Assignment 5

Fredrik Mattisson

May 10, 2020

For this assignment, I have chosen as my real world example network the western United States power grid. The data set was downloaded from `https://networkrepository.com`
I chose this because I think that both fitting models to it and assessing its robustness has real world utility. The network has 4941 vertices and 6594 edges, and is shown in fig. 1 along with its degree distribution. The distribution is somewhat similar to a Poisson distribution.

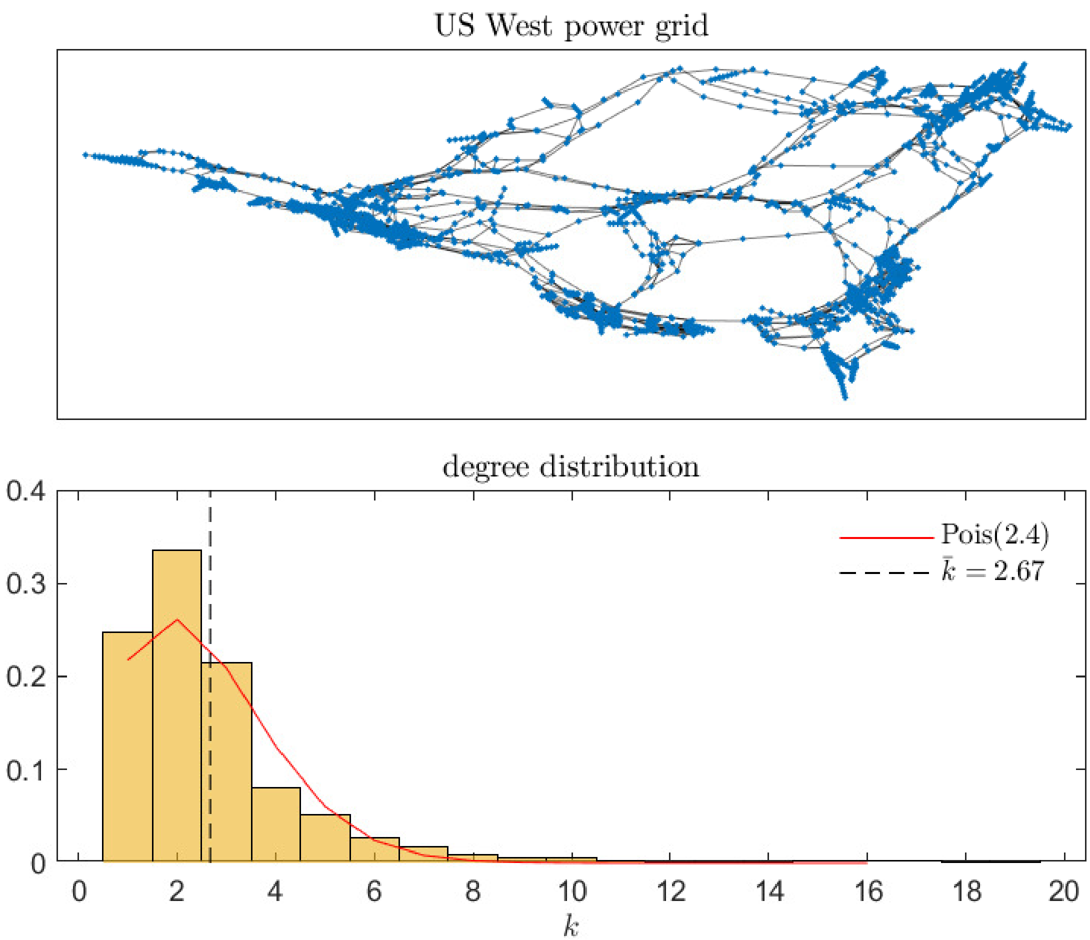

Figure 1: Graph plot and degree distribution of western US power grid. $N = 4941$, $E = 6594$

1 Null models

For an Erdös-Renyi random graph, we have $\mathbb{E}[E] = \binom{N}{2}p$. Due to the nature of the example network, we would also like the model to be connected. If we have $Np > 1$, we will typically get a large connected component containing a huge proportion of the vertices. Using that component as the model allows us generate a connected graph that is approximately the same size as the example, without restricting $p > \ln(N)N^{-1}$.

Taking $N = 5500$ yields $p \approx 0.000436$, given that the example has 6594 edges. With those parameters, I generate 30 random graphs, extract the largest connected component, and compute the average degree, global clustering coefficient, density and average shortest path length. The results are shown in fig. 2 and table 1. The average degree of the example is slightly below the 25th percentile of the sample, and the example density slightly below the sample minimum. As for global clustering coefficient and average shortest path length, the example has values well above the sample maximum. The median avg. shortest path is almost exactly half that of the example. Note that I have omitted the example values from the box plots in the in the cases where they were too far from the sample values.

The results are somewhat similar for the configuration model. As with the Erdös-Renyi samples, I have considered the largest connected component. This time both the example density and average degree are below the sample minima, and the discrepancy is larger than for the Erdös-Renyi samples. The global clustering coefficient is much higher in these samples compared to Erdös-Renyi, however the exaple value is still an order of magnitude larger. These results are shown in fig. 3 and table 2.

Overall, I would say that the Erdös-Renyi sample fits the example better. Both models produced degree distributions very similar to the example. The sizes of the discarded components from the samples were in both cases on the order of 1 – 8 vertices.

metric	min	median	max	example
N	4765	4840	4924	4941
E	6294	6517	6860	6594
$\bar{k}$	2.638	2.68	2.725	2.669
density	$5.44 \cdot 10^{-4}$	$5.57 \cdot 10^{-4}$	$5.67 \cdot 10^{-4}$	$5.4 \cdot 10^{-4}$
gcc	0	$1.98 \cdot 10^{-4}$	$1.1 \cdot 10^{-3}$	0.0935
spl	9.314	9.55	9.736	18.989

Table 1: Statistics for a sample of 30 Erdös-Renyi random graphs. Plotted in fig. 2.

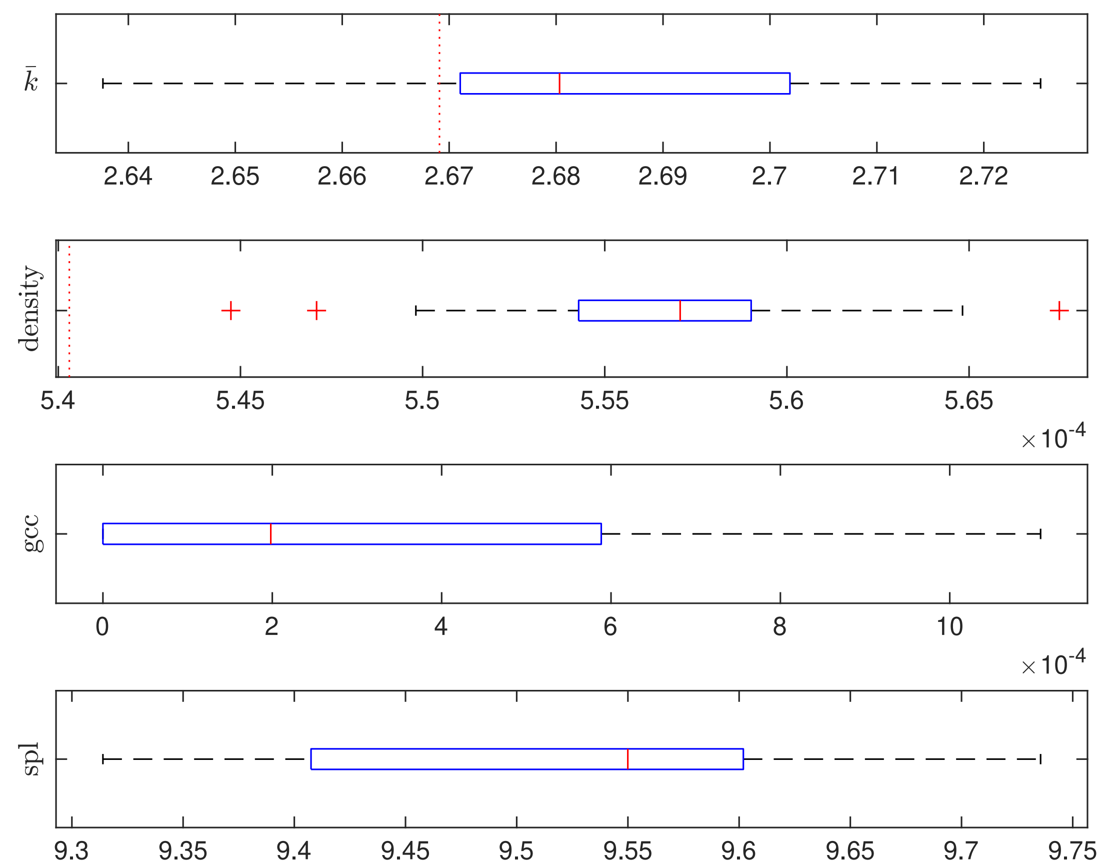

Figure 2: Statistics for a sample of 30 Erdös-Renyi random graphs. The plots show: average degree, density, global clustering coefficient, average shortest path length
The dotted red lines in the first two plots show the corresponding metric for the example network.
A comparison of these statistics to the example network is shown in table 1.

metric	min	median	max	example
N	4554	4613	4660	4941
E	6342	6384	6420	6594
$\bar{k}$	2.755	2.762	2.78	2.669
density	$5.91 \cdot 10^{-4}$	$5.955 \cdot 10^{-4}$	$6.07 \cdot 10^{-4}$	$5.4 \cdot 10^{-4}$
gcc	$3.86 \cdot 10^{-3}$	$6.4 \cdot 10^{-3}$	$7.776 \cdot 10^{-3}$	0.0935
spl	8.369	8.396	8.426	18.989

Table 2: Statistics for a sample of 30 configuration model graphs. Plotted in fig. 3.

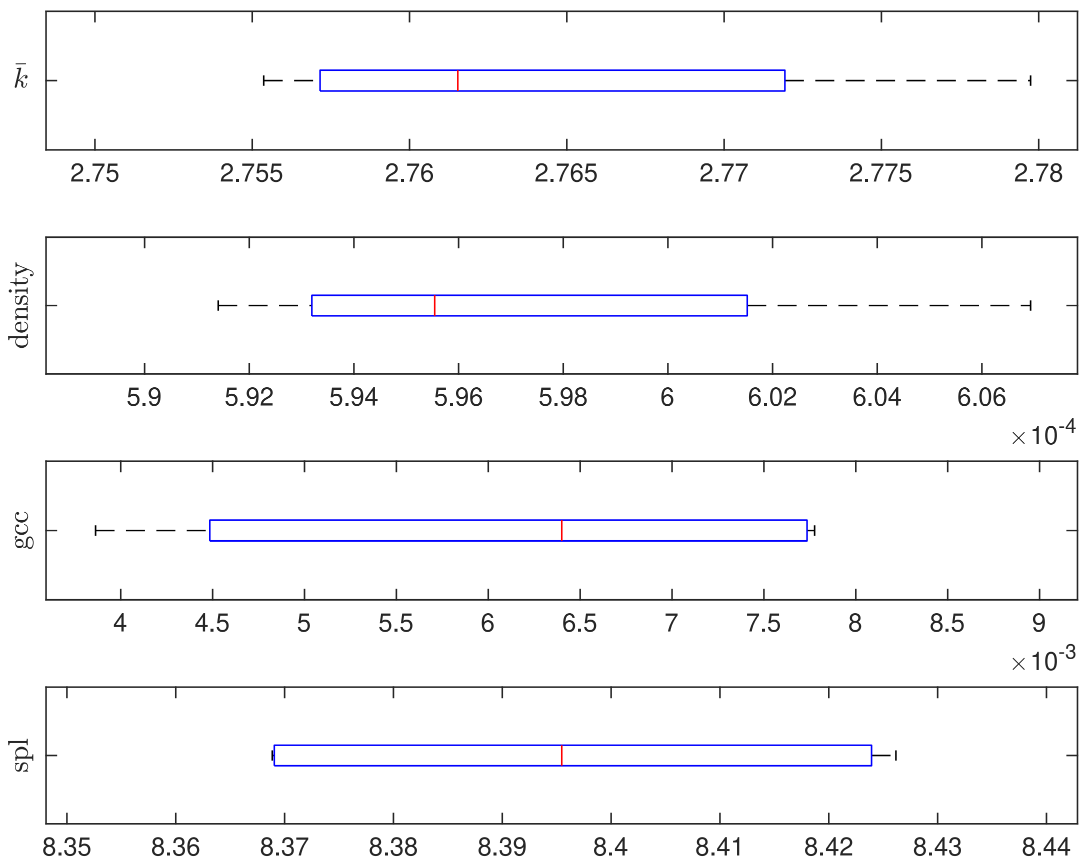

Figure 3: Statistics for a sample of 30 configuration model graphs. The plots show: average degree, density, global clustering coefficient, average shortest path length
A comparison of these statistics to the example network is shown in table 2.

2 Robustness

For these three simulations, I have used the Erdös-Renyi model with the parameters from 1. That is, I generate random graphs with $N = 5500$ and $p = 0.000436$. The third simulation is run exactly like the second, but nodes are removed in descending order of betweenness instead of degree. For the random percolation, I run the algorithm 10 times on one instance of the random graph. For the deterministic percolations (degree sequence and betweenness), I generate 10 random graphs.

The results are shown in fig. 4, fig. 5 and fig. 6. As one might expect, removing nodes with high degree or betweenness quickly has a large impact on a power grid. In all three cases, the random graph fails to accurately model the effect of percolation on the example, although the shapes of the curves are similar in the degree sequence case.

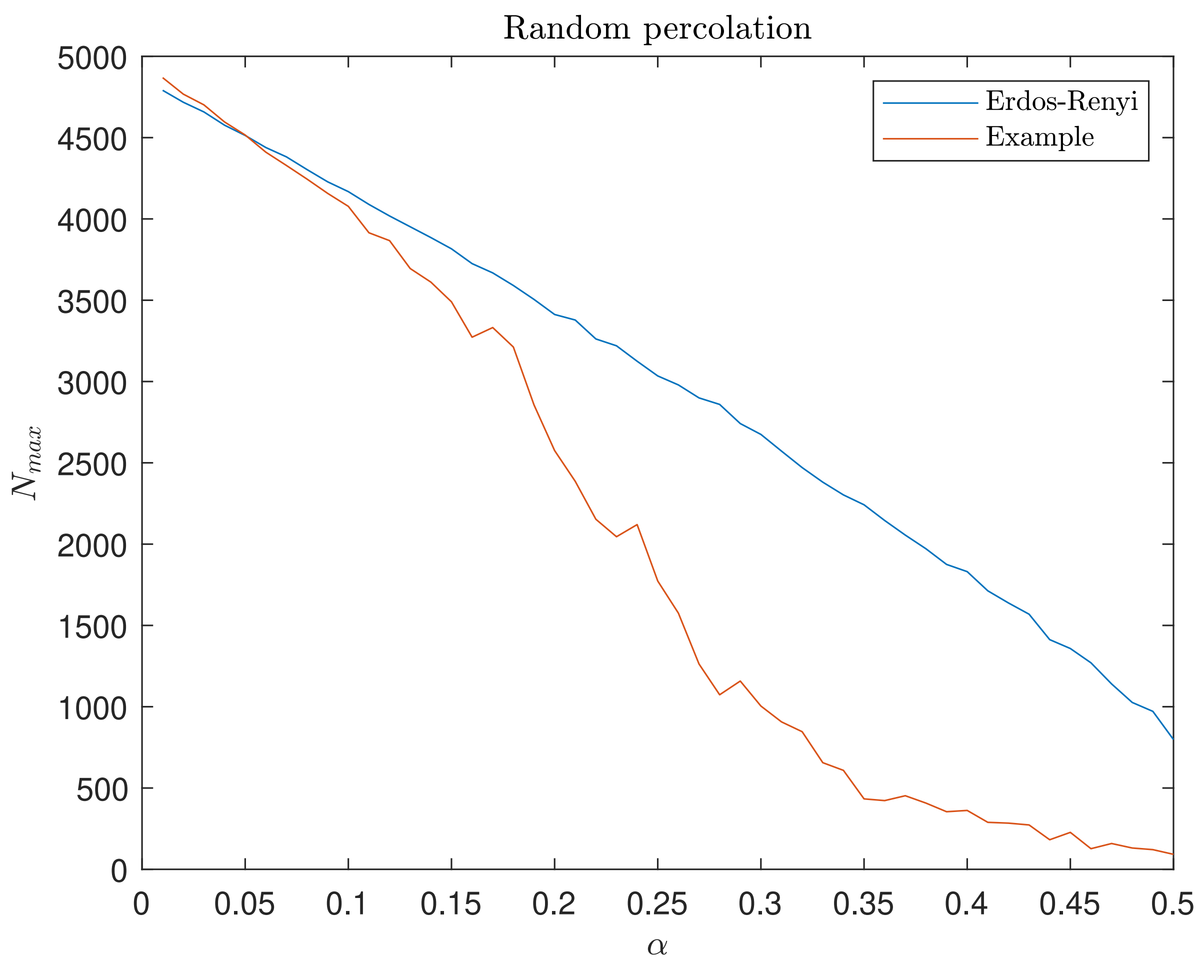

Figure 4: Average size of largest connected component over 10 trials of removing a proportion α of the nodes, selected uniformly randomly.

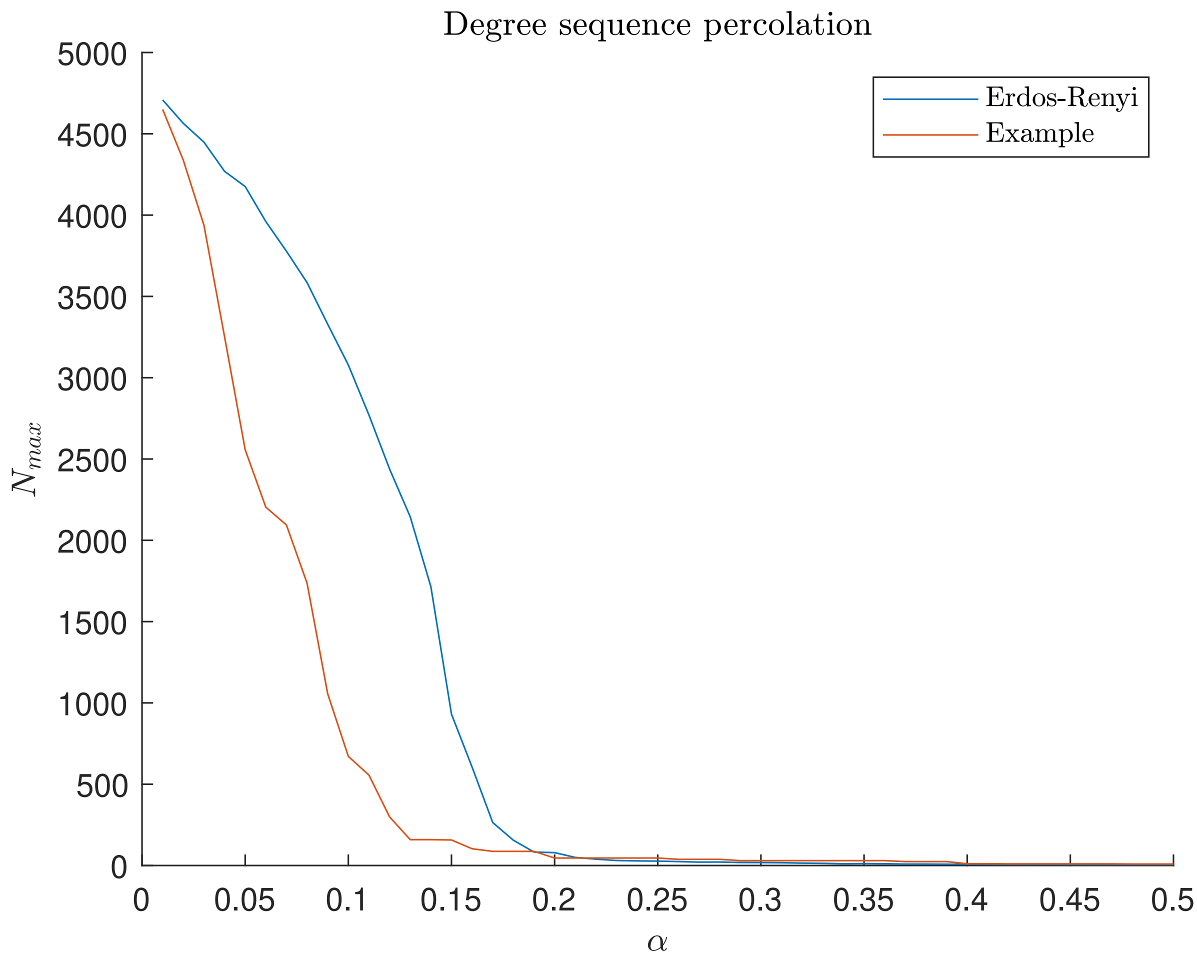

Figure 5: Average size of largest connected component after removing a proportion α of the nodes in degree sequence, from a sample of 10 random graphs.

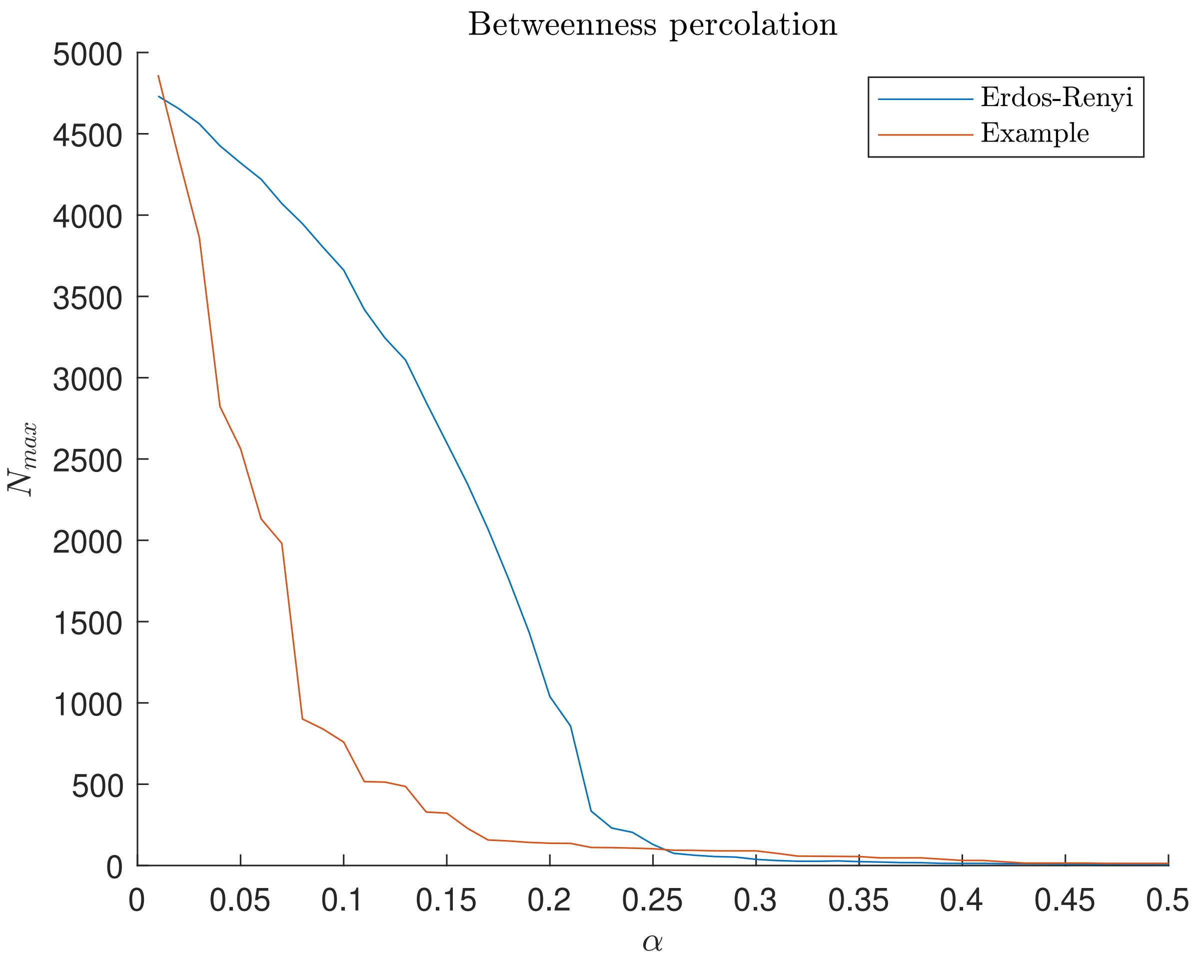

Figure 6: Average size of largest connected component after removing a proportion α of the nodes in descending order of betweenness, from a sample of 10 random graphs.

A Code

All the code can be checked out at
`https://github.com/Fredrik-M/BERV-MCS/tree/master/MCS/lab_5`