CS-E5740 - Complex Networks
Exercise set 4

Hugues Verlin (584788)
hugues.verlin@aalto.fi

December 5, 2016

1 Weight–topology correlations in social networks

1. a) Plot the complementary cumulative distribution (1-CDF) for node degree k, node strength s and link weight w

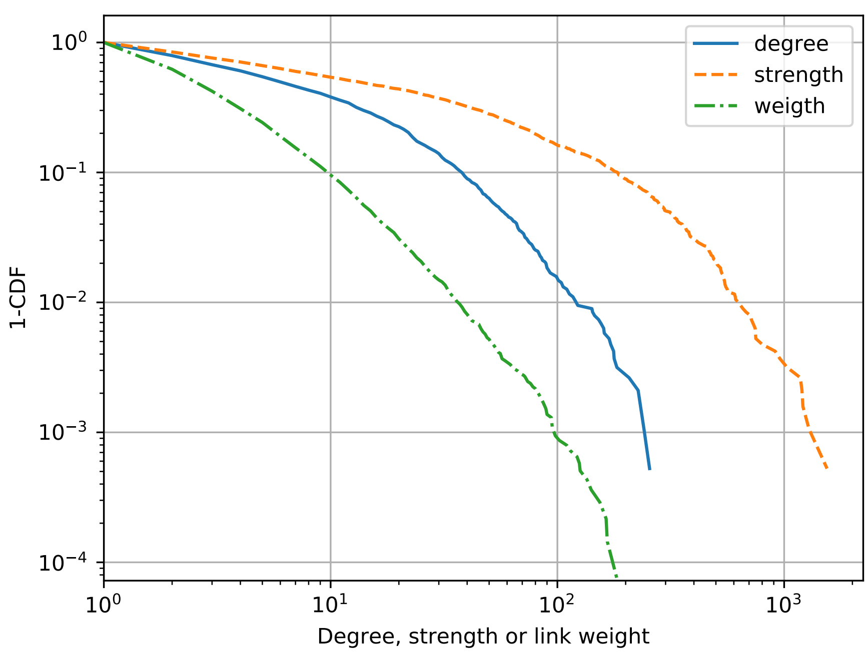

Figure 1.1: complementary cumulative distribution (1-CDF) for node degree k, node strength s and link weight w

- Briefly describe the distributions: are they Gaussian, power laws or something else?

We can see that the distributions looks like power laws with fat tails.

- Based on the plots, roughly estimate the 90th percentiles of the degree, strength, and weight distributions.

Estimates:

- degree: 40
- strenght: 190

- weight distribution: 10

1. b) Make a scatter plot that shows $\langle w \rangle$ as a function of k

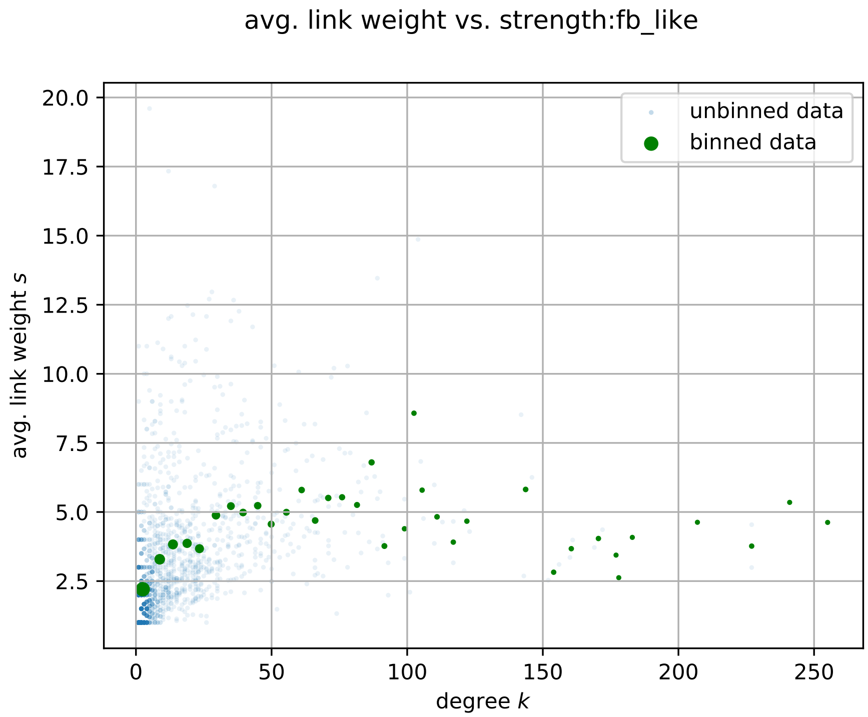

Figure 1.2: Average link weight versus degree

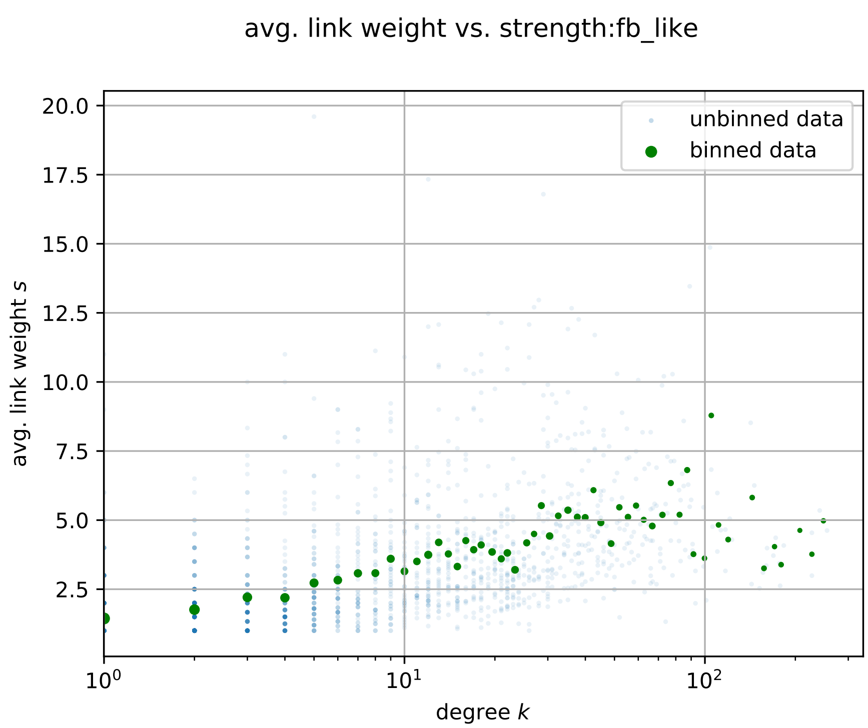

Figure 1.3: Average link weight versus degree

1. c) Questions

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

The log approche is better because it will show global trend way more easely.

- **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 conception? If not, how would you explain your finding?**

It is not that visible on the graph. There is a small decrease of the weight versus the degree at the end o f the graph, but only at the very end. Otherwise, the global trend is increasing.

The network of the exercise is not an accurate social network for small degree nodes: on Facebook for example is you have less than 10 friends, you are not using the website that much. That could explain the difference between this graph and a "real-life" social graph.

1. d) Link neighborhood overlap

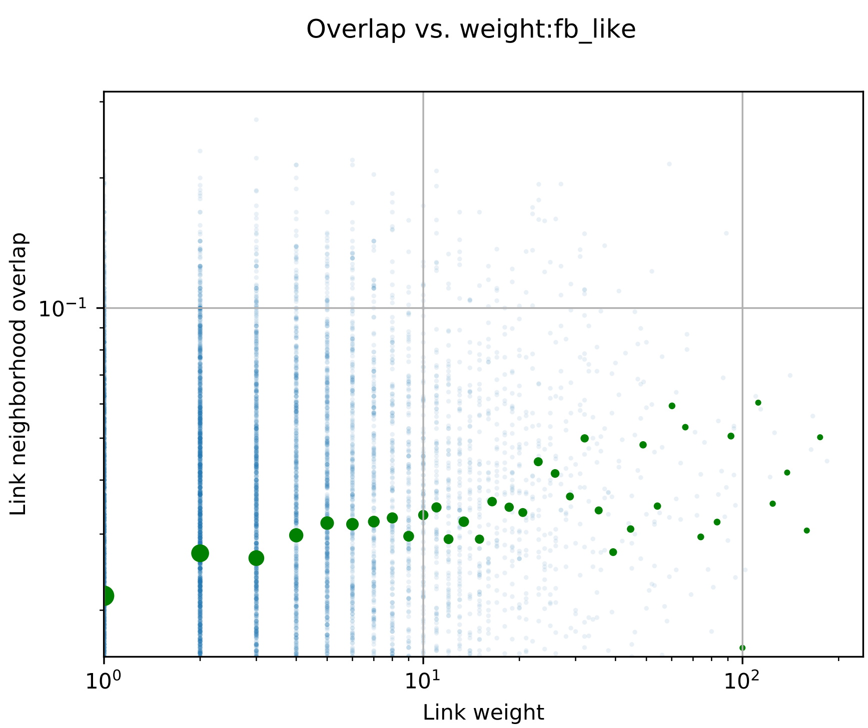

Figure 1.4: Overlap vs. weight

- **Is this trend in accordance with the Granovetter hypothesis? If not, how would you explain your findings?**

There is an increasing trend so it supports the hypothesis.

2 Error and attack tolerance of networks

2. a) Size of the largest component for different types of attacks

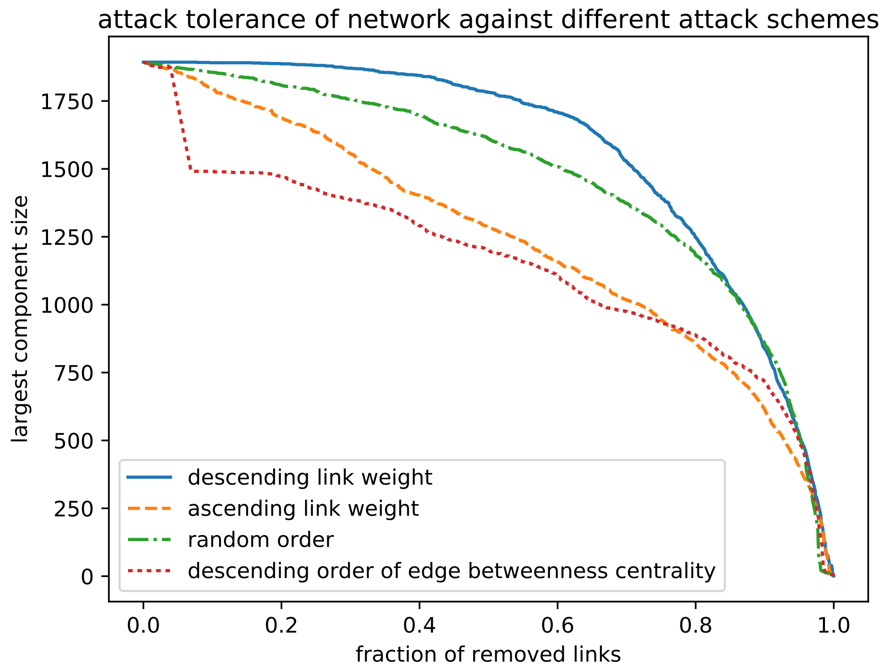

Figure 2.1: Attack tolerance of network against different attack schemes

2. b) For which of the four approaches is the network most and least vulnerable? In other words, in which case does the giant component shrink fastest / slowest? Or is this even simple to define?

- Most vunerable: descendig order of edges betweeness centrality
- Least vunerable: descending link weight

But it is not that easy to define as the network is not that vunerable against "descending order of edge betweness attack" when there is already a lot of nodes removed for example.

2. c) When comparing the removal of links in ascending and descending order strong and weak links first, which ones are more important for the integrity of the network? Why do you think this would be the case?

The most important for the integrity of the network is the weak links first as the one that have strong links have strong links between them.

2. d) How would you explain the difference between the random removal strategy and the removal in descending order of edge betweenness strategy?

If you attack using betweness you will remove the very important edge first, whereas if you attack the network randomly, it's not likely at all that you will hit an important edge at first. It could be but considering the size of the network, it is very unlikely.

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

3. a) Network properties

- Number of network nodes $N = 279$
- Number of links $L = 2088$
- Density $D = 0.0538$
- Network diameter $d = 4$
- Average clustering coefficient $C = 0779$

3. b) Visualize the full network with all links on top of the map of USA

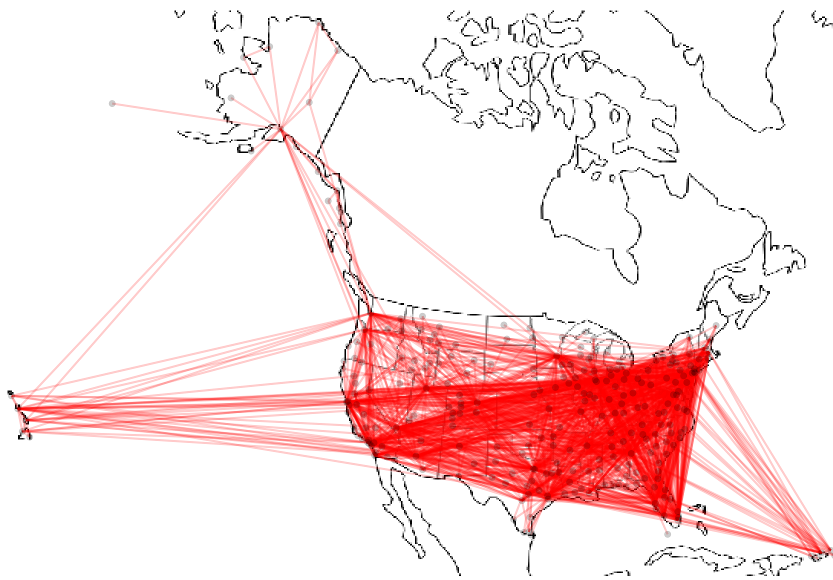

Figure 3.1: Full network with all links on top of the map of USA

3. c) ***Compute*** **both the maximal and minimal spanning tree (MST) of the network and visualize them**

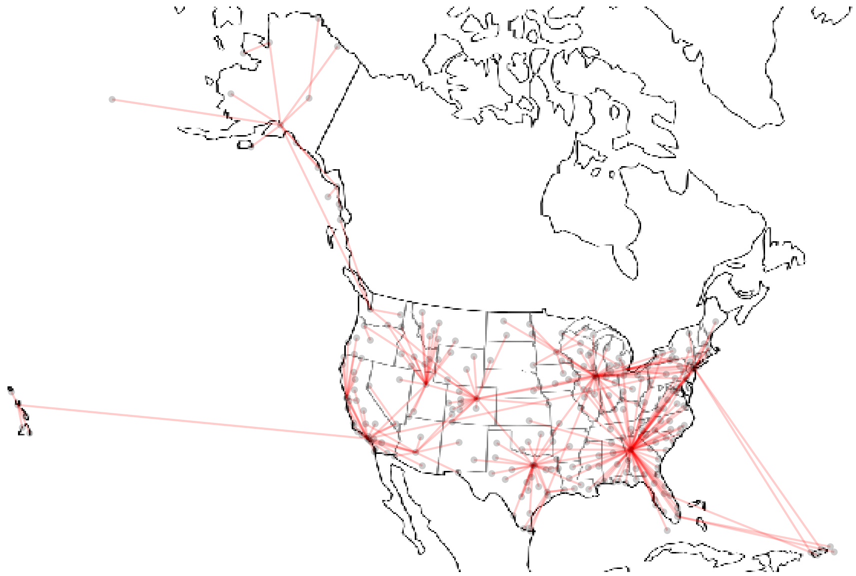

Figure 3.2: Maximum Spanning Tree

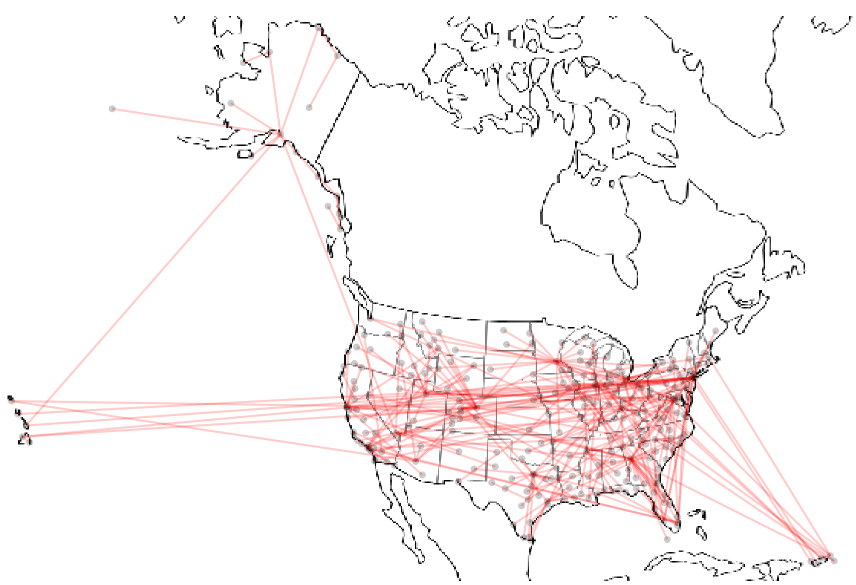

Figure 3.3: Minimum Spanning Tree

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

The minimum spanning tree show a great number of links to Hawai that have a long distance. It is because the strength of the link represents the number of flight during the time period. Indeed, there are not so many flight that go to Hawai from a such long distance. This is more likely that a flight will go to California first, and then to Hawai. This is what we can see using the maximum spanning tree.

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

I would use the maximum spanning tree as we can clearly see the hubs and the general directions of the air traffic. It is a lot more logical and meaningful.

3. d) Threshold and visualize the network by taking only the strongest M links into account, where M is the number of links in the MST

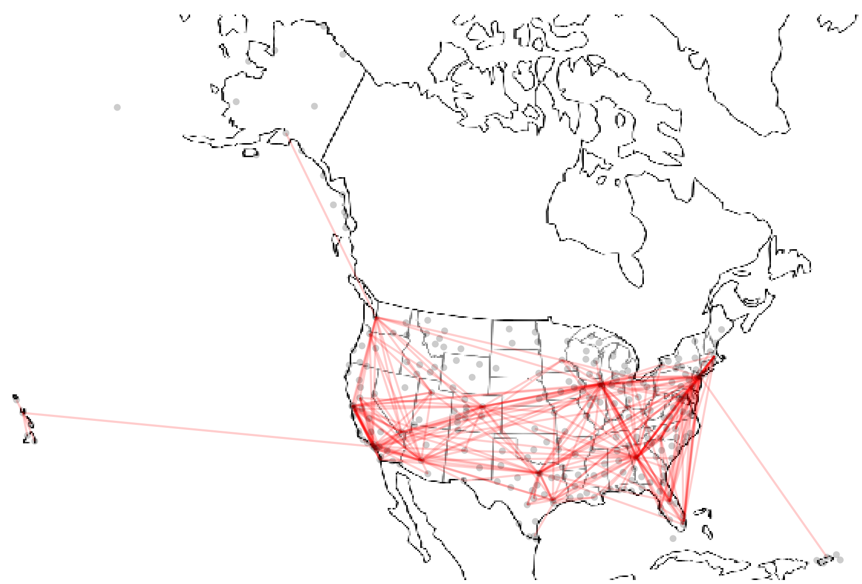

Figure 3.4: Threshold network with only the top 279 strongest edges

- **How many links does the thresholded network share with the maximal spanning tree?**

The intersection of the sets of edges of each network gives 98.

- **Given this number and the visualizations, does simple thresholding yield a similar network as the maximum spanning tree?**

Not exactly as only 35% of the edges are here in both cases.