Can Passenger Flow be Explained by Network Topology in Public Transport?

Luo, Ding; Cats, Oded; van Lint, Hans

Publication date
2018

Document Version
Accepted author manuscript

Published in
Proceedings of Conference on Advanced Systems in Public Transport (CASPT) 2018

Citation (APA)
Luo, D., Cats, O., & van Lint, H. (2018). Can Passenger Flow be Explained by Network Topology in Public Transport? In *Proceedings of Conference on Advanced Systems in Public Transport (CASPT) 2018: 23-25 July 2018, Brisbane, Australia* Article 67

CASPT 2018 Extended Abstract

Can Passenger Flow be Explained by Network Topology in Public Transport?

Ding Luo · Oded Cats · Hans van Lint

Abstract It has been rarely investigated in the field of public transport whether passenger flow can be explained by network topology. Based on the rich data sets from The Hague, The Netherlands, we conduct this study trying to shed light upon this question. The relation between passenger flow and topological measures in different public transport network representations are investigated in detail. Our preliminary results show promising evidence that the passenger flow is indeed correlated to topological measures in the case study network. Several linear regression models are also constructed to quantify the explanatory power of these topological measures.

Keywords Public transport · Passenger flow · Network topology · Betweenness centrality

1 Introduction

As a result of the interaction between demand and supply, passenger flow is a crucial element in public transport (PT) analysis and modeling. Substantial research effort has been dedicated to this subject in order to facilitate PT planning and operations. Based on the classic four-step modeling paradigm, many have studied origin-destination (O-D) demand estimation and transit assignment models to obtain passenger flow distribution across a public transport network (PTN). Two classes of assignment approaches prevail in the literature, namely the frequency-based and scheduled-based. The former considers the transit system in terms of service-segments and computes the average passenger flow based on service frequency (Nguyen and Pallottino, 1988; Spiess and Florian, 1989; Cepeda et al, 2006), while the latter explicitly represents

Ding Luo, Oded Cats, Hans van Lint
Department of Transport and Planning,
Faculty of Civil Engineering and Geosciences
Delft University of Technology, The Netherlands
E-mail: d.luo@tudelft.nl

individual vehicle trips and their departure and arrival times in the modeling process (Nuzzolo et al, 2001, 2012).

The abovementioned four-step modeling approach provides an intuitive and logical way to model passenger flow, especially in the situation where very limited amount and types of data can be collected. In this process, many assumptions about travelers' behavior have to be made. It, however, has not been fully explored whether more parsimonious approaches with fewer components can also be employed to explain and model the PT passenger flow to a sufficient extent. Thus, it still remains unclear whether passenger flow can be explained, and consequently potentially predicted, using solely the topological characteristics of the network itself. Some related attempts have been reported in the field of urban planning, which try to unravel the relation between urban traffic and the underlying street network topology, such as Borgatti (2005); Gao et al (2013). No consistent or definitive conclusion, nonetheless, has been reached within that community insofar. Performing an equivalent analysis for PT requires embarking on a challenging task given the multiplicity of the underlying network and possible graph representations, in particular the service layer. Moreover, information concerning PT passenger flow at a sufficient spatiotemporal scale is often not directly available.

The objective of this study, therefore, is to empirically investigate to what extent passenger flow distribution can be explained by solely information about PTN topology.

2 Methodology

2.1 Overview

An overview of the methodology is presented in Fig. 1. It consists of three parts: (i) the preparation of raw data (top layer), including automatic fare collection (AFC), automatic vehicle location (AVL) and general transit feed specification (GTFS); (ii) extraction and generation of information, including different graph representations of PTN, computation of topological measures, and passenger flow estimation (middle layer); and (iii) regression analysis (bottom layer). In what follows, further information is provided on the key modules.

2.2 Construction of PTN graph representations

GTFS data is employed as the data source to provide essential information about the underlying PTN, including the infrastructure layout and the service superimposed upon it. Based on the standard data structure of GTFS, the generation of required elements, such as stops, lines, and scheduled timetable, is automatically retrieved using MATLAB scripts.

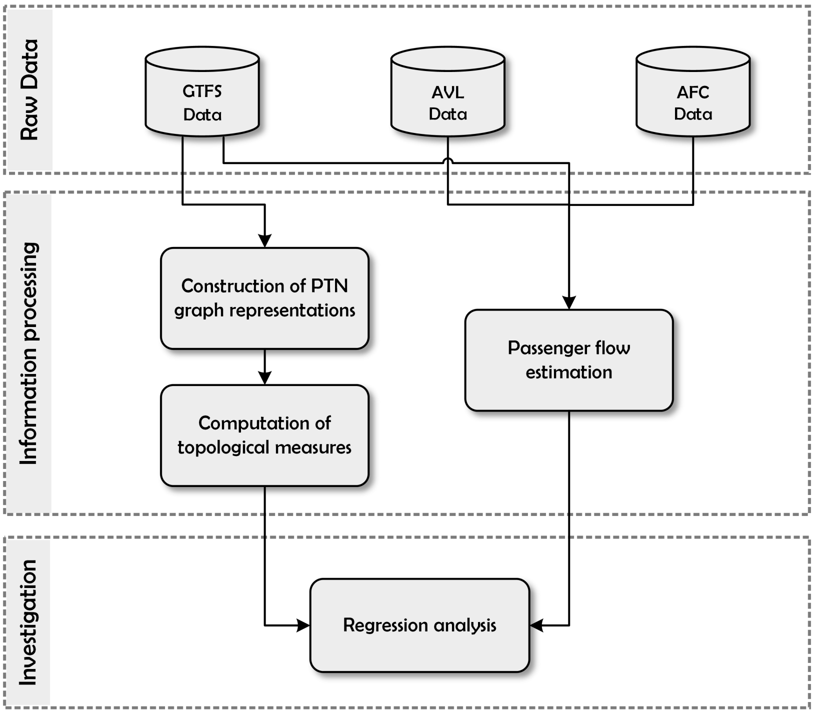


Fig. 1 An overview of the methodology.

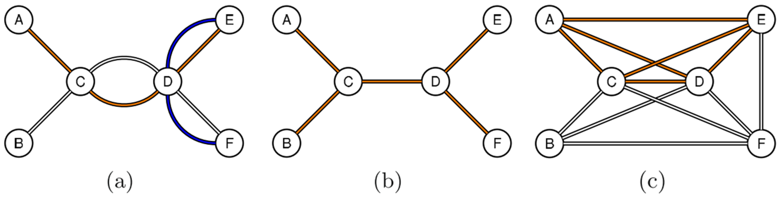


Fig. 2 (a) A simple PTN. Stations AF are serviced by routes No 1 (shaded orange), No 2 (white), and No 3 (dark blue); (b) 𝕃-space graph; (c) ℙ-space graph. Source: von Ferber et al (2009).

The PTN is further represented using two different types of graphs: 𝕃-space and ℙ-space. These two different graphs have been widely utilized to study PTN topological properties, such as von Ferber et al (2009); Yang et al (2014). In the 𝕃-space graph, each stop is represented by a node. Two nodes are connected by a link if these nodes are served successively on at least one service line. Duplication caused by common corridors are not permitted (i.e. one directed link per pair of nodes). In the ℙ-space graph, nodes still correspond to individual stops, whereas two nodes are connected only if they belong to at least one same service line, i.e. it is possible to travel between them using a single line. Simple graphical illustrations of these graphs can be seen in Fig. 2.

In our study, links are further labeled to allow obtaining more meaningful network indicators. In particular, scheduled in-vehicle travel times derived from the GTFS data are added to links in the $\mathbb{L}$-space graph, while planned service between each pair of nodes (joint headway of multiple lines) are added to links in the $\mathbb{P}$-space graph.

2.3 Passenger flow estimation

Both stop flow and links flow can be investigated. We hereby discuss the former which is defined as the total number of passengers who use or traverse stop v during a certain time period τ. Scaled by the total amount of flow in the network during the analysis period, the measure can be further defined as the stop throughput $p_\tau(v)$ as used by Ramli et al (2014). In fact, such a measure of the flow encompasses both the passengers who board or alight at a stop and the passengers who traverse a stop onboard a passing vehicle. It reflects the importance of the stop for flow distribution. This measure will be the dependent variable explained by a series of topological measures.

2.4 Computation of topological measures

A variety of measures has been proposed to investigate network topology. The betweenness centrality, first introduced by Freeman (1977), can potentially explain traffic flow as it represents the global importance of a node in connecting others. The betweenness centrality $b_c(v)$ is defined as follows:

$$b_c(v) = \sum_{s \neq v \neq t} \frac{\sigma_{st}(v)}{\sigma_{st}} \tag{1}$$

where σ_{st} represents the sum of shortest paths between stop s and t, while $\sigma_{st}(v)$ is the total number of shortest paths between stop s and t via the stop v.

The betweenness centrality can be applied to both $\mathbb{L}$-space and $\mathbb{P}$-space graph representations. In addition, in order to reflect the importance of nodes in terms of passenger demand, the betweenness centrality in $\mathbb{L}$-space is further extended by weighing each shortest path selection with the number of O-D trips (Cats and Jenelius, 2014). The extended measure is defined as follows:

$$b'_c(v) = \sum_{s \neq v \neq t} \frac{\sigma_{st}(v)}{\sigma_{st}} \times \frac{N_{st}}{N} \tag{2}$$

where N_{st} represents the total number of trips between origin stop s and destination stop t. N represents the total number of O-D trips within the entire network.

2.5 Regression analysis

The correlation coefficient between the passenger throughput at stops and all computed measures is first examined. Furthermore, linear regression models are constructed with the passenger throughput as the dependent variable, and all topological measures as the independent variables. By doing so, the power of each measure in explaining the passenger throughput is examined. This can shed light on the overall importance of behavioral factors without the consideration of individual behavioral choice determinants.

3 Case Study: The Hague Urban Rail Network

3.1 Network and data

The proposed methodology was demonstrated for the urban rail network of The Hague in The Netherlands (Fig. 3). The network includes a total of 12 tram and light rail lines in operation during the study period, which is the entire month of March 2015. Necessary data sets, including AFC, AVL and GTFS, are all available for this time period. Details about these data sets and related processing work are described in detail in our previous work (Luo et al, 2018).

The final tram network graph contains 247 nodes and 531 directional links. The morning peak period (7 AM - 10 AM) on weekdays was used in this analysis. Illustrations of the two different graph representations, $\mathbb{L}$-space and $\mathbb{P}$-space, can be seen in Fig. 4a and Fig. 4b, respectively. Stop throughput was derived from the resulting spatiotemporal load profiles constructed by Luo et al (2018).

3.2 Preliminary results

All variables, including passenger throughput and node betweenness centrality measures in different graph representations, were computed. They were further scaled based on the sum value (i.e. expressed in percentage-wise terms) for the sake of comparison, which are shown in Fig. 5. It can be seen that the central area of The Hague, where a lot of travelers are attracted and many lines intersect, stands out.

As displayed in Table 1, the stop throughput and all betweenness centrality measures in different graph representations exercise strong correlations. Fig. 6 further depicts the results of linear regression models with the scatter plot included. In particular, the measures from the $\mathbb{L}$-space exhibit strong correlations. Note that the results from unlabeled and labeled $\mathbb{L}$-space graphs turn out to be identical, which indicates that the addition of link in-vehicle travel times does not change the searching process of shortest paths in this

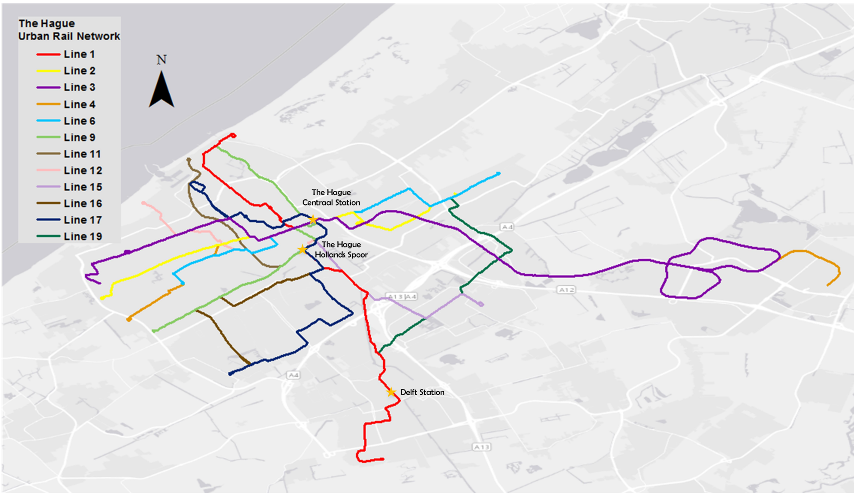
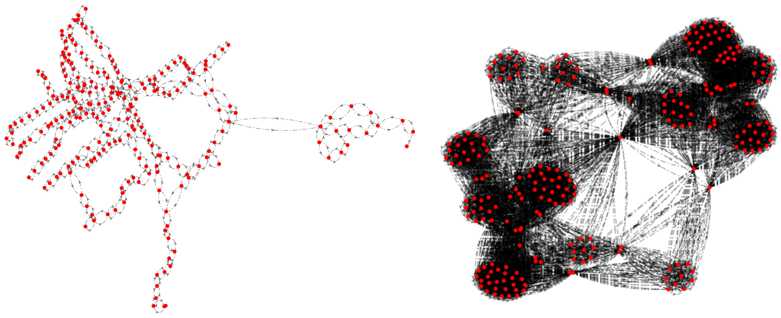


Fig. 3 The Hague urban rail network.



(a) (b)

Fig. 4 (a) $\mathbb{L}$-space representation of the network; (b) $\mathbb{P}$-space representation of the network.

case. Inspecting these plots, linear relation between the betweenness centrality measures and passenger throughput can be visually detected in the case of $\mathbb{L}$-space, but not for $\mathbb{P}$-space which contains information about the line service layer. The reported values of goodness of fit (R^2) are also in line with this observation. Such results are intuitive because of the essence of betweenness centrality in the $\mathbb{L}$-space graph: shortest paths are expected to be taken most frequently by travelers, especially in this relatively simple network. The largest discrepancies occur at the busiest places, such as the central station and station Hollands Spoor (see the yellow marks in Fig. 3) where connections to railway services are also available. Significant additional amount of travelers from/to the railway system might account for these large differences.

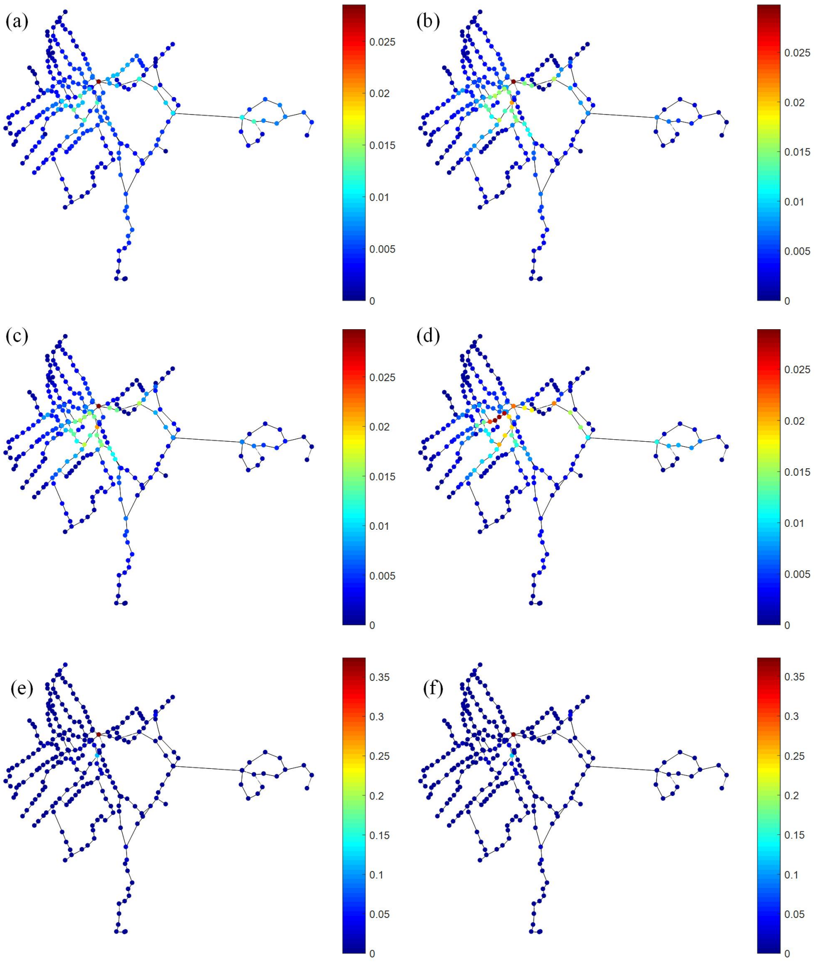


Fig. 5 Geographical distribution of the dependent and independent variables: (a) Stop throughput; (b) Betweenness centrality of the basic $\mathbb{L}$-space network without link label and O-D pair weight; (c) Betweenness centrality of link-labeled (in-vehicle travel time) $\mathbb{L}$-space network without O-D pair weight; (d) Betweenness centrality of link-labeled $\mathbb{L}$-space network with O-D pair weight; (e) Betweenness centrality of the basic $\mathbb{P}$-space network without link label and O-D pair weight; (f) Betweenness centrality of link-labeled (scheduled headway) $\mathbb{P}$-space network without O-D pair weight.

Table 1 Results of correlation coefficients.

Betw.	B^L	B^P	B^L_{tt}	$B^L_{tt,od}$	B^P_{hdwy}
Corr. Coef.	0.71	0.58	0.71	0.67	0.58

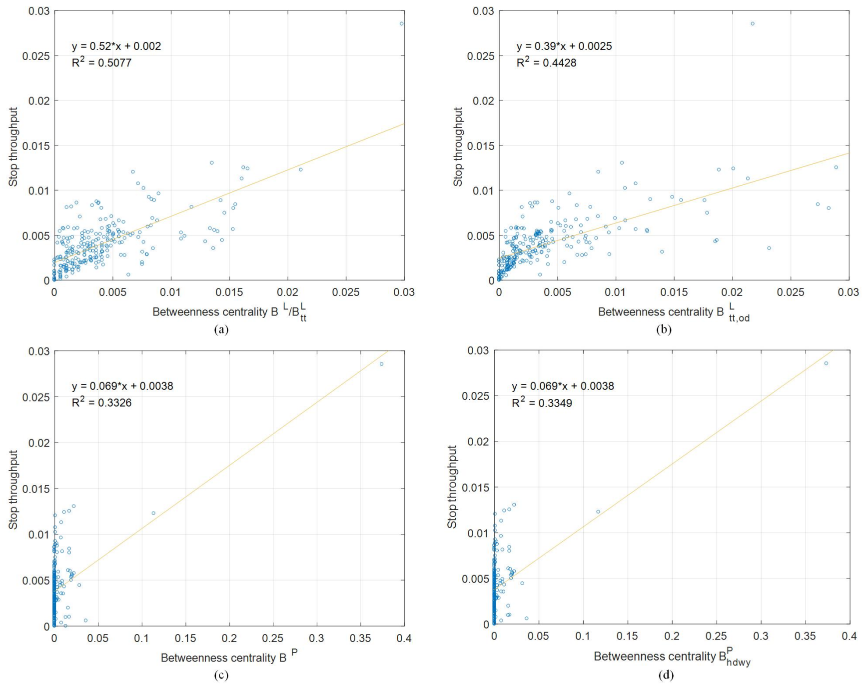


Fig. 6 Correlation between node throughput and betweenness centralities in different graph representation.

Table 2 Results of regression models.

No.	B^L		B^L_{od}		B^P		B^P_{hdwy}		Intercept		R^2
	Coef.	t	Coef.	t	Coef.	t	Coef.	t	Coef.	t	
1	0.52	15.9	-	-	-	-	-	-	0.001	10.42	0.51
2	-	-	0.39	13.86	-	-	-	-	0.002	13.59	0.44
3	-	-	-	-	0.07	11.04	-	-	0.004	23.73	0.33
4	-	-	-	-	-	-	0.07	11.11	0.003	23.77	0.33
5	0.41	11.04	-	-	0.03	5.08	-	-	0.002	11.98	0.55
6	0.41	10.97	-	-	-	-	0.03	5.07	0.002	11.98	0.55
7	-	-	0.3	11.57	-	-	0.05	8.63	0.003	16.32	0.57

Multiple regression models with betweenness centrality measures in both $\mathbb{L}$- and $\mathbb{P}$- spaces considered as explanatory variables are estimated. All results of regression models (simple and multiple) are summarized in Table 2. It can be seen that by combining two different betweenness centrality measures (models 5-7), the explanatory power of the regression model improves.

4 Conclusions and Future Work

This study explores the possibility of explaining passenger flow solely based on network topology in public transport. Owing to the increasing availability and quality of public transport data, outputs of such data-dependent investigation can be obtained. Our preliminary results attest to strong correlations between stop flow and betweenness centralities for different graph representations. Furthermore, maximally 57% of the spatial variations in stop flow can be explained by these betweenness centrality measures through constructed multiple regression models. Ongoing work includes the estimation of alternative regression models and their interpretation, including in relation to findings from behavioral studies. Also, the proposed method will be extended to link flow too. More importantly, the results of the obtained explanatory model will be validated by applying it to data from another network.

Acknowledgements The authors thank HTM and Stichting OpenGeo for providing the AFC and AVL data sets, respectively. We also acknowledge the support of the SETA project funded by the European Union's Horizon 2020 research and innovation program.

References

Borgatti SP (2005) Centrality and network flow. Social Networks 27:55–71

Cats O, Jenelius E (2014) Dynamic Vulnerability Analysis of Public Transport Networks: Mitigation Effects of Real-Time Information. Networks and Spatial Economics 14:435–463

Cepeda M, Cominetti R, Florian M (2006) A frequency-based assignment model for congested transit networks with strict capacity constraints: characterization and computation of equilibria. Transportation Research Part B: Methodological 40(6):437–459

von Ferber C, Holovatch T, Holovatch Y, Palchykov V (2009) Public transport networks: empirical analysis and modeling. The European Physical Journal B 68:261–275

Freeman LC (1977) A Set of Measures of Centrality Based on Betweenness. Sociometry 40(1):35–41

Gao S, Wang Y, Gao Y, Liu Y (2013) Understanding urban traffic-flow characteristics: A rethinking of betweenness centrality. Environment and Planning B: Planning and Design 40:135–153

Luo D, Bonnetain L, Cats O, van Lint H (2018) Constructing Spatiotemporal Load Profiles of Transit Vehicles with Multiple Data Sources. In: Transportation Research Board 97th Annual Meeting, 18-02399

Nguyen S, Pallottino S (1988) Equilibrium traffic assignment for large scale transit networks. European Journal of Operational Research 37(2):176–186

Nuzzolo A, Russo F, Crisalli U (2001) A Doubly Dynamic Schedule-based Assignment Model for Transit Networks. Transportation Science 35(3):268–285

Nuzzolo A, Crisalli U, Rosati L (2012) A schedule-based assignment model with explicit capacity constraints for congested transit networks. Transportation Research Part C: Emerging Technologies 20(1):16–33

Ramli MA, Monterola CP, Khoon GLK, Guang THG (2014) A method to ascertain rapid transit systems' throughput distribution using network analysis. Procedia Computer Science 29:1621–1630

Spiess H, Florian M (1989) Optimal strategies: A new assignment model for transit networks. Transportation Research Part B 23(2):83–102

Yang XH, Chen G, Chen SY, Wang WL, Wang L (2014) Study on some bus transport networks in China with considering spatial characteristics. Transportation Research Part A: Policy and Practice 69:1–10