Use of General Transit Feed Specification (GTFS) Data in Analyzing Transportation Barriers Faced by Housing Voucher Holders

Peter K. Kwok, Julia Rotondo
NORC at the University of Chicago, 55 East Monroe Street 20th Floor, Chicago, IL 60603

Abstract
General Transit Feed Specification (GTFS) is a standardized format for public transportation timetables and locating information. Since its introduction in the mid-2000's, GTFS has become a popular choice of data specification by nearly 200 transit agencies in the U.S. While it is long known that people with disadvantaged socio-economic status tend to rely on public transportation, the potential of using GTFS data in analyzing social problems has been relatively unstudied. In this case study, we attempt to assess the non-financial aspect of transportation barriers faced by housing voucher holders by analyzing the geographic coverage and service frequencies of bus routes. In particular, we use the Housing Choice Voucher program by the Housing Authority of Baltimore City (HABC) as an example. This program allows the voucher holders to rent eligible apartments located in "allowable" tracts, mainly in the neighboring counties of Baltimore City. By analyzing the locations of the current HABC residential facilities in relation to bus routes, we illustrate some of the challenges current HABC residents and potential voucher holders may encounter when re-locating.

Key Words: Public Transportation, Public Housing, Network Analysis, Transportation Connectivity

1. Background

The primary objective of this paper is to illustrate the use of public transportation[1] data in the analysis of housing problems. We demonstrate how General Transit Feed Specification (GTFS) data can be used to help predict relocation preference of participants of Housing Choice Voucher Program (HCVP). The U.S. Department of Housing and Urban Development (HUD) established HCVP in 1998 to subsidize public housing residents to rent in the private market. It was assumed that the residents would prefer private housing if they were financially empowered to make that choice. (Khadduri and Struyk 1982) However, it soon became obvious that people were reluctant to participate even when financial burden did not appear to be the main concern. As observed by Varady *et al.* (1999), these residents often do not have cars and are highly reliant on public transportation. If the prospective neighborhood does not have the transit services that they are familiar with, then some of them would not even visit the prospective housing unit, at least not without considerable counseling effort or even a ride for a house tour. When they moved, they did not move very far away from their old locations and remained spatially clustered. (Devine *et al.* 2003) Obviously, there are many considerations behind a decision to move residency. However, since transportation is such a priority for this special group of people, we will focus on the role of the non-financial aspect of public transportation. In particular, we seek to find out how to help voucher holders relocate through identifying and disseminating the relevant information.

2. Methods

We will analyze HCVP as implemented by Housing Authority of Baltimore City (HABC) and their agent Metropolitan Baltimore Quadel that manages the voucher program. Due to the experimental nature of this

[1] We will use the term "public transportation" interchangeably with "transit" (contraction for "mass transit").

study, we will make various modeling assumptions for operational purposes, as long as the primary objective of illustrating the use of transportation data is not compromised.

Our analysis looks for “hot spots,” namely, locations favorable for relocation, and “cold spots,” namely, locations unfavorable for relocation. The contrast between two such areas will serve as the basis of quantifying transportation barriers against residential relocation. We identify the prospective locations throughout Greater Baltimore to which voucher holders can potentially move. Then we obtain transportation statistics in those areas, and examine how they differ from the rest of Greater Baltimore. More specifically, we will use GTFS data to locate where services are provided, and then associate those locations with Census 2010 data to provide the measures needed for quantification. Properties of transit network will emerge as critical considerations when we evaluate individual movements on a technological network (Newman 2010) like the transportation system in our study. Finally, we will assess how the transportation barriers may be perceived, and make some suggestions on how to overcome them.

Baltimore offers two advantages for analyzing housing and transportation. First, Baltimore City has highways and railways radiating from the city center to its 6 neighboring counties (which, all together, define Greater Baltimore). Such geographic features make the “distance from downtown” a useful measure of resident dispersion. This is even more appropriate if the Concentric Zone Model is assumed. (Burgess 1924, 1925) This urban development model applies when the city has a single center, called a Central Business District (CBD), surrounded by various layers of residential areas of different characteristics.[2] Glaeser *et al.* (2008) observed that “older” cities that have a single CBD tend to have poverty concentrated within 3 miles from city center, while “newer” cities tend to have jobs dispersed to 3 to 10 miles. They also estimated that it costs people about 2 extra minutes of commute time for every mile moved away from CBD. Later we will use distance from downtown to define housing zones for analysis.

Second, Baltimore is a leader in GTFS data availability. Fixed route transit agencies from this area publish almost all of their transit service information in the GTFS format, which is essentially a standardized format of maps and time tables for the web application Google Transit. GTFS has been around since 2005, and, hence, is not in itself a new data source *per se*. What is new in this paper is the introduction of GTFS into housing analysis. GTFS informs transit network topology, and allows linking geocoded locations to auxiliary data. Together, a new combination of data is created to enable new utilization. Even though GTFS is static data in the sense that all locations and timing of transit vehicle or vessels arrivals are pre-determined and have zero variance across time, and though transit load information is not available for modeling passenger (or rider) demands, we consider the data to be adequate for our purposes. Furthermore, HABC happens to publish transportation information, such as bus routes near their housing facilities, which are useful indicators of their residents’ transportation habits.

3. Data

HCVP divides the 7 county-equivalent areas in Greater Baltimore into allowable and non-allowable tracts. A tract is a geographic unit defined by the U.S. Census Bureau. Due to boundary and numbering changes between 2000 and 2010, and due to an apparent time lag in official website update, there is ambiguity on the actual coverage. At the time of the study, the voucher applicant website used the documented tract numbers (Metropolitan Baltimore Quadel 2012) in conjunction with the “Census 2010” definitions. Since our primary purpose is to illustrate the use of GTFS data, we will use the available tract numbers as they are presented. The Appendix shows what a more thorough tract definition conversion may look like.

[2] Alternative models include the Sector Model (Hoyt 1939), which defines zones as wedge-like or fan-like shapes and can accommodate angle-based measures, and the multi-nuclei Model (Harris and Ullman 1945), which has more than one center and results in zones of more irregular shapes. Both models can be supported by GTFS data.

Figure 1 shows a map of Greater Baltimore. County boundaries (rather than shorelines) are drawn as thick, black lines; and tract boundaries, as thin, white lines. HABC will subsidize their residents for renting eligible housing units in the "allowable tracts" indicated by the solid red or in part of the red stripped areas. The "non-allowable tracts" are not eligible for subsidy, and are indicated by the blue areas, which cover some land, and the white areas, which are entirely on Chesapeake Bay. All HABC facilities are dispersed approximately in the blue areas within Baltimore City (unmarked in Figure 1). According to Census 2010, the allowable tracts have 1.2 million residents and 477 thousand housing units within 1,628 squared miles, whereas the non-allowable tracts have 1.5 million residents and 655 thousand housing units within 974 squared miles. The allowable tracts account for slightly over 40% of the general population and housing units, but slightly over 60% of land area. Thus, it can be inferred that the allowable tracts are, in general, less densely populated than their non-allowable counterpart.

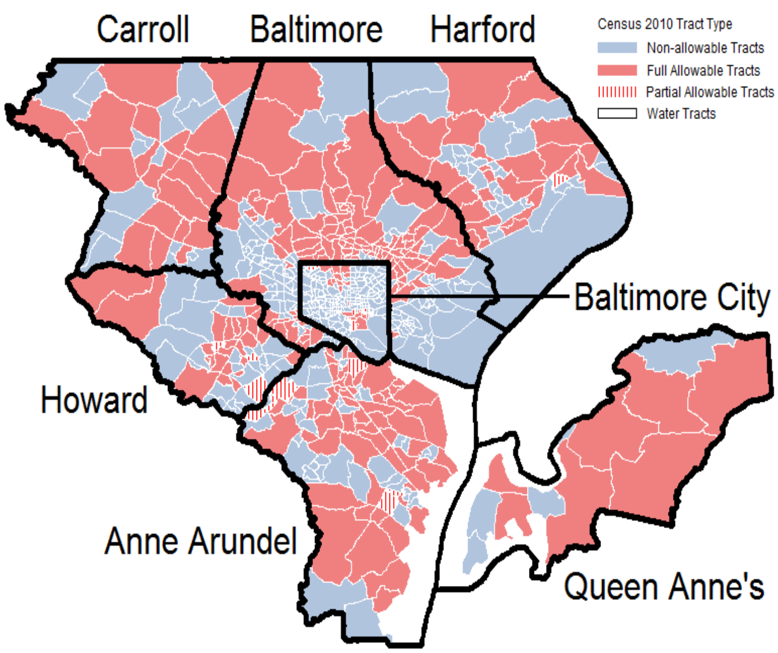

Figure 1: Census 2010 Tracts in Greater Baltimore (Assuming No Definition Change)

Next we retrieve all available GTFS data (dated 10/26/2011) at the route, trip, and stop levels. A "stop" can be any pick-up or drop-off point (*e.g.*, bus stop, train station, pier, *etc.*). A "trip" is an administrative unit that describes the movement of a transit vehicle or vessel from a starting terminal (first stop), through any number of intermediate stops, to an ending terminal (final stop). The ordering of all possible stops on a trip forms a "stop sequence." A "route" is a group of trips with similar service geographies and administrative purposes. In reality, a bus does not always make all the stops. However, since we are interested in the full scope of service coverage rather than just the actual services taken place or used, we will consider the "maximum times stopped" and practically assume all stops are made on all trips.

Table 1 covers almost all fixed route transit services touching Greater Baltimore except for a few bus routes in Baltimore City and Anne Arundel County (see the footnotes for Table 1). The overall number of stops counts each stop only once even when shared by multiple trips. We also note that the public transportation services in Carroll County are mainly "demand response" (*i.e.*, call service without any fixed schedule) running on "deviated fixed routes" (*i.e.*, allowing off-route trips besides pre-determined stop locations). Such service characteristics are deemed out of scope for this study. Hence, no transit agency covering Carroll County was included. As expected, the Maryland Transit Administration (MTA) dominates the services (regardless of transportation mode) provided in Greater Baltimore.

Table 1: GTFS Data (10/26/2011) of Fixed Route Services Touching Greater Baltimore

Transit Agency	*Transportation Mode*	*Routes*	*Trips*	*Stop Locations*	*Maximum Times Stopped*
Maryland Transit Administration	Light Rail	1	699	69	15,585
Maryland Transit Administration	Subway	1	557	26	7,405
Maryland Transit Administration	Heavy Rail§	3	93	79	870
Maryland Transit Administration	Bus	82	11,620	6,083	877,486
Central Maryland Regional Transit	Bus	10	332	417	11,542
Howard Transit	Bus	8	437	404	15,639
BWI Thurgood Marshall Intl Airport	Bus	1	690	4	2,070
Charm City Circulator	Bus	2*	498	56	14,442
Charm City Circulator	Ferry§	1	146	3	292
Annapolis Transit	Bus	3**	166	51	1,785
Frederick Transit Meet-The-MARC	Bus§	1	21	29	166
Overall		**113**	**15,259**	**7,017**	**947,282**

§Weekday service only. *Missing 3 minor routes. **Missing 5 minor routes.

4. Distributional Results

Our main results focus on identifying housing zones with higher population density and service availability, which are considered favorable conditions for relocation. We predict success (or a higher chance of success) for HCVP when their allowable areas exhibit such characteristics.

4.1 County that Predicts Success for HCVP

Table 2 cross-tabulates county-equivalent areas by demographic and transit service measures. Number 0.0 indicates a proportion less than 0.05%. Blank cells indicate no data available. Numbers in each column add up to 100%. Our demographic measures include counts of population and housing units, as well as land area in squared miles. Our transit service measures include counts of stop locations and maximum times stopped.

Table 2: Allowable ("Allow") and Non-allowable ("Non") Tracts by County-Equivalent

County-Equivalent	*Column Percent of Population in…*		*Column Percent of Housing Units in…*		*Column Percent of Land Area in …*		*Column Percent of Stop Locations in…*		*Column Percent of Max. Times Stopped in…*	
	Non	*Allow*	*Non*	*Allow*	*Non*	*Allow*	*Non*	*Allow*	*Non*	*Allow*
Baltimore City	35.1	6.7	38.7	9.1	7.2	0.7	69.3	27.0	80.5	34.7
Baltimore	26.7	33.8	25.8	34.9	22.5	23.3	18.9	44.5	16.4	52.2
Anne Arundel	14.4	27.0	14.0	25.3	15.3	16.3	5.5	13.7	1.9	8.0
Howard	9.5	12.0	8.2	11.6	13.0	7.6	5.6	13.1	1.2	5.0
Carroll	5.1	7.6	4.5	6.9	15.3	18.3				
Harford	8.2	10.1	7.8	9.4	20.7	14.5	0.7	1.6	0.0	0.1
Queen Anne's	1.0	2.8	1.0	2.8	6.0	19.3	0.0	0.1	0.0	0.0

If we uniformly randomly select one person from the non-allowable tracts, then we may expect a 35.1% chance that such person resides in Baltimore City. In contrast, if we randomly select one from the allowable tracts, then the chance drops to merely 6.7%. If the allowable tracts were randomly designated, then the two probabilities should have been very similar. The existence of such a discrepancy suggests a geographic "preference" since the allowable tracts were assigned in such a way to avoid Baltimore City.

A likely rationale for such differentiation is to encourage residential dispersion and, ultimately, integration into the new neighborhoods.

In spite of such preference, Baltimore City still “captures” 27.0% of the stop locations, and 34.7% of the maximum number of times stopped. By “capture” we mean that the residents live in the same county-equivalent area as where the stops are located. In reality, not all residents will live “close enough” to the stops to use the services (at least not without assistance from other modes of transportation), nor will all of them use all available services regularly even when they live close enough. The implication is that Baltimore City is a demographic “cold spot” in the sense that it is not extensively covered by HCVP according to various demographic measures. On the average, a voucher holder will likely encounter relatively fewer housing units on the allowable tracts. Nonetheless, it is a transportation “hot spot” in the sense that a large portion of transit services is available to local residents. A voucher holder who seeks to take advantage of public transportation will likely find this area convenient. Relocation within Baltimore City would impose the least inconvenience (or transportation barriers) and, hence, the least lifestyle changes for the public housing residents.

Similarly, Baltimore County is predicted with a high chance of success for HCVP. Anne Arundel and Howard Counties have somewhat proportional demographic and transportation measures, and do not show any obvious advantage or disadvantage. Carroll, Harford, and Queen Anne’s Counties account for over 50% of the land areas among the allowable tracts, but only 20% of the population and virtually none of the transit service. These areas have low population densities, and are highly suggestive of a relatively rural lifestyle. These areas can be considered cold spots, and would impose the biggest change in lifestyle to urban residents in terms of diminished freedom of movement by public transportation.

4.2 Distance Zone that Predicts Success for HCVP

Since the transportation preference of residents is strongly influenced by distances between their points of interest, it seems reasonable to speculate that a physical factor such as distance, rather than an administrative factor such as county, imposes the transportation barriers. Hence, we re-partition Greater Baltimore into 5 distance zones. The first two zones in the 0-to-3- and 3-to-10-mile ranges were suggested by Glaeser *et al.* (2008) Furthermore, their estimate of two extra minutes per mile moved away from the city center can be extrapolated to roughly half an hour extra time for every 15 miles moved. We can define three more zones accordingly. Results are summarized in Table 3 below. Number 0.0 indicates a proportion less than 0.05%. Blank cells indicate no data available.

Table 3: Allowable (“Allow”) and Non-allowable (“Non”) Tracts by Distance Zone

Distance from Down-town (Mile)	*Projected Extra Commute Time (Hour)*	*Column Percent of Population in…*		*Column Percent of Housing Units in…*		*Column Percent of Land Area in …*		*Column Percent of Stop Loca-tions in…*		*Column Percent of Max. Times Stopped in…*	
		Non	*Allow*	*Non*	*Allow*	*Non*	*Allow*	*Non*	*Allow*	*Non*	*Allow*
0 to 3	0 to ¼	14.7	3.4	17.6	5.1	2.2	0.2	34.1	11.9	44.0	19.7
3 to 10	¼ to ½	39.9	30.4	39.8	31.9	14.5	7.3	53.5	57.4	52.3	62.7
10 to 25	½ to 1	37.5	52.4	35.2	49.6	43.5	44.0	11.9	30.6	3.7	17.6
25 to 40	1 to 1½	7.8	13.4	7.3	13.0	37.1	41.6	0.5	0.1	0.0	0.0
Over 40	Over 1½	0.1	0.4	0.1	0.4	2.7	6.9				

The 3-to-10- and the 10-to-25-mile zones are hot spots according to both demographic and transit service measures. About 83% of the people in the allowable areas live in this city fringe. And they have access to 80 to 88% of the transit services. This is ground to speculate that housing seekers will prefer, or have better luck in finding housing that they like in, this zone, which corresponds mostly to Baltimore City, Baltimore and Howard Counties, and part of Anne Arundel County.

At 25 miles and further out, virtually no transit service is available. The low population and housing unit densities (per squared mile of land area) suggest that this outer zone is a lot more rural than the city center, and would impose a big change in lifestyle for the prospective residents. Together with the expectation that it will probably cost people an extra hour of commute time, we may anticipate less success for HCVP in this outer zone.

4.3 Brand that Predicts Success for HCVP

Besides geographical factors such as county characteristics and distance, psychological factors such as residents' familiarity with transit services may also contribute to transportation barriers (or their perception). After all, fare cards and benefits may not be transferrable between agencies. Service areas of routes, conditions on transit vehicles or vessels, accessibility of transit stops, and many other aspects of transit service may also be different. Uncertainty in whether a potential user can depend on the new agency can be a strong disincentive. Due to the constraints of this study, we did not interview the voucher holders to obtain psychometric measures. Instead, we took note of the transportation information advertised on the HABC website. The MTA bus routes 1, 3, 5, 7, 8, 10, 13, 15, 19, 20, 21, 22, 23, 27, 33, 35, 36, 51, 62, 64, 91, 3X, 5X, 10X, 15X, 19X, and 64X were mentioned in the web pages that introduce HABC facilities. Hence we know that such services are available to the housing residents. Since most of the voucher holders would likely be familiar with or even use such bus routes, we speculate that they would also prefer to have access to them in the prospective locations. For convenience, we will call this group of bus routes "HABC Bus." Table 4 contrasts the service frequencies (as measured by maximum number of thousand times stopped) of HABC Bus against that of other "brands."

Table 4: Maximum Times Stopped (unit:1,000 times)

Transit "Brand"	*Harford*		*Baltimore*		*Baltimore City*		*Howard*		*Anne Arundel*		*Queen Anne's*	
	Non	*Allow*	*Non*	*Allow*	*Non*	*Allow*	*Non*	*Allow*	*Non*	*Allow*	*Non*	*Allow*
MTA—HABC Bus			42.2	52.9	457.8	43.2				0.3		
MTA—Others	0.2	0.1	83.6	36.8	146.9	12.4	0.6	0.6	7.6	11.6	0.0	0.0
Charm City					10.7	4.0						
Central Maryland							1.7	0.3	3.2	1.1		
Howard Transit							7.0	7.7	0.6	0.0		
Annapolis Transit									1.2	0.6		
BWI Airport									2.0			

HABC Bus makes trips with up to about 457,800 stops in the non-allowable areas in Baltimore City, but only about 43,200 in the allowable areas. A large proportion of transit services familiar to the HABC residents are no longer available in the allowable areas. However, when compared with the 12,400 stops made by other MTA services and with the 4,000 stops made by Charm City services, HABC Bus is still the most available service in the allowable area in Baltimore City. In Baltimore County, HABC Bus accounts for up to 52,900 times stopped, but in a smaller proportion when compared with the 36,800 times stopped made by other MTA services. If familiarity is the only consideration, Baltimore City would be the clear favorite. However, MTA also offers other services in Baltimore and Anne Arundel Counties. If these places are promoted, then it seems reasonable to speculate that the psychological barriers can be overcome by way of brand name recognition. The same argument can be made for Howard Transit and Annapolis Transit, except that they do not have as big an advantage as MTA.

5. Network Results

The transportation barriers considered so far have been quantified for specific locales. Such analysis implicitly assumes that people will face little or no barriers when moving between locales once they get on the transportation network— regardless of where they get on. In practice, freedom of movement on a network is constrained by travel time and network topology. The former has already been briefly discussed in Section 4.2. The topological issues at stake here are 1) whether existing services separate areas, and thereby prevent riders from moving between certain parts of the network, and 2) whether riders have topologically appealing options. The first question is a "connectedness" issue, meaning we need to check whether the transportation network supports transfers between routes, agencies, or modes of transportation. The second question is a transportation "connectivity" issue, meaning we need to check what choices the riders have at the stops. Interested readers may refer to Mishra *et al.* (2012) for a sophisticated demonstration that buses in Baltimore have good connectivity compared to buses in Washington D.C. In the rest of this section, we will construct a network graph from GTFS data, and then obtain some rudimentary measures to help address our questions from a high level. We also refer to Newman (2010) for network and graph theoretic terminologies.

5.1 Construction of Transit Network Graph

Consider in Figure 2 a hypothetical bus route defined by 4 stop sequences. In one direction, it runs on one roadway (or one side of the street) and has a set of stops. In the opposite direction, it runs on another roadway (or the other side of the street) and has a largely different set of stops. Some trips may skip some of the stops. In Figure 2, each vertex represents a transit stop; and each edge represents a trip link, which indicates movement between successive stops in a stop sequence.

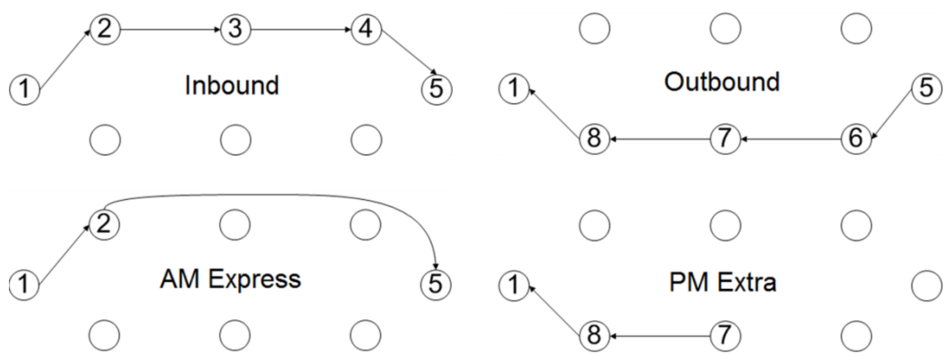

Figure 2: Trips (Stop Sequences) of a Hypothetical Bus Route

The network graph for this hypothetical bus route is then drawn according to the "**L**-space model" in Figure 3 by taking union of all trips upon removing any duplicated links. The graph is simple (meaning no multiple edges between stops) and undirected. Note that there is an edge between Stops 2 and 5, even though this trip link may geographically overlap inside the path connecting Stops 2 through 5. This rendition of **L**-space model is constructed from stop sequences to provide a more accurate description of passengers' experiences. It also allows the construction of network topology from time table data programmatically without any additional geographical de-duplication.

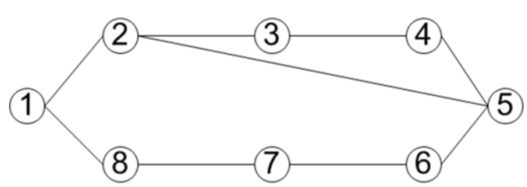

Figure 3: **L**-space Model of the Hypothetical Bus Route in Figure 2

5.2 Transit Network Connectedness

Since our network graph is undirected, network connectedness assessed from such a model would ignore direction of travel on a trip link. However, a typical transit route usually has stops on the opposite side of a street or platform. Hence, a rider can usually re-trace the route in the opposite direction along similar geographies. Even though transit vehicles and vessels move unidirectionally on the actual network, we may still use the undirected model to simplify computation.

The entire network is constructed by taking union of all transit routes. The overall result is shown in Figure 4. Vertices represent transit stops in Greater Baltimore and beyond, and are plotted approximately according to the stops' GPS coordinates. The edges indicate straightline distances between successive stops. The red edges, representing trip links inside the allowable tracts, are highly disconnected. Since people's movement will not be confined to only within the allowable tracts, it only makes sense to consider the entire transportation network as a whole. The blue edges represent trip links reaching out of allowable tracts, and show the potential places people can go to if they move there. Together with the gray edges (trip links entirely outside of allowable areas), the network graph gives a complete picture for the transportation system touching Greater Baltimore. The largest cluster near the upper right is located mostly within Baltimore City and Baltimore County. The cluster near the lower left touching the blue edges is located near Washington D.C.

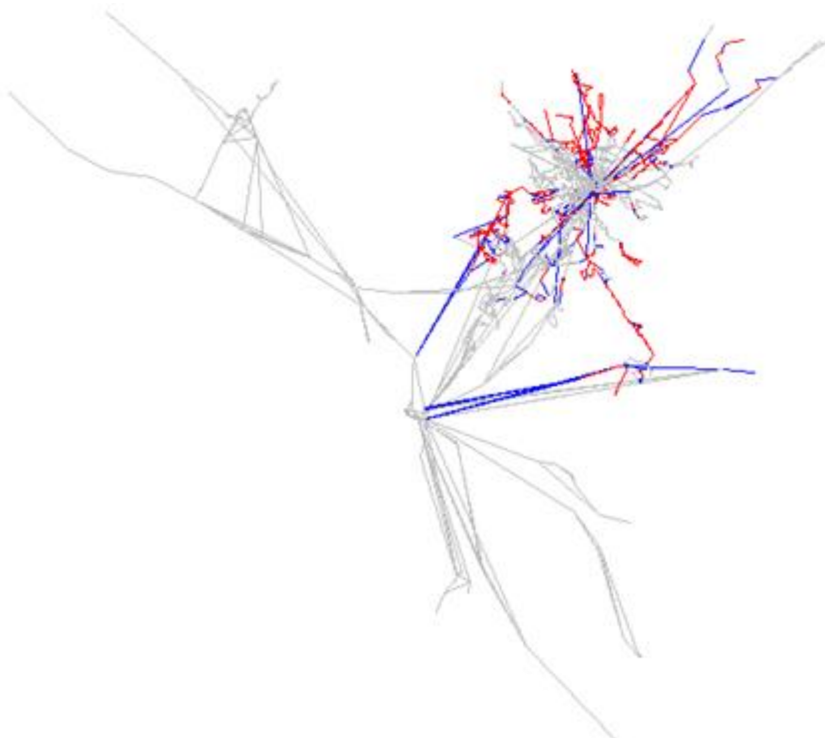

Figure 4: Network Graph of Transit Routes Touching Greater Baltimore

A further algorithmic verification confirms that, except for an experimental bus route "M" (Odenton Shuttle) of Central Maryland Regional Transit, all bus routes by all agencies considered so far have some overlapping stops. The rails and ferries have distinct stop numbers. But they all have stations and piers within walking distance from some bus stops. Thus, the entire transit network in Greater Baltimore can be considered "practically connected" albeit with some minor gaps.

5.3 Transportation Connectivity on Transit Network

One way to quantify transportation "connectivity" (as freedom of movement) at a transit stop is to count its number of out-going links. This measure corresponds to "average degree" in our network graph— the larger the average degree, the more possibilities for a rider to move to a different stop and, hence, the less transportation barriers. Note that graph-theoretic connectivity is defined by the minimum number of internally disjoint paths among all pairs of stops, and is different from the transportation connectivity being measured here. Also, average degree is more useful as a connectivity rather than connectedness measure. The graph of a transportation network must have an average degree at least 1, because any degree-0 vertex would absurdly represent a stop without any arrival or departure. According to graph theory, a connected simple undirected graph on $n \geq 2$ vertices must have at least $n-1$ edges and an average degree at least $d = 2 - \frac{2}{n}$. Thus, if a network graph has an average degree at least 1 and less than d, then it not only suggests a lack of transportation connectivity, but actually guarantees disconnectedness. The converse is not true in the sense that a disconnected network may still have an average degree at least d. (For example, a set of disjoint cycles has an average degree 2, but is disconnected.) Nonetheless, an average degree larger than 2 tends to suggest decent transportation connectivity (at least locally), even though it does not guarantee connectedness globally. Table 5 shows a summary of the network graph.

Table 5: Network Graph Parameters of the Transit Network Touching Greater Baltimore

Transit Agency	*Transpor-tation Mode*	*Vertices (Stops)*	*Edges¶ (Trip Links)*	*Average Degree*
Maryland Transit Administration	Light Rail	69	71	2.06
Maryland Transit Administration	Subway	26	32	2.46
Maryland Transit Administration	Heavy Rail§	79	123	3.11
Maryland Transit Administration	Bus	6,083	7,199	2.37
Central Maryland Regional Transit	Bus	417	456	2.19
Howard Transit	Bus	404	489	2.42
BWI Thurgood Marshall Intl Airport	Bus	4	4	2.00
Charm City Circulator	Bus	56	56	2.00
Charm City Circulator	Ferry§	3	2	1.33
Annapolis Transit	Bus	51	58	2.27
Frederick Transit Meet-The-MARC	Bus§	29	38	2.62
Rail Only		**174**	**226**	**2.60**
Bus Only		**6,840**	**8,195**	**2.40**
Ferry Only		**3**	**2**	**1.33**
Overall		**7,017**	**8,423**	**2.40**

§Weekday service only. ¶Undirected edge.

MTA heavy rail has an average degree larger than 3, suggesting that this mode of transportation probably has many express train trips jumping across stations. Airport and Charm City buses both have an average degree 2, implying that their routes probably follow loop-like patterns. The overall network has average degree 2.4, which is a moderate value for transportation connectivity. It shows that most of the stops either allow transfers between routes, or have various trips that skip stops on the same route.

This degree-based approach can also accommodate checking the so-called “scale-free property” (Barabási and Albert 1999). This property has been observed in many transit networks (Derrible and Kennedy 2011). It says that the order of transit stops linking to a certain order of other stops follows a power law pattern. The implication is that an ordinary transit network has relatively few stops connecting to many other adjacent stops, but relatively more stops connecting to only a few adjacent stops. While there is no universal standard on what degree distribution would indicate decent connectivity, this scale-free property provides some suggestive evidence for any anomaly in the degree distribution on a transit network. We expect an ordinary network to display a linear relationship between the order of degree frequency and the order of the degree. Degrees and their frequencies are plotted on a log-log scale in Figure 5.

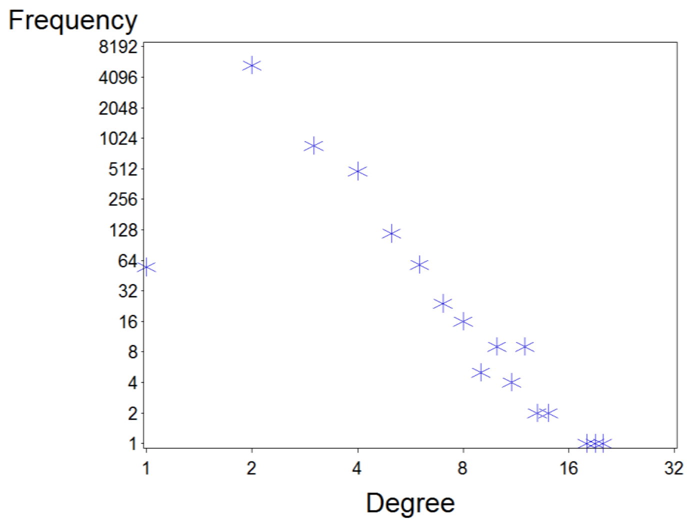

Figure 5: Degree-frequency Diagram of the Entire Transit Network Touching Greater Baltimore

A linear relationship is clearly present for most values except for degree 1. Derrible and Kennedy (2010) have shown that 14 out of 19 networks in their study displayed a power law patterns. So the anomaly observed here seems unexpected. However, we note that a similar "lack of fit" of degree-1 vertices has been observed in Los Angeles (von Ferber *et al.* 2009, Figure 4a), Beijing (Xu *et al.* 2007, Figure 2b), and throughout Poland (Sienkiewicz and Holyst 2005, Figures 4a and 4b). Since most bus routes run in loop-like patterns (due to two sets of stops on both sides of a street), there are fewer vertices of degree 1 in a network consisted of mostly bus trips than what the scale-free property requires. Thus, we may conclude that the transit network in Greater Baltimore is not out of the ordinary.

6. Conclusions and Outlook

We have shown how the GTFS data can be used to describe hot spots and cold spots in housing zones. The contrast between two such areas informs housing planners about where transit service promotion is needed. The GTFS data can also be used in various network analyses to further validate people's perception on any transportation barriers. We have predicted that Baltimore City and Baltimore County will likely be favored by the current public housing residents. Howard and Anne Arundel Counties can also be made appealing if properly promoted. Carroll, Harford, and Queen Anne's Counties will probably require a longer time period for voucher holders to adapt to lifestyle changes. Overall, the transportation network is connected and does not separate allowable and non-allowable tracts. It also has a decent degree distribution (hence, freedom of movement) comparable to many other transit networks.

Some of our analysis assumptions could have been changed without impacting the usefulness of our methodology. For city planning, socio-economic indicators such as income or history of residency with HABC may be used. More transit service statistics (such as service frequency, time of arrival, time traveled, or distance traveled) and other types of transportation barriers in addition to service-availability and network-connectedness (such as accessibility barriers due to public security concerns or disability compliance of transportation facilities) can also be considered. Our network has been modeled as a simple graph, giving no differentiation to the number of transit routes or the service frequency on a link. We can either use a multi-graph (permitting multiple edges between stops), or restrict to specific time periods if we are interested in spatio-temporal correlation rather than just network topology in general. For example,

Table 5 has included 148 (out of 8,423) trip links that are served on weekends only. While this is fine for capturing the maximum capacity of the network, we note that the network graph can be further restricted to only weekdays or even just the hours of interest to the analysis.

In summary, GTFS offers much flexibility to accommodate a wide range of housing and transportation network analyses, and hopefully will prove itself as a valuable data source for further research.

Acknowledgements

The authors would like to thank Ken Copeland and Steven Pedlow for comments that led to improvements in the abstract and main text.

References

Barabási, A.-L., Albert, R. (1999) "Emergence of Scaling in Random Networks" *Science* **286**: 509–512, October 15, 1999.

Burgess, E. W. (1924) "The Growth of the City: an introduction to a research project." Publications of the American Sociological Society, **18**: 85–97.

Burgess, E. W. (1925) "The Growth of the City" pp. 47–62 in *The City*, ed. by R. Part and E. Burgess. Chicago, IL: University of Chicago Press.

Derrible, S., Kennedy, C. (2010) "The complexity and robustness of metro networks" *Physica A* **389**: 3678–3691.

Derrible, S., Kennedy, C. (2011) "Applications of Graph Theory and Network Science to Transit Network Design" *Transport Reviews* **31**(4): 495–519.

Devine, D. J., R. W., Rubin, L., Taghavi, L. B. (2003) *Housing Choice Voucher Location Patterns*: Implications For Participant And Neighborhood Welfare. U.S. Department of Housing and Urban Development, Office of Policy Development and Research.

Glaeser, E. L., Kahn, M. E., Rappaport, J. (2008) "Why do the poor live in cities? The role of public transportation" *Journal of Urban Economics* **63**(1): 1–24.

Harris, C. D., Ullman, E. L. (1945) "The Nature of Cities" *Annals of the American Academy of Political and Social Science* **242**: 7–17.

Hoyt, H. (1939) *The Structure and Growth of Residential Neighborhoods in American Cities*. Washington, DC: Federal Housing Administration.

Khadduri, J., Struyk, R.J. (1982) "Housing Vouchers for the Poor*" Journal of Policy Analysis and Management* **1**(2): 196–208.

Metropolitan Baltimore Quadel (2012) "Allowable Census Tracts (Rev. 2/9/2012)". Baltimore, MD. http://www.mbquadel.com/AvailableProperties.aspx

Mishra, S., Welch, T. F., Jha, M. K. (2012) "Performance indicators for public transit connectivity in multi-modal transportation networks" *Transportation Research Part A* **46**: 1066–1085.

Newman, M. E. J. (2010) *Networks*: An Introduction. New York, NY: Oxford University Press.

Sienkiewicz, J., Holyst, J. A. (2005) "Statistical analysis of 22 public transport networks in Poland" *Physical Review E* **72**(4), Article 046127, 11 pages.

Varady, D. P., Walker, C. C., McClure, K., Smith-Heimer, J., Larkins, S. (1999) "Helping families move: Relocation counseling for housing-voucher recipients" *Netherlands Journal of Housing and the Built Environment* **14**(1) "Recent Housing Policies in the USA," pp. 33–59.

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

Xu, X., Hu, J., Liu, F., Liu, L. (2007) "Scaling and correlations in 3 bus-transport networks of China" *Physica A* **374**(1): 441–448.

Appendix: Allowable Tracts in Greater Baltimore Assuming Definition Changes

As an alternative, we can convert all Census 2000 block numbers into Census 2010 block numbers by using the Census Bureau's relationship file to resolve the area overlaps. A Census 2010 tract that covers more than a certain portion of allowable areas as defined by Census 2000 numbering may be classified as a full allowable tract; otherwise, it may be classified as non-allowable or partial allowable according to other thresholds. A possible conversion would look like Figure 6 below. County boundaries (rather than shorelines) are drawn as thick, black lines; and tract boundaries, as thin, gray lines. Four non-allowable tracts entirely on Chesapeake Bay that cover zero land area are colored white. Other tracts that are partially on water but cover some land areas are still colored according to their allowable status. Full allowable tracts re-defined with different Census 2010 tract numbers are colored dark red. They indicate areas that are allowable "geographically" according to the Census 2000 definitions, but have been assigned new tract numbers in 2010 and became out of sync with documentation available at the time of this study. We do not expect the changed definition would significantly alter the conclusions in the main analysis. Thus, any potential discrepancy is noted without further effort to reconcile.

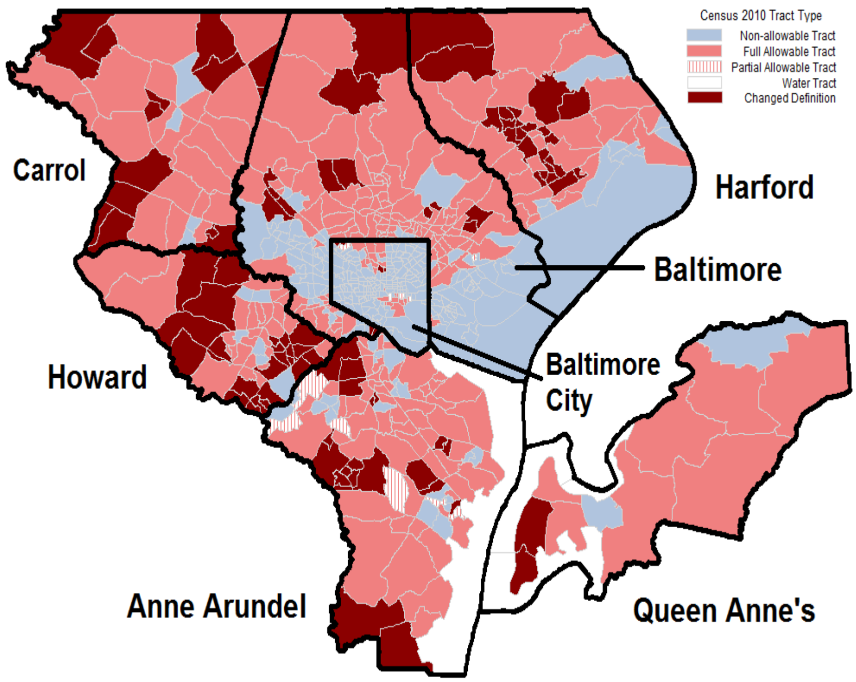

Figure 6: Census 2010 Tracts in Greater Baltimore (Assuming Definition Changes)