Mapping transit-based access: integrating GIS, routes and schedules

T. L. Lei & R. L. Church


To link to this article: http://dx.doi.org/10.1080/13658810902835404

Published online: 01 Mar 2010.

Article views: 1483

View related articles

Citing articles: 30 View citing articles
Accessibility is a concept that is not entirely easy to define. Gould (1969) once stated that it is a ‘slippery notion … one of those common terms that everyone uses until faced with the problem of defining and measuring it’. Considerable research over the last 40 years has been devoted to defining and measuring accessibility, ranging from access to jobs within an hour’s travel time to the ease at which given places can be reached. This article is concerned with the measurement of access provided by transit. It includes a review of past work on measuring accessibility in general and with respect to transit services in particular. From this overview of the literature, it can be seen that current methods fall short in measuring transit service access in several meaningful aspects. Based on this review and critique, we propose new refinements that can be used to help overcome some of these shortcomings. As a part of this, we define an extended GIS data structure to handle temporal elements of transit service. To demonstrate the value of these new measures, examples are presented with respect to mapping accessibility of transit services in Santa Barbara, California. Finally, we show how these measures can be used to develop a framework for supporting transit service analysis and planning.

Keywords: Public transit; accessibility; schedule and route information; geographic information systems; urban applications

1. Introduction

Accessibility is one of the most important concepts in the study of transportation systems, and providing access is without a doubt a major function of an urban system. Historically, access to jobs and city cores in the older cities of the United States was largely provided by public transit, and transit systems played a central role in shaping the form and scale of the urban landscape during the late 1800s to 1920s. Even after the advent of automobiles and highway expansion, public transit has remained an important alternative means of transportation to private cars. This is especially true in urban environments such as big cities and university towns where there are high concentrations of people. Furthermore, transit is widely deemed as a means for providing equal access, reducing congestion and alleviating environmental problems caused by the use of the automobile. However, transit ridership will remain small if the level of accessibility provided by public transit is low relative to that of the automobile. Improving public transit is linked to a number of factors such as improving service, increasing access and promoting...
greater safety while keeping costs to a reasonable level. Therefore, it is important to measure the level of accessibility provided by transit alternatives in order to help support the process of transit planning and decision making. For example, improvements to the level of service and access can be used as a basis for justifying and making choices in public transit investment. This article addresses the issue of measuring access provided by a public transit system.

Conceptually, accessibility is a measure of the ease with which people can reach their destinations or activity sites (Dalvi 1978). Although accessibility is a key concept in transportation, there is no consensus on how to operationally measure it. Researchers have investigated the measurement of accessibility from various perspectives such as access for those with disabilities (Church and Marston 2002), gender differences in access (Kwan 1999), access to jobs (Hanson and Schwab 1987, Kawabata 2003), access to city centres (O’Sullivan et al. 2000) and access to shopping and health facilities (Lou and Wang 2003). Compared with other modes of transportation, public transit influences accessibility in unique ways. In addition to important factors such as a user’s time budget and socioeconomic characteristics, travel by transit depends significantly on the routes (e.g. a direct route vs. a so called milk-run route), schedule (frequent vs. infrequent), the location of the user as well as the time of day the trip is made. Some of the past work (see, e.g., Hillman and Pool 1997, Murray et al. 1998, O’Sullivan et al. 2000, Gan et al. 2005) has focused on access provided by public transit. In the next section, we provide a review of access metrics in general and make specific comments associated with the use of these metrics within the context of public transit. Following this, we discuss shortcomings in existing metrics as well as how these metrics can be improved for use in modelling transit access. The metrics developed in this article represent an improvement over previous work in that they capture certain essential aspects of transit access such as the variation of access associated with time of day as well as travel direction. In particular, time tables are explicitly used and some of the commonly made simplifying assumptions are relaxed. In section 3, we discuss how such metrics can be used to enhance a GIS environment in terms of modelling access using routes and schedules. We also provide examples that have been developed in analyzing transit services in Santa Barbara, CA. In section 4, we provide details on how these techniques can be used to develop a framework for supporting transit service analysis and planning that can measure the impact on access associated with proposed system changes in meaningful ways. Finally, we conclude with a short summary and recommendations for future work.

2. Literature review

Accessibility can be generally defined as the ease with which people can reach their opportunities or services (Wachs and Kumagai 1973). It embraces three main elements – people, linkages and activities (Moseley 1979). A large number of measures of accessibility have been proposed. This can be attributed to a variety of reasons, ranging from the level of aggregation (Handy and Niemeier 1997) to the underlying behavioural model (Morris et al. 1979). In the following discussion, we have grouped accessibility metrics into six categories. This classification scheme helps to understand past work and how it relates to measuring transit access as well as the role GIS has played in transit service access.

The first category of access metrics is called system accessibility. This class of metrics deals with physical access to a system, based on the distance, time or effort
to reach a system network or a set of access points associated with that system. For example, Nyerges (1995) used GIS to analyze transit coverage in the Queen Anne Community of Seattle. He generated quarter-mile buffers around transit routes using GIS in order to identify the streets served by the current system. If a street intersects with one of the buffers, then people along that street were considered having good access to the transit system. Using this map, Nyerges proposed adding additional routes to serve ‘unserved’ streets. Similarly, Azar et al. (1994) assumed that employees who lived within a quarter-mile buffer of any transit line had good accessibility to medical institutions in Boston. For measuring access to point facilities, Aultman-Hall et al. (1997) measured the average and maximum walking distance for residents to local destinations such as school and transit stops with a distance standard of 400 m to evaluate pedestrian accessibility of local neighbourhoods. Hillman and Pool (1997) have discussed a software system for measuring accessibility for local transit operators in London. The authors calculated a local accessibility measure widely used in the United Kingdom, which measures the proximity of users to transit stops. Gan et al. (2005) developed a program called the Florida Transit Geographic Information System, which is designed to determine those areas that are transit accessible, as well as to calculate the proportion of a service region served by transit. Using demographic data, their system can identify areas that are underserved by transit but meet minimum levels of housing and employment density. Polzin et al. (2002) have developed a tool to calculate access to a transit system based on calculating the differences between the spatial and temporal dimensions of demand and the spatial positions of stops along with route headway times. Thus, they have expanded the definition of physical access by time of day. They assume that if headway times are large, then a certain proportion of people will be unwilling to wait for service, thus reducing access. Murray et al. (1998) and Murray (2001) proposed reconfiguring a transit system based on a physical access metric, by relocating bus stops in order to maximize the number of people who lived within a given distance of a stop. In further work, Matisziw et al. (2006) develop a model to extend routes and maximize access with the location of stops along the route extensions. While proximity to transit stops or routes may have an important impact on travel cost and choice, system accessibility measures do not consider the travel cost incurred in using the transit system to travel to a desired destination (Murray et al. 1998). Thus, such an analysis may significantly underestimate the travel cost as well as exaggerate the level of access provided.

The second accessibility metric is called system facilitated accessibility. As compared to system accessibility, system facilitated accessibility measures a user’s ability to get to their destination and takes into account the travel time or cost spent in the transportation network or the associated effort in making the trip. How the travel cost is determined depends on the availability of data and the application context. For example, Liu and Zhu (2004) calculated transit time by dividing distance with average speed(s). More complicated models take into account factors such as transfer time, wait time and schedule information. Hillman and Pool (1997) used realistic travel times from origins to destinations in order to calculate system facilitated access. They described the procedure for calculating the travel time as summing up the walk time from the origin to a stop, the waiting time at the stop, the onboard travelling time and waiting time at any interchanges. They reported using detailed information about the public transport network, including timing points along a route to figure out the cumulative times for stops. Wu and Murray (2005)
have developed a model to optimize an existing route structure in order to optimize system access and system facilitated access by determining which existing stops can be dropped. Thus, their model tries to streamline an existing system to improve transit times while reducing system coverage as little as possible.

O’Sullivan et al. (2000) developed a shortest path model based on the Dijkstra algorithm that identifies the least cost path in terms of travel time from any origin to a desired destination. They used a multimodal network of bus and rail. It was assumed that transfer times between different bus routes or between bus and rail equalled one-half of the headway time (i.e. this assumes that people on the average arrive at a stop or make a transfer halfway between the previous bus/train and the next bus/train). Travel times along a given transit route were based on published schedules. They used their shortest path model to estimate the time to travel to a central business district and mapped travel time isochrones.

Using a similar method for calculating travel time, Gent and Symonds (2004) proposed identifying areas that are within a 60-minute catchment area of a city centre. The work of Liu and Zhu (2004), O’Sullivan et al. (2000), Hillman and Pool (1997) and Gent and Symonds (2004) are important as they focus on the ability of the system to take the individual to a specific destination (such as the central business district). Unfortunately, most of this previous work is based on the assumption that travel and service times are the same regardless of time of day and day of week. To address this issue, Peng (1997) developed a trip planner, which uses schedules and routes, so that a user can find the quickest route to get to a desired destination when leaving at a specific starting time. Trip planners are now available on the Internet for many cities, allowing people to calculate the best transit route from a starting point to a designated destination. Although trip planners have been developed for many transit agencies, they have not been used to generate isochrones-based maps for a region, as they are principally designed to depict the best way to use a system to accomplish a specific trip.

The third measure of access is called integral accessibility. Whereas the first two types of metrics involve access to a network or access provided by a facility to travel to a destination, this third category is associated with calculating a measure of overall access associated with a number of possible destinations. The simplest integral measure is a count of the number of opportunities of some type of activity within a reasonable travel distance or time of a particular location (Wachs and Kumagai 1973, Talen and Anselin 1998). We can define this mathematically as

\[
A_{ik} = \sum_{j \in M_{jk}} O_{jk}.
\]

where

- \(i,j\) = indices used to represent locations
- \(k\) = index used to represent activity type
- \(A_{ik}\) = measure of overall access for location \(i\) with respect to activity type \(k\)
- \(d_{ij}\) = the distance between location \(i\) and location \(j\)
- \(s_k\) = the maximum distance or time in which a user is likely to consider activity type
- \(M_{ik} = \{ j \mid d_{ij} \leq s_k \}\)
- \(O_{jk}\) = the number of activities of type \(k\) available at location \(j\).
Counting available activities within a maximum travel distance or time does not depict relative closeness. However, one can extend the concept by depicting the aggregate accessibility for a series of cut-off distance values. Such a plot of total number of opportunities as a function of $s_k$ is called a location profile. Gertman and Ritsema van Eck (1995) and de Jong and Ritsema van Eck (1996) developed a GIS application to compute location profiles for selected areas. Ingram (1971) defined relative accessibility as the degree to which two places are connected and integral accessibility as the degree of interconnection of one place to all other places. Using distance as a surrogate for relative accessibility, Ingram defined this mathematically as

$$A_{ik} = \sum_{j \in M_k} d_{ij} O_{jk}$$

For this measure, the larger the value, the less accessible activity $k$ is to location $i$. If the above equation were divided by the sum $\sum_{j \in M_k} O_{jk}$, then access would be measured as the average distance to a given opportunity type $k$.

A more common way of describing integral accessibility is related to the trip generation concept and represented by the classic gravity model. Hansen (1959) was the first to propose discounting activities using a gravity-based model. Shen (1998) has made a further refinement of the gravity model by taking into account the demand side in modelling the competition between job seekers who are attracted to the same employment centre. This modified gravity model has been used by a number of researchers to measure job accessibility for low-income workers (Kawabata and Shen 2006), to evaluate access to local parks (Omer 2006) and to evaluate the accessibility impact of a proposed new rapid transit line (Liu and Zhu 2004). In terms of model specification, Lee and Goulias (1997) have presented an interesting approach to calibrate the parameters of a gravity model using data from a travel survey and evaluated the importance of accessibility in determining home-based shopping trip frequency. Thompson (1998), as another example, estimated each component of the gravity model, i.e. attraction, friction as well as the demand for transit based on a range of socioeconomic variables at the census tract level.

The fourth category of accessibility measurement is based on the notion of space–time geography (Hägerstrand 1970). This approach is based on the recognition that a person’s movement over space and the choice of activities is dependent on one’s mobility and limited by one’s time budget. Miller (1991), Miller and Wu (2000) and Kwan and Hong (1998) have derived space–time accessibility measures by using network distance/time derived from GIS network data.

Kwan (1998) compared space–time accessibility measures and integral accessibility measures (gravity and cumulative opportunity type) and showed that space–time
accessibility measures are statistically distinct from integral accessibility measures while integral accessibility measures are similar within its group. Kim and Kwan (2003) presented a space–time accessibility model that accounted for facility operating hours and transport network properties such as one-way streets and turn restrictions. Accessibility was calculated as the sum of accessible opportunities weighted by available service time. Kwan (1999) has also evaluated differential access by gender due to factors other than distance friction (e.g. time constraints). For a further overview of the space–time accessibility literature, the reader is referred to Kwan et al. (2003).

The fifth category of accessibility metrics involves the use of utility theory. Such metrics are based on viewing transport users as consumers and alternatives of travel as the choice set. The consumer is assumed to be rational and chooses the alternative with the maximum utility, which is dependent on the characteristics of the user, attributes of the transport options such as travel time, monetary cost and comfort, and the properties of the activity site. Koenig (1980) defined utility of an individual for visiting activity site from an origin as

\[ U_{ij} = V' - C'_{ij} \]

where \( V' \) is the gross random utility for visiting the activity site and \( C'_{ij} \) is the generalized travel cost. Assuming non-random generalized travel cost and using logit-like models, Koenig showed that the maximum utility on the average is \( U_i = \log A_i \), where \( A_i \) is a gravity-type accessibility measure with exponential distance decay. Generally speaking, all approaches could be viewed from a utility theory perspective when accessibility is treated as a proxy of the potential for spatial interaction. The previous categories of accessibility metrics emphasize the influence of distance, time or space–time conditions on travel. Utility-based methods, however, typically consider a wider range of variables. For example, Rastogi and Rao (2002, 2003) developed an accessibility index to measure access to transit stations in India using a logit-based utility model. They considered random generalized travel costs and estimated them using socioeconomic variables such as household income and mode availability (e.g. walk, bicycle and bus) and other variables such as distance to transit stations and environmental impacts of each mode. They also collected modal split data as well as stated preference for modal shift under hypothetical policy changes. Based on these data, they were able to derive an accessibility index reflecting the perceived utility of reaching the transit stations as a function of the input variables.

The final category of accessibility measures, relative accessibility, is based on comparing access between modes or types of users (Church and Marston 2002). If a consumer has a choice between using a personal vehicle and using public transit in travelling to a destination, the choice is often made as a function of cost, time, convenience and safety. When parking costs are high and transit times are competitive to the time needed to drive a personal vehicle, many will choose to use transit. That is, the choice to use transit for many people is based on the relative value of transit as compared to another mode. Thus, it makes sense to calculate transit accessibility in a form that considers cost or time relative to other modes. A classic example of such a measure is the travel time ratio used in constructing diversion curves for modal split (see for example, Sheppard 1995).

Each of the six types of access metrics can provide useful information in designing, analyzing, and managing a transit system. For example, the
Transportation Research Board report on Bus Routing and Coverage (Pratt and Evans 2004) lists walk time and transfer times as having the highest relative importance in terms of transit travel time components. There is also a distance limit beyond which most people will not walk to a bus stop. Thus, physical access to a system is an important metric for assessing transit system performance. In fact, virtually all of the types of metrics discussed above have been used in analyzing transit services. It also makes a great deal of sense to make transit system measurements within a GIS, as this allows for better demographic and geographic analysis. Because of this, GIS has been used in a number of instances to analyze transit access and service since the late 1990s. There are four good examples which should be mentioned here. The work of Liu and Zhu (2004) is a good example of this trend. They developed a GIS application tool called the Accessibility Analyst for transit analysis which was designed to measure system accessibility, system facilitated access, integral accessibility, system utility and time-constrained access but not relative accessibility. O’Sullivan et al. (2000) also provides a very good example of integrating a shortest travel time algorithm with geographic data on bus and rail to determine transit facilitated access to a downtown area. Their application presented travel time data in terms of time isochrones, representing the time it takes to travel to the city centre. Gan et al. (2005) have also developed a special GIS application to analyze transit systems in Florida. This system is designed to identify those areas that have physical access to transit services, calculate the proportion of people who have such services, as well as identify whether any demographic strata is not well served. Finally, Peng (1997) devised a GIS-based system to give route planning advice to system users. Even though the past research is quite substantial, there still exist shortcomings, especially in measuring system facilitated access.

In order to analyze many of the metrics, besides physical accessibility, it is necessary to estimate the time it takes to make a trip using transit. Many projects involving GIS to estimate access use distance as a proxy for access time or use distance divided by average travel speed (Ritsema van Eck and de Jong 1999, Liu and Zhu 2004). This is not surprising as many of the metrics were originally specified in distance rather than time. The reason why this is often done is that many GIS applications provide a shortest distance algorithm that can be applied to find shortest network distances without needing to develop specialized codes or requiring speed or velocity data. Unfortunately, transit times can vary considerably over a network depending upon schedules, transfer times, etc., so using network distance as a proxy for time may introduce considerable error. There are notable exceptions from ignoring time altogether. For example, Liu and Zhu (2004) use a shortest path algorithm to generate the length of the shortest path from an origin point to a destination. This path length is comprised of three elements: the travel distance to the network, the travel distance along the network and the travel distance from the network to the destination. Distances are converted to travel time by dividing each distance component by an average travel speed for that component. This approach does not explicitly handle waiting times at stops, transfer times between bus routes or varying headway times along routes depending on time of day, etc. To do so would require integrating timetable information on the network and handling transfers explicitly. O’Sullivan et al. (2000) recognized this problem and used timetables explicitly in their work. They solved a series of shortest paths, which involved the walking time to bus or rail stations, estimated route times, possible transfer times, and then the final walk time to the destination. For their defined
network, they found shortest travel times, involving either bus or rail or both in
travelling to the downtown area. Their application then presented travel time results
as a set of isochrones representing the time to get to the city centre. Unfortunately,
they did simplify the task of finding the quickest routes. For example, they took the
time taken for an entire route from the bus schedule and divided it by the length of
the route to derive a route speed. Then the time taken along a portion of a route was
calculated as the distance for that part of the route divided by the average speed of
the whole route. Note that this is an accurate method only when a bus travels at a
constant speed throughout an entire route, which is often not the case. Another issue
of possible error is that the waiting time in making a transfer was assumed to be
equal to one half of the headway time associated with the next boarded route. Such
simplifications help to ease the burden of making an application but may
underestimate the time needed to make a trip. This type of model also neglects
the time of day a trip may be made or the day in which a trip is made.

The simplifications used in previous applications of GIS and accessibility analysis
are unacceptable if such a system is to be used as a route guidance tool. Peng (1997)
discusses how a GIS-T application can be developed to provide route guidance.
Peng’s objective was to identify the optimal trip itinerary for a traveller leaving the
origin at a specific time and reaching the destination. To do this requires knowing
when the traveller plans to depart, using all route tables explicitly, etc. This is the
basis for most route guidance systems that are available today, including some web-
based applications made available by transit districts. However, it is interesting to
note that route guidance models have not been used as a basis for building an
accessibility analyst tool. The main reason for this is that many have thought that it
is not necessary to address the nuances of daily schedules when estimating access.
Unfortunately, such explicit detail is necessary in modelling facilitated access or
integral access in transit planning. Since many transit trips are made in order to get
to work or to school in a timely fashion, the times in which a route is available may
significantly impact use and utility. Thus, it makes sense to build an accessibility
analysis tool that is based on a traveller route guidance system approach. In fact,
this was one of the main objectives of this study. In the next section, we describe how
this was accomplished as well as how specific types of accessibility metrics can be
modelled.

3. Modelling transit services and access: the Transit Accessibility Planning Analyst

From the outset, it should be stated that many of the GIS-based tools that have been
developed to analyze accessibility address important issues. Many of the past efforts
have been concentrated at analyzing system access or average facilitated access. In
many circumstances, such tools address the problems of concern. However,
ridership decisions are based on the individual, and thus it is important to also
view transit services from the perspective of the potential rider.

Each potential user of a transit system makes a decision on whether to use transit
based on a number of criteria, including cost, access and service (Pratt and Evans
2004). We can often divide transit ridership into two groups: those who do not have
a choice (often called captive users) and those who do have a choice as to whether to
use the system or not (Beimborn et al. 2003). In either case, it is desirable to keep
levels of service as high as possible. This includes two main features, geographic
access to services or what is referred to as service area coverage and level of service
(travel time, safety, frequency, etc.). Designing a public transit system is often based
on striking a balance between system access (geographic coverage) and system facilitated access (travel times). Both types of access may vary over time of day and day of week depending on the design of a system (which is often budget constrained). Without addressing time in an explicit manner, it is impossible to model variations in service access as well as facilitated access.

Developing an explicit time model requires the use of transit schedules, a geographic road network and geographic locations of route stops/stations at the very least. Routing a specific trip on a transit system requires the use of a shortest path algorithm. For our work, we employed the Dijkstra algorithm as this is known to be one of the better performing algorithms in the literature (Zahn and Noon 1998; Zeng and Church in press). Given $X, Y$ coordinates for both the origin and destination of a trip as well as a departure time, the goal is to identify the most efficient path (i.e. path of least time in using transit). First, an entry point to the road network/street network is identified. This is the point on an arc that is closest to the $X, Y$ coordinate of the origin. The exit point from the network is defined as the point on an arc that is closest to the destination point. It is assumed that the traveller will walk directly to the network, enter at the network entry point and eventually leave the network at the exit point. The shortest path algorithm is used to identify the least elapsed time path from the network entry point to the network exit point, assuming that the traveller has arrived at the network entry point at a time based on the departure time plus the time to walk to the network entry point. Using the transit system requires the traveller to walk to a transit stop from the entry point. The traveller can then use any transit vehicle leaving that stop after the traveller arrives at the stop. This means if it takes 7 minutes to walk to a transit stop after leaving at 7 am and the transit vehicles depart every 10 minutes (7 am, 7:10 am, 7:20 am, etc.), then the traveller will need to wait 3 minutes for the next transit vehicle. Thus, waiting times are inherently built into the route. It should be noted that the closest transit stop may not be the timeliest, and the search for the least time path is not constrained to go to the closest stop. That is, the transit stop choice is made based on the one that yields the quickest feasible journey to the destination, including the walk times to arrive at the appropriate transit stop. Transfer times involve the elapsed time from when one arrives at a stop on one transit vehicle and the departure time of the transit vehicle associated with the desired transfer. Thus, the algorithm identifies the route with the shortest elapsed travel time possible when leaving an origin at a given starting time. This process represents the first type of shortest path algorithm that was developed for the Transit Accessibility Planning Analyst. We will call this shortest path routine A.

The implementation of routine A is not entirely straightforward within the context of integration with a GIS. The basic difficulty is that shortest path routines are designed to solve for the optimal path along a network without time constraints or timing issues. The fact is that a street segment that is served by a bus does not have a continuous operating tram service but intermittent service. That is, arriving at a link whose endpoint represents a bus stop does not guarantee immediate service. To account for scheduled service, access times for a route segment are stored as link attributes. Thus, the algorithm must check the arrival time at an arc to detect if a ‘wait’ for the next bus is needed. This approach is a bit novel. An alternative approach would be to develop a network of transitions and times (called time-expanded network) (Choi et al. 1988) and apply a shortest path routine to find the path taking the least time. But this network itself contains a temporal dimension.
that would be many times larger than the original network. Modifying the algorithm instead and storing time attributes along links allows the original network in GIS to be used rather than a temporally structured network of considerably higher dimensionality. Further, integrating bus times can be easily accomplished by developing an automated process to add attributes to each link, based on a published transit schedule.

The Dijkstra algorithm is designed to find the optimal path from a given origin to all possible destination points. Thus, the procedure described above could be used to calculate the time to travel from a point of origin, starting at a specific time, to all possible destinations. This type of model is useful in analyzing the difficulty in travelling from a major activity site, like an employment centre, to areas of a city or region. This information can then be used to generate an isochrone-based map of elapsed travel times in leaving the centre at a given time (say 5 pm) and arriving at locations away from the centre. For example, suppose that an employer wants to encourage employees to use transit. Employees could then use the map to identify the approximate time it would take to get to their home using a combination of transit and walking when leaving their job at 5 pm. This type of result could also be useful for setting flexible work schedules so that individual employees take advantage of the times when they are best served by transit.

As an example of using routine A, we have developed a GIS-based application using elements of ArcGIS™ and the specialized shortest path algorithm described above. Using geographic data for the Santa Barbara area of California and the routing and timetables of the Santa Barbara Metropolitan Transit District (SBMTD), we generated a map depicting the time it would take to leave the University of California, Santa Barbara (UCSB) campus, and arrive at virtually any destination along the south coast of Santa Barbara County. Figure 1 depicts the map that was generated when the specific departure time was set at 5 pm. The origin for all trips (UCSB) is marked by a triangle on the map and thicker lines represent routes of the current transit system. The value of each point on the map is the time

![Figure 1. Travel time for leaving UCSB at 5 pm and traveling to each location on the map.](image-url)
that it takes to travel to this location by transit, and travel times are capped by 120 minutes assuming that longer trips are not practical for frequent users. Note that the downtown area in the middle part of the map are rather well served by bus as there exists an express bus connecting the downtown area to the university campus that has a limited number of stops and makes most of the trip on two highway segments. Also, note that there are many areas of the south coast region served by SBMTD that cannot be reached in 45 minutes from the university using public transit, while most if not all of the region can be driven to in less than 30 minutes.

The above shortest path routine addresses only part of determining transit facilitated access times. The fact is that many users need to arrive at a given time and therefore need to adjust their departure time in order to get to a destination at or before that given fixed time. In analyzing transit facilitated access, it is necessary to be able to calculate when a person must leave a given location so that they can arrive at a destination by a desired arrival time. This type of analysis is becoming more common in on-line transit route guidance systems for individual trip planning but has not been developed for regional accessibility modelling and mapping. To do this requires a different type of shortest path routine. This can be approached by modifying the Dijkstra by starting at the destination. The objective for the routine is to label the network backwards in terms of decreasing latest arrival time (LAT) until the origin or starting location is reached, thereby identifying the path which can take the smallest amount of elapsed time and reach the desired destination by the fixed arrival time. For example, if the desired arrival time is 8 am and it takes 7 minutes to walk from a nearby transit stop to the destination, then in retrospection the latest time by which the user should arrive at this stop is 7:53 am. Therefore, this stop can be labeled with a LAT of 7:53 am assuming it is the fastest to walk from this stop to the destination. Again, if a transit vehicle arrives every 10 minutes (7:40 am, 7:50 am, etc), then the latest time the user could have got off the bus is 7:50 am. Subtracting the transit time that it took the user to arrive at this stop, one can then propagate the label for required arrival times backwards to the previous stop. A best (latest) temporary label can be made permanent at each step as is in Dijkstra’s algorithm and in Routine A, and the routine terminates when the origin of the trip is permanently labeled. Therefore, transfers and wait times are handled in a manner similar to routine A, except they represent delays based in terms of pushing back the LAT. This second shortest path algorithm is called routine B.

Routine B can be used to develop another version of a travel time isochrone-based map. Here, points along an isochrone would represent locations that share the same level of closeness to the destination, in terms of the time that one would need to leave a given point of origin in order to arrive at the destination at a desired time. Such a map would be useful in that it would allow employers as well as transit planners to identify those areas that the transit service provides good access for employees in order to get to work at a specific time. Figure 2 depicts a map of accessibility to the university based on arriving at 8 am. Travel times represent the amount of time before 8 am an individual would have to leave an origin, walk to a transit stop and take the most efficient trip to the university and arrive by 8 am. Note that there is also a large area of the south coast region that would require a trip of 45 minutes or longer in order to travel to the campus and arrive at or before 8 am. It can also be noted that travel times in Figure 2 are different from those of Figure 1 in certain areas as the direction of the trip as well as the transfer and waiting times, etc. are different.
Using ‘elapsed-time’ maps generated by routine A and ‘latest-departure’ time maps generated by routine B, one can begin to analyze the ability of a transit service in providing access in travelling to a given destination and then returning to the starting location, in terms of both trips (going to and arriving from the desired location) for specific time periods (e.g. arrive by 8 am and leave at 5 pm). For example, we define accessibility to destination $j$ from an origin point $i$ as follows:

\[
\bar{A}_{ij} = \frac{t_{ij}(\text{arrivaltime}) + t_{ji}(\text{departuretime})}{2}
\]

where

- $t_{ij}(\cdot)$ = the travel time from $i$ to $j$ for a desired arrival time at $j$ using transit
- $t_{ji}(\cdot)$ = the travel time from $j$ to $i$ with specified departure time at $j$ using transit
- $\bar{A}_{ij}$ = the average time in travelling to a given destination $j$ by a given arrival time in the first trip and returning at a given departure time in a second trip.

This type of accessibility measure is an extended form of that used in O’Sullivan et al. (2000) in that it takes into account travel time variances associated with both direction of travel and times of day (within the context of the time that one needs to arrive by and is available for departure at). For a large employer or business park, a map based on this metric depicts the attractiveness of using public transit in terms of a set time to arrive at work and a set time to leave work. If one were to consider flexible arrival and departure times, then one may want to consider what the best average trip time would be over a number of possible arrival and departure times. For instance, let us say that we identify several alternative travel time windows, e.g. in by 8 am and leave at 5 pm, in by 7:30 am and leave at 4:30 pm, in by 8:30 am and leave at 5:30 pm, etc. Then we can identify for each home location which time window yields the lowest average travel time in going to and from work. This can be
expressed as follows:

\[ A'_{ij} = \frac{\min\{t_{ij}(WA_r) + t_{ij}(WD_r)\}}{2} \]

where

- \( r \) = an index representing a given pair of departure and arrival times
- \( WA_r \) = the arrival time associated with the \( r \)th pair of arrival and departure times
- \( WD_r \) = the departure time associated with the \( r \)th pair of arrival and departure times
- \( A'_{ij} \) = the best combined average trip time over a number of possible arrival and departure time windows.

Figure 3 depicts a map of the best average travel times associated with the University as the destination, involving three possible pairs of arrival and departure times [(7:30 am, 4:30 pm), (8:00 am, 5:00 pm), and (8:30 am, 5:30 pm)] at each origin location on the map. Compared with Figure 1 and 2, Figure 3 shows a larger area one can reach within a fixed amount of time indicating that flexible time windows offer shorter travel times for transit users. Thus, it is important to develop maps using sets of departure and arrival times, so that employees may identify those times that transit may best serve them and attempt to negotiate flexible work hours. It can also be noted that in major areas of the region, travel from home to the university still cannot be made using transit in a timely manner.

The types of accessibility maps given in Figures 1–3 represent an improvement over past work. They are sensitive to both travel direction and time of day in addition to the fact that timetables are used explicitly rather than assuming an

Figure 3. Best average round-trip time for possible travel time windows: 7:30–4:30, 8–5 and 8:30–5:30.
average speed and assuming that wait times and transfer times are equal to half of the average headway time.

In large urban areas, headway times can be quite small and average 3 to 5 minutes, but in suburban communities headways might be as high as an hour. Given this range of possible times, it is necessary to ensure that this assumption is not made unless all headway times are always small. Eliminating such errors in estimating travel times is an important need in transit planning and analysis. There is also the need to understand the role of transit in supporting a wide variety of trips. For example, for the Santa Barbara area, we developed a list of eight major areas of the city, classified in terms of specific needs. This list included the Camino Real Market Place (a big box retail centre, Point 8 in Figure 4), Calle Real Market centre (a popular suburban strip retail centre, Point 6), La Cumbre Plaza (a shopping mall with major tenants, Point 3), the University (a major employment centre, Point 1), Castilion drive (a central location for a number of high-tech companies, Point 5), Cottage Hospital (a major centre for medical services, Point 7), the Santa Barbara courthouse (a point which represents the location of a number of public services as well as the courts themselves, Point 4), and the transit centre of SBMTD (a point close to a variety of stores, restaurants, etc., Point 2). We can develop a composite map of access by expanding the definition given above as

\[
A_i^{''} = \frac{\sum_{j \in M} t_{ij}(X) + t_{ji}(Y)}{2n}
\]

where

- \(X, Y\) = a pair of arrival and departure times
- \(A_i^{''}\) = the composite average accessibility in travelling to and from a variety of popular locations
- \(n\) = the number of destinations considered
- \(M\) = the set of destinations \(j\) that are considered in the analysis.

![Figure 4. Composite access time for eight destinations in the south coast of Santa Barbara.](image-url)
Given arrival and departure times \((X, Y)\), this calculation would represent the average time needed in travelling and returning from each of these destinations. The composite average travel time is representative of the average time to travel to popular destinations. This reflects the ability of the transit service in supplying many possible trip purposes as well as an average geographic access to a number of areas of the Santa Barbara area. Figure 4 gives a map of the average travel time (to and from) for the eight destinations for any point in the south coast region. Here, it can be seen that many of areas of the region are more than 45 minutes travel time away. Since the driving times to many of these locations in a personal automobile are less than 15 minutes, it can be seen that transit may provide a reasonable level of service coverage but not necessarily good levels of system facilitated access.

Back of the envelope comparisons of service (transit vs. auto) may be useful but should not necessarily be relied upon in making good estimates of the level of transit service. What is more helpful is to compare accessibility values. That is, to calculate accessibility values for transit and personal automobile and then compare the two. This suggests the use of a relative access measure (the last category of accessibility measures that were defined in the previous section of this article). Relative accessibility measures are not commonly used in transit accessibility analysis; however, they may form an important component of modelling mode choice for travellers. In previous work, Church and Marston (2002) have demonstrated the value of measuring access on a relative basis where they focused on people with disabilities. It seems somewhat obvious that transit service times relative to other forms of travel will in part dictate whether non-captive riders will choose to use transit. Past work (Barber 1995) has shown that the faster transit is relative to personal vehicles, the higher the percentage of people choosing to use transit. Thus, accessibility analysis should involve making relative measurements.

In a recent survey of those travelling to and from the university using a personal car, we obtained estimates of their commuting times. Although we did not ask for a specific household address (to keep their identity secret), we did ask that they give their addresses with the number rounded off to the nearest 100. Thus, we had approximate locations for each of the commuters using their car along with the time that it took to make the trip. The survey also entailed collecting detailed information in terms of the time of day for each day of the week that the commuting trip was made. In order to keep the survey to a manageable size, we asked for their average commute time. We geocoded virtually all addresses from the survey and created a map of self-reported commute times for those people who used their automobile. From this map of car driving times, we were able to compute a relative accessibility map of bus versus automobile by dividing estimated transit commuting times by auto travel times. The map of relative travel time ratios is given in Figure 5 for areas that the survey covers. Note that in most areas, automobile travel is quite fast in comparison to using transit. In fact, there are many areas in which the time to ride transit is four times or longer than what it would take to drive a personal automobile. This means that transit facilitated access is too slow to be a meaningful alternative for these areas. However, in a densely populated student residential area two miles to the west of the University which is served by several transit routes and in a small area near MTD transit centre in downtown Santa Barbara, the transit to automobile travel time ratio is actually below 1.
4. A framework for using accessibility maps in transit planning

Mapping accessibility based on time of day and trip destination allows one to create a virtual view of changing accessibility over the day. But this can be viewed as a major drawback as well, as one may be overwhelmed by a wide variety of maps and possible destinations. The real value is in how this information once generated can be used to estimate the impacts on changing transit services. Transit systems in the past have adjusted routes and times sparingly as there is a major concern for disrupting current users. The main objective of many transit managers is to protect the current user as they are the ones who use and benefit from the system. This means that changes are often made on an incremental basis, so as to keep the majority of the system operating without change, thereby disrupting as few of the existing users as possible. In times of increasing costs and flat revenues, transit officials have to either increase fares or cut back on service. Striking an optimal balance on adjusting fares and planning route changes can be a difficult task to say the least. In times of revenue growth, system planners often add new routes or increase frequency of service along routes. Increasing frequency of service along a route will not adversely impact existing ridership, and adding new routes can attract more riders or increase service coverage. But these changes can be incremental as well in that they often do not represent major changes in system operations. For example, in the SBMTD, recent changes have involved the addition of a morning and afternoon express commuter bus for an area that is about 40 miles away (representing adding service coverage for journey to work trips to core employment areas on the Santa Barbara south coast) and small changes to existing routes. Such tactics are typical of system planners.

The greatest concern is the problem of tracking levels of service for trip making across the service area. For all but a simple system of a few routes, this task can get somewhat complicated to do. It is difficult to track service levels for a large number of trip origin and destination pairs over a number of time periods throughout the day without the aid of a computerized model. The accessibility maps that were
described in the previous section provide a basis for performing such analyses, as well as the basis for mapping those areas that are affected, either negatively or positively, by changing routes and service times and frequencies. To explain this in some detail, consider the following notation:

\[ t^k_{ij} = \text{the travel time between origin } i \text{ and destination } j \text{ at time of day } k \text{ with existing service system} \]

\[ t'^k_{ij} = \text{the travel time between origin } i \text{ and destination } j \text{ at time of day } k \text{ associated with a modified service system plan.} \]

We can define the changes in service time for a given origin-destination pair \((i, j)\) at time of day \(k\) as

\[ \theta^k_{ij} = t'^k_{ij} - t^k_{ij} \]

Thus, total cumulative system change (CSC) can be calculated as

\[ \text{CSC} = \sum_{(i,j) \in \text{OD}} \sum_k \theta^k_{ij} \]

where \(\text{OD}\) is the set of origin and destination pairs \((i, j)\) that are used in service system planning.

The CSC represents the total sum of net changes in service times throughout the service region. If the sum is negative, then system changes have resulted in overall system improvement. If the sum is positive, then the net sum of net changes reflects that the system performance is degraded. It may also be important to relate the net changes in time based on a percentage change of current service times. This can be calculated as follows:

\[ \text{RCC} = \frac{\sum_{(i,j) \in \text{OD}} \sum_k \theta^k_{ij}}{\sum_{(i,j) \in \text{OD}} \sum_k t^k_{ij}} \]

Relative cumulative change (RCC) is a fraction that represents change as a ratio of the CSC to times associated with existing system operations. If we multiply the value, RCC, by 100 we can then express system changes as a percentage of existing system times. Negative RCC values represent improvement, and positive RCC values represent deteriorating levels of service.

To reflect the impact on users, it may be appropriate to weigh the travel between a given \((i, j)\) pair by the demand for travel along that route in the following manner:

\[ \text{WCC} = \sum_{(i,j) \in \text{OD}} \sum_k q_{ij} \theta^k_{ij} \]

where \(q_{ij}\) is the travel demand between origin \(i\) and destination \(j\). This now reflects a weighted cumulative change (WCC) on schedule and routing changes based on travel demands. Although we have not defined demand by time of day, that can be done as well in the above equation. Unfortunately, demand data for transit travel between specific origins and destinations is somewhat difficult to obtain and many systems lack origin-destination travel demand by time of day. Using the path routines A and B, one can generate the travel times necessary to calculate CSC, RCC and WCC values. These three measures can help a system planner as well as a transit board of directors in understanding the total and/or relative impact of any
proposed system changes. It may also be helpful to calculate a relative weighted cumulative change value (RWCC) which can be calculated in a manner that is analogous to that of calculating RCC.

We can also measure the impact on a specific area in terms of service changes. This can be calculated as

$$CNC_i = \sum_{(i,j) \in OD} \sum_k \theta_{ij}^k + \sum_{(j,i) \in OD} \sum_k \theta_{ji}^k$$

where CNC$_i$ is the cumulative net change in travelling to and from location i. Such changes may net out positive or negative. If we calculate this value for all areas i in the region, we can then map these cumulative values to show which areas will have service improvements, which areas will experience poorer service, as well as which areas will experience no net change in service. Again the above formula can be weighted to reflect travel demand between a given origin and destination, so that it reflects existing travel patterns.

Suppose that a standard is set by a transit board that states that any given trip should not be lengthened by more than $X$ minutes without specific board approval and discussion. This type of standard can be easily supported by identifying those trips ($i, j$), where

$$t_{ij}^k - t_{ij}^k \geq X$$

Any trip between a given origin $i$ and destination $j$ that satisfies the above inequality takes more than $X$ minutes longer (during time period $k$) to make than it did before making changes to the system. We can then identify areas that are origins or destinations where trip times are lengthened beyond the standard. These areas can then be mapped so that transit planners can focus on those areas to mitigate impacts by making further changes to the system or by acknowledging that these areas have the greatest impacts when making the planned system changes. Easily calculating such information and mapping, it can be useful in planning sessions, board meetings and public presentations.

The basic idea is that the shortest path routines, integrated with a GIS database of roads, transit routes and route time tables can be used to provide high-quality accessibility analyses beyond what has been accomplished to date. The framework discussed here can be used to map system facilitated access, calculate system impact metrics such as cumulative net change and WCC, as well as map those areas that experience impacts beyond some specified threshold.

5. Summary and conclusions

The subject of this article deals with the process of accessibility mapping within the context of transit services. Many researchers have struggled with the task of defining and measuring accessibility. Our short review classified past work into six main categories. These were (1) system access, (2) system facilitated access, (3) integral access, (4) time-constrained access, (5) access based on utility and value, and (6) relative access. We have shown that although there are numerous papers that deal with measuring access, many are concentrated on the physical access of a system (e.g. being close to a bus stop or transit station) and not on the time in which it takes to travel between a desired origin and a destination. Those who have modelled system travel times have usually made simplifying assumptions in a variety of
categories, e.g. transfer and waiting times, average travel speed to estimate times along routes or ignored scheduled arrival or departure times altogether. Most of the past work involving GIS has been devoted to the mapping access to an urban core, but not by time of day or by direction of trip, or with detailed schedule information. Although these details may not be necessary from a macro-level perspective, they can be particularly useful for transit users and transit planners. Route planners for transit users are now available on the Internet to help support trip planning with the type of detail necessary for the transit user. Although computer-based route planners have been developed with or integrated into GIS-based applications, they have not typically been used beyond identifying single routes. In this article, we proposed a system for accessibility modelling that uses a detailed route planning-based approach and presented results generated for an analysis of the services provided by the SBMTD. This approach is based on a novel method in integrating bus service times as arc attributes in a GIS. The results show that in many areas, service levels are low enough that users would not be encouraged to abandon their private vehicles, as routes take 2–5 times longer to accomplish using transit as compared to the personal car. Accessibility maps do show several areas in which competitive routes do exist (i.e. transit service times being competitive to trip times made by a personal car) for certain times of the day. We also describe how nuances of trip times by time of day can be used to help employers focus on developing flexible hour work schedules for their employees. Finally, we define a family of metrics for transit system performance (e.g. cumulative net change) that can be used to model the impact of proposed changes in routes and schedules. This forms the basis of a modelling framework, which can aid transit planners and directors in planning system improvements and evaluating changes.

Travel accessibility analysis for public transit has focused on average service-provided access. However, to track the impacts of system changes on users, it is important to track changes in access over different times of the day. Kwan (1999) and others have discussed the notion that people’s activities are limited by time space constraints. The framework described above can be expanded to model accessibility within time constraints, especially with respect to the ability to make several trips and return within a given time frame (Arentze et al. 1994). Another area of needed research involves the development of GIS-based tools that can be used to redefine existing routes, search for possible productive new routes, or fine tune time tables and departure times so that service levels can be improved. The overall objective would be to identify those changes that yield the best gains in efficiency with the least amount of expended cost or identify those changes which produce a desired level of savings without negatively impacting users beyond a standard.

Notes

1. The reason why we focus on the university in our examples is that the university is one of the largest employers in the county. It also has an enrollment in excess of 20,000 students. Although some of the students live relatively close to campus, many students and employees tend to live in surrounding areas concentrated along those routes that provide relatively good transit access (e.g. by the downtown express bus).

2. In an informal presentation of these maps to the SBMTD board of directors, directors expressed some surprise that service to the campus was not better, leading one to comment that the maps depicted valuable information.
References


MOSELEY, M.J., 1979, Rural accessibility – role of the motor car. Highway Engineer, 26(6), pp. 18–19.


