CS-E5740 Complex Networks, Answers to exercise set 7

Tommaso Brumani, Student number: 100481325

November 8, 2022

Problem 1

a) The complementary cumulative distribution function for node degree, strength, and link weight of the network described by the **OClinks_w_undir.edg** file was computed and plotted.
To better visualize the fat-tailed nature of the distributions, the CCDFs of exponential distribution with the same mean were plotted alongside them.
The results are reported in Figure 1.

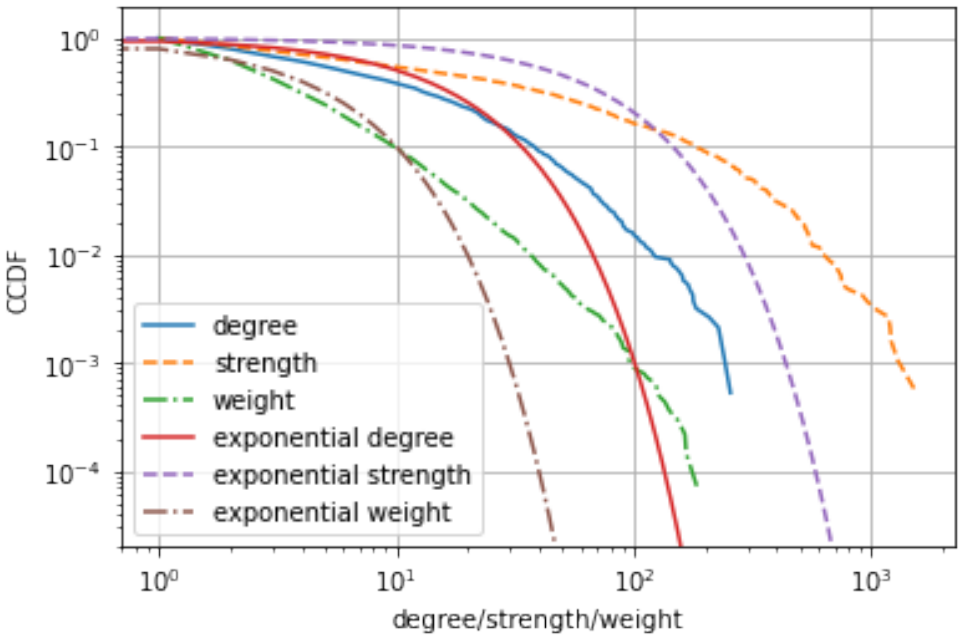

Figure 1: CCDF of k, s, and w.

b) The link weight per node was calculated for all nodes, and plotted against node degree both along linear and logarithmic horizontal axes.
The data was also binned based on the node degree, and the bin-averages for v were calculated and plotted.
The result are reported in Figures 2 and 3.
From observation, it appears that the logarithmic plot is a better representation of the relationship, as it more clearly shows the tendency of v to gradually increase as k grows.

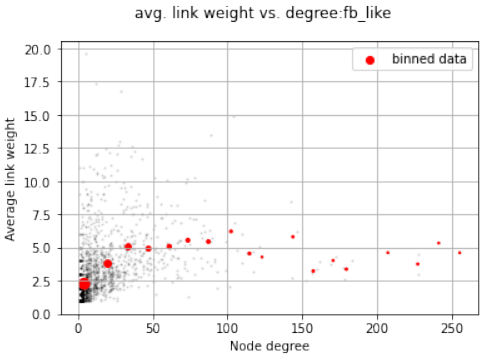

Figure 2: v vs k (linear).

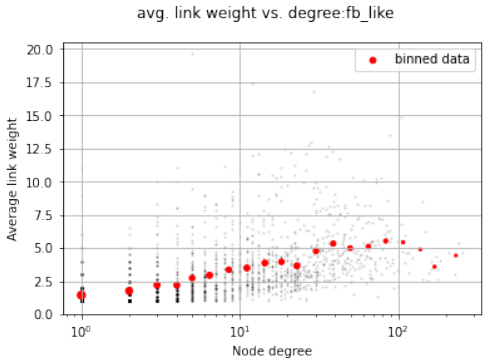

Figure 3: v vs k (logarithmic).

c) The neighborhood overlap measure was calculated for each link in the network and plotted as a function of link weight.
In the same plot were also indicated the bin-averages calculated over 20 bins to better grasp the trend of the data.
A logarithmic plot was determined to be the most suitable, and the results are reported in Figure 4.
The trend in the data is for the neighborhood overlap to slightly increase as a function of link weight, which is in line with the Granovetter hypothesis for social networks.

Problem 2

a) A number of quantities were computed over the network described by the **aggregated_US_air_traffic_network_undir.edg** file, and are reported in Figure 5.

b) The network links were visualized overlayed to a map of the USA, and the result is reported in Figure 6

c) In order to reduce the number of plotted links, the maximal and minimal spanning trees were computed and visualized over the same map, and the results are reported in Figures 7 and 8.

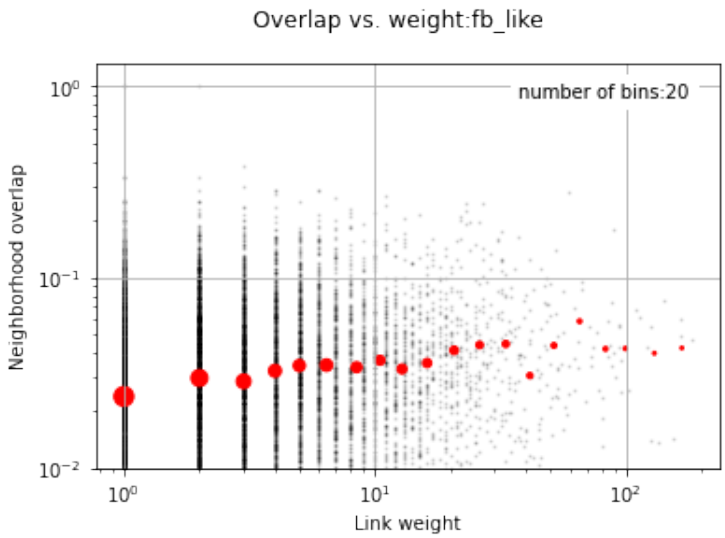

Figure 4: Neighborhood overlap vs link weight.

```
Number of nodes: 279
Number of links: 2088
Density: 0.05384079832907867
Diameter: 4
Average clustering coefficient: 0.6465167472774311
```

Figure 5: Network quantities.

Since the link weights represent the number of flights between the airports in the given time period, the maximal spanning tree is probably a more useful tool for understanding air traffic in the US.

Figure 7: Minimal spanning tree.

Figure 8: Maximal spanning tree.

Figure 6: Visualized full network.

d) The thresholded network keeping only the $N - 1$ strongest links was calculated and plotted over the US map, with the result being reported in Figure 9.
Of the edges present in the thresholded network, 97 overlap with those of the maximal spanning tree, which is about a third of the total links in the two networks.
From this and visual observation, we can determine that the results of the two processes are quite similar.

Figure 9: Thresholded network.