribution is found as the trace of the square adjacency matrix A^2 since this is the number of ways to go from node i, i and back using a step length 2 i.e. one edge. The degree distribution follows below.



Figur 2: *Erdõs-Rényi random graph degree distribution using $n = 1500$ and $p = 0.1$ plotted together with the theoretical prediction.*

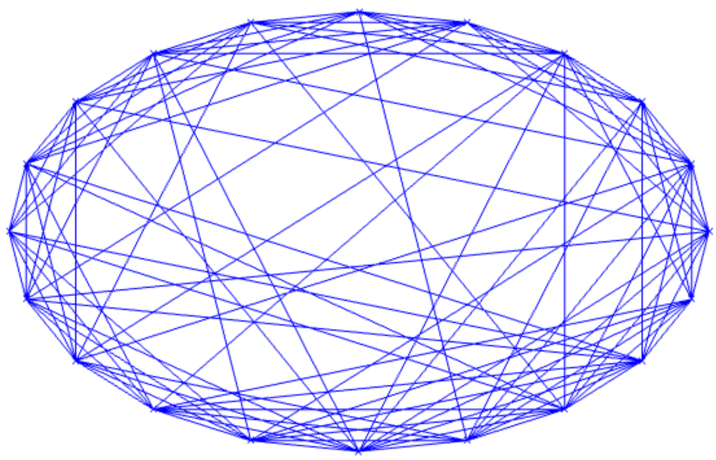
Watts-Strogatz small world model :

The positions are again generated on the unit-circle equally spaced. The networks follows below.



Figur 3: *Watts-Strogatz small world model using* $n = 20$ *and* $c = 8$ *i.e. each node connects to its eight nearest neighbors.*

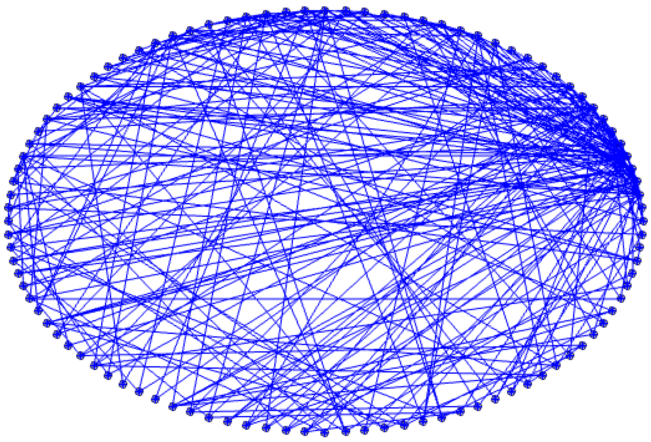
Now I randomly add new edges (without rewiring) and the results looks like below.



Figur 4: *Watts-Strogatz small world model using* $n = 20$, $c = 8$ *and* $p = 0.2$ *i.e. each node has a probability* p *to connect an edge to another node chosen randomly.*

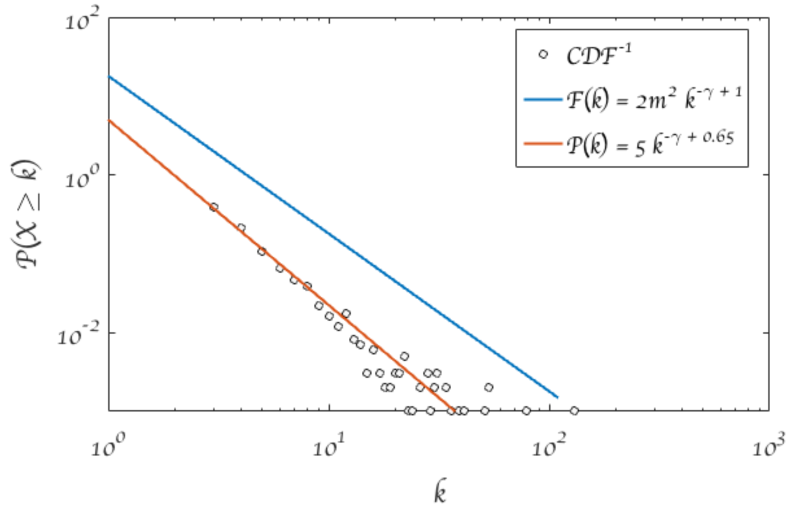
Albért-Barabási preferential growth model :

The positions are again generated on the unit-circle equally spaced. The networks follows below.



Figur 5: *Albért-Barabási preferential growth model using* $n_0 = 100$ *and* $m = 3$*. Each connection is chosen probabilistic in proportion to the cumulative sum of the number of connected nodes.*

The inverse cumulative distribution is found to counting the occurrences of a number of edges for each node and then dividing by the sum to get the probability that a node has equal or more than that number of edges.



Figur 6: *Inverse cumulative distribution of the Albért-Barabási preferential growth model using* $n_0 = 1000$ *and* $m = 3$*. The theoretical prediction* $F(k)$ *does not fit the curve well since in the implemented model connection to another node was picked uniformly from a list of nodes which means new nodes can connect multiple time to hubs which benefits these and therefore the slope is expected to be slightly steeper. This curve is shown as* $P(k)$*.*

Clustering coefficient :

As mentioned earlier the trace of the square adjacency matrix A^2 gives the degree distribution k_i. Applying the formula

$$\#\text{triples} = \sum_i \frac{k_i(k_i-1)}{2}$$

give the number of triples since this is the maximum number of edges having this degree distribution. Next the trace of the cube adjacency matrix A^3 give the number of ways going from node (i,i) and back in step length of 3 which is precisely a triangle. However, to not over count triangles by counting the clockwise and counter-clockwise paths I divide the number of triangles with 2. This gives the algorithm

```
ClusteringCoefficient = sum(diag(A^3))/( sum(diag(A^2).*(diag(A^2)-1)) )
```

which sums up the diagonal elements of A^3 divided by the diagonal elements of A^2 input to the formula above.

Using this algorithm on the smallWorldExample.txt gives me the clustering coefficient:

$$\text{Clustering coefficient} = 0.61127956$$

Average path length and diameter :

Analogously as in previous task the A^L determines the path lengths L between nodes. Finding the shortest path length between nodes means finding the smallest L for which $A(i,j)^L$ first is non-zero. My routine for doing this is to first generate a list of all the possible element positions (i,j) to check i.e. all the upper triangular positions $\forall i$ and $j > i$. These are stored in a list and for every order of L I check the positions found in the list and apply them to A^L. If this position is non-zero then (i,j) is set to L in another matrix that keeps track of the L:s corresponding element position (i,j).Index (i,j) is then removed from the list. This goes on until the list is empty. Finally I apply the average path length formula on the matrix storing all of the Ls for each distance between (i,j) (multiplied by two since the formula assumes you average over both lower and upper triangular part). The diameter is found as the maximum L in the same matrix. The algorithm follows as

```
list = [];
for i = 1:n
    tmp = zeros(n - i,2);
    for j = i+1:n
            tmp(j-i,1:2) = [i j] ;
    end
    list = cat(1,list,tmp);
end

l = 1; lengthMatrix = zeros(n);
while any(list)
    mat = A^l; r = [];
    for i = 1:size(list,1);
        if (mat(list(i,1),list(i,2)) ~= 0)
        lengthMatrix(list(i,1),list(i,2)) = l;
            r = [r i];
        end
    end
    list(r(:),:) = [];
    l = l + 1;
end
average_path_length = 2*sum(sum(lengthMatrix(:,:)))/(n*(n-1));
diameter = max(max(lengthMatrix));
```

Applying this algorithm to smallWorldExample.txt gives me the average path length and diameter

$$\text{Average path length} = 2.93232$$

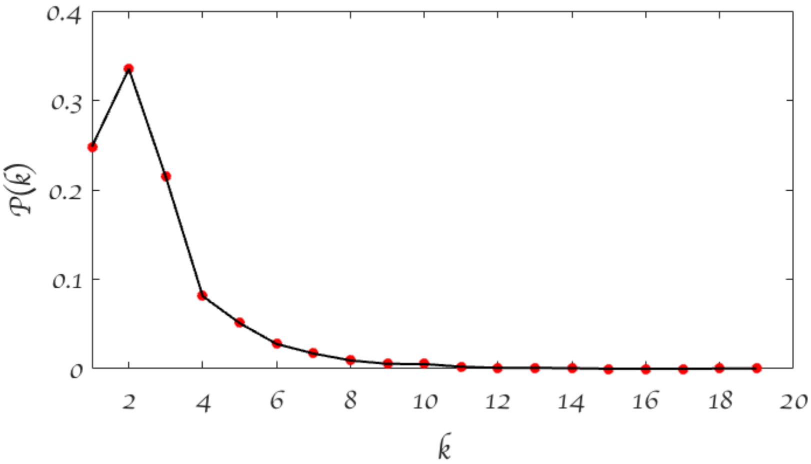
$$\text{Diameter} = 5$$

Three unknown networks :

Applying all of the tools: Degree distribution, clustering coefficient, average path length and diameter for each network should give some hint of what we are dealing with.

Network 1 :

The degree distribution follows as



Figur 7: $P(k)$ *is the probability of nodes with degree* k*. This distribution has a lot of nodes with degree* $1, 2$ *and* 3 *which indicates that there are a lot of closely connected nodes. However, there are also some that are further away from the rest which indicates possible separated cluster(s).*

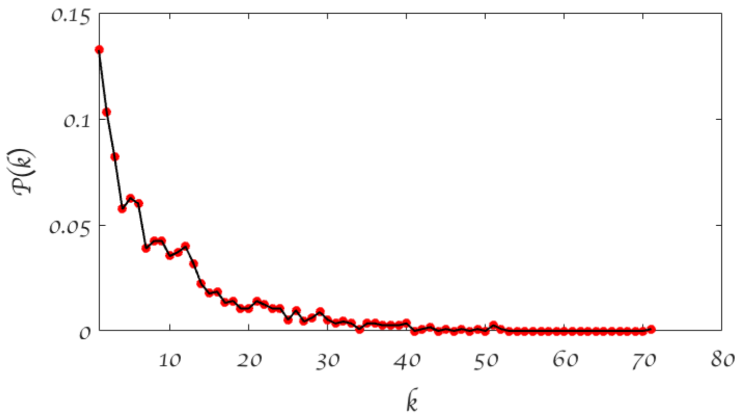
The clustering coefficient, average path length and diameter is found to equal

$$\text{Clustering coefficient} = 0.10315$$
$$\text{Average path length} = 18.9891$$
$$\text{Diameter} = 46$$

Since this is a social networks people in general tend to prefer connecting with more socially active people or these are simply the people they are most likely to meet. The clustering coefficient is large even though the average path length and diameter is large which indicates that many nodes are connected to a hub from far away which explains the large probability of have neighbors that are neighbors but at the same time nodes on average being far away from each other. So this network should be a preferential growth network because of the large clustering coefficient but large average path length and diameter.

Network 2 :

The degree distribution follows as



Figur 8: $P(k)$ *is the probability of nodes with degree* k. *The distribution falls of quickly which indicates that there are a lot nodes close to each other. However, if they are close to a hub or close to spread out neighbors is hard to tell from this plot.*

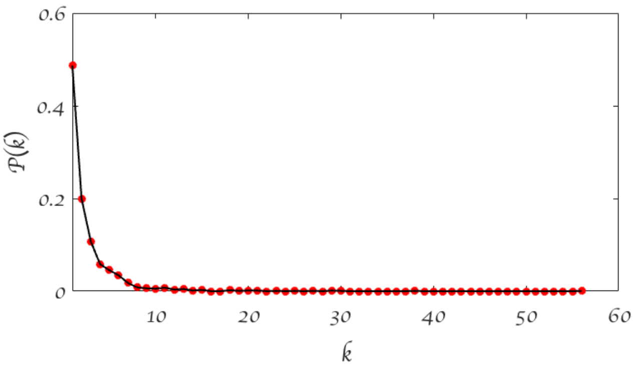
The clustering coefficient, average path length and diameter is found to equal

$$\text{Clustering coefficient} = 0.16627$$
$$\text{Average path length} = 3.6060$$
$$\text{Diameter} = 8$$

Since this is a power grid one expects that there are connections between houses neighboring houses which all finally are connected to a power station. The clustering coefficient is larger than the other networks and that tells us that neighbors of nodes are also likely to be neighbors just like a power grid is suppose to function. The average path length is rather small which agrees with that neighboring houses near each other are connected to each other. However, the diameter is large but this is defined as the largest distance which can be though of as the connection from the last house to the power plant which typically are stationed a bit away from the center of a city. So this network should be a small world network because of large clustering coefficient and small average path length.

Network 3 :

The degree distribution follows as



Figur 9: $P(k)$ *is the probability of nodes with degree* k*. The distribution falls of more rapidly than previous networks and has degrees in between that is zero hinting about random effects.*

The clustering coefficient, average path length and diameter is found to equal

$$\text{Clustering coefficient} = 0.07937$$
$$\text{Average path length} = 6.8124$$
$$\text{Diameter} = 19$$

Since this is interaction on microlevels in nature I expect it to be rather random without a preferred interaction patterns. However, yeast might group up to become separated hubs. The clustering coefficient which tells how likely your neighboring neighbors are to be connected is very low. The average path length is higher than network2 so they are generally not very close by each other. The diameter is large which is hard to explain but it could be a random effect of a random graph where one node happen to randomly get very disconnected from the rest by connection to only one other node that is also very far away and so on. So this networks has rather spread out nodes or nodes that simply are connected randomly. But, since the clustering coefficient is so low I would argue that this is a random network because of the small clustering coefficient and the high diameter.