Solution 5-11

This is an example solution from the forthcoming book Networks and Complexity.
Find more exercises at https://github.com/NC-Book/NCB

Ex 5.11: Computer network [4]

A computer network consists of 5 servers and 100 workstations. Each server is connected to all other servers and to 20 of the workstations. The workstations have no further connections. Find the exact value of the average path length, and compare it to approximations, with and without correction for clustering coefficient. (You will find that the approximation performs exceptionally poorly in this case. A brief explanation why this happens is given in the solution.)

Solution

We can compute the following distances (S=Server, W=Worstation):

Communication	distance	number of paths
S to S	1	$5 \cdot 4/2 = 10$
S to W on that server	1	$5 \cdot 20 = 100$
S to W on other server	2	$5 \cdot 80 = 400$
W to W on same server	2	$5 \cdot 20 \cdot 19/2 = 950$
W to W on different server	3	$5 \cdot 20 \cdot 80/2 = 4000$

So the diameter is

$$D_{]rmExact} = \frac{1 \cdot (10 + 100)) + 2 \cdot (400 + 950) + 3 \cdot 4000}{10 + 100 + 400 + 950 + 4000} = \frac{14810}{5460} \approx 2.712 \qquad (1)$$

A good way to make sure that we have not forgotten any type of path, is to compute the the total number of shortest paths in a different way. There are in total 105 computers in the network, so the total number of source-destination pairs is

$$105 \cdot 104/2 = 5460 \qquad (2)$$

which is also the denominator that we found by adding up all the individual types of shortest paths considered in the table. This shows that we dd not forget to consider paths.

To use our diameter estimate we first need the mean degree. There are 10 links between the 5 servers and 100 links between servers and workstations, so our network has 110 links in total between its 105 nodes. So the mean degree is

$$z = \frac{220}{105} \approx 2.095 \qquad (3)$$

Using our simple diameter estimate we would expect that the average path length is

$$D_{\text{Est}} = \log_{2.095}(105) \approx 6.44 \qquad (4)$$

which more than twice the actual value. So let's try the formula with the clustering coefficient correction. To do so let's first count the triangles in the network. The number of three-node-chains between servers is $5 \cdot 4 \cdot 3/2 = 30$ as all of these are part of triangles we can simply find the number of triangles by dividing by 3 to obtain

$$n_\Delta = 10. \qquad (5)$$

To find the total number of three-node chains we have to add the all the three node chains between servers and workstations connected to other servers ($5 \cdot 4 \cdot 20 = 400$) and chains between workstations on the same server ($5 \cdot 19 \cdot 20/2 = 950$), hence

$$n_{--} = 30 + 400 + 950 = 1380. \tag{6}$$

Thus the clustering coefficient is almost zero,

$$c = \frac{3n_\triangle}{n_{--}} = \frac{30}{1380} \approx 0.02 \tag{7}$$

At this low value we don't expect the clustering coefficient to expect the diameter estimate very much. Indeed we find

$$D_{\text{Est2}} = \log_{(1-c)z((1-c)N)} \approx \log_{2.05}(103) \approx 6.46 \tag{8}$$

So the correction due to the clustering coefficient is very minor and takes our estimate even a bit further from the truth.

Why does the estimate perform so poorly in this case? Consider that in our derivation of the estimate we assumed that the networks are reasonably randomly connected. However, our computer network is a very non-random network. In fact, the network that we considered is not a typical example of a network with 105 nodes and 110 links. Instead among the gazillion networks that can be constructed out of 105 nodes and 110 links it is among a very small number of topologies that have a very low diameter.

Of course this is no accident, we build computer networks in this way precisely because having central highly connected servers brings down the diameter substantially and thus allows data to be exchanged efficiently.

We will understand the mathematics of how this reduction in diameter is achieved in the next chapter and discover how to account for it for it in our equations.