Analyzing Employment Accessibility in a Multimodal Network using GTFS:

A Demonstration of the Purple Line, Maryland

By

Ting Ma
Ph.D. Candidate
Department of Urban and Regional Planning and Design
University of Maryland, College Park, MD 20742
tingma@umd.edu

Gerrit Jan-Knaap
Professor and Director
National Center for Smart Growth Research and Education
University of Maryland, College Park, MD 20742
gknaap@umd.edu

Submitted to 2014 The Association of Collegiate Schools of Planning (ACSP) Annual Conference
Philadelphia, Pennsylvania
October 30th – November 1st, 2014
Introduction
Transit has a lot economic and social benefits. Among them is the increased accessibilities to jobs. By connecting the labor forces to jobs, transit has the potential to improve the overall employment rate. Sanchez found in Portland, Oregon, and Atlanta, Georgia, that access to transit is a significant factor to labor participation rate (Sanchez, 1999). Kawabata found in Los Angeles, California, that improving transit-based job accessibility significantly enhances the employment probability (Kawabata, 2002). Previous quantitative research demonstrated the transit’s effect on job access. But few zoomed into a smaller scale and explored transit’s impact on job access at the neighborhood level. The neighborhood level analysis is important. It provides opportunities for residents to understand transit’s impact on the neighborhood they live in. In addition, planners can evaluate the job access by transit for specific neighborhood and make economic development and labor training policy suggestions.

Modelling employment accessibility at neighborhood scale requires detailed input data. Luckily, with the availability of two open data, the General Transit Feed Specification (GTFS) and OpenStreetMap, as well as ArcGIS Network Analyst program, planners are able to map out accurate commuteshed in a multimodal network, which takes both the real transit schedules and street network into account. These data and tool enables the analysis of existing transit system, but more importantly, they could help the planning of new transit investment.

In this paper, I will use the proposed Purple Line in the State of Maryland as an example, to demonstrate how to build a multimodal network using GTFS and OpenStreetMap, how to analyze employment accessibility change, and how to interpret and to respond to the findings through planning. The first section of this paper is an introduction to the Purple Line light rail and research question. Then I review time-based, space-based, and space-temporal based accessibility measurements. The method I use is the third one, but with input of open data. The third section is a summary of modeling input and steps. Two examples of results, Langley Park and Bethesda, as well as policy implications are discussed later.

Modelling Accessibility: A Literature Review
Time-based method
There are three methods measuring accessibility, time-based, space-based, and a combined method. The first type, time-based measurement, is often in an index format. Hansen defines accessibility as “a measurement of the spatial distribution of activities about a point, adjusted for the ability and the desire of people or firms to overcome spatial separation” (Hansen, 1959). As shown in equation (1), accessibility (A) of point 1 to zone 2 is calculated as the entropy of the size of activities (S1) in Zone 2, and travel time (T) between point 1 and zone 2 with X the effect exponent.

\[ A_{1-2} = \frac{S_1}{T_1^{X-2}} \]  

Hansen’s travel time-based accessibility index has been widely accepted by scholars in transportation-land use research field, and several variances have been developed. For example, Shen integrated transportation modes, and proposed equation (2a), (2b) and (2c) (Shen, 2000). Travel impedance function \( f(C_{ij}) \) is expressed as \( e^{-\beta C_{ij}} \), the exponential function of travel time \( C_{ij} \).

\[ A_{i}^{auto} = \frac{\sum_{j} E_{j} f(C_{ij}^{auto})}{\sum_{k} W_{ik} f(C_{ik}^{auto}) + (1-\alpha) W_{ik} f(C_{ik}^{trans})} \]  
\[ A_{i}^{trans} = \frac{\sum_{j} E_{j} f(C_{ij}^{trans})}{\sum_{k} W_{ik} f(C_{ik}^{auto}) + (1-\alpha) W_{ik} f(C_{ik}^{trans})} \]  
\[ A_{i}^G = \alpha_k A_{i}^{auto} + (1-\alpha_i) A_{i}^{trans} \]
Time-based method has also been used in mapping. For example, the Institute on Race and Poverty (IRP) conducted a commute shed analysis in the Twin Cities. Data is the 1990 and 2000 journey-to-work data of Census Transportation Planning Packages at the Traffic Analysis Zone (TAZ) level. A travel time matrix between each pair of TAZ is created. 0-20 minute, 20-30 minutes, 30-40 minutes, and > 40 minutes commute time benchmarks are used to determine commute shed (Institute on Race and Poverty, 2006). Transit accessibility index is useful for planners and engineers, but may be too abstract for the ordinary people to understand. On the other side, one advantage of mapping out where transit can get people to the high visibility. With maps, planners, citizens, and transit users can easily interpret their accessibility and identify the impact of proposed transit investment.

**Space-based method**

In space based mapping, service areas are stimulated in a network model. Cheng and Agrawal (2010) summarized four types of space-based mapping mechanisms based on literature, as shown in Figure 1. The first type is the immediate area or buffer area along the two sides of the streets/transit routes (Wirasinghe and Vandebona 1987; Chapleau et al.1987). This corridor method is inaccurate because the segments between transit stations are not accessible to riders. So service areas are then measured as concentric polygons surrounding transit stations (Dufourd et al. 1996; Bruno et al. 1998). Ways drawing polygons vary. The Euclidian metric creates circle buffers at transit stations, assuming that space surrounding transit stations are equally and freely accessible. The Manhattan metric assumes a perfect perpendicular east-west and north-south street grid, and people only make right turns. Service areas in the Manhattan metric is typically diamonds. However, since in most cases street grids are not perfect, network analysis is used to stimulate service areas reachable along the actual street network. Impedance is usually set to be the maximum distance one can walk. ArcGIS is the major package for realizing this network analysis (Kimpel et al. 2006) (Cheng & Agrawal, 2010).

Figure 1. Four methods to measure transit service areas

(Source: Cheng 2010)
A time-space combined method
The third method takes into consideration both time and space constraints by creating a “time-space prism.” The term was proposed by Hagerstrand in 1970 as “the set of locations in space-time that are accessible to an individual given the locations and duration of fixed activities, a time ‘budget’ for flexible activity participation, and the travel velocities allowed by the transportation system” (Hägerstraand, 1970; Miller, 1999).

Cheng and Agrawal (2010) found the space-based transit service area mapping methods ignore the travel time parameter, and propose a series of shrinking service areas as the distance from travel origins increases. His idea is illustrated in Figure 2. The proposed method is named TTSAT. The mechanism of TTSAT is this: (1) mapping transit network including stations and routes, (2) calculate three travel time (travel time between origin and transit stop, wait time, and in vehicle time) and identify all transit stations accessible, (3) subtract the travel time for each station from the travel time budget, (4) mapping areas accessible to riders within the remaining travel time away from transit stations they get off (Cheng & Agrawal, 2010).

Figure 2. Transit Service Areas Generated by Traditional Methods vs. TTSAT.

Tribby and Zandbergen (2012) also explored a spatial-temporal method to evaluate transit network. To create a multimodal network, they combined a pedestrian network and a bus transit network using GIS, and attached bus schedule tables. Travel time is measured as the sum of walking time between origin/destination and bus stops, travel time along bus routes between two stops, and waiting time. Similarly to Chen and Agrawal’s work, waiting time is assumed to be the half of the headway time. Applying their model to test two new rapid bus lines in Albuquerque, New Mexico, they found the model suitable to assess the impact of new transit investment, such as how much travel time would be saved (Tribby & Zandbergen, 2012).

Two improvements can be made to the precedent work. First, the above two models only include bus. But in reality, riders may use a combination of transit modes rather than using one. For example, in Washington D.C. metropolitan area, the combination of pedestrian, bus, and heavy rail is not uncommon for commuters. Thus, a multimodal transit network is highly preferred from a transit user’s perspective. The other potential improvement is better input data, including street network and transit schedule. Cheng pointed out that TIGER/Line data is not detailed enough for pedestrian and bicycle analysis because it includes only vehicle-accessible streets.
Purple Line in Maryland
Purple Line is a proposed light rail located in central Maryland, running through Montgomery County and Prince George’s County. It is close to Washington, D.C. and connect to the Washington Metropolitan Area Transit Authority (WMATA)’s Metro rail system at four transfer stations. On March 19, 2014, a Record of Decision (ROD) was issued for the Purple Line Project by the Federal Transit Administration (FTA), showing that all requirements of the National Environmental Policy Act have been satisfied. At this moment of this paper, the Maryland Transit Administration (MTA), who is the project sponsor, is waiting for the FTA’s decision on federal funding for the Purple Line project (Maryland Transit Administration, 2014).

The 16-mile light rail line is designed to connect Bethesda, Maryland, and Carrollton, Maryland; along the transit corridor are neighborhoods with diverse characteristics. For example, Langley Park on the University Boulevard is a cluster of Hispanic population, University of Maryland in College Park accommodates more than 30,000 students, faculty and staff, and Downtown Silver Spring and Bethesda serve as two major regional employment hubs of Maryland. Based on their different characters, five subareas were designated, which are: Bethesda - Chevy Chase, Silver Spring, International Corridor, University of Maryland, and Riverdale – New Carrollton.

Figure 3. Purple Line Transit Corridor and Subareas

The Purple Line, from an urban planner’s perspective, is a valuable addition to the existing rail transit system in Washington, D.C. Currently, there are six rail transit lines in Washington, D.C. and its inner suburbs, operated and maintained by the Washington Metropolitan Area Transit Authority (WMATA), making D.C. a transit-accessible and walkable urban place. However, the existing transit system is monocentric, with all lines running into/from downtown D.C. This transit system pattern has greatly helped suburban residents commute to their downtown workplaces, however, it is obviously ignoring the transit needs to travel between suburban destinations.
However, in reality, the concept of the Purple Line received different, sometimes contrary opinions. Many residents take the Purple Line an opportunity to increase their accessibility and mobility, some find it a potential greener and sustainable transport alternative to commute/casual travel, while the others feel the Purple Line not beneficial since “we’re an automobile culture” (Action Committee for Transit, 2014). In terms of the Purple Line’s impact on housing, some residents and interest groups see it a chance to connect to more housing options, while the others fear about losing housing affordability and being priced out. Similarly, some business owners predict that the Purple Line would bring them more customers while some else, mostly local small businesses, worry about surviving the construction period and competing with the new comers, particularly the national chain stores.

These discussions are rooted in one question: what do people know about the Purple Line’s impacts on their neighborhood and themselves? To reveal the Purple Line’s impacts, and to maximize the benefit of the Purple Line, planners and researchers at the National Center of Smart Growth Research and Education, University of Maryland, College Park, established the Purple Line Corridor Coalition (PLCC) as a platform to engage organizations active in the corridor and to provide valuable information to help assure that investments in the Purple Line. Since its establishment in June 2013, the PLCC is able to attract about 30 members, hold two workshops, and publish several data-driven maps and reports. Mapping and analyzing the current and future employment accessibility is one of the research that PLCC has been interested.

**Purple Line Employment Accessibility**

To map out how the proposed Purple Line will change corridor neighborhoods’ employment accessibility, I built two models and stimulated the current situation and the future one. The two models will be called “before” model and “after” model in this paper. Two travel modes, transit (which includes bus and rail) and walking, have been considered.

**Data**

GTFS and OpenStreetMap, which are both open data, are two major input data. They will be introduced separately in below. The employment data is from the Longitudinal Employer-Household Dynamics (LEHD). Since it is a relatively familiar data to the most planners, introduction to LEHD is omitted on purpose in this paper.

**GTFS**

The General Transit Feed Specification (GTFS), originally known as the Google Transit Feed Specification, is a format for transit schedules and geographic information. Formatted into a relational database, GTFS consists of key elements of transit: (1) administration information, such as operator and service calendar, (2) spatial information, including stops and routes, (3) schedules written by trips, and (4) others, such as fares (which is optional). GTFS was co-designed by TriMet and Google Maps, originally to create transit trip planner. The first Google Transit Trip Planner was launched on December 7th, 2005, and TriMet was the only operator available on Google Maps (Roth, 2010). As an open data, GTFS is voluntarily prepared and maintained by regional/local transit agencies and shared with public. Till March 2014, there are about 250 GTFS available in the Unites States, according to Google Transit’s list; it has become an industrial standard. GTFS is revolutionary; prior to GTFS, software engineers had to collect pieces of information from operator’s websites or submit Freedom of Information Act requests to obtain data. Data is mailed on a CD and thus is limited in access. Sometimes operators charged on sharing data to the public (Roth, 2010). With GTFS and other open data, planners have access to more accurate transit information (World Bank, 2013).

Recently, planners and urban researchers have found GTFS helpful and started to include it in urban studies. The beauty of GTFS is that it provide an opportunity to compare transit services of different
places and regions and thus enables study at large scales. For example, the Brookings Institute used GTFS to conduct a national study comparing transit access to jobs. Centroids of block groups located within ¾ mile from any transit stations are identified as transit-served origins, then connectors are created to link origin and the nearest transit stations. Similarly, centroids of each census tracts, which are the destinations, are connected to transit stations within ¾ mile or the nearest one if distance is larger than ¾ mile. Three time, origin connector walk time, in-vehicle time, and destination connector walk time, are added up as the total travel time. Therefore, for each block group, its accessible census tracts within certain total travel time budget can be mapped, and thus the total number of accessible jobs can be calculated using LEHD data (Tomer, Kneebone, Puentes, & Berube, 2011). Using GTFS, transportation engineers developed the Delaware Valley regional transit forecasting model (Puchalsky, Joshi, & Scherr, 2012). The University of Arizona Transit Research Unit (UATRU) uses GTFS data for detecting alighting and/or transfer stop locations with farecard transaction data (Nassir, Khani, Lee, Noh, & Hickman, 2011), and for aggregating transit stops (Sang Gu Lee, Hickman, & Tong, 2012). (S. G. Lee, Tong, & Hickman, n.d.)

The potential using GTFS in transit planning is well recognized in the report Expanding the Google Transit Feed Specification to Support Operations and Planning, prepared for the Federal Department of Transportation. In its conclusion part, one potential found is to used GTFS “to measure accessibility based on time and location.”(Catala, Dowling, & Hayward, 2011) Antrim and Barbeau (2010) provided an overview of opportunities to leverage GTFS, and pointed out that “, GTFS data can power many other different types of transit and multimodal software applications, including multimodal trip planning, timetable creation, mobile apps, visualization, accessibility, analysis tools for planning, real-time information, and interactive voice response (IVR)” (Antrim & Barbeau, 2013).

Besides its high accuracy and detail level, another benefit of GTFS is that it includes different transit types. Transit types are assigned different codes in GTFS. For example, rail is “2” and bus is “3”. Therefore, GTFS help to save the time connecting routes and schedules of different transit types.

GTFS for the Purple Line commuteshed models are listed in Table 1. In the “before” model, GTFS of four existing transit systems, the WMATA system, MTA system, Montgomery County’s Ride-On bus system, and Prince George’s County’s TheBus bus system, were included. The first three GTFS data are available online. The Prince George’s county bus was not provided, and I created from scratch based on information on the website.

Table 1. GTFS Input of Two Models

<table>
<thead>
<tr>
<th>“Before” Model</th>
<th>“After” Model</th>
<th>Data Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Transit Schedule</td>
<td>WMATA GTFS</td>
<td>WMATA</td>
</tr>
<tr>
<td>MTA GTFS</td>
<td>MTA GTFS</td>
<td>MTA</td>
</tr>
<tr>
<td>Montgomery RideOn GTFS</td>
<td>Montgomery RideOn GTFS</td>
<td>RideOn</td>
</tr>
<tr>
<td>Prince George’s TheBus GTFS</td>
<td>Prince George’s TheBus GTFS</td>
<td>Self-Made</td>
</tr>
<tr>
<td></td>
<td>Purple Line GTFS</td>
<td>Self-Made</td>
</tr>
</tbody>
</table>

In the “after” model, besides the four transit GTFS, I also included the Purple Line transit GTFS. The Purple Line GTFS was created based on schedules provided in the Federal Environment Influence Statement (FEIS) and station and route shapefiles from MTA. As Table 2 illustrates, during peak hours, the Purple Line is planned at a 6-minute frequency, and 10-minute frequency during the non-peak hours. The Purple Line will run seven days a week with different schedule to meet the first and last Metrorail at each of the four transfer stations.
Table 2. Planned Schedule of Purple Line

<table>
<thead>
<tr>
<th>Day of Week</th>
<th>Hours of Operation</th>
<th>Peak-Hour Headway</th>
<th>Off-Peak Headway</th>
</tr>
</thead>
<tbody>
<tr>
<td>Monday – Thursday</td>
<td>5:00 am – 12:00 am</td>
<td>6 minutes</td>
<td>10 minutes</td>
</tr>
<tr>
<td>Friday</td>
<td>5:00 am – 3:00 am</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Saturday</td>
<td>7:00 am – 3:00 am</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sunday</td>
<td>7:00 am – 12:00 am</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(Source: Purple Line Federal Environment Influence Statement. MTA. 2013)

OpenStreetMap

OpenStreetMap is an open data co-built by a community of mappers, including GIS professionals and engineers. Compared with other street data source, such as the stimulated street network of the Maryland State Transportation Model (MSTM), it has more details and captures the real road shape. Compared with the centerline data prepared by the State Highway Administration with illustrates road lanes, the OpenStreetMap has one line for each road and reduces the calculation.

For the “after” model, important planned connections between the proposed stations and their adjacent neighborhoods were added to the future street network. Information on planned street connections are from the FEIS Conceptual Engineering Plans by MTA. This step is crucial for the Purple Line “after” model. Part of Purple Line will run on the existing Georgetown Trail, which has limited access to its surroundings. Thus, some proposed stations lack connections to street grid and any relevant designs need to be captured in the model.

Modeling steps

Figure 4 illustrates the modeling steps, as well as input data and output products. ArcGIS 10.1 is the main package building up the two models.

Figure 4. Modelling Steps

1. Collect, clear-up, and create GTFS and street networks. GTFS come from different sources; during this step, it is important to ensure that all GTFS share the same service period. For the Purple Line modeling, all GTFS are set to provide service between June 1st, 2014 and June 30th, 2014.

---

1 One can read more on detailed information on step 2, 3, 4, 5 from [http://www.transit.melindamorang.com/AddGTFSstoND_UsersGuide.pdf](http://www.transit.melindamorang.com/AddGTFSstoND_UsersGuide.pdf). The author has used the toolkit Add GTFS to a Network Dataset version 0.3 beta created by Melinda Morang and Patrick Stevens, Esri for these steps.
2. Generate transit routes and stations. The latitude/longitude information of transit stations in GTFS is read in ArcGIS and point shapefiles are created to store the spatial information and other fields. Straight lines are generated to connect two adjacent stations; lines are converted to line shapefiles as the transit routes. In total, there are 24,127 transit stops and 28,030 transit route segments in the “before” model. 21 more stops and 20 more segments are added to the “after” model.

3. Create connectors between transit stations to street networks. Due to the different sources, transit stations are not always mapped to the street network. Therefore, connectors, short straight lines which are perpendicular to streets, are generated to connect transit system and street network. This step is very vital. By creating snaps and connectors, layers are connected and only connected at stops, which prevents walking along transit lines.

4. Create a multimodal transit network. With “creating a multimodal network dataset” toolkit provided in ArcGIS Network Analyst, a multimodal transit network is created. There are two assumptions, that each connector between stop and street takes 15 seconds, and people walk at the speed of 180 feet/minutes.

5. Add GTFS transit schedule to network. Convert GTFS to ArcGIS recognizable transit schedule table, and attach the table to the multimodal network.

6. Create service area for areas of interest. Service area, according to Esri’s definition, is a region that “encompasses all accessible streets” under certain impedance condition (Esri, 2014). For the Purple Line modelling, travel time is the impedance. For example, a 30-minute commuteshed/service area of place A is all areas that one can access to within 30 minutes travel time (including walking and transit time) starting from A. For the Purple Line research, three layers of commutesheds were created, which are 30-minute, 45-minute, and 60 minute; since transit schedules vary by day and time, for analysis, 8:00 am of the morning of June 18th, 2014, Wednesday, was selected to output results.

7. Overlay with LEHD employment data and sum by commutesheds. LEHD data includes employment data by residence/workplace, wage level, education level, industry, and so on. For this research, employment data by education attainment (skill level) was analyzed.

Results
Langley Park
Langley Park neighborhood accommodates a large number of immigrant population, among which Latino is the largest. About 80 percent of its 17,300 population are Hispanic. 15 percent of population lives at or below the Federal poverty guidelines. In terms of transport, 33 percent households have no vehicle, and 28 percent of workers commute by public transportation (MTA, 2013). Riggs Road station is one of the two stations serving Langley Park. Figure 5 shows the Purple Line’s impact on its transit accessibility. On the left are the 30-, 45-, and 60- minute commutesheds (in dark purple, purple, and light purple separately) under the current transit conditions, and on the right are the commutesheds with additional Purple Line service. Riggs Road station is marked as a yellow star.
As the maps show, with the Purple Line, areas that residents in Langley Park can commute to expand significantly along the Purple Line corridor. Within 30 minutes, starting from Riggs Road, one can travel to downtown Silver Spring, which is one mile west away from where he can travel to before the Purple Line. Within 45 minutes, one can travel to the two end destination of the Purple Line, Bethesda to the west and New Carrollton to the east. Within 60 minutes, downtown D.C. is accessible.

Table 3 shows the number of jobs located within commutesheds before and after the Purple Line, as well as the change. Total number of jobs within 30 minutes increased from 28,800 to 51,217 with the Purple Line. The change is as high as 78 percent. Similarly, total number of jobs within 45 minutes and 60 minutes increase a lot. In terms of job skill levels, within 30-minute transit commute distance, low-skill jobs (defined as jobs requiring an education degree less than high school) and medium-skill jobs (high school degree or associate degree) see a significant increase (80 percent) after the Purple Line. In the contrary, high-skill jobs will increase 100 percent in the “after” model.

The neighborhood of Langley Park has a much higher proportion of low-skilled workers. According to the 2011 LEHD data, 27 percent of workers living in Langley Park have less than a high school diploma,
which is double the regional (D.C., Montgomery County, and Prince George’s County) proportion, 14 percent. The expansion of commutesheds after the completion of the Purple Line is a good news to these low-skilled workers in Langley Park. With 80% more level-matched jobs accessible within 30 minutes, they are exposed to more opportunities and could save commute time.

**Bethesda**

The Langley Park modeling and analysis example shows the current and future employment accessibility of residential neighborhood. In addition, the model is useful for mapping accessibility to employment centers.

Bethesda is part of Bethesda-North Bethesda, one of the 23 Maryland statewide employment centers. About 200,000 jobs are provided in the less than 5,000-acre area, which makes a high job density of 38.4 jobs/acre. With abundant high-skill, high-wage jobs, the average annual wage was more than $65,000 in 2007. In addition, Bethesda has a high job diversity level. Currently, Bethesda is served by Red Line Metro Rail, WMATA Metro bus, and county bus.

As Figure 6 demonstrates, Bethesda is much more accessible with Purple Line. 60 minute commuteshed expands to 176 square miles and reaches to Beacon Heights neighborhood in Prince George’s County. The expansion would connect additional 70,000 workers to Bethesda.

**Conclusion and Discussion**

How the new transit investment will impact corridor neighborhoods? With the model introduced above, this question can be answered by maps and numbers. For Langley Park neighborhood in the planned
Purple Line corridor, 80 percent more low-skill jobs can be accessed after the Purple Line, which will benefit the neighborhood since it has a high proportion of low-skilled workers. For Bethesda, a regional employment center, the Purple Line would help more than 70,000 workers to access it within one hour. How could Bethesda take advantage to this potential labor force growth? This is the question left for local planners and researchers.

Research could be further extended in two ways. First, currently only two modes, walking and transit (including rail and bus), have been included. Ideally, biking could be integrated into the stimulation model. Washington, D.C. and Maryland has a culture of bicycling, and has a great bicycle sharing program, Capital Bikeshare. It is not unusual to see young professionals or students biking between transit stations and home/workplace. Bikeable lanes, either designated or shared, would be added to the model as a new layer, and connect to street layer at the bicycling docking stations. Second, compared to the static maps shown above, one thing my colleagues and I have been exploring is to create a Web GIS tool, so we can upload our data online for the audience to model employment accessibility for their own area of interest. The web tool could be used in public participation meetings, to inform and to educate them about how new transit would change the neighborhood’s future.

Acknowledgement
The authors thank Melinda Morang at Esri Network Analyst development team for GTFS integration tool, and NCSG summer interns, Max Kittner and Stanley Y. Wang, for their assistance work.

Reference
Ma and Knaap


