Available online at www.sciencedirect.com

ScienceDirect

Transportation Research Procedia 46 (2020) 237–244

Transportation Research Procedia

www.elsevier.com/locate/procedia

City Logistics 2019

A matching algorithm to study the evolution of logistics facilities extracted from GPS traces

Nadia M. Trent[a], Johan W. Joubert[a]*, Győző Gidofalvi[b], Behzad Kordnejad[c]

[a]Centre for Transport Development, University of Pretoria, Private Bag X20, Hatfield, 0028, South Africa
[b]Department of Urban Planning and Environment and Integrated Transport Research Lab, KTH Royal Institute of Technology, Teknikringen 10A, 114 28 Stockholm, Sweden
[c]Division of Transport Planning, Department of Civil and Architectural Engineering, KTH Royal Institute of Technology, Binellvägen 23, 100 44 Stockholm, Sweden

Abstract

The ubiquity of anonymous GPS data has opened the door to a promising data source in the field of city logistics modelling. From this data the locations of logistics facilities can be extracted and their evolution studied over time. The minimum edge cover problem, weighted by the Hausdorff distance, is used as a basis for a matching algorithm to study the location of facilities across longitudinal datasets. The efficacy and validity of the algorithm is assessed through the visual inspection of results in three urban areas across five time instances. Prevalent errors are unpacked and algorithm modifications suggested. This paper makes a methodological contribution to the handling of GPS data for the purpose of city logistics modelling.

Keywords: GPS data; cluster evolution; logistics facilities; Hausdorff distance; minimum weight edge cover; concave hull clustering

1. Introduction

Gridlocked roads, ever-present smog curtains, noise and irritation are only some of the symptoms of urban transport systems buckling under the pressure of rapid urbanisation. Public agencies have come to appreciate that better policy decisions related to city logistics must be undergirded by appropriate quantitative impact assessments (de Jong et al., 2013). These assessments cannot be blanket extensions of passenger transport models, but must come from

* Corresponding author. Tel.: +27-12-420-2843; fax:+27-12-362-5103.
E-mail address: johan.joubert@up.ac.za

2352-1465 © 2020 The Authors. Published by Elsevier B.V.

10.1016/j.trpro.2020.03.186

appropriately designed freight transport models that reflect the dynamics and decision-making in the logistics industry (Joubert et al., 2010).

There are different approaches to freight transport modelling and the interested reader is referred to de Jong et al. (2004, 2013) for classifications and examples. When it comes to urban logistics, agent-based modelling has proven superior to aggregate and gravity-based models in its ability to capture the autonomous daily decisions that drive logistics behaviour (Anand et al., 2014). These models are disaggregated, defining individual actors (shippers, carriers, receivers) that interact on a daily basis using defined locations on an established transportation network. Logistics behavioural modelling (characterising the interactions between actors) is an emerging field of research parallel to this paper (Bean and Joubert, 2018). Meanwhile, this paper focuses on another element that is important in agent-based modelling – knowing where the logistics facilities are and how this changes as time progresses. Apart from its use in agent-based modelling of city logistics, tracking facility location over time is also important in studies related to logistics sprawl. This has become a very prevalent topic in city logistics circles (Aljohani and Thompson, 2016), nonetheless identifying facility locations and tracking these over time is not as trivial as it seems.

Data, or the lack thereof, is one of the primary limiting factors in freight transport modelling. In urban planning circles, a number of different approaches have been used to identify logistics facilities. When available, researchers can make use of officially recorded business registers (Cidell, 2010; Giuliano and Kang, 2018). Some studies combine business registers, archives and contextual information to triangulate facility locations (Dablanc and Rakotonarivo, 2010). In the absence of official registers, systematic searches of business directories can be combined with targeted surveys (Coetzee and Swanepoel, 2017). Another approach is to leverage off comprehensive commodity/freight flow surveys like the ones in Tokyo (Sakai et al., 2015, 2017) and Sweden (Abate et al., 2018). Each of these approaches have benefits and limitations. In general, there are trade-offs between five data attributes: geographic disaggregation, data richness (e.g. identifying facility function or commodity type), coverage, standardisation & repeatability, and longitudinal potential (Viljoen and Joubert, 2019).

The advent of Global Positioning System (GPS) tracking, in vehicles, on mobile phones and even in smart watches, now opens up many possibilities for alternative sources of data. It is possible to extract the location of logistics facilities starting with nothing more than the anonymous GPS traces of thousands of commercial vehicles. From these traces, activity chains that identify where and when logistics activities were performed by commercial vehicles can be extracted (Joubert and Axhausen, 2011). Density-based algorithms can then be employed to identify the location of a concentration of logistics activities as a logistics facility (Joubert and Meintjes, 2015). Once the locations of the facilities have been extracted and the logistics activity between these facilities mapped, researchers take a step further to study the complex network of supply chain interactions (Joubert and Axhausen, 2013; Joubert and Meintjes 2015; Viljoen 2018; Viljoen and Joubert, 2019).

Developing the data science tools required to wrangle GPS “big data” is ongoing. The tools already exist to identify logistics facilities from anonymous GPS traces. The next step is the longitudinal study of these facilities over time. Using a density-based clustering approach to identify facilities means that each distinct dataset could identify the same shopping centre with a unique polygon that has a unique centroid. Fig. 1 shows an example of this phenomena. Without a methodology which identifies that unique, yet similar polygons actually point to the same facilities in different datasets, each polygon would incorrectly be identified as a unique facility.

This paper makes a methodological contribution to the field of city logistics. A custom matching algorithm is developed to match facilities identified from a time series of GPS traces. This algorithm is tested using data from three urban areas in South Africa at five distinct points in time. The next section lays the contextual foundation for the algorithm followed by a detailed description of the methodology. Thereafter, a validation of the algorithm’s performance is presented. The paper concludes with proposed algorithm modifications to improve performance.

2. Conceptual description

Consider the polygons shown in Fig. 2. The left pane contains all the logistics facilities around the corner of Maltzan and Charlotte Maxeke Streets, Pretoria (South Africa) extracted from the March 2010 GPS traces. Each red polygon represents a facility. Similarly, the right pane contains all the logistics facilities extracted from the GPS traces in the same area during March 2012.

The objective is to track the locations of logistics facilities over time. To do so, polygons from the two panes that point to the same facility must be matched. For example, in Fig. 2 there are three matches. Polygon 2 (left) refers to

Fig. 1. An example of three logistics facilities identified by the concave hull clustering algorithm (Joubert and Meintjes, 2015b). The red polygons map the shape of each cluster and the index in the black circle shows which dataset produced that cluster (1: March 2010, 2: March 2011, 3: March 2012, 4: March 2013, 5: March 2014).

the same facility as polygon 4 (right), therefore these are matched on a one-to-one basis (A) and the facility is *preserved* over time. Polygons 1 and 2 (right) both point to the same logistics complex as polygon 1 (left) and therefore these are matched (B) as a *split*. (A *merge* is the opposite of a split, where multiple polygons in the earlier dataset refer to the same polygon in the latter dataset.) Finally, polygon 3 (right) does not have a corresponding polygon to match to and thus is not matched but regarded as a *birth* (C) – a facility that showed no logistics activity in the earlier dataset. (A *death* is the opposite of a birth, where a facility that did exhibit logistics activity in an earlier dataset no longer does so in the latter dataset.) These matchings must be determined objectively for large datasets.

Spatial-temporal analysis of polygons is an important field of inquiry in geospatial data science. Sadahiro and Umemura (2001) characterised four ways in which an immobile polygon could evolve based on two-way area overlap (TWAO). Liu et al. (2018) identified limitations of TWAO approaches and instead proposed an algorithm based on the overlap of minimum bounding rectangles that rivals the accuracy of TWAO approaches. However, both these studies assume immobile polygons. While logistics facilities are indeed immobile, the polygons extracted from the dataset that identify the facilities are not. The concave hulls emanating from vehicle parking behaviour are mobile around the boundaries of a facility resulting in no overlap at all.

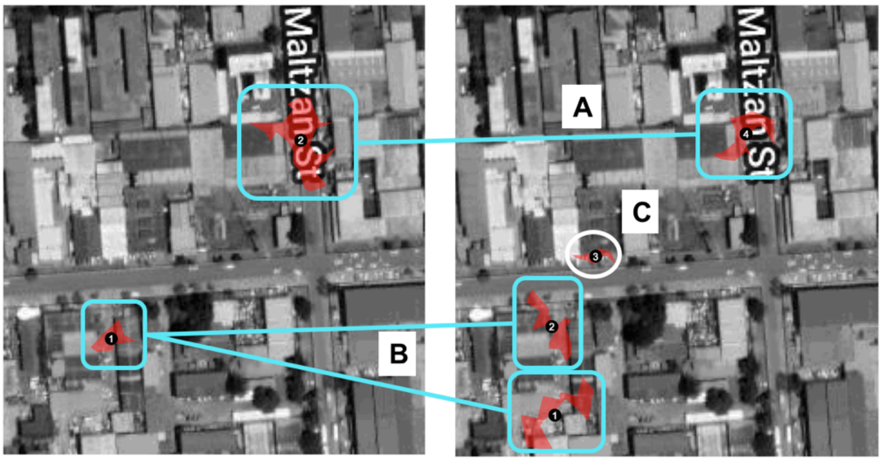

Fig. 2. *Left*: All the logistics facilities identified around the corner of Maltzan and Charlotte Maxeke streets, Pretoria from the March 2010 GPS traces. *Right*: All the logistics facilities identified in the same area from the March 2012 dataset. A: *preservation*; B: *split*; C: *birth*.

Robertson et al. (2007) expanded the work of Sadahiro and Umemura (2001) by developing the Space Time Analysis for Moving Polygons (STAMP) method for application to moving polygons such as wildfires (Robertson et al., 2007) and plague regions (Morris et al. 2013). STAMP allows the matching of non-overlapping polygons that are closer to each other than some threshold d and are not matched to other polygons. While this approach is closer to what is required for our application, the polygon events of both Sadahiro and Umemura (2001) and STAMP are unnecessarily detailed to answer the simple question of whether polygons in different datasets refer to the same facility.

Huh et al. (2014) use a simpler weighted bi-partite graph as a starting point where the graph edges are weighted by an overlap-based metric. They then use agglomerative hierarchical clustering to match polygons. The weighted bipartite graph provides a simpler representation appropriate to matching polygons in this application.

While overlap and Euclidean distance are commonly used in polygon matching studies, an alternative metric exists. The Hausdorff distance is a similarity metric that takes into account both shape similarity and displacement. The original metric and its extensions are considered by Min et al. (2007) as more sophisticated in spatial object matching and are used by Tong et al. (2014) to successfully match databases of road networks in Shanghai, China. Therefore, our proposed approach uses a weighted bi-partite graph representation where edges are weighted by the Hausdorff distance. Our matching is executed by solving the minimum edge cover problem on this graph.

3. Methodology

Consider that D1 and D2 are two independent sets of vertices in a weighted bi-partite graph where each vertex represents a polygon. Every edge of the graph connects a vertex in D1 to a vertex in D2. The weight of the edge is the Hausdorff distance between the incident vertices. Thus, the higher the weight of the edge, the more dissimilar those two polygons are and the less likely it is that they refer to the same logistics facility. The formulation must make it possible for every vertex in D1 and D2 to be matched to zero, one or more vertices in the other dataset to allow for births/deaths, preservations, and splits/merges as previously described. Therefore, the cardinality of the weighted bi-partite graph is $\mathrm{N:M}$ where N and M are whole numbers. To find the most appropriate matching of facilities between two datasets, the sum of the edge weights (Hausdorff distances) of the matching between vertices in D1 and D2 within the weighted bi-partite graph must be minimised.

3.1. Partial matching produced by the Hungarian method

Let $\mathrm{G = (U, V, E)}$ be the weighted bi-partite graph where $\mathrm{U}$and $\mathrm{V}$are two independent sets of vertices (for example $\mathrm{U = D1}$ and $\mathrm{V = D2}$) and $\mathrm{E = \{e_{uv}\}}$ is the set of binary variables that indicates whether vertex $\mathrm{u \in U}$ and vertex $\mathrm{v \in V}$ are matched ($\mathrm{e_{uv} = 1}$) or not ($\mathrm{e_{uv} = 0}$). Further, let $\mathrm{H_{uv}}$ be the Hausdorff distance between each vertex u and v where $\mathrm{u \in U}$, $\mathrm{U = \{1,2,3 \dots, n\}}$ and $\mathrm{v \in V}$, $\mathrm{V = \{1,2,3 \dots, m\}}$. Then the objective function of minimising the cumulative weight of the graph's edges is represented by equation (1):

$$min\ \sum_{u \in U} \sum_{v \in V} e_{uv} H_{uv} \tag{1}$$

To reduce computational time, H_{uv} is only calculated for vertex pairs which have centroids that are closer than 5km. For pairs further apart, a value is recorded that is orders of magnitude greater than the largest Hausdorff distance in the matrix. A partial matching is produced by an implementation of the Hungarian method (Munkers, 1957) in R using the clue package (Hornik, 2018).

The partial matching includes only one-to-one matches. But the datasets are not of equal size ($\mathrm{|U| \neq |V|}$) and therefore dummy vertices need to be inserted. If a vertex is matched to a dummy vertex, it basically means that it is a birth or a death. The Hausdorff distances between each dummy vertex and all the vertices in the adjacent dataset must be set to some predetermined value. These distances cannot be calculated as the dummy vertices have no geographic coordinates nor shape. This parameter has a pertinent impact on the final solution and defining an appropriate value is non-trivial. Preliminary testing led to the suggestion of using percentiles of the distribution of the Hausdorff distances between all the other vertex pairs. This parameter is henceforth referred to as the birth-death parameter, or bd, where bd is the percentile of the Hausdorff distance distribution that determines the distance between a dummy vertex and every vertex in the adjacent dataset.

3.2. Inserting merges and splits using a greedy algorithm

The partial matching includes only preservations and births/deaths. The possibility of merges/splits is not accommodated. Thus an iterative greedy algorithm is used to replace preservations and births/deaths with merges/splits in cases where this would improve the objective function in equation (1). In each iteration, the 5% of edges that have the highest weight are selected and their matching reconsidered. If there is an alternative matching (i.e. merge or split) that would have a lower weight, the current matching is replaced by this alternative. This algorithm converges to a local optimum within 20 iterations. A post-execution procedure is applied that identifies any vertices that are part of both a merge and a split matching. Such a situation does not make intuitive sense. In these cases, one of the edges is removed so that the vertex is only party to either a merge or a split matching.

In the final dummy modified weighted bi-partite graph, each vertex participates in exactly one type of cluster evolution matching, therefore it is still a minimum weight edge cover solution.

4. Algorithm validation

To study the validity of this algorithm, GPS datasets from three urban areas in South Africa namely the Gauteng Province, eThekwini Metropolitan Municipality and the City of Cape Town are used. The original data set, courtesy of *Digicore Technologies*, provided the detailed vehicle movements of approximately 16 000 commercial vehicles for the period January 2010 to May 2014 across the entire southern Africa. Each record contained typical GPS data: anonymous vehicle id, a time stamp, longitude and latitude, and an ignition on/off signal. Extracting activity chains from the GPS records is similar to that described in Joubert and Axhausen (2011). For each activity extracted we have an anonymous vehicle id, the location of the activity, and its start and end time (date and time of day). Indeed, there is a selection bias in the data as we only have vehicles of companies who subscribe to *Digicore*'s telemetry service, *Ctrack*. Still, the focus in this paper is not on interpreting any absolute values of the chain extraction process. Instead, the data is valuable in demonstrating the matching algorithm over a longitudinal data set.

Readers are referred to Viljoen and Joubert (2019) for a description of the development of the city logistics networks in the three study areas. Five data snapshots are taken for March 2010, March 2011, March 2012, March 2013 and March 2014. For each of these months, activity chains are constructed and a database of logistics facilities is generated using concave hull clustering (Joubert and Meintjes, 2015). Without a customised algorithm, there would be no way to match facilities in one dataset to another.

In the absence of any benchmark results, algorithms or gold standards, validation is based on visual inspection by the authors. Each metropolitan area is divided into a tessellation of hexagons. Each hexagon is assigned a weight based on the sum of facilities over all five datasets identified in that hexagon. Weighted sampling is then used to select 20 hexagons (called Validation Areas or VAs) for validation. Each VA is assigned a random year-combination. The algorithm was executed three times with $bd \in \{0.05\%, 0.07\%, 0.09\%\}$. Thus each VA had three sets of results, one for each value of bd, resulting in 60 sets of results to evaluate visually. Evaluation took between 15 minutes and 1 hour per result. The sample size of 20 VAs and three values of bd was admittedly limited. However, at this prototypical stage of algorithm development, this sample size was sufficient to identify design amendments for the next iteration of algorithm development.

With the caveat of a limited sample size in mind, the objective of the validation was threefold: Determine the error rate of the algorithm; test the sensitivity of the results to bd, and identify prevalent errors.

4.1. Error rate of the algorithm

Each VA was visually scrutinised to determine the number of incorrect matches. In Fig.3 the cumulative number of incorrect matches are plotted against the total number of matches in each VA. The mean error rates were 9%, 13% and 19% when using $bd \in \{0.05\%, 0.07\%, 0.09\%\}$, respectively. The higher the value of bd, the lower the likelihood of births and deaths occurring and the higher the likelihood of incorrect merge, split and preservation matches.

However, in some cases the higher bd values yielded more accurate results. Thus, the mechanism that caused an error in one VA was often the same mechanism that caused a more accurate match in another. What is conclusive is that the algorithm is definitely sensitive to the value of bd and investing time to refine this parameter is warranted.

4.2. Prevalent errors and design implications

In evaluating the VAs, prevalent errors emerged. In this section three of these errors are discussed and potential algorithm improvements suggested.

Polygon similarity trumps geographic distance. The Hausdorff distance, by definition, measures both shape similarity and displacement in two planes. However, in this context the shape of the polygons has far more influence in the metric than the distance between centroids. This can result in errors such as the one displayed in Fig. 4 a) where

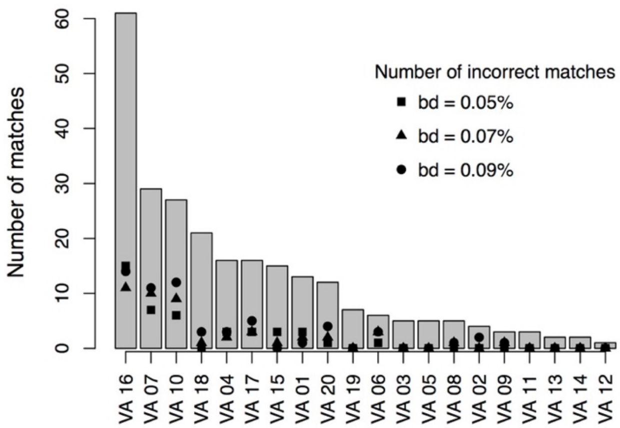

Fig. 3. Incorrect matches plotted against the actual number of matches for each VA.

the larger blue polygon should have been matched to the red polygon that underlies it, but instead it was matched to the red polygon that is further away but more similar in shape. The suggested improvement is to either add relative overlap or the geographic distance between centroids to the measure of similarity or to investigate the extended Hausdorff distance proposed by Min et al (2007) that dampens the effect of shape dissimilarity.

Ignorance of underlying land use. What is an acceptable geographic distance between two polygons depends on whether these occur in industrial or suburban areas as facilities are much larger in the former. Fig. 4 b) shows an example of the errors that can occur when the algorithm doesn't take land use into account. In the top left corner, two polygons are correctly matched as they lie on either side of a large industrial building. Yet in the bottom right corner, an incorrect matching occurs between polygons in a suburban area that are separated by roughly the same geographic distance. The only way to prevent such errors would to add parameters to the algorithm that indicate the underlying land use of a polygon. Not only would this be greatly cumbersome computationally, but the underlying land use data is not readily available in two of these three metropolitan areas.

Spill-over errors from the concave hull algorithm. The concave hull algorithm that defines a polygon as a logistics facility remains an approximate approach. Errors made by this algorithm in defining the polygons cannot be corrected by the matching algorithm. A very good matching is shown in Fig. 4 c). However, looking at the underlying map imagery it is clear that both the red and blue polygons – each supposed to represent only one facility – cover a number of facilities in real life. The concave hull algorithm should have split each of these large polygons into smaller ones. Conversely, Fig. 4 d) shows a scenario where the three red polygons should have been combined in the earlier dataset and the two blue polygons in the latter as they all pertain to different sides of the same facility in reality. These errors should ideally be prevented by refinements to the concave hull algorithm.

Based on the outcome of the validation process, it was decided that the algorithm is not yet ready for deployment. However, in its current state it provides a solid and practically implementable foundation for future modifications.

5. Proposed algorithm modifications

Proposed algorithm modifications are discussed in order of increasing perceived difficulty to implement. Firstly, the sample size of VAs could be increased and the evaluation procedure automated to obtain a more generalisable view of the algorithm's performance. Once this is in place, more extensive experimentation should be executed.

The second group of modifications relate to the formulation and algorithm implementation. The concept of similarity should be refined, either by creating a single metric that balances polygon similarity and overlap/geographic distance or through a multi-objective formulation. The current greedy algorithm could be replaced by an algorithm that is less likely to get stuck in local optima.

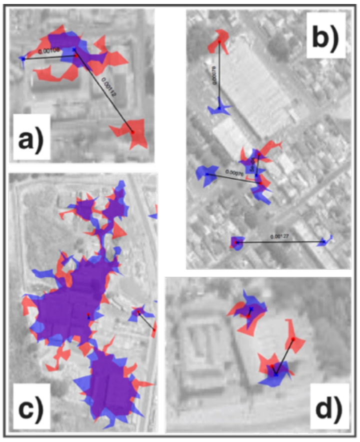

Fig. 4. Prevalent errors discovered during the validation process.

Thirdly, a consensus approach could be used by solving multiple bi-partite graphs simultaneously. For example, three bi-partite 2010-2011, 2011-2012 and 2010-2012 could be solved interdependently using some measure of consensus to validate matchings.

The final group of modifications relate to factors currently external to the matching algorithm. The density-based algorithm could be improved or land use data could be included as an additional parameter in the matching algorithm.

6. Conclusion

The ability to identify and track logistics facilities has long been problematic in transport geography in general and city logistics in particular – especially in studies relating to logistics sprawl. GPS traces offer a complementary source of data for modelling purposes. This anonymous data can be used to identify logistics facilities in urban areas using density-based clustering. Even though GPS data does not offer the contextual richness of business archives or surveys, it is a standardised data source with good repeatability and transferability.

Any longitudinal study of facilities identified from GPS data requires an algorithm to match facilities over time. This paper presented a custom matching algorithm that uses the Hausdorff distance as similarity metric and a minimum weight edge cover formulation to match facilities. The results were visually validated across three urban areas. It was found that this first prototype of the algorithm is not ready for deployment, but noteworthy observations led to proposed algorithm modifications. This paper is a significant methodological contribution to the ongoing development of data science tools that support research addressing logistics sprawl and agent-based modelling in city logistics.

Acknowledgements

The authors gratefully acknowledge the sponsorship of the Volvo Research Foundation (Grant SVG-2018-02) and a special thanks also to Prof. Erik Jenelius whose suggestions improved the algorithm.

References

Abate, M., Vierth, I., Karlsson, R., de Jong, G., Baak, J., 2019. A disaggregate stochastic freight transport model for Sweden. Transportation 46.3, 671-696.

Aljohani, K. and Thompson, R.G., 2016. Impacts of logistics sprawl on the urban environment and logistics: Taxonomy and review of literature. Journal of Transport Geography 57, 255-263.

Anand, N., van Duin, R., and Tavasszy, L., 2014. Ontology-based multi-agent system for urban freight transportation. International Journal of Urban Sciences 18.2, 133-153.

Bean, W.L., and Joubert, J.W., 2018. A systematic evaluation of freight carrier response to receiver reordering behaviour. Computers & Industrial Engineering 124, 207-219.

Cidell, J. (2010). Concentration and decentralization: the new geography of freight distribution in US metropolitan areas. Journal of Transport Geography 18, 363-371.

Coetzee, P. J. and Swanepoel, P. A., 2017. Spatial relationships and movement patterns of the air cargo industry in airport regions. Journal of Transport and Supply Chain Management 11, 297.

Dablanc, L. and Rakotonarivo, D., 2010. The impacts of logistics sprawl: How does the location of parcel transport terminals affect the energy effciency of goods' movement in Paris and what can we do about it? Procedia – Social and Behavioral Sciences, 2:6087-6096.

de Jong, G., Gunn, H., and Walker, W., 2004. National and international freight transport models: An overview and ideas for future development. Transport Reviews 24.1, 103-124

de Jong, G., Vierth, I., Tavasszy, L., and Ben-Akiva, M., 2013. Recent development is national and international freight transport models within Europe. Transportation 40, 347-371.

Giuliano, G. and Kang, S., 2018. Spatial dynamics of the logistics industry: Evidence from California. Journal of Transport Geography 66, 248-258.

Hornik, K., 2018. *clue: Cluster ensembles*. R package version 0.3-56, https://CRAN.R-project.org/package=clue.

Joubert, J. W. and Axhausen, K. W., 2011. Inferring commercial vehicle activities in Gauteng, South Africa. Journal of Transport Geography 19, 115-124.

Joubert, J. and Axhausen, K. W., 2013. A complex network approach to understand commercial vehicle movement. Transportation 40.3, 729-750.

Joubert, J.W., Fourie, P.J., and Axhausen, K.W., 2010. Large-scale agent-based combined traffic simulation of private cars and commercial vehicles. Transportation Research Record 2168, 24-32.

Joubert, J. W. and Meintjes, S., 2015. Computational considerations in building inter-firm networks. Transportation 42.5, 857-878.

Joubert, J.W., de Beer, D. and de Koker, N., 2016. Combining accelerometer data and contextual variables to evaluate the risk of driver behaviour. Transportation Research Part F 41, 80-96.

Liu, L., Zhu, X., Zhu, D. and Ding, X., 2018. M:N Object matching on multiscale datasets based on MBR combinatorial optimization algorithm and spatial district. Transactions in GIS 22,1573-1595.

Min, D., Zhilin, L. and Xiaoyong, C., 2007. Extended Hausdorff distance for spatial objects in GIS. International Journal of Geographical Information Science 21.4, 459-475.

Morris, L.R., Blackburn, J.K., Talibzade, A., Kracalik, I., Ismaylova, R. and Abdullahyev, R., 2013. Informing surveillance for the lowland plague focus in Azerbaijan using a historic dataset. Applied Geography 45, 269 – 279.

Munkres, J., 1957. Algorithms for the assignment and transportation problems. Journal of the Society for Industrial and Applied Mathematics 5.1, 32-38.

Robertson, C., Nelson, T.A., Boots, B. and Wulder, M.A., 2007. STAMP: spatial-temporal analysis of moving polygons. Journal of Geographical Systems 9, 207-227.

Sadahiro, Y. and Umemura, M., 2001. A computational approach for the analysis of changes in polygon distributions. Journal of Geographical Systems 3,137-154.

Sakai, T., Kawamura, K., and Hyodo, T., 2015. Locational dynamics of logistics facilities: Evidence from Tokyo. Journal of Transport Geography 46, 10-19.

Sakai, T., Kawamura, K., and Hyodo, T., 2017. Spatial reorganization of urban logistics system and its impacts: Case of Tokyo. Journal of Transport Geography 60, 110-118.

Tong, X., Liang, D. and Jin, Y., 2014. A linear road object matching method for conflation based on optimization and regression. International Journal of Geographical Information Science 28.4, 824-846.

Viljoen, N.M., 2018. Quantifying supply chain vulnerability using a multilayered complex network perspective. PhD Thesis, chapter 7. University of Pretoria, South Africa.

Viljoen, N. M. and Joubert, J. W., 2019. Supply chain micro-communities in urban areas. Journal of Transport Geography 74, 211-222.