Complex Networks

Master of Science in Electrical Engineering

Erivelton Geraldo Nepomuceno

Department of Electrical Engineering
Federal University of São João del-Rei

May 16, 2019

4. Small-World Networks

- "It's a small world!" This is the typical expression we use many times in our lives when we discover, for example, that we unexpectedly share a common acquaintance with a stranger we've just met far from home.
- This happens because social networks have a small characteristic path length.
- Social networks also have a large clustering coefficient, i.e. they contain a large number of triangles.
- The small-world model proposed in 1998 by Watts and Strogatz to construct graphs having both the small-world property and also a high clustering coefficient.

4.1 Six Degrees of Separation

- Stanley Milgram, a Harvard social psychologist, conducted a series of experiments in late 1960s to show that, despite the very large number of people living in the USA and the relatively small number of a person's acquaintances, two individuals chosen at random are very closely connected to one another.

Chain length	2	3	4	5	6	7	8	9	10
Number of completed chains	2	4	9	8	11	5	1	2	2

Table 1: Length distribution of the completed chains in Milgram small-world experiment. The length of the chains varied from 2 to 10 intermediate acquaintances, with the median of the distribution being equal to 5.5 and its average to 5.43. Rounding up, the experiment suggests that we are all at six steps of distance, the so-called six degrees of separation.

Box 1 (The Erdös Number)

Paul Erdös,one of the fathers of random graph theory, was an incredibly prolific mathematician. He wrote thousands of scientific papers in different areas of mathematics, many of them in collaboration with other mathematicians. Erdos number is based on a concept of co-authoring. Erdös co-authors have Erdös number 1. Mathematicians other than Erdös who have written a joint paper with someone with Erdös number 1, but not with Erdös, have Erdös number 2, and so on. Finding the Erdös number: https://www.csauthors.net/distance

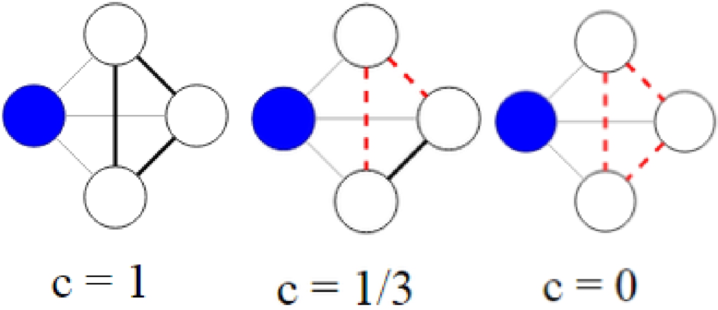

Figure 1: The local clustering coefficient of the blue node is computed as the proportion of connections among its neighbours which are actually realised compared with the number of all possible connections. The blue node has three neighbours, which can have a maximum of 3 connections among them. In the first figure, the local clustering coefficient is 1. In the middle part of the figure only one connection is realised (thick black line) and 2 connections are missing (dotted red lines), giving a local cluster coefficient of 1/3. Finally, the clustering coefficient value is 0.
Adaptaded from: https://en.wikipedia.org/wiki/Clustering_coefficient

Table 2: Characteristic path length and clustering coefficient of three real networks (two social networks and a biological one) and of the correspondingER random graphs with same number of nodes and links.

	N	$\langle k \rangle$	L	L^{ER}	C	C^{ER}
Movie actors	225226	73.71	3.65	2.87	0.79	0.00033
ArXiv	44337	10.79	5.99	4.50	0.62	0.00024
C. elegans	279	16.39	2.44	2.01	0.34	0.06

4.2 The Brain of a Worm

- In 1972 Sydney Brenner, a biologist then at Cambridge University, decided to work out the connections of every cell in the nervous system of the C. elegans. Dr. Brenner has won the Nobel Prize in 2002.
- He picked this animal because its body has only 959 cells, of which 302 are nerve cells.
- In DATA SET Box 4.2 we introduce and describe the basic features of the neural network of the C. elegans

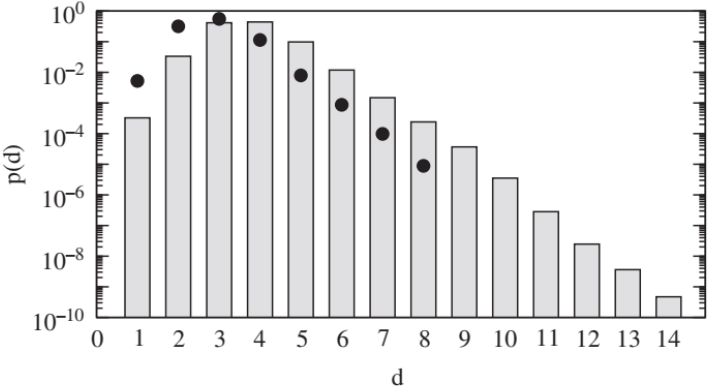

Figure 2: The normalised number $p(d)$ of pairs of actors connected by a shortest path of length d is plotted as a function of d (histogram). The normalised number of actors at distance d from Bacon is also reported for comparison (circles).

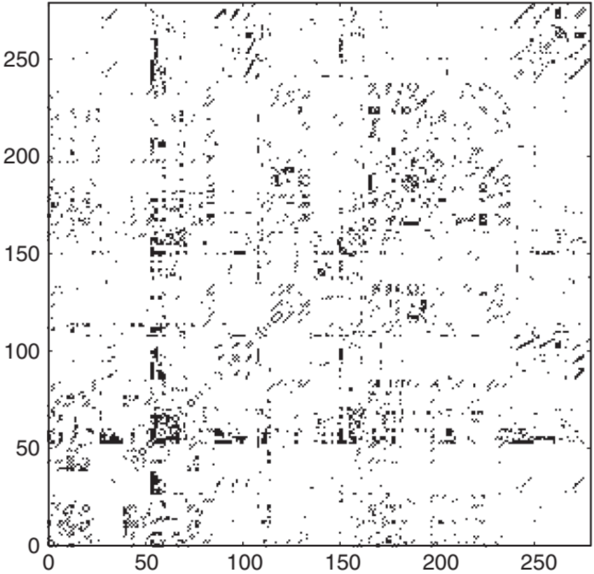

Figure 3: Graphical representation of the connections among the 279 nodes

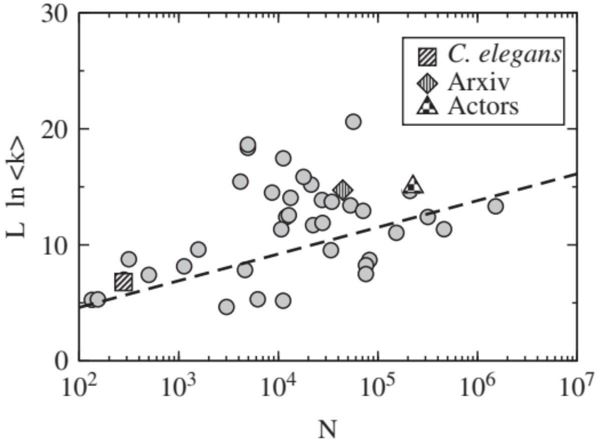

Figure 4: The average path length of real networks (symbols) is compared with the prediction for ER random graphs of different orde(dashed line).

Definition 1 (Small-world behaviour)

We say that a network exhibits a small-world behaviour if the characteristic path length L grows proportionally to the logarithm of the number of nodes N in the network.

4.3 Clustering Coefficient

- Important property of social networks: the presence of a large number of triangles.
- The clustering coefficient of Erdös–Renyí random graphs reads: $C^{ER} = p = \langle k \rangle / N$.
- We notice that in real networks the quantity $C/\langle k \rangle$ appears to be independent of N, instead of decreasing as N^{-1}. Thus random graphs are inadequate to describe some of the basic properties observed empirically, and we will therefore need to explore other models.

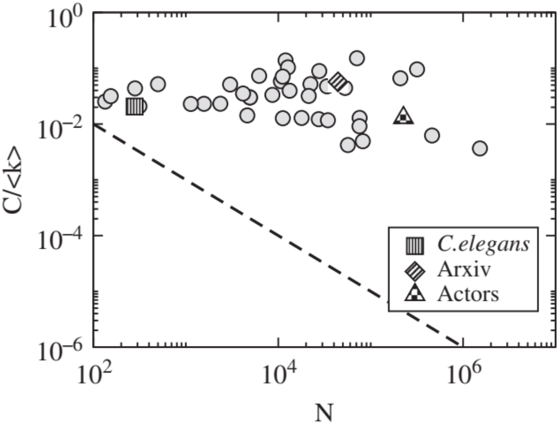

Figure 5: The clustering coefficient of real networks (symbols) is compared to the prediction for ER random graphs (dashed line).

4.4 The Watts–Strogatz (WS) Model

- ER random graph models have the small-world property, but their clustering coefficient is negligible.
- The first model to construct graphs having, at the same time, a high clustering coefficient and a small characteristic path length was proposed by two applied mathematicians at Cornell University, Duncan Watts and Steven Strogatz, in a paper published in Nature in 1998.
- We will refer to the model as the Watts and Strogatz small-world model, or, in short, the WS model.
- Basically, the idea of Watts and Strogatz is to introduce a certain amount of random connections in a lattice graph.

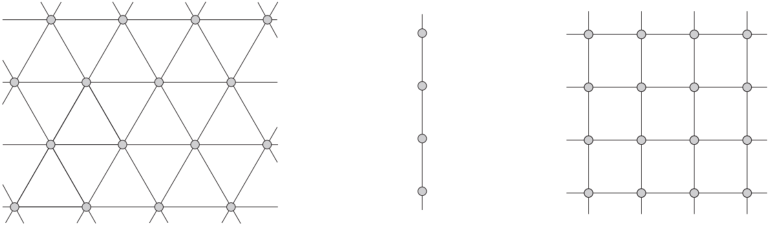

Figure 6: Various examples of lattices: a two-dimensional triangular lattice, a one-dimensional lattice and a two-dimensional square lattice

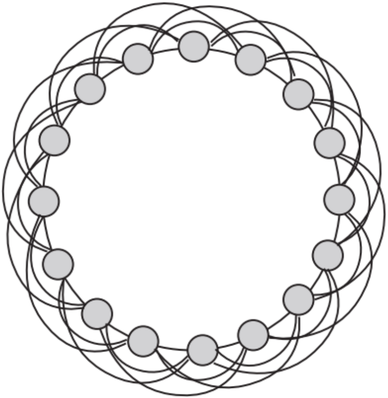

Figure 7: Circle graph with $N = 16$ and $m = 3$. Consider a circular lattice of size N, i.e. N nodes uniformly distributed on a circle, each node connected to its two first neighbours. $A(N, m)$ circle graph, with $1 < m < N/2$, is obtained by additionally linking each node of a circular lattice to the $2m - 2$ closest nodes in the lattice. This graph has a total of $K = Nm = 48$ links.

Computational Exercises

1. Study the functions: `clust`, `largest_component`, `ws`
2. Develop a function that could be able to distinguish a model from WS and other from ER with the same number of nodes. Use $N = 1000$.