



Final Project: Network scientific analysis about the vulnerability of the Swiss Railway network

Mathias Lüthi (15-707-037), Nino Scherrer (12-729-216), Peter Giger (14-915-383)

Network Science for Business, Economics, Informatics and Social Sciences

Faculty of Business, Economics and Informatics

09.12.2019

ABSTRACT

Railway networks are of great importance for every economy (**Resilience**). Thus, it is all the more important to understand the reasons why and how they can fail. Yet, due to the size and complexity, their susceptibility to failures is not completely understood (**Resilience**). This research project investigates the vulnerability of the Swiss railway network using network scientific models.

1 INTRODUCTION

Railway systems need to perform well even under suboptimal circumstances (**Resilience**). Delays and cancellations are daily challenges for both passengers and railway companies (**Resilience**). Existing research showed that the network topology has an effect on the performance of a railway system (**ComplexTopology**). More recently, **Resilience** analyzed the resilience and robustness of a UK railway network. They showed that poor performance correlates more with cascade effects than failures (**Resilience**).

Needless to say, there are many reasons why railway systems can fail, but network effects are one of the less studied subjects. For this reason, the focus of this study is the Swiss railway system for which only few studies exist. In 2008, the authors **GraphSwiss** published a graph-theoretical analysis of the Swiss railway network over time. They have used measures such as degree/betweenness centrality to characterize the growth of the Railway network (**GraphSwiss**). However, they have suggested further research to explain the robustness of the network. In 2009, the authors **VulnerabilitySwiss** proposed a generalized linear model (GLM) to assess the vulnerability of the Swiss road network. The downside of their approach is that they assumed the failures to be mutually exclusive (**VulnerabilitySwiss**).

This research project extends the work of **GraphSwiss** and **VulnerabilitySwiss** using network scientific models. The primary focus is to analyze the Swiss railway network regarding graph characteristics and vulnerabilities.

1.1 Transport Networks

Transport networks are essential for people and economy (**GraphSwiss**). Early research on transport networks focused on geometric and topological properties (**GraphSwiss**). Later, with the availability of computational power, the research shifted towards network structures (**GraphSwiss**). Recently, papers such as **Resilience** and **ComplexDelay** used graph properties to model robustness and delay dynamics.

There are several differences to other networks such as social networks (**GraphSwiss**). Transport networks represent physical objects such as train tracks and railway stations (**GraphSwiss**). These physical objects have constraints themselves which influences the degree distribution (**GraphSwiss**). Additionally, monetary constraints apply as well, thus restrict the ability for them to be scale-free (**GraphSwiss**).

1.2 The Swiss Railway Network

According to the "European Railway Performance Index", Switzerland has one of the best performing railways (**RailwayPerformanceIndex**). Yet, the increased load over the past years is a constant challenge. Every delay or cancellation might result in cascade failures.

Every railway network is slightly different. This is due to their history and how they grew over the years (**GraphSwiss**). The Swiss railway network is special because its early growth was purely driven by economic values (**GraphSwiss**). This means that cities were prioritized and urban areas were not considered in the planning process (**GraphSwiss**). Moreover, due to competition of private companies, the networks grew more or less independent of each other (**GraphSwiss**). Today, most railways belong to the Swiss Federal Railways (SBB).

2 MEASURES

This section describes the basic measures used in this research project.

2.1 Degree Centrality

The degree centrality is a fundamental measure based on the degree of the node (**ComplexTopology**). The degree represents the number of connections a node has (**ComplexTopology**). Hence, the degree centrality assumes that nodes with a large number of connections are more important (**ComplexTopology**). It was first introduced by **Freeman** and **Freeman2**:

$$C_i^D = \frac{k_i}{N - 1} \quad (1)$$

where

$$k_i = \sum_{j \in N} a_{ij} \quad (2)$$

In railway networks, the degree centrality is constrained by spacial and economic limits (**GraphSwiss**).

2.2 Betweenness Centrality

The betweenness centrality is the number of shortest paths passing through the node (**ComplexTopology**). It was introduced by **Freeman** and **Freeman2** as well:

$$C_i^B = \sum_{j \neq k} \frac{n_{jk}(i)}{n_{jk}} \quad (3)$$

where, for the nodes j and k , the shortest path is defined by $n_{jk}(i)$. In transportation networks, the betweenness centrality can be used to measure the impact of a node. (**ComplexTopology**)

2.3 Closeness Centrality

The closeness centrality measures how close the node is to all other nodes (**ComplexTopology**). It is therefore the inverse of the mean shortest path to all other nodes and defined as: (**ComplexTopology**)

$$C_i^C = \frac{1}{\sum_{j \neq i} d_{ij}} \quad (4)$$

The closeness centrality depends on the geographic position of the nodes as well as the size of the network (**GraphSwiss**). This restricts the comparability of different sized networks (**GraphSwiss**).

2.4 Eigenvector Centrality

Eigenvector centrality is an extension of the degree centrality. Other than in the degree centrality measure, eigenvector centrality also accounts for the importance of the adjacent nodes. This means not all connections in the network are equal. Eigenvector centrality is calculated by assessing how well connected an individual is to the part of the network, which is also highly connected. This can be written as the Eigenvector equation (**GraphSwiss**)

$$v_i = \frac{1}{\lambda} \sum_{j=1}^n A_{ij} v_j \quad (5)$$

2.5 Edge Betweenness Centrality

The edge betweenness centrality is defined as the number of the shortest paths that go through an edge in a network. A high edge betweenness centrality score represents a bridge-like connector between two parts of the graph. The removal of such a node might disconnect these two parts (**Betweenness**)

3 DATA

3.1 Data

The data set analyzed in this research project 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 set is a snapshot from the 20.12.2017 and includes GIS information and data about the different public transport companies in Switzerland. (**Dataset**)

The data about the railway stations includes 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.

xtf-id	Betriebspunkt-Name	geometry
ch14uvag00092584	Wabern, Eichholz	POINT (2600987.93 1197507.6371)
ch14uvag00092599	Bern, Sandrain	POINT (2600494.741999999 1197843.540100001)
ch14uvag00092576	Basel, Rheingasse	POINT (2611516.495999999 1267823.809999999)
ch14uvag00092992	Zürich, Schiffbau	POINT (2681631.868000001 1249127.912)
ch14uvag00092591	Bern, Kursaal	POINT (2600835.989999998 1200203.734000001)

Table 1. Railway station data

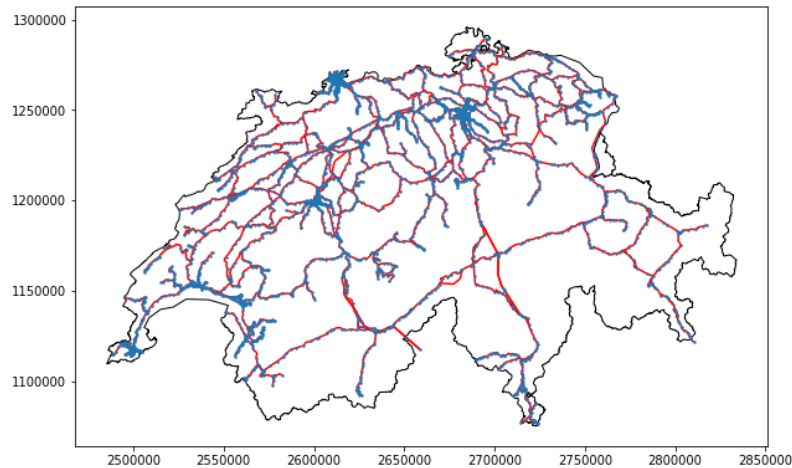


Fig. 1: Railway and tram network mapped on Switzerland

3.2 Data Preparation

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.

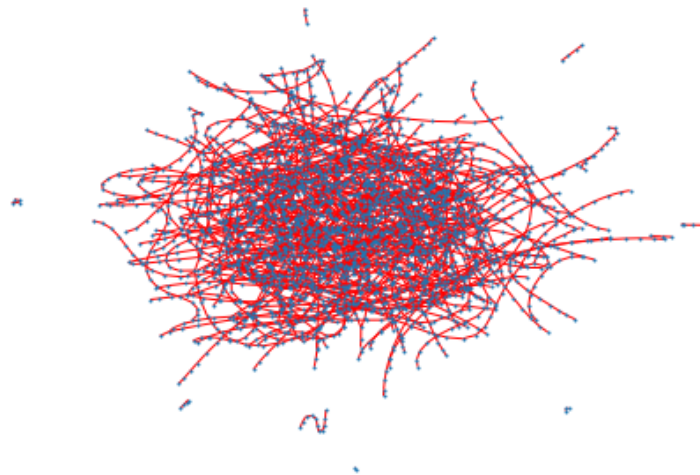


Fig. 2: Railway and tram network of Switzerland not geographically mapped

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 3'139 nodes and 3'344 edges. The number of connected components is reduced to five, from which the giant connected component contains 99% of all nodes in the network (3'109/3'139). This new better-connected network can be seen in Figure 3.

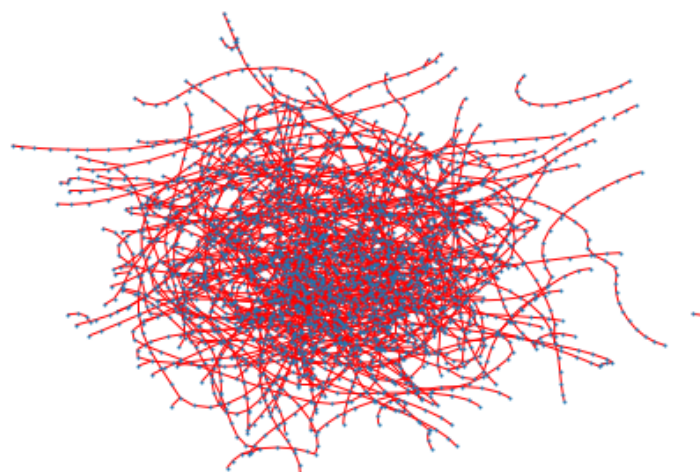


Fig. 3: Modified railway network with only 5 connected components

3.3 Characteristics Of The Network

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.g. 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 ($\langle k \rangle = 2.13$).

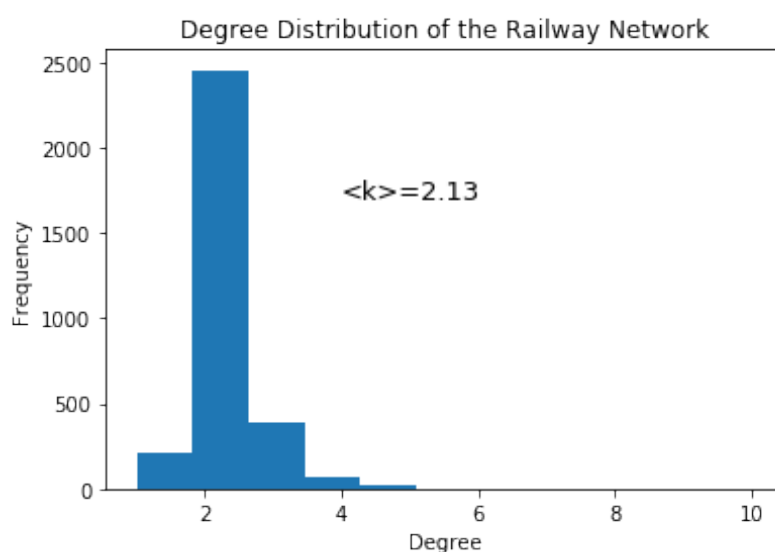


Fig. 4: Degree distribution of the swiss railway network

4 ANALYSIS

The goal of our analysis is to test the robustness of the Swiss Railway network, or in other words, we measure the connectedness of the network in case of failures or attacks on the network. For this purpose we remove nodes or edges from the network using random or targeted attacks. This simulates failures or disasters on certain routes or in certain train stations. Then we observe the effects of this simulation on the development of the biggest connected component in the network.

4.1 Random And Targeted Attacks

Testing the robustness of the network by removing random nodes shows the stability of the network. When removing nodes randomly one needs to remove over 200 nodes to half the size of the giant component. When removing over 500 nodes, the network is almost completely broken down and less than 10% of the network is connected in one giant component. A visualization of the random attack is provided in Figure 5.

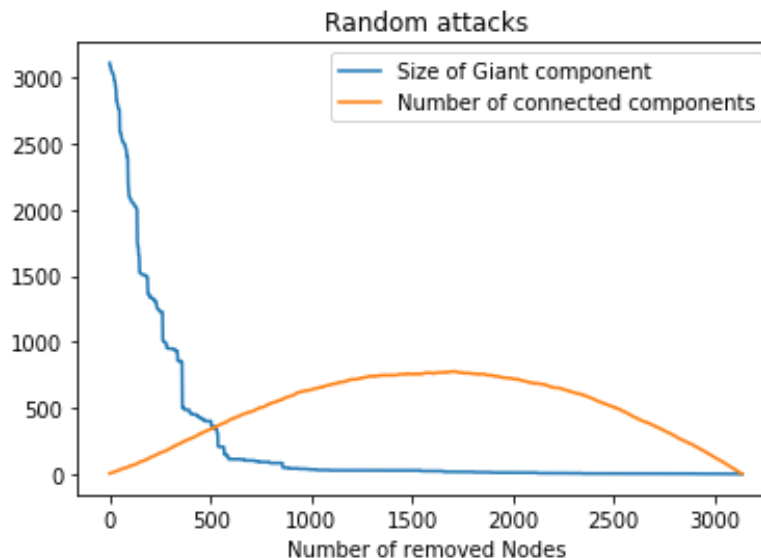


Fig. 5: Random attacks on the Swiss Railway network

Comparing this to targeted attacks on the network, as seen in Figure 6. The giant connected component breaks down a lot faster. After removing the 100 most central nodes, based on degree centrality, the biggest connected component is only about one sixth of the giant connected component from the original network.

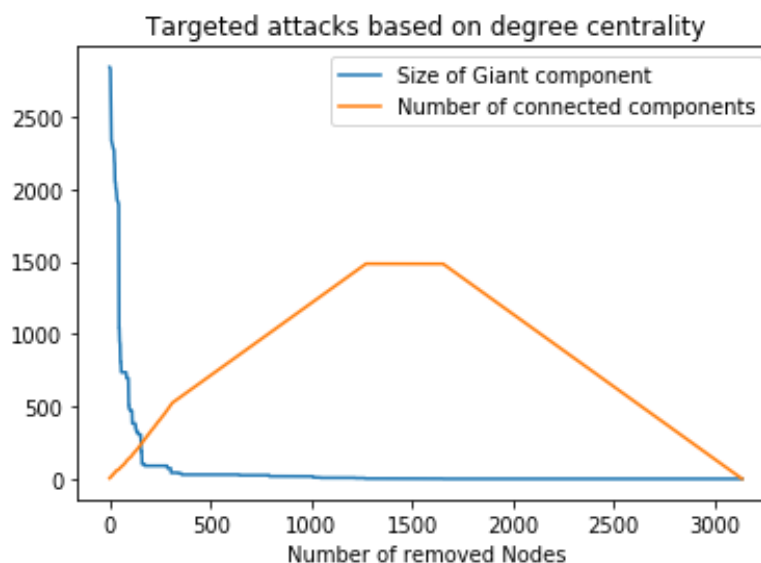


Fig. 6: Targeted attacks on the Swiss Railway network

4.2 Different Targeted Node Attacks

If we compare different centrality measures on the network, we find very different results. Degree centrality and eigenvector centrality show “Zürich, Altstetten” and “Zürich, Langstrasse” as the most central nodes. Both of those are very close together and in close proximity of the “Zurich, Main station”, which is the biggest train station in Switzerland and one of the most frequented train stations of the world. Compared to those measures the most central node according to the closeness centrality measure is the station “Olten, Bhf”, which lies in the center of the rail network and right in the center of the cities of Zurich, Bern and Basel. The most central node, according to the betweenness centrality, is the station “Wanzwil” near Herzogenbuchsee. This is a tiny station that does not hold much significance as a station but it lies on one of the most important geographic routes through Switzerland.

Comparing the importance of these different measures, we simulated a targeted attack on the network by removing 100 nodes according to these different centrality measures. As seen in Figure 7, removing nodes based on the betweenness centrality was the most effective way of breaking down the giant connected component. Followed by the strategy of removing the nodes based on closeness, degree and eigenvector centrality.

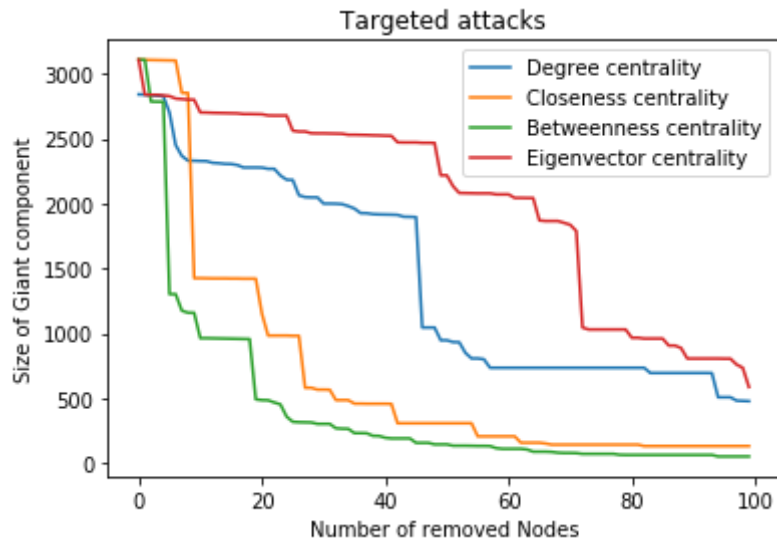


Fig. 7: Targeted attacks on 100 nodes of the Swiss Railway network

4.3 Attacks On Edges

While failures that break down an entire train station and stop any trains from passing through it do happen, they are rarer than failures on a certain track that stop all traffic from using that connection. This leads us to also simulate the failures of edges on the railway network. As before with the attacks on nodes, we also simulated random and targeted attacks on the edges of the network. As a measure of centrality we used the edge betweenness centrality and similar to the previous simulation the targeted attacks were much more effective in breaking down the giant connected component. Figure 8 shows the results of a random edge attack on the railway network. Figure 9 compares the effectiveness of removing random edges to removing edges based on the edge betweenness centrality.

Targeted edge attacks in particular show that after only removing the 20 most important edges, the biggest connected component is only about one sixth of what it was before the attack. This means that failures on the most important routes have the biggest impact on the overall connectedness of the network in our analysis.

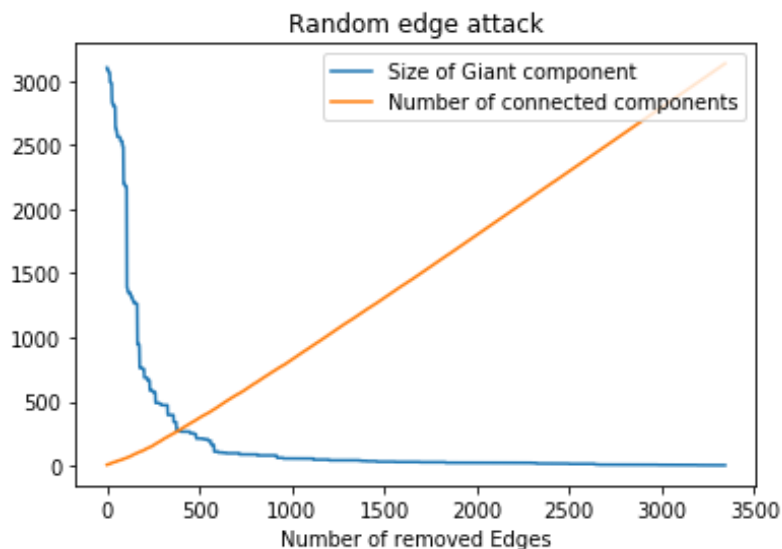


Fig. 8: Random attacks on edges instead of nodes

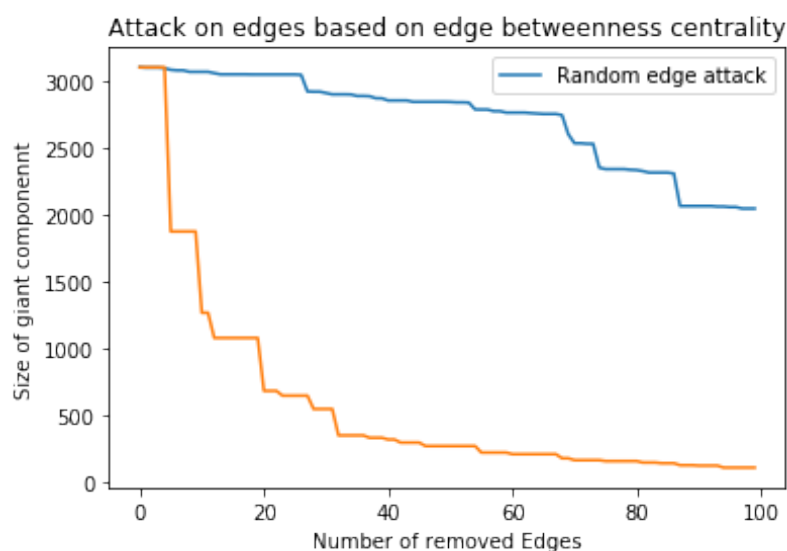


Fig. 9: Targeted vs random attacks on 100 nodes of the network

4.4 Comparison To Random Networks

Random graph models are frequently used to predict the behavior of networks with pretended characteristics. These characteristics are, for example, the degree distribution or the global clustering in the network. To compare the behavior of the Swiss railway network to failures, we used an Erdős-Renyi (ER) random model as well as a Barabási-Albert (BA) random model.

When creating an ER random model, a graph with a given number of nodes is generated. Between every pair of nodes, with probability p , an edge is added to the graph. These randomly generated models are characterized by a degree distribution, which follows a Poisson distribution with $\langle k \rangle = n * p$ as well as a clustering coefficient close to the edge creation probability p (**Barabasi**)

The degree distribution of many networks observed in reality does not follow a Poisson distribution. Therefore the need for random models with different characteristics arises. Often observed networks follow a power-law distribution of node degrees. The power-law distribution is characterized by the existence of a very high number

of low-degree nodes and the existence of few nodes with very high degree (**Barabasi**) Since the Swiss Railway network has a high number of nodes with a degree equal to two and only a few nodes that have a degree up to 7, it might be worth to compare the network to a random graph following a power-law distribution. An example of such a random graph is the BA model. The nodes in the BA random graph are created one after another and every new node is connected to a given amount of existing nodes, where nodes with higher degree are preferred (**Barabasi**)

The theoretical characteristics of the two random graph models suggests that ER graphs are more vulnerable to random attacks than BA graphs. This is because when the nodes fail randomly, the probability is high, that a low degree node fails in the BA graph. As you can see in Figure 10, the giant component of the ER graph decreases much faster with random failures than in the BA graph. Nevertheless, the Railway network seems to be more vulnerable than the random models to random failures. The same vulnerability ranking is true for targeted attacks on nodes based on the degree centrality, as seen in Figure 11.

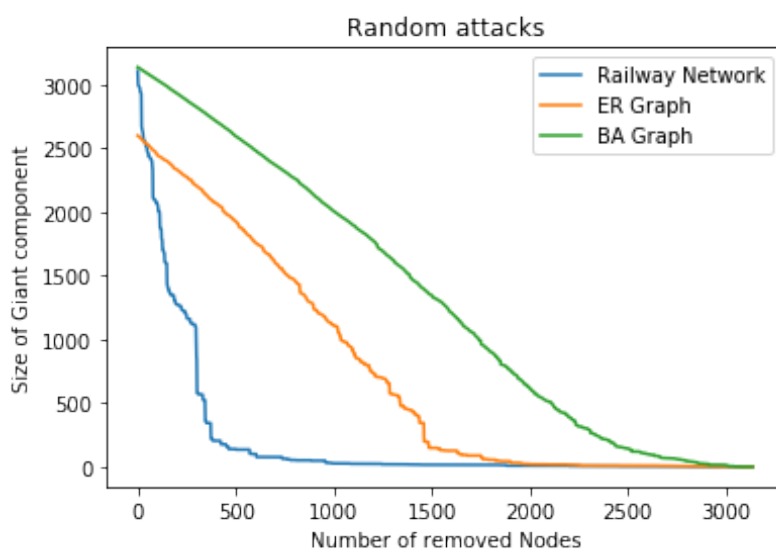


Fig. 10: Random attacks on nodes for different networks

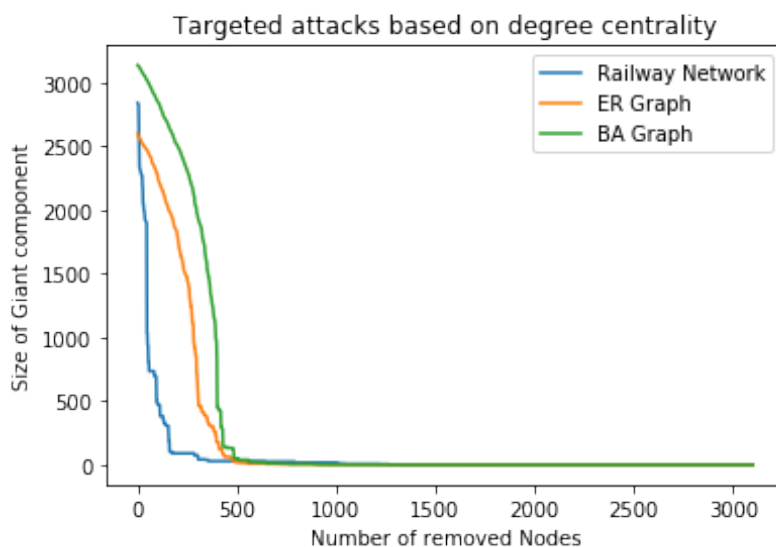


Fig. 11: Targeted attacks on nodes for different networks

5 CONCLUSION

The Swiss Railway network is important to the Swiss citizens as well as to the Swiss economy. Yet, increasing demand and tight schedules are a constant challenge. However, there is a limit to the network especially because few new lines are added. This means that it is all the more important to understand how delays or failures influence the network. This research project investigates the vulnerability of the Swiss Railway network.

We found that the network is quite robust against random attacks. Even though small parts of the network begin to disconnect from the giant component, more than 200 nodes need to be removed to half the size of the giant component. However, the capacity of trains might be impacted if an important node or edge of the network is defective. This leads to delays and rerouting, but the overall connectedness of the network only starts to break down if hundreds of nodes or edges fail at the same time.

Furthermore, we found that the Swiss railway network is at risk of targeted attacks. Breaking down the traffic in 10 selected stations or on 10 important routes impacts the network significantly. Even the biggest connected part of the remaining network would be less than half as big as it was before. An appropriate countermeasure to improve the network could be to split important stations into independent sub-networks such that targeted attacks do not break down the whole system. One countermeasure against the failure of edges could be to add redundant edges on the most important routes.

Further research could look into different properties of the edges. For example the amount of parallel train tracks, the capacity of the tracks or the distance of the edges. This would help in finding out which edges are more prone to failures and congestions or how the network can react if a high capacity edge fails and the traffic has to be rerouted over low capacity edges.