Human Environment and Transport Inspectorate
Ministry of Infrastructure and Water Management

Universiteit Leiden

Understanding Dynamics of Truck Co-Driving Networks

Gerrit Jan de Bruin [1,2], Cor J. Veenman [1],
H. Jaap van den Herik [1], and Frank W. Takes [1]

[1]Leiden University, The Netherlands

[2]Ministry of Infrastructure and Water Management, The Netherlands

Complex Networks, Lisbon, December, 2019

Table of Contents

Introduction

Problem statement

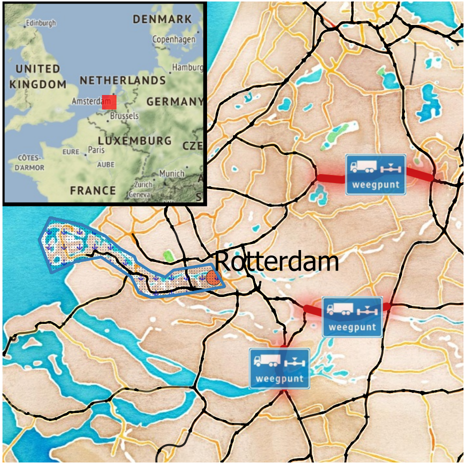
Learn dynamics of truck co-driving behaviour.

Motivation

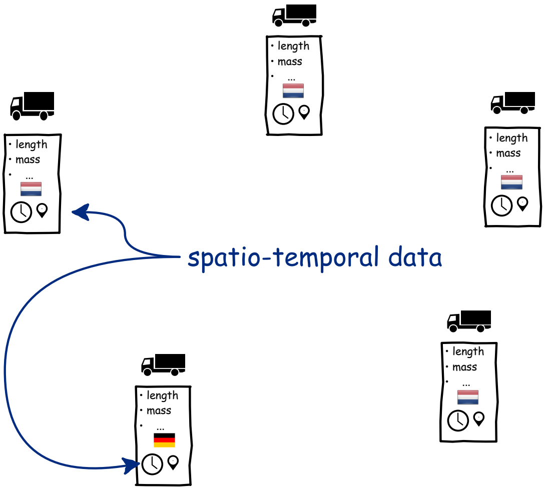
- Positive environmental impact.
- Innovative forms of transportation may have major implications for this behaviour.

Data

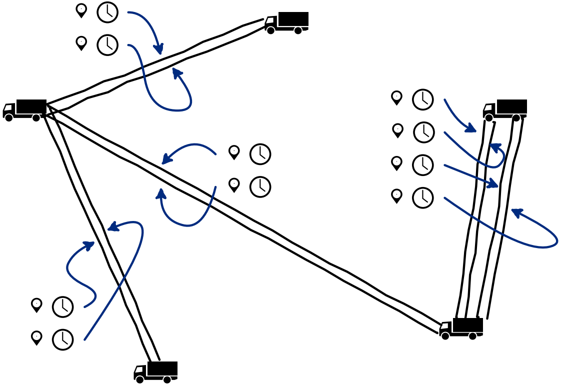
- 9,202,764 measurements
- 905,442 trucks
  - length
  - no. axes
  - mass
  - country
- collected in 2016 and 2017



Co-Driving Network



Co-Driving Network



Co-Driving Network

Co-driving is the activity where two trucks drive together, i.e., are frequently at the same place at the same time.

[1] Determined empirically.
[2] Idem.
[3] Idem.

Co-Driving Network

Co-driving is the activity where two trucks drive together,
i.e., are <u>frequently</u> at the same place at <u>the same time</u>.

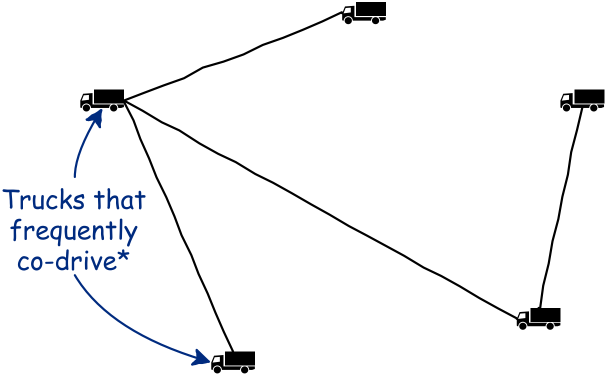
Criteria:

- Trucks appear at the same place within at most $\Delta t = 8\ s$[1].
- This happens at least twice[2], with more than 2 hours difference[3].

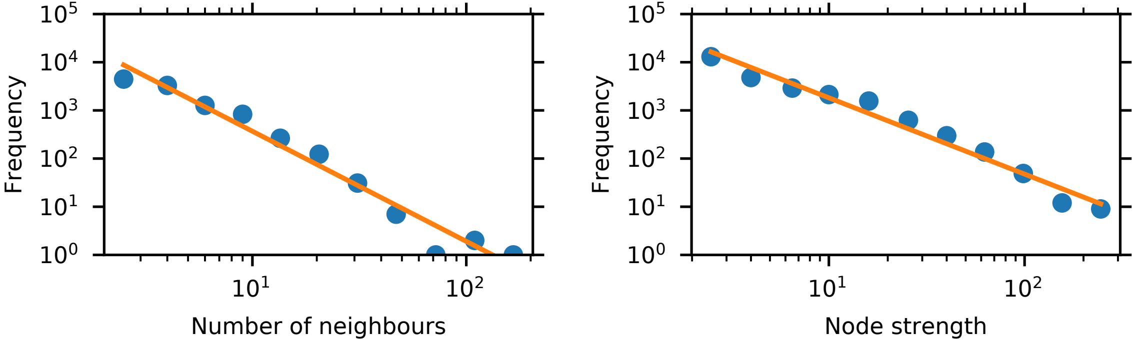
These criteria should select only **intentional** *co-driving behaviour.*

[1] Determined empirically.
[2] Idem.
[3] Idem.

Co-Driving Network



Degree and strength distribution



Network Statistics

Common characteristics:

- Dominant giant component
- Low average shortest path length
- Power-law degree distribution

Table: Statistics of the truck co-driving network.

Metric	Value
Number of nodes	25,553
Number of links	73,059
Number of node pairs connected	27,986
Fraction nodes in giant component	62%
Fraction links in giant component	79%
Density	2.2×10^{-4}
Average shortest path length	7.8
Diameter	24

Previous work

Understanding Behavioral Patterns in Truck Co-Driving Networks

Gerrit Jan de Bruin[1,2,3], Cor J. Veenman[1,4], H. Jaap van den Herik[2], and Frank W. Takes[1]

[1] Leiden Institute of Advanced Computer Science (LIACS), Leiden University
g.j.de.bruin@liacs.leidenuniv.nl
[2] Leiden Centre of Data Science (LCDS), Leiden University
[3] Human Environment and Transport Inspectorate (ILT), Netherlands Ministry of Infrastructure and Water Management
[4] Data Science Department, TNO

Abstract. This paper examines the co-driving behavior of truck drivers using network analysis. From a unique spatiotemporal dataset encompassing more than 10 million measurements of trucks passing 17 different highway locations in the Netherlands, we extract a so-called *co-driving network*. In this network, nodes are truck drivers and edges represent pairs of trucks that are systematically driving together. The obtained co-driving network structure has various properties common to real-world networks, such as a dominant giant component and a power law degree distribution. Moreover, network distance metrics and community detection reveal that the network has a highly modular structure. We furthermore propose a method for understanding the network community structure through attribute assortativity. Results indicate that co-driving links are mostly established based on geographical aspects: truck drivers from the same country or the same region in the Netherlands are more inclined to drive together. The resulting improved understanding of co-driving behavior has important implications for society and the environment, as trucks coordinating their driving behavior together help reduce traffic congestion and optimize fuel usage.

Keywords: co-driving networks, infrastructure networks, network analysis, community detection, assortativity

1 Introduction

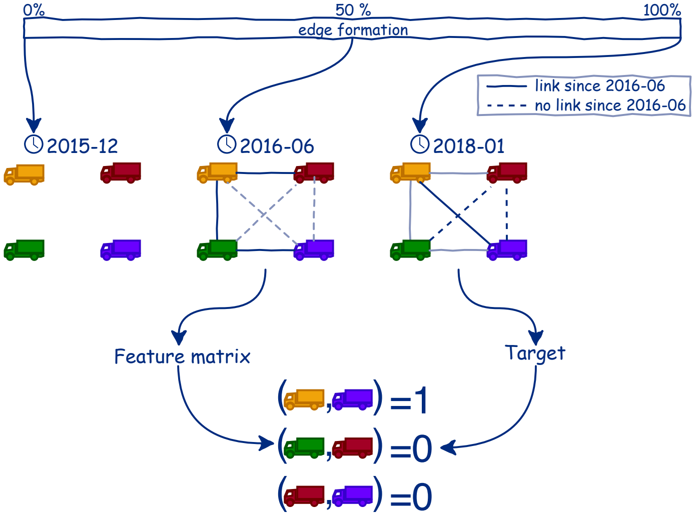
Techniques from the field of network science are used in a broad range of domains to extract knowledge from the network structure of real-world systems [1,12]. In this paper we use a network approach to investigate the factors that stimulate co-driving behavior of truck drivers. We furthermore look at what patterns can be found in groups of truck drivers who systematically drive together. To do so, we use network community detection [5] as well as various metrics related to assortativity (also known as mixing patterns, see [10]).

Making sense of communities using the attributes available in the data.

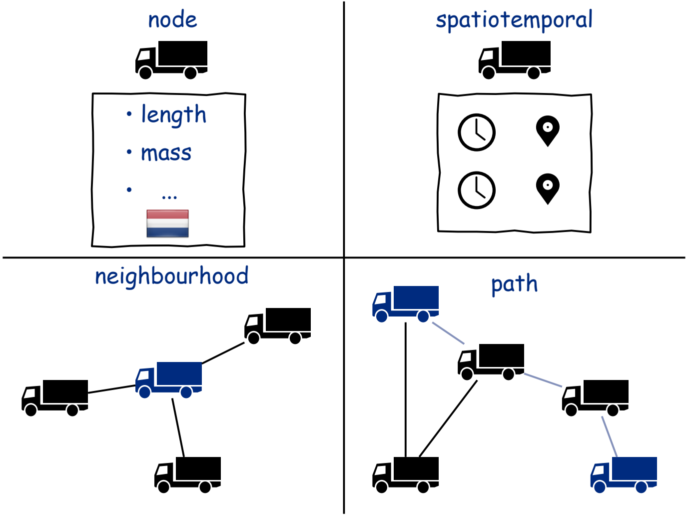
Available at:

`gerritjandebruin.nl/CN2018.pdf`

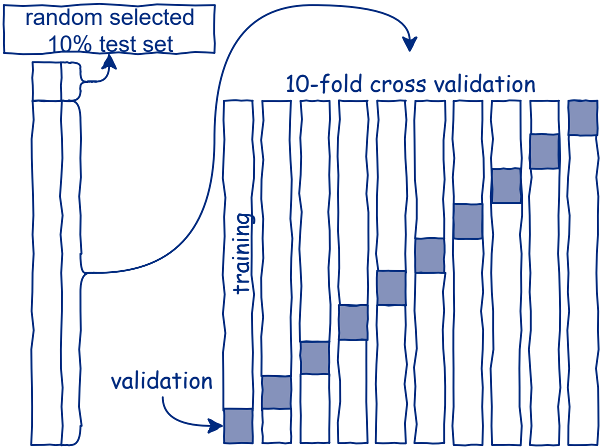
Link prediction



Feature types



Random forest



Random forest

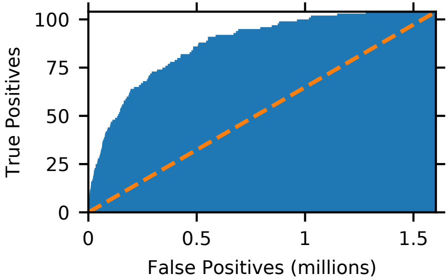


Figure: ROC curve of the random forest link prediction classifier. The AUROC is 0.84.

Features in link prediction

Table: Features for truck pair (a, b) used in the link prediction model. The rightmost column lists the feature importance calculated using Gini importance as provided by the random forest classifier.

Feature	Type	Importance
$\lvert\Gamma(a)\rvert + \lvert\Gamma(b)\rvert$	neighbourhood	0.117
shortest path length in G	path	0.111
$\lvert\Gamma(a) \cup \Gamma(b)\rvert$	neighbourhood	0.093
$last\_year_l\,(a+b)$ for $l = 1, \ldots, 8$	spatio-temporal	0.010–0.060
$last\_month_l\,(a+b)$ for $l = 1, \ldots, 8$	spatio-temporal	0–0.057
$s_a + s_b$	neighbourhood	0.056
$\lvert truck\_len\,(a) - truck\_len\,(b)\rvert$	node	0.040
$\lvert truck\_mass\,(a) - truck\_mass\,(b)\rvert$	node	0.030
$\lvert driving\_hours\,(a) - driving\_hours\,(b)\rvert$	node	0.030
$last\_week_l\,(a+b)$ for $l = 1, \ldots, 8$	spatio-temporal	0–0.027
$\lvert\Gamma(a) \cap \Gamma(b)\rvert$	neighbourhood	0.021
$\lvert weekend\_driver\,(a) - weekend\_driver\,(b)\rvert$	node	0.019
$\lvert s_a - s_b\rvert$	neighbourhood	0.017
$truck\_len\,(a) + truck\_len\,(b)$	node	0.017
$truck\_mass\,(a) + truck\_mass\,(b)$	node	0.016
$driving\_hours\,(a) + driving\_hours\,(b)$	node	0.016
$weekend\_driver\,(a) + weekend\_driver\,(b)$	node	0.014
$\Big\lvert \lvert\Gamma(a)\rvert - \lvert\Gamma(b)\rvert \Big\rvert$	neighbourhood	0.013
$\lvert truck\_ax\,(a) - truck\_ax\,(b)\rvert$	node	0.008
$truck\_ax\,(a) + truck\_ax\,(b)$	node	0.006
$truck\_country\,(a) = truck\_country\,(b)$	node	0.005

Gini Feature Importance

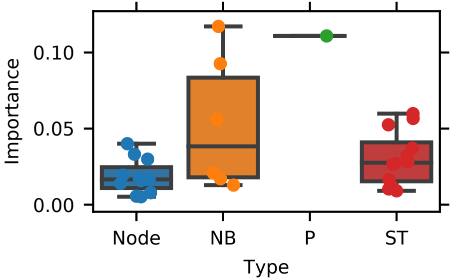


Figure: Gini feature importance. NB, P and ST are the neighbourhood, path and spatiotemporal features resp.

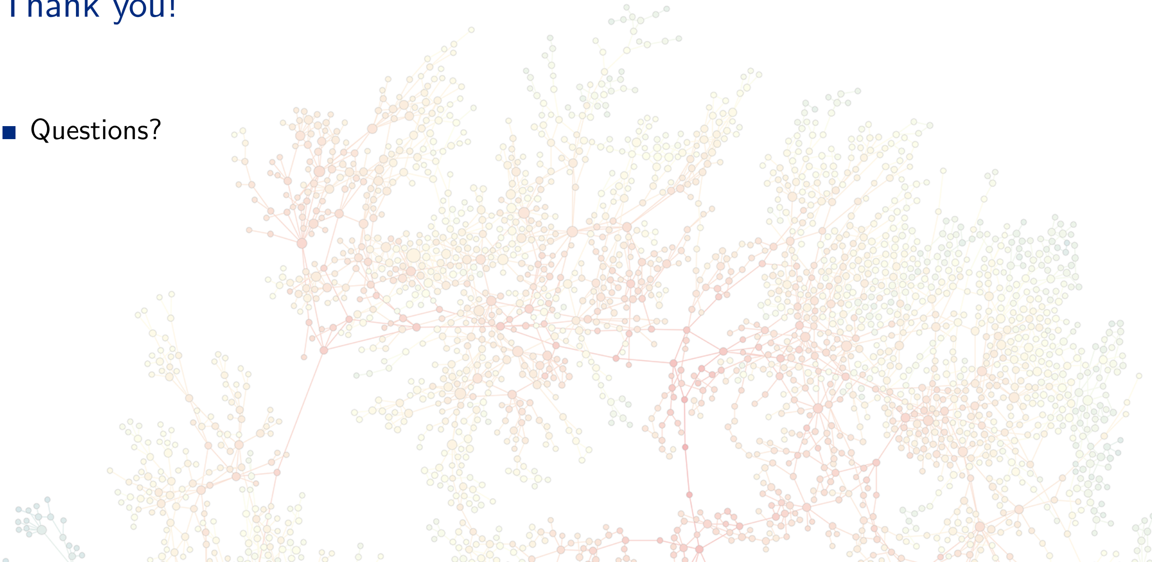
Conclusion

- Link prediction for understanding network dynamics
- Four types of feature sets
- Feature importance for each feature
- Network effects are present

- Future work: interpret results with infrastructure domain experts

Thank you!

- Questions?



Thank you!

- Questions?

- E-mail: `g.j.de.bruin@liacs.leidenuniv.nl`
- Leiden Computational Network Science Lab
- `https://cns.liacs.nl`

This presentation is available at `http://gerritjandebruin.nl/Talk-CN2019.pdf`.

Class imbalance

We use three measures to limit class imbalance:

- use class weights (`class_weight=balanced`)
- use only active nodes
- use only nodes in GC

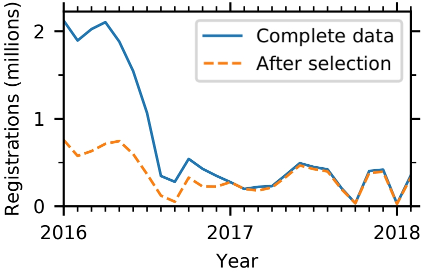


Figure: Number of registrations over time.

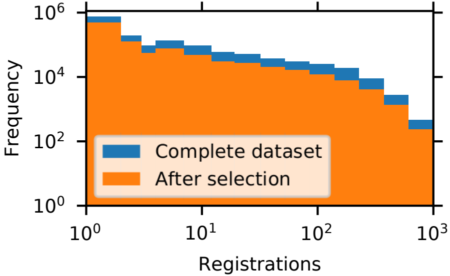


Figure: Histogram of registrations per truck. Note logarithmic axes.

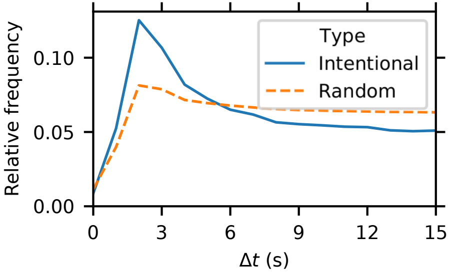


Figure: Frequency distribution of time gap measured between the two trucks in a co-driving event.

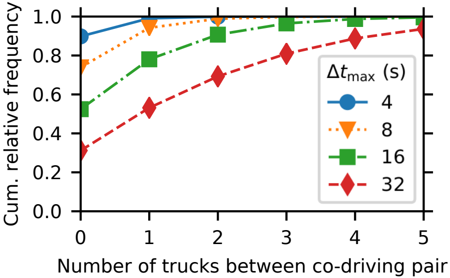
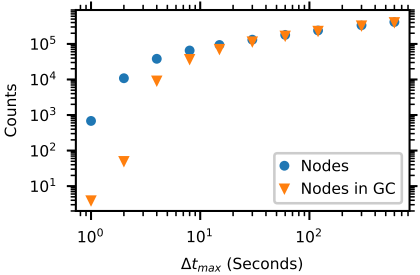
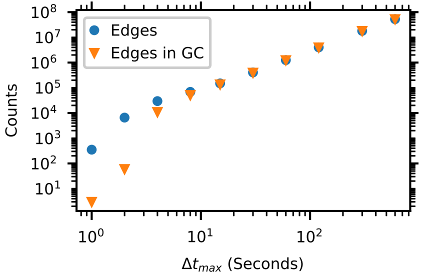


Figure: Frequency distribution of trucks driving between the two trucks in a co-driving event.

Table: Information available about truck u, collected from its registrations $\mathcal{D}_u$.

Property		Description	Type
$truck\_country_u$		country of registration	string
$truck\_axes_u$	$\underset{\{x_i \in \mathcal{D}_u\}}{\text{Median}}\ axes_i$	number of axes	number
$truck\_length_u$	$\underset{\{x_i \in \mathcal{D}_u\}}{\text{Median}}\ length_i$	length	number
$truck\_mass_u$	$\underset{\{x_i \in \mathcal{D}_u\}}{\text{Median}}\ mass_i$	mass	number
$driving\_hours_u$	$\underset{\{x_i \in \mathcal{D}_u\}}{\text{Mean}}\ \lvert t_i(\text{h}) - 12\text{h}\rvert$	usual driving hours	number (0–12)
$weekend\_driver_u$	$\underset{\{x_i \in \mathcal{D}_u\}}{\text{Mean}} \begin{cases} 0 \text{ if } t_i = \text{weekday} \\ 1 \text{ if } t_i = \text{weekend} \end{cases}$	fraction driving in weekend	number (0–1)





(a) Number of nodes in the network and in its giant component (GC) for increasing Δt_{max}.

(b) Number of edges in the network and in its giant component (GC) for increasing Δt_{max}.