Assessing public transport systems connectivity based on Google Transit data

Yuval Hadas*

The Department of Management, Bar-Ilan University, Ramat Gan 5290002, Israel

A R T I C L E   I N F O

Keywords:
Connectivity
Network analysis
Public transport
Spatial analysis

A B S T R A C T

A PT system consists of various physical features such as roads, railways, routes, and stops which are represented by a complex network of spatial and temporal data. Since these networks are usually very large and include millions of entities, it is difficult to assess PT systems. Assessment in this context is defined as the ability to extract and analyze data in an automated and recurring process so as to enhance decision making and to make it possible to compare between PT networks over time. The unified methodology that this work presents for extracting, storing and analyzing PT data enables relatively easy spatial analysis with GIS techniques based solely on: (a) Google Transit feeds and (b) Transportation networks. In order to implement this new methodology for analyzing a PT system, five connectivity indicators are introduced: (a) transportation network coverage level; (b) average speed; (c) intersection coverage level; (d) stop transfer potential; and (e) route overlap. This work demonstrates the proposed methodology by analyzing PT systems in Auckland (New Zealand), Vancouver (Canada), and Portland (Oregon, USA).

1. Introduction

A PT system consists of various physical features such as roads, railways, routes, and stops which are represented by a complex network of spatial and temporal data (Ceder, 2007; Vuchic, 2005). Since these networks are usually very large and include millions of entities, it is difficult to assess PT systems. Assessment in this context is defined as the ability to extract and analyze data in an automated and recurring process so as to enhance decision making and to make it possible to compare between PT networks over time. The large number of different PT software systems and data sources makes extraction and analysis time consuming tasks that require a dedicated application for each data extraction and analysis.

Most assessment models rely heavily on multiple sources of data and consequently extensive efforts are required in order to obtain the data. In this paper, we present a framework for assessing PT networks based on the minimal data required, namely, PT and the underlying infrastructures (road network, rail network, ferry, seaways, etc.).

The Transit Capacity and Quality of Service Manual (Kittelson and Associates et al., 2003) describes four availability factors of PT systems: (i) spatial – where the service is provided, (ii) temporal – when the service is provided, (iii) information – how to use the service, and (iv) capacity – space available for the passenger. These factors delineate the scope of any PT system analysis. The first two are particularly important in assessing PT connectivity since the connectivity within a PT system is both spatial (routes coverage, stops locations, transfer availability, etc.), and temporal (waiting time, travel time, transfer time, etc.).

This work presents a unified methodology for extracting, storing and analyzing PT data as derived from these four availability factors. Our approach enables relatively easy spatial and temporal analysis with geographic information (GIS) techniques that use the topological, geometric, or geographic properties that characterize an entity (Hadas and Ceder, 2010; Thill, 2000). Our approach is based solely on Google Transit feeds and any available infrastructure layers with no need for additional data.

In 2006, Google introduced Google Transit, a supplemental service to Google maps. This service enabled users to plan public transport trips from origin to destination. In order to easily implement the service and encourage as many agencies as possible to participate, Google established a unified specification, General Transit Feed Specification (Google Transit, 2010), or GTFS (previously called Google Transit Feed Specification). This specification, which enables public transport providers to upload relevant information to the Web, allows users to plan trips from any web-browser. The availability of such data provides researchers with the opportunity to conduct PT analysis with relative ease. Road network layers are also relatively easy to acquire, whether commercially from NAVTEQ (NAVTEQ, 2012), or freely from OpenStreetMap project (OpenStreetMap, 2012). Other modal networks can also be acquired from online sources, such as the US...
National Transportation Atlas Database (The Bureau of Transportation Statistics, 2013), or by digitizing the networks from published maps.

Many connectivity measures for general networks exist, such as: connectivity and strong connectivity of graphs (Ahuja et al., 1993); the cyclomatic number, which is essentially a measure of the number of circuits in a graph; the alpha index, which is the ratio between the number of existing circuits and the maximum of circuits possible; and other similar measures (Black, 2003; Rodriguez et al., 2006). For transportation networks, it is also possible to define accessibility measures such as the longest shortest path of a network (which is the longest distance possibly traveled among all shortest paths in a network). There is also the degree of a node, which can take the form of the number of arcs connected or the form of the sum of shortest paths to all other nodes (Black, 2003). Another possible measure is the ratio between the network-based shortest path and the direct line between node pairs. An extensive overview of connectivity measures can be found in Mishra et al. (2012).

There are other definitions and models for public transport networks which are far more complex than these measures. In these more complex models, arcs represent roads and routes while nodes represent intersections and stops. Thus such measures account for the routes as well as timetables, access, transfer, etc. Vuchic (2005) presented a set of measures such as transfer permutation (for routes sharing the same stop or station); network complexity (ratio of arcs and nodes); line overlapping; directness of service, etc.. Hadas and Ceder (2010) and Hadas and Ranjitkar (2012) developed connectivity measures that integrate demand forecast and transformation of arcs and nodes; line overlapping; directness of service, etc. Figueiredo et al. (2012). Among these models are Wu and Hine’s model for measuring changes in bus service accessibility (Wu and Hine, 2003); a model developed by Currie (2010) for quantifying spatial gaps in PT supply based on social needs (Currie, 2010); connectivity measures for multi-model PT networks (Mishra et al., 2012); and a method to define PT opportunity space (Mamun et al., 2013).

Based on the literature review it is evident that: (1) It is difficult to acquire the data these models require, as the data are from different sources, usually local, and require tailored extraction tools. (2) It is difficult to use them to perform benchmarks since all of the data must share the same characteristics (scale, resolution, context, etc.). These findings support the aim of this work – the use of a single source of PT data – the general transit feed specification.

This work is organized as follows. Section 2 presents the principles of data acquisition and construction; Section 3 provides detailed formulation of connectivity indicators and is followed by a case study in Section 4. Conclusions are presented in Section 5.

2. Data acquisition and construction

Calculating the connectivity indicators requires constructing a database with data from such elements as the transport network, operational transport information, and the stops where vehicles pick up or drop off passengers. Fig. 1 presents a framework composed of the following steps.

(a) Add and connect stops to the network: create a PT stops layer and connect it to the transportation layer. A stop in this context refers to any PT boarding and alighting location.
(b) Build routes based on a sequence of stops: Google Transit does not include route layers, but rather provides the sequence of stops visited by each route. It is necessary to reconstruct these routes as a GIS layer. This is done by locating the shortest path between all adjacent stops of a route.
(c) Calculate headways: headways and frequencies are calculated based on the scheduled departures associated with each route.

![Fig. 1. General modeling framework.](image-url)
Network analysis: in this step the stored spatial and temporal data is used to calculate the different indicators for assessing PT system connectivity.

2.1. Data sources

The model derives information from two sources: (a) Google Transit as a source for public transport network attributes, and (b) the GIS-based road network.

2.1.1. Google Transit data

In 2006, Google introduced to Google Maps an additional feature, Google Transit. This service enables users to plan public transport trips from origin to destination. For easily implementing the service and encouraging more agencies to participate, Google established a unified specification called the General Transit Feed Specification (GTFS). [Prior to 2010, it was referred to as the Google Transit Feed Specification]. This specification enables public transport providers to upload relevant information to the Web, allowing users to plan trips from any web-browser. The availability of such data provides researchers with the opportunity to conduct public transport analysis, with relative ease of access. The GTFS includes the data summarized in Table 1.

2.1.2. Transport network data

A transport network in a GIS format is required for constructing public transport routes, the spatial representation of the network, and spatial analysis. [Freeware of reliable road networks, such as CloudMade (CloudMade.com, 2010), is available.]

2.2. Public transport network construction

The raw data must be processed before the connectivity indicators can be calculated. The first step is to convert the public transport data into a spatial network with routes and stops. The second step is the calculation of headways and frequencies for each route.

2.2.1. Public transport network creation

The GIS road layer is represented by the centerline of the actual transport network. However, the location of each stop or station is

---

Table 1
The general transit feed specification data tables (Google Transit 2010).

<table>
<thead>
<tr>
<th>Data table</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Agency</td>
<td>This file contains information about one or more transit agencies that provide the data in this feed</td>
</tr>
<tr>
<td>Stops</td>
<td>This file contains information about individual locations where vehicles pick up or drop off passengers</td>
</tr>
<tr>
<td>Routes</td>
<td>This file contains information about a transit organization’s routes. A route is a group of trips that are displayed to riders as a single service</td>
</tr>
<tr>
<td>Trips</td>
<td>This file lists all trips and their routes. A trip is a sequence of two or more stops that occurs at specific time</td>
</tr>
<tr>
<td>Stop times</td>
<td>This file lists the times that a vehicle arrives at and departs from individual stops for each trip</td>
</tr>
<tr>
<td>Calendar</td>
<td>This file defines dates for service IDs using a weekly schedule. Specify when service starts and ends, as well as days of the week where service is available</td>
</tr>
<tr>
<td>Calendar dates</td>
<td>This file lists exceptions for the service IDs defined in the calendar.txt file. If calendar_dates.txt includes ALL dates of service, this file may be specified instead of calendar.txt</td>
</tr>
<tr>
<td>Fare attributes</td>
<td>This file defines fare information for a transit organization’s routes</td>
</tr>
<tr>
<td>Fare rules</td>
<td>This file defines the rules for applying fare information for a transit organization’s routes</td>
</tr>
<tr>
<td>Shapes</td>
<td>This file defines the rules for drawing lines on a map to represent a transit organization’s routes</td>
</tr>
<tr>
<td>Frequencies</td>
<td>This file defines the headway (time between trips) for routes with variable frequency of service</td>
</tr>
<tr>
<td>Transfers</td>
<td>This file defines the rules for making connections at transfer points between routes</td>
</tr>
</tbody>
</table>

Extracted from the specification document.

---

Fig. 2. Stop connectors to the transport network centerline.
usually at the sidewalk, curb, platform, or pier. Relocating the stop on the centerline can simplify the network and the calculations, but prohibit the realization of the actual distance between stops, which is essential for assessing transfer potential. Thus the model incorporates a connector (a small distance that can be also illustrated on the map) from each stop to the transport centerline. This is illustrated in Fig. 2. This illustration also demonstrates the incorporation of all PT modes within the network.

The GTFS does not include a spatial representation of the routes corresponding to the transportation layer. Based on each route’s sequence of stops, it is possible to create such a representation by taking into account the transport network characteristics and the shortest path between each two adjacent stops.

### 2.3. Data quality check

During the network’s construction, it is important to perform quality check of the transport network (direction, restricted turns, network connectivity, etc.), and the GTFS data (stops’ location, routes layout, and time-tables). The spatial accuracy can be carried by selecting a route set and comparing the layout with the officially advertised routes. As for the GTFS data, it is officially published first and foremost for trip planning by PT passengers, hence it is likely to be accurate. Nevertheless, a sample of time-tables can be easily compared with the operators or authorities’ official web-sites.

### 3. Connectivity indicators

The PT network, like other networks, is based on nodes and arcs. The transport layer is composed of road sections, railways, etc. (arcs) and intersections (nodes), while the route layer is composed of route segments (arcs) and stops (nodes). Accordingly, five connectivity indicators are introduced for measuring four attributes:

(a) transport network coverage and accessibility level indicators measure the transport network arcs (flow and speed);
(b) an intersection coverage level indicator measures the transport network nodes;
(c) a stop transfer potential indicator measures the route nodes; and
(d) a route-overlap indicator measures the route arcs.

Fig. 3 illustrates these measures.

#### 3.1. Transport network coverage and accessibility level indicators

Based on the PT system’s timetables, it is possible to calculate the flow and the average speed of PT vehicles along a transport network section.

##### 3.1.1. Transport network coverage level indicator

This indicator, which captures the flow of PT vehicles along a transport section, is represented by the sum of frequencies

\[
RC_{(i,j)} = \sum_{r \in A(r)} f_r
\]

where \((i,j)\) is an arc in a network \(G\); \(A(r)\) is the set of the arcs that route \(r\) traverses; and \(f_r\) is the frequency of route \(r\).

The following example illustrates the network coverage indicator by time of day. Thus, the coverage can change throughout the day since it is represented by the vehicle flow (total PT vehicles serving during a time interval), as Fig. 4 illustrates. The bold line...
width represents the hourly PT vehicle flow (intensity), and it is evident that the flow (hence the coverage) between 08:00 and 10:00 is higher than the flow between 20:00 and 22:00 in Auckland CBD.

3.1.2. Transport network speed indicator

Given that the PT system’s timetables reflect traffic conditions, it is possible to estimate PT vehicle speed throughout the network. Based on the arrival and departure times $T_{r,x}^{s}$ at stop $s$, and route $r$, of a trip $x$, it is possible to calculate the average speed between consecutive stops $s_1,s_2$ and the distance $l_{s_1,s_2}$

$$S_{s_1,s_2} = \frac{1}{|T|} \sum_{(s_1,s_2) \in \text{network}} \frac{1}{X} \sum_{x=1}^{X} (T_{r,x}^{s_2} - T_{r,x}^{s_1}) / l_{s_1,s_2}$$  \hspace{1cm} (2)

And $X$ is the number of trips occurred within a specified time window.

Hence, the average speed along a network section $(i,j)$ which overlaps with $(s_1,s_2)$ is:

$$RS_{i,j} = \sum_{\text{all } s_1,s_2} \sum_{o_{(i,j)}/o_{s_1,s_2} > 0} S_{s_1,s_2}$$  \hspace{1cm} (3)

Fig. 6. Auckland’s PT network.

Fig. 7. Portland’s PT network.
3.2. Intersection coverage level indicator

This indicator provides the same information as the transport network coverage level indicator, but emphasizes the flow at the intersection level, thus enabling the node-level assessment of the network (setting PT priorities at intersection, for instance).

\[ IC_n = \sum_{r \in N(r)} f_r \]  \hspace{1cm} (4)

where \( n \) is a node in a network \( G \), and \( N(r) \) is a set of nodes traversed by route \( r \).

3.3. Stop-transfer potential indicator

Transfers between routes are a common practice in modern PT networks, even though they detract from the convenience and
smoothness of trips. Since a passenger is apt to associate a high service level with ease of transfer, assessing the transfer potential of a PT system is crucial. Since transfers are attributes of both space and time (Hadas and Ranjitkar, 2012), the assessment of transfer potential is based on the possible departures within a specified time window and a walking distance, as illustrated in Fig. 3(c). The following equations formally define the potential.

\[
X_{x} = \sum_{s_{1} \in S(r_{1})} \sum_{s_{2} \in S(r_{2})} \sum_{r_{1} \in S(r_{1})} \sum_{r_{2} \in S(r_{2})} \sum_{x_{s_{1}}} \sum_{x_{s_{2}}} \sum_{x_{r_{1}}} \sum_{x_{r_{2}}} \sum_{x_{x_{s_{1}}}} \sum_{x_{x_{s_{2}}}} \sum_{x_{x_{r_{1}}}} \sum_{x_{x_{r_{2}}}} \sum_{x_{x_{x_{s_{1}}}}} \sum_{x_{x_{x_{s_{2}}}}} \sum_{x_{x_{x_{r_{1}}}}} \sum_{x_{x_{x_{r_{2}}}}} \sum_{x_{x_{x_{x_{s_{1}}}}}} \sum_{x_{x_{x_{x_{s_{2}}}}}} \sum_{x_{x_{x_{x_{r_{1}}}}}} \sum_{x_{x_{x_{x_{r_{2}}}}}} \sum_{x_{x_{x_{x_{x_{s_{1}}}}}}} \sum_{x_{x_{x_{x_{x_{s_{2}}}}}}} \sum_{x_{x_{x_{x_{x_{r_{1}}}}}}} \sum_{x_{x_{x_{x_{x_{r_{2}}}}}}}
\]

where \( r_{1}, r_{2} \) are two routes, \( x \) a trip of route \( r_{1} \); \( s_{1}, s_{2} \) are two stops; \( S(r) \) is the set of stops for route \( r \); \( T_{x}^{r, s} \) is the arrival or departure time of route \( r \) at stop \( s \) and trip \( x \); \( ws \) is the walking speed; \( \Delta T_{\text{max}} \) is the maximal walking and waiting for a transfer; \( d(s, t) \) is the distance between stop \( s \) and stop \( t \); and \( d_{\text{max}} \) is the maximal walking distance between two stops.

Eq. (5) calculates the possibility of a transfer from trip \( x \) of route \( r_{1} \) at stop \( s_{1} \) to all other routes within a maximal walking distance. Such a transfer is possible if the time between arrival and departure is not smaller than the walking time and not larger than the maximal waiting time. Eq. (6) aggregates all possible transfers per trip \( x \) and then averages for the stop. Thus, for an average stop-transfer potential of 3, a passenger alighting will have 3 transfers available.

Given a multi-modal transport network, the stop-transfer potential reflects multi-modal transfers, such as bus-rail.
3.4. Route overlap indicator

In terms of network as well as operation efficiency, route overlap should be minimized. On the other hand, road network structure and route constraints tend to reduce the possibilities for minimizing route overlap. When overlap is minimal, passengers are forced to cross streets in order to perform transfers in contrast to a transfer between two routes sharing a stop or overlapping along a road section, as Fig. 5 illustrates. Thus it is important to analyze the route-pair overlap of a PT network in order to assess efficiency versus ease of transfer. Clearly, this indicator is most relevant for road transport, as bus systems rely on the existed road network, hence high-level of route choice exists.

The following indicator calculates the overlap ratio (percentage) between a pair of routes.
\[ OP_{r_1, r_2} = \frac{ol_{r_1, r_2}}{l_r} \]  

where \( r_1, r_2 \) are two routes; \( ol_{r_1, r_2} \) is the overlap between the two routes; and \( l_r \) is the length of route \( r \).

With Eq(7), average and variance as well as other statistics can easily be calculated.

It is worth noting that for a pure grid network the route overlap indicator will be zero.

4. Case study

The proposed methodology is demonstrated with analysis of the following PT systems: Auckland (New Zealand), Vancouver (Canada), and Portland (Oregon, USA). The Auckland PT network was the testing ground for the concept of using Google Transit data (Hadas and Ranjitkar, 2012). The methodology was then extended to Portland and Vancouver to test the concept further. Portland is cited as a city very similar to Auckland (Mees et al., 2010), and the Vancouver PT network is regarded as an efficient, well run system (Mees, 2010). For the analysis, GTFS as well as road layer data were extracted for April-May 2010 (CloudMade.com, 2010). For this demonstration, the following guidelines were set: (a) time of day period – 7:00–9:00 a.m. (b) day of the week – Tuesday; and (c) routes – buses only (excluding rail, commuter rail, light-rail, cable cars, ferries, etc.). Figs. 6–8 provide an overview of those networks. The data was analyzed with TransCAD (Caliper, 2010), a transportation-oriented GIS package. It is important to note that the analysis is based on “as-is” data sets (PT and road networks) retrieved independently and as automated procedures.

4.1. Transport network coverage level analysis

Fig. 9 summarizes the results of the transport network coverage level indicator. Each bar represents the percentage of total network size (in kilometers) as covered by a certain flow of vehicles per hour (bidirectional). It is evident that the Auckland network suffers...
from a low flow level of PT vehicles compared to Portland and Vancouver, with Vancouver’s low level surpassing that of Portland.

From the spatial representation of the indicator presented in Fig. 10 for Vancouver, it is possible to assess the coverage (higher in city center and along arterials crossing the city).

4.2. Average speed analysis

Based on the published timetables, it was possible to estimate the average weighted commercial-speed (including dwell time). The average speed was weighted by road section lengths. Auckland and Portland have a slower average speed than Vancouver (30, 32, 48 km/h respectively). Figs. 11–13 present the average speed of each PT system respectively.

4.3. Intersection coverage analysis

Intersection coverage level provides similar results to those derived from the road coverage analysis. The difference is the calculation of the percentage based on the total number of intersection versus the percentage of total network length (transport network

Fig. 18. Auckland’s Northern Busway stop-transfer potential at Akoranga station.
coverage level). From Fig. 14 the same conclusions (with comparison to Fig. 9) can be reached.

4.4. Stop-transfer potential

For the stop-transfer potential indicator, the following parameters were used: (a) a maximal walking distance of 100 m, (b) a waiting time of 5 min, and (c) a walking speed of 4 km/h. All available transport modes were analyzed. Both Auckland and Vancouver have a higher level of stop-transfer potential (2.03, 2.02 respectively) than Portland (1.44). Figs. 15–17 present a spatial analysis of the networks. In each map, the PT network is laid out with point size representing the stop-transfer potential. It is evident that Auckland’s high potential is due to multiple routes running along the same road sections. Portland, on the other hand, is a classical grid network with potential transfers at intersections. Vancouver seems to have better transfer potential (spatially speaking), with pivot points at central locations and along major roads. From these figures, it is also possible to reveal the cities’ topologies effect on the PT system, specifically on the stop-transfer potential. High stop-transfer potential is evident around the city centres, as the transport network structures are radiating from the centers to the surrounding districts.

The stop-transfer potential indicator can assist with the detailed analysis of central transfer stations in terms of stop locations, parking, passenger safety, etc. For instance, Fig. 18 (up) presents combined transport network coverage and stop-transfer potential around the Akoranga station which is served by the Northern Busway (Auckland’s BRT) and feeder routes. Fig. 18 (down) presents a combined map and aerial map from which accessibility, parking, etc. can be assessed. From the aerial map it is possible to assess the accessibility – the overlap over the motorway, the location of parking, and the ease of transfer. As the analysis is for the morning peak, the southbound BRT station (towards the city CBD) has a higher potential than the northbound station (drive on left transport system). Furthermore, feeders’ stops are closely located.

4.5. Route overlap

Route overlap statistics (Fig. 19) show that Portland’s PT network has less overlap than Auckland’s and Vancouver’s. This may be due to the grid topology of the road system. Auckland’s overlap is slightly higher than Vancouver’s and this may possibly be related to the Auckland’s road network which is radial/diametrical and where overlap is a side-effect. Vancouver’s PT network is similar in structure to Portland’s, but the overlaps provide a better connectivity (when investigating the stop-transfer potential analysis).

5. Conclusions

The model that was developed is an easy-to-use tool enabling decision-makers to analyze the connectivity of PT networks and to perform benchmarking. Data acquisition is simple, enabling researchers to easily construct a data repository for analyzing networks. Since the connectivity indicators are calculated automatically within a GIS package, it is possible to analyze large public transport networks, as was demonstrated with the analysis of three large cities.

As a GIS based approach, analysis can be carried from a city-wide scale down to block or street section scale. Furthermore, GIS generic tools such as buffering, overlaying, and clipping can be used, as well as performing geo-statistical analysis.

It is possible to perform what-if analysis, such as altering the PT network, changing frequencies, relocating stops, and merging stops for better accessibility. It is also possible to check the effect of different stop-transfer potential parameters in terms of policy-making (maximum walking distance, maximum transfer time, etc.).

The analysis can be carried out according to dynamic properties of the network, specifically the time of day and day of the week.

Travel time analysis, based on published schedules and timetables, is easy to perform. If we assume that the published timetables are based on estimated speeds, then it is possible to calculate travel time by PT mode, time of day, etc., without the need of AVL data. Furthermore, comparing different PT modes speeds and travel times to similar destinations is possible (i.e. bus versus rail).

As the process of constructing a depository is efficient, resources can be allocated to add other layers such as demand (Hadas and Ranjitkar, 2012), or socio-economic dimensions (Currie, 2010).

References

Caliper, 2010. TransCad, 5.0 ed, Newton MA.
Earthican, London.

Fig. 19. routes overlap statistics.


