CASPT 2018 Full Paper

Using passenger flows to determine key interchange connections for public transport synchronization

Menno Yap · Ding Luo · Oded Cats

Abstract For large urban networks and hubs, optimizing transfer synchronization becomes computationally challenging. The objective of this paper is therefore to develop a generic, data-driven methodology to determine the key line/direction-combinations to synchronize based on passenger flows. We developed an approach to detect communities of directional lines based on passenger transfer flows, by calculating modularity using a C-space inspired network representation. Our results show intuitive clusters to prioritize for synchronization on a network level for tactical planning, and on the hub level for real-time coordination.

Keywords: Clustering · Hubs · Passenger flow · Synchronization · Transfers

1 Introduction

Transfers are an inevitable part of public transport journeys, since it is not economically viable to directly connect all origin-destination pairs in a network. Empirical studies however show that transfers are perceived as one of the most negative components in the public transport journey (e.g. Schakenbos et al., 2016; Van Oort et al., 2016). Therefore, several studies focus on improving the transfer experience at hubs, for example by transfer synchronization (e.g. Goverde, 1998) or by improving the waiting experience (e.g. Van Hagen and Sauren, 2013).

Menno Yap
Department of Transport and Planning, Delft University of Technology
Delft, the Netherlands
Email: M.D.Yap@TUDelft.nl

Ding Luo
Department of Transport and Planning, Delft University of Technology
Delft, the Netherlands
Email: D.Luo@TUDelft.nl

Oded Cats
Department of Transport and Planning, Delft University of Technology
Delft, the Netherlands
Email: O.Cats@TUDelft.nl

Although optimizing transfer synchronization has been studied by several researchers, there are limits in terms of scalability and complexity for optimizing synchronization for both tactical planning and real-time operations. For example, Lee et al. (2014) consider the impact of synchronizing two lines during tactical planning on service reliability, whereas Gavriilidou et al. (2016) study real-time synchronization of two tram lines based on passenger data. Nesheli and Ceder (2014) compare the effects of different control tactics on optimal synchronization, applied to a case study network of three bus lines with two transfer locations. However, for large, real-world urban public transport networks with multiple lines and transfer locations, optimizing transfer synchronization becomes mathematically expensive, if not infeasible. Optimizing transfer synchronization is considered NP-hard due to the combinatorial nature of the problem (Desaulniers and Hickman 2007). For practical problems in larger real-world PT networks, computation time for solving this problem can rise substantially, making it infeasible to solve. A hub with $|l|$ lines provides $(2 * |l|) * (2 * |l| - 2)$ transfer possibilities, excluding transfers to the same line in either direction. Thus, a hub with 15 lines already provides 840 transfer possibilities, which makes optimizing real-time coordination between all lines simultaneously computationally challenging. Enumerating all transfer possibilities for the urban metro, tram and bus lines of a large and high-density real-world network such as London (Figure 1) would make a network-wide optimization of the synchronization of all transfer possibilities infeasible within reasonable computation times.

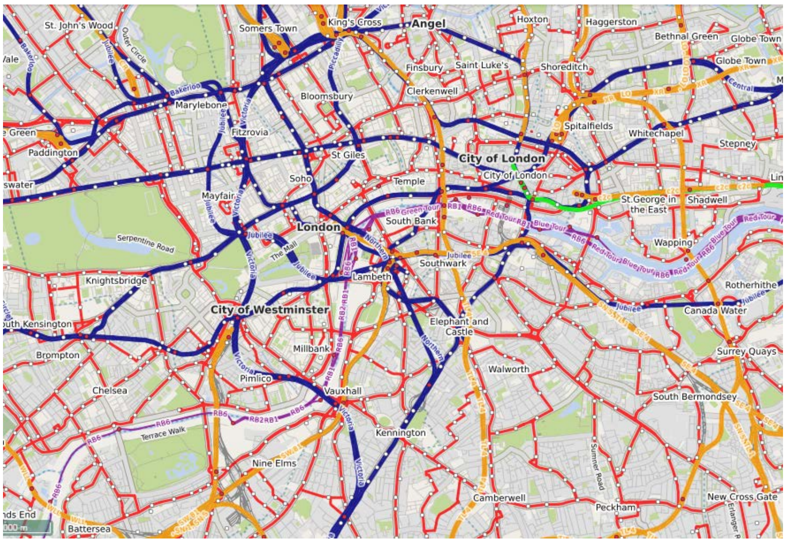


Fig. 1 Illustration transfer possibilities for the London public transport network (Map memomaps.de CC-BY-SA, map data Openstreetmap ODbL)

Our research objective is therefore to develop a generic, data-driven methodology to determine the most important line connections based on passenger flow data, in order to select key line/direction-combinations to synchronize on a network level for tactical planning and specifically for a hub for real-time coordination. We detect communities of lines with strong transfer relations by applying modularity to a public transport network or hub using a C-space inspired representation.

2 Methodology

In this section our methodology is explained. First, the inference of transfer flow data is explained (2.1), after which the C-space network representation is addressed (2.2) to which modularity is applied to identify line communities to synchronize (2.3).

2.1 Passenger transfer flow input

As input we use passenger transfer flows obtained from Automated Fare Collection (AFC) systems for the AM peak and PM peak respectively, in order to investigate whether different key interchange connections exist for different time periods. Table 1 provides the format of the AFC data, where each row represents a separate AFC transaction. Only AFC transactions of morning and evening peaks without disruptions are included in our dataset, to make sure that synchronization priorities are determined based on regular passenger flow distribution patterns. Each AFC transaction represents a passenger journey leg. To determine whether a passenger alighting is considered a transfer or final destination, we apply the transfer inference algorithm detailed in Yap et al. (2017), which is an extension of the algorithm developed by Gordon et al. (2013). This results in a transfer flow matrix with the number of transferring passengers between each line/direction combination for the AM and PM peak period. This matrix is obtained for the whole urban public transport network considered for tactical planning purposes, as well as for a specific hub for real-time synchronization purposes.

Table 1 Illustration format AFC data

Tap-in date and time	Tap-in stop	Tap-in line	Tap-out date and time	Tap-out stop	Trip-ID	Vehicle ID	Smart-card ID
4-3-2018 11:42:37	35309	6	4-3-2018 12:03:19	34997	3423	3050	81675688
4-3-2018 12:15:57	30091	18	4-3-2018 12:23:04	32857	6545	187	81675688

2.2 Network representation in C-space

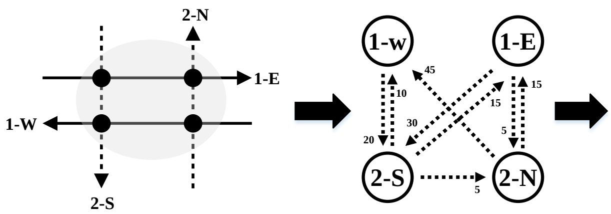
Inspired by the public transport C-space representation where individual lines are represented as nodes and are connected (via a link) only if they share common transfer stops, the transfer topology is represented as a directed graph $G = (V, E)$. Each node $v \in V$ corresponds to a public transport line $l \in L$ in a certain direction, whereas each link $e \in E$ represents an observed transfer possibility between two lines. An illustration is presented in Figure 2. Links are also weighted by two different types of attributes, either the passenger transfer flow w^q or passenger transfer waiting time w^t. By applying two different link weights, clustering results can be compared when only passenger transfer flows are incorporated, or when passenger transfer flows and the expected transfer waiting time are incorporated. The first type of link weight corresponds to the number of passengers transferring between two lines in a certain direction, which is defined as follows:

$$w_{ij}^q = q_{v_i,v_j} \tag{1}$$

In this equation w_{ij}^{q} denotes the flow-based weight between node i and j (line i and j in a specific direction) and q_{v_i,v_j} denotes the observed transfer flow between line i and j. The second type of link weight w^t relates to passenger transfer waiting time which is calculated as follows:

$$w_{ij}^{t} = \frac{h_j * q_{v_i,v_j}}{2} \tag{2}$$

In this equation w_{ij}^{t} denotes the passenger transfer time based weight between node i and j (line i and j in a specific direction). q_{v_i,v_j} and h_j, respectively, denote the observed transfer flow between line i and j, and the planned headway of line j. For the tactical planning purpose of our study, the entire public transport network is represented in C-space. Given the direction-specific network representation, the graph consists of $|l|$*2 nodes. The link label represents the total transfer flow or transfer waiting time between two lines. These transfers can occur anywhere within the considered network, hence the link label does not have a direct geographical meaning. For the second study purpose of real time synchronization, a specific hub is represented in C-space. In this case, the number of nodes equals twice the number of lines serving that specific hub, whereas the link label represents the intra-hub passenger transfer flow or transfer waiting time.



	1-W	1-E	2-S	2-N
1-W	0	0	20	0
1-E	0	0	30	5
2-S	10	15	0	5
2-N	45	15	0	0

Fig. 2 Illustration of network representation for a hub transfer pattern. The original layout of an identified hub (shaded area) is presented on the left with four directed lines marked, i.e. 1-E, 1-W, 2-S and 2-N. The transfer pattern is then represented as a graph (middle). The graph label matrix is displayed on the right.

2.3 Modularity

Based on the constructed networks (graphs) labelled with observed transfer patterns, a *community detection* technique from the field of complex network science is applied to identify line bundles. In essence, the problem that community detection intends to address is to partition a network into communities of densely connected nodes, with the nodes belonging to different communities being only sparsely connected. In our application, line bundles will thus become the partitioning result, in which within-transfer flows/waiting time are maximized.

Given our aforementioned objective, an optimization-based method called the Louvain method is adopted to identify line bundles. Proposed by Blondel et al. (2008), the Louvain method is a heuristic method based on modularity optimization. As a class of community detection method that has received the greatest attention from researchers, the optimization technique aims at finding an extremum - usually the maximum - of a function indicating the quality of a clustering, over the space of all clustering possibilities (Fortunato and Hric, 2016). The most popular quality function is the *modularity* proposed by Newman and Girvan (2004), which estimates the quality of a partition of the network in communities. The essential idea of this measure is to reveal how non-random the network structure is by comparing the actual structure and its randomization where network communities are destroyed. The value of modularity varies between −1 and 1, which measures the density of links inside communities as opposed to links between communities. Its general expression is formulated as follows:

$$Q = \frac{1}{2m}\sum_{ij}(a_{ij} - p_{ij})\delta(C_i, C_j) \quad (3)$$

In this equation *m* represents the number of edges of the network. The sum runs over all pairs of nodes i and j, in which a_{ij} and p_{ij} denote the element of the adjacency matrix and the null model term, respectively. Derived by randomizing the original graph, the term p_{ij} indicates the average adjacency matrix of an ensemble of networks to preserve some of its features. C_i indicates the community to which node i is assigned. The Kronecker delta function is as follows:

$$\delta(C_i, C_j) = \begin{cases} 1, & if\ C_i = C_j \\ 0, & otherwise \end{cases} \quad (4)$$

The modularity measures thus how different the original graph is from such randomizations. Since weighted networks (links are weighted by transfer flows/waiting time) are used in our application, the modularity is reformulated as follows (Newman, 2004):

$$Q = \frac{1}{2m}\sum_{ij}(a_{ij} - \frac{k_i k_j}{2m})\delta(C_i, C_j) \quad (5)$$

where $k_i = \sum_j a_{ij}$ denotes the sum of the weights of the edges attached to node i.

The Louvain method is adopted because it has been recognized as one of the best-performing clustering algorithms after a comparative evaluation (Lancichinetti and Fortunato, 2009). The Louvain method has several advantages. First, the algorithm is quite intuitive and easy to implement. Second, the outcome is unsupervised and computationally light, which requires the link label matrix as the only input. The essence of this method is a greedy optimisation of Q in a hierarchical manner. It assigns each node to the cliques of their neighbours that can yield the largest Q, and thus creates a smaller weighted super-network whose nodes are the clusters already found. Therefore partitions found on this super-network consist of clusters that contain previous ones as well, resulting a higher hierarchical level of clustering. This procedure is not stopped until the largest possible modularity value is reached.

3 Case study

We applied our methodology to the urban public transport network of The Hague, the Netherlands, operated by HTM. The network consists of 12 tram lines and 10 urban bus lines (Figure 3). For the tactical planning study purpose, the C-space network representation thus consists of 44 nodes. For the real-time coordination study purpose we consider the main interchange hub, the central train station which is served by 9 tram lines and 6 bus services, resulting in a graph of 30 nodes.

We used all AFC transactions on the network for the 20 working days between November 2 and November 29, 2015, as demand input. After removing days where one or more disruptions occurred anywhere on the network, AFC transactions of 10 working days were finally used in the analysis. Table 2 provides an overview of the 8 experiments for which modularity is applied to identify community structures for different analysis units, time periods and using different link labels.

Table 2 Overview of experiments for which modularity is applied

	Tactical planning: entire urban network		**Real-time coordination: Central Station hub**	
Link label	*AM*	*PM*	*AM*	*PM*
Transfer flow	Experiment 1	Experiment 2	Experiment 5	Experiment 6
Transfer waiting time	Experiment 3	Experiment 4	Experiment 7	Experiment 8

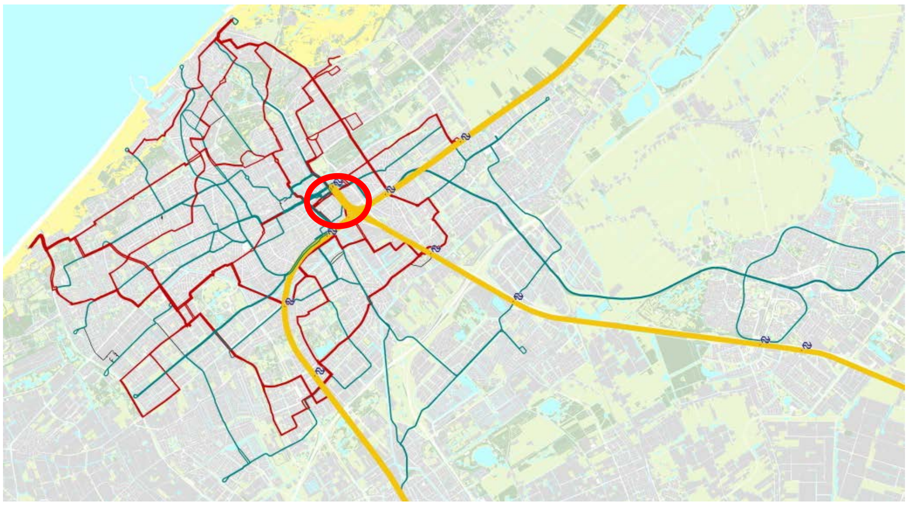


Fig. 3 Overview of urban tram (green) and urban bus services (red) in The Hague. The train services (yellow) are not incorporated in this study. The main train station hub is marked red.

4 Results

4.1 Results summary

Table 3 summarizes the main results of community detection for all 8 experiments: the number of communities, modularity], and the percentage of intra-community flows compared to total flows: the latter two are partitioning evaluation measures, where a higher value indicates a stronger partitioning.

Table 3 Summary of performance of community detection algorithm for all 8 experiments

Experiment	**Number of communities**	**Modularity**	**% intra-flow / total flow**
Experiment 1 *network; AM; transfer flow*	5	0.225	47%
Experiment 2 *network; PM; transfer flow*	4	0.258	51%
Experiment 3 *network; AM; transfer wait time*	4	0.223	51%
Experiment 4 *network; PM; transfer wait time*	4	0.258	51%
Experiment 5 *hub; AM; transfer flow*	2	0.276	85%
Experiment 6 *hub; PM; transfer flow*	2	0.290	83%
Experiment 7 *hub; AM; transfer wait time*	3	0.285	67%
Experiment 8 *hub; PM; transfer wait time*	2	0.288	83%

For each scenario the heuristic for modularity optimization is performed 1,000 times, after which the results of the run with the highest modularity are used. For the network level, results are quite robust over experiments 2-4: four communities are identified, resulting in 51% of the transfer flow / transfer waiting time to be intra-communal. In experiment 1 five cliques are detected, with a slightly lower intra-communal percentage of 47%. For the Central Station hub, experiments 5, 7 and 8 identify two communities, whereas experiment 6 identifies three communities. In general, the percentage intra-communal flow compared to the total flow decreases with an increasing number of detected communities.

4.2 Tactical planning: results for the entire network

For all four experiments it can be observed that public transport lines heading into the same direction are clustered together (Figure 4). In experiments 2-4 in each of the obtained communities one direction - eastbound, westbound, northbound or southbound – dominates. In experiment 1 the five communities are dominated by lines bounded north-west, south-east, south, west and north. These results are intuitive, indicating dominant transfer flows between lines in the same overall direction. In experiment 1 the fifth community shows a strong transfer flow from tram line 9 (eastbound from a residential area) to bus lines 69 and 79 (northbound towards an employment area). This is a very concentrated passenger stream particularly for the AM peak resulting in a separate community, while these transfer flows in the PM are not sufficient to be identified as separate community.

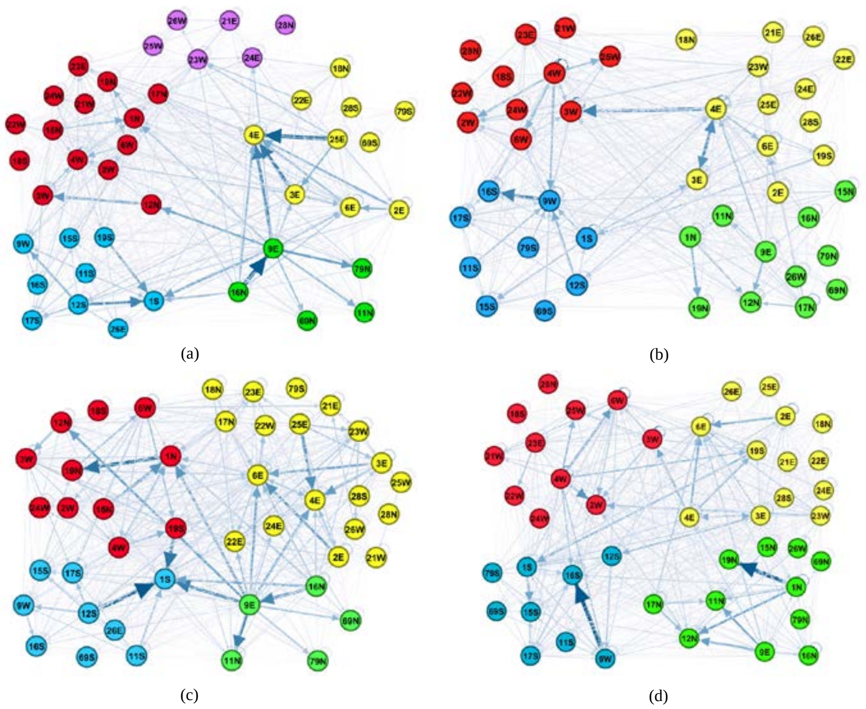


Fig. 4 Community structure for entire network based on AM transfer flows (a), PM transfer flows (b), AM transfer waiting time (c) and PM transfer waiting time (d). Each node represents a service line-direction; each link with arrow shows the directional transfer flows.

When modularity is based on transfer flows (experiments 1-2), one can see that no community consists of the same line in the opposite direction. This could be expected, since the largest transfer flows occur between high frequent, radial tram lines, and thus it is plausible that a same line in the opposite direction is not clustered in the same group. However, when the clustering is based on transfer waiting time incorporating service frequencies (experiments 3-4), it can be observed that in the morning peak bus lines 21, 22 and 23 in both directions are part of the same community. This can be explained, since these bus lines are relatively less frequent, tangential or circular services. Transfers between the radial tram lines and these tangential or circular bus lines can occur in both directions (see the green and red lines in Figure 4). Although these transfer flows are relatively small compared to transfer flows between radial tram lines, incorporating the relatively low frequency of these bus lines does result in a separate community for the AM peak, in which average passenger demand is higher for our case study network compared to the PM peak.

4.3 Real-time coordination: results for the Central Station hub

For the Central Station hub, in experiments 5, 6 and 8 two clear communities can be observed with lines heading into northern and eastern directions, and with lines heading into southern and western directions (Figure 5). In experiment 7 northbound and eastbound lines are clustered into separate communities. For all experiments 5-8, lines in opposite direction are not part of the same community. Since almost all lines serving this hub are radial lines, this result shows the partitioning is plausible. Results are quite robust over all four experiments. Since almost all radial lines serving this hub operate with relatively high frequencies, differences between applying transfer flow (experiments 5-6) and transfer waiting time (experiments 7-8) as link label are limited.

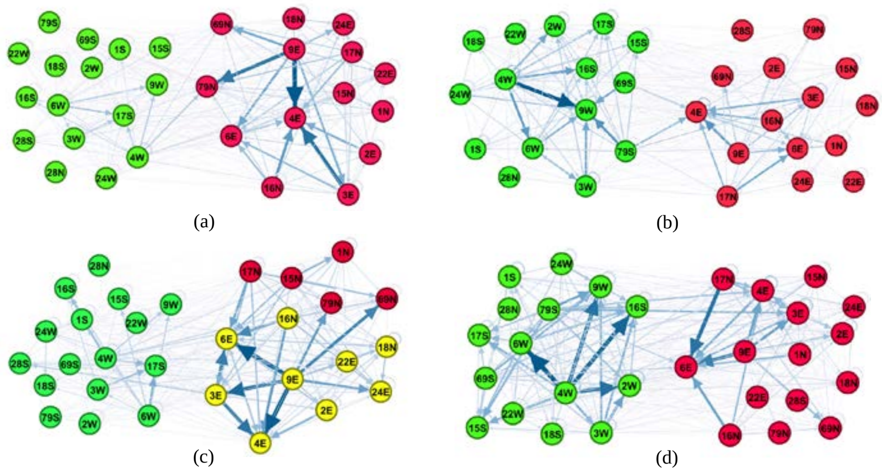


Fig. 5 Community structure for main station hub based on AM transfer flows (a), PM transfer flows (b), AM transfer waiting time (c) and PM transfer waiting time (d). Each node represents a service line-direction; each link with arrow shows the directional transfer flows.

5 Conclusions

We develop a data-driven, generic and passenger-oriented methodology to determine line bundles to be prioritized in devising transfer synchronization measures. The proposed non-supervised learning technique enables identification of line bundles based on passenger transfer flows, independent of local knowledge. The application of a modularity-based community detection technique to a public transport hub represented in C-space shows intuitive lines grouped together to prioritize during transfer synchronization. Our results illustrate the necessity of synchronizing different line bundles during different periods of the day, depending on the travel patterns prevailing during the relevant time period. Our methodology and study results support public transport operators in timetable design and when performing real-time control decisions, such as holding, to determine where and which lines to synchronize. Moreover, public transport agencies can select hubs and line bundles in determining where to invest in measures for improving the design of a seamless transfer experience (e.g. amenities, physical environment, island vs. side platforms).

We recommend extending the line bundle identification in our study by applying a link-based clustering technique, rather than node-based clustering. In our modularity-based community detection technique the nodes – i.e. lines in a certain direction – are clustered. However, when the transfer links between nodes would be clustered, one would be able to distinguish between transfer flows from line l_i in direction a to l_j in direction b, and flows from l_j in direction b to l_i in direction a. Incorporating the transfer direction between two lines, next to the lines itself, enables deriving further recommendations for timetable planning and real-time coordination by specifying the desired sequence of arrivals.

Acknowledgements

This research was performed as part of the TRANS-FORM (Smart transfers through unravelling urban form and travel flow dynamics) project funded by NWO grant agreement 438.15.404/298 as part of JPI Urban Europe ERA-NET CoFound Smart Cities and Communities initiative. The second author acknowledges the support of the SETA project funded by the European Union's Horizon 2020 research and innovation program. The authors thank HTM, the urban public transport operator of The Hague, the Netherlands, for their valuable cooperation and data provision.

References

Blondel, V.D., Guillaume, J., Lambiotte, R., Lefebre, E. (2008). Fast unfolding of communities in large networks. Journal of Statistical Mechanics: Theory and Experiment, 2008, 1-12.

Desaulniers, G., and M.D. Hickman. (2007). Public Transit. In C. Barnhart and G. Laporte (Eds.), Handbook in OR & MS (pp. 69-127). Amsterdam, the Netherlands: Elsevier.

Fortunato, S., and D. Hric. (2016). Community detection in networks: A user guide. Physics Reports 659:1–44.

Gavriilidou, A., Cats, O., Leffler, D., Corman, F., Hoogendoorn, S.P. (2017). Real-time transfer synchronization of public transport services using passenger data. Proceedings of the 96th Annual Meeting of the Transportation Research Board 2017, Washington D.C.

Gordon, J.B., Koutsopoulos, H.N., Wilson, N.H.M., Attanucci, J.P. (2013). Automated inference of linked transit journeys in London using fare-transaction and vehicle location data. Transportation Research Record: Journal of the Transportation Research Board, 2343, 17-24.

Goverde, R.M.B. (1998). Synchronization control of scheduled train services to minimize passenger waiting times. Proceedings of the 4th TRAIL year congress, Part 2, TRAIL Research School, Delft, the Netherlands.

Lancichinetti, A., and S. Fortunato. (2009). Community detection algorithms: A comparative analysis. Physical Review E - Statistical, Nonlinear, and Soft Matter Physics 80:56117.

Lee, A., Van Oort, N., Van Nes, R. (2014). Service reliability in a network context: impacts of synchronizing schedules in long headway services. Transportation Research Record: Journal of the Transportation Research Board, 2417, 18-26.

Nesheli, M.M., Ceder, A. (2014). Optimal combinations of selected tactics for public transport transfer synchronization. Transportation Research Part C, 48, 491-504.

Newman, M. E. J. (2004). Analysis of weighted networks. Physical Review E - Statistical Physics, Plasmas, Fluids, and Related Interdisciplinary Topics 70:9

Newman, M. E. J., and M. Girvan. (2004). Finding and evaluating community structure in networks. Physical Review E - Statistical, Nonlinear, and Soft Matter Physics 69:26113.

Schakenbos, R., La Paix, L., Nijënstein, S., Geurs, K. (2016). Valuation of a transfer in a multimodal public transport trip. Transport Policy, 46, 72-81.

Van Hagen, M., Sauren, J. (2013). Influencing the train experience: using a successful measurement. Transportation Research Procedia, 41st European Transport Conference 2013, Frankfurt, Germany.

Van Oort, N., Brands, T., De Romph, E., Yap, M.D. (2016). Ridership evaluation and prediction in public transport by processing smart card data: a Dutch approach and example. Chapter 11, public transport planning with smart card data, eds. Kurauchi, F., Schmöcker, J.D., CRC Press.

Yap, M.D., Cats, O., Van Oort, N., Hoogendoorn, S.P. (2017). A robust transfer inference algorithm for public transport journeys during disruptions. Transportation Research Procedia: 20th EURO Working Group on Transportation Meeting, EWGT 2017: Budapest, Hungary.