Unraveling the Hierarchy of Public Transport Networks

Wang, Ziyulong; Luo, Ding; Cats, Oded; Verma, Trivik

DOI
10.1109/ITSC45102.2020.9294342

Publication date
2020

Document Version
Final published version

Published in
2020 IEEE 23rd International Conference on Intelligent Transportation Systems (ITSC)

Unraveling the Hierarchy of Public Transport Networks

Ziyulong Wang*, Ding Luo†, Oded Cats‡, Trivik Verma§

*Dept. of Transport and Planning, Delft University of Technology, Delft, The Netherlands, wangziyulong@gmail.com
†Dept. of Transport and Planning, Delft University of Technology, Delft, The Netherlands, D.Luo@tudelft.nl
‡Dept. of Transport and Planning, Delft University of Technology, Delft, The Netherlands, O.Cats@tudelft.nl
§Dept. of Multi-Actor Systems, Delft University of Technology, Delft, The Netherlands, T.Verma@tudelft.nl

***Abstract*—Hierarchy is regarded as a natural phenomenon of public transport networks (PTN). The imbalanced distribution of passenger flow result in a hierarchical structure of PTN and it is also related to the development of technology and the introduction of new modes. However, there is still a lack of knowledge on how to identify the hierarchical structure of the multi-layer PTN. This study proposes a three-step passenger transfer flow based methodology for separating and ranking the PTN: (1) using passenger journey data to derive transfer flow matrix; (2) applying network representation with Louvain method of community detection to separate the PTN layers; (3) performing ranking method, separating inner-transfer and inter-transfer flow. To demonstrate our method, we use one-month smart card data of The Hague, the Netherlands provided by the PTN operator HTM. Our results show that our method is able to, regardless of the geographic location and the mode of transportation, identify the hierarchy of PTN based on the passenger transfer flow pattern. Temporal attributes are also discussed to illustrate how hierarchy is time-dependent, e.g. with respect to the day of the week and the time of the day. Our method supports public transport (PT) operators during design and optimization of PTN and in determining which sets of higher-level service to prioritize during different time periods.**

***Index Terms*—public transport network, hierarchy, community detection, data-driven**

I. Introduction

Hierarchy is a feature of many domains with the meaning of order, inclusion, control or level [1]. For a given public transport network (PTN) with its transport service, the imbalanced distribution of passenger trips will result in a hierarchical structure [2]. Understanding the hierarchy of PTN can facilitate prioritizing transfer synchronization decisions and give priority to higher-level networks. This will help in promoting the attractiveness of public transport (PT) and increase the ridership as empirical research indicate that transfer is perceived as the most unfavorable part of a PT journey [3].

Many researchers have investigated multi-level PTN and various ways of defining the networks developed correspondingly. For example, one way is to define the multi-level PTN as a two-level PT system, consisting of two interconnected subsystems: an urban network and an interurban network [4]. Also, constructing multi-level PTN by categorizing different route hierarchies (e.g. mass, feeder and local) or other specific linkages (e.g. feeder route for rail or bus network) has been widely adopted and acknowledged [5]–[8]. Similarly, based on route hierarchies, trip distance can be added up to determine the optimal hierarchy of PTN [9]. In addition, multi-modality can be used in describing the multilevel PTN [10], [11].

Remarkably, most of the studies are based on either mode or other qualitative features. This means that the consideration of the relationship between PTN and passenger flow is still lacking.

Complex network science emerged as an effective and quantitative method to study PTN and is able to capture the topological properties of PTN structures [12], [13]. Several studies have applied network science indicators to analyze PTN, including estimation of passenger flow [14], identification of hubs [15] and accessibility analysis [16]. Besides, complex networks theory can also potentially offer an approach for unraveling network hierarchy since it is witnessed in several domains such as metabolic networks [17], social networks [18] and internet networks [19]. Notwithstanding, hitherto no study has performed such an analysis on PTN to the best of our knowledge.

Therefore, we conduct this study to unravel the hierarchy of PTN based on passenger transfer flow by using complex network theory. The temporal attributes are also discussed in order to represent the time-dependent characteristics of this hierarchy. The developed methodology, regardless of the geographic location and the mode of transportation, could unveil the comparatively important sets of service to support PT operators decisions in design and optimization of PTN.

The remainder of this paper is organized as follows: the second section describes the proposed methodology which is followed by the description of the case study in the third section. The fourth section presents the results and conclusions are drawn in the fifth section including a discussion of the future research direction.

II. Methodology

In this paper, we propose a data-driven, geography-independent and mode-agnostic methodology that is based directly on passenger flows rather than service design to identify the hierarchy of PTN. The methodology consists of three steps: deriving the topological representation of passenger transfer flow patterns, identifying clusters of PT lines and ranking PT line clusters.

The methodology of this study follows the flow diagram shown in Fig. 1. The Automatic Vehicle Location (AVL) data and Automatic Fare Collection (AFC) data are already processed to passenger journey data and passenger ride data. Based on these input data, we derive transfer flow matrices. Then, a topological representation is applied to represent the PTN with the consideration of transfer flow. Next, in order to cluster the line bundles, the Louvain Method of community detection is adopted. Furthermore, the time-dependent hierarchy of PTN is identified based on the inner-transfer and inter-transfer flow.

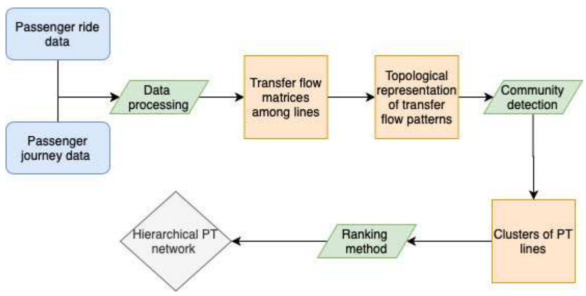

Fig. 1. Framework for unraveling the hierarchy of PTNs based on passenger transfer flow

A. Data

We use the pre-processed passenger journey data and passenger ride data as input to our analysis. Examples of these two types of data are presented in Tables 1 and 2, respectively. Passenger journey data contains information of the passenger ride(s) record(s) while the passenger ride data contains information on the route (i.e. line) traversed. We hereby select the journeys with at least one transfer. Finally, the origin tap-in time is used to match the start and the end of the selected time period for the investigation of temporal properties.

TABLE I
EXAMPLE OF PASSENGER JOURNEY DATA

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

TABLE II
EXAMPLE OF PASSENGER RIDE DATA

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

B. Topological Representation of Passenger Transfer Flow Patterns

A weighted directed graph $G = (V, E)$ is used to represent passenger transfer flows, where nodes $v \in V$ represent a certain PT line $l \in L$ and links $e \in E$ between nodes are built only when two lines have common stop (s) $s \in S$ where passengers can transfer [20]. The graph G is thus formulated as a weighted adjacency matrix A where a_{ij} represents the weight of the edge between i and j, which is the transfer flow f_{ij} between line i and j. A simple example of PTN with its corresponding topological representation is shown in Fig. 2.

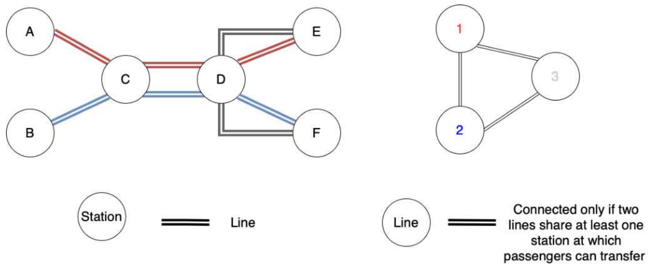

Fig. 2. A simple example of PTN with the C-space representation of it

C. Identifying Clusters of PT Lines

Based on the proposed topological representation weighted with the corresponding transfer flow, the concept of community detection is implemented to identify line bundles. Community detection is able to partition the network into communities with densely connected nodes and nodes in different communities being sparsely linked. In this paper, the weights are the transfer flow and thereby the network is partitioned into communities, where transfers between lines within the same cluster are maximized while transfer flows between lines belonging to different clusters are minimized.

The Louvain method is selected as the community detection technique to apply with the above-mentioned objective. It is a heuristic method based on modularity (Q) optimization and the most popular quality function is the modularity proposed by Newman and Girvan [21], which measures 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 nonexistent. The value of modularity (Q) is in the range $[-1, +1]$, which measures the density of links inside communities as opposed to links between communities. Its general expression is formulated using Eq. (1):

$$Q = \frac{1}{2m} \sum_{ij} \left(A_{ij} - \frac{k_i k_j}{2m} \right) \delta\left(\sigma_i, \sigma_j\right), \quad (1)$$

where A_{ij} is the adjacency matrix and m is the total number of edges. The summation term pertains to over all pairs of nodes i and j, in which k_i and k_j denotes the sum of weights of the edges attached to node i and j. σ_i represents the community of node i and the Kronecker delta function $\delta(\sigma_i, \sigma_j)$ is defined as shown in Eq. (2).

$$\delta(\sigma_i, \sigma_j) = \begin{cases} 1, & if\ \sigma_i = \sigma_j \\ 0, & otherwise. \end{cases} \quad (2)$$

The Louvain method consists of two repeated steps for maximizing the modularity: Initially, it assigns the nodes to communities with favoring the local optimization of modularity, i.e. when no individual node moving can improve the modularity; Next, it builds a new network based on the communities found in the first step where nodes are the communities now and then reapplies the algorithm in the first step until there is no improvement in the modularity value [22]. Hence, the partition found in the second step will contain the previous ones as well and the procedure will not terminate until the largest modularity is found.

The Louvain method is selected as it has two advantages: First, it is easy to apply and understandable; Second, it is computationally efficient since it only requires the edge dataset as an initial input. Moreover, it was found to be one of the best-performing clustering algorithms after a comparative analysis of community detection algorithms [23].

D. Ranking PT Line Clusters

Given the communities identified, which correspond to line bundles, we utilize two types of transfer flow - inner-transfer and inter-transfer - to identify the hierarchy of PTN. The inner-transfer flow is the transfer flow within a community. Suppose the PTN have been divided into n clusters, which means that n communities have been detected. The inner-transfer flow of a specific community c is calculated as follows:

$$f_c^{inner} = \sum_{i,j \in c} f_{ij}, \quad (3)$$

where f_{ij} represents the transfer flow between line i and j in the same community c while f_c^{inner} stands for the total inner-transfer flow volume within this community.

The inter-transfer flow corresponds to the total of transfer flow between a specific community c and all other $n-1$ communities. The inter-transfer flow of a specific community c is thus calculated as follows:

$$f_c^{inter} = \sum_{i \in c, k \notin c} f_{ik}, \quad (4)$$

where f_{ik} represents the transfer flow between line i from community c and k from all other $n-1$ communities except for c and therefore f_c^{inter} represents the frequency of communication between community c and other communities.

The higher level PTN community has frequent interchanges within the community and lesser communication between itself and other communities. We therefore identify PTN hierarchy in this study based on self-sufficiency as measured by the ratio between inner-transfer and inter-transfer flow. The magnitude of this ratio of a certain community reflects thus the importance of this community. The ratio is formulated as:

$$\theta_c = \frac{f_c^{inner}}{f_c^{inter}}, \quad (5)$$

where θ_c represents the ratio between inner-transfer and inter-transfer flow of a community c. A higher ratio of a community means a more important PTN community, corresponding to a higher-level set of services in the obtained PTN hierarchy.

III. Application

A. Case Study Description

We apply the proposed methodology to the urban PT network of the city of The Hague, the Netherlands, operated by HTM. The Hague is situated in the west of the Netherlands and has more than 500,000 residents. The urban PT network in the dataset consists of 12 tram lines and 8 urban bus lines as shown in Fig. 3.

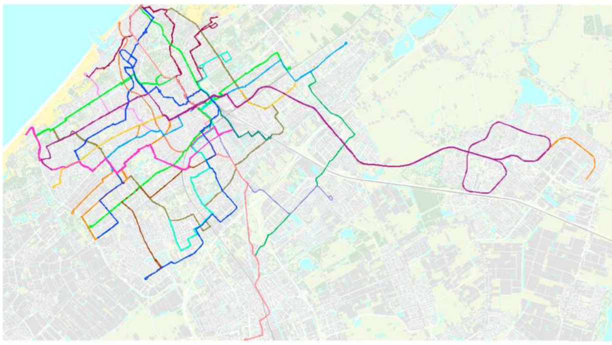

Fig. 3. Overview of urban public transport (bus and tram) in The Hauge

B. Selected Period Description

The pre-processed passenger journey and passenger ride data from March 1st to March 31st, 2015, was available as the initial input. The data cover 5 Sundays, Mondays and Tuesdays and 4 Wednesdays, Thursdays, Fridays and Saturdays. Since passenger transfer flow patterns vary for different times of the day and for weekdays versus weekends, we average the number of transfer by the type of day (weekday or weekend) for each 30-minute time slice throughout the day. The resulting trend in terms of the average transfer volume on weekday and weekend per half an hour is shown in Fig. 4.

The weekday passenger transfer pattern shows a two-peak trend while the weekend passenger transfer pattern exhibits only one spike. In the following, we illustrate the time-dependency of the PTN hierarchy by focusing on three time periods, namely AM peak (06:30-08:30), PM peak (16:30-18:30) and off-peak hour (11:30-13:30).

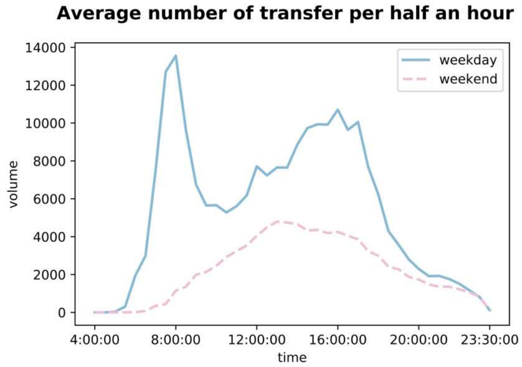

Fig. 4. Average transfer volume on weekday and weekend for each 30-minute time slice

IV. Result and Analysis

In the following we focus on the results of the weekday morning period for demonstration purposes. We select both weekday morning and weekday evening to establish a comparison for the identification of PT line clusters with topological representation thereof. Furthermore, in the subsection of the discussion of time-dependent hierarchy, all six selected periods are included to allow for a comparison.

A. Transfer Flow Derivation

The dataset consists of 40 line-directions. All 5,945,118 passenger journey data and 7,351,929 passenger ride data have been selected based on periods and processed into 40 $\times$ 40 transfer flow matrices.

A distinctive spatial-temporal pattern can be observed for different times-of-the-day and for weekday versus weekends. Conventionally during the weekdays, the morning peak travel pattern is almost the opposite of that of evening peak and the travel pattern in the off-peak hour is less pronounced. Moreover, the weekends' travel pattern is different from weekdays due to changes in the dominant trip purpose from commuting to leisure activity-oriented. The derived transfer flow on weekday morning peak is presented in Fig. 5. The line is denoted as line number with direction (e.g. 3-E means line 3 eastbound.)

It can be seen from Fig. 5 that several lines are the incoming line of a relatively large transfer flow in the morning peak period, namely line 3 eastbound, line 6 westbound, line 9 eastbound, and line 16 eastbound. These lines are connecting residential areas with the Hague center and traverse either The Hague central station or The Hague HS station or both stations.

B. Separation of network layers

Based on the derived transfer flow matrices, we construct the proposed directed weighted graph representation. We then apply for this graph the Louvain method of community detection. Figures 6 and 7 illustrate the graph for PTN on the weekday morning and evening where each community is shown using a different color. The community detection method yields four different and distinctive communities with different size for

Fig. 5. Chord chart of transfer flow on weekday morning peak

each analysis period. In addition, the link width represents the magnitude in terms of transfer flow where high-level line bundles have more inner-transfer flow within the community while low-level ones have more inter-transfer flow between communities, regardless of the time periods.

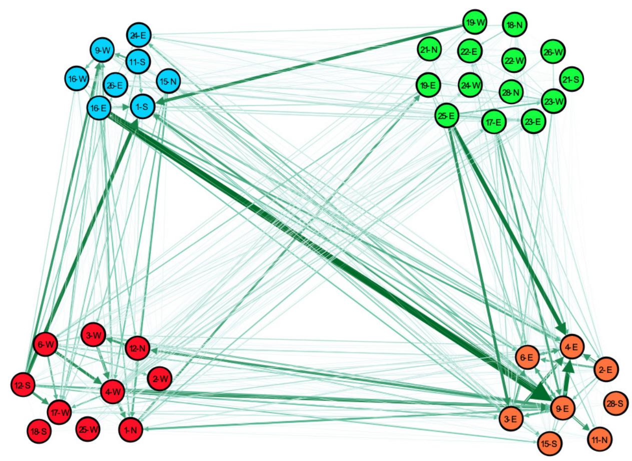

Fig. 6. C-space representation of PTN with community detection on weekday morning

C. Hierarchy of PTNs

Figure. 8 shows the obtained hierarchy of The Hague PTNs on weekday morning peak (WKDY_AM), weekday off-peak (WKDY_OP), weekday evening peak (WKDY_PM), weekend morning peak (WKED_AM), weekend off-peak (WKED_OP) and weekend evening peak (WKED_PM), respectively. Table 3

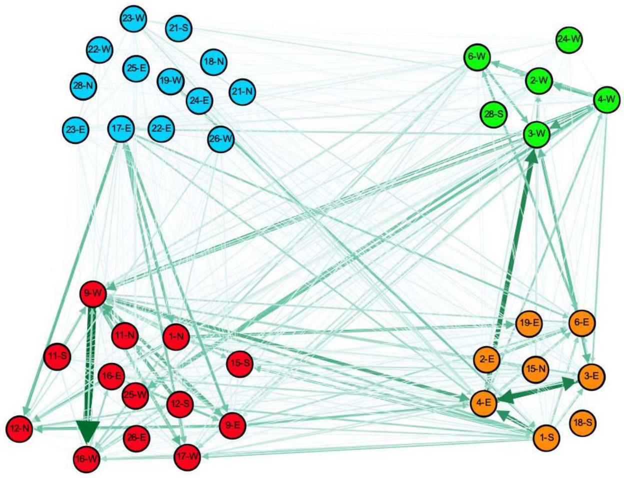

Fig. 7. C-space representation of PTN with community detection on weekday evening

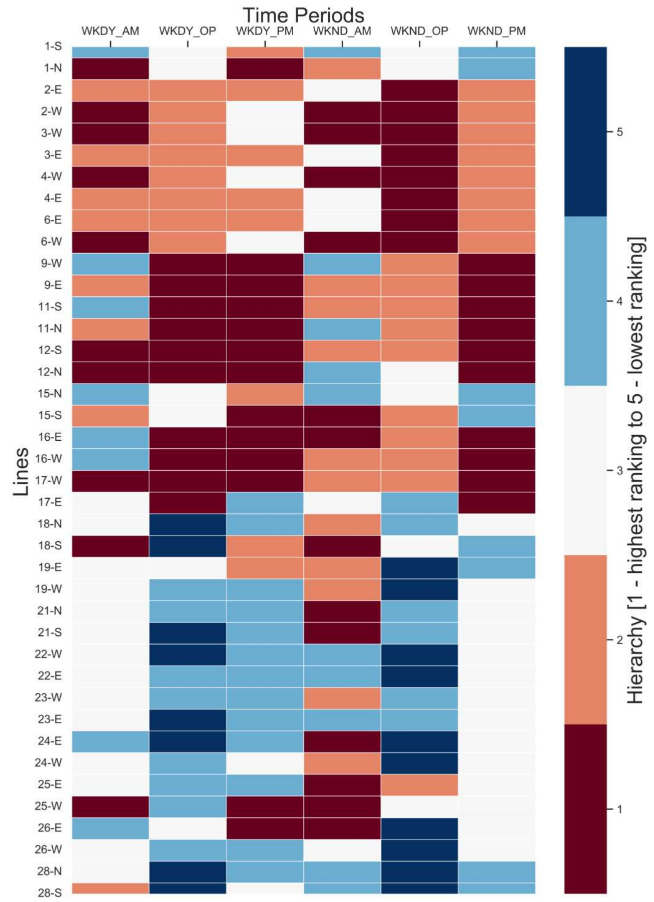

Fig. 8. Hierarchy of The Hague PTNs during different time periods

reports the ratio between inner-transfer and inter-transfer flow for each of the above-mentioned time periods.

TABLE III
THE RATIO OF INNER-TRANSFER AND INTER-TRANSFER FLOW OF THE HAGUE PTNS DURING DIFFERENT TIME PERIODS

Hierarchy	WKDY_AM	WKDY_OP	WKDY_PM
1	0.573	0.919	0.695
2	0.441	0.876	0.389
3	0.342	0.378	0.363
4	0.335	0.279	0.354
5	-	0.171	-
Hierarchy	**WKED_AM**	**WKED_OP**	**WKED_PM**
1	0.467	0.895	1.106
2	0.435	0.733	0.756
3	0.206	0.239	0.565
4	0.127	0.232	0.325
5	-	0.218	-

In general, it can be seen from Fig. 8 that the hierarchy of The Hague PTNs varies for different time periods due to changes in passenger transfer patterns. Trams (all with numeric label below 20) are always clustered into higher hierarchy layers, with some notable exceptions, such as bus lines 25 and 26 that have remarkable transferability with other lines. Tram lines traverse the two main railway stations (The Hague central station and The Hague HS station) are often identified as members of the higher hierarchy layer, for instance, line 9, 16 and 17. This concurs with the notion that hubs in PTN form a well-knitted self-contained set of transferring connections and consequently lines with the connection to the hubs are forming the highest level of PTNs. Interestingly, the extent of hierarchy exhibited in the network - as measured in terms of the gap in ratios between the higher hierarchy layer and the lower hierarchy layer - is lower in the morning periods on both weekdays and weekends than in other time periods as shown in Table 3 .

Figures 9 and 10 show the hierarchy map of The Hague on weekday morning and evening. During weekday mornings, the higher level of PTN consists of the lines where people travel from residential areas to the city center. As can be seen, the highest hierarchy layer consists of light rail lines connecting the large residential city of Zoetermeer (at the eastern edge of the network) to the city center, as well as high-frequency tram services that traverse the business areas and the major train stations. In contrast, during weekday evening, the lines in the city center are clustered into high hierarchies and several lines are no longer in the high hierarchy. This could be due to the non-uniform off-peak travel patterns. Moreover, both directions of a line are often grouped into the same level of hierarchy during the off-peak hours, whereas in the peak hours the more directional demand results in asymmetric flows and hence line direction rankings.

V. CONCLUSION

In this paper, a data-driven and passenger transfer flow-based methodology for identifying the hierarchy of PTN is proposed. The methodology consists of three-step: (1) using the passenger journey data to derive transfer flow matrix; (2) applying network representation with Louvain method of community detection to separate the PTN layers; (3) performing self-sufficiency ranking method to determine the hierarchy,

Fig. 9. Hierarchy of The Hague PTN on weekday morning

Fig. 10. Hierarchy of The Hague PTN on weekday evening

considering inner-transfer and inter-transfer flow. The mode-agnostic and geography-independent method is applied to the case of multi-modal network of The Hague, the Netherlands. Results show that the hierarchy varies for different time periods. The proposed data-driven method allows clustering lines into hierarchical layers without relying on any prior-knowledge. Future research may investigate the possibility to consider segments of lines as members of different hierarchical layers since segments of the same line may vary in their functionality. Another direction for further research is the potential incorporation of direct trips. While the proposed method only utilizes information from passenger journeys involving a transfer, direct trips may also provide information on line's position in the network. Our study though paves the way for further developments of topological approaches for analyzing passenger service network hierarchy.

REFERENCES

[1] Lane, D., "Hierarchy in Natural and Social Sciences", ed Pumain, D., Springer, Dordrecht, 81-119, 2006
[2] Van Nes, R., "Design of Multimodal Transport Networks: A Hierarchical Approach," Ph. D. Thesis, Technische Universiteit Delft, 2002
[3] Schakenbos, R., La Paix, L., Nijenstein, S., Geurs, K., "Valuation of a transfer in a multimodal public transport trip," Transport Policy, 46, 72-81, 2016
[4] Van Nes, R., "Multilevel Network Optimization for Public Transport Networks," Transportation Research Record, 1799 (1), 50–57, 2002
[5] Salzborn, F.J.M., "Scheduling bus systems with interchanges," Transportation Science 14 (3), 211–220, 1980
[6] Knoppers, P., Muller, T., "Optimized transfer opportunities in public transport," Transportation Science 29 (1), 101–105, 1995
[7] Shariat, A., Gholami, A., "Multimodal Feeder Network Design Problem: Ant Colony Optimization Approach," Journal of Transportation Engineering, 136 (4), 323-331, 2010
[8] Bagloee, S., Ceder, A., "Transit-network design methodology for actual-size road networks," Transportation Research Part B: Methodological, 45 (10), 1787-1804, 2011
[9] Gao, J., Zhao, P., Zhuge, C., Zhang, H., "Research on Public Transit Network Hierarchy Based on Residential Transit Trip Distance," Discrete Dynamics in Nature and Society, 2012, 1-14, 2012
[10] Gallotti, R., Barthelemy, M. "The multilayer temporal network of public transport in Great Britain," Sci. Data 2, 140056, 2015
[11] Aleta, A., Meloni, S. and Moreno, Y, "A Multilayer perspective for the analysis of urban transportation systems," Sci. Rep. 7, 44359, 2017
[12] Newman, M. E. J., "Networks: An introduction," Oxford University Press, New York, 2010
[13] Lin, J., Ban, Y., "Complex Network Topology of Transportation Systems," Transport Reviews, 33(6), 658–685, 2013
[14] Luo, D., Cats, O., van Lint, H., "Can passenger flow distribution be estimated solely based on network properties in public transport systems?" Transportation, 1-20, 2019
[15] Yap, M., Luo, D., Cats, O., van Oort, N., Hoogendoorn, S, "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, 433–448, 2019
[16] Luo, D., Cats, O., van Lint, H., Currie, G., "Integrating network science and public transport accessibility analysis for comparative assessment," Journal of Transport Geography, 80, 102505, 2019
[17] Ravasz, E., Somera, A. L., Mongru, D. A., Oltvai Z. N., Barabási A. L., "Hierarchical organization of modularity in metabolic networks," Science, 297(5586), 1551–1555, 2002
[18] Lu, C., Yu, J. X., Li, R. H., Wei, H., "Exploring hierarchies in online social networks," IEEE Transactions on Knowledge and Data Engineering, 28(8), 2086-2100, 2016
[19] Sun, Y., Wen, X., Zhao, Z., Li, Y., "Using complex network theory in the Internet engineering," 2012 7th International Conference on Computer Science & Education (ICCSE), Melbourne, VIC, 390-394, 2012
[20] Von Ferber, C., Holovatch, T., Holovatch, Y., Palchykov, V., "Public transport networks: Empirical analysis and modeling," European Physical Journal B, 68(2), 261–275, 2009
[21] Newman, M. E. J., and M. Girvan, "Finding and evaluating community structure in networks," Physical Review E - Statistical, Nonlinear, and Soft Matter Physics, 69(2), 026113, 2004
[22] Blondel, V. D., Guillaume, J. L., Lambiotte, R., Lefebvre, E., "Fast unfolding of communities in large networks," Journal of statistical mechanics: theory and experiment, 2008(10), P10008, 2008
[23] Lancichinetti, A., Fortunato, S., "Community detection algorithms: a comparative analysis," Physical review E, 80(5), 056117, 2009