A Generalized Computational Framework for Accessibility: From the Pedestrian to the Metropolitan Scale

Fletcher Foti and Paul Waddell
Department of City and Regional Planning
University of California Berkeley
Dennis Luxen
Institute of Theoretical Informatics
Karlsrühe Institute of Technology

February 22, 2012

Abstract

Travel models, including activity-based travel models developed in recent years, still generally use traffic analysis zones and ignore local streets in their network representation. In short, they ignore walking scale access and movements. This is a well-known and very problematic limitation in current travel models, and by extension, in integrated land use and transportation models, even if the land use models are at a parcel level.

This paper describes a new project that involves generating a graph of the urban region that unifies information on parcels and the full street network, including all local streets, into a topological graph, and creates very efficient algorithms to compute point to point accessibilities, as well as accessibility trees to activities, using a set of weights on the edges of the graph to allow shortest path computations from parcels to activities.

The objective of this project is to create software infrastructure that can provide an interface between parcel-level land use models, which maintain information on what kinds of businesses and households are located on parcels, and emerging activity-based travel models that are attempting to move down to the local street and parcel level of detail.

Current benchmarks have region wide aggregation results (355K queries with a search radius of .5km) performed on a 4 cpu-core machine in 2.8 seconds. An "OpenWalkscore" is computed on the same network in 1.2 seconds. This research is part of a project funded by the National Science Foundation and the Metropolitan Transportation Commission and is being applied to the 9-County Bay Area in California.

Contents

1 Introduction

This paper describes a new project that involves generating a graph of the urban region that unifies information on parcels and the full street network, including all local streets, into a topological graph. Efficient algorithms operate on this graph to compute point to point accessibilities, as well as accessibility trees to activities, using a set of impedances on the edges of the graph to allow shortest path computations from parcels to activities.

This is not the first work of its type, but this paper differentiates itself in its general framework and special attention given to high performance network algorithms, in particular the use of contraction hierarchies, which can be orders of magnitude faster than previous techniques.

The paper will be organized as follows. First, the concept of accessibility will be outlined as is familiar in transportation-land use modeling. Then the scope will be broadened to introduce an urban theoretical conceptualization of urban form as a pattern of land uses within a transportation network. Next, data structures and algorithms are outlined that are required to compute the proposed accessibility measures, and performance benchmarks are given from a case study in the San Francisco Bay Area. Finally, the current software implementation will be assessed, including potential uses and future improvements.

2 Theory and Measurement of Accessibility

Accessibility is a well explored area of urban theory. Kevin Lynch in *Good City Form* states, "activities are assumed to locate according to the relative cost of reaching materials, customers, services, jobs, or labor. Other values are simply subsidiary constraints in this struggle for access" (24). Lynch traces the original concepts to Wingo and Alonso (35; 4), but the first explicit discussion is provided by Hansen (20).

Operationalization of accessibility begins with the Weibull axiomatic framework (33). Current accessibility frameworks are considered to be gravity-model based (defined by attractions and discounted by distance), cumulative-opportunity (summations within a set impedance measure) and space-time (limited by the opportunity prism of an individual's activity skeleton) (22; 25). Dong and others expand on the space-time prisms by creating a logsum-based measure within a travel model (12).

To measure access, one must first choose a basic unit of space to use. The majority of transportation models in use today still rely heavily on zone-based geography for its simplicity and computational tractability (21). Zones can vary in size, but are usually a few city blocks at their smallest. Drawbacks to this method include: the zones must be defined manually, they are arbitrary in scope, and they are too large to model micro-land use measures and walkability (which often vary on a block-by-block basis).

In recent years, land use modeling has evolved to incorporate the natural micro-measure of land use: parcels (32). In fact, the computing power and quantitative methods now exist to enable the use of the smallest units of interest across several dimensions, including buildings (from county assessors data), households (from population synthesizers), and jobs (from *NETS* and *CoStar* data).

Some early work has attempted to reach this level of precision in a computationally efficient way. Chen computes accessibility variables at the block level for the SCAG activity model with a focus on employment accessible through driving and transit (9). The popular "Walk Score" captures excellent block-by-block walkability metrics but focuses only on amenities like coffee shops, groceries, and parks and does not include aggregate measures of the built environment (2). And commercial products also have similar tools, most notably ESRI's Network Analyst and Citilabs' Accession. Again, this work differentiates itself with its use of contraction hierarchies for extremely fast network queries.

3 The City as a Hierarchical Graph

This work unifies information on parcels and the full street network into a topological graph. Urban design has predated many of the concepts in this work by decades; Kevin Lynch, in his seminal book *Image of the City*, defined urban geography with the abstractions: "paths, edges, districts, nodes, and landmarks" and discussed their use in forming mental maps (23). Although early work by Von Thünen and Christaller introduced the idea of a monocentric city (31; 10), much recent work explores the distribution of land use in the polycentric city (19; 18; 17). In addition, there are numerous articles by Castells and others exploring the idea of city as network (6; 34; 5).

This accessibility framework uses a new modeling conceptualization - street node geography, which places the set of parcel boundaries in a geography delineated by the graph of transportation options. The street network has a naturally varying density - more edges in denser areas and fewer edges in less dense areas - and has a well defined geometric relationship with the parcel boundaries. In other words, there is a direct link between a parcel's size and the street network that surrounds it, and this link is an immanent and ubiquitous property of the form of the city (notable exceptions being areas like large campuses).

Each parcel is connected to the agents of the city - whether persons, households, buildings, land, or businesses, and the parcels have access to each other through an explicitly modeled transportation network; no "as the crow flies" distances are used. Simply put, each parcel is connected to the street network, variables are aggregated to the minimal level of street node, and then accessibility queries are performed on the loaded data. It should be noted that 2 million parcel centroids can be linked to their respective nearest street nodes using a K-D tree in about 1 second.

One way of thinking of this is that the street network defines very small TAZs, where each TAZ is the set of parcels that are closest to a given street node. This conceptualization has none of the previously discussed limitations of zone-based models: there is no manual intervention, zones are not arbitrary in scope (they have a direct relationship to the street network), and are easily small enough to measure pedestrian-scale changes in the built environment. Of course street node geography is not required to use contraction hierarchies - parcels can be tied directly to the street network - but another order of magnitude in performance is gained using this approach.

4 Types of Accessibility Variables

One of the primary reasons for the creation of this framework is to capture a broad array of variables that can be used to predict non-auto travel as a mode choice (or towards location choices which facilitate this). It has been well established that the proportion of non-automobile travel is correlated to the built environment (8; 13; 26). This work characterizes the built environment using Kockelman and Cervero's original 3Ds of travel demand - density, design, and diversity (7). Additional"D"s have been added in intervening years, including distance to transit, destination, and demographics, and a recent literature review by Cervero and Ewing has been updated to reflect this (14).

This work focuses on the original 3Ds as aggregate measures of the built environment and uses contraction hierarchies for an additional speedup when dealing with specific destination data (point data for available activities). The use of 3Ds-style variables is nearly ubiquitous in the literature, so only a brief description is provided here and the reader is referred to urbansim.org for the constantly updated detailed API specification. Although 3Ds variables are not novel to this work, its contribution is in defining a general cloud-based software framework based on contraction hierarchies for increased performance.

4.1 Operationalizing the 3Ds

DENSITY - Density variables can be "cumulative-opportunity" (or isochrones) which are simple summations within a radius, "gravity-based" where attractions are discounted by distance, and "utility-based" which require a fully estimated activity model and are based on the logsum of available activities. This work focuses on gravity-based measures and allows a choice of aggregation and decay functions.

DESIGN - Work by Michael Southworth captures aspects of street network design which are valuable in predicting walkability in adjacent areas (30; 28; 29). Variables include street node density, lineal street feet, average block length, and others. Kockelman and Cervero list a number of other design variables in the original 3Ds paper: proportion of 4-way intersections, angle of street intersection (to capture grid vs. curvilinear crossings), freeway length per unit area, number of cul-de-sacs, speed limits, street width, presence and type of parking, etc (7). These are all implemented as data permits.

DIVERSITY - Diversity metrics are typically captured using Shannon entropy, which can be computed for any categorical variable. Shannon entropy is the central tenet of information theory (27), and measures the expected value of the information contained in a sequence of draws from a random variable, ranging from 0.0 when a single category dominates the sample to 1.0 when there is a perfectly even allocation of the sample to each category. In this work, entropy is used to detect balance in jobs/housing, employment sectors, building types, and others.

4.2 Online API

The end product of this project is an API that can be used to contact a remote server and compute various dimensions of accessibility in a way that is trivial to the end user. The client-side API is currently available in the Python programming language, and the server-side functionality is written in C++ (the C++ code can also be accessed directly). In the future, this framework will support XML and web services, as this is the standard technology for an online API of this sort. Full and updated API documentation will be made available on urbansim.org.

5 Contraction Hierarchies

Contraction hierarchies (CH) is the underlying technology which allows very efficient network range and POI queries. It is used for all network queries in the current implementation. Below, the theory and application of contraction hierarchies is discussed, followed by performance comparisons between CH and Boost Graph Library (BGL).

Finding shortest paths in a graph is a problem that can be solved in polynomial time by Dijkstra's seminal algorithm (11). However, Dijkstra's algorithm does not scale well with network size, and the computer science community has only recently made substantial advances in speedup techniques. For this discussion, the road network is modeled as a weighted graph $G = (V, E)$ where the vertices are intersections and the edges are street segments. The weight of an edge gives the impedance, or average travel time to traverse that edge.

Geisberger et al.'s Contraction Hierarchies (CH) (16) is a speedup technique to Dijkstra's algorithm especially tailored to exploit the hierarchical properties of road networks (3). The technique utilizes the property that sufficiently long routes will enter the long-distance network, i.e. the sparse sub-network of highways and regional or national roads. CH offer a trade-off between preprocessing and query times: continental sized networks can be processed on commodity hardware within a matter of minutes and queries run in the order of one hundred microseconds. The Bay Area network used in this work is preprocessed in about 4 seconds.

CH computes an heuristic ordering of the graph nodes by some measure of importance, e.g. the endpoint of a dead-end street is less important than a highly frequented junction. The ordered nodes are replaced by *shortcuts*, where shortcutting means that a node is (temporarily) removed from the graph and as few as possible shortcuts edges are inserted to preserve shortest path distances. The resulting data structure consists of all original edges and nodes together with the generated shortcut edges. The query is a modified bi-directional Dijkstra search and only needs to consider edges to more important nodes, which makes the search data structure a directed acyclic graph. The search space usually consists of a few hundred nodes only, even for long-range queries from one end of the continent to the other.

5.1 Contraction Hierarchies and POI

Querying for points of interest (POI) can be done with breadth-first search in the unit-distance case or a unidirectional Dijkstra search for arbitrary edge weights (15). Consider the case where one is interested in only the k-nearest of a set of of categorized POI (i.e. distinct categories for restaurants, gas stations, ATM locations, etc). The actual locations of the POI are mapped to the road network and thus the input is a list of vertices $\mathcal{L}_i$.

To index the POI locations, the backward CH search space for each of the input vertices $w \in \mathcal{L}_i$ is explored. Each encountered vertex contains a list ordered by distance that saves the shortest distances to the POI as they are encountered by the search. Since k POI are considered at most, each list is of length k.

When searching for the set of closest POI, a query enumerates the forward search space and checks the list of every encountered vertex. Each list is merged with the (sorted) result list and the lowest k entries are kept. The search can be aborted as soon as the k-furthest POI in the result list is closer to the source vertex than any remaining vertex from the search space, resulting in the lightning-fast search times described in this paper.

6 The MTC Project

The Metropolitan Transportation Commission (MTC) is responsible for regional transportation planning in the San Francisco Bay Area. This work is funded by a grant from MTC to the University of California Berkeley with Paul Waddell as Principal Investigator. The grant was designed to coincide with the MTC Sutainable Communities Strategies and Regional Transportation Plan Outreach Process. Deliverables include generating visualization of alternative growth scenarios while engaging local governments.

For the bay area, parcel shape files were attained by MTC for the 9-county bay area. The road network is freely available through the OpenStreetMap (OSM) project. *Table 1* contains the number of objects used in this case study. All data are stored in a PostgreSQL (PostGIS enabled) database. After being read into memory, OSM data is stored in Boost Graph Library (BGL) data structures, and Boost Geometry is used for most computational geometry algorithms.

7 Performance Metrics

Using the framework described in this paper, a set of performance metrics are computed which demonstrate the computational advances of the current implementation. Unless otherwise stated, benchmarks are performed on an Intel I3 with 4GB of RAM using Windows 7 (generally considered to be commodity hardware at the time of this writing).

The performance of the algorithm is not sensitive to the aggregation method or decay function, but is highly sensitive to the search radius used. This is because the number of nodes which are aggregated is of order radius-squared since a range query is closely related to the area of a circle. The preliminary implementation is done using the Boost Graph Library (BGL) and is compared to the updated version which uses Contraction Hierarchies (CH).

The performance numbers for a single range query in the dense street network of downtown San Francisco are given in *Table 2*. These numbers are for the graph range query (all nodes within a distance) and do not include aggregation of any kind. As can be seen, CH is 8-15 times faster, and the factor increases with radius.

Table 3 shows the performance for a typical accessibility query for a single radius for the whole San Francisco Bay Area. This benchmark performs a buffer aggregation query for each of the 355K nodes in the Bay Area network. Note that the performance of the average query in this benchmark is faster than that given in *Table 2*. This is because *Table 2* gives the search time in the densest part of the Bay Area street network - in downtown San Francisco - and range queries on less dense networks run much faster. Since many spatial metrics will be computed for each node in the Bay Area, it is important to find the average performance across the distribution of street node densities. It can be seen that 355K aggregations (the complete Bay Area network) can be multi-threaded to finish in only 2.8 seconds (with a radius of .5km).

Table 4 gives the performance benchmarks for each of the 355K nodes using CH but varying the search radius. As this work moves forward, it will be important to compute these metrics for different impedances (using travel time instead of distance), and for those values of the search radius deemed appropriate (radii will be larger for auto travel than for walking). *Table 4* gives the search radii applicable to walkability.

As a final benchmark, the system is run in a server environment (12 core Intel Xeon - X5690 @ 3.47GHZ - with sufficient RAM). *Table 5* shows results for an aggregation run on the entire United States network using a radius of .5km. There are 16.4 million edges and 13.1 million nodes in the OSM network for the United States, and these 13 million range queries are computed in between 15.7 and 21.1 seconds, depending on the exact variable being aggregated.

8 A Simple Example: "OpenWalkscore"

"Walkscore" is a new commercial project created by FrontSeat Software. It's popularity is becoming widespread in both the real estate industry and academia, and fortunately its implementation details are available directly on the website (2). Walkscore is a weighted average of 23 different POI queries for 9 different POI categories (using the distance to the first 10 restaurants, 1 grocery store, 2 coffee shops, etc). Land use data for this computation can be obtained from OSM or commercial data sources like Navteq, and thus recreating Walkscore as an "OpenWalkscore" is an ideal proof-of-concept for this framework.

Additionally, POI queries don't have to touch every node in the search radius (as described above), and so can run much faster than the typical density aggregation queries. In fact, for 355K node sources (one per node in the Bay Area network) and 23 POI queries per node, the 8.17 million individual POI queries can be run in 1.2 seconds (on the 12 cpu-core server hardware described above). Not only does this framework allow for modifications of Walkscore to tailor it to indexing different concepts (Bikescore, Coffeescore, etc), and to allow removing socioeconomic bias (bookstores as included in Walkscore being a typical destination for higher income patrons), this computation time allows near-interactive speeds for looking at the results of modifications. *Figure 1* shows a color-coded representation of the results (green being high OpenWalkscore), where each node is displayed individually so as to emphasize the point-based nature of the analysis. This map is nearly identical to the one available on walkscore.com (after converting to a rasterized image).

9 Assessment

As has been shown, this study is a framework for a generalized accessibility engine which captures a broad array of 3Ds and POI variables. This by no means is a complete set, and additional variables and types of variables can be added as the need arises. In addition, the framework is generalized so that it can be applied to any data that is spatially distributed.

This specific implementation of the framework contributes to previous work in its increased performance and higher level of geographic detail. Its architecture as an online server which can respond to simple client API calls makes it widely available to users without high levels of programming experience. The framework is usable by academics and practitioners working on travel models, land use models, public health projects, GHG emissions studies, real estate development comparables, and others.

9.1 Limitations and Next Steps

One major limitation of this work is that it is only a distance and gravity-measure based framework. Ideal accessibility measures would be derived from the utility of a logsum in a choice model. Additionally, queries have a single node of origin; to capture space-time prism-style accessibility, the shortest path between two points and a maximum deviation would need to be implemented.

Second and perhaps more importantly, this work is a distance-based implementation. It should be expanded to include multi-modal travel times, using congested road network travel times, transit schedule-based travel times, and bike network travel times. These travel times can then be used in lieu of distance in the radius of the accessibility computation. Note that the framework already supports the use of other impedances in lieu of distance, but the network mapping has not yet been done.

Future work for this study is to create a simple travel model framework which can be used to derive utility-based accessibility measures. The travel model will take into account the specific multi-modal travel times and activity locations available to an agent, as well as deriving specific preference structures based on the agent's demographic attributes. A simple supply-side traffic simulator will be used to create congested travel times on the road network. In total, the implementation will include *Urbansim* as a land use model, as well as the upcoming travel demand and traffic congestion models, which will comprise a full set of regional modeling tools.

References

[1] Steps for computing the dissimilarity, entropy, Herfindahl-Hirschman and accessibility (Gravity with competition) indices, 2008.

[2] Walk score methodology, July 2011.

[3] Ittai Abraham, Amos Fiat, Andrew V. Goldberg, and Renato F. Werneck. Highway Dimension, Shortest Paths, and Provably Efficient Algorithms. In Moses Charikar, editor, *Proceedings of the 21st Annual ACM–SIAM Symposium on Discrete Algorithms (SODA'10)*, pages 782–793. SIAM, 2010.

[4] W. Alonso. *Location and land use: toward a general theory of land rent.* Harvard University Press, 1964.

[5] Jonathan V. Beaverstock, Richard G. Smith, and Peter J. Taylor. World-City network: A new metageography? *Annals of the Association of American Geographers*, 90(1):123–134, 2004.

[6] M. Castells. *The rise of the network society.* Blackwell Publishers, 2000.

[7] Robert Cerveo and Kara Kockelman. Travel demand and the 3Ds: density, diversity, and design. *Transportation Research Part D: Transport and Environment*, 2(3):199–219, 1997.

[8] Robert Cervero and Jin Murakami. Effects of built environments on vehicle miles traveled: evidence from 370 US urbanized areas. *Environment and Planning A*, 42(2):400–418, 2010.

[9] Yali Chen, Ravulaparthy, Pendyala, and Bhat. Development of opportunity-based accessibility indicators. *TRB 2011 Annual Meeting*, 2010.

[10] Walter Christaller. *Die zentralen Orte in Suiddeutschland.* 1968.

[11] Edsger W. Dijkstra. A Note on Two Problems in Connexion with Graphs. *Numerische Mathematik*, 1:269–271, 1959.

[12] X. Dong, M.E. Ben-Akiva, J.L. Bowman, and J.L. Walker. *Moving from trip-based to activity-based measures of accessibility.* Elsevier, 2006.

[13] Reid Ewing and Robert Cervero. Travel and the built environment: A synthesis. *Transportation Research Record: Journal of the Transportation Research Board*, 1780:87–114, January 2001.

[14] Reid Ewing and Robert Cervero. Travel and the built environment: A meta-analysis. *Journal of the American Planning Association*, 76:265–294, 2010.

[15] Robert Geisberger. *Advanced Route Planning in Transportation Networks.* PhD thesis, Karlsruhe Institute of Technology, February 2011.

[16] Robert Geisberger, Peter Sanders, Dominik Schultes, and Daniel Delling. Contraction Hierarchies: Faster and Simpler Hierarchical Routing in Road Networks. In Catherine C. McGeoch, editor, *Proceedings of the 7th Workshop on Experimental Algorithms (WEA'08)*, volume 5038 of *Lecture Notes in Computer Science*, pages 319–333. Springer, June 2008.

[17] A. Getis. Second-order analysis of point patterns: the case of chicago as a multi-center urban region. *Perspectives on Spatial Data Analysis*, pages 83–92, 2010.

[18] P. Gordon, H. W Richardson, and H. L Wong. The distribution of population and employment in a polycentric city: the case of los angeles. *Environment and Planning A*, 18(2):161–173, 1986.

[19] D. L Greene. Recent trends in urban spatial structure. *Growth and Change*, 11(1):29–40, 1980.

[20] W. G Hansen. How accessibility shapes land use. *Journal of the American Institute of Planners*, 25(2):73–76, 1959.

[21] J. D Hunt, D. S Kriger, and E. J. Miller. Current operational urban land-use-transport modelling frameworks: a review. *Transport Reviews*, 25(3):329–376, 2005.

[22] M. P Kwan. Space-time and integral measures of individual accessibility: a comparative analysis using a point-based framework. *Geographical Analysis*, 30(3):191–216, 1998.

[23] K. Lynch. *The image of the city*. MIT Press, 1960.

[24] K. Lynch. *Good city form*. MIT Press, 1984.

[25] H. J Miller. Measuring space-time accessibility benefits within transportation networks: basic theory and computational procedures. *Geographical Analysis*, 31(1):1–26, 1999.

[26] James Parks and Joseph Schofer. Characterizing neighborhood pedestrian environments with secondary data, 2006.

[27] Claude Shannon. *The mathematical theory of communication*. The University of Illinois Press, paperback ed. edition, 1948.

[28] M. Southworth. Walkable suburbs?: an evaluation of neotraditional communities at the urban edge. *American Planning Association. Journal of the American Planning Association*, 63(1):28–44, 1997.

[29] M. Southworth. Designing the walkable city. *Journal of urban planning and development*, 131:246, 2005.

[30] M. Southworth and P. M Owens. The evolving metropolis: studies of community, neighborhood, and street form at the urban edge. *Journal of the American Planning Association*, 59(3):271–287, 1993.

[31] J. H Von Thunen and P. G Hall. *Isolated state: an English edition of Der isolierte Staat*. Pergamon Press, 1966.

[32] P. Waddell. UrbanSim: modeling urban development for land use, transportation, and environmental planning. *Journal-American Planning Association*, 68(3):297–314, 2002.

[33] J. W Weibull. An axiomatic approach to the measurement of accessibility. *Regional Science and Urban Economics*, 6(4):357–379, 1976.

[34] Mark Wigley. Network fever. *Grey Room*, (4):83–122, July 2001.

[35] L. Wingo. *Transportation and urban land*. Resources for the Future, 1961.

Object Type	*Count*
street edges	463K
street nodes	355K
parcels	2.1M

Table 1: Object counts in the Bay Area

	$.5km$	$1.5km$	$2.5km$	$3.5km$
BGL	39ms	431ms	1.254s	2.015s
CH	5ms	26ms	81ms	120ms

Table 2: Performance of a single range query

	No threading	*Threading − 4 cores*
BGL	295.88s	213.8s
CH	6.39s	2.8s

Table 3: Performance for a density query for the full 355K nodes

Radius	*Time*
.5km	2.8s
1.5km	18.9s
2.5km	52.8s
3.5km	103.3s

Table 4: Performance by radius for 355K density queries using CH

Item	*Amount*
street edges	16.4M
street nodes	13.1M
preprocessing	22min
search time	15.7-21.1s

Table 5: Benchmark on US network - 13.1M queries, .5 km radius using 12 cpu-cores

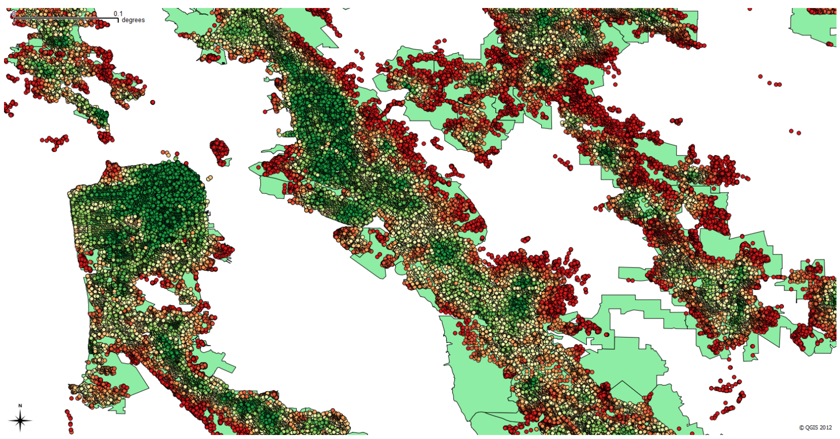

Figure 1: OpenWalkscore for 355K nodes computed in 1.2s