



Universiteit
Leiden
The Netherlands

Network analysis methods for smart inspection in the transport domain

Bruin, G.J. de

Citation

Bruin, G. J. de. (2023, November 16). *Network analysis methods for smart inspection in the transport domain*. SIKS Dissertation Series. Retrieved from <https://hdl.handle.net/1887/3656981>

Version: Publisher's Version

License: [Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)

Downloaded from: <https://hdl.handle.net/1887/3656981>

Note: To cite this publication please use the final published version (if applicable).

Samenvatting

Inspecteurs zijn onmisbaar voor toezicht op veilig en schoon goederenvervoer. Echter, het is moeilijk om al het gevaarlijk gedrag op te sporen met een beperkt aantal inspecteurs. De huidige personeelstekorten maken deze situatie nog actueler. Daarom zoeken inspectiediensten naar innovatieve methoden om gevaarlijk gedrag op te sporen en naleving te verbeteren. Wij onderzoeken een data-gedreven aanpak om te komen tot een *slimme* inspectie. Onder een slimme inspectie verstaan wij een (1) nauwkeurige, (2) geautomatiseerde, (3) onbevooroordeelde, en (4) verklaarbare manier om voertuigen te beoordelen op overtredingen.

Modellen, die gebruikt worden om voertuigen te beoordelen, kunnen onbedoeld vooringenomen zijn tegen bepaalde voertuigkenmerken. Deze kenmerken worden verdeeld in *statische* en *dynamische* kenmerken. Voorbeelden van statische kenmerken zijn voertuigtype, grootte, verzekeraar, en land van registratie. Sommige van deze kenmerken (zoals verzekeraar en land van registratie) kunnen door eigenaars veranderd worden om het model te beïnvloeden, zonder dat er daadwerkelijk vermindering plaatsvindt van ongewenst gedrag. Daarom gebruiken we dynamische kenmerken (zoals routes die genomen worden) die daadwerkelijk iets zeggen over het *gedrag* van een voertuig (en zijn exploitant) en moeilijk te vervalsen zijn. Om het voertuiggedrag te coderen gebruiken we *netwerken*. De *overkoepelende probleemstelling* van dit proefschrift is hoe netwerkmethoden deze gedragsgegevens kunnen benutten voor de slimme inspectie van voertuigen.

We starten Hoofdstuk 1 met het uiteenzetten van de context van slimme voertuig inspecties en de methoden waarmee dit bereikt wordt. Zoals genoemd, gebruiken we daarbij gedragsgegevens die we coderen met behulp van netwerken. De netwerkwetenschap is een jong en multidisciplinair onderzoeksgebied waar veel aandacht besteed wordt aan de universele eigenschappen van netwerken. Veel van deze eigenschappen zijn ook aanwezig in onze temporele transportnetwerken, d.w.z. netwerken waarbij van elke link bekend is wanneer deze ontstaan is. Een taak die vaak wordt uitgevoerd op temporele netwerken is de linkvoorspellingstaak, waarbij het doel is om *nieuwe* links te voorspellen tussen bestaande knopen. De linkvoorspellingstaak komt veel terug in ons werk, namelijk in Hoofdstukken 2 tot en met 4.

Uit eerder onderzoek is gebleken dat bij een verwante taak, het voorspellen van *missende* links, er een relatie bestaat tussen de structuur van het netwerk en de prestatie van de taak. In ons werk, hebben we deze onderzoekslijn uitgebreid naar de linkvoorspellingstaak op *temporele netwerken*. We zijn in het bijzonder geïnteresseerd om de relatie te begrijpen tussen (1) de netwerkstructuur en (2) de prestatie van de linkvoorspellingstaak.

In Hoofdstuk 2, analyseren wij de linkvoorspellingstaak op 26 temporele netwerken. We doen dit met behulp van een machinaal geleerd linkvoorspellingsmodel dat gevoed wordt met topologische kenmerken. Het model leert zelfstandig welke knopenparen waarschijnlijk verbinden (en welke niet). We noemen vier resultaten verkregen uit experimenten. *Ten eerste*, laten wij zien dat de prestatie van de linkvoorspellingstaak meer precies is wanneer er rekening gehouden wordt met het temporele aspect. *Ten tweede*, tonen wij aan dat er een relatie is tussen de *algehele structuur* van een netwerk en de mate waarin links voorspeld kunnen worden. In het bijzonder, valt op dat de prestatie van de linkvoorspellingstaak meer precies is in netwerken die een negatieve graadassortativiteit hebben, d.w.z. netwerken waarin lage-graad knopen voornamelijk linken met hoge-graad knopen (en vice versa). *Ten derde*, tonen we aan dat in netwerken met discrete gebeurtenissen, de prestatie van de linkvoorspellingstaak verhoogd kan worden bij het adequaat coderen van de discrete gebeurtenissen. *Ten vierde*, vinden wij geen duidelijke prestatieverschillen tussen knoop-georiënteerde en link-georiënteerde kenmerken. De enige uitzondering hierop lijkt te bestaan voor netwerken in het informatiedomein, maar verder onderzoek zal moeten uitwijzen hoe deze vondst kan worden verklaard.

Bij modellen die machinaal geleerd zijn op basis van tabelgegevens is het gebruikelijk dat de prestatie gevalideerd en getest wordt met gegevens die *apart gehouden zijn* en *onafhankelijk zijn* van de gegevens die gebruikt zijn om het model te leren. De onafhankelijkheid valt echter niet te waarborgen bij relationele gegevens zoals die geïmplementeerd zijn in netwerken. Meer in het bijzonder, is het niet-triviaal om schattingen te verkrijgen van de prestaties van linkvoorspellingsmodellen, zelfs niet wanneer gebruik gemaakt wordt van adequate splitsingen in *train*, *validatie*, en *test sets*. In Hoofdstuk 3 onderzoeken we daarom twee benaderingen om de splitsing uit te voeren: (1) de willekeurige splitsing en (2) de temporele splitsing. We vergelijken de prestaties van deze twee benaderingen

op de linkvoorspellingstaak en vinden dat de *willekeurige* splitsing te optimistische resultaten geeft. De *temporele* splitsing geeft een meer realistische indicatie van de prestaties en verder blijken onze resultaten robuust te zijn voor een brede selectie aan parameters.

In de laatste drie hoofdstukken, richten we ons expliciet op het doen van *slimme voertuiginspecties*. We starten met samenrijdgedrag, de activiteit waarbij twee vrachtwagens ‘samen’ rijden op dezelfde tijd en dezelfde plaats. Het onderzoek naar samenrijdgedrag van vrachtwagens is belangrijk omdat het een positieve bijdrage aan het klimaat kan leveren. Als voorbeeld noemen we dat samenrijden resulteert in verminderde luchtweerstand, wat dan kan resulteren in verminderd brandstofverbruik. We beginnen ons onderzoek met de vraag hoe netwerkstructuur en voertuigkenmerken relateren aan samenrijdgedrag. Als zodanig is het *samenrijden van vrachtwagens* het hoofdonderwerp in Hoofdstuk 4. In dat netwerk is elke knoop een vrachtwagen en bestaat er een link als twee vrachtwagens consequent samenrijden. *Systematisch* samenrijden gebeurt wanneer twee vrachtwagens frequent samenrijden. De gegevens voor deze studie zijn verzameld vanuit 18 miljoen vrachtwagenbewegingen in Nederland. We kunnen de inzichten van de linkvoorspellingstaak gebruiken om het gedrag van voertuigen te begrijpen. In dit model worden kenmerken gebruikt die we categoriseren in vier typen: (1) tijd-ruimtelijk, (2) topologisch, (3) knoop-, en (4) pad-georiënteerd. We hebben gevonden dat het samenrijdgedrag het best gecodeerd wordt met de topologische kenmerken en in mindere mate door de pad en tijd-ruimtelijke kenmerken. Onze bevindingen suggereren dat de dynamiek van het samenrijdnetwerk duidelijk sociale netwerk-effecten heeft.

Om het netwerk nog beter te begrijpen, kunnen we ook de *gemeenschappen* bestuderen die in het netwerk aanwezig zijn. Deze gemeenschappen worden gevonden met behulp van een zogenaamd gemeenschapsdetectie algoritme, die de structuur van het netwerk gebruikt om tot een goede partitionering in gemeenschappen te komen. Specifiek in ons geval, hebben we aanvullende kenmerken van de vrachtwagens beschikbaar, die we ook kunnen gebruiken om tot een adequate partitionering in gemeenschappen te komen. We onderzoeken hoe knoopkenmerken benut kunnen worden om automatisch tot een goede partitionering van het samenrijdnetwerk in gemeenschappen te komen.

In Hoofdstuk 5 hebben we een nieuwe maat voorgesteld, de *gemiddelde maximale gemeenschapsassortativiteit*, die helpt bij het begrijpen van de gemeenschapsstructuur in een netwerk door middel van de assortativiteit van knoopkenmerken. In het bijzonder stellen we voor om die splitsing in gemeenschappen te gebruiken waarbij de hoogste gemiddelde maximale gemeenschapsassortativiteit gevonden wordt. Een hogere assortativiteit voor een bepaald kenmerk geeft aan dat er hier een betere partitionering van die gemeenschap is gebruikt. In het geval van samenrijdende vrachtwagens, hebben we opgemerkt dat er vooral gemeenschappen bestaan die gekarakteriseerd worden door geografische kenmerken.

Het laatste onderwerp van deze thesis houdt verband met zowel *slimme voertuiginspecties* als *de netwerkwetenschap*. Het betreft de vraag hoe scheepsgedrag benut kan worden om slimme inspecties van vrachtschepen mogelijk te maken.

In Hoofdstuk 6 komen we tot een dergelijke aanpak voor het slim inspecteren van vrachtschepen. We maken gebruik van een *interpreteerbaar en onbevooroordeeld* machinaal geleerd model. Het model is niet alleen in staat om *statische* administratieve eigenschappen van schepen te gebruiken in de risicovoorspelling maar ook specifiek het *gedrag* van schepen. Het gebruik van gedragsgegevens stelt ons vervolgens in staat om betekenisvolle kenmerken van schepen af te leiden. Daarna kan het resultaat gebruikt worden om tot een *eerlijke* en *betere risico-inschatting* van schepen te komen. Dit betekent dat onze benadering inspectiediensten in staat stelt om specifiek de overtredende vrachtschepen op te sporen. Daarmee draagt dit hoofdstuk bij aan *maritieme veiligheid* en *milieubescherming*.

In meer algemene zin, hebben we laten zien hoe netwerkwetenschappen en gedragsgegevens benut kunnen worden om te komen tot een *slimme inspectie* van voertuigen. Daarmee voorzien we in een antwoord op de overkoepelende probleemstelling van dit proefschrift.

