Additional Graduation Work, Research Project

Identification of the Hierarchy in Public Transport Networks based on Passenger Flow Patterns

By

Ziyulong Wang 4890973

Supervised by

Dr. O. (Oded) Cats

Dr. T. (Trivik) Verma

Ding Luo

TU Delft

Contents

1 Introduction 3
1.1 Background 3
1.2 Challenge 1: The representation and separation of multi-layer network 3
1.3 Challenge 2: The ranking of multi-layer network 4
1.4 Aim of study and research questions 4
1.5 Structure of the report 4
2 Literature Review 5
3 Methodology 6
3.1 Derivation of passenger transfer flow matrix 6
3.2 Separation of network layers 8
3.3 Identification of the hierarchy of PTNs 9
4 Case study 10
5 Results & Analysis 12
5.1 Transfer flow matrices 12
5.2 Weighted C-space with Louvain community detection method 16
5.3 Hierarchy of PTNs 20
6 Conclusion 22
Bibliography 24

1 Introduction

In this chapter, the background of the hierarchy of public transport networks (PTNs) is introduced. Following the aim of this study and the research questions. The chapter wraps up with the brief introduction of the structure of the paper.

1.1 Background

Public transport (PT) is a vital part of urban mobility, in terms of reducing congestion, air pollution and the consumption of energy. In the field of PT, one of the most important feature is the public transport networks (PTNs) as it has profound influence on supporting planning and management of PT industries. Complex network theory, is a science that studies the connection and interaction between components in a system (Lin and Ban, 2013). This emerging field of study has provided powerful tool to study the characteristics of the topology of PTNs (Newman, 2010). Since the travel demand is unevenly distributed on a PTN, this results in a hierarchical structure, which means an unequal importance. The hierarchical network is figured as a natural phenomenon resulting from the interaction between demand and supply and is related to the technological developments, modern decision process, etc. (van Nes, 2002).

With the advancement of computation tools, a large number of studies has been conducted to study the multi-layer transport network. The urban transportation systems have been partitioned based on modes and investigated by (Aleta et al., 2017) with the aim to reproduce the real-world fact of network improvements. The same separation method to construct a multi-layer PTN can be also seen in other researches (van Nes, 2002; Zhang et al., 2014; Gallotti and Barthelemy, 2015). By this way, each layer is only consisted of the lines of the mode, aggregated as a sort of super layer (Aleta et al., 2017). Except considering the multimodality of PTNs, other separation methods have also been applied. Feng et al. (2017) offered a method that divided the network into three parts, namely subway infrastructure layer, train flow layer and passenger flow layer. Moreover, to identify the community structure of an urban city, the multi-layer network can be represented by the movement of cars, buses and passengers (Yildirimoglu et al., 2017). Therefore, the construction of layered PTNs is based on the study aim.

There are a large number of studies that go depth into the multi-layer network construction, however, there is still a lack of knowledge towards a uniform and clear ranking method for multi-layer PTNs in a data-driven method. With the consideration of the hierarchical structure of PTNs is critical for optimizing the PTNs, a methodology with high potentials shall be developed. Two main challenges are needed to be addressed:

1.2 Challenge 1: The representation and separation of multi-layer network

Transportation network is often represented as the nodes and edges embedded in space. Understanding and characterizing the structure of spatial networks is crucial (Barthélemy, 2010). The fundamental representation of network can have substantial influence on the later network analysis as the topologic representation construction is the very basic stage of it. Different definition for the node and edge can subsequently yield different representation of networks. A boarding/alighting station can be defined as a node, and the connection between two stations can therefore be regarded as a link. However, a route can also be seen as a node whilst the link is now becoming when the two routes have transferability (von Ferber, 2009). The weighted and directed graph study has been conducted in the work done by Háznagy et al. (2015). In this study, the weight of the links represents the capacity of vehicles, which facilitates the identification of sensitive routes and stations. Therefore, the graph representation is based on the study aim of research and weighted graph is a way that can incorporate more information to facilitate the study.

The separation of network can be troublesome, namely the construction of multi-layer network. Previous study showed a number of qualitative method. Aleta *et al.* (2017) constructed the multi-layer network based on transportation mode. Gallotti and Barthelemy (2015) also divided the whole network into mode-oriented so that the connection among modes can be investigated, including airports, bus/metro/coach/railway stations.

The same methodology can be also seen in many other studies as it is rather intuitive and simple but straightforward. Other qualitative methods like, infrastructure layer plus traffic and passenger flow layer, regional separation have the similar pros and cons. Moreover, quantitative way to derive the multi-layer network has been used in several studies in a data-driven method. Verma *et al.* (2016) developed a pruning process based on removal of underutilized links and redistribution of loads, leading to the appearance of core-peripheries structure of airline network. In the most of the networks, a community structure is existing, namely their vertices are organized into groups (Fortunato & Hric, 2016). Yap *et al.* (2019) identified the urban PT hubs and their key synchronization priorities under the consideration of not only tram but also bus. In this study, the network construction and line bundle identification follow a bottom-up strategy, regarding only the revealed preference from data. The unsupervised learning technique is inspiring to cluster the lines and hubs. Examples of studies applying this technique can be also seen in Cats *et al.* (2015) to identify and classify the PT activity centers and Luo *et al.* (2017) to construct transit origin-destination matrices. As a result, community detection becomes an increasingly powerful tool to study the multi-layer network and thereof drives the direction of this study.

1.3 Challenge 2: The ranking of multi-layer network

The notion of the hierarchy of a multi-layer transport network has been widely discussed. Van Nes (2002) is almost the first one who proposed this notion. This study considered a two-level PTN, consisting of an urban network and an interurban network with the aim to maximize the welfare of two operators. Van Nes (2007) connected the traveler behavior to the network characteristics, which results in a multi-class urban transit network design. Stephen (2006) proposed a transit-oriented hierarchy which is relatively qualitative, considering only the mode and routes. Bagloee & Ceder (2011) divided the transit network into three degrees: mass route, feeder route and local route and the method to determine the hierarchy of a route is developed. This mass transit plus feeder network definition could also been seen in many other studies with different modes like railway or bus (Chien *et al.*, 2000; Verma & Dhingra, 2006; Mohaymany & Gholami, 2010). In addition to the three type abovementioned, the hierarchy identification can be based on residential transit trip distance (Jian *et al.*, 2012). However, those ranking methods are not very satisfying as they are either based on conventional thinking or with less consideration. This means that there is still a knowledge gap towards a systematical approach to identify the hierarchy of PTNs. Given weighted PTNs, no uniform or clear strategy to set an importance order of them with data supporting.

1.4 Aim of study and research questions

In this study, the separation of PTNs is determined by the value of passenger transfer flow and the ranking is based on the value of intra-flow and inter-layer flow. Overall, the aim of the study is twofold: 1) Choosing a community detection method to define the layers of PTNs; 2) Proposing a ranking method to order the different layer.

Particularly, due to the fact that the objective of this study is to identify the hierarchy of the PTNs based on the transfer flow patterns, the main research question is set as:

How to identify the hierarchy of public transport networks based on passenger transfer flow pattern?

Sub-questions are:

1. *What should be the raking method for the multi-layer public transport networks and how to testify it?*
2. *How to construct the multi-layer public transport networks based on passenger transfer flow pattern?*
3. *How to derive the passenger transfer flow pattern, given the passenger journey data and ride data?*

1.5 Structure of the report

The structure of this report is as following: section 2 introduces the literature review that is related to the hierarchical identification of PTNs. In section 3, methodology of the paper has been presented, including a general definition of the PTNs based on graph theory, the derived transfer flow with temporal attributes, the topological representation of PTNs, the construction of multi-layer PTNs and the ranking approach to set the order of multi-layer PTNs. In section 4, the case study of The Hague has been displayed, using the data from PT enterprise HTM in The Hague to illustrate the results of the methodology. Section 5 discusses and analyzes the result. The whole report warps up by the conclusion, namely section 6.

2 Literature Review

In order to rank the PTNs, one of the early works (van Nes, 2002) has defined two-level public transport system, consisting of two interconnected subsystems: an urban network and an interurban network. In this separation, the urban network is dedicated for local trips to the city center and offer access to interurban transport network while interurban transport network is serving the passengers between cities. In another work by van Nes (2002), he proposed a way of analysis by dividing the PTNs in spatial scale and four important factors that influence hierarchical network patterns, including demand pattern, importance of travel time, costs of building and operating transfer facilities and scale benefits due to concentration of flows. As PTN is a multimodal network per se, consisting of different modes, the faster the mode means the shorter travel time, indicating a higher-level network. Stephen (2006) proposed a transit-oriented hierarchy which is relatively qualitative, considering only the mode and routes. Bagloee & Ceder (2011) divided the transit network into three degrees: mass route, feeder route and local route and the method to determine the hierarchy of a route is developed. This mass transit plus feeder network definition could also been seen in many other studies with different modes like railway or bus (Chien *et al.*, 2000; Verma & Dhingra, 2006; Mohaymany & Gholami, 2010). In addition to the three type abovementioned, the hierarchy identification can be based on residential transit trip distance as well (Jian *et al.*, 2012). Therefore, the network can be separated based on mode and the ranking is based on the trip distance and the speed of the mode.

The same line of reasoning can be seen in many works later. Gallotti and Barthelemy (2015) with the aim to focus on the features of multi-modality, they used a coarse graining procedure and define explicitly the coupling between different transport modes, containing the connections at airports, ferry docks, rail, metro, coach and bus stations. This is equally meaning a separation of PTNs based on mode and the resulting multilayer network has the attributes of weight, direction and temporal, which is combined with dataset. In addition, the ranking in this paper is from National Public Transport Access Nodes scheme, in which air > ferry > rail > metro > coach > bus. Further, *superlayer* network has been introduced by Aleta *et al.* (2017). In this *superlayer* network, each line of each mode is regarded as a single layer to gain an insight into network properties in different cities. However, this study preserves the ranking and is not able to quantify the importance. The separation and ranking methods above-mentioned are quite intuitive but rather qualitative, which means it is lack of data supported and needs to be testified.

Quantitative way to derive the multi-layer network has been used in several studies in a data-driven method. Verma *et al.* (2016) developed a pruning process based on removal of underutilized links and redistribution of loads, leading to the appearance of core-peripheries structure of airline network. In the most of the networks, a community structure is existing, namely their vertices are organized into groups (Fortunato & Hric, 2016). Yap *et al.* (2019) identified the urban PT hubs and their key synchronization priorities under the consideration of not only tram but also bus. In this study, the network construction and line bundle identification follow a bottom-up strategy, regarding only the revealed preference from data. The unsupervised learning technique is inspiring to cluster the lines and hubs. Examples of studies applying this technique can be also seen in Cats *et al*. (2015) to identify and classify the PT activity centers and Luo *et al.* (2017) to construct transit origin-destination matrices. As a result, community detection becomes an increasingly powerful tool to study the multi-layer network and thereof drives the direction of this study.

Moreover, instead of considering the network as a whole, namely considering all the nodes and links into account, the main characteristics of a PTN can be derived and preserved based on the aim of study. This will bring the efficiency but still capture the desired features. Various topologies of PTNs have been studied, including L-space, P-space, C-space and B-space. C-space represents each route as a node, a link between nodes means that they share common stops (von Ferber, 2009). An example of a simple PTN is shown in the Fig.1 with a C-space representation of it. In this study, only C-space is focused. For other topological representations, it can be found in the study conducted by von Ferber (2009). This facilitates the study into the topology of network and meanwhile with the help of C-space, the network has been simplified, resulting in the characteristics with needed. Different representation will certainly result in different details of the conclusion. An example of this way of conducting a study in PTNs can be seen in the paper by Sui *et al.* (2019) where they constructed a layered network model with two layers. The bottom layer is a bus infrastructure network while the top layer is a traffic flow network to investigate into the interaction between PTN and traffic dynamics. Thus, in this study, the main focus is the characteristics of passenger transfer flow, for the goodness of simplification and effectiveness, the C-space representation has stood out.

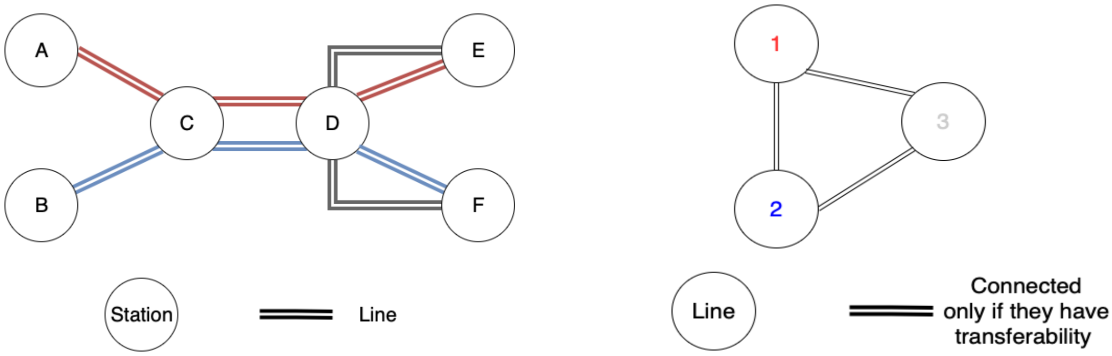

Fig.1 A simple example of PTN with the C-space representation of it

3 Methodology

In this chapter, the illustration of methodology is conducted, mainly in three parts: the derivation of passenger transfer flow matrix, the weighted C-space network construction and the identification of PTN hierarchy.

To demonstrate the methodology of this report, a flow diagram has been built shown in Fig.2 below. The first stage of the preparation is highly related to the data processing. The raw data is Automatic vehicle location (AVL) data and automatic fare collection (AFC) data. But in this report, the data used has already been processed. The passenger journey data includes the information of the passenger ride record while the passenger ride data has the knowledge of the route traversed, which facilitates the further derivation of transfer flow matrices and network construction. After the preparation, the topological representation can be construct. In this step, the temporal attribute of network has been considered and one kind of network representation has been applied. This network representation will be discussed in details in the latter part. Next, in order to build the separate network layer, community detection has been employed. Furthermore, the hierarchy of PTN has been identified based on the intra-transfer and inter-transfer flow.

3.1 Derivation of passenger transfer flow matrix

The first step of this study is to derive the passenger transfer flow matrix, as visualized in the first phase. As input for the study, the raw data is PT vehicle position data obtained from Automated Vehicle Location (AVL) systems and passenger data from Automated Fare Collection data for urban tram and bus services. However, the input data has been already processed into passenger journey and passenger ride data, which facilitates the study. Each passenger journey data contains the information of journey ID, operating date, day of week, origin stop ID, destination stop ID, origin stop check in time, destination check out time, number of rides and ride IDs. While each passenger ride data contains the data of ride ID, operating date, line ID, trip ID, direction ID, check in and check out Stop ID, check in and check out time. Examples of passenger journey data and ride data are shown in Tab.1 and 2.

The aim of processing data is to find out the passenger transfer flow, and further to construct the C-space network representation. In a C-space representation, the spatial attributes of a line have been simplified. Thus,

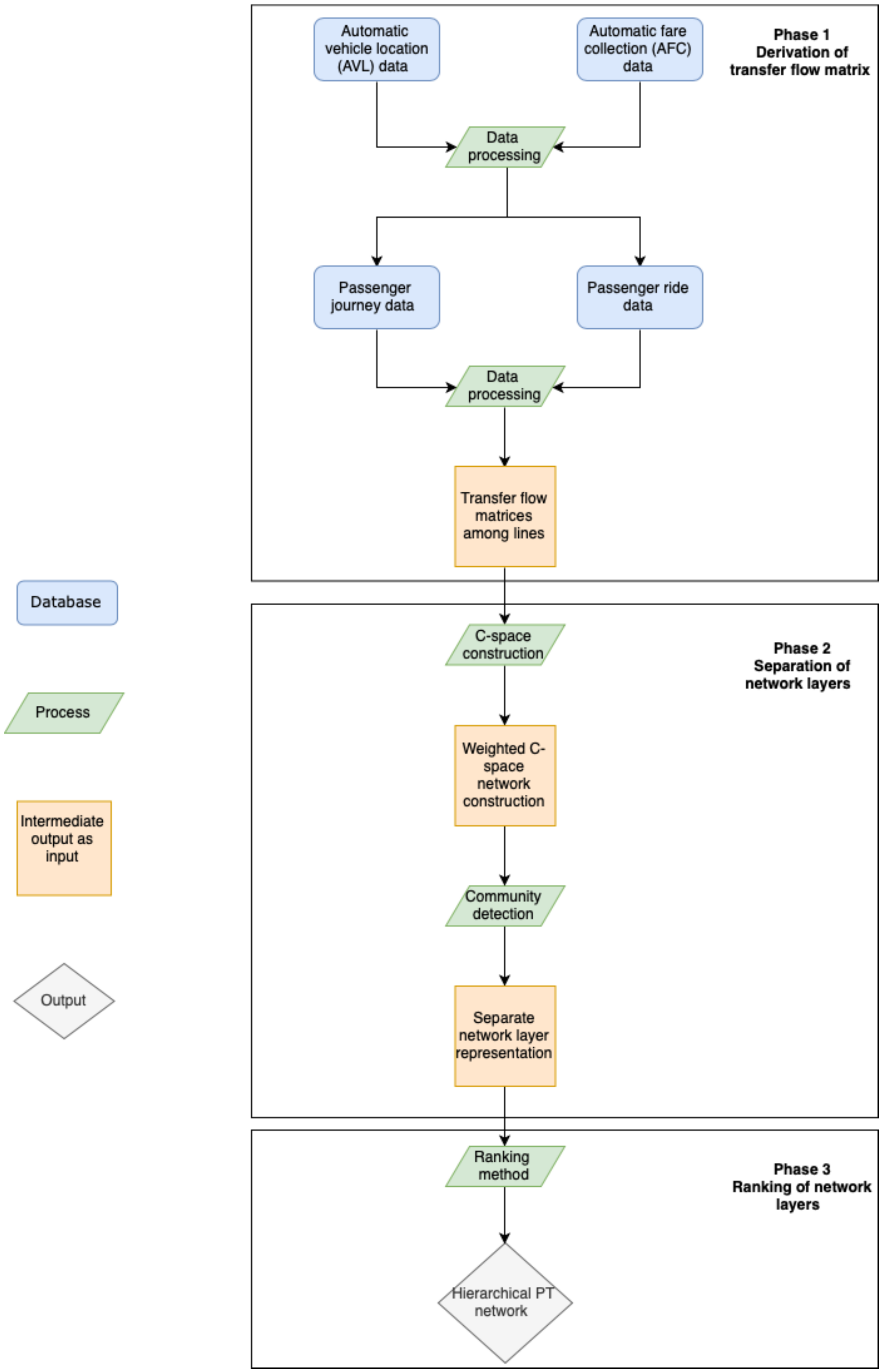

Fig.2 The flow diagram

Tab.1 Example of passenger journey data

Journey ID	Date	Day of week	Origin stop ID	Destination stop ID	Origin check in time	Destination check out time	Number of rides	Ride ID
129	20150301	0	2917	5413	46767	49076	3	2957721; 2969; 5378523

Tab.2 Example of passenger ride data

Ride ID	Date	Line ID	Trip ID	Direction ID	Check in stop ID	Check out stop ID	Check in time	Check out time
2969	20150301	1	2273	0	2832	2730	47667	47876

the only thing needs to do is to fuse the passenger journey and ride data based on the ride ID in journey data and then refer to the ride in passenger ride data. In order to derive the transfer flow matrix, the number of ride ID must be more than 2. This means that at least one transfer existing.

By this way, the line ID, direction ID and check in and check out time attributes can be added. In a C-space network representation, the lines are regarded as nodes. So, the line is vital to the further parts where C-space construction needs. Moreover, generally speaking, a line of PT is always bi-directional. This means that a transfer of passenger from line to line needs the direction information to specify. Otherwise, the direction and temporal characteristics could be lost. Because one direction of a line could be busy in the morning peak while the other direction of the same line can be busy in the evening peak. Also, the temporal attribute is taken into account when selecting a certain time period to investigate, the origin check in time is used to match the start and the end of the selected time period.

3.2 Separation of network layers

A community detection method has been applied to fulfill the aim of separating the network layers. The separation of network layers is actually consisting of two steps: the first one is to construct the weighted C-space network representation. This is to depict the characteristics of network properties. And the second step is to identify the clusters of the lines which are the communities and then the network layers are separated based on the communities. This data-driven method is automated, intuitive and scalable. The details of these two steps are illustrated in the sections following:

3.2.1 Weighted C-space network construction

For a PT system, PTNs is consisted of two important layers: the first one is the physical network layer, in terms of infrastructure, including stations and roads, etc. The second one is the service layer on the top of the infrastructure layer, namely the routes. A graph can be featured as G= (V, E, R), with V, E $\subseteq$ V × V, R representing the set of nodes, links, routes, respectively, building the connections among the elements. In the C-space network graph G, each node v $\in$ V represents a line with direction specified while each link e $\in$ E is constructed only if two lines (u, v) have transferability. In here, u and v, separately, denote the source and target nodes.

3.2.2 Community detection—Louvain community detection

Louvain community detection method is a heuristic method that is based on modularity optimization (Blodel *et al.*, 2008). This method has The method performs a greedy optimization of Q in a hierarchical manner, by assigning each vertex to the community of their neighbors yielding the largest Q, and creating a smaller weighted super-network whose vertices are the clusters found previously. Partitions found on this super-network hence consist of clusters including the ones found earlier, and represent a higher hierarchical level of clustering. The procedure is repeated until one reaches the level with largest modularity (Fortunato & Hric, 2016). This method has many advantages: first of all, the method is easy to apply and understandable. Second of all, the outcome is unsupervised and computationally efficient, which requires the edge dataset as the only input. The method itself is a greedy optimization algorithm of the modularity q. The procedure of the method is first assigning the nodes into communities with the discovery of the largest q. Thus, it creates the cluster of nodes. Further, the network is partitioned into clusters and becomes an entire network with communities. The details of the modularity are discussed below:

Modularity proposed by Newman and Girvan (2004) estimates the quality of a partition of the network in communities. The basic idea of this measurement is to discover how non-random the network structure is by comparing the actual structure and its randomization where network communities are destroyed (Yap *et al.*, 2018). The value of modularity q is varied between -1 and 1, which measures the density of links within communities in opposite of that of among communities. The general formulation is shown in Eq. (1) below:

$$q = \frac{1}{2|E|}\sum_{ij}(a_{ij} - p_{ij})\delta_{c_i,c_j} \tag{1}$$

In Eq. (1), $|E|$ stands for the total number of edges in the graph G. a_{ij} represents the element of adjacency matrix while p_{ij} indicates the randomized null model term. The summation of $a_{ij} - p_{ij}$ of all node pairs ij measures the quantity of non-random network structure. c_i and c_j are the communities to which nodes i, j have been assigned to. δ_{c_i,c_j} is the Kronecker delta function and is formulated in Eq. (2):

$$\delta_{c_i,c_j} = \begin{cases} 1, if\ c_i = c_j \\ 0, otherwise \end{cases} \tag{2}$$

The normal modularity measurement of the Louvain method only refers to the adjacency matrix to apply the community detection. This means that it only looks into the connection status of two nodes in the graph. But in the C-space graph of this study, the edge weight, namely the passenger transfer volume also needs to be considered. The weight of edge attribute can represent for the passenger transfer flow derived from passenger travel information. The Eq. (1) has thus been modified to Eq. (3) with weighted edges. w_i, w_j denote the weight of edges attached to nodes i, j (Newman, 2004).

$$q = \frac{1}{2|E|}\sum_{ij}\left(a_{ij} - \frac{w_i w_j}{2|E|}\right)\delta_{c_i,c_j} \tag{3}$$

3.3 Identification of the hierarchy of PTNs

The ranking of the layer of PTN is based on the community detection performed in the last step. Since the community detection has gathered the nodes into communities which in C-space are the line bundles that have been clustered into a community. Each community has a line bundle and thus naturally forms a layer in a PTN no matter the mode type. The main topic of this study is based on the volume of passenger transfer flow and thus the identification is also totally counted on that. The transfer flow can be defined as two types. One of them is inner-transfer flow, which is the transfer flow within a layer. f_{ij} represents the transfer flow between line i and j in the same community while f_{c_l} stands for the total inner-transfer volume. This in turn shows the frequency of travel and usage of the layer. The inner-transfer flow is formulated as Eq. (4) shown:

$$f_{c_l} = \sum_{i,j \in c_l} f_{ij} \tag{4}$$

The other type is intra-transfer flow that is the transfer flow among layers. f_{ik} represents the transfer flow between line i and k in the different community l and m while $f_{c_{lm}}$ stands for the total intra-transfer volume between community l and m. This equally means the communication among different layers. The intra-transfer flow is formulated as Eq. (5) shown:

$$f_{c_{lm}} = \sum_{i \in c_l, k \in c_m} f_{ik} \tag{5}$$

The absolute value of inner-transfer can partially show the importance of a layer because the communication between layers is also important. If the inner-transfer flow is high but also the intra, this can only show the importance of being a transfer cluster but not as a critical line bundle. Therefore, the higher level of a PTN layer should have more frequent interchange within the layer, compared to the flow between itself to other layers. This means a stronger ability of self-sufficiency, namely the inner-transfer flow within a layer divided by the intra-transfer flow among layers should be higher. Eq. (6) formulates the ratio ϑ of inner- and intra-transfer flow:

$$\vartheta = \frac{\sum_{i,j \in c_l} f_{ij}}{\sum_{i \in c_l, k \in c_m} f_{ik}} \tag{6}$$

4 Case study

The case study is taken place on the urban PT network of the city of The Hague, the Netherlands, operated by HTM (Fig.3). The Hague is regarded as one of the important cities in Randstad, which is the most developed economic region in the west of the Netherlands and has more than 500,000 residents. The urban PT network in the dataset is consisted of 12 tram lines and 8 urban bus lines (March 2015). In this case study, on the urban PT network is focused, meaning that the services on the scale of regional level is not included, in terms of train, tram and bus.

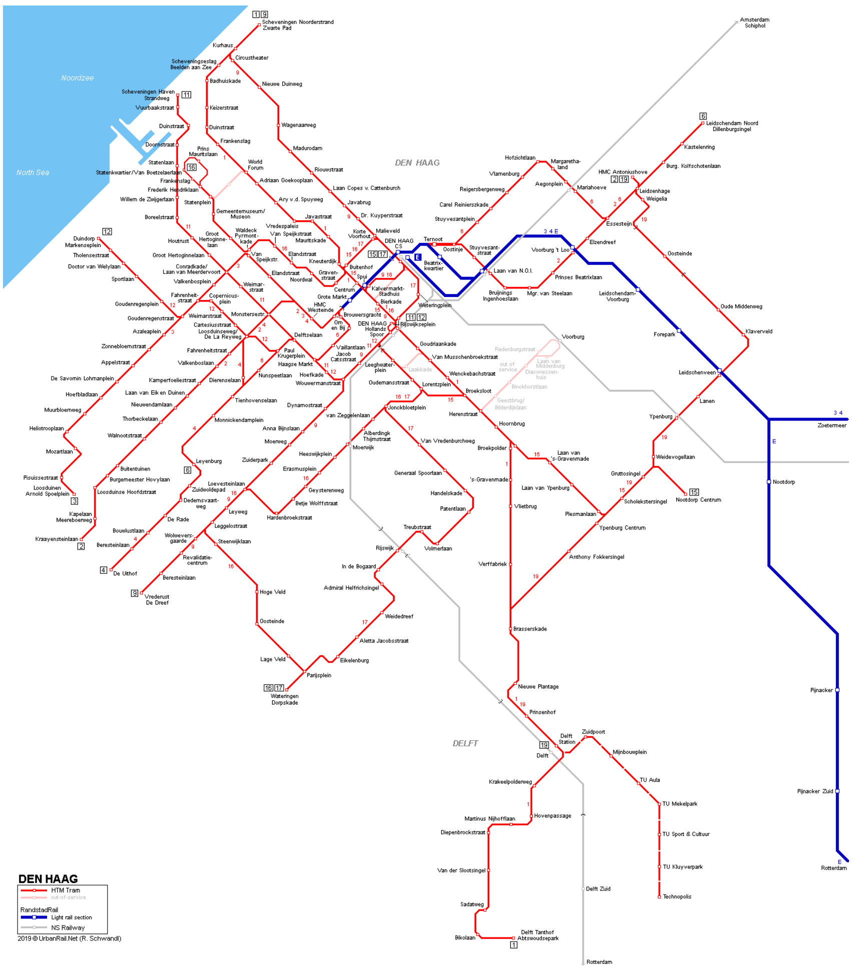

Fig.3 Overview of HTM networks in The Hague

As initial input, the case study uses the data from processed passenger journey and passenger ride data from March 1st to March 31st, 2015. In order to have an overall understanding to the March, 2015 ridership data of the case study, a line chart of the number of rides is presented in the Fig.4. In this figure, the ridership has been averaged in half an hour. Every time represents half an hour onward from the timestamp. The ridership starts to have a number from 4:00 AM onward and thereof the plot starts at the same time. From the figure we can clearly see a difference lies in the weekdays and weekends. During the weekdays, there are two peaks, started from approximately 6:30 to 9:30 and the other from around 15:30 to 18:30, respectively. This shows a morning and an evening travelling pattern. Whereas, the peak hours on the weekends are totally different from the weekdays, the average rides start to increase during the noon and the total number of rides is lower than every weekday along every time period of the day. And the ridership on Tuesdays has the highest amount. Intuitively this is relatively high and quite weird to see this phenomenon. However, when going depth into the data, there is no evidence of a wrong recording in the 5 Tuesdays in that month.

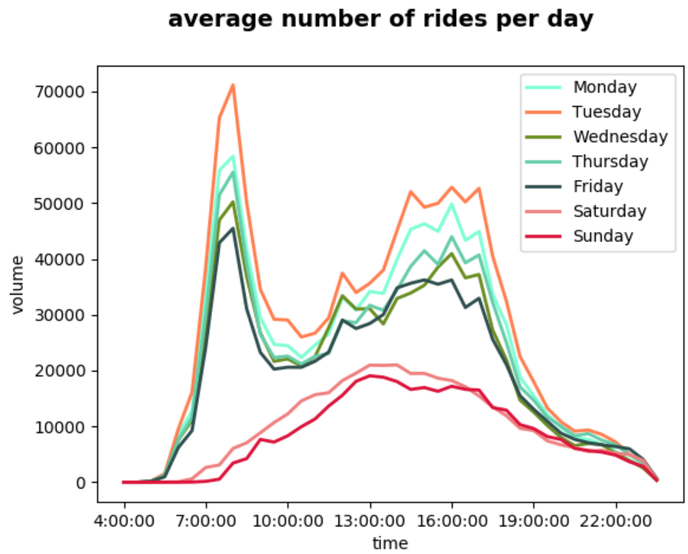

Fig.4 Average number of rides per day

Then, a same type of line chart is shown in the Fig.5 with the aim to understand the pattern of average number of transfer. With the same line of reasoning and plot logic, from Fig.5 we can see that the transfer rate is quite stable in every day of week and hence the trend of lines is more or less the same as the figure of ridership. Two peaks in the weekdays and a single peak on the weekends.

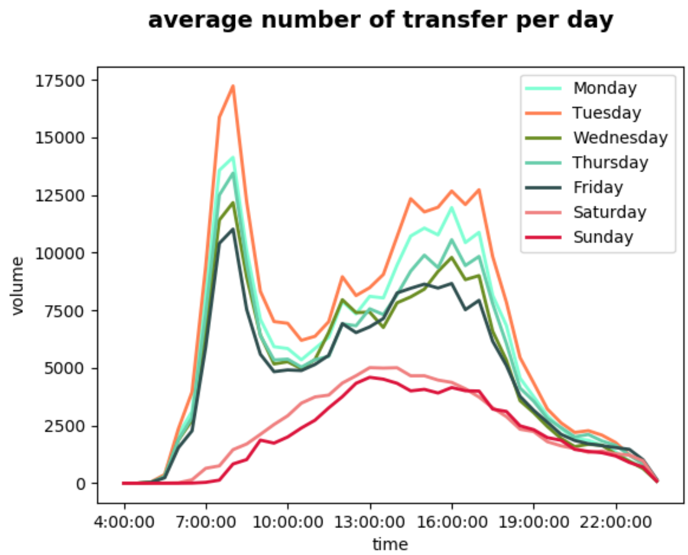

Fig.5 Average number of transfers per day

Moreover, in order to show the difference trend of weekdays and weekend, the transfer volume has been averaged through the type of day, namely weekday or weekend. Fig.6 shows the trend of two-peak weekday and the movement of one-peak weekend transfer flow pattern.

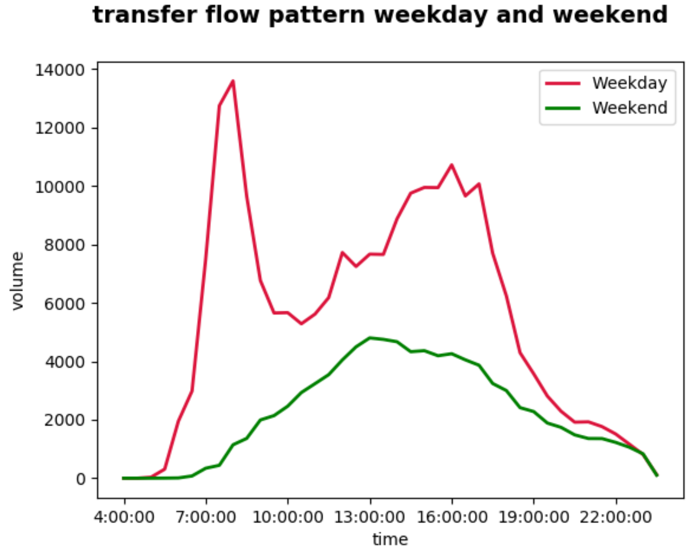

Fig.6 Comparison of the transfer flow pattern between weekdays and weekends

Therefore, for the peak hours, in this case study are defined as 6:30—9:30 as morning peak, 15:30—18:30 as evening peak while 12:30—15:30 as an off-peak period in order to make comparison. The pattern of travel is different from weekdays to weekends. Furthermore, for the latter part of the report, the selected instances are Monday and Sunday as the trend is obvious but not significantly different from the average transfer flow pattern of weekday and weekend.

5 Results & Analysis

In this chapter, the results are shown in three different sections with the sequence depicted in flow diagram, including derivation of transfer flow matrices, separation of PTNs and ranking of PTNs. Each section is using 6 sub-case, i.e. the Monday morning peak, off-peak, evening peak and Sunday morning peak, off-peak and evening peak.

5.1 Transfer flow matrices

Passenger travel is always combined a certain spatial-temporal pattern. Conventionally during the weekdays, the morning peak travel pattern is almost the adverse to that of evening peak and the travel pattern in the off-peak hour may reveal differently. Moreover, the weekends travel pattern is different from weekdays as the trip purpose has changed from commuting mainly to leisure activity-oriented. Therefore, in order to construct the C-space network and display community detection in the later chapter. This report has selected Monday morning peak, off-peak and evening peak, Sunday morning peak, off-peak and evening peak as the case study. In here, morning peak has been defined as from 6:30 to 9:30, evening peak is from 15:30 to 18:30 while off-peak is defined as from 12:30 to 15:30 as above-mentioned. All the timing has been selected as the time of the origin check-in time of data.

The transfer flow matrix heat maps of Monday morning peak, off-peak and evening peak are shown in Fig.6, Fig.7 and Fig.8, respectively. And the transfer flow matrix heat maps of Sunday morning peak, off-peak and evening peak are shown in Fig.9, Fig.10 and Fig.11, separately.

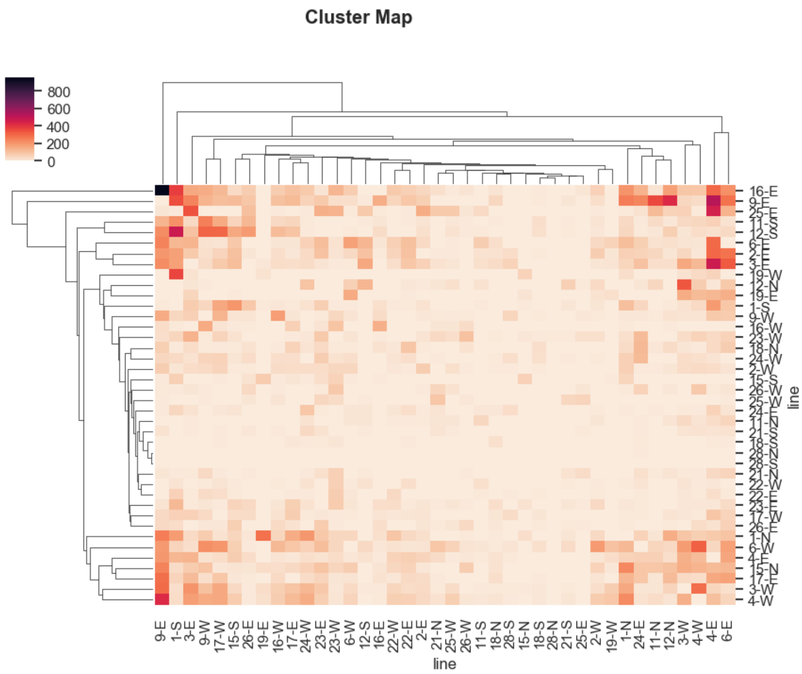

Fig. 6 Transfer flow matrix on Monday morning peak

As shown in Fig.6, on Monday morning, line 9 east direction and line 16 east direction has the highest transfer flow with more than 800 people. Next, line 1 south direction and 16 east, line 1 south and line 9 east direction, line 4 east direction and line 9 east direction, line 4 east direction and line 3 east direction and so forth with the dark color in the map have the closer connection. Consider the map, this in the real world shows that the transfer hub is in The Hague central station where each line combination has linked to each other. This is in line with the reality since on weekday morning, people travel from resident area to the central station where most lines confluence and then change to the line with their desired destination. In addition, line 1, line9 and line 16 also have the transferability at The Hague HS station, which is another important hub in The Hague area.

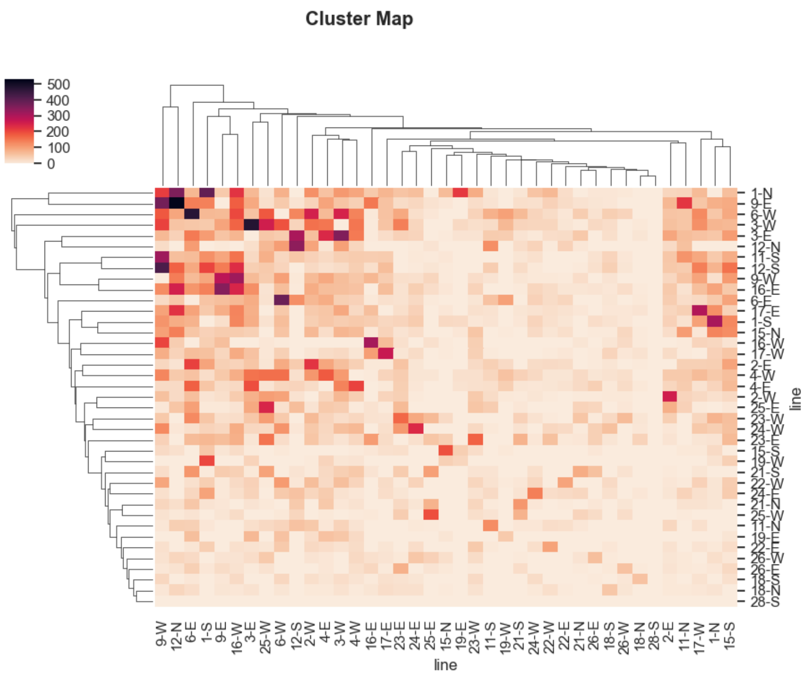

Fig. 7 Transfer flow matrix on Monday off-peak

The transfer flow pattern on Monday off-peak hour is dissimilar with Monday morning peak. In general, the volume of transfer is lower and thus reflect to the total number of trips decreases. This is because during the weekdays, the main trip purpose is commuting and hence the number of trips during the off-peak correspondingly decreases. The high connection line bundles are line 12 north direction and line 9 east direction, line 9 west direction and line 12 south direction, etc. Line 12 and line 9 has the transferability at The Hague HS station, which is regarded as an important hub in The Hague area.

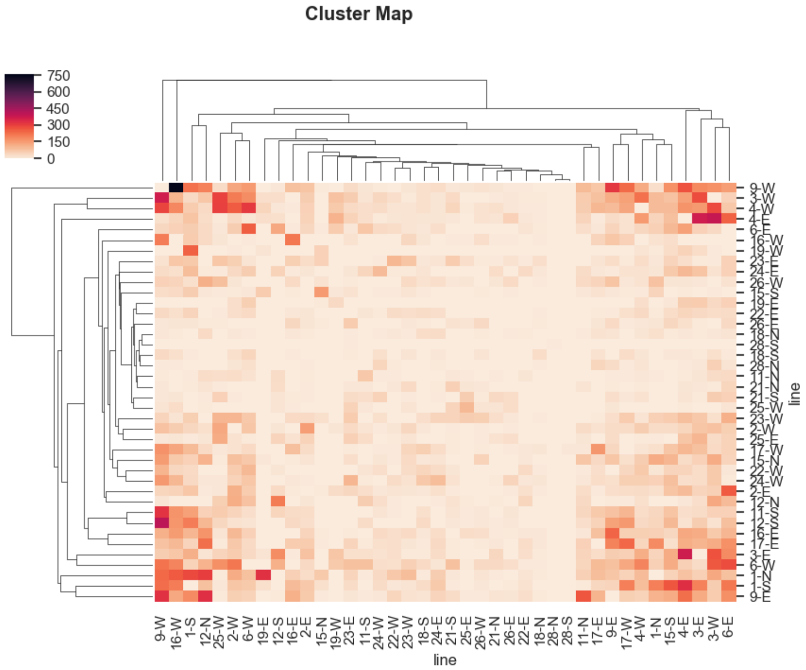

Fig. 8 Transfer flow matrix on Monday evening peak

The transfer flow pattern on Monday evening peak is in very big contrast of that on the morning peak. For example, the highest transfer happens between line 16 west and line 9 west, which is totally the other way around of that on the morning peak. Other line combinations with closer connection are line 3 west and line 9 west, line 9 west and line 12 south, etc. The high transfer volume exists on the line from central station area to the residential areas.

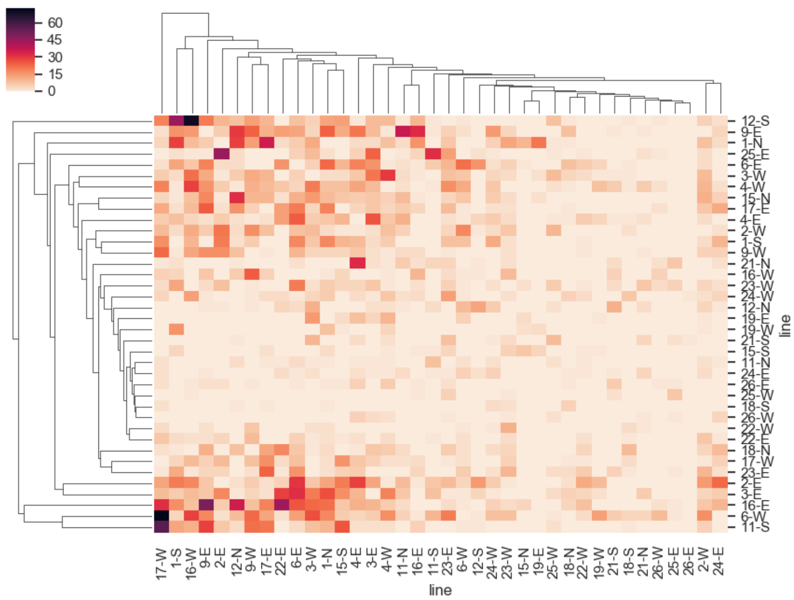

Fig.9 Transfer flow matrix on Sunday morning peak

The overall transfer volume on Sunday morning peak is extremely low with the highest transfer value of around 80 passengers. There is almost no commuting on weekends and only line 17 west and line 6 west, line 17 and line 11 south, line 16 west and line 12 south have close connection. This could be visiting friends or family in another town or other purposes.

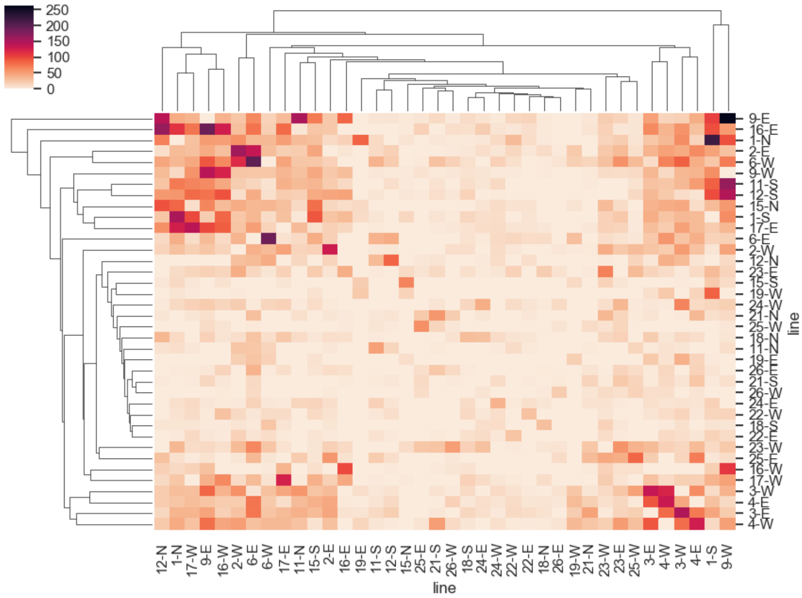

Fig.9 Transfer flow matrix on Sunday off-peak

The total number of trips starts to increase around noon. Line 1 north and line 4 east, line 16 east and line 9 east, line 6 east and line 2 east and so on have the higher transfer volume. These lines are associated to The Hague central area where shopping mall, catering service and entertainment agglomerate. The main travel purpose on weekends is about to enjoy leisure activities and therefore people travel from towns or resident areas to the commercial district.

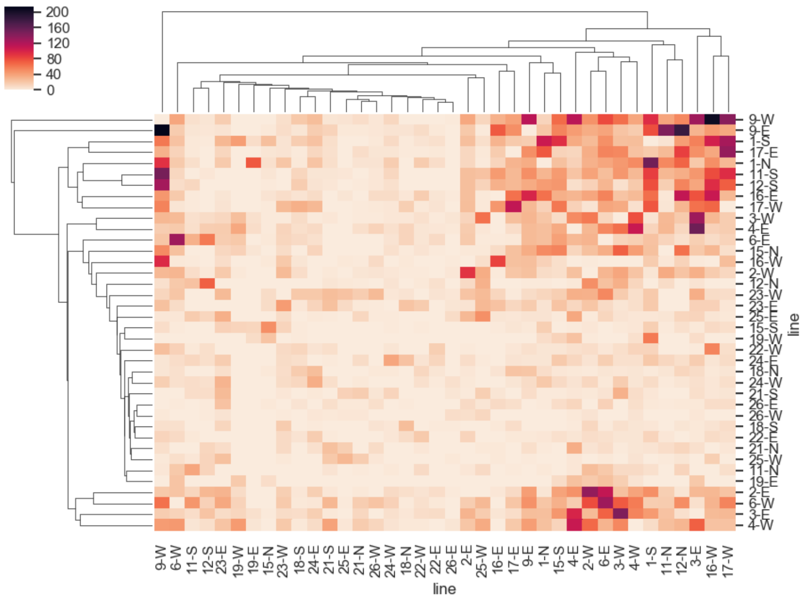

Fig.10 Transfer flow matrix on Sunday evening peak

During the evening peak on Sunday, the total number of trips still remain at a higher level compared to the morning. As shops are normally closed at that time, returning home or catering could be the main intention of travelling. The strongly connected line bundles are line 9 west and line 16 west, line 9 west and line 12 north, line 9 west and line 1 south and etc. This is approximately the opposite direction of travelling as to the noon. After the entertainment of the day, people tend to be back home around evening or going the central area for dinner.

In a word, on the weekdays, people tend to travel from residential areas to the central station area for either working or accessing to train in the morning. But in the evening, the travel pattern shows an adverse pattern from central business area to the residential area. The number of trips during the off-peak hour is lower than peak hours as expected. On the weekends, the total number of trips is lower than during the weekdays. The object of travelling can be inferred from the travel pattern as during the weekdays people travel for commuting while during the weekends people travel for entertainment. However, the strong association remains the same from residential area to the business area no matter the day of week and time period of a day.

5.2 Weighted C-space with Louvain community detection method

With the transfer flow matrices, the weighted C-space representations have been directly derived from those. Initially, map data was used to construct the topological representation. However, there is an ambiguity existing that the name of node could represent a certain boarding/alighting station or a group of adjacent stops that serves for the whole area. Therefore, the method of first construct and then assign the transfer flow has been discarded but the direct derivation has been deployed.

The Louvain community detection is unsupervised and automatic. The network layer has thus been separated by the community detection method with nodes in different color as depicted in the figures. The community clusters the nodes based on the attribute of edges and in this case study, since C-space network representation is introduced, the nodes are the lines. Hence, the separation of a PTN has naturally formed. The details of clustering and the interpretation of results are discussed in the next section while this section is dedicated for the visualization of C-space network representation with communities. The link color and length within a graph represent the volume of transferring and the same node color is for the same community. Note that different graph is lack of comparison as the benchmark has been changed. The C-space network representation with communities of Monday morning peak, off-peak and evening peak are shown in Fig.11, Fig.12 and Fig.13, respectively. And the C-space network representation with communities of Sunday morning peak, off-peak and evening peak are shown in Fig.14, Fig.15 and Fig.16, separately.

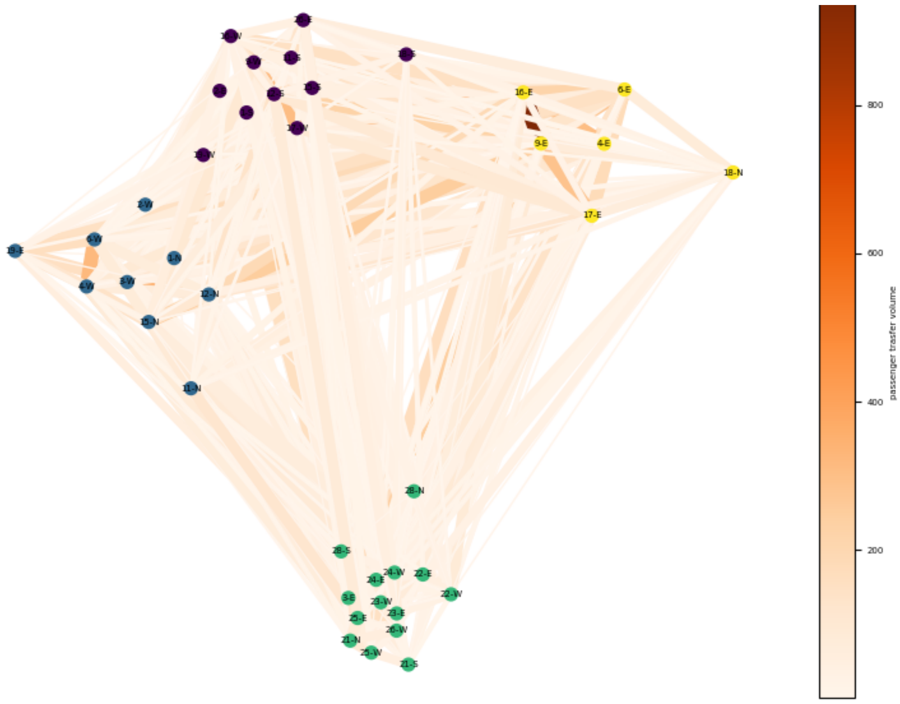

Fig. 11 C-space network representation with communities on Monday morning peak

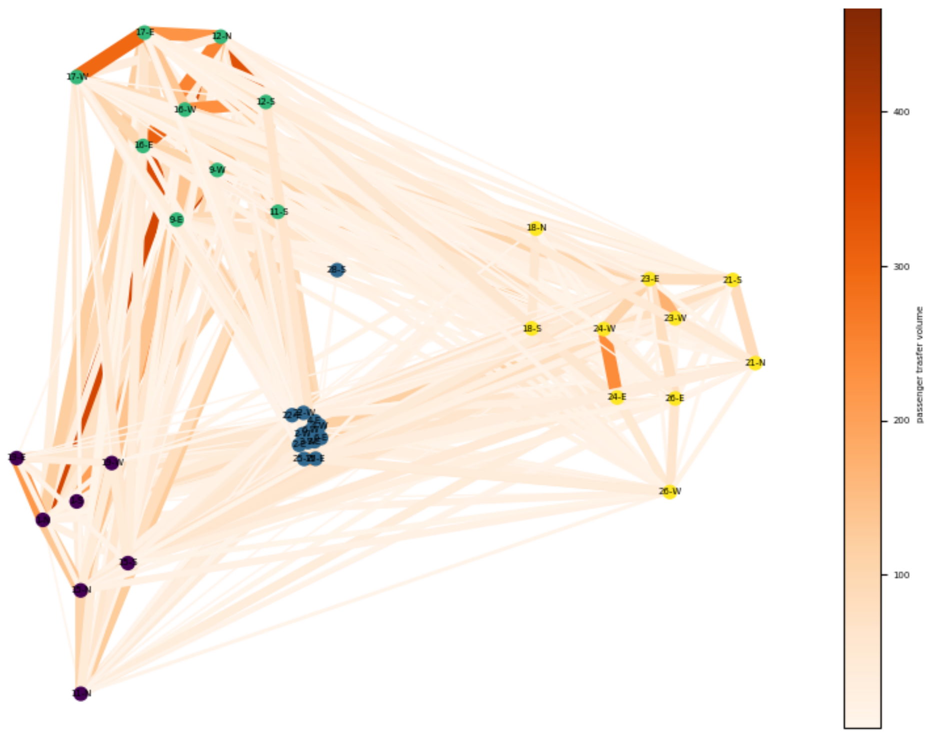

Fig. 12 C-space network representation with communities on Monday off-peak

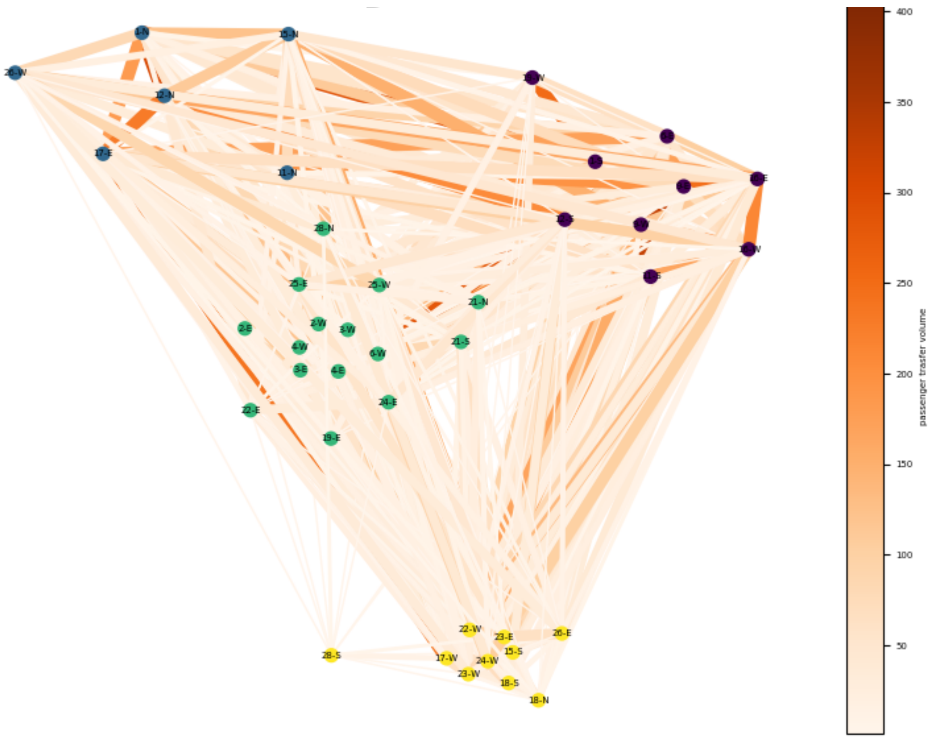

Fig. 13 C-space network representation with communities on Monday evening peak

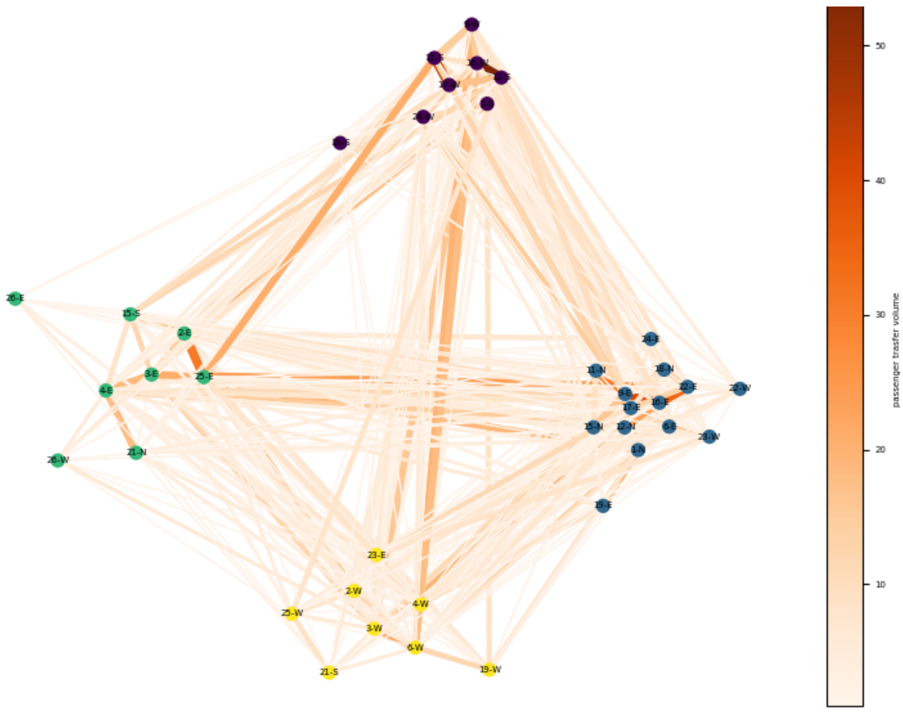

Fig.14 C-space network representation with communities on Sunday morning peak

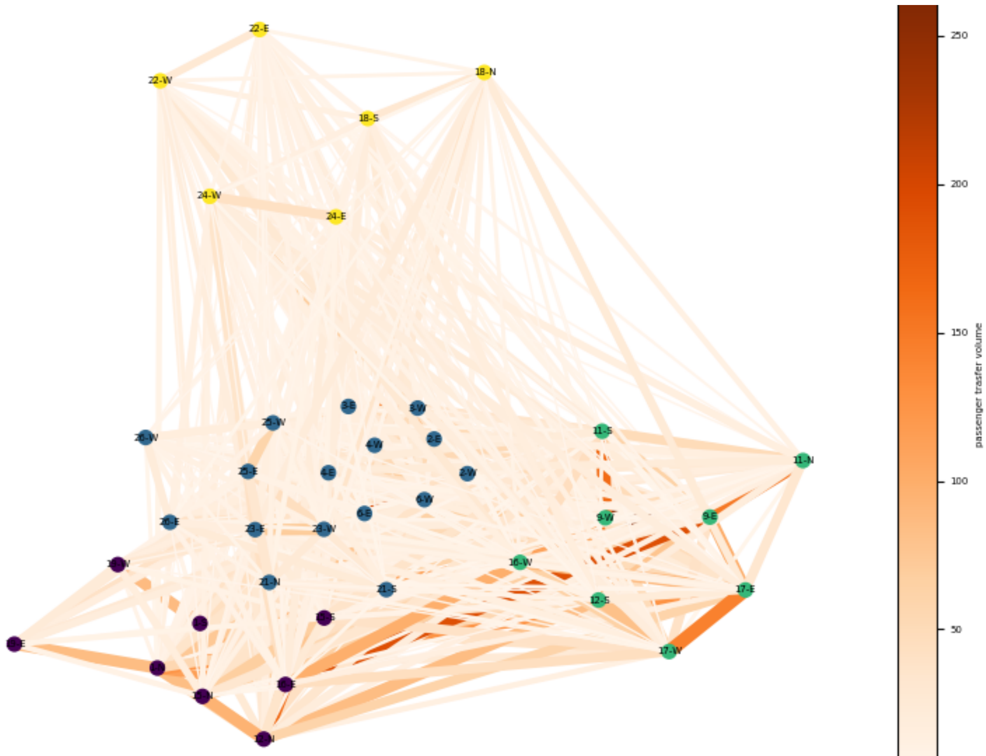

Fig.15 C-space network representation with communities on Sunday off-peak

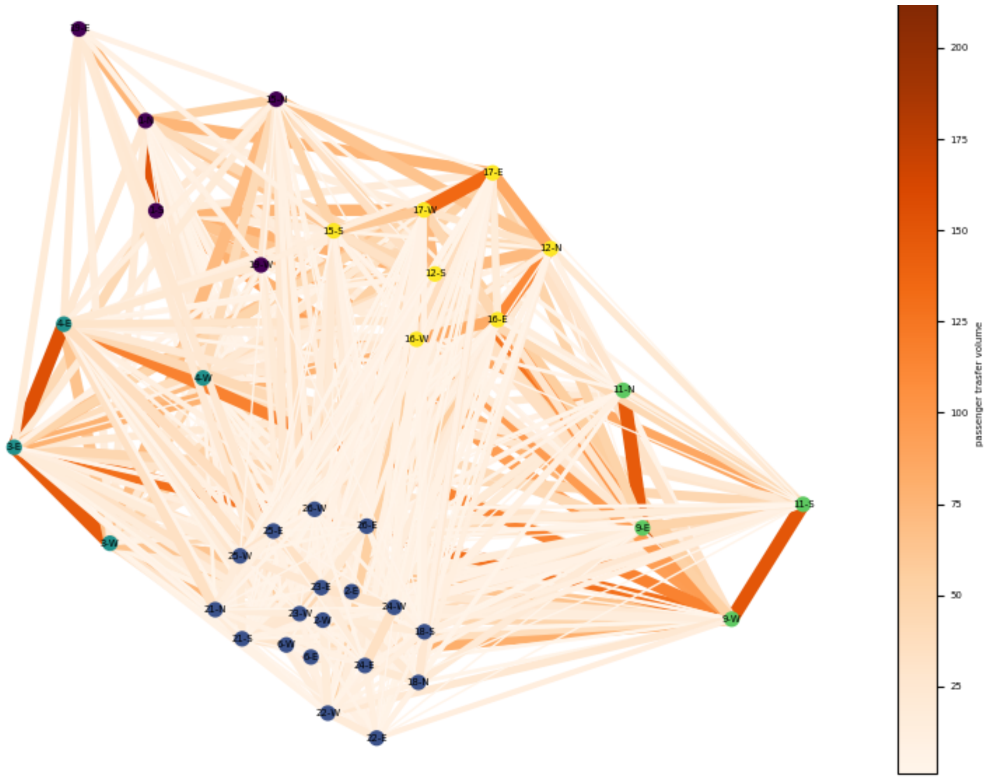

Fig.16 C-space network representation with communities on Sunday evening peak

To conclude, the community detection and the C-space network representation facilitate the separation of network layers with quantitative approach. The separation is objective and incorporate the data no matter the mode of transportation, which is considered as an improvement on top of the previous works.

5.3 Hierarchy of PTNs

The identification of the hierarchy of PTNs is using the methodology, regarding the inner- and intra-transfer flow volume. The higher of a hierarchy should have the higher ratio of inner transfer flow within a community and the intra transfer flow between this community and all other communities. The ratio is actually reflecting the ability of meeting self-sufficiency, which is the importance of connection among several lines in the community.

The resulting hierarchy of the case study with time periods is shown in Tab.3 to 8 below:

Tab.3 Hierarchy of PTN on Monday morning peak

Level	Lines	Sum of inner transfer flow	Sum of intra transfer flow	Ratio
1	1-N, 2-W, 3-W, 4-W, 6-W, 11-N,12-N, 15-N, 19-E, 21-S	5442	6726	0.81
2	1-S, 2-E, 9-W, 11-S, 12-S, 15-S, 16-W, 17-W, 19-W, 26-E	5188	9178	0.57
3	4-E, 6-E, 9-E, 16-E, 17-E, 18-N	4739	9527	0.50
4	3-E, 18-S, 21-N, 22-W, 22-E, 23-W, 23-E, 24-E, 24-W, 25-E, 25-W, 26-W, 28-N, 28-S	3741	9269	0.40

The identification of hierarchy on Monday morning peak is in line with the conventional understanding towards morning peak in a weekday which means that the higher level of PTN is where people travel from residential area to CBD area. In the first level, all lines except line 11 north, line 12 north, line 19 east are connected to The Hague central station while line 11 north, line 12 north, line 1 north are connected via The Hague HS station. This shows that on Monday morning, the highest hierarchy of a PTN is connected to the business zone or to the central railway station where lines agglomerate and interchange. This also indicates that the hub in a PTN is vital and the lines with connection to the hub are forming the highest level of a PTN during a weekday morning peak. The highest level layer with a ratio of inner- and intra- transfer flow with 0.81 while the second highest is 0.57. It in turn represents a relatively bigger gap between the first level and other levels. In contrast, the direction of a line to the residential area or most of the bus line with less capacity for passengers are regarded as lower level.

Tab.4 Hierarchy of PTN on Monday off-peak

Level	Lines	Sum of inner transfer flow	Sum of intra transfer flow	Ratio
1	2-E, 2-W, 3-W, 3-E, 4-W, 4-E, 6-E, 6-W, 22-W, 22-E, 25-E, 25-W, 28-S	10045	9442	1.06
2	9-W, 9-E, 11-S, 12-N, 16-E, 16-W, 17-W, 17-E, 12-S	8274	8362	0.99
3	18-N, 18-S, 21-N, 21-S, 23-W, 23-E, 24-E, 24-W, 26-E, 26-W	2569	3901	0.66
4	1-S, 1-N, 11-N, 15-N, 15-S, 19-E, 19-W	2401	5942	0.40

From Tab.4 we can see that during the off-peak hours on Monday, the margin of the first level and the second level is not that significant, which indicates a small gap between the two. The off-peak hour hierarchy is totally different from the morning peak. Although the bus is with less capacity, compared to tram, some of the lines including 22 and 25 both direction, 28 south have been into the first level as these lines travers through center. And line 2, 3, 4, 9, 16, 17 both direction have been clustered into higher level. Those lines are run through The Hague area via central station area from east to west with different start and end stop. Thus, during the off-peak of a weekday, some bus lines through central business area are upgraded into higher level and lines with the ability to cross the area with rather long distance are clustered into higher level as well.

Tab.5 Hierarchy of PTN on Monday evening peak

Level	Lines	Sum of inner transfer flow	Sum of intra transfer flow	Ratio
1	1-S, 9-W, 9-E, 11-S, 16-E, 16-W, 12-S, 19-W	5945	9380	0.63
2	1-N, 12-N, 17-W, 17-E, 26-W, 28-S	6244	10092	0.62
3	2-E, 2-W, 3-W, 3-E, 4-W, 4-E, 19-E, 25-E, 25-W, 28-N	5518	9654	0.57
4	6-E, 6-W, 11-N, 15-N, 15-S, 18-S, 21-S, 22-W, 22-E, 23-W, 23-E, 24-E, 24-W, 26-E, 18-N, 21-N	1619	6046	0.27

Overall, the first level of Monday evening peak shows an adverse from that of Monday morning peak. Line 1 south, line 11 south, line 12 south and line 19 west are the opposite direction from the first level of Monday morning. This shows the end of a weekday, people from commuting to work becomes to travel home. Line 9 both directions and line 16 both directions are in the first level as they run through two railway stations, namely The Hague central and The Hague HS. Thus, the flow between two stations is high and thus promotes the line 9 and line 16 into highest level. However, the ratio margin among first three levels are not significant, especially the first two. Line 1 and line 17 are also lines via two railway stations while line 12 north is the only line with the destination to The Hague Duindorp area from HS station. Most of the bus lines are once again regarded to the lowest level, except 25,28 traversing central areas and 26 west traversing from station Voorburg via HS station to the west residential area.

Tab.6 Hierarchy of PTN on Sunday morning peak

Level	Lines	Sum of inner transfer flow	Sum of intra transfer flow	Ratio
1	6-E, 9-E, 11-N, 12-N, 15-N, 16-E, 17-E, 18-N, 22-W, 22-E, 23-W, 24-E	556	681	0.82
2	1-N, 2-W, 3-W, 4-W, 6-W, 19-E, 19-W, 21-S, 23-E, 25-W	362	548	0.66
3	2-E, 3-E, 4-E, 15-S, 21-N, 25-E, 26-E, 26-W	224	452	0.50
4	1-S, 9-W, 11-S, 12-S, 16-W, 17-W, 18-S, 24-W	305	663	0.46

On Sunday morning, as shown in the Tab.6, the transfer flow is at a very low volume with the highest level approximately ten times less than that of Monday morning. The trip purpose changes on the weekends and people do not need to commute in the morning. Line 9 east, line 11 north are connected to Scheveningen, which is the famous beach in The Hague area and travelers may want to go on a hiking. Line 6 east, line 12 north are connected to towns near The Hague. While all other tram and bus lines at the highest level are connected to the center.

Tab.7 Hierarchy of PTN on Sunday off-peak

Level	Lines	Sum of inner transfer flow	Sum of intra transfer flow	Ratio
1	2-E, 2-W, 3-W, 3-E, 4-W, 4-E, 6-E, 6-W, 21-N, 21-S, 23-W, 23-E, 25-E, 25-W, 26-E, 26-W	6400	4097	1.56

2	9-W, 9-E, 11-S, 11-N, 12-S, 16-W, 17-W, 17-E	2772	4236	0.65
3	1-S, 1-N, 12-N, 15-N, 15-S, 16-E, 19-E, 19-W	1804	3399	0.53
4	18-N, 18-S, 22-W, 22-E, 24-E, 24-W	378	910	0.42

During the off-peak time period on Sunday, which is the noon, interchange of passenger flow starts to increase. The first level of hierarchy shows a remarkable gap with other levels. Line 2, 3, 4, 6 traverse through The Hague area via center. Line 3 and 4 are from a relatively large residential city Zoetermeer to The Hague. 4 bus lines, including 21, 23, 25 and 26 are in the first level. Those lines are running from the suburbs of The Hague to either the center area or HS station area where entertainment service agglomerates. The hierarchy of PTN on Sunday off-peak shows that people start taking on a trip till noon and the high hierarchy existing among the lines from towns or dwelling district to the downtown area.

Tab.8 Hierarchy of PTN on Sunday evening peak

Level	Lines	Sum of inner transfer flow	Sum of intra transfer flow	Ratio
1	2-E, 2-W, 6-E, 6-W, 18-N, 18-S, 21-S, 21-N, 22-W, 22-E, 23-W, 23E, 24-E, 24-W, 25-E, 25-W, 26-E, 26-W	3081	3402	0.91
2	3-W, 3-E, 4-W, 4-E	1127	2648	0.43
3	12-S, 12-N, 15-S, 16-E, 16-W, 17-W, 17-E	1539	3758	0.41
4	1-S, 1-N, 15-N, 19-E, 19-W	638	1636	0.39
5	9-W, 9-E, 11-S, 11-N	796	2186	0.36

Surprisingly, all the bus lines during the Sunday evening peak are regarded as the highest hierarchy. This could be resulted from the trip chain of a trip. People combine the mode of bus and tram to travel at this time period, starting their trip from the suburban area or towns to the center or the other way around. Use the connection of tram and bus to fulfill a trip. Only tram line 2 and line 6 with almost the same direction and intermediate stops with line 2 extended to the west of The Hague are clustered into the highest level. These two lines run through the center area and start from Leidschendam. The gap between levels are quite remarkable as all other time period on Sunday.

6 Conclusion

In this study, a data-driven, generic and transfer-based methodology for separation and ranking the PTNs has been put forward. With the hierarchy of a network, this is beneficiary for the management and operation of operators for focusing on the higher level network layer and in turn provide better service for passengers. The study introduces three steps to rank the hierarchy of a PTN: (1) using the passenger journey and ride data to derive transfer flow matrix; (2) applying C-space network representation with community detection method to separate and visualize the PTN layer; (3) performing ranking method, regarding inner- and intra-transfer flow. To this end, the hierarchy of a PTN could be presented with temporal attributes. Different day of week and various time period of a day could potentially yield different hierarchy. The proposed unsupervised learning algorithm is based on passenger transfer flow data, independent from geographic location and the mode of transportation. The study shows that the level is changing based on the selected time slot and can be a mixture of different modes, which is dissimilar from the hierarchy purely based on qualitative method.

The approach applied in this study is able to be transferred into any case study as long as the requirement of data meets. Thus, the identification of the PTNs hierarchy is more flexible and quantitative. The study shows that during the peak hours of weekday, the highest hierarchy of a PTN is connected to the business zone or to the central railway station where lines agglomerate and interchange. This also indicates that the hub in a PTN is vital and the lines with connection to the hub are forming the highest level of a PTN. While the off-peak time period bus lines with shorter operation distance through central business area are upgraded into higher level and lines with the ability to cross the area with rather long distance are clustered into higher level as well. However, during the weekends, interchange of passenger flow starts to increase till noon and the high hierarchy existing among the lines from towns or dwelling district to the downtown area. Bus lines are regarded as the highest hierarchy in the evening probably due to the shorter distance of travel and resulted from the trip chain from suburban area to the center areas. Moreover, the result also shows that the gap between the first level and other levels after weekends noon and the morning peak hours in weekdays are significantly higher than other time period or day of week, indicating a larger margin of the priority of the usage.

Although the data-driven method shows the strong ability to identify the hierarchy of PTNs with quantitative approach, there are still several potentials in this report and the methodology. First of all, the C-space network representation neglects the attributes of spatial development. This means that the line has been degraded into nodes without the spatial characteristics. When moving the identification of hierarchy, the whole line is regarded to a certain level. However, maybe only a section or two of this line are at higher level but not the whole line. This can be further addressed by other network representation. Second of all, the topic is based on the information of passenger transfer data and hence overlooks the direct trip. For example, if the line is a direct line from residential area to the CBD area, there could be less transfer flow but high direct flow. But in the method applied in this report, this line could be regraded at a low hierarchy. The incorporation of direct trips could be the further research and study direction. Third of all, the data from HTM is useful but the amount could be more. The sample data is able to reflect the travel pattern and the hierarchy of the network but more data could yield more significant or different result.

Bibliography

Beyond connectivity: multiscale and enriched topological representation of public transport networks based on service properties. (2019).

MT-ITS_2015_IEEE_Haznagy-London.pdf. (n.d.).

Aleta, A., Meloni, S., & Moreno, Y. (2017). A Multilayer perspective for the analysis of urban transportation systems. *Scientific Reports*, *7*(July 2016), 1–9. https://doi.org/10.1038/srep44359

Asadi Bagloee, S., & Ceder, A. A. (2011). Transit-network design methodology for actual-size road networks. *Transportation Research Part B: Methodological*, *45*(10), 1787–1804. https://doi.org/10.1016/j.trb.2011.07.005

Barthélemy, M. (2011). Spatial networks. *Physics Reports*, *499*(1–3), 1–101. https://doi.org/10.1016/j.physrep.2010.11.002

Cats, O., Wang, Q., & Zhao, Y. (2015). Identification and classification of public transport activity centres in Stockholm using passenger flows data. *Journal of Transport Geography*, *48*, 10–22. https://doi.org/10.1016/j.jtrangeo.2015.08.005

Chien, S. I., Spasovic, L. N., Elefsiniotis, S. S., & Chhonkar, R. S. (2007). Evaluation of Feeder Bus Systems with Probabilistic Time-Varying Demands and Nonadditive Time Costs. *Transportation Research Record: Journal of the Transportation Research Board*, *1760*(1), 47–55. https://doi.org/10.3141/1760-07

Feng, J., Li, X., Mao, B., Xu, Q., & Bai, Y. (2017). Weighted complex network analysis of the Beijing subway system: Train and passenger flows. *Physica A: Statistical Mechanics and Its Applications*, *474*(January), 213–223. https://doi.org/10.1016/j.physa.2017.01.085

Feng, Y., Zhuang, Y., & Pan, Y. (2003). *A hierarchical approach*. https://doi.org/10.1145/564437.564484

Fortunato, S., & Hric, D. (2016). Community detection in networks: A user guide. *Physics Reports*, *659*, 1–44. https://doi.org/10.1016/j.physrep.2016.09.002

Gallotti, R., & Barthelemy, M. (2015). The multilayer temporal network of public transport in Great Britain. *Scientific Data*, *2*, 1–8. https://doi.org/10.1038/sdata.2014.56

Gao, J., Zhao, P., Zhuge, C., & Zhang, H. (2012). Research on public transit network hierarchy based on residential transit trip distance. *Discrete Dynamics in Nature and Society*, *2012*. https://doi.org/10.1155/2012/390128

Lin, J., & Ban, Y. (2013). Complex Network Topology of Transportation Systems. *Transport Reviews*, *33*(6), 658–685. https://doi.org/10.1080/01441647.2013.848955

Luo, D., Cats, O., & van Lint, H. (2017). Constructing Transit Origin–Destination Matrices with Spatial Clustering. *Transportation Research Record: Journal of the Transportation Research Board*, *2652*(1), 39–49. https://doi.org/10.3141/2652-05

Marshall, S. (2006). Transit-Oriented Hierarchy : a hierarchical network structure for street-based public transport. *Flux Journal*, 1–26. Retrieved from https://www.academia.edu/5488983/Transit-Oriented_Hierarchy

Mohaymany, A. S., & Gholami, A. (2009). Multimodal Feeder Network Design Problem: Ant Colony Optimization Approach. *Journal of Transportation Engineering*, *136*(4), 323–331. https://doi.org/10.1061/(asce)te.1943-5436.0000110

Newman, M. E. J., & Girvan, M. (2004). Finding and evaluating community structure in networks. *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*, *69*(2 2), 1–16. https://doi.org/10.1103/PhysRevE.69.026113

Pagani, A., Mosquera, G., Alturki, A., Johnson, S., Jarvis, S., Wilson, A., … Varga, L. (2019). Resilience or robustness: Identifying topological vulnerabilities in rail networks. *Royal Society Open Science*. https://doi.org/10.1098/rsos.181301

Sui, Y., Shao, F., Yu, X., Sun, R., & Li, S. (2019). Public transport network model based on layer operations. *Physica A: Statistical Mechanics and Its Applications*, *523*, 984–995. https://doi.org/10.1016/j.physa.2019.04.269

van Nes, R. (2002). Multilevel Network Optimization for Public Transport Networks. *Transportation Research Record: Journal of the Transportation Research Board*, *1799*(1), 50–57. https://doi.org/10.3141/1799-07

van Nes, R. (2007). Multiuser-Class Urban Transit Network Design. *Transportation Research Record: Journal of the Transportation Research Board*, *1835*(1), 25–33. https://doi.org/10.3141/1835-04

Verma, A., & Dhingra, S. L. (2006). Developing Integrated Schedules for Urban Rail and Feeder Bus Operation. *Journal of Urban Planning and Development*, *132*(3), 138–146. https://doi.org/10.1061/(asce)0733-9488(2006)132:3(138)

Verma, T., Araújo, N. A. M., & Herrmann, H. J. (2014). Revealing the structure of the world airline network. *Scientific Reports*, *4*(May). https://doi.org/10.1038/srep05638

Verma, T., Russmann, F., Araújo, N. A. M., Nagler, J., & Herrmann, H. J. (2016). Emergence of core-peripheries in networks. *Nature Communications*, *7*, 1–7. https://doi.org/10.1038/ncomms10441

Von Ferber, C., Holovatch, T., Holovatch, Y., & Palchykov, V. (2009). Public transport networks: Empirical analysis and modeling. *European Physical Journal B*, *68*(2), 261–275. https://doi.org/10.1140/epjb/e2009-00090-x

Yap, M., Luo, D., Cats, O., van Oort, N., & Hoogendoorn, S. (2019). Where shall we sync? Clustering passenger flows to identify urban public transport hubs and their key synchronization priorities. *Transportation Research Part C: Emerging Technologies*, *98*(March 2018), 433–448. https://doi.org/10.1016/j.trc.2018.12.013

Yildirimoglu, M., & Kim, J. (2018). Identification of communities in urban mobility networks using multi-layer graphs of network traffic. *Transportation Research Part C: Emerging Technologies*, *89*, 254–267. https://doi.org/10.1016/j.trc.2018.02.015

Zhang, L., Yang, H., Wu, D., & Wang, D. (2014). Solving a discrete multimodal transportation network design problem. *Transportation Research Part C: Emerging Technologies*, *49*, 73–86. https://doi.org/10.1016/j.trc.2014.10.008