

# UrbanFootprint Technical Documentation

## Walk and Transit Accessibility Analysis

---

### Overview

The Walk and Transit Accessibility modules highlight opportunities and deficiencies in accessibility by measuring the proximity of each canvas geometry (parcel or census block) in a project canvas to employment, amenities, population, and housing.

The modules are loaded with up-to-date network data to support robust analysis of existing conditions and future scenarios. The Walk and Transit Accessibility modules use data for walkable roads and walk paths. Additionally, the Transit Accessibility module incorporates an existing transit network composed of the transit service providers that operate routes in the project area.

The proximity to amenities and accessibility to opportunities are measured based on the walk network, in the Walk Accessibility module, or in combination with the transit network in the Transit Accessibility module. The modules report two categories of accessibility metrics:

- Travel time, in minutes, from a parcel or census block to the nearest point of interest (POI). It is the time necessary to reach a particular POI based on the shortest distance within the network that is provided to the module.
- Number of amenities or opportunities one can reach within a specified amount of time.

Metrics include, for example, the average time needed to walk from a given parcel to the nearest hospital, or the number of jobs that can be reached on foot or by transit from that parcel within 30 minutes.

The Accessibility modules can be run on alternative scenarios as well. Currently, scenario results are based on existing roads, walk paths, and transit, and reflect changes in land use patterns as indicated by scenarios as painted. Future enhancements will allow for the modeling of synthetically applied representative transit and road networks to highlight how new and modified transportation networks can impact general and region-specific accessibility metrics.

Analysis allows users to assess how scenarios would change a region's overall accessibility, how their project area performs compared to the entire region, and how scenarios compare to each other.

---

## Transit Accessibility Methodology

The Transit Accessibility module measures proximity to amenities and accessibility to specific features or opportunities by foot and transit. To do this, the module first compiles an integrated network, composed of both walk path and transit service data. The segments of the network are determined based on attributes that are used to determine, in terms of time, how long it takes to traverse a given segment, based on reported schedule times (for transit) or walk speed and segment length (for walk portions).

Then location data, such as the location of schools, parks, and hospitals are extracted from national reference datasets and resolved against the users painted scenarios (as represented in their scenario canvas). In addition to this information, the canvas is also used to generate data related to locations of jobs, households and population. These datasets are all then layered on top of the network based on their locations in relation to nodes of the network (Figure 2). Lastly, a set of algorithms are performed to produce key proximity and accessibility metrics.

### Requirements

The Transit Accessibility module takes the following data and parameters:

- The project canvas (automatically included with the project scope)
- General Transit Feed Specification (GTFS) zip files of schedule and stop data for transit operators (automatically generated)
- (Optional) A synthetic transit JavaScript Object Notation (JSON), which features metadata that describes the performance of a planned network extension

### Calculations

First, an integrated network of walk and transit paths, one of the key inputs to Transit Accessibility module, is built using two sources of data. The walk network is built based on data sourced from OpenStreetMap (OSM)<sup>1</sup> and the transit network is assembled based on transit data that are retained in the General Transit Feed Specification (GTFS)<sup>2</sup> format and sourced from TransitLand<sup>3</sup>, the largest transit data repository on the web.

To compile a transit network for a project area; first, transit routes by different transit operators are identified for a project area. UrbanFootprint assembles and regularly updates a dataset of all transit operators and their transit service coverages in the United States and this data is extracted from GTFS feeds that contain transit routes information. With that national reference, transit routes and schedules by all transit operators in a project area can be easily determined by finding all operators who have a transit service coverage area that overlaps with the target project area.<sup>4</sup>

---

1 OpenStreetMap is a free, editable map of the whole world that is being built by volunteers largely from scratch and released with an open-content license. Learn more here: [https://wiki.openstreetmap.org/wiki/About\\_OpenStreetMap](https://wiki.openstreetmap.org/wiki/About_OpenStreetMap)

2 The General Transit Feed Specification (GTFS) defines a common format for public transportation schedules and associated geographic information. Learn more here: <https://developers.google.com/transit/gtfs/>

3 TransitLand is a community-edited data service aggregating transit networks across metropolitan and rural areas around the world. Learn more here: <https://transit.land/>

4 Note that, in the future, users will be able to submit their own transit data and customize the included list of transit routes that are included in each scenario. This will be useful if users replace or update an existing service provider's transit network with, for example, new routes or increased headways. Simulating the removal of a specific transit network will also be possible.

Figure 1. Accessibility Analysis Flow

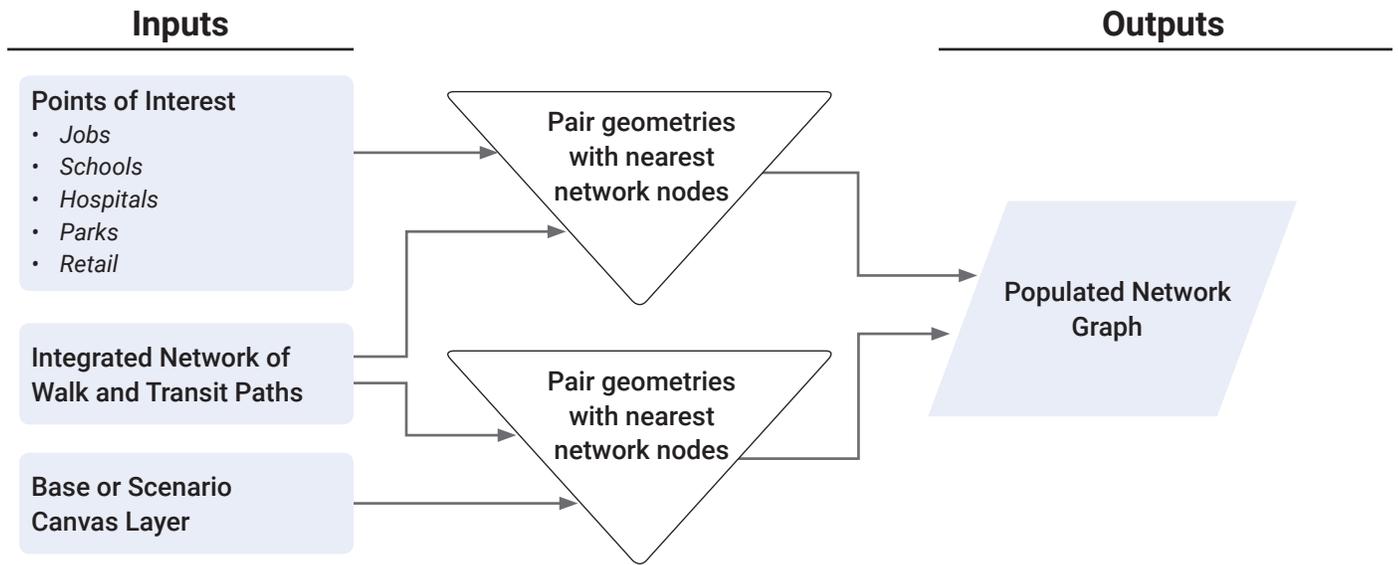
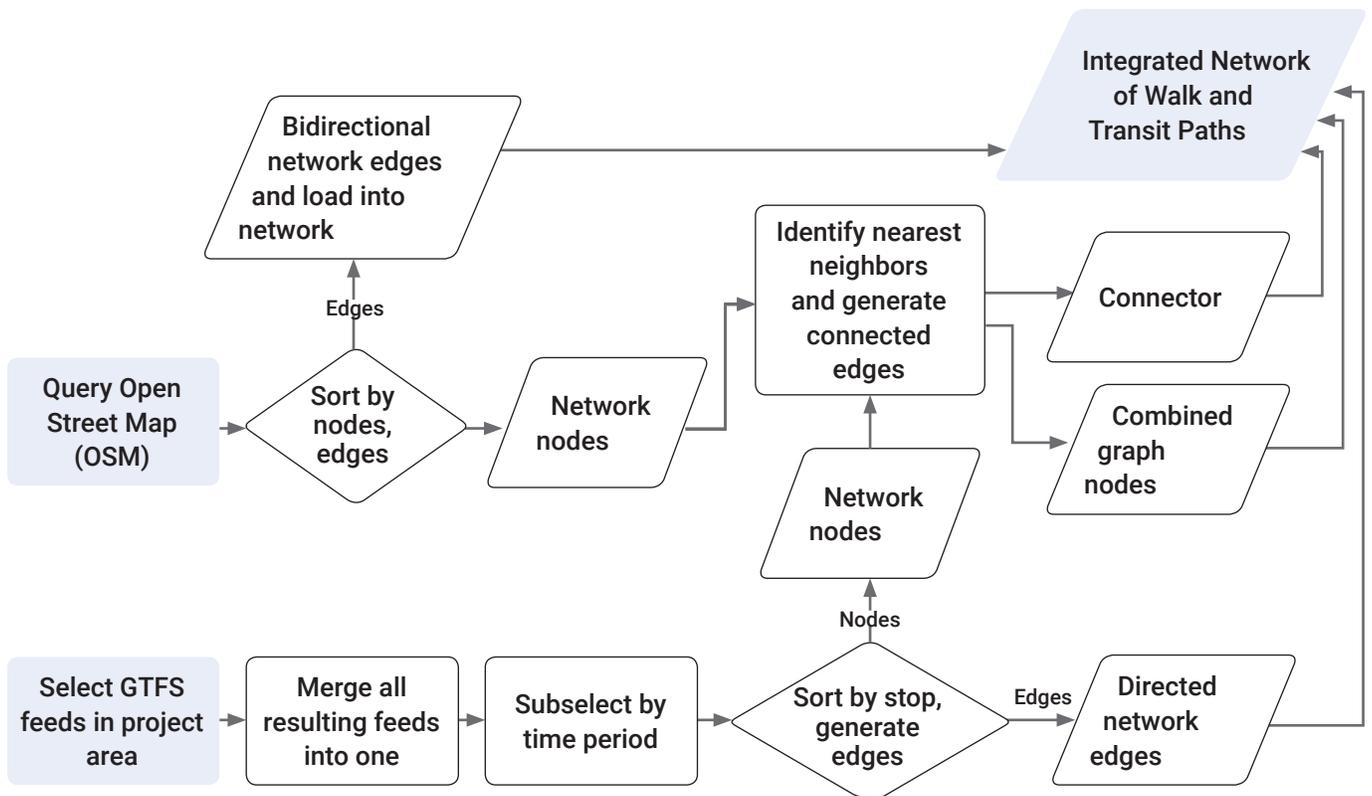


Figure 2. Generation of the Integrated Walk and Transit Network



After specific transit operators are identified for a project area, service routes and stop times are then selected from each transit operator schedule given a parameterized target period. The default target time frame is 7:00 AM–10:00 AM. The day with the busiest typical service schedule pattern is selected. This is determined by finding the day with the most bus service scheduled, and typically corresponds with a Tues-Thurs weekday schedule.

Based on the identified transit service routes and schedules in that specified time period, a transit network is then built by calculating travel time cost between stops and wait time cost at stops. In a transit network, a node is a transit stop and an edge is a path between two transit stops if there is any transit service operating in between. This step of the analysis module estimates:

1. Edge cost, which is the time required to travel from one stop to the next via transit based on the published schedule data, and
2. Stop cost, which is the half of the average time between arrivals for a specific line at a specific stops

With summarized stop and edge costs, a network graph that represents the transit system in the project area can then be constructed.<sup>5</sup> Two sets of information are stored in this network graph. Stop locations are stored as nodes, with attributes including locations, in latitudes and longitudes, and waiting costs. An edge between a pair of nodes, if there is transit service between these two nodes, is stored in the network graph, with attributes including direction<sup>6</sup> and cost, measured in terms of time.

The walk network is constructed similarly. Nodes along known walk paths and roads represented in the OpenStreetMap (OSM) data are converted to nodes in the walk network, along with locations. OSM “ways” (paths) between two given nodes are introduced to the network as two distinct directional edges with attributes of costs, which are derived based on distance between two nodes, and an assumed walk speed (which is a parameter that the user can adjust).

To generate the integrated network, walk and transit networks are connected. To connect the transit network to the walk network, each node in the transit network is paired with its nearest node in the walk network. Direct paths between these two nodes, one from the transit node to the walk node and another from the walk node to the transit nodes are then created, and such paths create connections from the transit network to the walk network. Traversing through these paths resemble boarding and alighting transit system in reality. However, if a transit stop is too far from its nearest walk node<sup>7</sup>, then no connector path is built between two.

By default, if it takes more than 5 minutes to walk from a transit node to its nearest walk node<sup>8</sup>, then no such connector edge is built. These connector edges are stored in the network graph with attributes of costs, just like other edges in transit and walk networks. Edges from a walk node to a transit node have the same cost as the boarding cost for the transit node (half that of the average headway). The intent of this is to model the cost of

---

5 The method for summarizing edge and stop costs parallels those used in the open source tool Peartree (<https://github.com/kuanb/peartree/>). With this tool, schedule information is aggregated from the target time period and converted to a direct multigraph representation of that time frame’s scheduled service patterns. Impedance along each segment of each route is a function of the service level (headways, route segment speed) for that segment.

6 An edge from Node A to B is different from an edge from Node B to A.

7 This may occur, for example, when the transit network happens to be bigger than the walk network.

8 Consider crow-fly distance.

moving from the walk network onto the transit network (e.g. the cost to board the transit service in terms of the amount of time that one would on average need to wait).

The return edge, from a transit node to a walk node, is “free.” That is, no cost is assessed there as no wait time is modeled for disembarking a transit service. The creation of these connector edges allows the two networks to be merged into a single integrated network.

Once this integrated network has been developed, points of interest (POIs), including parks, schools, and hospitals, can be added to the network. These POIs are associated with one or more nodes on the network on the basis of either locations. If they are point locations, like hospitals and schools, a single point is used. If they are large polygons, like parks, then the shape is disaggregated into a representative “cloud” of points which are assigned to the network, each representing a subset of the total park (in terms of, for example, acreage).

Base canvas geometries are similarly associated with nodes on the integrated network. Specifically, the closest walk node<sup>9</sup> is identified for each census block or a parcel based on the distance between the centroid of the canvas geometry and any walk node. By associating POIs and canvas geometries to network nodes, one can then quickly summarize number of amenities, such as hospitals, and opportunities, such as jobs, that are accessible from any given node in the network (within a given threshold of time).

With the fully specified network and the POIs information loaded onto the network, costs from any given census block or parcel to any other census block or parcel or any given amenity can be generated via the graph.

Pandana, an open source software library, is used to perform accessibility calculations. Pandana utilizes optimizations from Open Source Routing Engine (OSRM) to enable performant traversal of the computed graph for holistic, project-wide analyses. To learn more about the software, please review the journal article cited in the footnotes and in the citations section at the end of this document.<sup>10</sup>

## Model Logic Assumptions

- Because schedule headways are averaged, headways between trips on a given transit route are assumed equal during the GTFS subschedule time period.
- Because elevation has not yet been included as a factor in network walk traversal, walk speed is considered static without regard for the slope of a given network edge.
- The cost of transfers between two transit routes is always assumed to be half the headway of the later route. Timed transfers and awareness of schedule synchronization has not yet been accounted for.
- Programmatically added links between transit network stop nodes and walk network nodes are assumed to be traversable, but OSM data for barriers (e.g., low walls) have not yet been considered.
- The maximum travel time to be considered is 60 minutes. Network analysis does not consider paths between an origin and a destination that take longer than 60 minutes to traverse.

---

<sup>9</sup> This implies one always has to be on a walk network first in order to get to a transit network.

<sup>10</sup> <http://onlinepubs.trb.org/onlinepubs/conferences/2012/4thITM/Papers-A/0117-000062.pdf>

- The decay method used in the network analysis is flat, i.e. amenities nearer to the reference census block or parcel are given the same weight as amenities further to it.

---

## Walk Accessibility Methodology

The Walk Accessibility module has the same structure as the Transit Accessibility module, except that the Transit Accessibility module incorporates the transit network via GTFS transit schedule data. Refer to the methodology description for the Transit Accessibility section for more details.

The walk module does not import transit schedules to construct the transit network, or integrate the transit network and the walk network<sup>11</sup>. Instead, it begins network analysis with only the walk network that is built upon the OSM data.

---

## Input Parameters

Inputs to the Walk and Transit Accessibility modules include General Transit Feed Specification (GTFS) transit and OpenStreetMap (OSM) street network data, along with user-modifiable parameters for transit service hours and walk speed.

## Transit Network

UrbanFootprint source GTFS data contain relevant transit operator information (including stop locations and route schedules) from TransitLand<sup>12</sup> via its transit feeds API<sup>13</sup>. UrbanFootprint extracts, cleans, and stores these feeds on a regular basis and the Transit Accessibility module constructs the transit network based on these feeds. UrbanFootprint's loaded transit network thus represents the known transit network that is accessed from all available GTFS feeds published by operators in North America.

## Street Network

OpenStreetMap (OSM) data is downloaded from the OSM Overpass API<sup>14</sup>. As a result, the network is representative of the latest network available from OSM. OSM data includes two key elements that are used to construct the network graph: nodes<sup>15</sup>, which are key points along a given path, and "ways." Ways<sup>16</sup> (also known as edges) are used to create directional links from one node to another; the links form the edges of the network graph.

Three layer of POIs are used in both Accessibility modules.

- Hospitals: US Homeland Infrastructure Foundation, Federal Geographic Data Committee
- Parks: Esri, TomTom North America, Inc.; parks, gardens, and forests within the US

---

<sup>11</sup> The analysis does currently rely on importing GTFS to extract transit stop locations so that minutes to nearest transit stop can be generated for Walk Accessibility module.

<sup>12</sup> <https://transit.land/documentation/>

<sup>13</sup> TransitLand collects existing geographic and temporal data from authoritative sources. They have connected the dots with common identifiers, and allowed crowdsourced editing and validation of their federated database via a "community datastore."

<sup>14</sup> <http://overpass-turbo.eu/>

<sup>15</sup> <http://wiki.openstreetmap.org/wiki/Elements>

<sup>16</sup> <http://wiki.openstreetmap.org/wiki/Way>

- Schools: US National Center for Education Statistics (NCES), School Attendance Boundary Information System (SABINS)

## User-Modifiable Parameters

Users can adjust the timeframe considered for transit access, and the assumed average walk speed. By default, the hours during which transit accessibility is assessed is 7 AM to 10 AM. The default average walk speed is 3 miles per hour.

Users can also specify the  $n$ th (e.g., 1st, 2nd, 3rd, 4th, etc... where  $n$  is user-specified) nearest retail location considered for the Nearest Retail Access metrics. For example, if  $n$  is set to 5, then the module will measure the minutes it takes to access the fifth nearest retail location. By default,  $n$  is set to 1. Users can set the value up to 10.

---

## Output Metrics

The Walk and Transit Accessibility modules generate mapped spatial output layers and corresponding data tables; all can be used within UrbanFootprint for mapping and data exploration, and exported. Separate sets of spatial layers are generated for walk and transit accessibility. Results for both include three categories of accessibility metrics: aggregate access, nearest point of interest access, and regional proportion of amenities accessible, as described in the following sections.

The module also reports individual and comparative scenario results via summary charts, and generates spreadsheet summaries in Excel format. The attributes of the spatial output/data tables are summarized in Table 1.

## Aggregate Access

These results indicate the number of opportunities and the amount of certain amenities each parcel of census block can reach in the allotted time.

- Opportunities include jobs, households, and population.
- Amenities currently includes acres of parks space.

For example, from a given parcel in a project, one might see that the parcel has, via walking, access to 120 jobs in 5 minutes and 280 jobs in 10 minutes. In addition, if users paint land uses to reflect job growth in a scenario, aggregate access will increase to reflect such changes (assuming a specific canvas geometry is able to reach within the given time frame the other parcel of census block that was painted to feature an increase, or decrease, in employment for example). These values are visible as a mapped layer in which amount of opportunities/amenities accessible is represented by color variation (e.g. a choropleth map).

## Nearest Point of Interest Access

These results indicate the amount of time, measured in minutes, required to reach the nearest of a given type of amenity (or in the case of retail locations, the nearest  $n$ th location as set by the user). For example, the “Nearest Hospital By Mode” results show how easily any place in a project can access a hospital. These values are visible as a mapped layer as well (again, a choropleth map). With the default coloration, the lighter colors correspond to smaller amount of time (relative to other sites) to reach the nearest hospital. Accessibility measures also reflect changes in destinations. For example, the module outputs will change if a new school, hospital, park, or retail<sup>17</sup> destination is added to the scenario’s base canvas. These results indicate the amount of time on foot and/or transit that is required to reach the nearest hospital from each parcel or census block in that base canvas.

## Regional Proportions with Amenities Nearby Metrics

These results are prepared as summary charts and show percentages of the total population in the project area that can reach their nearest amenity within a given period of time. For example, the walk module determines the portion of the total population that can reach their nearest park on foot within 5, 10, 15, and 20 minutes.

## Regional Proportion of Amenities Accessible

These results are prepared as summary statistics on the right-hand side of the UrbanFootprint application window (as well as in the analysis module reporting charts). They reflect accessibility-related impact of revised land use patterns or existing conditions (e.g. when run in the base canvas). The resulting figures indicate what portion of the population can reach equal to or greater than the target percentage of all jobs in the project area within the value of the set bucket of time. For example, these values can be reported in the following manner:

*“42% of the population of the project area can reach at least 25% of all jobs in the project area within 30 minutes.”*

There are three charts in this category. Each assumes a different percentage of the project area’s jobs that should be considered (that is, 10%, 25%, and 50% of all jobs in the project area). For each assumed percentage of jobs considered, the portion of the project area’s population that can access these jobs is then assessed against four time buckets: 15, 30, 45, and 60 minutes.

Please note that this data is only compiled for the transit accessibility module, and not the walk module.

The Accessibility module generates a mapped spatial output layer and corresponding data table; both can be used within UrbanFootprint for mapping and data exploration, and exported. The module also reports individual and comparative scenario results via summary charts, and generates a spreadsheet summary in Excel format. The attributes of the spatial output/data table are summarized in Table 1, with an indication of attributes that are mappable or not.

---

<sup>17</sup> A location is considered retail if there is any retail employee at the location.

**Table 1: Walk and Transit Accessibility Module Outputs**

<b>Attribute(s)</b>	<b>Description</b>	<b>Walk Times</b>	<b>Transit Times</b>
<b>Aggregate Access Metrics (mappable)</b>			
Employment Access by Mode	Number of jobs that can be reached	5, 10, 15, 20, 30, 45 & 60 minutes	15, 30, 45, & 60 minutes
Population Access by Mode	Number of individuals that can be reached	5, 10, 15, 20, 30, 45 & 60 minutes	15, 30, 45, & 60 minutes
Household Access by Mode	Number of households that can be reached	5, 10, 15, 20, 30, 45 & 60 minutes	15, 30, 45, & 60 minutes
Park Acreage Access by Mode	Acres of park area that can be reached	5, 10, 15, 20, 30, 45 & 60 minutes	15, 30, 45, & 60 minutes
<b>Nearest Point of Interest Access Metrics (mappable)</b>			
Nearest School by Mode	Amount of time to nearest school		
Nearest Hospital by Mode	Amount of time to nearest hospital		
Nth nearest Retail by Mode	Amount of time to <i>n</i> th nearest retail, where <i>n</i> is user-specified		
Nearest Park by Mode	Amount of time to nearest park		
Nearest Transit Stop by Mode (for walk accessibility only)	Amount of time to nearest transit stop		
Nearest High Quality Transit Stop by Mode (for walk accessibility only)	Amount of time to nearest high quality transit, which includes types of fixed-route transit		
<b>Regional Proportion of Jobs Accessible Metrics (summary statistics only, not mappable)</b>			
Access to 10+% of project area jobs	Portion of base canvas from which at least this percentage of area jobs within the set time frames		15, 30, 45, & 60 minutes
Access to 25+% of project area jobs	Portion of base canvas from which at least this percentage of area jobs can be reached within the set time frames		15, 30, 45, & 60 minutes
Access to 50+% of project area jobs	Portion of base canvas from which at least this percentage of area jobs can be reached within the set time frames		15, 30, 45, & 60 minutes

Attribute(s)	Description	Walk Times	Transit Times
<b>Regional Proportions with Amenities Nearby Metrics (summary statistics only, not mappable)</b>			
Access rates to schools	Portion of residents that can access a school within the set time frames	5, 10, 15, & 20 minutes	10, 20, & 30 minutes
Access rates to hospitals	Portion of residents that can access a hospital within the set time frames	5, 10, 15, & 20 minutes	10, 20, & 30 minutes
Access rates to retail	Portion of residents that can access the <i>n</i> th nearest retail location within the set time frames	5, 10, 15, & 20 minutes	10, 20, & 30 minutes
Access rates to parks	Portion of residents that can access a park within the set time frames	5, 10, 15, & 20 minutes	10, 20, & 30 minutes
Access rates to all transit stops	Portion of residents that can access a transit stop within the set time frames	5, 10, 15, & 20 minutes	10, 20, & 30 minutes
Access rates to high quality transit stops	Portion of residents that can access a high quality transit stop within the set time frames	5, 10, 15, & 20 minutes	10, 20, & 30 minutes

---

## Final Citations

In an effort to highlight the utility of a core component of the accessibility model, the graph engine for computing accessibility in a performant manner, it's important to highlight and acknowledge the use of Pandana, an open-source tool maintained within the Urban Data Science Toolkit organization on Github (<https://github.com/UDST>). The repository for the Pandana source code (<https://github.com/UDST/pandana>) requests the following paper, describing the underlying methodology, be included when referring to the library's methodology. It is included, in-line, in the footnotes, as well as here, in the below citation:

Foti, F., Waddell, P., & Luxen, D. (2012, April). A generalized computational framework for accessibility: from the pedestrian to the metropolitan scale. In Proceedings of the 4th TRB Conference on Innovations in Travel Modeling. Transportation Research Board.

Link: <http://onlinepubs.trb.org/onlinepubs/conferences/2012/4thITM/Papers-A/0117-000062.pdf>

In addition to using Pandana to hold the initialized network graph, it is also necessary to acknowledge Dennis Luxen and the Open Source Routing Machine project (OSRM). This open source project's contraction hierarchies logic is employed to process the initial network graph and compress it to enable higher performance network queries when Pandana aggregates accessibility metrics tethered to graph nodes. This project can also be found on Github, at the following address: <https://github.com/Project-OSRM/osrm-backend>.