GIS Based Methodology to Analyse the Public Transport Supply – Hungarian Case Studies

Martin Báráta
Received: February 13, 2022 | Revised: April 13, 2022 | Accepted: May 14, 2022
doi: 10.5937/gp26-36423

Abstract
The paper aims to introduce a new way of comparing the efficiency of public transport operations based on publicly available data. It draws on four main sources, the Hungarian Central Statistical Office, public transport provider data, GTFS and OSM map layers. Methodologically, it combines the use of the GTFS format and corresponding static timetable component files, Thiessen polygons and empirical selection of relative indicators. As places of research, three comparable Hungarian cities have been selected by their population size; Pécs, Szeged and Miskolc. The comparison consists of 8 quantitative indicators that cover six major geographical aspects of transport operation (accessibility in terms of proximity, capacity, connectivity, density, frequency and velocity of vehicles). The analysis does not consider the mode of public transport, thus opening up the possibility of an independent comparison of efficiency regardless of various infrastructure characteristics. The results show that Miskolc and Pécs achieved the best values in four indicators. On the contrary, the city of Szeged, despite its diverse structure of transport modes, does not have an advantage in any aspect. The relatively loosely anchored methodology leaves room for an extension to include economic, environmental, and other specific factors.

Keywords: public transport; GTFS; Thiessen polygons; indicators; accessibility; GIS

Introduction
The analysis of the efficiency of public transport operations in a selected city can be approached from the perspective of many different disciplines. Within the exact disciplines, economic, environmental and geographical (spatial) factors are essential. In terms of the possibilities for inter-city comparison and visualisation of specific differences, the most appropriate tools for analysis appear to be geographical ones, which include not only the basic aspect of the distribution of the selected characteristic in space but also significant interconnectedness with downstream factors. Examples include the issue of transport accessibility, the quality of which depends, among other things, on the stop distance from the origin or destination, the capacity and condition of the vehicles used for transport, or the cost of the transport itself. This fundamental ability of geography to link several otherwise independent sectors makes it possible to create a relatively comprehensive analysis from purely geographical indicators. For the reasons mentioned above, this study thus works with a selection of rather geographic indicators only, leaving open the possibility of a possible extension to other more advanced tools from economics or environmental science.

Geographical factors can be further subdivided into individual spatial aspects of transport infrastructure related to the accessibility of the selected area in general. These include the connectivity of stops by lines of connections, the density of the transport network concerning the area and population of the territorial unit, the frequency of connections at stops, the capacity of transport vehicles, the average oper-
The detailed structure of operation time of vehicles deals (Matulin et al., 2011). (Birr et al., 2014) extends the topic to predictive models of vehicle arrival time at a stop. Accessibility in the sense of proximity is primarily addressed by GIS-oriented methodologies using network analyses in (Háznagy et al., 2015) and (Luo et al., 2019). (Mavoa et al., 2012) uses both proximity and frequency aspects to assess accessibility. A comprehensive geographical view of a transport system's efficiency, quality, and accessibility and its sub-elements always represent a specific selection of the indicators or functions used. By its very subjective nature, the nature of the area of interest does not have the ambition to include everything that interferes with the system's functioning.

Nevertheless, more and more works deal with the complex synthetic concepts of the efficiency of urban public transport. Their approach to processing differs in many respects. Typical examples are quantitative vs qualitative studies, hard or soft data methodologies, comparative or case studies, etc. Another critical factor is the purpose of the study, which determines whether it will be more theoretical or practical. (Bajčetić et al., 2018) and (Ušpalytė-Vitkūnienė et al., 2020) work in their methodology with the perception of the user and the direct participant of the traffic, while (Zhou et al., 2021) links transport quality assessment with educational activities. The theoretical framework and the complexity of defining adequate indicators are described by (Išoraite, 2005). The strategic use of comprehensive urban transport assessment for the potential development of an area is found in the works of (Gaal et al., 2015) and (Hawas et al., 2016). More and more studies can be found concerning large urban municipal or even regional units and their comprehensive comparisons with the GIS application.

In the case of this paper, we can talk rather about a new combination, though belonging closest to the last type. The objectives of this thesis can be summarised in several basic points, conditions and the resulting research questions:

1. to develop a comparable system for assessing the geographical efficiency of public transport operations based on selected indicators
2. applying only publicly available data sources to maximise the transferability of the methodology to any city
3. comparison of three case studies, Hungarian cities (Szeged, Miskolc and Pécs) according to the resulting values of the indicators

The research questions are then mainly related to the comparison itself. According to geographical factors, which of the three examined cities shows the highest vs lowest public transport efficiency? Which geographical factors most significantly influence the resulting overall efficiency rating? Based on an essentially arbitrary selection of indicators, can relevant results be obtained to reveal strengths and weaknesses and possible optimisation?

**Study area**

Regarding the availability of public transport data within Hungary and comparable population size, three cities, Szeged, Miskolc and Pécs, have been selected for comparison. According to the Hungarian Statistical Office, their population on 1 January 2021 was 159 074 (Szeged), 150 695 (Miskolc) and 140 237 in the case of Pécs (KSH, 2021). They occupy the third, fourth and fifth positions in the ranking of the largest cities in Hungary. The relative position within Hungary is shown in Figure 1.

In terms of public transport, cities vary considerably in transport modes and main types of transport. In Pécs, only the bus network is currently in operation. In Miskolc, there is the bus network and, to a small extent, the tram network, and finally, in Szeged, buses, trolleybuses and trams are used. The choice of these three cit-
ies for the analysis thus reflects the three most common combinations of public transport modes. An overview of the basic attributes of the selected cities and their public transport is given in Table 1. All of these characteristics are included in more detail in the final analysis.

In addition to knowledge of population, area and transport network, it is also necessary to describe the character of cities in terms of the distribution of the major land cover classes. Looking at Figure 2, it is clear that the diverse composition of the landscape significantly differentiates the conditions for the functioning and accessibility of transport infrastructure. While agricultural areas dominate Szeged, Pécs and Miskolc are mainly forests and semi-natural areas. Urban artificial surfaces then form the main population centres and thus the densest concentration of the transport network. Of course, the need for transport links is not only associated with urban development but occurs in almost any land cover class where there is some form of housing, employment offer or tourist sites. Thus, in Miskolc, for example, urban transport extends relatively far beyond the boundaries of the typical urban areas into the centre of Bük National Park, or in Pécs, a network of lines connects the peripheral sites of industrial extraction and settlement there with the city centre. In Szeged, on the other hand, there is virtually no area outside the main centre with adjacent neighbourhoods where either accommodation, work, or tourist functions meet. That may consequently lead to a large area without transport services, although it administratively falls within the city borders.

Table 1. Basic characteristics of public transport in Szeged, Miskolc and Pécs by 2021

<table>
<thead>
<tr>
<th>City</th>
<th>Population</th>
<th>Area (km²)</th>
<th>Length of transport network (km)</th>
<th>Number of stops</th>
<th>Number of routes (lines)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Szeged</td>
<td>159 074</td>
<td>281.00</td>
<td>132.2</td>
<td>290</td>
<td>48</td>
</tr>
<tr>
<td>Miskolc</td>
<td>150 695</td>
<td>236.67</td>
<td>164.7</td>
<td>263</td>
<td>46</td>
</tr>
<tr>
<td>Pécs</td>
<td>140 237</td>
<td>162.78</td>
<td>153.4</td>
<td>265</td>
<td>75</td>
</tr>
</tbody>
</table>

Source: own elaboration of data from (KSH, 2021), (Transitfeeds, 2021)
Methods of analysis and GTFS

The methodology is based on four broad types of publicly available sources. The basic one is the statistical office, which provides input data for analysing the actual population and the area of the administratively defined units. The core part of the data for most aspects of accessibility is obtained from the GTFS database of files on Hungarian cities (Transitfeeds, 2021). Each city contains data from a different provider, Szeged (DAKK, 2021), Miskolc (MVKZRT, 2021), Pécs (Biokom, 2021).

A GTFS timetable is a set of text (CSV) files packaged in a ZIP archive. It is a format for timetables and related geographical information. In addition to static information, it also supports the feeding of real-time data from public transport operations. The files must be encoded in UTF-8. The static component consists of five required files defining agency, stops, routes, trips and stop times. Two filenames describing calendar and calendar dates are conditionally required. The other ten optional files may provide, e.g. information on frequencies, continuity or fares (GTFS, 2021). The GTFS format significantly facilitates the processing of detailed timetable information. Due to regular updates and archiving of old data, extensive analyses and comparisons of urban transport operations can be performed. It is thus the cornerstone of many studies, such as (Bok & Kwan, 2016), (Prommaharaj et al., 2020). However, the nature of static timetables may not always correspond to actual travel times, as the creation of the timetable cannot consider all relevant factors that randomly or even regularly influence (ex. congestion) the performance of the transport operation. (Wessel et al., 2017) examines the differences between scheduled and observed services. A balanced critique and detailed explanation of the individual components of GTFS is further presented by (Kujala et al., 2018).

The primary data from the transport companies themselves are then used as a source for vehicle capacities aspect and potential extension of the limited source of GTFS files. The transport providers in these cases always manage the entire public transport network, Szeged (SZKT, 2021), Miskolc (MVKZRT, 2021), Pécs (Tüke Busz, 2021). Compared to the previous types, this is relatively the least reliable source because of the different approaches to managing and operating the published data. Map visualisation and spatial distribution of the population and traffic data collected are essential parts of the analysis. For this purpose, the universal map source OSM (OpenStreetMap) is used. All data used are based on the period before the end of July 2021.
A system of 8 indicators was then compiled that shows both the absolute and relative values of 6 major geographical aspects by using these four primary sources. It covers accessibility in the meaning of proximity, connectivity, density, frequency, occupancy (capacity) and velocity (speed). The complete list of examined indicators shows in Table 2. The indicator formulas have been empirically designed and arithmetically adjusted to range their values from 0 to 200. Further attenuation of the incommensurability of different results is achieved by standardising the modified Bennet method. It is the sum of the relative deviations (percentage distances) from mean values per indicator. Except for indicators 3 and 5, all others are based on a combination of GTFS and public transport companies’ data. The critical aspect of accessibility (proximity) is represented by indicator 5, which attempts to measure differences in the size of un-served

Table 2. List of 8 main indicators used for the analysis

<table>
<thead>
<tr>
<th>Num.</th>
<th>Indicator</th>
<th>Min value</th>
<th>Max value</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Average number of connections at a stop during one working day per number of inhabitants multiplied by 100 000</td>
<td>0</td>
<td>–</td>
<td>A, F</td>
</tr>
<tr>
<td>2</td>
<td>Average number of connections at a stop during one Sat., Sun., or public holiday per number of inhabitants multiplied by 100 000</td>
<td>0</td>
<td>–</td>
<td>A, F</td>
</tr>
<tr>
<td>3</td>
<td>Density of public transport network by number of inhabitants and area of the municipality</td>
<td>0</td>
<td>–</td>
<td>A, D</td>
</tr>
<tr>
<td>4</td>
<td>Weighted mean of total capacity (occupancy) of one vehicle</td>
<td>0</td>
<td>–</td>
<td>A, O</td>
</tr>
<tr>
<td>5</td>
<td>Percentage difference of the 10th percentil of the largest Thiessen polygon areas by total area of the city</td>
<td>0</td>
<td>100</td>
<td>A</td>
</tr>
<tr>
<td>6</td>
<td>Number of lines per number of stops multiplied by 100</td>
<td>0</td>
<td>–</td>
<td>A, C</td>
</tr>
<tr>
<td>7</td>
<td>Average number of connections (WD + HOL) per lines</td>
<td>0</td>
<td>–</td>
<td>A, C, F</td>
</tr>
<tr>
<td>8</td>
<td>Average velocity of public transport according to operated lines (km/h)</td>
<td>0</td>
<td>–</td>
<td>A, V</td>
</tr>
</tbody>
</table>

(A = accessibility, C = connectivity, D = density, F = frequency, O = occupancy (capacity), V = velocity)

Figure 3. Cities partitioning using Thiessen polygons for set of stops

Source: own elaboration of data from (OSM, 2021), (Transitfeeds, 2021)
areas with the Thiessen polygon method. Its range of values is governed by the percentage difference, i.e. from 0 to 100.

The method of Thiessen polygons (also Voronoi diagram) works on the principle of partitioning the space according to the distances between a set of given points so that each point falls into a polygon whose boundaries are always formed by half the distance between each point. The design of this method is discussed in detail and practically used for transport analysis (Wang et al., 2014). It has also been applied for multicriteria analysis of bus and tram stops (Bárta & Masopust, 2020). An illustration of what such partitioning looks like in the example of all examined cities is shown in Figure 2. Here, two basic factors come into play: the density of stops and their distribution within the city limits. More regularly located stops mean fewer above-average polygons, with the optimal state of accessibility corresponding to the same area per polygon. However, it is important to note that another equally important aspect is the degree of arbitrariness in the delimitation of the city borders concerning the nature of the settlement and the need of citizens.

Results

The calculations of the individual indicators provided a clear picture of the distribution of values for the three selected cities. A concise statistic of this distribution of values gives Table 3. The capacity aspect of the average vehicle most closely describes the difference in using different combinations of transport modes. The relatively most significant deviation from the average is associated with synthetic indicator number 7. The explanation lies in its dual meaning. The higher the connectivity in the form of more lines, the fewer connections per line, which means it is impossible to have above-average connection frequency and connectivity for a better relative result. The aspect of an average speed of service between two stops included in the last indicator deviates only minimally from the average due to the similar attributes of the transport infrastructure and the size of the cities in question. The map visualizations of the absolute values of the number of connections at stops within Figure 3 extend the first two frequency indicators and the third network density indicator with a more detailed spatial distribution and associated localization of the disparities between the core and peripheral parts of the cities.

The intervals for the number of connections have been kept the same except for the upper limit of the highest. The busiest stops occur in the centre of Pécs, although, unlike the other two cities, they are served only by bus lines. The urban design and its effect on the layout of the transport network and the arbitrariness of the administrative demarcation and density also significantly influence the Thiessen polygon method results. The city of Szeged, which is designed compactly with a radial street network, a smaller distance of peripheries from the centre and a large area, suffers from this effect to a considerable negative extent. That also corresponds to the lowest population density compared to Miskolc and Pécs.

Specific differences in absolute values by indicator can be seen in Figure 4, which also refines the range of values from Table 3. A more convenient idea of comparing indicators between cities is given by the relative format in Figure 5. Working with a modified Bennet method based on percentage distances from the mean value. According to this method, it is apparent to see which cities perform below or above average, considering the range of a given indicator. The city of Pécs,

<table>
<thead>
<tr>
<th>Num.</th>
<th>Indicator</th>
<th>Min value</th>
<th>Mean value</th>
<th>Max value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Average num. of connec. at a stop during one working day per number of inhab. multipl.by 100 000</td>
<td>120,15</td>
<td>139,12</td>
<td>150,29</td>
</tr>
<tr>
<td>2</td>
<td>Average num. of connec. at a stop during one Sat., Sun., or public holiday per num. of inhab. multipl. by 100 000</td>
<td>81,39</td>
<td>88,25</td>
<td>99,50</td>
</tr>
<tr>
<td>3</td>
<td>Density of public transport network by number of inhabitants and area of the municipality</td>
<td>19,78</td>
<td>26,49</td>
<td>32,11</td>
</tr>
<tr>
<td>4</td>
<td>Weighted mean of total capacity (occupancy) of one vehicle</td>
<td>102,82</td>
<td>121,18</td>
<td>146,11</td>
</tr>
<tr>
<td>5</td>
<td>Percentage difference of the 10th percentil of the largest Thiessen polygon areas by total area of the city</td>
<td>26,61</td>
<td>41,66</td>
<td>53,46</td>
</tr>
<tr>
<td>6</td>
<td>Number of lines per number of stops multiplied by 100</td>
<td>16,55</td>
<td>20,78</td>
<td>28,30</td>
</tr>
<tr>
<td>7</td>
<td>Average number of connections (WD + HOL) per lines</td>
<td>18,91</td>
<td>45,91</td>
<td>66,91</td>
</tr>
<tr>
<td>8</td>
<td>Average velocity of public transport according to operated lines (km/h)</td>
<td>23,34</td>
<td>23,79</td>
<td>24,59</td>
</tr>
</tbody>
</table>
in this comparison, achieved the absolute most significant negative deviation within the aspect of connectivity and frequency. See the explanation above. Nevertheless, overall it significantly outperforms Szeged. Finally, for Miskolc, the sum of the difference between positive and negative deviations results in the highest values on average and the position of the most efficient public transport operation.

Figure 4. Number of connections (trips) by public transport within examined cities in 2021
*Source: own elaboration of data from (OSM, 2021), (Transitfeeds, 2021)*

Figure 5. Results of 8 indicators in absolute values
Discussion and conclusion

Evaluating the efficiency of traffic draws on a selection of quantitative relative indicators means making an inevitable compromise between the availability of reliable data, including as many influencing factors as possible and the objective setting of weights for individual indicators. This study is based on a combination of only hard, verifiable data, a relatively small number of indicators using only geographical aspects, and a relatively objective level of assessment of the importance of each indicator for inter-comparison. The incomplete listing of significant factors and their very general inclusion in the formulas can be considered a significant shortcoming in the choice of indicators. A certain simplification of an otherwise complex issue has been applied for several quite different reasons. For any comparison, it is necessary to obtain sufficient data to allow two or more selected examples to be evaluated in the same way. In the context of transport analysis, the data source issue is complicated by the dependence on publicly available materials or the need to collect data from own observations. However, far from all necessary aspects can be obtained from public databases or own field research. Therefore, a certain narrowing of perspective was inevitable. The reason for the overall triviality of the formulas can be explained by the desire for a more accessible methodology to be extended to include economic or environmental factors, which could be loosely linked to an interdisciplinary geographical angle. The objective assessment of the weight of a given indicator and the subsequent results was partly guaranteed by using the modified Bennet method, a form of standardisation that accounts for the percentage distance from the mean value. Of course, the chosen method does not, even so, provide absolute independence from the inherently subjective selection of indicators.

The methodology is, to some extent, intertwined with (Bárta, 2020) and (Bárta & Masopust, 2020). However, unlike the previous ones, it attempts inter-city comparisons while incorporating more precisely defined geographical aspects. The GTFS format as a systematic way of accessing traffic data appears in an increasing number of studies. More advanced command formation and the most efficient use of static component files have been inspired by (Bok & Kwan, 2016) and (Kujala et al., 2018) in particular.

In terms of methodology and results, three main points have been met in line with the objectives, so we can answer the research questions. The analysis shows that Miskolc has the most efficient public transport operation in absolute and relative data. Due to the higher diversity of transport modes, Szeged surprisingly comes out worst in almost all indicators. However, the difference in efficiency between the cities is not significant enough to draw critical conclusions. The more complex question of the relevance of specific factors for the analysis results can be viewed from several angles. Because of the arbitrariness of the chosen indicators, it is necessary to choose the most unbiased guide. For this purpose, the modified Bennet method of percentage distances from the mean value of each indicator was used. According to this meth-
GIS Based Methodology to Analyse the Public Transport Supply – Hungarian Case Studies

od, the combination of accessibility and connectivity in the form of the average number of connections per number of lines clearly stands out, the extent of which most influences the final result of the evaluation. Beyond purely numerical differences, however, the significance of some indicators, particularly the accessibility derived from Thiessen polygons, is closely tied to the nature of the administrative delimitation. Therefore, cities whose borders do not correspond to the settlement pattern with the population distribution may exhibit significantly lower efficiency, even though they do not stand out in other areas. The answer to the key question of the relevance and applicability of the arbitrary indicator methodology to specific disparities and optimisation options is broadly positive, given the possibility of focusing on the most detailed urban districts and consistent comparability with the corresponding average for the higher unit. Although this paper does not directly examine the detailed structure of the selected cities, its methodology based on a detailed GTFS timetable format, combined with possible network analysis, offers the possibility of an accurate analysis of gaps in spatial and time accessibility that is crucial for any optimization. Yet, it is essential to note that the selection of geographical factors alone without economic, environmental and other factors does not provide a comprehensive view of the issue of evaluating a complex public transport system.

In conclusion, this paper can be classified as another attempt at a geographical comparative evaluation of selected factors interacting with the daily operation of urban public transport. Despite a specific generalisation and simplification of a complicated issue, it can serve as a tool for future analysis of strengths and weaknesses and possible extension for economic and environmental factors of public transport operations.

Acknowledgements

This study was supported by grant IGA_PrF_2021_027 „Landscape as a palimpsest of space and places”

References


