Network analysis methods for smart inspection in the transport domain

Bruin, G.J. de

Summary

Inspectors are indispensable for monitoring essential regulations that protect the safe and clean transport of goods. However, finding all dangerous behavior with a limited number of inspectors and increasing personnel shortages is challenging. That is the reason inspectorates are looking for innovative methods to find dangerous behavior and improve compliance. We consider a data-driven approach to arrive at *smart* inspection. Smart inspection is performed when we assess compliance of vehicles in a (1) accurate, (2) automated, (3) fair, and (4) interpretable manner.

Models that assess vehicle compliance can be unintentionally biased against certain vehicle features. These vehicle features are divided into two classes, being *static* and *dynamic* features. Examples of static features are the type of vehicle, the size, the insurer, and the country of registration. A vehicle owner can change some of these features (such as insurer and country of registration) to influence an automated model without any actual reduction in the vehicle's dangerous behavior. Therefore, we choose to use dynamic features of the vehicles (such as the routes to be chosen), which say something about the *behavior* of a vehicle and its operator. We use *networks* to encode the vehicle's behavior, allowing us to model a particular part of the transport system as a whole. The *main problem* that the thesis aims to address is thus how network methods can leverage this behavioral data for the smart inspection of vehicles.

We start in Chapter 1 with establishing the context of smart vehicle inspection and the methods used to achieve it. As mentioned, we use behavioral data that we encode with the help of networks. Network science is a young multidisciplinary field of study in which much attention has been paid to the universal properties of networks. Many of those properties are also present in our temporal transport networks. A task often applied to temporal networks is link prediction, aiming at predicting *new* links between existing nodes in a temporal network. A temporal network is a network where the creation time

of edges is known. Link prediction is also a key aspect of our work; Chapters 2 to 4 relate to this task.

Previous research has shown a relationship between a network's structure and performance in a related task, being *missing* link prediction. Our work extended this line of research by applying it to the link prediction task in *temporal networks*. We are particularly interested in uncovering the relationship between network structure and model performance in link prediction.

In Chapter 2, we, therefore, analyze the link prediction task in 26 temporal networks. We do so using a machine-learned classification model fed with topological features. The model independently learns which pairs of nodes likely connect (and which do not). We mention four results obtained from experiments. *First*, we show that the performance of link prediction is higher when the temporal aspect is considered. *Second*, we find a relation between the *overall structure* of a network and the extent to which links can be predicted. In particular, the link prediction model performs well on networks exhibiting negative degree assortativity, i.e., networks wherein low-degree nodes primarily link to high-degree nodes (and vice versa). *Third*, we find that in a network with discrete events, we can improve link prediction performance further by adequately encoding discrete events. *Fourth*, we do not find any apparent performance differences between node-oriented and edge-oriented features except for networks from the information domain. Further research should reveal how this finding can be explained.

In machine learning on tabular data, it is common practice to validate and test model performance by applying the model to data that is *disjoint* and *independent* of the data used to train the model. However, independence cannot be guaranteed with relational data as they occur in networks. Specifically, it is a nontrivial task to estimate rather precisely the performance of link prediction models even when using adequate splits into train, validation, and test sets. In Chapter 3, we, therefore, compare two common approaches from the literature: (1) the *random* split, and (2) the *temporal* split. We compare the performances of these two approaches on the link prediction task and find that the random split gives overly optimistic results. The temporal split does give a more realistic indication of performances. Furthermore, our results prove robust for a wide selection of model parameters.

In the last three chapters, we explicitly focus on *smart vehicle inspection*. We start with *co-driving*, the activity where two trucks drive "together", i.e., pass by the same location and time. Investigating the co-driving behavior of trucks is important because it can positively impact the environment. As a case in point, co-driving may reduce aerodynamic drag and, therefore, may result in optimized fuel usage. We investigate how network structure and vehicle characteristics relate to co-driving behavior. As such, the main topic of Chapter 4 is the *truck co-driving network*. In this network, every node is a truck, and a link exists when two trucks are *systematically* co-driving. *Systematic co-driving* is when two trucks *frequently* drive together. Data for such a study were collected from 18,000,000 truck movements in the Netherlands. We have used insights gained by applying link prediction to this network to understand truck co-driving behavior. The model uses features that are categorized into (1) spatiotemporal, (2) topological, (3) node-, and (4) path-oriented features. We found that truck co-driving behavior is best encoded using *topological* features and, to a lesser extent, the *path-oriented* and *spatiotemporal* features. Our findings indicate that the dynamics of the co-driving network exhibit significant social network effects.

We also looked at its communities to better understand the truck co-driving network. A so-called community detection algorithm can use the structure of a network to arrive at a good partitioning into groups of densely connected nodes. In our specific case, however, we also have information on the truck (i.e., the network's nodes) that we use to arrive at a proper partitioning into communities. We investigated how node attributes can be exploited to automatically create a good partitioning of a co-driving network into communities.

In Chapter 5, we propose a new metric, *the average maximal community assortativity*, to better understand the structure of communities in a network using node attribute assortativity. More specifically, we propose to select solutions to the community detection problem that maximizes the average maximal community assortativity metric. A high assortativity for a particular feature then indicates a better community representation. In the case of the truck co-driving network, we observe that *geographical* node attributes especially characterize communities.

This thesis's final topic relates to smart vehicle inspection and network science. It concerns the question of how ship behavior can be utilized to enable smart inspection of cargo ships.

In Chapter 6, we provide such an approach to smart cargo ship inspection. We use a model that is *interpretable* and *fair*. The model cannot only use static administrative ship properties in its prediction but, in particular, utilizes features describing the ship's *behavior*. By incorporating ship behavior, meaningful characteristics can be derived and utilized as input for the model. It leads us to a smart risk assessment of cargo ships. Our approach allows inspectorates to trace specifically noncompliant cargo ships. Thereby, this chapter contributes to improved maritime safety and environmental protection.

In general, we demonstrate how network science and behavioral data can be utilized to arrive at a *smart inspection* of vehicles. With this explanation and interpretation of smart inspection, we are sure to have addressed the overall problem statement of the thesis.