# 3 Dataset and Data Preparation

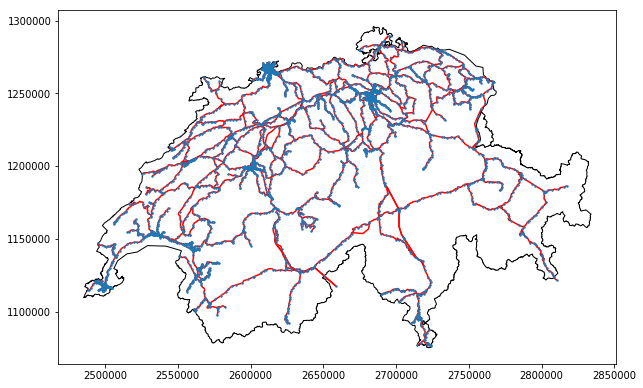
The Dataset analyzed in this research represents the Swiss Railway and Tram Network. It contains 3’190 railway and tram stations all over Switzerland with 3'353 connections between them. The data set we used was provided by geo.admin.ch the official geo information portal of the swiss confederacy. The data is from the 20.12.2017 and includes GIS information and data about the different public transport companies in Switzerland. [1]

The data about the railway stations included an id, a station name and geographic data, as shown in table 1. The connection data included an id, a start and end node, track operator, electrification, track width, distance and geographic data. All stations and connections mapped on Switzerland can be seen in figure 1.

Table 1: Railway station data.

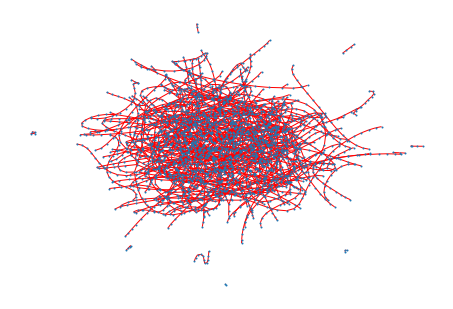
|  |  |  |  |
| --- | --- | --- | --- |
|  | **xtf\_id** | **Betriebspunkt\_Name** | **geometry** |
| **0** | ch14uvag00092584 | Wabern, Eichholz | POINT (2600987.93 1197507.6371) |
| **1** | ch14uvag00092599 | Bern, Sandrain | POINT (2600494.741999999 1197843.540100001) |
| **2** | ch14uvag00092576 | Basel, Rheingasse | POINT (2611516.495999999 1267823.809999999) |
| **3** | ch14uvag00092992 | Zürich, Schiffbau | POINT (2681631.868000001 1249127.912) |
| **4** | ch14uvag00092591 | Bern, Kursaal | POINT (2600835.989999998 1200203.734000001) |

Figure 1: Railway and tram network mapped on Switzerland.



While creating the graph for the network, we assigned the names of the stations as well as their location coordinates to the nodes. After that we added all the edges connecting to their respective stations. A visualization of this network can be seen in Figure 2.

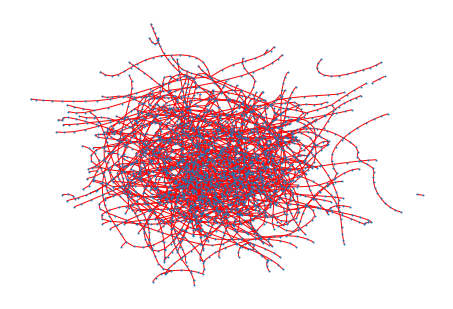
Figure 2: Railway and tram network of Switzerland not geographically mapped.



Before starting with our analysis some minor data cleaning and preparation had to be done. 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»). Also, some tram and independently managed train routes are not connected to the main network because stations that are right next to each other used different nodes («Bern SBB», «Bern RBS»). This leads to the existence of nodes, which are geographically located next to each other, but nevertheless are not connected. The network is divided in 47 connected components, from which the biggest 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.

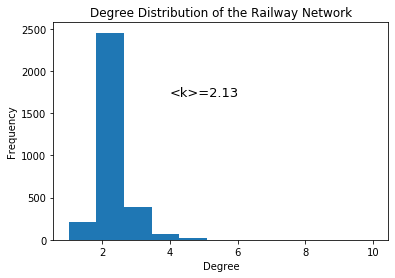
Because the network is divided in many connected components and the existence of close but not connected nodes, we decided to make some modifications to the network. Nodes whose spatial distance was less than 300 meters and no path between them existed, were merged to one node that includes the connections 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 close 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). This new better-connected network can be seen in Figure 3.

Figure 3: Modified railway network with only 5 connected components.



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, they are located in the economically most important cities of Switzerland (e.q. Basel, Bern, Geneva and Zurich). Most nodes do have only two connections to other stations. This follows the characteristic that most railway lines are going through stations in a sequenced fashion, with some start and end nodes with a degree of one and some connecting stations with a higher degree. The degree distribution can be seen in Figure 4. 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).

Figure 4: Degree distribution of the swiss railway network.



[1]

https://data.geo.admin.ch/ch.bav.schienennetz/