CS-E5740 Complex Networks, Answers to exercise set 6

Yangzhe Kong, Student number: 765756

November 9, 2019

1 Weight–topology correlations in social networks

a) The image is shown in Figure 1

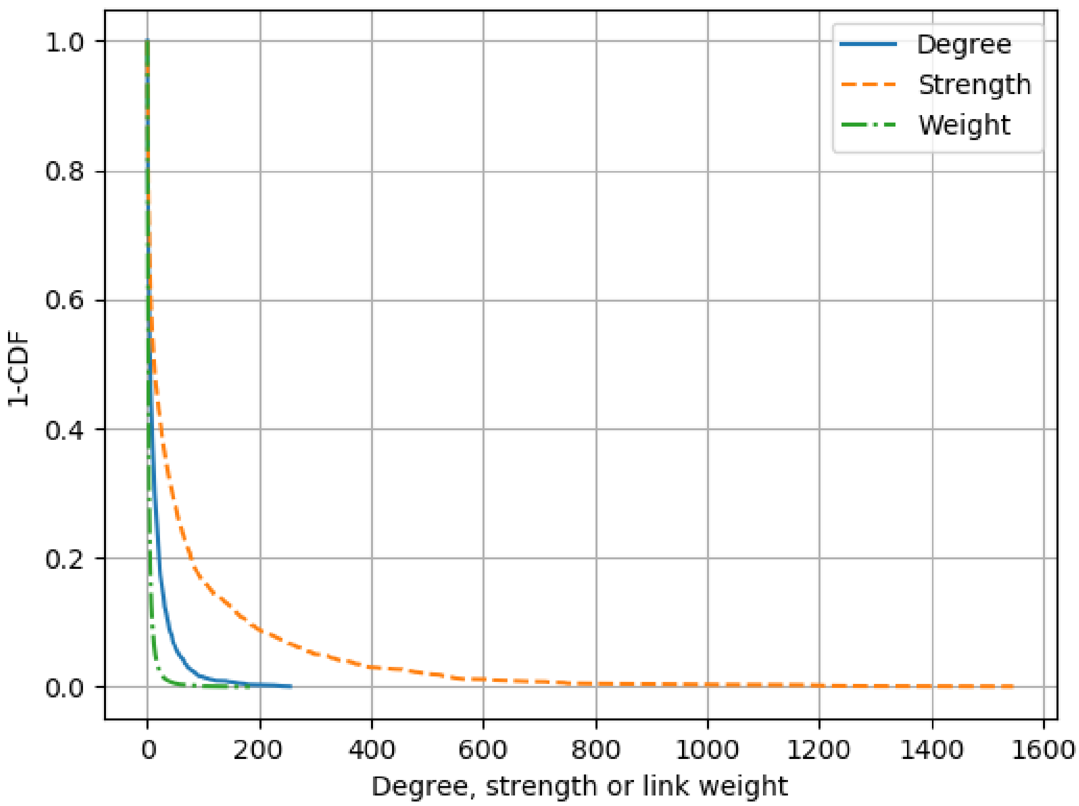

Figure 1: 1-CDF for node degree, node strength and link weight

As we can see, they follow the Gaussian Distribution as the graph is concave. The 90^{th} percentiles for Degree, strength or link weight is as follows:

Degree: ≈ 40 Strength: ≈ 160 Link weight: ≈ 10

b) and c) The scatter plot of $\langle w \rangle$ as a function of k with linear x-axis and its binned version is shown in Figure 2 and the scatter plot of $\langle w \rangle$ as a function of k with log x-axis and its binned version is shown in Figure 3.

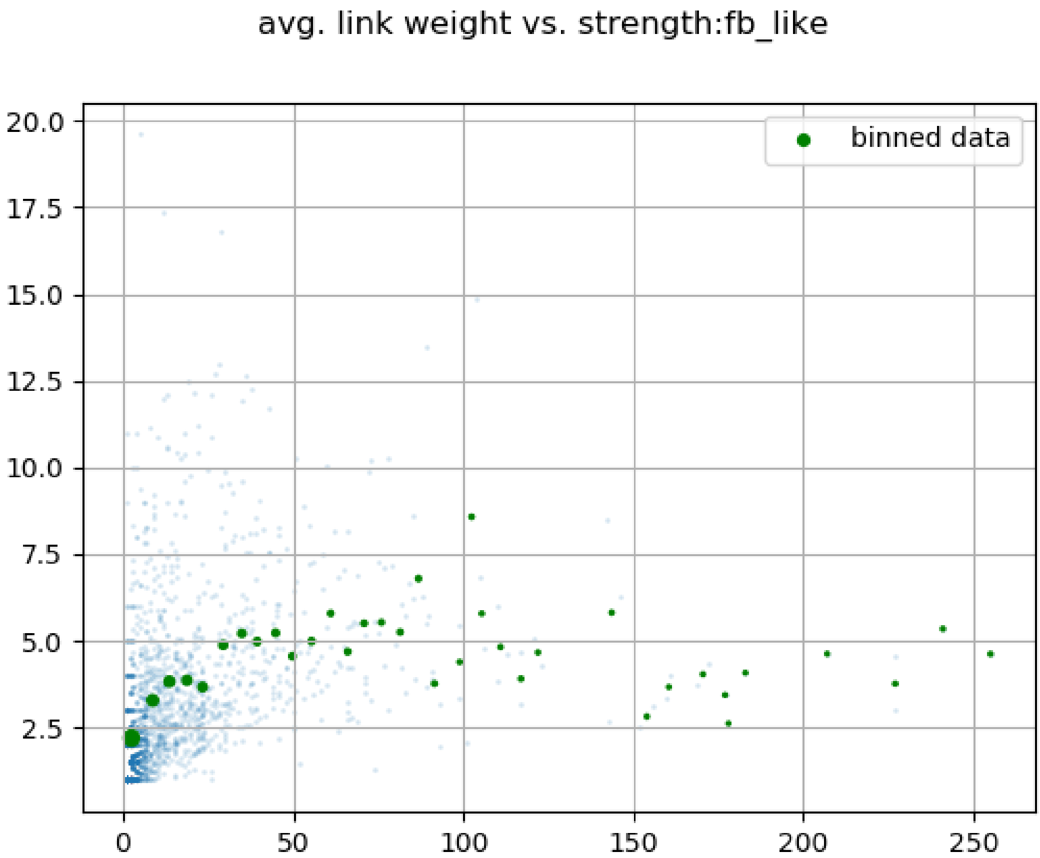

Figure 2: The scatter plot of $\langle w \rangle$ as a function of k with linear x-axis and its binned version

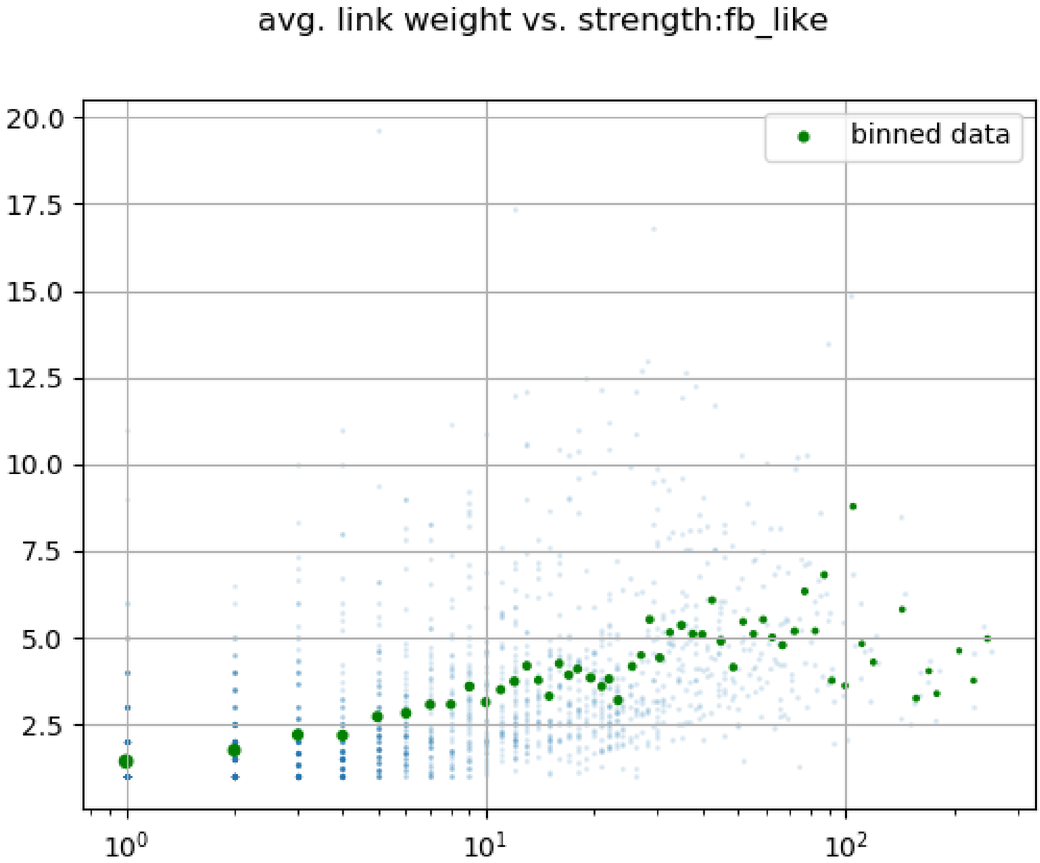

Figure 3: The scatter plot of $\langle w \rangle$ as a function of k with log x-axis and its binned version

d) **Question** Which of the two approaches (linear or logarithmic x-axes) suits better for presenting $\langle w \rangle$ as a function of k? Why?

   Answer For $\langle w \rangle$ as a function of k, we would choose logarithmic x-axes because it is a better representation since most of binned data points degrees are between 1 and 100.

   Question In social networks, $\langle w \rangle$ typically decreases as a function of the degree due to time constraints required for taking care of social contacts. Are your results in accordance with this observation? If not, how would you explain this?

   Answer My results are in accordance with this observation.
   First as the degree k increases, average link weight increases. We can explain this via homophily based on the assumption that people who have more friends spend more time on Facebook sending more messages to friends who has many friends has well.

After some point around $k = 100$, the average weight link decreases, due to time constraints as the question explains. And that's exactly what the plots show.

e) The result is shown in Figure Figure 4 and it's using log binning strategy.

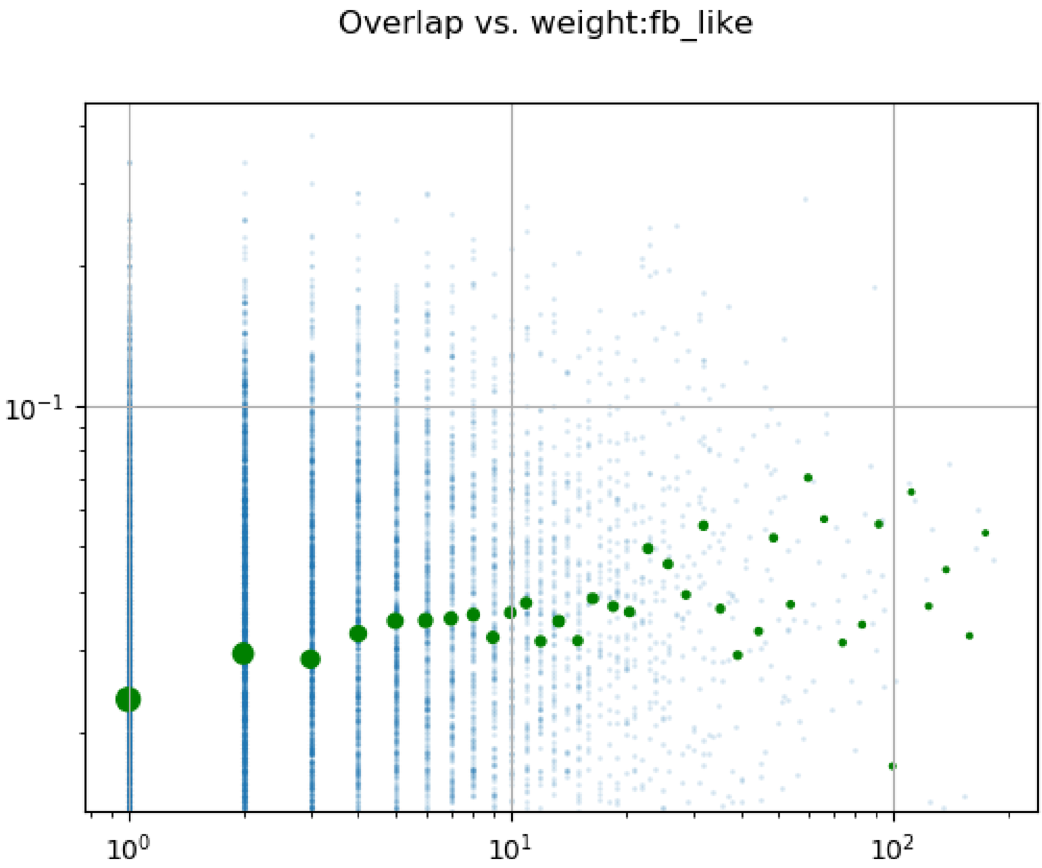

Figure 4: The scatter plot of the overlaps as a function of link weight

Since there's a log linear relation between link weight and link neighborhood overlap as the figure shows. A log increase in link weight will relate to linear increase in link neighborhood overlap. Thus this trend is in accordance with the Granovetter hypothesis.

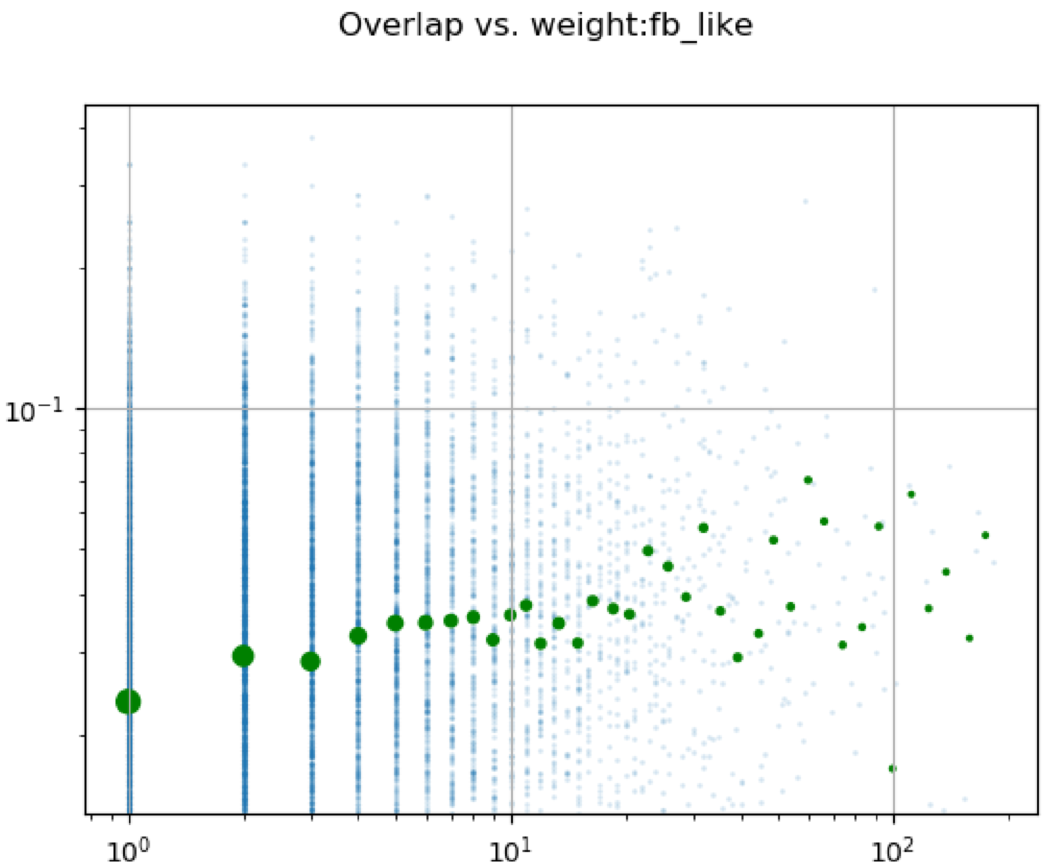

Figure 5: Bin-averaged scatter plot of $\langle w \rangle$ as a function of k with log x-axis

2 Network thresholding and spanning trees: the case of US air traffic

a) Number of nodes: 279
Number of edges: 2088
Density: 0.05384079832907867
Network diameter: 4
Average clustering coefficient: 0.6465167472774311

b) The image of the full network is shown in Figure 6.

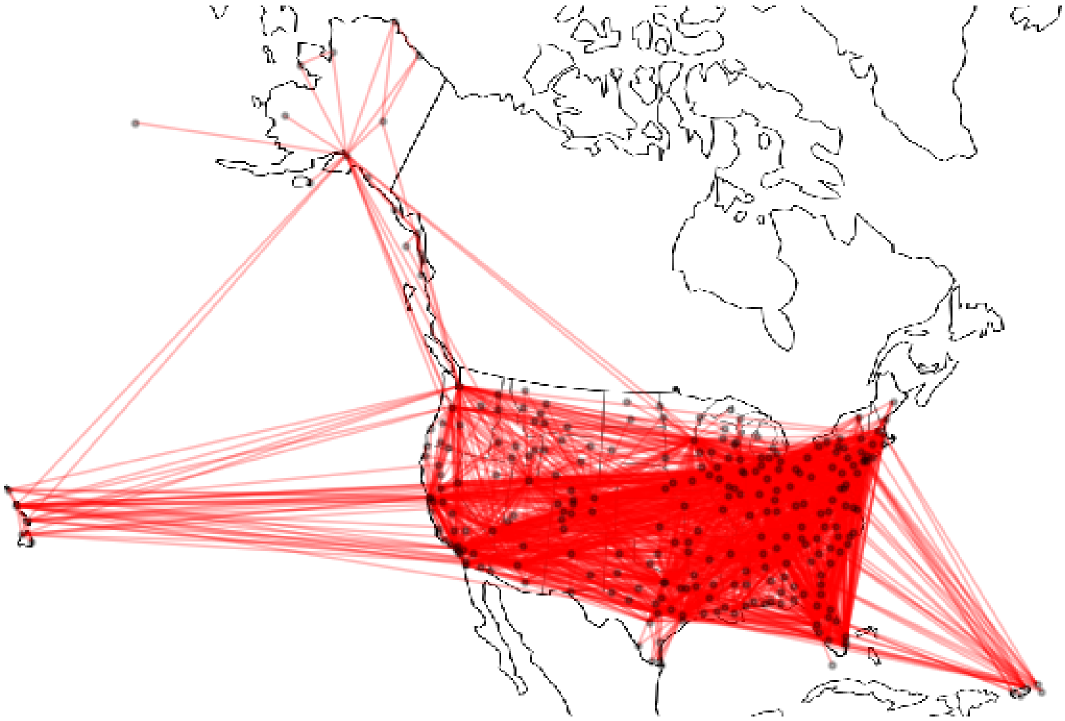

Figure 6: The full network

c) The minimal spanning tree is shown in Figure 7 and maximal spanning tree is shown in Figure 8.

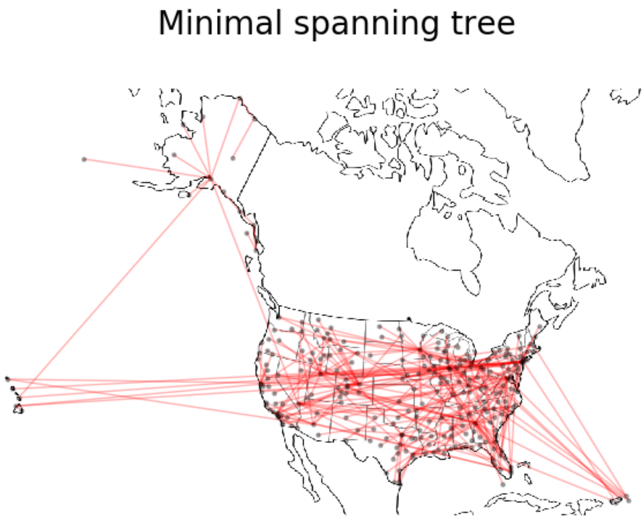

Figure 7: Minimal spanning tree

Figure 8: Maximal spanning tree

Question If the connections of Hawai'i are considered, how would you explain the differences between the minimal and maximal spanning trees?

Answer For the maximal spanning tree, we can observe that there is only one flight to Hawai'i which is from Los Angeles. This means between all flights to Hawaii, Los Angeles - Hawaii has the most flights.

For minimal spanning tree, it looks more messy. There are many flights to Hawai'i which means these flights are not only least frequent flights to Hawai'i but also least frequent flights in US.

Question If you would like to understand the overall organization of the air traffic in the US, would you use the minimal or maximal spanning tree? Why?

Answer I would like to use Maximal spanning tree in the sense that the visualized flights are the most frequent ones, in other words, people use these flights the most. In addition, we can view more of the domestic flights that are from a larger airport to a smaller airport

d) The result is shown in Figure 9

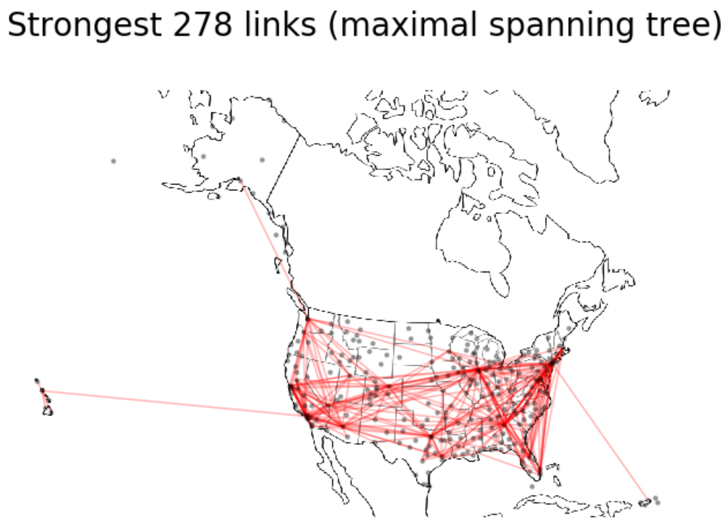

Figure 9: The network with only the strongest M links

There are only 97 links that the thresholded network share with the maximal spanning tree. From this number and the visualizations we can't really say that simple thresholding yield a similar network as the maximum spanning tree. We can take Alaska as an exmaple. In maximal spanning tree since it's required to use all the nodes to form the tree, so it can catch the local strong links. But for the thresholded network, it will only choose the top M strongest links, so it can potentially miss links in some regions that air traffic is not so frequent compared to other regions, and that's why the maximal spanning tree misses the local strong links in Alaska.