Introduction to Data Analytics Serie 4.

Exercise 1. Network analysis.

In this exercise, we generate (i) Erdős-Rényi and (ii) Barabási-Albert networks and analyze some of their properties.

To build an Erdős-Rényi network, begin with a set of N isolated nodes and iterate over all possible node pairs. Two nodes are connected with probability p.

Barabási-Albert networks are constructed via preferential attachment after starting from a connected network with m_0 nodes (e.g., a dyad). Preferential attachment means that a new node is connected to $m \le m_0$ existing nodes and the probability p_i of being connected to an existing node i is proportional to the degree k_i of that node (i.e., $p_i = k_i / \sum_i k_j$).

- (a) You can start with networks that have N=100 nodes and later extend your analyses to networks with 1000 and 10000 nodes. Vary the connection probability p for Erdős-Rényi networks and the parameter m for Barabási-Albert networks. What is the influence of these variations on the degree distribution?
- (b) Implement Dijkstra's algorithm and determine the average path length between two nodes for networks (i) and (ii) as a function of the number of nodes N. What do you observe?

Exercise 2. Visualizing networks with graph-tool.

This exercise will provide a short introduction to working with the network analysis and visualization library graph-tool. (A commonly used alternative is the library NetworkX.)

Hint. A graph-tool tutorial can be found at https://graph-tool.skewed.de/static/doc/quickstart.html.

(a) Create a Delaunay triangulation network and its minimum spanning tree using the following commands.

```
g, pos = triangulation(random((500, 2)) * 4, type="delaunay")
tree = min_spanning_tree(g)
```

- (b) Visualize the resulting network that contains only those nodes that belong to the minimum spanning tree.
- (c) Compute geodesic edge and node betweennness centralities and again visualize the network. Color nodes and use edge widths according to their centralities. (Edge with large centrality should have a large edge width in your plots.)