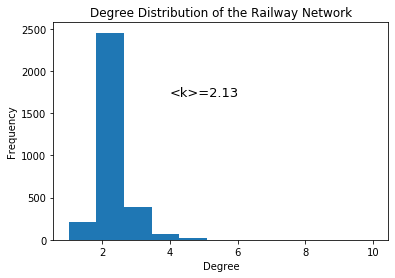
# 3 Dataset

The Dataset analysed in this research represents the Swiss Railway and Tram Network. It contains 3’190 Railway and Tram stations all over Switzerland with 3'353 connections between them. While creating the graph fort he network, we assigned the names of the stations as well as their location coordinates to the nodes.

Some of the bigger stations are subdivided into several data points, for example the station «Interlaken Ost» is sectioned by platforms («Platform 1-2», «Platform 3-4», «Platform 5-8»). This leads to the existence of nodes, whose spatial distance is rather small nevertheless they are not connected. The network is divided in 47 connected components, from which the giant connected component includes 1650 nodes, which is around 53% of the whole network. There are four nodes, which are not connected to any other node in the network. These nodes were excluded for the analysis of the network.

Because of the fact that the network is divided in many connected components and the existence of close but not connected nodes, we decided to take modifications in the network. Nodes whose spatial distance was less than 300 meters and no path between them existed, were merged to one node that includes the connections of all the merged nodes. This modification relies on the assumption that these nodes belong to the same station or are at least located in walking distance. This results in a network with 3'146 nodes and 3'351 edges. The number of connected components is reduced to five, from which the giant connected component contains 99% of all nodes in the network (3’116/3’146).



The emergence of the Railway network characterizes its degree distribution. There are few nodes with a relatively high degree, which represent the biggest traffic hubs, which are located in the economically most important cities of Switzerland (e.q. Zurich and Basel). Most nodes do have two connections to other stations. This follows the characteristic that most routes are straight lines going through nodes and ending in a node with degree one. This also explains the low clustering discovered in the network (C = 0.016). This distribution results in an average degree close to two (<k> = 2.13).