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]).

In this paper we analyze a unique dataset that was gathered over the period of one year, detailing the presence of at least 900,000 trucks. The dataset consists of more than 10 million measurements and encompasses all trucks which were driving at one of 17 highway locations in the Netherlands. This system is able to measure (a) the weight of each axle of a truck as well as (b) the speed and (c) the license plate. The latter allows trucks to be uniquely identified.

This paper investigates the above mentioned spatiotemporal data as a so-called *co-driving network*, wherein the nodes represent trucks, which essentially identify the truck drivers whose behavior we are interested in understanding. A *co-occurrence* of trucks takes place when two trucks are at the same location. More specifically, we refer to these truck co-occurrences as truck *co-driving* if two trucks are both measured within a small time window. Those pairs of co-driving trucks that occur for a certain number of times (e.g., more than once), are defined as *systematic co-driving trucks*. In the co-driving network, the edges represent this systematic co-driving behavior. We will explain the precise definitions and thresholds to derive these networks in more detail in Section 3.

The results of this work contribute to topics related to understanding human behavior, autonomous driving and environmental sustainability. Using network metrics, we derive what factors may influence the decision of truck drivers to systematically drive together. These findings can prove useful for research on innovative forms of transportation, such as autonomous driving. The expectation is that by co-driving trucks can save up to 15% on fuel as a result of reduced aerodynamic drag [13]. In addition, co-driving trucks reduce traffic congestion. This highlights the potential environmental implications of understanding co-driving behavior.

The co-driving network turned out to have at least three properties that are often encountered in real-world networks. First, our co-driving network is scale-free, i.e., the degree distribution follows a power law [1]. Second, the network has a large giant component which contains 37,858 nodes (trucks) and the majority of the co-driving links of the network. Third, the average shortest path in the network is around 9 edges, which, given the large number of nodes, is relatively small and hints at a small-world like structure [8]. We also show that the network has a highly modular structure with a clear division into communities.

Apart from the network structure, we have access to additional meta-data on the the trucks. This enables us to investigate node attribute assortativity, resulting in insights into what factors contribute to the network structure and more importantly, explain co-driving behavior. Subsequently, we will use this meta-data to better understand the discovered communities. This allows us to understand how local groups of co-driving trucks emerge and contribute to the global network structure. Furthermore, the proposed approach for evaluating and understanding the results of community detection using node attribute assortativity are broadly applicable in other types of networks, providing a methodological contribution to the field.

The remainder of this paper is organized as follows. After discussing related work in Section 2, Section 3 explains how the network was constructed from the raw data. Section 4 is concerned with the proposed approach and techniques to understand the network structure. Then, Section 5 provides details on the results obtained. Conclusions and suggestions for future work are provided in Section 6.

2 Related Work

Below we will particularly focus on three network science studies that infer behavioral patterns from the underlying network structure.

The first work is [2], in which face-to-face contacts were recorded with a 20 second interval using measurement infrastructure at several social settings. One of the results was that aggregated network topology and temporal behavioral properties are strongly related. Additionally, they showed that community detection was able to make a sensible division of the network that was explainable by various properties of the nodes. This paper employs a similar approach, where the network topology and community structure are also explained by properties of the individual nodes and their assortative linking patterns.

Second, in a more recent study, researchers handed over 1000 phones to students who agreed to have their communication and spatiotemporal activities traced. The work showed that network metrics are more informative indicators of university performance than individual characteristics such as personality [7]. It underpins the added value that network metrics can provide over more classical data aggregates.

Third, research on the Brazilian Federal Police criminal intelligence network [4] used network science techniques to uncover behavioral patterns amongst criminals. Similar to our data, their network also featured a large giant component, as well a degree distribution that follows a power law. Their observed low density and high average shortest path length were explained as ‘no trust among thieves’. Additionally, they showed that their giant component had a highly modular structure, which was explained by the necessity of being both efficient running criminal activities within the group while at the same time also being obscure to the outside world.

Throughout this paper we will employ techniques similar to the ones pointed out in the works mentioned above, aiming to extract behavioral insights. To the best of our knowledge, this paper is the first to investigate the phenomenon of truck co-driving using network science methods and techniques.

3 Network Construction

This section explains how the network has been constructed. We start in Section 3.1 with the characteristics of the data. In Section 3.2 we describe how to construct the network, leading to the structure described in Section 3.3. Section 3.4 reports two robustness checks, and finally Section 3.5 details a particular regional co-driving network for which additional attributes could be obtained.

3.1 Truck Observation Data

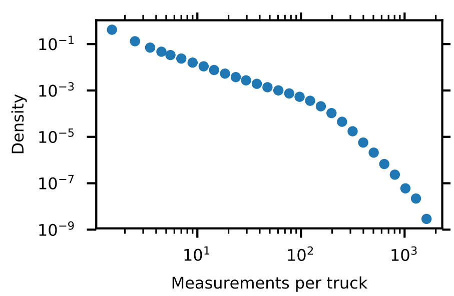
The data was created by observations in 2016 from the Weigh-In-Motion system (see https://international.fhwa.dot.gov/pubs/pl07028 as well as Section 1 for details on this system). The system is maintained by the Dutch Ministry of Infrastructure and Water Management. The data contains over 16 million observations of trucks passing at one of 17 measurement points situated at evenly distributed locations in the Netherlands. For each observation, the following data was available:

- license plate (serving as a unique identifier)
- location ℓ (either one of 17 highway locations)
- lane h, indicating which of the (at most 2) lanes the truck was in
- speed v (in km/h)
- timestamp t at a 10 millisecond resolution
- *country* (using the ISO-2 country code)

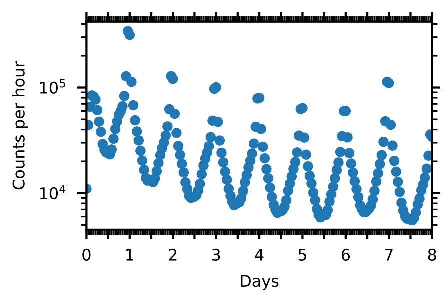
We briefly investigate several properties of the truck observation data. A frequency distribution of how often each distinct truck (identified by its license plate) is measured, is given in Figure 1a. The distribution is highly skewed to the lower values, meaning that most trucks are only measured a few times. In Figure 1b the distribution of the interval between two successive measurements of the same truck at the same location is shown. It demonstrates how most trucks return at similar times at the same location, visible from the peaks at multiples of 24 hours. Similarly, the peak at 7 days indicates weekly driving patterns of trucks. In general, this figure indicates that most individual trucks have regular driving patterns.

3.2 Construction from raw data

In the co-driving network, nodes are trucks and edges represent systematically co-driving trucks. To determine which truck pairs are systematically driving



(a) Distribution of the number of measurements of a truck in the dataset.



(b) The counts per hour between two successive observations of the same truck at the same location over 8 days.

Fig. 1: Characteristics of truck observation data (note various logarithmic axes).

together, we employ the following definitions. When applied in sequence, they serve as the steps to reduce the candidate set of all node (truck) pairs to the set of systematically co-driving node pairs that form the links of the co-driving network.

Definition 1. Co-occurrence *of trucks a and b takes place if two trucks are at the same place, i.e., their location attribute is identical, so* $\ell_a = \ell_b$.

Definition 2. Co-driving *trucks are those co-occurrences* (a, b) *of trucks that take place within a time window of at most* Δt_{max}*, so* $|t_a - t_b| \leq \Delta t_{max}$.

Definition 3. Systematically co-driving *trucks are those co-driving trucks* $(a, b) \in E$ *occurring at least* $\Theta > 0$ *times.*

Thus, to derive the co-driving network, we must set parameters Δt_{max} and Θ.

An intuitive guess would say that Δt_{max} could range anywhere from five seconds up to one minute, depending on how close to each other trucks drive. We derive the right parameter setting in a data-driven manner. In Figure 2 network characteristics are shown for increasing values of Δt_{max}. Definitions of these metrics, all common in the field of network science, can be found in [1]. Recall that a high value for Δt_{max} will result in a high probability that a pair of co-occurring trucks is added by chance. Therefore we choose to keep the value relatively low, namely at $\Delta t_{max} = 8$ seconds. At this value the density of the resulting network is lowest, while the giant component's size compared to the full network (in terms of both nodes and edges) has become stable. Other network metrics such as the diameter and the average distance of the giant component also stabilize around this value, as can be seen in Figure 2.

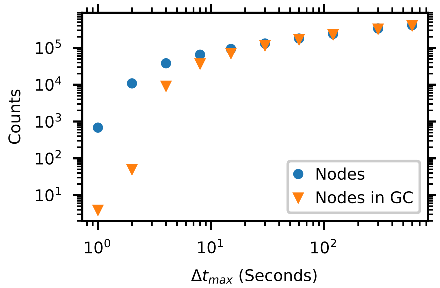
We expect the probability that two trucks randomly co-drive twice is sufficiently small. Therefore, we identify non-random and thus systematic co-driving by setting $\Theta = 2$. Section 3.4 reports various robustness checks for the parameter settings.

3.3 Co-driving Network

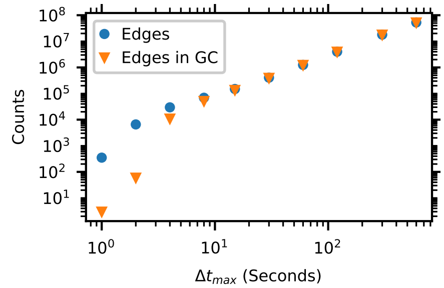
The co-driving network is an undirected weighted network $G = (V, E, w)$, where V is the set of all trucks that is involved in a co-driving activity at least once. For a truck pair $(a, b) \in E$ the weight indicates the number of times the two trucks drove together. For this value holds that $w_{a,b} \geq \Theta$ as required by Definition 3. We furthermore computed several attributes for each truck from the raw measurement data, resulting in the following set of attributes per node:

- *country*, directly derived from the license plate
- median truck speed ($\tilde{v}$)
- the number of different locations where the truck was observed (n_ℓ)
- the location where the truck was most frequently observed (ℓ_{max})

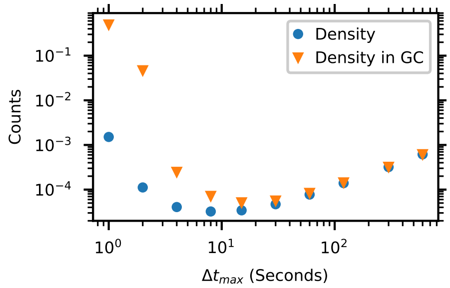
In the remainder of this paper we investigate the structure of the co-driving network and the extent we observe assortative behavior with respect to these attributes.



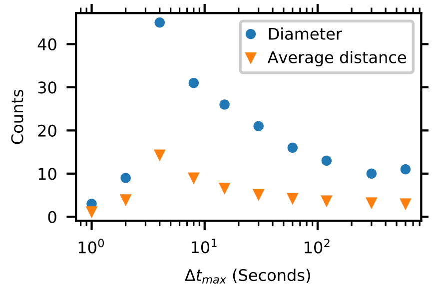
(a) Number of nodes in the network and in its giant component (GC) for increasing Δt_{max}



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



(c) Density of the network and its giant component (GC) for increasing Δt_{max}



(d) Diameter and average shortest path length in the GC for increasing Δt_{max}.

Fig. 2: Co-driving network statistics for increasing values of time window Δt_{max}.

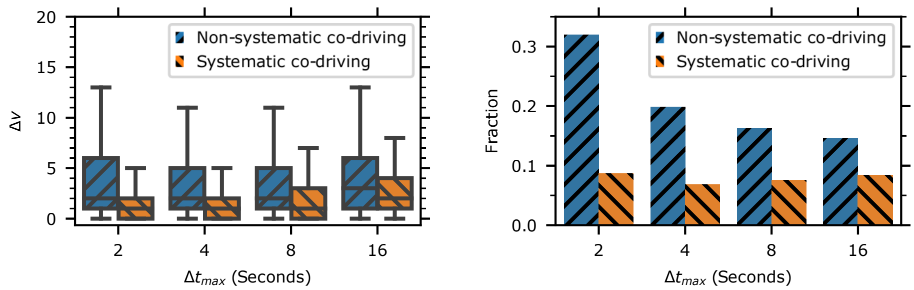
3.4 Robustness Checks

We validate our choice of $\Delta t_{max} = 8$ seconds by assessing whether two metrics from the raw truck measurement data differ between the non-systematically ($w_{a,b} < 2$) and systematically ($w_{a,b} \geq 2$) co-driving truck pairs.

The first metric is Δv: the speed difference $|v_a - v_b|$ between two co-occurring trucks within Δt_{max}. We would assume that trucks who drive systematically together for longer distances would have a lower value of Δv as their speed needs to be aligned. In Figure 3a we observe that this is indeed the case, where the result is most obvious for smaller values for Δt_{max}, up to 8 seconds.

The second validation metric is whether the considered pair of trucks is driving on the same lane or not, i.e., whether for a truck pair (a, b) it holds that $h_a = h_b$. In case of systematic co-driving behavior it is assumed to be more likely that two trucks are present in the same lane, since they do not have to overtake each other to drive together. Figure 3b shows that indeed the fraction of trucks driving on a different lane ($h_a \neq h_b$) is more than two times higher for non-systematic co-driving than for systematic co-driving trucks.

Concluding, these robustness checks convince us that the derived co-driving network captures true systematic co-driving behavior.



(a) Distributions of the difference in speed Δv for increasing Δt_{max}.

(b) The fraction of trucks driving on a different lanes for increasing values of Δt_{max}.

Fig. 3: Robustness checks for establishing systematic co-driving.

3.5 Regional Co-Driving Network

Although trucks from various countries are observed in our data, we have additional information on Dutch trucks, obtained from the Netherlands Vehicle Authority (RDW). Therefore we also consider the Dutch *regional co-driving network* consisting of trucks for which the country attribute was equal to NL (59% of the nodes) and all systematic co-driving links between these trucks, having the following additional attributes:

- *city* where the truck is registered
- empty mass m_{empty} of the truck
- maximum mass m_{max} of the truck
- *capacity* of the truck
- *company* that owns the truck
- registration date (*regdate*)
- the 4-digit (postal) zip code of where the vehicle is registered. Attributes zip_1, zip_2, zip_3 and zip_4 each indicate the location with a higher geographic precision.

4 Approach

Here we describe the techniques used to understand systematic co-driving behavior from a network perspective. We will start with outlining how node assortativity can explain the driving forces in edge formation in Section 4.1, followed by the approach to detect and understand communities within the co-driving network in Section 4.2.

4.1 Network-Driven Understanding of Co-Driving Behavior

We will use assortativity to investigate what type of common attributes explain the formation of links in the co-driving network. *Assortativity* is a measure of the preference of nodes in a network to connect with other nodes that are alike in some way [12]. The assortativity metric r_a can be computed for each nominal and numerical attribute a of the network using the definitions in [11]. It should be noted that degree assortativity is the assortativity computed for the (numerical attribute) degree, so the number of connections of the node.

An assortativity value r_a closer to 1 indicates that nodes have more links to nodes that have equal attribute a. A value closer to -1 indicates disassortivity, meaning that nodes with different values for attribute a are more likely to be connected. An assortativity of 0 for an attribute means that there is no preferential attachment of edges between nodes based on the value of attribute a.

4.2 Community-Driven Understanding of Co-Driving Behavior

To better understand the structure of the co-driving network, we investigate the communities in this network, which can provide insights in the different groups of truck drivers that together form the network. As we will show in Section 5, our network has a modular structure with a clear division into communities. To detect these communities, we use the well-known Louvain algorithm [3]. It takes as input the structure of a weighted network, and outputs an assignment of each node to a community. It furthermore has a resolution parameter γ that dictates whether a more fine-grained or coarse-grained division into communities should be found [9].

The Louvain algorithm uses heuristics to optimize the so-called *modularity* value Q, indicating the quality of the division of the network into communities. A modularity value close to 1 indicates that there are more edges within communities and fewer edges between communities. When adjusting the aforementioned resolution parameter, the value of modularity and with that the number of discovered communities C changes. At different resolutions γ, very similar values of Q can be measured, each with a different number of communities C. This so-called modularity landscape must be explored to obtain the division of the network into communities that best explains the formation of groups in the underlying system [6].

To explore these solutions automatically, we propose to use the available node attribute information. We then determine for each community and for each attribute, the assortativity within the subgraph of nodes induced by that community. Subsequently, for each community we take the highest attribute assortativity. We average this value over all communities, obtaining the proposed metric of average maximal community assortativity R, defined as follows:

$$R = \frac{1}{C} \sum_{c} \max_{a} r_a^{G(c)}$$

Here, C is the number of communities, c is one of the communities (defined as the subset of nodes in this community), a is an attribute, $G(c)$ is the subgraph induced on the nodes in community c and $r_a^{G(c)}$ is the assortativity of attribute a in subgraph $G(c)$. Based on the value of R for different divisions of the network into communities as a result of varying the resolution parameter γ, we select the division into communities for which R is highest.

5 Results

We start this section with an analysis of the co-driving network structure in Section 5.1. The results of applying the two approaches to understanding the formation of links outlined in Section 4 are discussed in Section 5.2 and Section 5.3.

5.1 Network Statistics

Network metrics, of which definitions can be found in [1], were computed using NetworkX (networkx.github.io), whereas distance metrics were computed using teexGraph (github.com/franktakes/teexgraph). The python-louvain package (github.com/taynaud/python-louvain) was used for community detection.

In Table 1 we list basic network statistics for the complete network and the regional co-driving network of measured Dutch trucks. We note that the majority of activity is captured in the giant component. The degree distribution for both networks is given in Figure 4, showing a power law distribution, suggesting that the co-driving network is scale-free. This means that a few truck drivers drive with a large number of other trucks, whereas the majority only does so with a relatively small number of others. In addition, the weight distribution in Figure 4 shows that some co-driving trucks actually very frequently drive together. The diameter (which is affected by distant outliers) is quite high, whereas the average distance is not as low as 6 as is common in many real-world networks, but is with a value of 9 still substantially low given the size of the network. The power law exponent of the degree distribution is 3.6. Together, these three metrics indicate that although the network has a very skewed degree distribution, nodes are not that as close to each other as in other real-world networks, hinting at a more modular structure, which we will investigate further in Section 5.3.

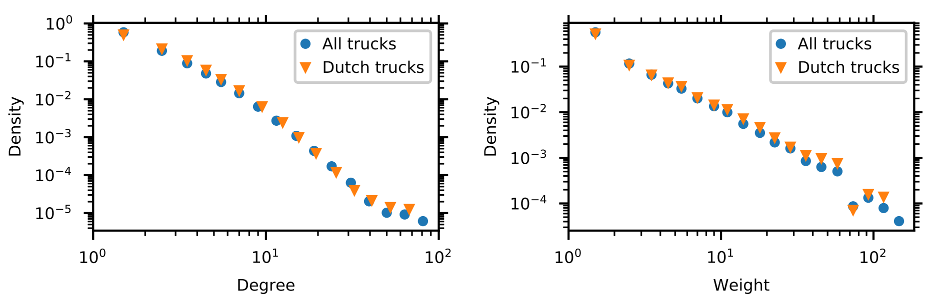


Fig. 4: Degree (L) and weight (R) distribution of the full and regional network.

Table 1: Statistics for full and regional network and their giant component (GC).

Metric	Full Network	Regional Network
Number of nodes	65,290	35,706
Number of nodes (GC)	37,858	22,511
Number of edges	68,958	36,885
Number of edges (GC)	51,730	30,851
Density	3.2×10^{-5}	5.8×10^{-5}
Density (GC)	7.2×10^{-5}	1.2×10^{-4}
Diameter (GC)	31	28
Average distance (GC)	9	9
Clustering coefficient	0.06	0.07
Power law exponent	3.58	3.61

5.2 Attribute Assortativity

Applying the metric of node assortativity discussed in Section 4.1 resulted in the values reported in Table 2. The positive value for degree assortativity may indicate that actively co-driving trucks are preferentially connected to other active co-driving trucks. Geographic information available about the trucks appears to perform best in explaining systematic co-driving behavior, with substantially high assortativity metrics for the *zip code* attributes in the regional network, but

Table 2: Assortativity of node attributes introduced in Section 3.3 and Section 3.5.

Attribute	Type	Full Network	Regional Network
degree	numeric	0.12	0.12
country	17 categories	**0.56**	-
$\tilde{v}$	numeric	0.55	0.34
n_ℓ	numeric	0.45	0.40
ℓ_{max}	17 categories	0.25	0.21
city	1,319 categories	-	0.33
m_{empty}	numeric	-	0.30
m_{max}	numeric	-	0.35
capacity	numeric	-	0.32
company	numeric	-	0.29
regdate	numeric	-	0.13
zip_4	1,975 categories	-	0.32
zip_3	718 categories	-	0.33
zip_2	90 categories	-	0.35
zip_1	9 categories	-	0.41

in particular a value of 0.56 for the *country* attribute in the full network. So, truck drivers from the same country, but also from the same city (e.g., because they live there, or are based there) are more likely to systematically drive together.

5.3 Average Maximal Community Assortativity

The results of applying community detection to the giant component of the full network are shown in Figure 5. It shows for increasing resolutions the number of communities and the modularity value. A maximum value of $Q = 0.86$ is found for resolution $\gamma = 1$. This high value is a second piece of evidence that our co-driving network is highly modular. We observe how there are a number of solutions with a similar modularity value, but a very different number of communities.

To better understand these findings, we look at the average maximal community assortativity R (see Section 4.2) shown in the bottom right of Figure 5. Although at $\gamma = 1$ the highest modularity is found, at we see that for $\gamma = 2$ (as opposed to lower values of γ), the best community division is obtained in terms of explainability using node attribute assortativity. For this value of the resolution, we find that 52 of the total 120 communities are best described using the *country* attribute, whereas the remaining attributes $\tilde{v}$, n_ℓ, and ℓ_{max} explain 30, 29 and 9 communities respectively.

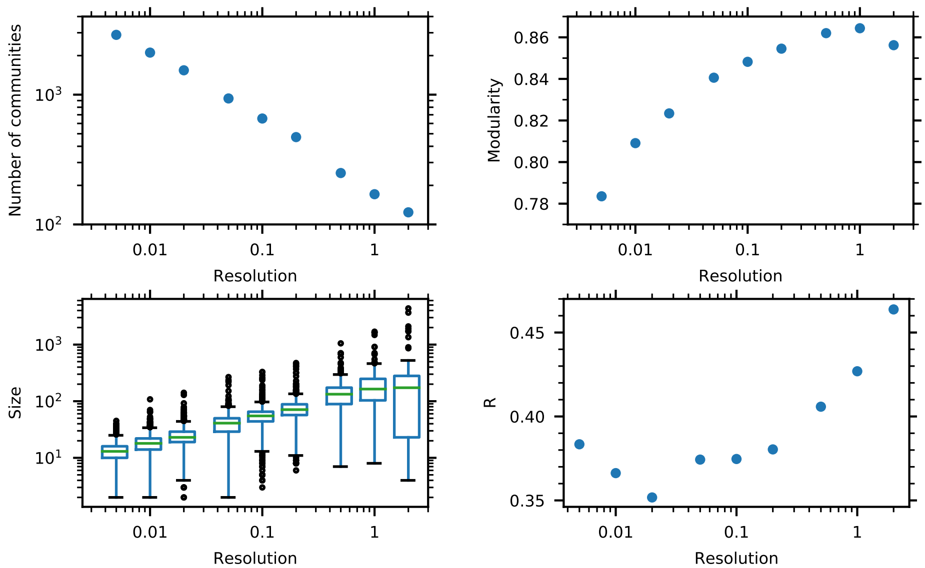


Fig. 5: For increasing values of resolution parameter γ, figures show the number of communities C (top left), modularity value Q (top right), average community size (bottom left) and average maximal community assortativity R (bottom right).

6 Conclusion

In this paper we described how a co-driving network can automatically be extracted from truck measurement data. The network exhibited typical real-world network properties such as a giant component and a scale-free degree distribution. Positive degree assortativity hinted at actively co-driving trucks that are preferentially connected to other trucks. Distance metrics as well as community detection showed a highly modular structure. Node attribute assortativity in turn showed how geographical attributes such as country and region of origin best explained co-driving links. Furthermore, using the proposed metric of average maximal community assortativity, we found that the network's highly modular community structure can be explained using different attribute's assortativity in each community, again dominated by geographical attributes.

In future work, we plan to further investigate the relation between the observed network characteristics and the considered domain, in particular in relation to similar network studies. In addition, we will extend our work to incorporate timestamps to investigate the co-driving network's dynamics. This could shed light on which truck drivers initiate co-driving behavior and under what conditions behavior diffuses to other nodes. A better understanding of how fast and through which drivers specific behavior spreads would enable interventions to educate drivers with best practices, saving fuel and reducing traffic congestion.

References

1. Barabási, A.L.: Network Science. Cambridge University Press (2016)
2. Barrat, A., Cattuto, C.: Temporal networks of face-to-face human interactions. In: Temporal Networks, pp. 191–216. Springer (2013)
3. Blondel, V.D., Guillaume, J.L., Lambiotte, R., Lefebvre, E.: Fast unfolding of communities in large networks. J. Stat. Mech. Theory Exp. **2008**(10), 1–12 (2008)
4. da Cunha, B.R., Gonçalves, S.: Topology, robustness, and structural controllability of the brazilian federal police criminal intelligence network. Appl. Netw. Sci. **3**(1), 36 (2018)
5. Fortunato, S., Hric, D.: Community detection in networks: A user guide. Phys. Rep. **659**, 1–44 (2016)
6. Good, B.H., de Montjoye, Y.A., Clauset, A.: Performance of modularity maximization in practical contexts. Phys. Rev. E **81**(4), 046,106 (2010)
7. Kassarnig, V., Mones, E., Bjerre-Nielsen, A., Sapiezynski, P., Dreyer Lassen, D., Lehmann, S.: Academic performance and behavioral patterns. EPJ Data Sci. **7**, 1–16 (2018)
8. Kleinberg, J.: The small-world phenomenon: An algorithmic perspective. In: Proc. 31st Annual ACM Symposium on Theory of Comput., pp. 163–170. ACM (2000)
9. Lambiotte, R., Delvenne, J.C., Barahona, M.: Laplacian dynamics and multiscale modular structure in networks. IEEE Trans. Netw. Sci. Eng. **1**(2), 76–90 (2014)
10. Newman, M.E.J.: Assortative mixing in networks. Phys. Rev. Lett. **89**(20), 208,701 (2002)
11. Newman, M.E.J.: Mixing patterns in networks. Phys. Rev. E **67**(2), 026,126 (2003)
12. Newman, M.E.J.: Networks: An Introduction. OUP Oxford (2009)
13. Tsugawa, S., Kato, S.: Energy its: another application of vehicular communications. IEEE Commun. Mag. **48**(11) (2010)