Homework 6 - Theory/Laboratory

Dainese Fabio, 857661

April 12, 2020

1 Load, Visualize and Analyse Network Data

After loaded the '*VIPD.mat*' file and generated the relative sub-graph without isolated nodes starting from the first 2 columns (*source, target*), we have obtained the following results:
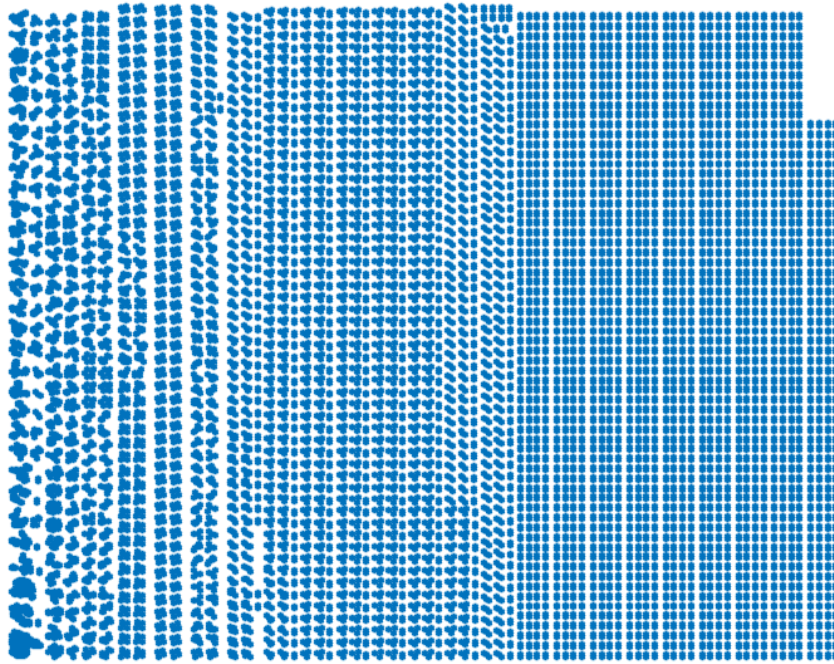

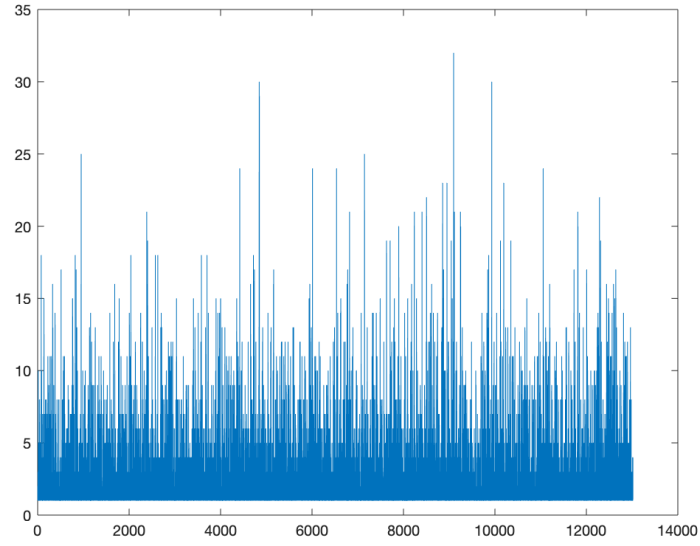Figure 1: Plot of the sub-graph using the '*force*' layout

Figure 2: Plot of the degree distribution of the sub-graph (unsorted)
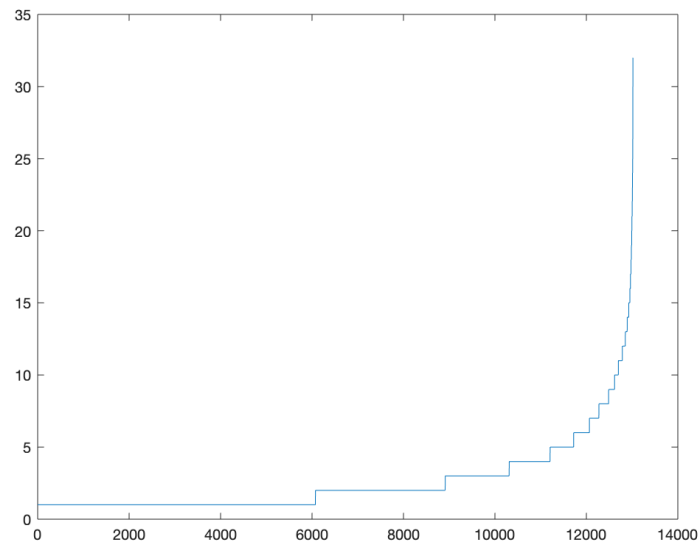

Figure 3: Plot of the degree distribution of the sub-graph (sorted)

- The average degree is 2.595591

- The average clustering coefficient is 0.421364

- The average path length is 0.034427.

  To precise that it has been used the following formula to calculate the APL:
  $$APL(G) = \frac{1}{n \cdot (n-1)} \sum_{i \neq j} d(v_i, v_j)$$

  where n is the number of nodes in the graph G and also assume that $d(v_1, v_2) = 0$ if v_2 cannot be reached from v_1.

  This clarification is due since the MatLab function '*distances(G)*' doesn't put 0 whenever a specific node cannot be reached, instead it puts '*Infinite*' (so using '*mean(distances(G))*' did returned the expected value).

- The number of nodes of the sub-graph are 13019

These previous results have been calculated using the following *MatLab* script:

```matlab
clear all;
clc;

% ====================================================================
% ======================= 1 - Load the data ========================
% ====================================================================

load VIPD.mat;

s = table2array(VIPDedgelist(:,1));
t = table2array(VIPDedgelist(:,2));

G = graph(s,t,[],max(s));

% ====================================================================
% ==================== 2 - Generate a subgraph =====================
% ====================================================================

cnt = 'degree';
d = centrality(G,cnt);

G_New = subgraph(G,d>=1);

% ====================================================================
% ==================== 3 - Plot the sub-graph ======================
% ====================================================================

figure(1)
plot(G_New,'Layout','force','iterations',10);

```

```matlab
31  % ===================================================================
32  % ============== 4 - Plot the degree distribution ==================
33  % ===================================================================
34
35  deg = degree(G_New);
36
37  figure(2)
38  plot(deg)
39  plot(sort(deg));
40
41  % ===================================================================
42  % ================ 5 - Find the average degree =====================
43  % ===================================================================
44
45  fprintf('%f \n',mean(deg));
46
47  % ===================================================================
48  % ========= 6 - Find the average clustering coefficient ===========
49  % ===================================================================
50
51  cc = clustering_coef_bu(G_New.adjacency);
52  fprintf('%f \n',mean(cc));
53
54  % ===================================================================
55  % ============== 7 - Find the average path length =================
56  % ===================================================================
57
58  pl = distances(G_New);
59  % Replace all the 'Infinite' with '0' in the matrix
60  pl(~isfinite(pl)) = 0;
61  n = numnodes(G_New);
62  fprintf('%f \n',(sum(pl,'all')/(n*(n-1))));
63
64  % ===================================================================
65  % ========== 8 - Find the number of nodes of the graph ============
66  % ===================================================================
67
68  fprintf('%d \n',numnodes(G_New));
```

2   Calibrate Random Network Models

The value of the parameters of the different models (*ER,WS,BA*) which make
the average node degree of each of them close to the one of the interlocking
directorate (ID) network are:

- **ER model** $G_{ER}(n,p)$

  With the number of nodes equals to $n = 13019$ (the same of the ID sub-
  graph) and the edge probability equals to $p = 0.0002$, we have an average
  node degree of 2.5941.

- **WS model** $G_{WS}(n,K,p)$

With the number of nodes equals to $n = 13019$ (the same of the ID sub-graph), the number of neighbours in each side for each node equals to $K = 1$ ($2K$ = the number of neighbours for each node = node degree) and the re-wiring probability q (in this case is indifferent the value of q since the re-wiring procedure doesn't alter the number of edges for each node, thus also the node degree), we have an average node degree of 2.

- **BA model** $G_{BA}(n, m, m_0)$ (with $1 < m < m_0$)

  With the number of nodes equals to $n = 13019$ (the same of the ID sub-graph), the initial number of connected nodes equals to $m = 1$ and the number of edges generated at each iteration m_0 (indifferent also in this case), we have an average node degree of 2.

In the following table you can see the comparison between the three calibrated models and the ID data over some statistics:

Model Name	Average CC	Average PL	Average Degree
ID sub-graph	0.421364	0.034427	2.595591
ER	0.000156	7.8931	2.5941
WS ($q = 0.5$)	0	22.584	2
BA ($m_0 = 2$)	0	7.122	2

In the following histogram ('Figure 4'), you can see a direct comparison between the different models (ER, WS, BA) and the ID data over the degree distribution. To point out that in the plot the range of degrees have been limited up to 20 in order to understand more clearly the data (zoom in the range that contains the peaks).
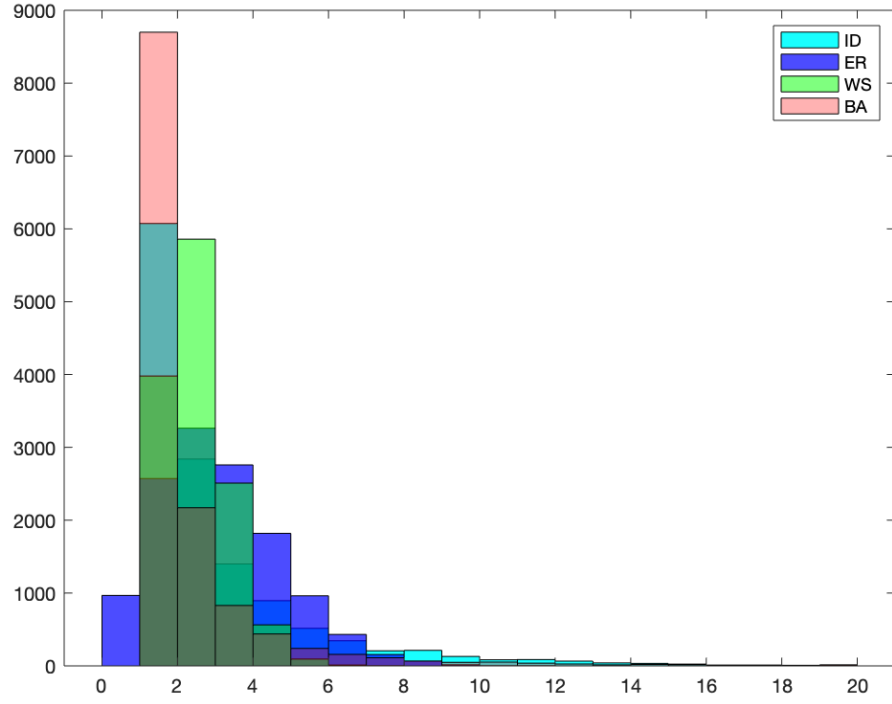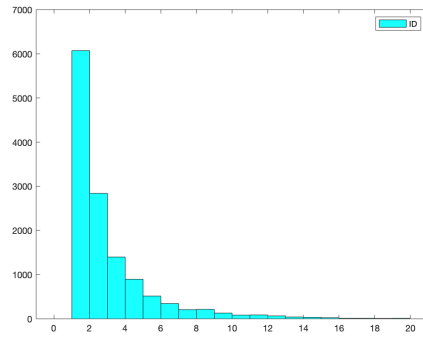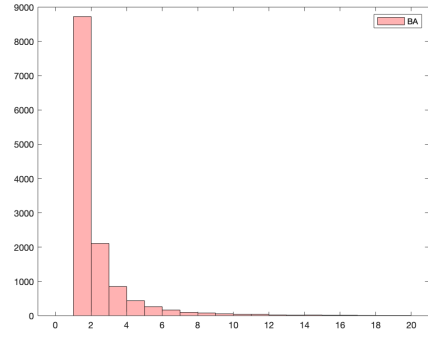
Figure 4: The degree distribution for the models and the ID data

The best model for the ID network is the BA (*Albert-Barabasi*) one, since it's the most similar starting from the previous experiment results and clearly visible by looking its degree distribution that fits the original one (see the next 2 plots).

(a) Isolated zoomed degree distribution of ID

(b) Isolated zoomed degree distribution of BA