Access Across America: Transit 2021 Methodology

Prepared by the

Accessibility Observatory at the University of Minnesota

July 18, 2023

ACCESSIBILITY
OBSERVATORY

UNIVERSITY OF MINNESOTA
Driven to Discover℠

Authors

Andrew Owen
Lead Researcher

Shirley Shiqin Liu
Researcher

Saumya Jain
Researcher

Eric Lind
Director

Accessibility Observatory
Center for Transportation Studies
University of Minnesota

Acknowledgements

The development of this report was made possible by sponsorship from:

- Arkansas State Highway and Transportation Department
- California Department of Transportation
- District Department of Transportation
- Federal Highway Administration
- Florida Department of Transportation
- Illinois Department of Transportation
- Iowa Department of Transportation
- Maryland Department of Transportation
- Massachusetts Department of Transportation
- Minnesota Department of Transportation
- North Carolina Department of Transportation
- Tennessee Department of Transportation
- Virginia Department of Transportation
- Washington State Department of Transportation

Contents

1 Summary

This document describes the methodology used by the Accessibility Observatory at the University of Minnesota to produce the accessibility metrics and related data that are presented in *Access Across America: Transit 2021*. An overview of the methodology for the Observatory's 2021 reports and calculations is provided below, and detailed descriptions can be found in the following sections.

- **Data Sources**
    1. U.S. Census TIGER 2010 datasets: blocks, core-based statistical areas (CBSAs)
    2. U.S. Census Longitudinal Employer-Household Dynamics (LEHD) 2019 Origin-Destination Employment Statistics (LODES)
    3. OpenStreetMap (OSM) North America extract, retrieved January 2021
    4. General Transit Feed Specification (GTFS) schedule data from transit operators, various dates
- **Data Preparation**
    1. Divide the geographical United States into analysis zones for efficient parallelization
    2. Construct unified pedestrian-transit network graph for each analysis zone
- **Accessibility Calculation**
    1. For each Census block in the United States, calculate travel time to all other blocks within 60km for each departure time at 1-minute intervals, over 7 - 9 AM period
    2. Calculate cumulative opportunity accessibility to jobs for each block and departure time, using thresholds of 5, 10, 15, …, 60 minutes
    3. Average accessibility for each block over 7 – 9 AM period
    4. Average accessibility for each included CBSA over all blocks, weighting by number of workers in each block
    5. Calculate weighted ranking for each included metropolitan area

2 Data Sources

2.1 Geography

All calculations and results in this project are based on geographies defined by the U.S. Census Bureau. Census blocks are the fundamental unit for on-network travel time calculation, and calculations are performed for every census block (excluding blocks that contain no land area) in the United States - this is a change in scope relative to *Access Across America: Transit 2014*, and aligned the data and calculations with the goals of the Observatory's National Accessibility Evaluation Pooled Fund project. This national scope was implemented for *Access Across America: Transit 2015*, and continues through the current year. Block-level accessibility results are then aggregated across core-based statistical areas (CBSAs) for metropolitan-level analysis. These geography definitions are provided by the U.S. Census Bureau's Topologically Integrated Geographic Encoding and Referencing (TIGER) program.[1] This project uses the geography definitions established for the 2010 decennial census.

2.2 Employment and Worker Population

Data describing the distribution of labor and employment in the region are drawn from the U.S. Census Bureau's Longitudinal Employer-Household Dynamics program (LEHD).[2] The LEHD Origin-Destination Employment Statistics (LODES) dataset, which is updated annually, provides Census block-level estimates of employee home and work locations. This project uses LODES data from 2019, the most recent available as of the performance of the 2021 accessibility calculations.

Note: The LODES dataset used in this report does not include job location data from the states of Alaska, Arkansas, or Mississippi. These states did not report these employment statistics to the Census for the 2019 year.[3] None of the top 50 metropolitan statistical areas reported in the *Access Across America: Transit 2021* lie within those states; however, Memphis (49th by total employment) borders both Arkansas and Mississippi, and thus access to jobs from areas within Memphis is likely to be underreported given the absence of those nearby job locations in the dataset.

2.3 Pedestrian Network

Data describing the pedestrian network across the country were obtained from OpenStreetMap,[4] an open-access online database of transportation network structures, maps, and other spatial information. OpenStreetMap, like Wikipedia, is composed of contributions from many individuals. In urban areas, it typically provides a much more detailed and up-to-date representation of pedestrian networks than datasets available from federal, state, regional, or local sources. The data used in this project were retrieved from OpenStreetMap on January 1, 2021. Specifically, the pedestrian network is composed of features with the "footway," "pedestrian," and "residential" tags.

---

[1] `https://www.census.gov/geographies/mapping-files/time-series/geo/tiger-line-file.html`

[2] `http://lehd.ces.census.gov/data/`

[3] a detailed LODES data release notes is here: `https://lehd.ces.census.gov/data/lodes/LODES8/LODESTechDoc8.0.pdf`

[4] `http://openstreetmap.org`

2.4 Transit Schedules

Detailed digital transit schedules in a consistent format are a critical component of this project, and the widespread availability of such data is a relatively recent phenomenon. The General Transit Feed Specification[5] (GTFS) was developed by Google, Inc. and Portland TriMet as a way to provide transit schedules for use in traveler routing and information tools.

Despite their importance and digital nature, the collection of GTFS datasets can be frustratingly inconsistent and error-prone. While the format of GTFS data itself is standardized, there are no standards for the digital publication of the datasets, and practices vary widely across transit operators. A majority of operators (at least among medium and large metropolitan areas) provide GTFS datasets via a direct website link. However, even among these, variations in URL naming conventions pose challenges for systematic retrieval. Other operators allow GTFS dataset downloads only after users interactively submit a form or agreement. Still others generate GTFS datasets and provide them directly to Google, Inc. for use in their popular online routing tool, but release them to the public only in response to direct requests with licensing.

To address some of these challenges, all GTFS datasets for this project were sourced through Transitland,[6] an open data platform that collects transit schedule data from transit providers worldwide. Utilizing Transitland's features, users can explore the actual routing and service information within a GTFS dataset on their website. This allows us the opportunity for certain levels of ground truth validation and quality control checks. Furthermore, we implemented an additional step of data quality verification using a GTFS validator[7] to ensure compliance with network routing calculation standards. Nevertheless, it remains crucial to acknowledge that validating whether a GTFS dataset obtained from the website originates from the actual transit operator, or if it has undergone any alterations, poses a continuing challenge.

Travel time calculations for this year's project are based on schedules valid for January 13, 2021 (a Wednesday with normal, non-holiday service). When a schedule for that date is not available for a given transit operator, the schedule which comes closest to including it is used.

---

[5] https://developers.google.com/transit/gtfs/

[6] https://www.transit.land/

[7] https://gtfs.org/schedule/validate/

3 Data Preparation

3.1 Analysis Zone Definition

This project relies on the efficient calculation of shortest paths between a very large number of origin–destination pairs given the national scope, repeated for many departure times. In order to efficiently parallelize these calculations across multiple computers, the geographical USA is divided into 4879 "analysis zones" each including no more than 5,000 Census blocks. Figure 1 shows the Census block and CBSA boundary structure for the Minneapolis–St. Paul region, and figs. 2 and 3 illustrate the process of constructing analysis zones on the national and local scales, respectively.

To simplify the calculation of local time, which is necessary to determine appropriate transit service for a given minute of the day, time zone geometries based on U.S. Census data[8] were used as parent geometries of the analysis zone areas. This way, each analysis zone is guaranteed to have a single associated time zone, whereas the use of non-time zone parent geometries would complicate local time lookup when calculating transit schedules and accessibility.

---

[8] `http://efele.net/maps/tz/world/tz_world.zip`

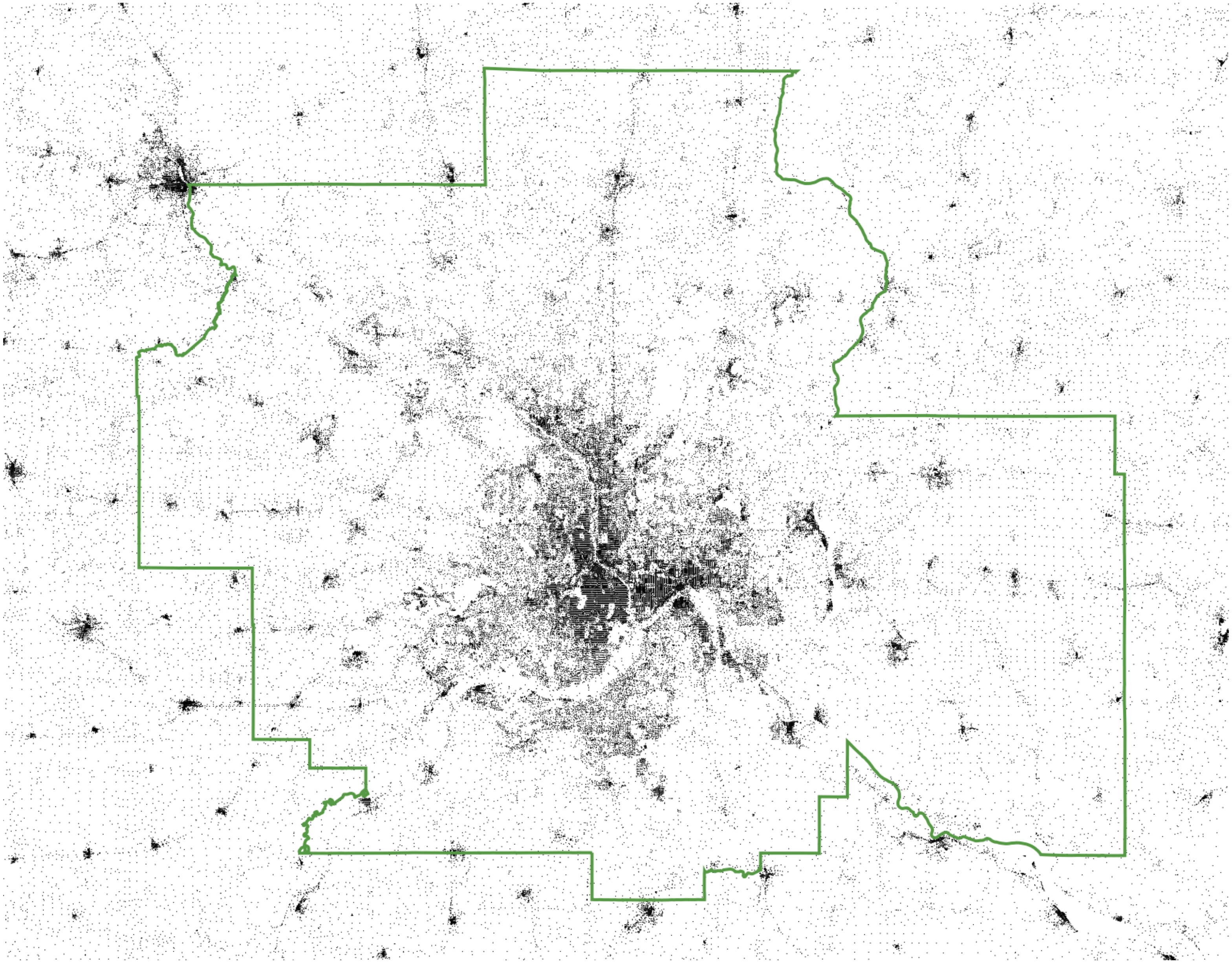

Figure 1: Boundary and Census blocks for the Minneapolis–Saint Paul, MN CBSA. Each dot represents the centroid of a single Census block.

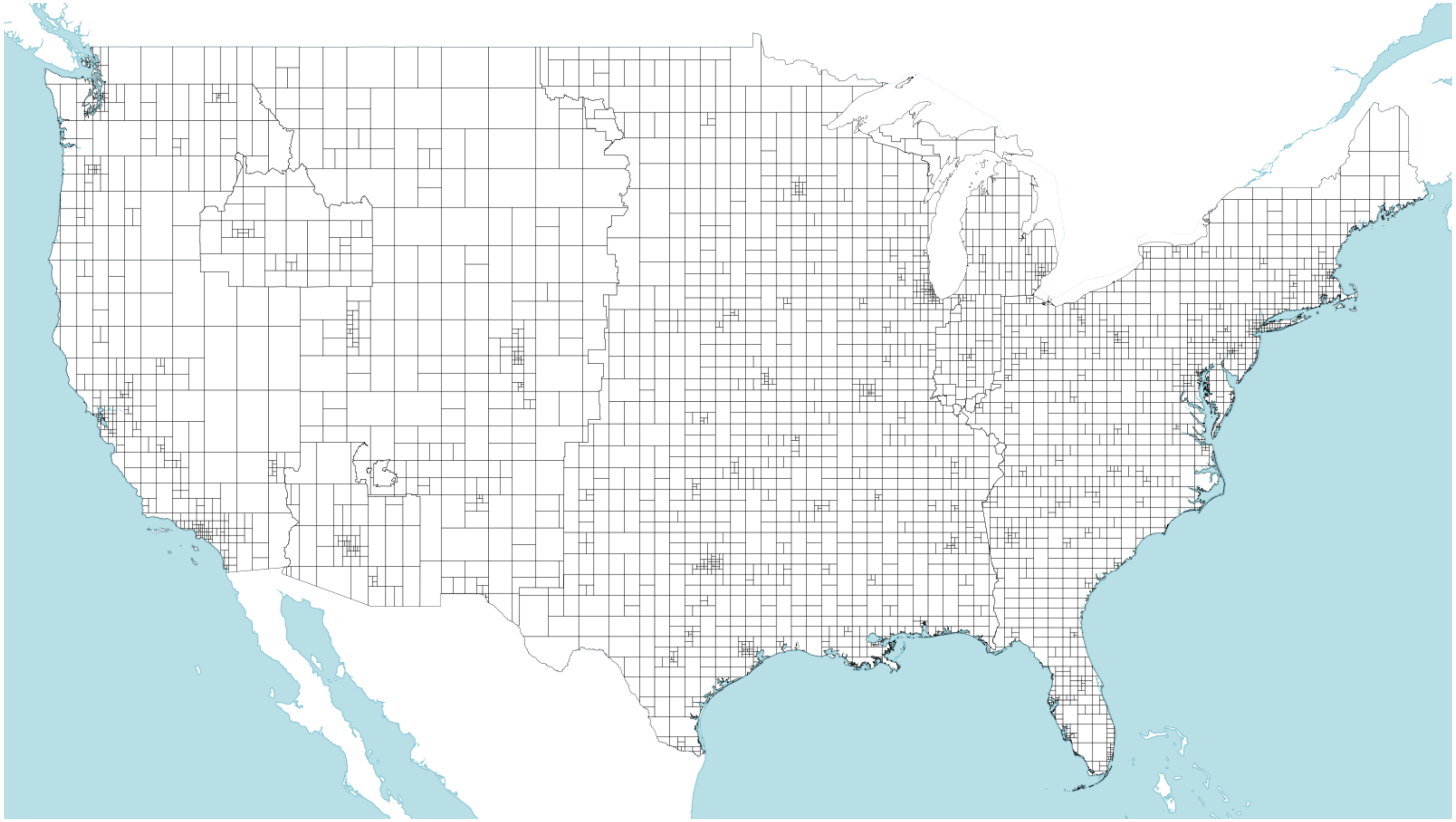

Figure 2: The United States divided into analysis zones. Each zone contains a maximum of 5,000 Census block centroids.

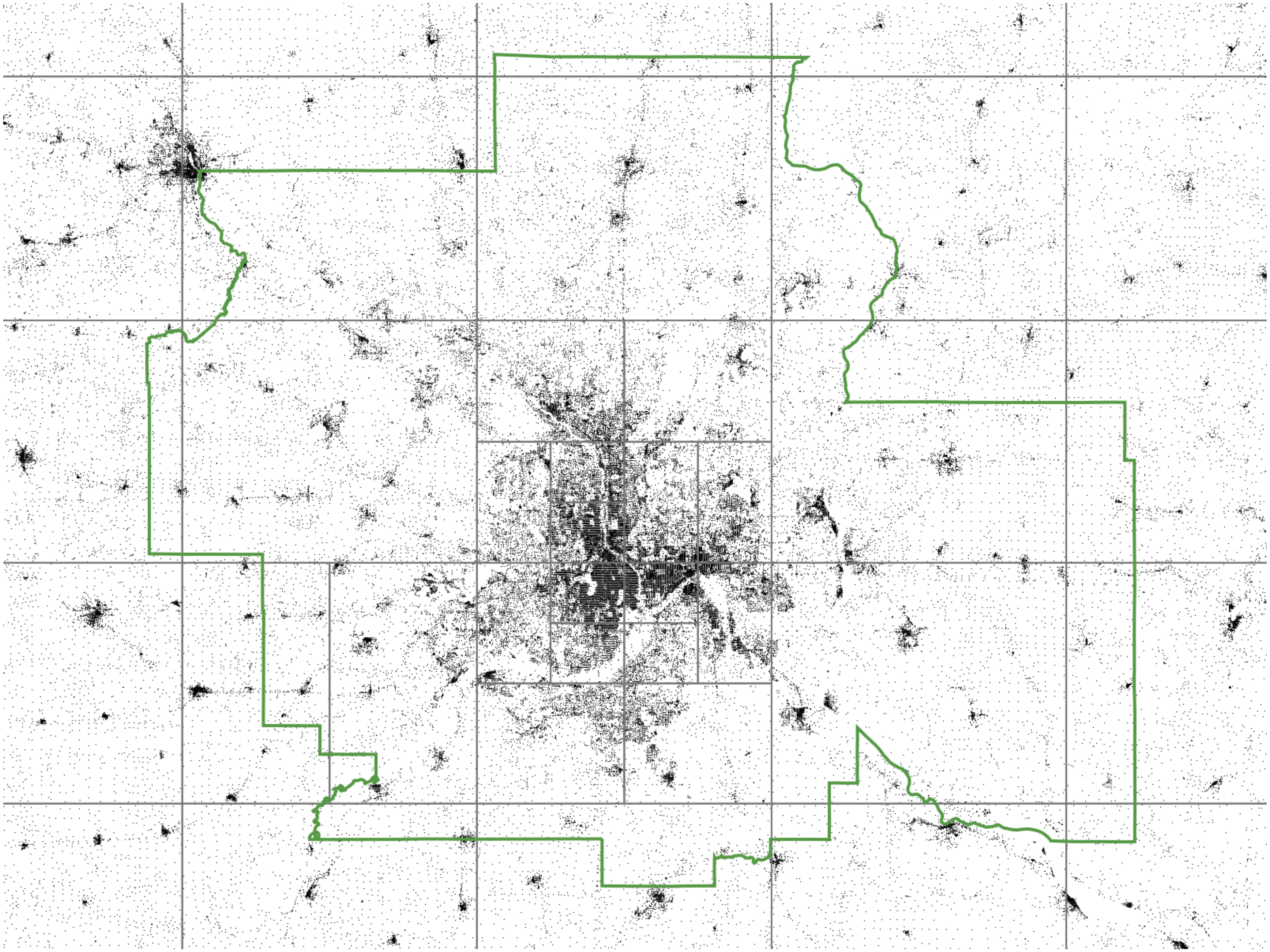

Figure 3: Example of the analysis zone structure within an urban area - Minneapolis & St. Paul, Minnesota

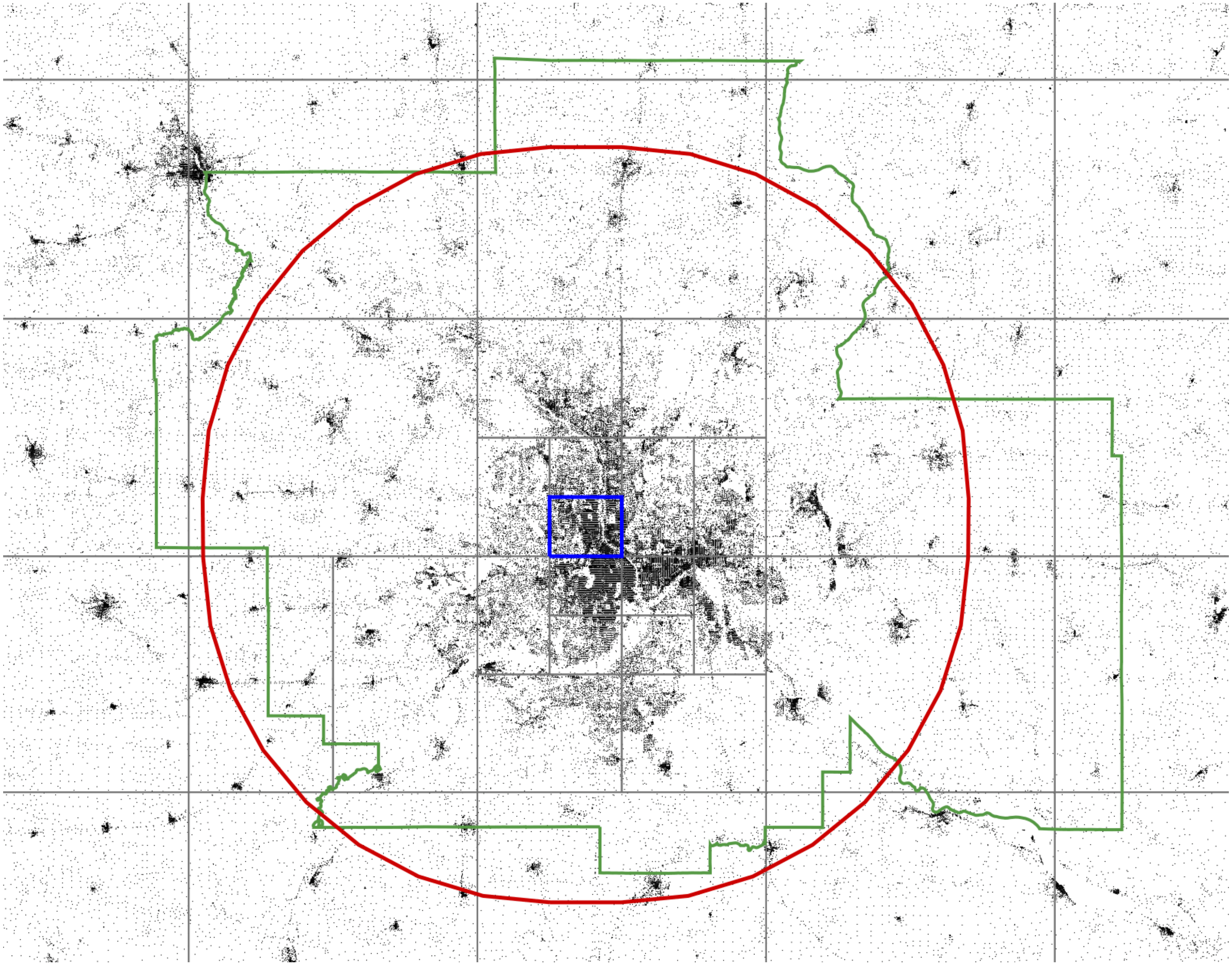

Figure 4: A single origin zone (blue) and its corresponding 60-kilometer destination zone (red). Travel times are calculated from each centroid in the origin zone to each centroid in the destination zone.

Each analysis zone defines a set of origins and a set of destinations. The origins for an analysis zone are simply those Census blocks whose centroids fall within the zone. All Census blocks whose centroids lie within 60km of the boundary of the analysis zone are included as destinations. This corresponds to an average speed of 60 km/hour; in 2011, U.S. bus service operated at an average speed of 20.4 km/hour, heavy rail operated at an average speed of 32.2 km/hour, and commuter rail operated at an average speed of 52.6 km/hour. (Dickens et al., 2013) Figure 4 provides an example of origin and destination selection for a single analysis zone in the Minneapolis area.

3.2 Graph Building

Travel time calculations in this project are performed using the OpenTripPlanner (OTP) software, described in more detail in Section 4.2. OTP includes a graph building function that combines pedestrian network data from OpenStreetMap and transit network and schedule data in GTFS format into a single unified graph. A graph is built for each analysis zone, including all relevant transit schedules as described above. This is combined with origin and destination locations to create a single analysis bundle that contains all data necessary to calculate accessibility values for the blocks in a single analysis zone. These analysis bundles are then easily transmitted for remote computation on computer clusters.

4 Accessibility Calculation

4.1 Overview

Accessibility evaluations rely on an underlying calculation of travel times. Here, transit travel times are evaluated from each Census block centroid based on a detailed pedestrian network and published transit schedule data. Travel time calculations are repeated for every departure time between 7 and 9 AM at one-minute intervals. These travel times are the basis of a cumulative opportunities accessibility measure which counts the number of opportunities (in this case, jobs) reachable from each origin within 10, 20, 30, 40, 50, and 60 minutes. The accessibility values for all departure times are averaged to indicate the number of jobs that are reachable, on average, within a given travel time threshold between 7 and 9 AM.

This block-level dataset provides a *locational* measure of accessibility—it indicates how many jobs can be reached from different points in space. This location measure is then weighted by the number of workers residing in each Census block and averaged across the entire metro area to produce *worker-weighted* accessibility. This metric indicates the accessibility that is experienced by the average worker in the metropolitan area.

Finally, the worker-weighted average accessibility values across the 10 through 60 minute thresholds are averaged for each metropolitan area to produce a weighted accessibility ranking.

Earlier evaluations of transit accessibility across multiple cities include Tomer et al. (2011) and Ramsey and Bell (2014). This evaluation incorporates four key advances relative to earlier work. First, it calculates accessibility for multiple departure times, rather than assuming a single departure time. This allows the final metrics to reflect the effects of service frequency, which is a critical determinant of transit's usefulness. Second, it calculates travel times at the block rather than the block group level, providing a significant increase in spatial resolution. This is important because most transit access and egress trip segments occur by walking; distances easily traveled by pedestrians are short relative to the size of block groups, which can distort travel time calculations. Third, it provides accessibility metrics for multiple travel time thresholds, rather than selecting a single threshold. And finally, the 2015 evaluation introduced a national scope, yielding data for every census block in the United States.

The following sections describe the specific tools, algorithms, and parameters that were used to produce the data presented in *Access Across America: Transit 2021*.

4.2 Travel Times

4.2.1 Software

Transit travel time calculations are performed using OpenTripPlanner (OTP), an open-source multimodal trip planning and analysis tool. OpenTripPlanner is a graph-based transit routing system that operates on a unified graph including links representing road, pedestrian, and transit facilities and services. OTP is available at `http://opentripplanner.org` and is described and evaluated in Hillsman and Barbeau (2011). OTP's Analyst extension provides efficient and parallelized processing of many paths from a single origin based on the construction of shortest path trees using Dijkstra's Algorithm. Additionally, locally-developed extensions to OTP allow automated batch processing of accessibility calculations for multiple departure times.

4.2.2 Transit Trip Parameters

The time cost of travel by transit is composed of several components. *Initial access time* refers to the time cost of traveling from the origin to a transit stop or station. *Initial wait time* refers to the time spent after reaching the transit station but before the trip departs. *On-vehicle time* refers to time spent on-board a transit vehicle. When transfers are involved, *transfer access time* and *transfer wait time* refer to time spent accessing a secondary transit station and waiting there for the connecting trip. Finally, *destination access time* refers to time spend traveling from the final transit station to the destination. All of these components are included in the calculation of transit travel times.

This analysis makes the assumption that all access portions of the trip—initial, transfer(s), and destination—take place by walking at a speed of 5 km/hour along designated pedestrian facilities such as sidewalks, trails, etc. On-vehicle travel time is derived directly from published transit timetables, under an assumption of perfect schedule adherence.

An unlimited number of transfers are allowed. This is somewhat unusual among evaluations of transit accessibility. In many cases travel times are limited to trips involving no more than one or two transfers; this is justified by the observation that in most cities a very large majority (often over 90%) of observed transit trips involve no more than two transfers. However, the shortest-path algorithms typically employed in these evaluations are single-constraint algorithms: they are guaranteed to find the shortest path only when given a single constraint (typically, travel time). When the path search tree is pruned based on an additional constraint such as number of transfers (or, in some cases, transfer wait time), these algorithms provide no insurance against a shorter trip, requiring additional transfers, remaining undiscovered in the pruned space (Korkmaz and Krunz, 2001; Kuipers et al., 2002).

Given the realities of transit networks, it likely that cases where (for example) a three-transfer itinerary provides a faster trip than a two-transfer itinerary are relatively rare. However, given the goal of evaluating the full accessibility provided by a transit system rather than simply the accessibility that is likely to be utilized, this analysis prefers the algorithmically correct approach of using travel time as the single routing constraint and leaving the number of transfers unconstrained.

Just as there is no upper limit on the number of vehicle boardings, there is no lower limit either. Transit and walking are considered effectively a single mode. The practical implication of this is that the shortest path by "transit" is not required to include a transit vehicle. This allows the most consistent application and interpretation of the travel time calculation methodology. For example, the shortest walking path from an origin to a transit station in some cases passes through potential destinations where job opportunities exist. In other cases, the shortest walking path from an origin to a destination might pass through a transit access point which provides no trips which would reduce the origin–destination travel time. In these situations, enforcing a minimum number of transit boardings would artificially inflate the shortest-path travel times. To avoid this unrealistic requirement, the transit travel times used in this analysis are allowed to include times achieved only by walking.

4.3 Cumulative Opportunities

Many different implementations of accessibility measurement are possible. El-Geneidy and Levinson (2006) provide a practical overview of historical and contemporary approaches. Most contemporary implementations can be traced at least back to Hansen (1959), who proposes a measure where potential destinations are weighted by a gravity-based function of their access cost and then summed:

$$A_i = \sum_j O_j f\left(C_{ij}\right) \tag{1}$$

A_i = accessibility for location i

O_j = number of opportunities at location j

C_{ij} = time cost of travel from i to j

$f\left(C_{ij}\right)$ = weighting function

The specific weighting function $f\left(C_{ij}\right)$ used has a tremendous impact on the resulting accessibility measurements, and the best-performing functions and parameters are generally estimated independently in each study or study area (Ingram, 1971). This makes comparisons between modes, times, and study areas challenging. Levine et al. (2012) discuss these challenges in depth during an inter-metropolitan comparison of accessibility; they find it necessary to estimate weighting parameters separately for each metropolitan area and then implement a second model to estimate a single shared parameter from the populations of each. Geurs and Van Wee (2004) also note the increased complexity introduced by the cost weighting parameter.

Perhaps the simplest approach to evaluating locational accessibility is discussed by Ingram (1971) as well as Morris et al. (1979). *Cumulative opportunity* measures of accessibility employ a binary weighting function:

$$f\left(C_{ij}\right) = \begin{cases} 1 & \text{if } C_{ij} \leq t \\ 0 & \text{if } C_{ij} > t \end{cases} \tag{2}$$

t = travel time threshold

Accessibility is calculated for specific time thresholds and the result is a simple count of destinations that are reachable within each threshold. Owen and Levinson (2012) demonstrate this approach in an accessibility evaluation process developed for the Minnesota Department of Transportation. Using the results of the travel time calculations described in Section 4.2, cumulative opportunity accessibility values are calculated for each Census block in each CBSA using thresholds of 5, 10, 15, 20, …, 60 minutes.

4.4 Time-Averaged Accessibility

Accessibility by transit is strongly dependent on departure time because of the scheduled nature of transit service. For example, if a transit route's service frequency is 20 minutes, then immediately after a vehicle departs all destinations become 20 minutes "farther away." Figure 5 illustrates the fluctuations in accessibility measured at a single Census block in the Minneapolis–Saint Paul metropolitan area

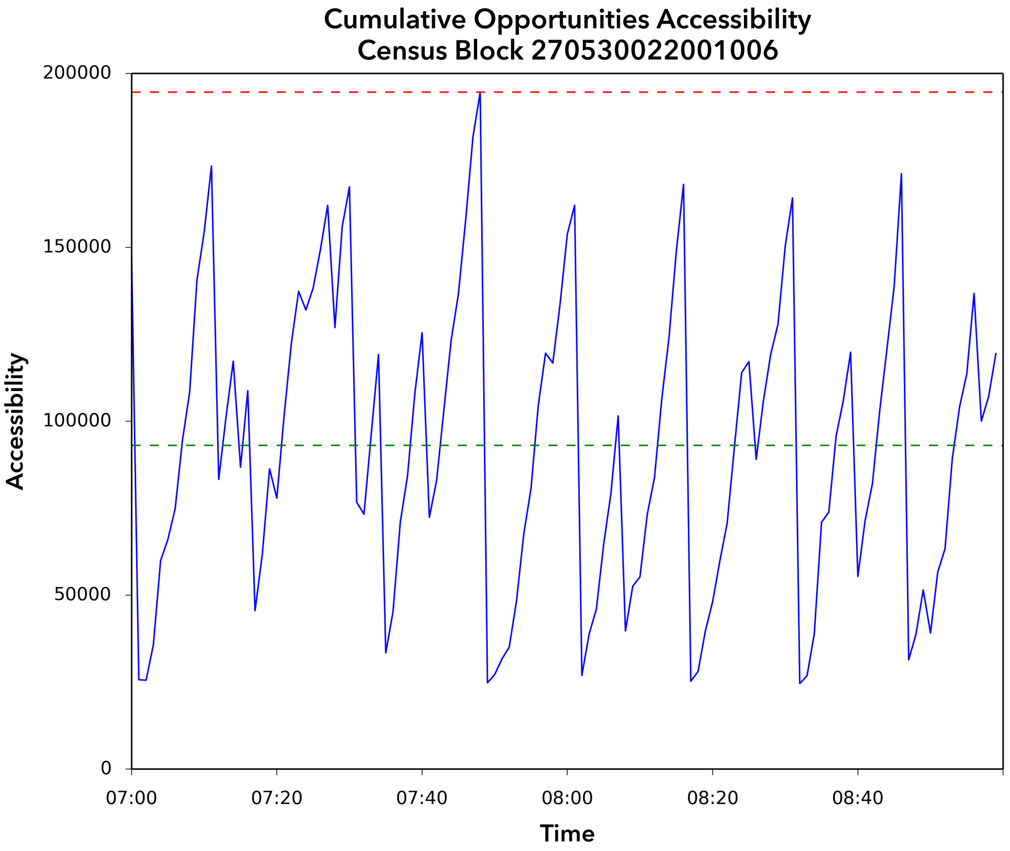

Figure 5: Transit accessibility between 7:00 AM and 9:00 AM for a single Census block. Red line indicates maximum accessibility value; green line indicates average accessibility value.

between 7:00 AM and 9:00 AM.

To address this and to reflect the influence of transit service frequency on accessibility, travel times are calculated repeatedly for each origin-destination pair using each minute between 7:00 and 9:00 AM as the departure time, and an accessibility value is calculated using each travel time result. The accessibility results are averaged to represent the expected accessibility value that would be experienced by a traveler departing at a random time in this interval.

4.5 Person-Weighted Accessibility

The accessibility calculation methods described in the sections above provide a *locational* accessibility metric—one that describes accessibility as a property of locations. The value of accessibility, however, is only realized when it is experienced by people. To reflect this fact, accessibility is averaged across all blocks in a CBSA, with each block's contribution weighted by the number of workers in that block. The result is a single metric (for each travel time threshold) that represents the accessibility value experienced by an average worker in that CBSA.

4.6 Weighted Accessibility Ranking

Metropolitan area rankings are based on an average of person-weighted job accessibility for each metropolitan area over the twelve travel time thresholds. In the weighted average of accessibility, destinations reachable in shorter travel times are given more weight, as they constitute more attractive destinations. A negative exponential weighting factor is used, following Levinson and Kumar (1994). Here time is differenced by thresholds to get a series of "donuts" (e.g. jobs reachable from 0 to 10 minutes, from 10 to 20 minutes, etc.).

$$a_w = \sum_t (a_t - a_{t-10}) \times e^{\beta t}$$

$a_w =$ Weighted accessibility ranking metric for a single metropolitan area
$a_t =$ Worker-weighted accessibility for threshold t
$\beta = -0.08$

References

Dickens, M., Neff, J., and Grisby, D. (2013). 2013 public transportation fact book. Technical report, American Public Transit Association.

El-Geneidy, A. M. and Levinson, D. M. (2006). Access to destinations: Development of accessibility measures. Technical Report MN/RC-2006-16, Minnesota Department of Transportation.

Geurs, K. T. and Van Wee, B. (2004). Accessibility evaluation of land-use and transport strategies: review and research directions. *Journal of Transport Geography*, 12(2):127–140.

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

Hillsman, E. L. and Barbeau, S. J. (2011). Enabling cost-effective multimodal trip planners through open transit data. Technical Report USF 21177936, National Center for Transit Research.

Ingram, D. R. (1971). The concept of accessibility: A search for an operational form. *Regional Studies*, 5(2):101–107.

Korkmaz, T. and Krunz, M. (2001). Multi-constrained optimal path selection. In *Twentieth Annual Joint Conference of the IEEE Computer and Communications Societies*, volume 2, pages 834–843. IEEE.

Kuipers, F., Van Mieghem, P., Korkmaz, T., and Krunz, M. (2002). An overview of constraint-based path selection algorithms for qos routing. *Communications Magazine, IEEE*, 40(12):50–55.

Levine, J., Grengs, J., Shen, Q., and Shen, Q. (2012). Does accessibility require density or speed? A comparison of fast versus close in getting where you want to go in us metropolitan regions. *Journal of the American Planning Association*, 78(2):157–172.

Levinson, D. M. and Kumar, A. (1994). Multimodal trip distribution: structure and application. *Transportation Research Record*, pages 124–124.

Morris, J. M., Dumble, P., and Wigan, M. R. (1979). Accessibility indicators for transport planning. *Transportation Research Part A: General*, 13(2):91–109.

Owen, A. and Levinson, D. (2012). Annual accessibility measure for the Twin Cities metropolitan area. Technical Report 2012-34, Minnesota Department of Transportation.

Ramsey, K. and Bell, A. (2014). The smart location database: A nationwide data resource characterizing the built environ-ment and destination accessibility at the neighborhood scale. *Cityscape: A Journal of Policy Development and Research*, 16(2).

Tomer, A., Kneebone, E., Puentes, R., and Berube, A. (2011). Missed opportunity: Transit and jobs in metropolitan america. Technical report, Brookings Institution, `http://www.brookings.edu/~/media/research/files/reports/2011/5/12%20jobs%20and%20transit/0512_jobs_transit.pdf`.