

Identifying contributing factors of train delay through Social Network Analysis

Fabian Becker

June 15, 2024

Abstract

The goal of this document is to demonstrate how to write a scientific paper. We walk through the process of outlining, writing, formatting in L^AT_EX, making figures, referencing, and checking style and content. Source files are available at: <http://hoffman.physics.harvard.edu/example-paper/>.

1 Introduction

Germany is home to the largest railway network in Europe, with a total length of over 39,000 kilometers. France is the next largest, with a network of 27,000 kilometers (Directorate-General for Mobility and Transport European Commission, 2023, p.80). However, the railway service in Germany is also known to be chronically unreliable and unpunctual. A review of the official data reveals that the punctuality rate of long-distance travelers is approximately 70 percent (Deutsche Bahn AG, 2024; German Press Agency, 2023). This figure may appear satisfactory at first glance, but it is important to consider that delay in terms of the German Railway Service (Deutsche Bahn, DB) means arriving at least 15 minutes late to your destination.

To provide context, Switzerland, a neighboring country to Germany, has a punctuality rate of 90 percent for long-distance train (SBB, 2024). The reliability of Deutsche Bahn trains is so poor, that the Swiss Federal Office of Transport proposed to stop German trains from entering too deep into the country in order to prevent cascading delays on their own trains caused by the DB trains (Magill, 2024).

Delay is one of the most significant reasons why the German government introduced a 47 billion euro rail-spending program in order to overhaul the railway network, improve reliability and most importantly to reduce CO₂

emissions by making the train an equally attractive mean of transport (Sally, 2023). One of the first measures to make traveling by train more appealing was to introduce affordable train tickets in the form of a 49 Euro Ticket (Restle, 2023), however if German trains lack general punctuality, the reduced travel costs will not necessarily result in a shift in customer preference towards rail travel over private vehicles.

In order to spend this money purposefully, an in-depth analysis of the German railway system is needed in order to determine plausible causes for unreliability and train delay. Hence to this event I want to discover if Social Network Analysis is a fitting tool in order to find structural causes for train delay that can only be eliminated by targeted investments.

In the first part of this paper, I will construct a network from historical delay data, and also gather metadata about the stations that form the network. I will then conduct a descriptive analysis of the network, before using linear regression to identify features from the network that may be correlated with train punctuality.

2 Related Work

The act of predicting train delay has received a lot of a from science over the years, with the primary objective of enhancing customer satisfaction by reducing the unpredictability of travel times. Train delay prediction can also assist in more accurately planning the utilization of the train network by train traffic control (Spanninger et al., 2022). Despite the fact that this paper does not focus on train delay prediction, determining factors for train delay represent a vital component in the production of precise prediction model.

Spanninger et al., 2022 reviewed and classified a variety of prediction models, into event-driven and data-driven approaches. Event-driven model incorporate train-events, such as departure, arrival and pass-through in order to mimic the railway operation dynamics. On the other hand data-driven approaches generate their predictions utilizing historic train movement data, most of the models therefore are machine learning models. In this paper also relies on historic delay data derived from arrival and departure times whereas it could be classified as a data-driven approach.

Chen and Harker, 1990 for example created a delay estimation system based on non linear equations. They specifically introduced train timetables as a part of their model instead of assuming that trains are uniformly distributed over time. this paper also utilizes timetable data in order to calculate the required delay data.

Murali et al., 2010 created a linear regression model to predict railroad delay. However instead of historic train data they generated their training data by using a simulation and later applied the resulting model to subsections of the Los Angeles train network. The main part of this paper is also employing linear regression, nevertheless here the aim is find important factors that determine train delays and would therefore assist in the improvement of such a linear regression model.

Another academic paper written by Hauck and Kliewer, 2020 makes use of the same linear regression model, however they used a dataset that was provided by the german railway service and is structured in similar fashion to the one that will be constructed as a part of this research. The dataset contains timestamps of the excepted events and the actual time of the event to calculate the delay out of the difference. This at first glance seems almost identical to what will be executed in this paper, however the mentioned research was again only interested in providing an accurate prediction system.

Research that aims to determine factors that cause train delay has been conducted by Dingler et al., 2010. The goal of this paper was gain a better understanding for delay-causing factors. They obtained their data from a railway simulation software and observed the relationship of delay and factors such as train volume or train speed.

This overview of related works helps to draw the conclusion, that the topic of train delays is quite recognized in science. Hence this paper embeds ideas that have already been observed in different perspectives. But ultimately the application of Social Network Analysis on historic german railway data in order to observe possible factors of train delay is an idea that has been absent from research until now.

3 The dataset

The most time-consuming aspect of this research was the creation of a realistic delay data set. In this chapter I will review the data gathering and preprocessing, as well as providing a description of the obtained dataset by using methodology from Social Network Analysis.

3.1 Data gathering

References

- Chen, B., & Harker, P. T. (1990). Two moments estimation of the delay on single-track rail lines with scheduled traffic. <https://doi.org/10.1287/trsc.24.4.261>, 24, 261–275. <https://doi.org/10.1287/TRSC.24.4.261>
- Deutsche Bahn AG. (2024). Erläuterung pünktlichkeitswerte für den april 2024. https://www.deutschebahn.com/de/konzern/konzernprofil/zahlen_fakten/puenktlichkeitswerte-6878476
- Dingler, M., Koenig, A., Sogin, S., & Barkan, C. P. L. (2010). Determining the causes of train delay. <http://railtec.illinois.edu/wp/wp-content/uploads/2019/01/Dingler-et-al-2010-AREMA.pdf>
- Directorate-General for Mobility and Transport European Commission. (2023). Eu transport in figures - statistical pocketbook 2023. <https://doi.org/https://doi.org/10.2832/319371>
- German Press Agency. (2023). Fernverkehr: Jeder dritte bahnreisende verspätet am ziel - zdfheute. <https://www.zdf.de/nachrichten/wirtschaft/unternehmen/bahn-reisende-verspaetung-auswertung-puenktlichkeit-100.html>
- Hauck, F., & Klierer, N. (2020). Data analytics in railway operations: Using machine learning to predict train delays, 741–747. https://doi.org/10.1007/978-3-030-48439-2_90
- Magill, S. (2024). Why swiss transport authorities want to ban german trains. <https://www.thelocal.ch/20230703/why-swiss-transport-authorities-want-to-ban-german-trains>
- Murali, P., Dessouky, M., Ordóñez, F., & Palmer, K. (2010). A delay estimation technique for single and double-track railroads. *Transportation Research Part E: Logistics and Transportation Review*, 46, 483–495. <https://doi.org/10.1016/j.tre.2009.04.016>
- Restle, B. (2023). Germany’s new €49 monthly travel pass explained – dw – 04/28/2023. <https://www.dw.com/en/germanys-new-49-monthly-travel-pass-explained/a-65421776>
- SBB. (2024). Facts and figures - punctuality. <https://reporting.sbb.ch/punctuality?=&years=1,4,5,6,7&scroll=0&highlighted=26949388b33a48d2c18c597dbb0ead74>
- Scally, D. (2023). Germany plans €47bn rail overhaul amid climate fight – the irishtimes. <https://www.irishtimes.com/business/2023/07/21/germany-plans-47-billion-rail-overhaul-amid-climate-fight/>
- Spanninger, T., Trivella, A., Büchel, B., & Corman, F. (2022). A review of train delay prediction approaches. *Journal of Rail Transport Planning*

E Management, 22, 100312. <https://doi.org/10.1016/j.jrtpm.2022.100312>