CSCI 688 Homework 1

Sidi Chang

Sunday 7th February, 2016

Answers of HW 1 will be shown here, code and readme file will be submitted in another file. If you have any problems with running the code, please let me know. Some of the MATLAB code cannot be run in function because I do not set return value, you can just paste it and it should be fine. I will also attach the workspace so you can see all the variables and their values there.

1. Data
   I cleaned my own data but the node is slightly different from Van's data. So all my answer is based on Van's data, if you want my data and the C code, I can send it to you.

2. 1(a)
   Total number of edges is 111632, self-loops is 393 and number of edges between distinct nodes separately is 111239.
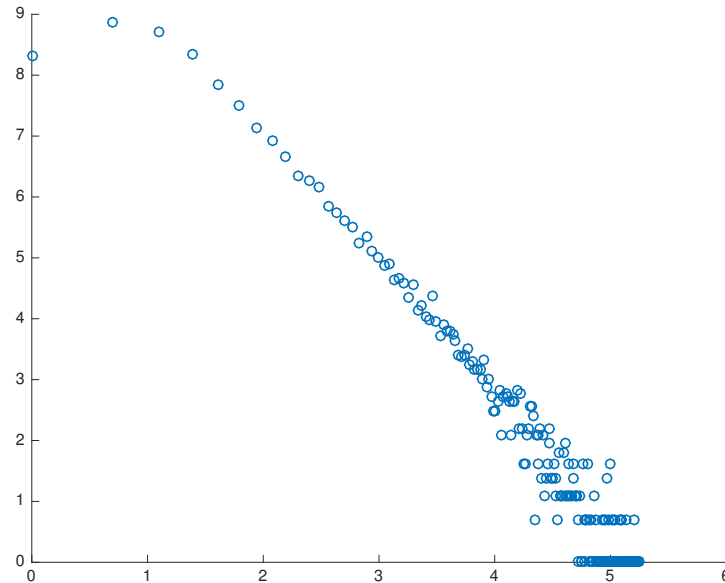
3. 1b(ii)
   The figure [1] is shown below:

Figure 1: log-log degree distribution of real world data

4. 1c

n* = 27837, n=34740, so the ratio is $n*/n = 27837/34740 = 0.80$. Then Table [1] will report j and k_j: Figure [2] is the scatterplot of the log-log relationship for Table [1]:

Table 1: connected component size j versus number of connected components k_j

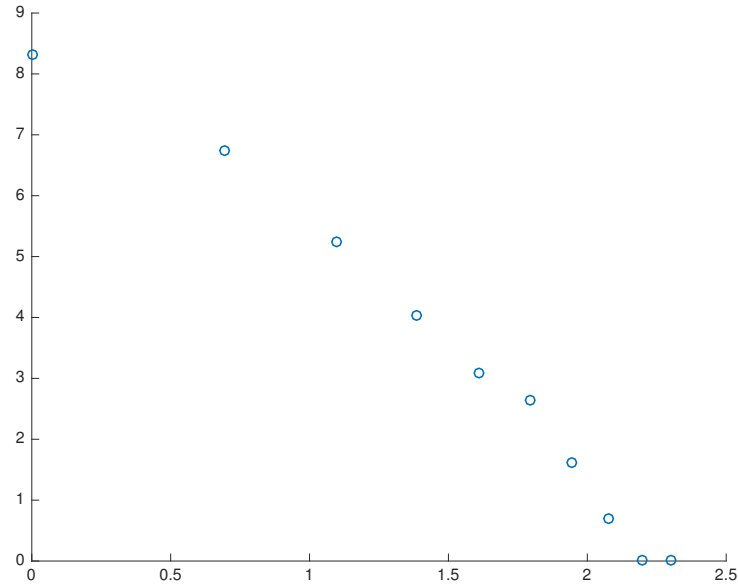| j	1	2	3	4	5	6	7	8	10	27837
k_j	4145	856	189	56	22	14	5	2	1	1

Figure 2: log-log relation of size j and number k_j

5. 1d

Use BFS and from BFS connected node, I got the maximum distance between node 3439 to others is 8. Similarly, for node 7881, I got maximum distance is 12.
So I assume the upper and lower bound on the diameter of the largest connected component is [8,12].

6. 2a

Random Network and Small-world Network generation code will be attached in another file.
3-4 Self-loops in expectation by probability which is much smaller than real world data.
Figure [3] will show three combination of the log-log degree distribution of three different data: Real world data log-log distribution is of course linear relation.
Small World is kind of like reverse-J shaped.
Random Network is kind of like mounded.
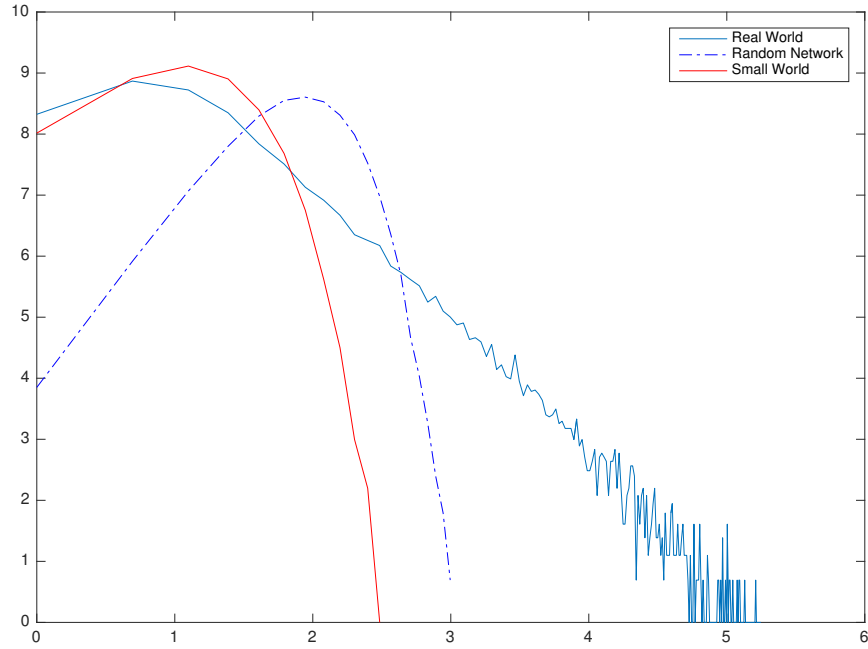Real world data has the most nodes with higher degree compared with small world data the

Figure 3: log-log degree distribution of three different data

least.

Small world data and random network data seem to be more clustered than real world data, real world data has a few nodes with degree 0 and 1, which means they are almost isolated or exactly isolated. This situation is not common on random network data, impossible on Small World data.

7. 2b
Random Network size of largest connected component is 34693 which is extremely high, while Small-World network is 34740(Inf in MATLAB) which covers the whole nodes. That makes sense to me because we connect all the nodes when we construct Small-World Network.

8. 2c
For Small-World network, I picked up [1:100] nodes and calculated their maximum distance each and maximum again, I got 9 for Small World Network. Similarly, I got 12 for Random Network data.