UC Davis
Research Reports

Title
BikewaySim Technology Transfer: City of Atlanta, Georgia

Permalink
https://escholarship.org/uc/item/23n9389j

Authors
Passmore, Reid
Watkins, Kari E.
Guensler, Randall

Publication Date
2021-12-01

DOI
10.7922/G2CF9NDV

Data Availability
The data associated with this publication are available at:
https://doi.org/10.5281/zenodo.5750140

eScholarship.org
Powered by the California Digital Library
University of California

BikewaySim Technology Transfer: City of Atlanta, Georgia

December 2021

A Research Report from the National Center for Sustainable Transportation

Reid Passmore, Georgia Institute of Technology
Dr. Kari E. Watkins, Georgia Institute of Technology
Dr. Randall Guensler, Georgia Institute of Technology

National Center for Sustainable Transportation
Georgia Tech School of Civil and Environmental Engineering
College of Engineering

TECHNICAL REPORT DOCUMENTATION PAGE

1. Report No. NCST-GT-RR-21-26	**2. Government Accession No.** N/A	**3. Recipient's Catalog No.** N/A

4. Title and Subtitle BikewaySim Technology Transfer: City of Atlanta, Georgia	**5. Report Date** December 2021
	6. Performing Organization Code N/A
7. Author(s) Reid Passmore, https://orcid.org/0000-0001-6602-2702 Kari Watkins, PhD, https://orcid.org/0000-0002-3824-2027 Randall Guensler, PhD, https://orcid.org/0000-0003-2204-7427	**8. Performing Organization Report No.** N/A
9. Performing Organization Name and Address Georgia Institute of Technology School of Civil and Environmental Engineering 790 Atlantic Dr, Atlanta, GA 30332	**10. Work Unit No.** N/A
	11. Contract or Grant No. USDOT Grant 69A3551747114
12. Sponsoring Agency Name and Address U.S. Department of Transportation Office of the Assistant Secretary for Research and Technology 1200 New Jersey Avenue, SE, Washington, DC 20590	**13. Type of Report and Period Covered** Draft Final Report
	14. Sponsoring Agency Code USDOT OST-R

15. Supplementary Notes

DOI: https://doi.org/10.7922/G2CF9NDV

Dataset DOI: https://doi.org/10.5281/zenodo.5750140

16. Abstract

Bicycle transportation is often excluded from travel demand and route choice models. Even when bicycle modes are incorporated, models may use a simplified network that does not contain all streets and bicycle paths that a cyclist could feasibly take. These models may also only use trip distance and travel time when modelling a cycling trip; research on revealed route choice preferences of cyclists has shown that cyclist routing is influenced by other factors, such as the presence of a bicycle facility or road elevation gain. The City of Atlanta plans to triple its mileage of protected bicycle infrastructure in the next two years, and needs a tool to be able to plan and prioritize these projects based on the estimated effects on bicycle accessibility, bicycle mode share, energy usage, and emissions, to make the best use of the limited funding. The objective of this project is to develop this analytical tool and an associated network that includes all possible bicycle paths (i.e., roads, bicycle paths, cut-through paths, etc.) for a 12 square mile study area in the City of Atlanta that can be expanded later to the Atlanta Metro area. The tool, BikewaySim, is a shortest path calculator that uses Dijkstra's shortest path algorithm to find both the preferred route from any origin to any destination within the study area using lowest travel time and lowest total impedance cost. The BikewaySim network was created by conflating network data from the Atlanta Regional Commission (ARC), OpenStreetMap (OSM), and HERE into a whole road and pathway for BikewaySim and future use in the ARC's activity-based travel demand model. The methods for conflating networks and developing the shortest path model are publicly available resources. The final model is destined to include all viable pathways and incorporate cyclist preferences for use in planning and modelling bicycle travel for research, planning, and design. The framework allows other organizations and researchers to contribute to the project over time.

17. Key Words Bicycle route choice, network conflation, bicycle facility preference	**18. Distribution Statement** No restrictions.

19. Security Classif. (of this report) Unclassified	**20. Security Classif. (of this page)** Unclassified	**21. No. of Pages** 70	**22. Price** N/A

Form DOT F 1700.7 (8-72) Reproduction of completed page authorized

About the National Center for Sustainable Transportation

The National Center for Sustainable Transportation is a consortium of leading universities committed to advancing an environmentally sustainable transportation system through cutting-edge research, direct policy engagement, and education of our future leaders. Consortium members include: University of California, Davis; University of California, Riverside; University of Southern California; California State University, Long Beach; Georgia Institute of Technology; and University of Vermont. More information can be found at: ncst.ucdavis.edu.

Disclaimer

The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the information presented herein. This document is disseminated in the interest of information exchange. The report is funded, partially or entirely, by a grant from the U.S. Department of Transportation's University Transportation Centers Program. However, the U.S. Government assumes no liability for the contents or use thereof.

Acknowledgments

This study was funded, partially or entirely, by a grant from the National Center for Sustainable Transportation (NCST), supported by the U.S. Department of Transportation (USDOT) through the University Transportation Centers program. The authors would like to thank the NCST and the USDOT for their support of university-based research in transportation, and especially for the funding provided in support of this project. The authors would also like to thank staff from the City of Atlanta and Atlanta Regional Commission for providing network data and assistance. Additionally, the project team would like to thank Ziyi Dai for her work in the development and application of network conflation process used in this research.

NCST

BikewaySim Technology Transfer: City of Atlanta, Georgia

A National Center for Sustainable Transportation Research Report

December 2021

Reid Passmore, PhD Student, Civil and Environmental Engineering, Georgia Institute of Technology, Atlanta

Dr. Kari E Watkins, Associate Professor, Civil and Environmental Engineering, Georgia Institute of Technology, Atlanta

Dr. Randall Guensler, Professor, Civil and Environmental Engineering, Georgia Institute of Technology, Atlanta

[page intentionally left blank]

TABLE OF CONTENTS

List of Tables

List of Figures

BikewaySim Technology Transfer: City of Atlanta, Georgia

EXECUTIVE SUMMARY

This report presents a methodology that reconciles multiple GIS transportation network datasets into an efficient all-paths network for bicycle trip routing. This report also introduces BikewaySim, Georgia Tech's newest shortest-path calculator for cycling trips. The BikewaySim network is generated using automated Python scripts. These scripts reduce the time required to construct an all-paths network from multiple networks. These scripts also ensure that all available network data on link attributes (i.e., road classification, number of travel lanes, bicycle facility presence, etc.) are included in the new network.

All-paths networks are crucial to accurately modelling bicycle travel, as many cyclists have preferences for links that are not represented in a traditional travel demand model network like the one used in the Atlanta Regional Commission's (ARC's) activity-based travel demand model (ABM). Popular multi-use path facilities like the Atlanta BeltLine are not included in the ARC's ABM network, which is focused on major arterials, connectors, and freeways. Even those road links that do have bicycle facilities and are included in the ARC ABM network do not have attached attribute information indicating the presence of a bicycle facility or the design characteristics of these facilities.

The BikewaySim network presented in this report was developed for a 12 square mile study area that includes the neighborhoods surrounding the Atlanta Beltline, a multi-use rail to trail conversion. Three GIS transportation network datasets were available for use in this area: the ARC's Activity Based Model (ABM) network, HERE's HERE Map Data network (HERE), and the OpenStreetMap (OSM) network. The ABM network is ARC's travel demand model network. It includes major roads and contains detailed information about the roads. The HERE Map Data network is used primarily for navigation and includes all roads. The OSM data are a volunteer-built network that includes all roads and paths. We examined each network thoroughly and developed new data definitions or modified existing data definitions for link and node attributes. The three networks were filtered into four different subnetworks of composed of roadway links that are of similar typology. The base subnetwork is the unfiltered version of the data that contains all the original links with only minor data cleaning modifications. The road subnetwork contains all public roads that are legal for bikes to traverse: arterials, local roads, and collectors. Interstate links were not included in this subnetwork because bikes are not allowed to use them. The bicycle subnetwork contains all off-street bicycle paths and roads that are closed to automobile travel. The service subnetwork contains all access-type roads: parking lot roads, driveways, and alleyways.

Each subnetwork is composed of links to simplify the conflation of subnetworks across networks. After examining each subnetwork and a preliminary assessment of its coverage and detail, a combination of the ABM road, HERE road, and OSM bicycle subnetworks was deemed to be the best approach to creating an all-paths network for the case study area. A conflation and subsequent network graph generation process was applied to these three networks, which

ultimately resulted in the BikewaySim network. The advantage of the chosen approach is that every link and node in the final network can be linked back to its network of origin (and the attributes carried in each original network). It is important to note that no new network links were created during this process. In some cases, new nodes were inserted during network generation when a link needed to be broken into smaller links to allow reconciling data across the three networks. The final simplified ABM network contains 1661 links and 1519 nodes, but after reconciling networks and adding links carried in the other two networks, the final all-paths network carries 7138 links and 5666 nodes. This final network, however, carries far fewer links and nodes than were originally present in the other two models, making the final network much more efficient for use in shortest path processes. Without this conflation process, the result from the combination of these three networks would have been between 9,000-13,000 links and between 10,000-14,000 nodes depending on the tolerance accepted.

The advantages and disadvantages of the all-paths network were demonstrated by comparing shortest path routing between the simplified network and the all-streets-all-path network using BikewaySim. BikewaySim utilizes Dijkstra's algorithm to calculate the shortest path on a network from any origin to any destination (*1*). The shortest path between every possible traffic analysis zone (TAZ) pair combination in the study area was identified using both the ABM and BikewaySim network. More than 180 TAZs yielded 25,000 TAZ pair combinations (excluding intra-zonal trips). The research team compared the aggregate distance and run time difference associated with using these two networks. The runtime for the more complex BikewaySim network was more than two hours longer than the runtime for the ABM network. However, the amount of trip miles on the ABM links decreased by 30% when the whole road network was employed. That is, the addition of road links in the BikewaySim network that were not in the ABM network provided more efficient bicycle routes that would likely be utilized by cyclists. This is an important concept in bicycle transportation planning, because the assessment of construction or repair of bicycle infrastructure needs to include realistic paths to optimize resource allocation.

The BikewaySim shortest path algorithm was then modified to include additional impedance elements translated to time penalties, so that perceived costs would not be based solely on travel time, where the travel time is calculated using an assumed bicycle speed. This perceived travel time was calculated by incorporating distance modification factors that changed a link's distance (and therefore the calculated travel time) based on its attributes. These distance modification factors are meant to reflect the preferences of cyclists. For example, travel uphill is penalized by adding additional travel distance to the physical distance, which increases the calculated travel time on that link as a penalty. These distance modification impedance factors were calculated using interpretations of multi-nomial logit coefficients from previous studies (*2*, *3*). The model results indicate that trips created using the perceived travel time method were about 12% longer on average than the shortest path calculated using time alone. That is, the physical shortest path was being penalized with additional impedance associated with cyclist preferences to avoid uphill travel, avoid traffic, etc. Because a number of the perception penalties apply to higher functional class roadways, the additional penalties reduced the bicycle trip miles on the ABM links to 40%, pushing even more trips onto the local road network.

The code and documentation for this process will be made available on a Georgia Institute of Technology GitHub Enterprise Server. All the methods used in this report are designed to be directly transferable to other regions. In addition, the use of the GitHub open source approach will allow researchers, planners, and engineers to not only use these methods but also contribute to the project over time.

Introduction

Nearly half of all vehicle trips are under three miles in the U.S., a distance feasibly traveled by bicycle, yet the percentage of trips taken by bicycle remains low in many U.S. cities (*2*).However, some U.S. cities are achieving relatively high bicycle use; Minneapolis and Portland have achieved commuting mode shares of up to 4.1% and 6.4% respectively (*4*). What separates these cities from a typical US city like Atlanta, which has a relatively low ridership of 1.0%, is their approach to planning for bicycle travel, quantity and quality of existing bicycle infrastructure, and funding commitment to bicycle infrastructure. Portland and Minneapolis have constructed 2.5 and 4.8 miles of bicycle infrastructure per square mile respectively. Whereas Atlanta provides only 0.7 miles of bicycle infrastructure per square mile (*5*). Even if mode share is only increased by a small percentage, the population-level health benefits from increases in moderate to vigorous physical activity and reduction in carbon dioxide emissions could be significant (*6*).

While increasing bicycle mode share is a long- term objective of many cities, investing in bicycle infrastructure can be an uncertain process given the low quantity of current bicycle trips. For a city and its stakeholders to invest in creating better bicycle networks, they need to have a better idea of the potential impacts of proposed infrastructure (e.g., projected ridership increase, projected CO2 savings, increase in bicycle connectivity, or increase in available low stress routes). Additionally, recent research on the revealed preference of cyclists has shown that not all bicycle infrastructure is equal, and the placement of new infrastructure in relation to existing bicycle infrastructure matters (*7*).

Travel Costs for Automobile Trips Compared to Bicycle Trips

In route choice modeling, travelers are assigned a route from their current location to their chosen destination as a function of relative time, monetary, and other costs, which establish the relative utility of potential alternative pathways. In travel demand modeling, it is assumed that travelers seek to minimize their travel costs. This cost is usually assumed to be time cost, and travelers are assumed to choose routes that minimize their travel time to their chosen destination. For automobile travel, route choice depends greatly on the amount of congestion on the links that compose a route. Because of congestion, the shortest travel time route for motor vehicles is not always the shortest distance path, and the shortest travel time route is constantly changing as network conditions change. In travel demand modelling, this generally means that travel demand models employ an iterative process until they converge or meet some specific criteria for assigning shortest routes to all travelers. This process introduces increased computational burden and can result in travel demand models taking days to weeks to model travel in a major metropolitan region.

Yet, time is not the only cost that can be considered in route choice modeling. Tolls are routinely introduced as a monetary cost converted to time. There are also potential factors affecting route choice that may vary significantly across travelers, based on their sociodemographic background or experience level. For example, a brand-new driver might be more hesitant to drive on an Interstate, given their inexperience. For automobile travel, these secondary factors are typically not incorporated into the route choice model because it is assumed that these factors do not tend to dominate vehicle route choice. For bicycle travel however, such factors may be much more important in influencing route choice. Recent research into the revealed preferences of cyclists on route choice has demonstrated that costs other than time and distance are needed to properly predict cyclist route choice. Studies indicate that cyclists routinely deviate from the shortest distance path and will sometimes travel up to 15-30% further than the shortest distance path (*7*, *8*, *9*). These revealed preference studies utilize GPS trace data from apps, instrumented bikes, or bicycle share programs to compare a cyclist's chosen route to a set of alternative routes in a choice set. These types of studies also attempt to explain the deviations from the shortest distance path by developing a set of route attribute preferences that can be employed in route choice models. These studies have shown that cyclists prefer travelling on dedicated bicycle facilities, streets with low vehicle traffic volumes, and level (less-steep) terrain (*7*, *10*). In all these studies, cyclists were still most sensitive to distance, and preferences for bicycle facilities seem to be less pronounced in places with a high amount of existing and connected bicycle infrastructure like in the Netherlands (*8*). This finding suggests that as the quality and connectivity of cycling infrastructure increases, the number of alternative bicycle routes also increases. This makes it less likely that cyclists will have to deviate from the shortest distance path to find routes with bicycle facilities or preferred attributes.

Preferences can vary based on a cyclist's sociodemographic background and level of cycling experience. Some studies have segmented monitored bicycle travel GPS traces by user type, age group, gender, race, and income level (*9*, *11*). While preference differences are noted across demographic characteristics, these studies have generally found that there is always a preference for routes with protected bicycle infrastructure.

The Need for an All-Streets Network

To assess cyclist revealed preferences, it is necessary to compare their chosen paths to a set of alternatives paths. Analyses must also account for detailed information about what distinguishes each alternative path. Each path is composed of links, and each link needs to carry information on bicycle facility presence, number of lanes, median presence, adjacent land use, or on-street parking presence.

According to NCHRP 08-36 Task 141, about 60% of regional MPOs in some way model bicycle demand, but less than 30% use an all-streets network. Even fewer of these MPOs are using considering cyclist preferences for route choice in their models (*12*). NCHRP 08-36 Task 141 also points out the need for methods and software used to model bicycle trips to be "accessible and transferable" to different regions. The research presented in this works towards this goal.

Automobile trips can span an entire metropolitan region. All-streets networks are rarely used in travel demand models because they can dramatically increase model computational burden by expanding the number of potential vehicle routes between origin-destination pairs and the need to account for changes in congestion and travel time as trips are allocated to routes during model iterations that define congested travel times. For bikes, however, trips are not likely to span more than four miles (limiting the number of potential routes) and the effect of congestion on route choice can be relaxed because vehicle congestion is generally seen as having a minor impact on bicycle travel time. This may not be strictly true on narrow facilities where bicycles and motor vehicles are weaving and interact, but bicycles are generally assumed to be able to move around the vehicle congestion at near-normal speed. Hence, an all-streets network for bicycle route assignment can be accommodated in modeling without unreasonably increasing computational burden (*12*). With the increasing use of parallel computing in transportation engineering, computation speed continues to improve over time.

Project Objective

The Atlanta Regional Commission (ARC), the MPO for the Atlanta Metropolitan Area, is one such MPO that models bicycle trips but does not incorporate an all-paths network or consider preferences for cycling infrastructure when modelling cycling trips. The objective of this report was to leverage existing GIS transportation network datasets to create an all-paths network for a 12-mile study area in the City of Atlanta. This network will be used to perform shortest path routing using both traditional time based costs and perceived time based costs that serve as an indicator of a cyclist's preferences for certain road attributes. This shortest path calculator is called BikewaySim.

The results of this study will show the potential advantages and disadvantages of this specific approach to network construction and all-paths networks in general. The overall objective of this project is to develop a test network in the City of Atlanta with a proof-of-concept bicycle shortest path calculator and open source Python-based network reconciliation and generation methods that are readily transferable to any region.

This project uses a simplified approach to network conflation that reconciles multiple transportation networks using Python GIS packages such as GeoPandas. The automation of several specific steps helps to eliminate a lot of the time required to manually conflate links across networks. The goal is to develop procedures that are repeatable (obtaining the same results with each run) and transferrable to other areas using their available datasets. All the code and documentation are posted on a Georgia Tech GitHub Enterprise server.

Ultimately, the goal of BikewaySim is to provide a coded network and link impedance costs for use in shortest path routing using Dijkstra's algorithm (*1*). The results from BikewaySim can be used in assessing how improvements to the bicycle network can provide pathways that reduce cyclist travel time or otherwise provide preferable routes, and BikewaySim can also be implemented as a simplified bicycle route choice model within a larger travel demand model.

Case Study Area within the City of Atlanta for BikewaySim

The 12 square mile case study area used in this research includes Midtown, Piedmont Park, Morningside, Virginia-Highlands, Old Fourth Ward, Ansley Mall, Atlantic Station, and the Georgia Institute of Technology campus. This area was selected because it contains a wide variety of bicycle facilities and street types, multiple networks were available that will require reconciliation, the area was of manageable size, and the resulting network would support a good demonstration of the conflation and routing methods. The area was also a familiar area to the researchers, which made it easier to verify links and network data validity.

The left-side map in Figure 1 shows the study (highlighted by green boundary) in relation to its location within the City of Atlanta boundary (cross-hatched). The right-side map shows a zoomed in view of the BikewaySim Study with OSM used as the base map for reference. Interstates are shown in red for reference.

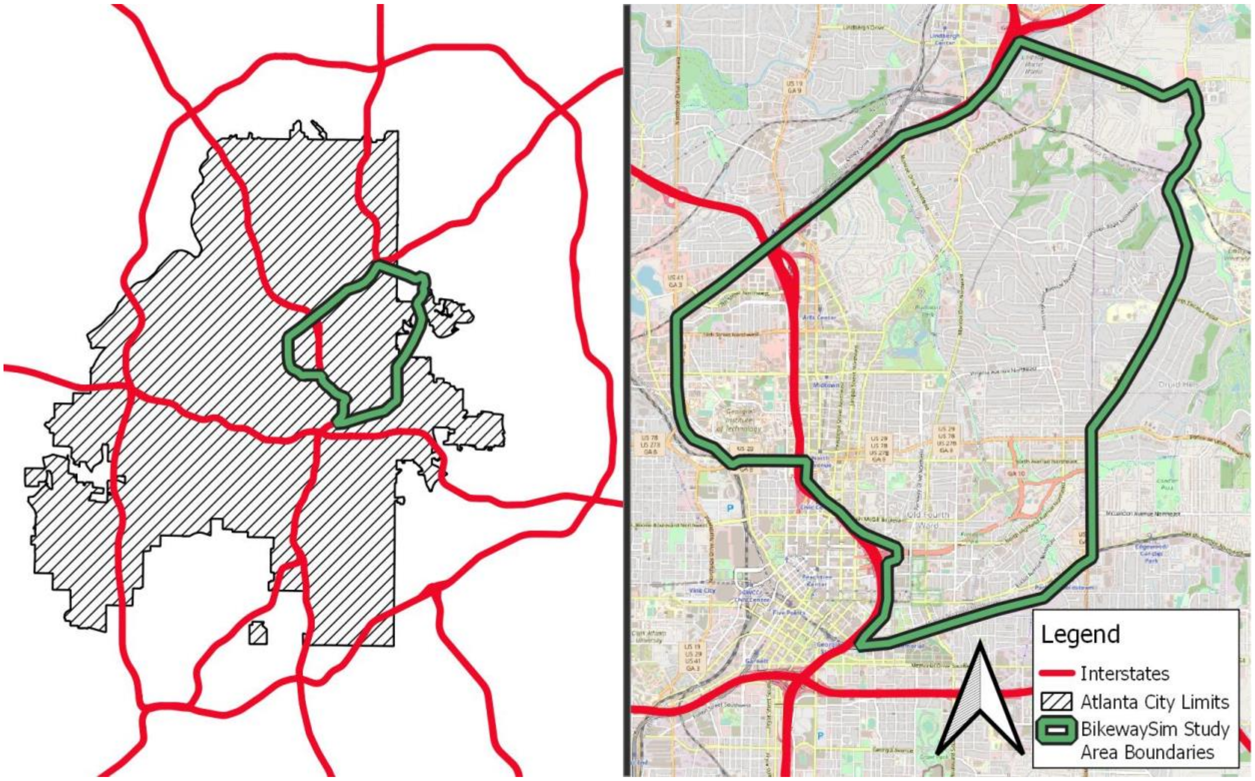


Figure 1. BikewaySim study area

Expanding the BikewaySim Network Graph

In this case study area, the goal is to take a base network that is generally considered to be spatially accurate, with good attribute detail, but incomplete with respect to potential bicycle pathways (i.e., the activity-based travel demand model network) and add roads and traversable bicycle paths from other network sources. This project starts with the ARC's activity-based model (ABM) road network as the base network graph. This is also the same network used for Georgia Tech's RoadwaySim shortest path calculator. Although the ABM network was originally intended for model representation rather than accurate spatial representation, the ARC has made changes in recent years to improve the accuracy of node locations at intersections and to add shape points to match up to existing roads. The ABM network also contains detailed information about links. Because the ABM network is used in RoadwaySim and TransitSim, the network nodes serve as reference points that support transfers within multi-modal shortest path processes (*13*). This will be important in later work when the research team will move trips from bicycle mode to transit modes at these transfer nodes. Hence, another goal of this project was to retain all the links used in the ABM network, adding links that can be traversed by bicycles that are not present in the current network, conflating links from other networks to reconcile the addition of new nodes and links (and carry the relevant attribute data). At the same time, researchers do not want to add unnecessary nodes, subdividing links into pieces when these new links do not provide access to alternative routes. Hence, part of the modeling network development process is to ensure that the final network only includes the links that are really needed (i.e., establish an efficient network).

It is of note that the ARC ABM network is primarily intended for model use, while the other networks such as HERE and OSM are really designed to support navigation and mapping displays. This is a key distinction, because it means that OSM and HERE can over-specify the network in certain areas where it may not make sense to retain links. This means that the conflation process can run into problems in certain areas where the ABM network employs a simplified road geometry that does not sit directly on top of the relevant physical road.

This report presents the development and application of procedures designed to reduce the number of hours needed to develop an all-streets network. No manual editing of geometry information was performed for this report. Even with the implementation of a solid automated process, a manual QA/QC check will be needed to resolve any node placement and other network errors. This report also identifies and discusses problem areas where such errors generally occur.

Network Acquisition

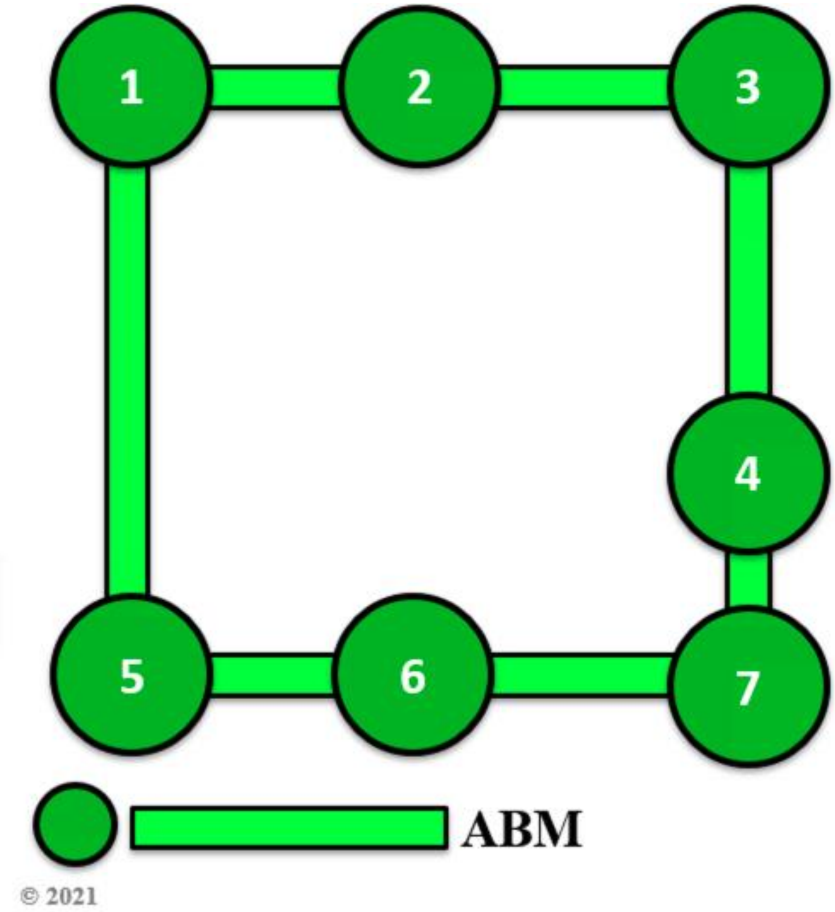
The networks used in this project were ABM, HERE, and OSM. The ARC's ABM network and data are managed by the ARC, the MPO for the Atlanta Metropolitan area. HERE is created and maintained by HERE. The Georgia Tech team licensed the HERE data for use in this project (the licensed data cannot be freely shared under the license agreement). OpenStreetMap is an open data source that is free to share and includes crowdsourcing mechanisms that allow users to edit the data.

OSM data can be obtained using a variety of methods, interfaces, and APIs. One widely used source is from the company, Geofabrik. Geofabrik cleans crowdsourced OSM data and provides both free and paid downloads, but they may remove some link and node attributes that we might want for the all-paths network in the process (*14*). So OSM data were instead retrieved by using Overpass API (*15*). This API can be used to query any OSM features and download them as a GeoJSON file. The advantage to using this method to retrieve OSM data is that it retrieves all keys and tags associated with OSM features. With these networks acquired, the next step was to clean and convert each network into a common network structure.

Network Structure

In this report, mathematical network graph “edges” and “nodes” will be referred to as “links” and “nodes” respectively. In OSM, links are referred to as “ways,” but this term will not be used in this report. Links in this network will represent paths (e.g. streets, separated bike paths, parking lots) that are traversable by bike. Nodes in this network will represent intersections of paths or changes in a path’s characteristics. Intersections of paths can represent an intersection of two roads or a driveway/parking lot access point. Nodes that represent a change in a path’s characteristics can occur when the number of lanes or speed limit changes on a road.

In the ABM network, all links are defined and identified with their starting node ID in the “A” column and the ending node ID in the “B” column. These start and end nodes will be referred to as reference nodes. Additionally, all links in the ABM are directional. If a specific link represents a two-way road, then the roadway will be represented by two links stacked on top of each other; the “A” node and “B” node for the northbound link will be the “B” node and the “A” node in the southbound link, as shown in Figure 2 below. ABM nodes are all numbered and each node is assigned a set of attributes for spatial location, whether the node represents an intersection or a TAZ centroid, the control type assigned to the intersection, etc.



A	B
1	2
1	5
2	1
2	3
3	2
3	4
4	7
5	1
5	6
6	5
6	7
7	6

Figure 2. Example of ABM network representation.

HERE and OSM are not structured in the same way as the ABM network, but both networks can be manipulated into this format. HERE employs reference nodes but does not use duplicate links to represent two-way streets. Additionally, HERE carries no node-specific attributes.

The OSM links retrieved using Overpass API did not include reference nodes in its data fields, but OSM nodes can be retrieved separately by running a query through the Overpass API. Once the nodes are retrieved, the nodes can be assigned to OSM links by performing a spatial join on the OSM link's start and end node geometries (establishing the reference nodes for each link).

As links and nodes are added to the BikewaySim network, it is important to keep track of the source of these links and nodes. All reference nodes and node IDs will be retained throughout BikewaySim's development so any link or node can be traced back to its origin network. This will be important if future edits to an origin network need to be pushed into the BikewaySim network (or vice versa).

All of the networks represent nodes numerically, meaning that there is a chance that two nodes from different networks could share the same ID. Hence, modified IDs are needed. The convention adopted in this research was to add three numbers at the beginning of each node ID that would represent the data source. The first number represents the network of origin. For ABM this would be 1, for HERE 2, and for OSM 3. The second number indicates how the node ID was created. Throughout the BikewaySim network generation process, there will be times when links are broken apart to create additional links and nodes. In this case, the current method for generating node IDs is to sequentially number all of the nodes added. If the node ID was from the original network then the third number is set to 0. However, if the node ID was generated after a certain process it is numbered with a higher value. These node ID conventions are summarized in Table 1 below. Note that these conventions are subject to change as the BikewaySim network develops in the future.

Table 1. Node ID Codes

1st Number: Network Name	2nd Number: Node Origin
1 = ABM	0 = Existing
2 = HERE	1 = Created from breaking links apart
3 = OSM	2 = Created from splitting links with another network

A node with an ID of 20458 would represent a node from HERE with an original HERE ID of 458. The second number, zero, indicates that the number 458 came from the original network and was not new node.

Network Attribute Information

Each network dataset being used in this project was reviewed for link and node attribute information. Table 2 below shows a generalized checklist of the type of information that each network carries and an initial evaluation of how complete the data were for each attribute. In Table 2, each attribute, shown in the left most column, was given a dash, empty circle, half circle, and full circle that indicated that the data for that attribute was empty, had less than 1% of data available, had 1-50% of data available, or had greater than 50% available respectively. Completeness for an attribute is the percentage of links with non-empty values.

Table 2. Attribute data available for links

Attribute	ABM	HERE	OSM
Bicycle Facilities	-	-	-
Bicycle Facility Width	-	-	-
Bicycle Facility Blockage	-	-	-
Bicycle Traffic Volume	-	-	-
Sidewalks	-	-	-
Sidewalk Width	-	-	-
Number of Lanes	+	+	-
Travel Lane Width	-	-	-
Shoulder Width	-	-	-
Automobile Traffic Volume	+	-	-
Directions Allowed	+	+	-
Road Classification	+	+	+
Posted Speed Limit	+	+	-
Observed Speed	+	-	-
Median Presence	+	+	-
On Street Parking	-	-	,
Road Grade/Slope	-	-	-
Bridge/Tunnel	-	+	,
Traffic Collisions	-	-	-

Legend: ‘-‘: Attribute not available. ‘,’: Less than 1% data available, ‘-’: 1%-50% data available, ‘+’: > 50% of data available.

As shown in Table 2, these three networks cover a lot of the desired attributes. It also makes a clear case for the inclusion of all three networks in the conflation process. The ABM network contains attributes for metrics like automobile traffic volume and observed speed which are not present in HERE or OSM. HERE and OSM data may not contain attributes for these metrics, but they are needed because they have links that are not represented in ABM. Because there is a lot of overlap between attributes in HERE and OSM, HERE data can cross examine matching OSM attributes. This will help increase the accuracy of the conflated network. However, there are still missing attributes for traffic collisions, width of features, and volume of bicycle/pedestrian travel. These attributes may require the acquisition of additional network data, satellite/road imagery data, or field measurements.

The completeness of the attributes for OSM is generally low. Unlike ABM and HERE, which are prepared by organizations for modelling and navigation respectively, OSM depends on a volunteer userbase to enter attribute data. Because of this, both the number and completion of OSM attributes will vary across different regions. OSM’s low completion rate does not dismiss the quality of the included information. When we compared ARC’s bicycle facility inventory (note that this bicycle facility inventory is separate from the ABM network) to the links with bicycle facilities indicated, they were similar in length; in fact, some of the bicycle facilities included in OSM were new facilities that were not in the ARC bicycle facility inventory. So, in some cases, OSM’s low completion rates could just be a result of it not being standard practice among volunteers to include the absence of an attribute. Additionally, completed data in ABM and HERE could represent imputed data rather than accurate data.

As for node attribute data, ABM’s nodes carried attribute data on signalized intersections. OSM’s nodes also carried data on signalized intersections, but there were only a limited number of signalized intersections that had been coded in. HERE nodes do not carry any attribute data. Thus, we did not include an attribute table for nodes in this report.

Bicycle Lanes, Protected Bicycle Lanes, and Mixed Use Paths

One issue to address in developing a complete network of bicycle-accessible facilities within any region is the general lack of availability of bicycle facility data. Neither the ABM nor HERE data contain information about bicycle facilities, as was shown in Table 2; however, the OSM network does. In this section, we compare OSM’s available bicycle facility data to that of the ARC’s bicycle facility inventory to verify OSM’s bicycle facilities.

Note that the ARC bicycle facility inventory is held in a separate data set and is not coded directly into the ABM network. Additionally, the ARC bicycle facility inventory only includes data for dedicated bicycle infrastructure. It does not contain an inventory of shared use roadways with painted “sharrows”. Sharrows are painted road markings that indicate that vehicles and bicycles are sharing a roadway lane.

OSM data can be filtered to obtain bicycle facilities. Figure 3 provides a side-by-side comparison of the ARC’s bicycle facility inventory (shown on the left and color coded by type of facility) with the bicycle facilities generated from the filtered OSM network (shown on the right and colored yellow).

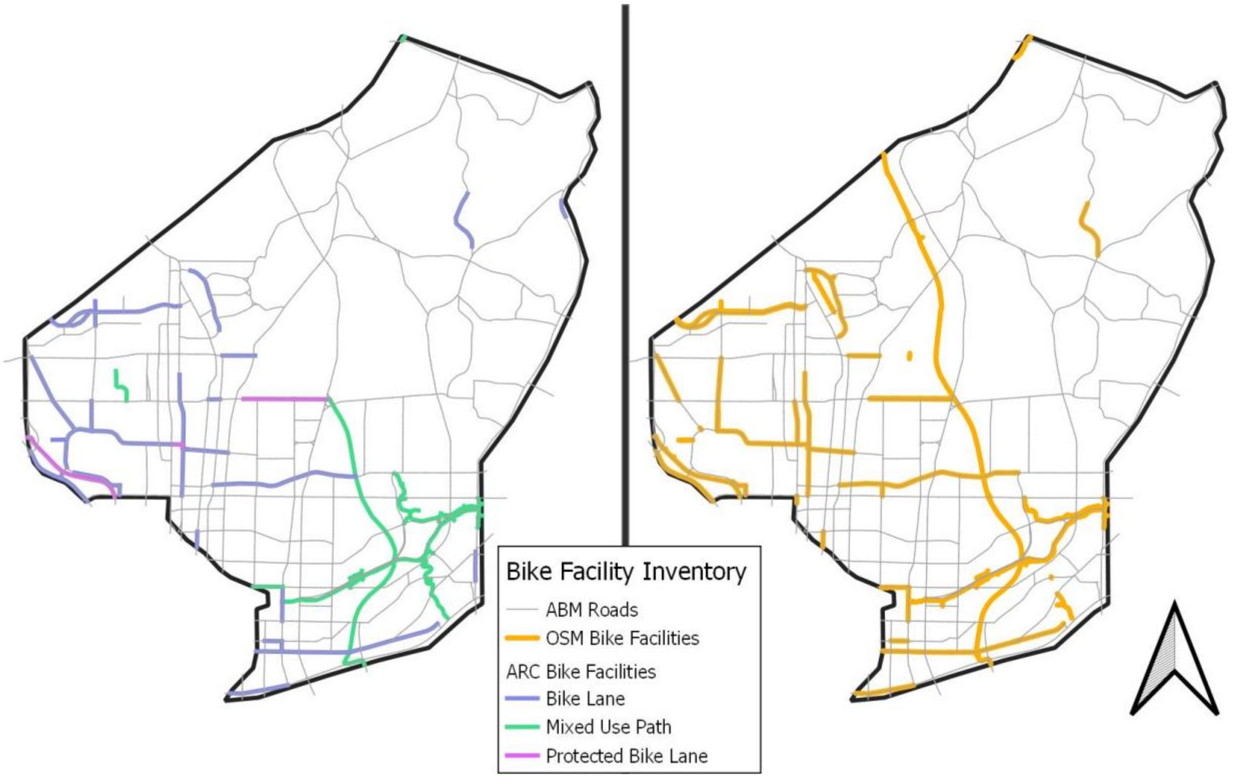


Figure 3. Bicycle facility inventory for ARC and OSM overlaid on ABM network.

Even though the OSM network contains an inventory of bicycle facilities, the tagging (set of feature attributes) of the facilities makes it difficult to interpret what type of facility is present on the road (e.g., a grade-separated facility vs. dedicated bicycle lane vs. shared lanes). The ARC bicycle facility inventory uses a NACTO-inspired bicycle facility classification scheme where bicycle facilities are either type of painted bicycle lane, mixed-use path, or protected bicycle lane (also called a cycletrack). Protected bicycle lanes are different from painted bicycle lanes in that they have some type of barrier separating cyclists from automobiles. However, unlike mixed use paths, they are still on the road. Carrying more detailed data associated with facility design characteristics will facilitate the incorporation of cyclist preferences for use in shortest path routing in the future.

In one previous study conducted across six major cities about the coverage and quality of OSM data compared to city maintained data on bicycle facilities, researchers found an average difference of about 30% in facility inventory miles (*16*). This study also noted that OSM tags typically don't distinguish between more nuanced facility types like protected bicycle lanes. In the BikewaySim study area, the bicycle facility network contained about 20 miles of bicycle facilities and the OSM network contained 23 miles of bicycle facilities, thus the OSM network contained about 3 more miles of bicycle facilities than the ARC bicycle facility inventory.

Most of this difference is a result of the ARC data having more stringent definitions for bicycle facilities. For instance, the ARC data omit a two-mile portion of the Atlanta BeltLine trail because it is unpaved. Some errors were spotted in the OSM data as well, such as a bicycle facility shown as a block longer than the actual length shown in satellite imagery. On a positive note, the crowdsourced OSM data set does contain a recently installed protected bicycle lane on Spring Street in Midtown Atlanta. In contrast, the ARC bicycle facility inventory does not include this facility because it is updated less frequently. Ultimately, this section has shown that OSM does have good coverage of bicycle facilities within the study area.

Filtering Methodology

Because each network contains varying levels of detail, it made sense to group links with similar topologies so that the three analysis networks could be more easily compared. The base network includes all the original network links and nodes. The base network is comprised of all the original links contained in each raw network from the provider, with only minor modifications made in a data cleaning process. These base networks were then filtered into the following subnetworks:

- The **road** subnetwork encompasses all publicly accessible roads that can be traversed by both vehicles and bicycles. Roads where bicycles are not permitted, such as Interstates, like I-85, and their access ramps were not included in the road subnetwork. This subnetwork will also carry attribute information on on-street bicycle facilities such as bicycle lanes and protected bicycle lanes.
- The **bicycle path** subnetwork contains all bicycle traversable paths that are separate from roads. This includes roads that are closed to motor vehicles, dedicated bicycle paths, and multi-use trails. The bicycle subnetwork does not include sidewalks but does include bicycle-pedestrian shared use paths. Bicycles are generally not allowed to use sidewalks in the state of Georgia (unless there are children under the age of 12), but there are people that still use them. There is an adjacent research team working on developing a complete sidewalk network for the study area, and once this network is complete it will be included in the BikewaySim network.
- The **service road** subnetwork contains all other links that are not public roads or bicycle specific paths. This includes service roads, alleys, parking lot roads, and driveways. The key distinguishing mark between these service links and bicycle links is that vehicle travel is still allowed on these non-roads.

The network data structure and documentation for each of the three networks defined the formats and differences in data structures that would be needed to develop methods to split each network into these subnetworks. The filtering process is depicted in Figure 4 below. The process starts with ensuring that all networks are in projected to UTM NAD 83 Georgia West. After this, each network undergoes a specific filtering method that was defined after examining a networks documentation. For instance, the ABM network has directional links that represent each direction for a road. The ABM specific filtering method removes these double links. The OSM specific filtering method adds in OSM reference nodes and any polygon features to lines. HERE did not require any prefiltering. Once the prefilter is done, additional network specific filtering methods are applied to each network to sort links into road, service, and bicycle links.

Nodes are then created from these links. The filtered links and nodes are then ready to go into the network conflation process.

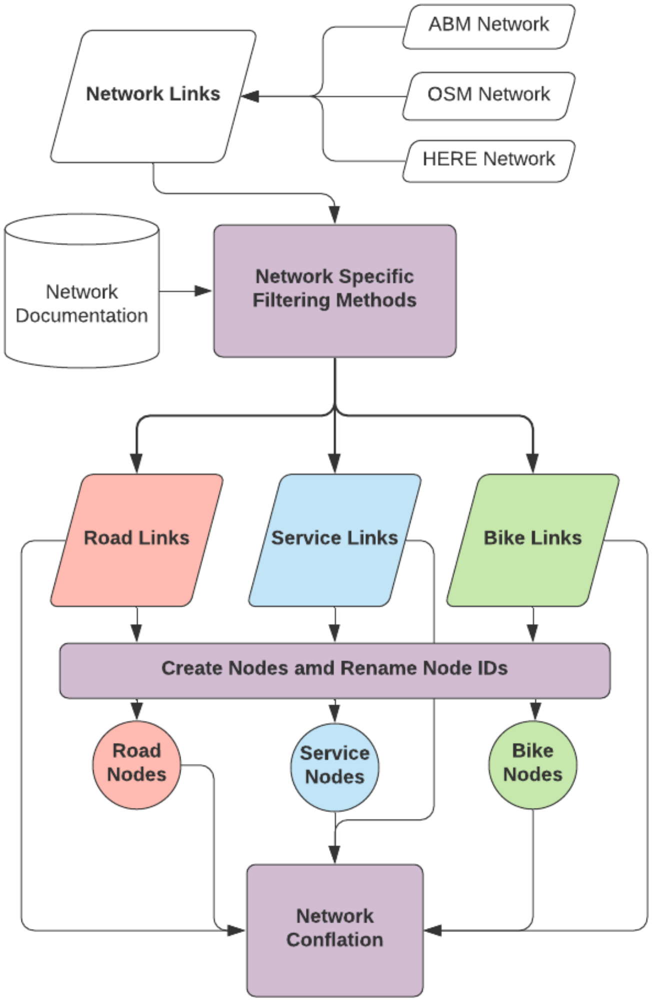


Figure 4. Network filtering process flow.

The bicycle subnetwork contains all bicycle traversable paths that are separate from roads. This includes roads that are closed to motor vehicles, dedicated bicycle paths, and multi-use trails such as the Atlanta Beltline. This subnetwork does not include on-street bicycle facilities, such as bicycle lanes or protected bicycle lanes. On-road facilities such as bicycle lanes, cycletracks, and shared lane markings will be carried in the road subnetwork instead. The bicycle subnetwork does not include sidewalks. Neither the ABM nor the HERE networks include sidewalk links. OSM does include sidewalk links, but the sidewalk data can be inconsistent by area because their inclusion is dependent on whether a volunteer OSM user drew them in. In the Atlanta metro area, a sidewalk network independent of the roadway network is currently being developed. Bicycles are generally not allowed to use sidewalks, so including these pathways in a shortest path analysis is not currently recommended.

The service roads subnetwork contains all other links that are not public roads or bicycle specific paths. This includes service roads and alleys, parking lot roads, and driveways. The key distinguishing mark between these service links and bicycle links is that vehicle travel is still allowed on these non-roads. This subnetwork does not contain sidewalks or streets that are closed to automobiles. These service road and access links can be useful as they can provide shortcuts or informal connections between roads or other bicycle links. However, it is also true that many of these links dead-end or are not practical for bicycle use (e.g., closed by gates or composed of an unpaved rocky surface). As BikewaySim further develops, discussions will be needed on adding cut-through and parking lot links, given that CycleAtlanta and other monitored bicycle activity data sets clearly show the use of these alternative pathways (*9*). Bicycle facility inventories on their own are difficult enough to maintain as new facilities are added every year but keeping track of which cut thru paths are legitimate, accessible, open, legal, or safe may require a large resource commitment. Ongoing bicycle monitoring data collection efforts may be needed to identify popular paths that are missing from the latest network.

The road network contains all of the roads that are accessible for bicycle travel. The bicycle network and service subnetworks contain links that need to be added to the road network to generate a complete streets network for bicycle travel and to reflect actual route paths that are chosen by cyclists in the study area.

Filtering Results and Discussion

This section presents the filtering logic used to filter HERE, ABM, and OSM into each subnetwork and the results of this filtering processes. The initial cleaning processes performed on each network to generate their respective base network and the processes to filter each base network into the road, bicycle, and service road subnetwork is described in detail for each network, and the results of these processes are also presented.

The networks were filtered into subnetworks using Python. For a small area like this study area, it takes less than 6 minutes to run this process with most of that time going into filtering the OSM data, due to the large amount of accompanying data fields employed by OSM. Processing the OSM data took around 60 minutes for the entire metro Atlanta area on a desktop with 32 Gbs of RAM and a 10 core 3.70 GHz processor.

As each network is filtered, a before and after filtering map will be used to visually demonstrate how the base network is filtered into subnetworks. The area that will be used in this section to demonstrate the effects of the various filtering process steps and resulting networks covers the area near Ponce City Market in Old Fourth Ward. This area contains several large shopping centers with large parking lots, the eastside BeltLine trail, parts of the Freedom Park Pathway, and many neighborhood streets. Satellite imagery was provided by Bing maps and the base map used was provided by OpenStreetMap. All the figures that follow in this section will use this area.

ABM

Before filtering the ABM network, the double links representing two-way streets were removed to provide a more accurate network mileage estimate and to reduce amount of memory needed to process the links. This new version of the ABM network served as the base network for ABM.

The ABM network is simple to filter as only centroid connectors, Interstates, and Interstate ramps need to be removed. Centroid connectors are unique to travel demand model networks and are shown in Figure 5. They represent the theoretical connection within the model from the transportation analysis zone centroid to various locations on the network where traffic is modeled as entering or leaving the zone. Centroid connectors also carry information about the generalized cost of travelling into and out of a TAZ (these links are also directional given that the time it takes to enter a zone may be different than the time it takes to leave a zone due to local roadway configurations or use restrictions). Because centroid connectors are not routes, they are removed, simplifying the ABM network structure.

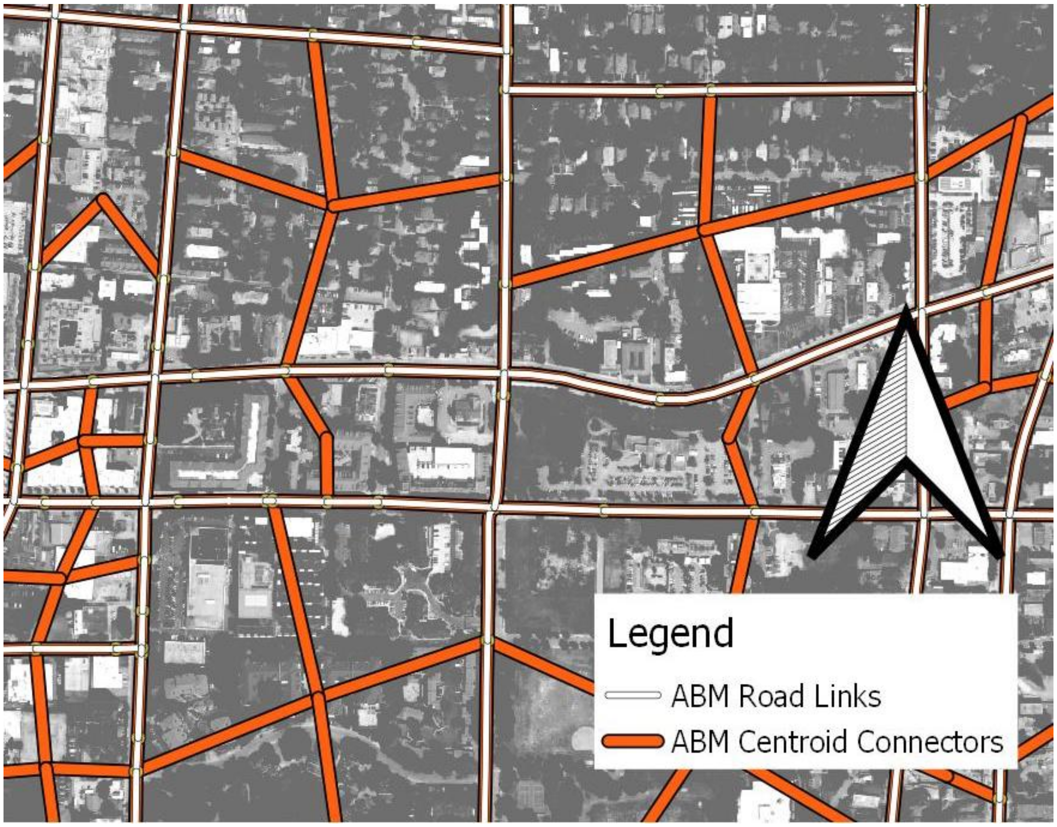


Figure 5. TAZ centroid connectors in the ABM network.

The ABM network was filtered into a road subnetwork by only including links with the values of ‘Principal Arterial’, ‘Minor Arterial’, or ‘Collector’ for its ‘FACTYPE’ attribute. This removed centroid connectors, Interstates, and Interstate ramps from contention. This “FACTYPE” attribute represents a links road functional classification.

The base ABM links are shown in the left-hand map of Figure 6 in purple and the filtered ABM road links are shown in the right-hand map of Figure 6 in blue. After filtering, all centroid connectors have been removed and a basic road network remains. There are no service or bicycle links in ABM, so no further filtering methods were needed.

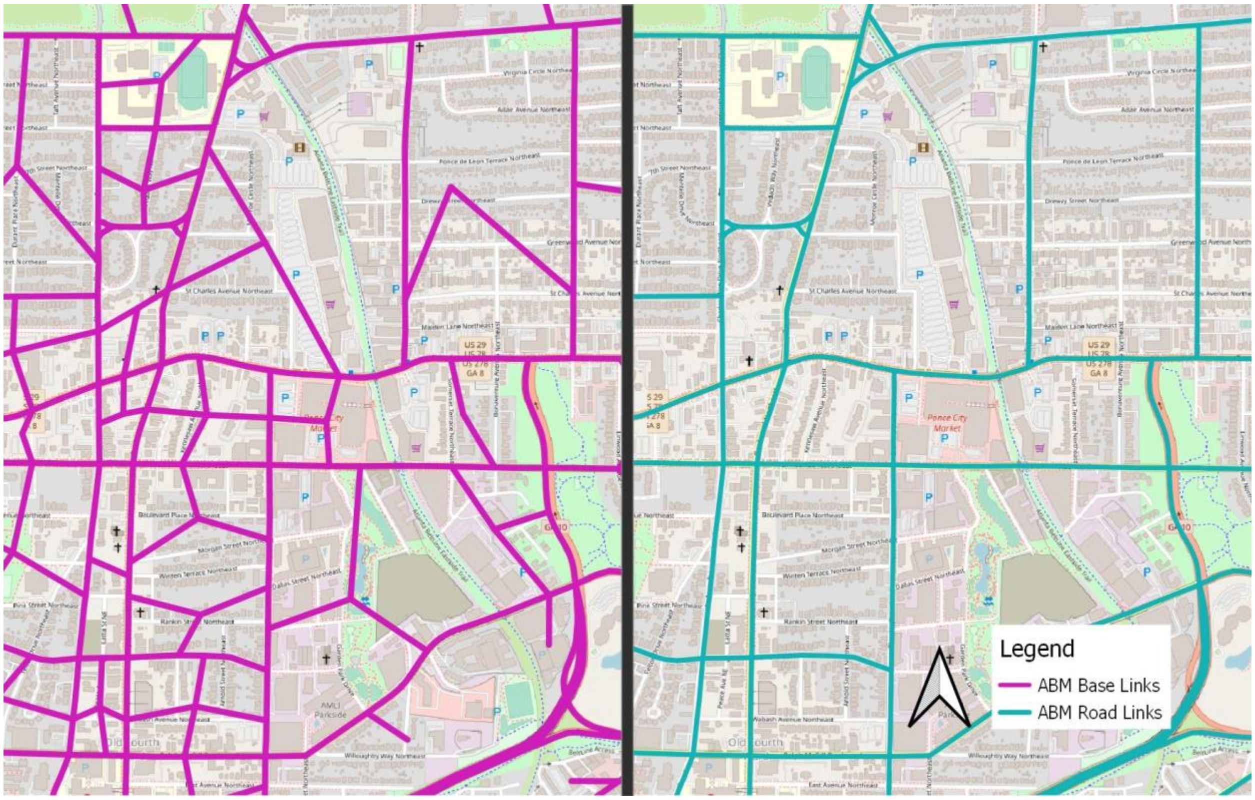


Figure 6. ABM filtering process.

HERE

The HERE network is a detailed all-streets dataset containing many more links than are included in the ABM. HERE has data for these links on road type, speed limits, controlled access roads, ramps, automobile access, etc. HERE includes roads, bicycle facilities, and service road links, so a filter method is needed for each.

The HERE network was filtered to a road subnetwork by first removing links with a value of “yes” for its “CONTRACC” or “RAMP” attribute. “CONTRACC is short for “controlled access” and is used to identify Interstates. “RAMP” is used to identify any roads that access the Interstates. Lastly, any links with a speed limit category of “Less than 6 MPH” were removed.

The base HERE network was filtered to a bicycle subnetwork by only selecting links that had a value of ‘No’ for its ‘AR_AUTO’ attribute. This attribute is short for “Automobile Access” and is used to identify if automobile traffic is allowed on a link.

The base HERE network was filtered to a service subnetwork by only selecting links with a value of “Yes” for its “AR_AUTO” attribute and “Less than 6 MPH” for its “SPEED_CAT” attribute. This means that the only HERE links included in the service subnetwork were links with a speed limit of less than 6 MPH that still allowed vehicle travel. The “less than 6 MPH” streets are usually parking lot roads or driveways.

The results of the HERE filtering process are shown in Figure 7 below. The base HERE links are shown in the left-hand map of Figure 6 in purple. The filtered HERE road, bike, and service links are shown in the right-hand map of Figure 7 in blue, green, and red respectively.

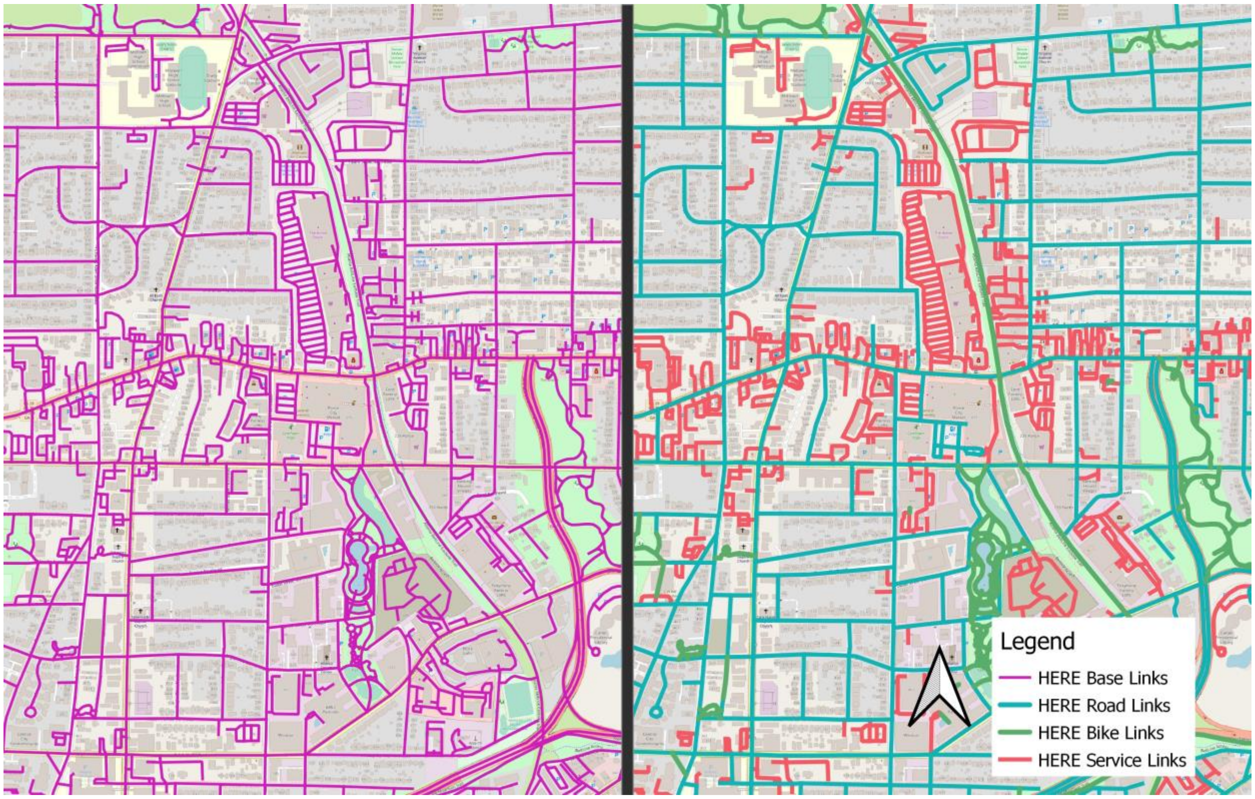


Figure 7. HERE filtering results.

OSM

The OSM network is also a detailed all-streets dataset, but because the OSM network is created through crowdsourcing by many users from different regions and backgrounds, it is not as consistent as HERE. It should be noted that with the right queries, OSM can be filtered without having to download the full dataset from the Overpass API. However, the syntax for downloading OSM data is complex, and it made more sense to just download all the data to examine first.

The OSM network needed some cleaning before usage. If any links in the OSM network form a closed loop, they form a polygon. Figure 8 below shows some examples of polygon formation (shaded in white). Most of the time, the polygon was associated with sidewalks, circular park paths, or crosswalks. This issue was resolved by treating the borders of these polygons as linestrings.

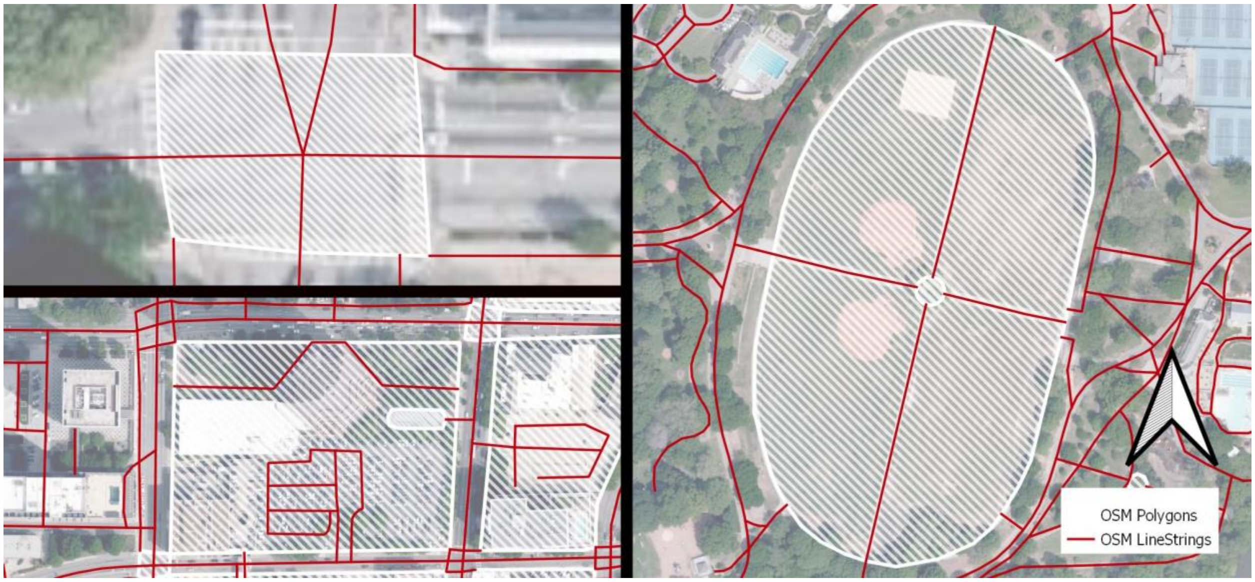


Figure 8. OSM polygons.

Many links in their default state in OSM are not broken into logical segments as one would see with a typical network graph as seen in Figure 9. In this figure, different colored links represent different links. On the left side of this figure, there are streets (one in red and one in purple) that intersect at the intersection highlighted by the red circle, but these links are not split at this intersection. Similarly, in the right side of the figure, the yellow path should be split into multiple links when it intersects with other paths (circled in red).

Figure 9. OSM network not split at intersections.

Links must be broken wherever they intersect with other links for use in network navigation apps and for shortest path calculation with Dijkstra's algorithm (*1*). Both programs need to be able to recognize that turns are allowed between these intersecting links. When links are not broken up at intersections then they are treated like they are a bridge or tunnel. A simple method can be used to break up lines at link intersections that works in both QGIS and ArcGIS. All the links are dissolved and converted from multipart to singlepart features. This splits links based on line intersections. Note, however, this method will create intersections where intersections should not exist, like between a bridge and a road that passes underneath. This occurrence was minimized by processing each filtered subnetwork separately, instead of all the data at once, but a manual QA/QC check will have to be done to ensure these splits occurred correctly. This final QA/QC process can also be supported by reconciling the resulting network with other data sets (such as a bridge asset management inventory).

Because OSM is maintained by volunteers, there are some inconsistencies in feature inclusion. Figure 10 shows examples of OSM's sidewalks, driveways, and paths features. In the top left map, the sidewalks depicted in purple are drawn in for two streets, but they are not drawn on the other street that clearly has sidewalk on it (circled in red). In the bottom left map, each driveway shown in red is meticulously drawn for this neighborhood, but most other neighborhoods in OSM do not have driveways. The detail of OSM's data varies from area to area depending on the efforts of volunteers. In the map on the right in Figure 9, Georgia Tech's campus paths are all drawn in full detail, but this full level of detail isn't entirely necessary for creating a shortest path model network (in fact, in most cases, driveways are not needed in the shortest path network).

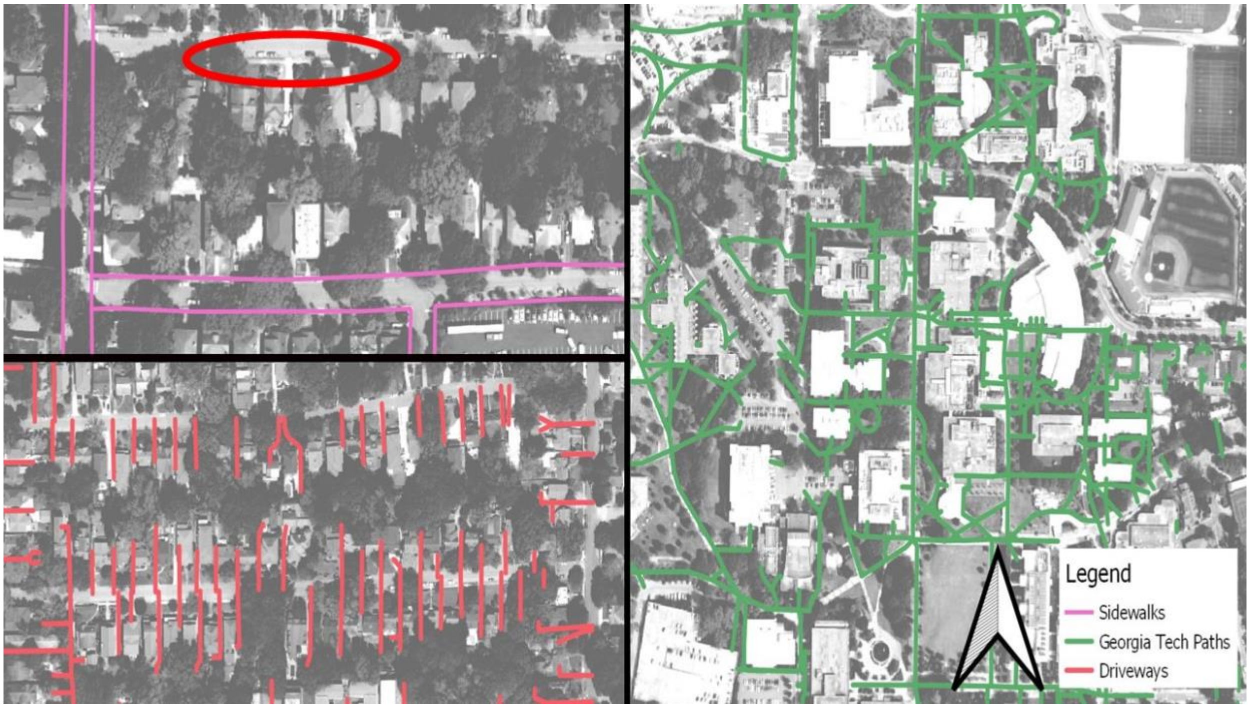


Figure 10. OSM sidewalks, driveways, and paths.

OSM was split into its road subnetwork by using its ‘highway’ attribute, which is similar to what functional classification is, to select road types that should kept. The values of interest from this attribute were ‘primary’, ‘primary_link’, ‘residential’, ‘secondary’, ‘secondary_link’, ‘tertiary’, ‘tertiary_link’, ‘trunk’, and ‘trunk_link’. Links containing these values for their ‘highway’ attribute were used for the OSM road subnetwork. This removed any Interstate or Interstate ramp links from consideration.

The OSM base network was filtered into a bicycle network by using its ‘highway’ and ‘footway’ attributes. The values of interest from the ‘highway’ attribute were ‘cycleway’, ‘footway’, ‘path’, ‘pedestrian’, and ‘steps’. The base network was filtered down to links containing one of these values for ‘highway’. After this, the sidewalks and crosswalks were removed by finding links with the attribute values ‘sidewalk’ or ‘crossing’ for its ‘footway’ attribute. These links served as the bicycle subnetwork.

The OSM base network was filtered into a service network by only including links with the value ‘service’ for its ‘highway’ attribute. The before and after results of this filtering process for OSM are shown in Figure 11 below. The base OSM links are show in purple on the left map. The OSM road, bicycle, and service road subnetworks are shown in blue, green, and red respectively. This figure shows that while sidewalks were removed (note the drop in thickness on some of the road links), all of the driveways in the service road network are still in the network. This is because these driveways did not have a feature distinguishing them from other service links like alleys and parking lot roads.

Figure 11. OSM before and after filtering.

Network Comparisons

In this section, the results of using the filtering methods on the networks are presented and discussed. The base network statistics are presented in Table 3, which resulted from the initial processing of the data to remove certain functional classes and clean the data, but before any links were filtered. For ABM this means that stacked links that represent two-way links were removed, which should give a more accurate reading of the total mileage represented, but centroid connectors are still included. Just from this data, it is clear that OSM carries the most spatial information due to the amount of total road mileage. However, HERE contains more links and has a shorter average link length which indicates that it may be more a more refined network.

Table 3. Base link statistics.

Network Name	Number of Links	Number of Nodes	Total Length (mi.)	Average Links Length (feet)
ABM	2,541	2,064	182	379
HERE	12,286	9,535	401	172
OSM	10,223	13,868	627	324

Table 4 shows road link statistics. Road links were generated after the base links were filtered using the specified road filter methodology for each network. Because HERE and OSM contain service roads and bicycle paths, which tend to be short in length, the average link length was expected to increase, given that road segments would be longer. This was largely what was observed, with the exception of ABM, which decreased its average link length.

Once centroid connectors were filtered out of ABM, it was clear that there was a wide gap in spatial detail between ABM and the other two all-streets-and-paths networks. The ratio in total network length is about a 3:1 increase in detail going from ABM to OSM or HERE. HERE, while having less total network mileage, still has a smaller average link size compared to OSM. This makes sense when considering the number of nodes present in the HERE network. From these initial results, HERE was identified as a better candidate for initial conflation with the ABM road network than the OSM road network

Table 4. Road link statistics.

Network Name	Number of Links	Number of Nodes	Total Length (mi.)	Average Links Length (feet)
ABM	1,661	1,519	88	278
HERE	5,575	4,771	225	213
OSM	3,008	2,154	230	403

Based on the numbers in Table 5, OSM was deemed to be the best source for adding in bike-specific links. While HERE does contain many links that would be eligible for inclusion in the bicycle subnetwork, OSM has about 20 more miles of bicycle links in its subnetwork and a smaller average link length. While HERE has the trails in Piedmont Park, it does not have any of the Atlanta Beltline multi-use path.

Table 5. Bicycle link statistics.

Network Name	Number of Links	Number of Nodes	Total Length (mi.)	Average Links Length (feet)
ABM	-	-	-	-
HERE	2,025	1,791	47	123
OSM	3,049	2,936	68	117

As discussed earlier in this report, service subnetworks were not used in the current reconciliation conflation process to generate the new network, but as Table 6 shows, there are a lot of service links in the study area. These links are worth keeping in case they need to be added later. Service links in some cases can serve as important connections or shortcuts to cyclists. However, in many cases the service links are dead ends. For this stage of BikewaySim though, we just wanted to demonstrate how a routable network could be constructed with just the road and bicycle subnetworks.

Table 6. Service link statistics.

Network Name	Number of Links	Number of Nodes	Total Length (mi.)	Average Links Length (feet)
ABM	-	-	-	-
HERE	4,480	4,813	114	134
OSM	6,329	7,296	151	126

Figure 12 demonstrates the results of the filtering process. In the first column on the far left side, the unfiltered base network is shown in grey. It is difficult to distinguish features in this state, and any conflation attempts on using the base network could result in the conflation of road links with bicycle links or service links. This base network is then shown broken up into individual subnetworks in the next three columns. This separation highlights the differences between each network. For example, in the second column for the road subnetwork, it becomes clear that the ABM network does not have nearly as much coverage as OSM or HERE. In the fourth column, it appears that OSM shows more service links than HERE does. The last thing that Figure 12 demonstrates is that conflation should occur within subnetworks as this is where there are links shared between networks, but links from different subnetworks can be combined. In the next step, we will use these filtered networks and reconcile them into a combined network.

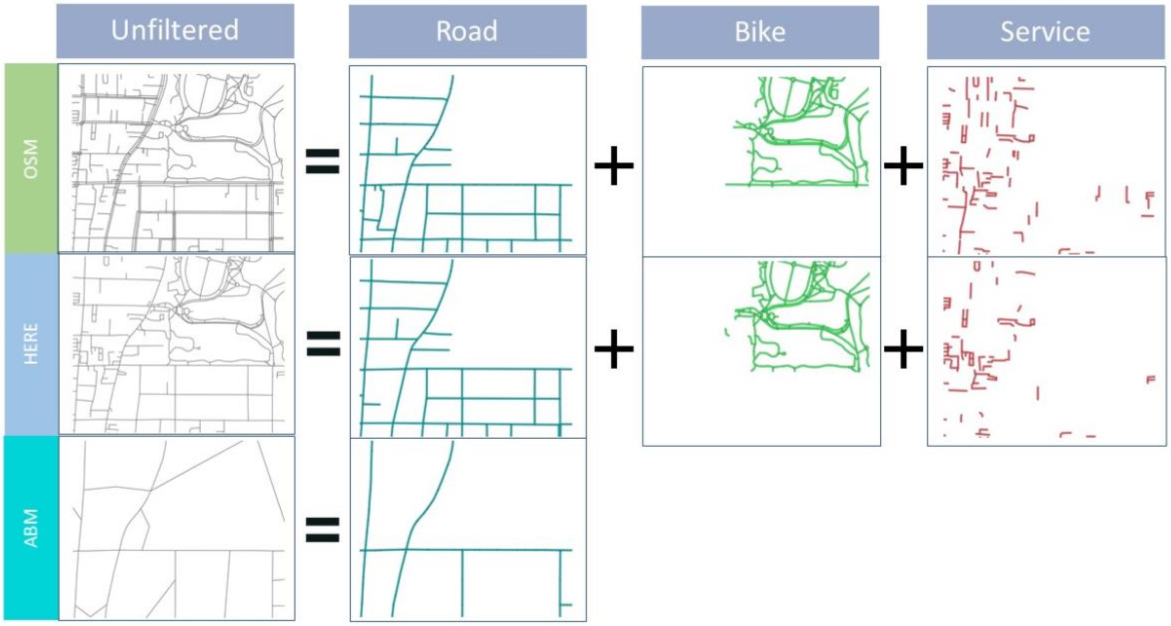


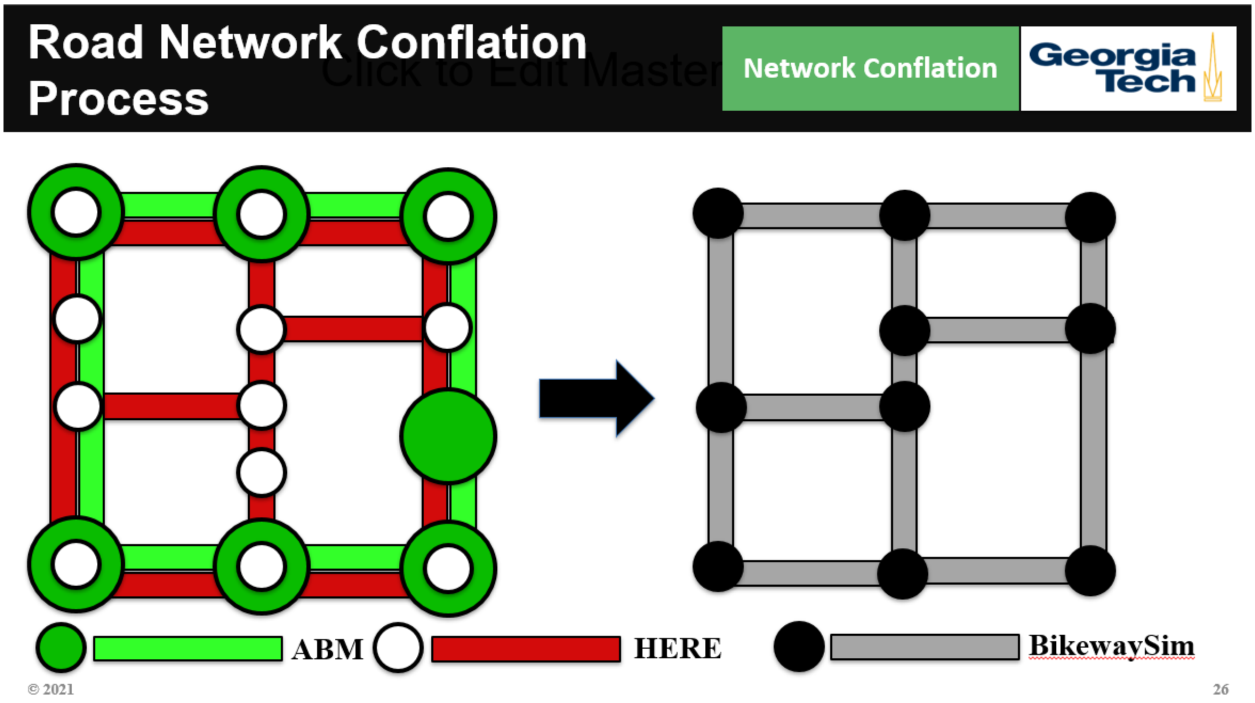
Figure 12. Visual example of subnetwork combining and shared links within subnetworks.

BikewaySim Network Generation and Conflation Methodology

As discussed in the previous section, filtering processes were used to first generate the cleaned base networks for each of the three original network sources, and then to break each of these three base networks into subnetworks for road, bike, and service facilities. The next step is to reconcile and combine the filtered subnetworks to create a comprehensive BikewaySim network.

In looking at the coverage by facility type, network densities, and overlap between the ABM network (whose nodes needs to serve as a connective structure between travel demand models, simulation models, and monitored activity), the decision was made to take the ABM network as the starting point, integrate ABM roads into the starting structure, and then integrate OSM roads and bicycle facility networks to construct the BikewaySim network.

When more than one of the same type of subnetwork is chosen for creating a combined network, these subnetworks will need to be conflated. Road subnetworks from different sources do not generally employ the same nodes and links. Even if two networks place a node at the same roadway intersection, the numbering scheme is different, the latitude/longitude position is different, and the nodes carry different attributes. Even when the same attributes are carried, such as road classification, the variables names are different, the coding is different, and the data values are often different. These differences can be fairly pronounced with the crowdsourced OSM network, where accuracy issues are a function of data input source and an unlimited number of tags are available to carry attribute information. The ABM network and HERE road subnetwork share many links (in part because the ARC generated the ABM 2020 network from 2018 licensed HERE data). However, HERE carries many additional road links that the ARC elected not to include as ABM network links. Shortest path modeling would take weeks to run if every local road were included in the regional travel demand modeling process, so the ABM network removes most of the local roads and represents travel on these smaller functional class links to and from the origin or destination and the main network, as centroid connectors with an average travel time and distance to get onto or off of the main network via the local roads. In addition, both networks have broken longer stretches of roadway links that definitely share the same two end nodes into different sets of smaller links (inserting new nodes along the way and breaking the longer links into shorter links). For example, this may be done to insert a node location to represent where traffic may enter or leave the network from a parking lot; these insertions were made by different personnel and for different reasons. Hence, in joining data from these two networks to create a more comprehensive bicycle-accessible infrastructure network, the spatial data need to be conflated, a process in which overlapping spatial data reconciled (node locations are modified, existing nodes are adjusted, unneeded nodes are removed, new nodes are added, and all nodes are renamed) to create a more accurate and comprehensive network. Figure 13 illustrates the conflation process in which ABM and HERE nodes are first matched to a reference spatial position, ABM network nodes and links (green) are defined as base nodes, new HERE network nodes and links are added, and then excess ABM and HERE nodes that do not provide network connectivity are removed. The grey network represents the resulting BikewaySim network.



Road Network Conflation Process

Network Conflation

Figure 13. Graph network representation of road subnetwork conflation process.

Figure 14 provides a visual example of how the network creation process will occur. ABM serves as the starting point, and HERE nodes and streets are added and reconciled to a new roadway network. Because OSM is the only bicycle subnetwork being used, the bicycle subnetwork needs to be added to the roadway network, which requires another conflation process to connect the bicycle network with the road network, so that the bicycle links can also be utilized in the shortest path calculator. Note that service links were not included in this version of BikewaySim and thus are not included in the network generation process.

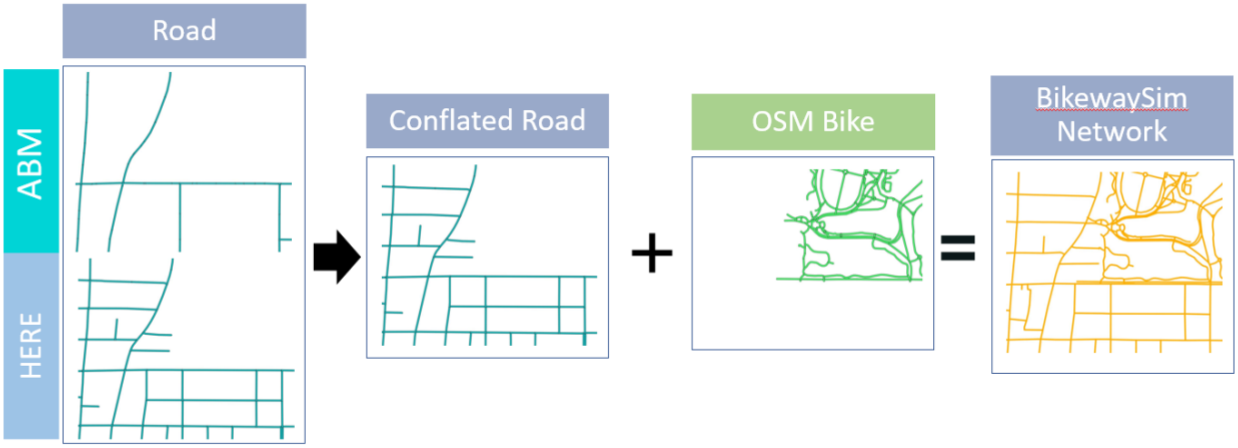


Figure 14. BikewaySim Network Generation Diagram.

Below in Figure 15 is a generalized process flow for how the BikewaySim network graph was generated from the ABM road, HERE road, and OSM bicycle subnetworks. After the filtering process is complete, the ABM road and HERE road subnetworks go through an intersection matching process in which the road intersections that are shared between the networks are matched. Then the remaining HERE road nodes are used to split up the ABM links when there should be an intersection. Next, the links that HERE has that don't overlap with the ABM network are added in. Next, the OSM bicycle subnetwork is brought in and added to the current set of conflated links. Lastly, on-street bicycle infrastructure data from the ARC bicycle facility inventory are attached to the relevant road links.

The following sections of this report go into detail on how each step was performed. There will be a subsection for each gray box in Figure 15. A set of instructional PowerPoint slides were developed for this process too. Some of these slides are reproduced below to give visual examples of the conflation steps.

Note that the OSM road subnetwork can also be included in this stepwise process and will be included in future versions of the BikewaySim network. For the study area, we were confident that HERE was not missing any road links, and that the difference in road links shown in Table 4 was due to how OSM was cropped into the study area rather than additional road links.

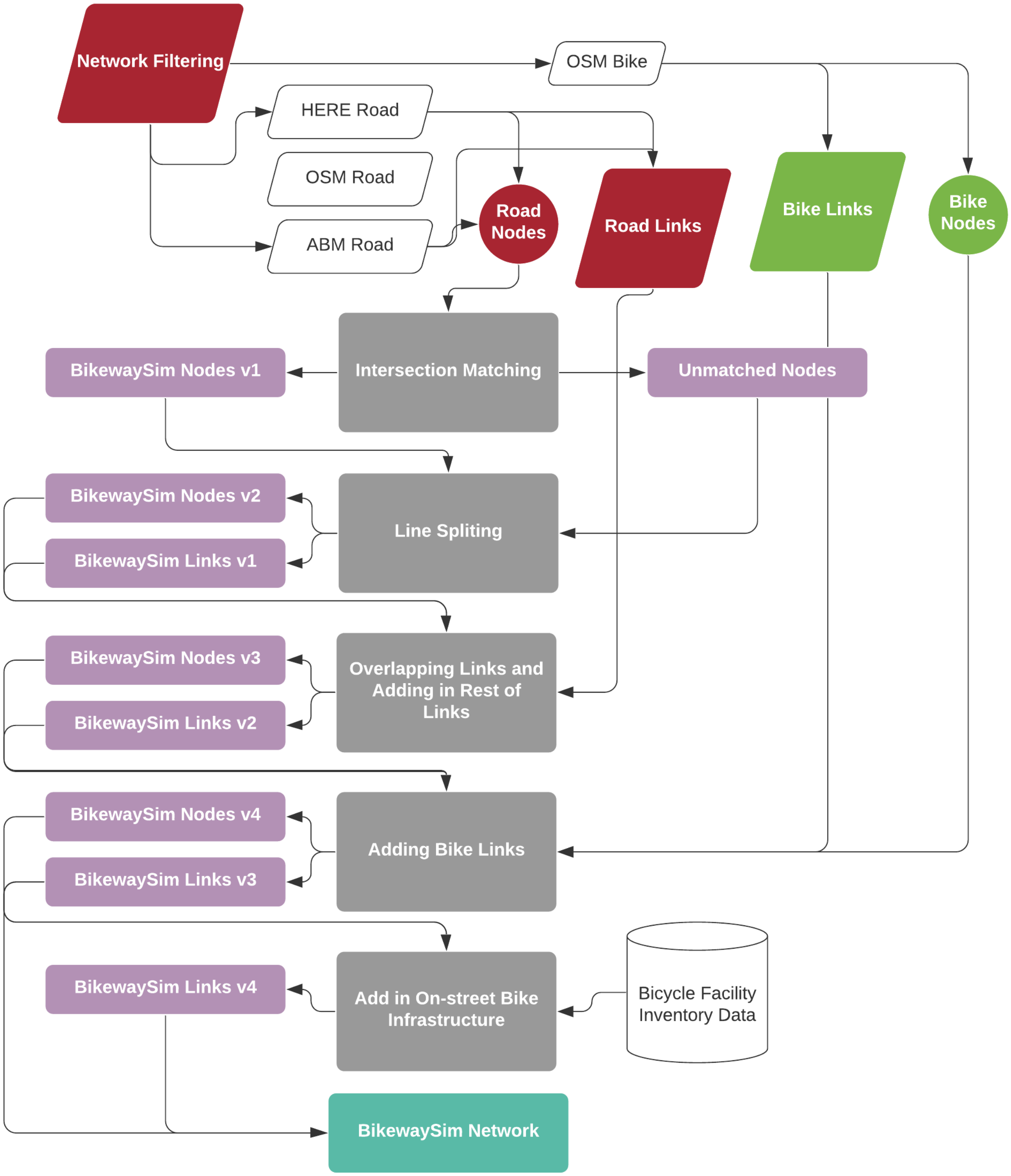
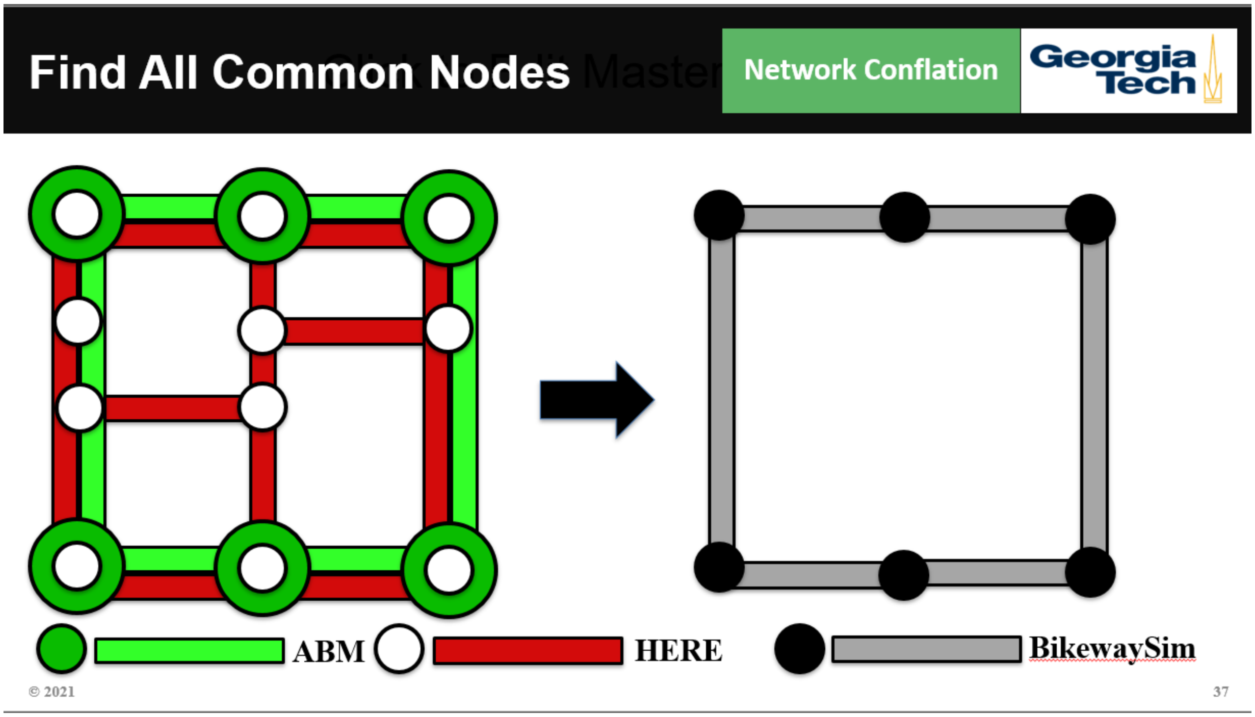


Figure 15. Process flow for network conflation process.

Intersection Matching

BikewaySim uses node-to-node connectivity and link impedance along each edge or connection to perform shortest path routing. The location of any of these nodes in physical space (i.e., their latitude and longitude or UTM coordinates) is irrelevant in these models, only the specification of which nodes connect to other nodes. Even the link distance between nodes is irrelevant, because these models use a link cost (typically all costs converted to time) to represent the impedance along that route. Hence, the shortest distance path from New York to San Francisco will be through Sydney Australia, if the links between New York and Sydney and Sydney and San Francisco are each coded with 10-minute travel times. In conflating a network, node position in space is desired, but not required for a shortest path models to function properly. Nevertheless, spatial accuracy of the networks is important because the networks will likely be used in routing apps and map displays in the future

The first step in the conflation process was to find all common nodes across the ABM and HERE road subnetworks that serve as road intersections. A visualization of this process is shown in Figure 16. Road intersections include any intersection of two or more public roads. This includes signalized and stop controlled intersections, but it does not include intersections where a public road meets a service road, such as a parking lot access road or a driveway. Both HERE and ABM contain many nodes that do not represent an intersection of two or more public roads. In ABM's case, this could be because modelers chose to put in specific high-traffic generating points like shopping center entrances, school entrances, and apartment complex entrances. In HERE's case, nodes also represent locations where the service and bicycle subnetworks meet with the road subnetworks. Additionally, for both the ABM and HERE links, nodes that represent access ramps to Interstates still remain after the previous filtering process removed the connecting freeway ramps.


Find All Common Nodes

Network Conflation

Georgia Tech

© 2021

37

Figure 16. Find common intersections.

Nodes that serve as intersections between multiple roadway links are easy to recognize in the network. By reviewing all of the links in a network and counting the number of times a node appears in the list, the result indicates the number of links that connect through that node. A node representing a four-way intersection would appear four times. By extension, a node representing a parking lot entrance or some other land use access point and connects only two links would only appear two times. This number can be calculated for all nodes in each subnetwork, as shown in Figure 17.

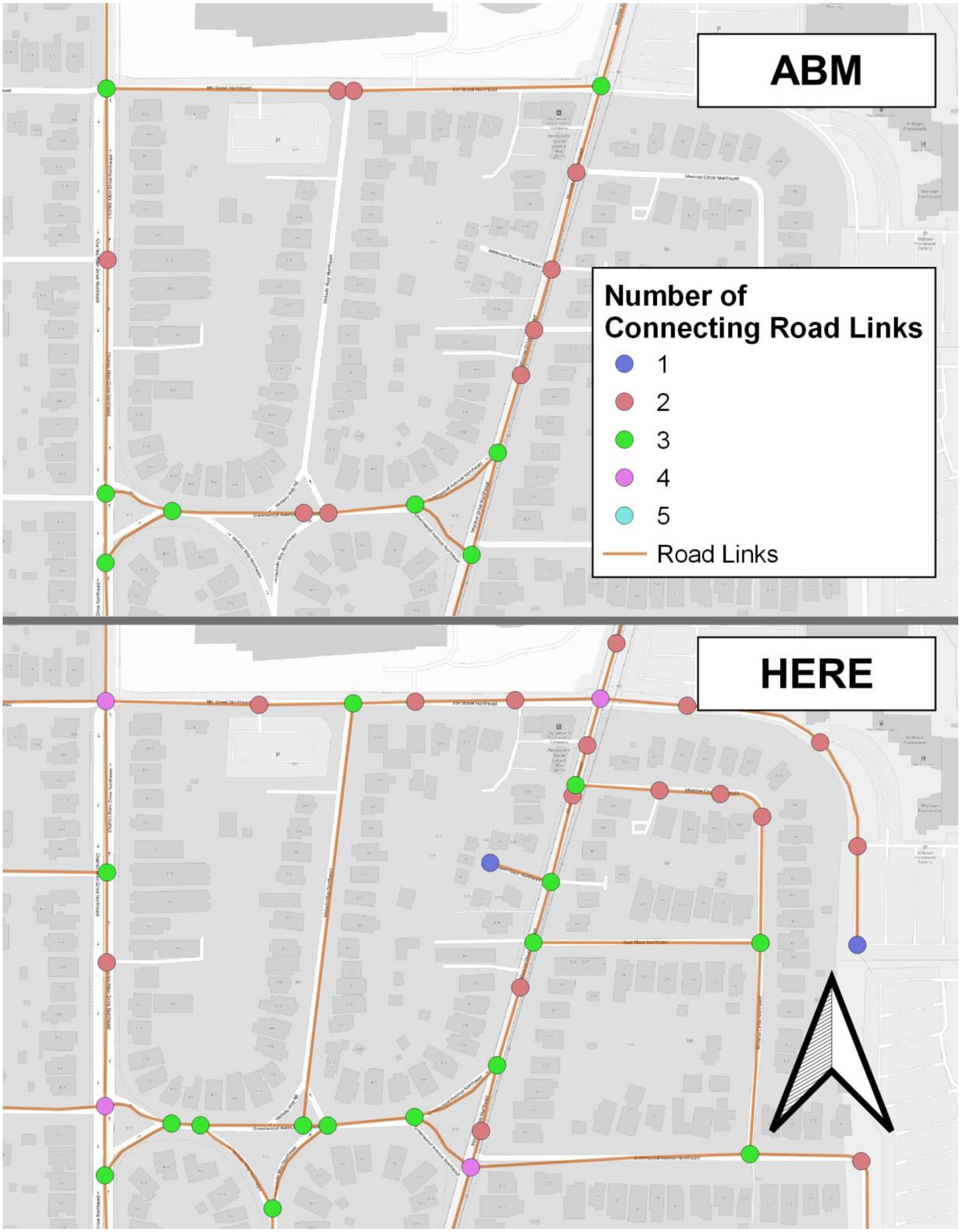


Figure 17. Number of connecting links demonstration.

Nodes that only connect to two links were filtered out of the high-density HERE road subnetwork using this calculated value. However, nodes connecting only two links were retained in the ABM network. ABM is a simplified network, which means that some of the nodes that represent intersections may only have two connecting links because one of the missing connecting links is a local road that was excluded from the ABM network (or may have been clipped in a cropping process). As coded in the ABM model, there is no way of telling which ABM nodes represent intersections vs parking lot access points. On the chance that some of these nodes have a HERE analog, they were retained in this step.

After filtering the HERE nodes, the next step matches each ABM node to its closest node in the higher density HERE network. When a nearest HERE node has been found for each ABM node, the distance separating the node infers the match. If any two or more ABM nodes would be matched to the same HERE node based upon a specific separation distance threshold, the match with the lower matching distance was accepted. Once an ABM and HERE node was matched, neither node was considered for future matching.

The results of matching based upon different tolerance distances are shown in Table 7 below. Each row of this table starts with stating what combination of network was used (either ABM+HERE or ABM+OSM). Then it lists the match distance tolerance in feet. A match distance tolerance of 10 feet would mean that an ABM node would only match to an OSM or HERE node if the distance between them was less than or equal to 15 feet. The next column reports the number of unique matching node pairs. Using ABM and HERE, for a tolerance of 15 feet there are 808 matches. The next columns report the number of duplicate matches, the remaining ABM nodes that have not been matched, and the remaining ABM nodes that have either less than or greater than 2 connecting links. This last column, the remaining nodes column, and a visual inspection of the matches were how a tolerance distance was selected.

As seen in Table 7, this process is able to match more ABM nodes with HERE nodes than it is able to match ABM nodes with OSM nodes at each match distance threshold. As the matching tolerance distance increases, the number of duplicate matches increases. Additionally, the number of matches seems to plateau after 25 feet. Setting the tolerance distance too high could mean inaccurate matches. This observation accompanied with viewing the matches in QGIS led to the selection of a match tolerance of 25 ft for ABM and HERE.

Table 7. Nearest HERE node from each ABM node results.

Network Combination	Match Distance Tolerance (feet)	Matched Nodes (no duplicates)	Duplicate Matches	ABM Remaining Nodes Total	ABM Remaining Nodes (2 links filtered out)
ABM+HERE	5	783	0	878	40
ABM+HERE	15	808	6	853	35
ABM+HERE	25	825	32	836	31
ABM+HERE	35	835	73	826	29
ABM+OSM	5	99	0	1420	247
ABM+OSM	15	601	8	918	108
ABM+OSM	25	716	36	803	70
ABM+OSM	35	758	83	761	56

After matching, the remaining unmatched nodes were examined graphically to see how many true intersections remained in QGIS. The remaining nodes did not represent real roadway intersections with a few exceptions. As shown in Figure 18, most of the unmatched nodes with 1 or 3 connecting links (shown as yellow and pink respectively) were either on the extents of the study area or at complex intersections. Most of the unmatched nodes had 2 connecting links, which means that they likely represent modeling nodes placed into the network by the ARC for travel demand modeling purposes. These nodes are not likely to have a HERE analog.

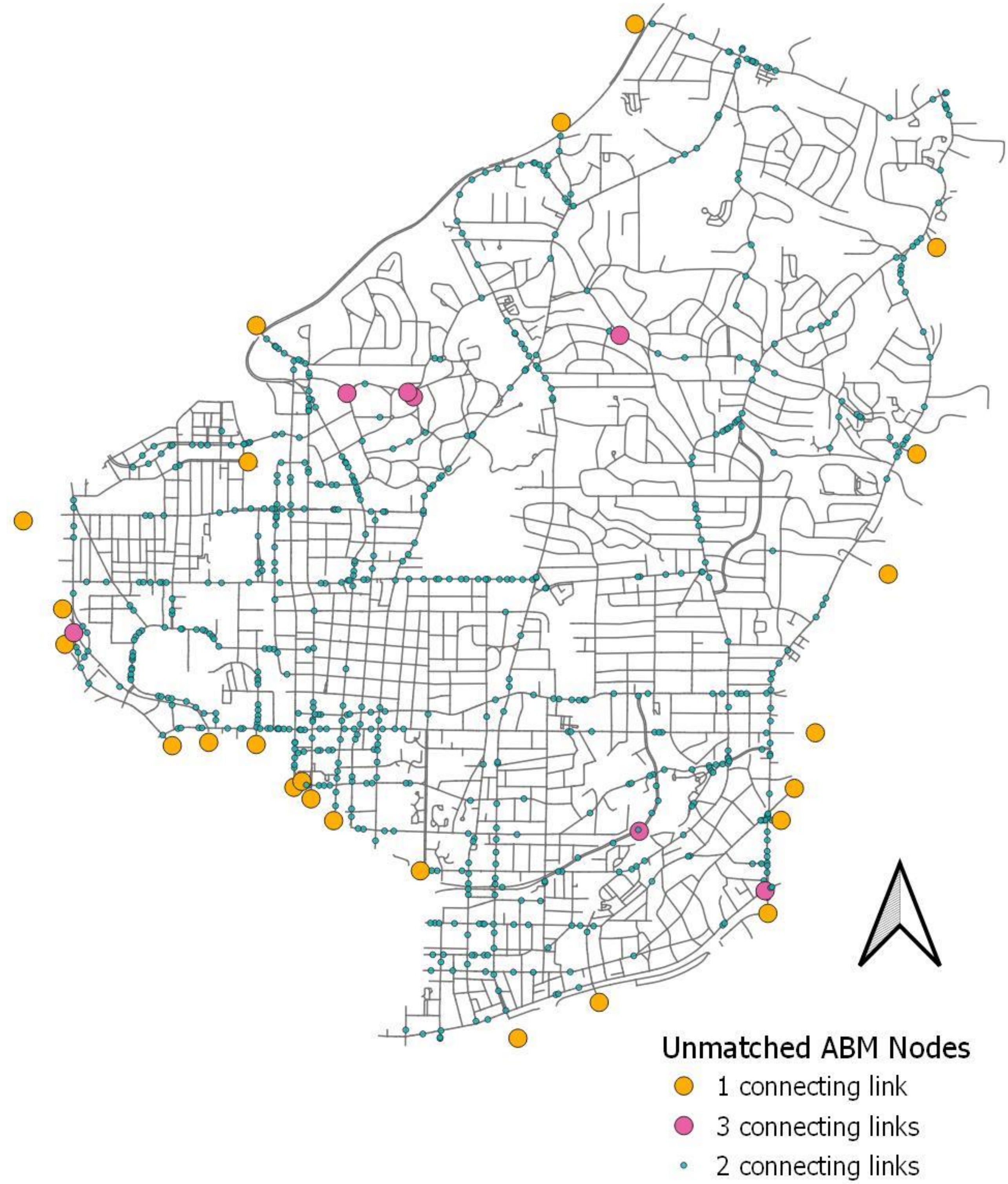


Figure 18. Unmatched ABM nodes by number of connecting links.

In a few cases though, the ABM network contained real intersections that did not match with HERE. These nodes were often found as part of roundabouts, which are represented differently across the three networks, as seen in Figure 19 below. The other case was when a road diverged, as seen also seen in Figure 19 below. These types of issues need to be resolved through a manual QA/QC process.

The nodes that did match correctly represent the first version of network nodes for the BikewaySim network. These nodes are shown in Figure 20. At this stage, BikewaySim is composed of 1519 nodes (no links between have been generated thus far).

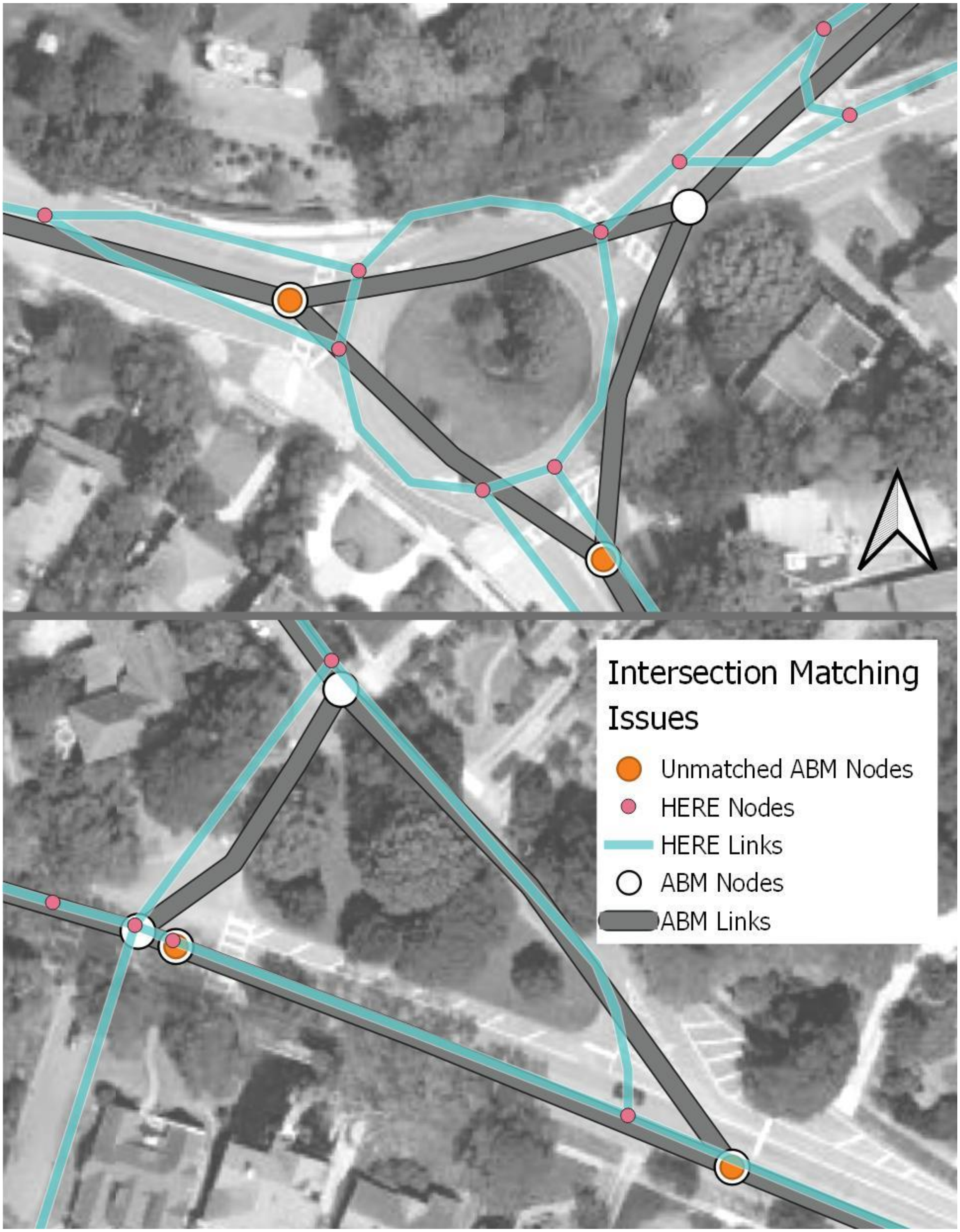
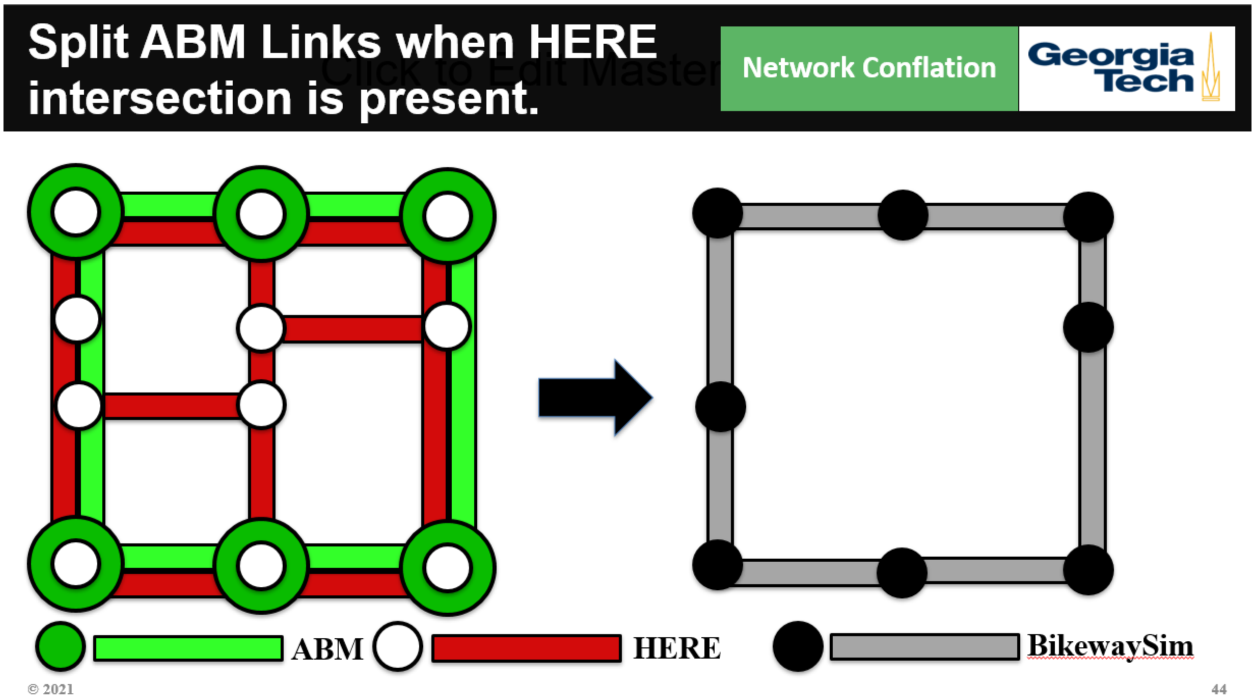


Figure 19. Intersection matching issues. Orange shows ABM nodes that didn't match to HERE nodes.

Figure 20. BikewaySim Nodes Version 1.

Link Splitting

The next step in the process is to split ABM links where an underlying HERE node represents a real intersection. This issue is shown in Figure 21 below, where some HERE nodes represent intersections that ABM does not include. This is rectified in this next step.



Split ABM Links when HERE intersection is present.

Figure 21. Splitting ABM links to accommodate HERE links representing real intersections.

The unmatched HERE nodes from the intersection matching step are used to split ABM links. These unmatched HERE nodes have already been filtered to remove any nodes with only two connecting links. From each HERE node, the nearest shape point in ABM link is identified, using the same tolerance distance of 25 feet that was used in the previous step. A new BikewaySim node is generated at the location of the ABM shapepoint node, and inserted into the BikewaySim data set, splitting the existing ABM link into two links that now connect at the inserted node location. Because new nodes are created in this process, they need a unique ID associated with them. Using the convention from Table 1, the first three numbers were 1, 1, and 3. The numbers following this were generated sequentially starting from 1. This allows these new nodes be matched back to their origin ABM link by referencing the nearest nodes file created in the process of finding the nearest point on a link.

Overall, 117 HERE nodes were able to match to a shape point on 94 ABM links. Splitting these ABM links added a total of 202 BikewaySim links. Figure 22 shows some interesting case examples of how splitting the ABM links works. The ABM roundabout nodes that did not match to anything back in Figure 19 now match to the HERE roundabout nodes because they were joined to shape points in the ABM network file. The bottom image in Figure 22 demonstrates how nodes are successfully added to the ABM network for road intersections that were not represented in ABM before. The red diamonds represent HERE nodes that matched to ABM links, and the blue circles represent the point at which a HERE node matched to an ABM link. The distance between the blue circles and red diamonds shows how close the match was. In the roundabout of Figure 22, there is one blue circle that is far away from its red diamond. This shows that roundabouts will need to be manually edited after this process. The white circles represent the original ABM nodes, and the small red nodes represent the HERE nodes.

The new split links and nodes are then added to the BikewaySim network. This represents the first version of the BikewaySim links and the second version of the BikewaySim nodes. This network is shown in Figure 23 below. There are 1,772 links and 1,636 nodes in this network.

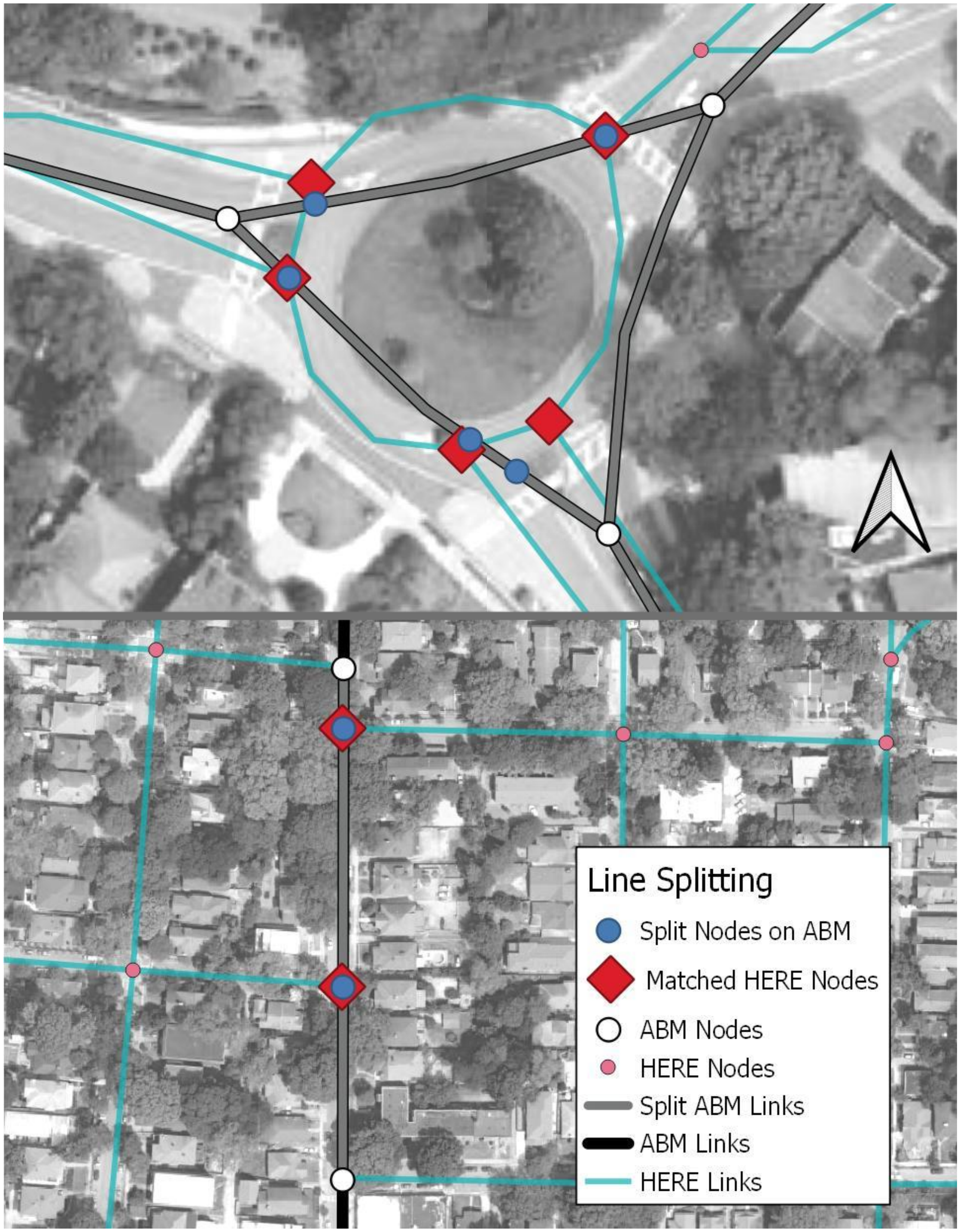


Figure 22. Results of line splitting process.

Figure 23. BikewaySim Network with added nodes and split links.

Overlapping Links and Adding in Rest of Links

At this point, the commonalities shared between the ABM and HERE road nodes have largely been addressed. The unmatched nodes remaining in HERE generally represent intersections with service links. The unmatched nodes in ABM are generally nodes representing travel demand model specific treatment locations for parking lots, or they are nodes serving as placeholders for future integration of dynamic traffic assignment routines.

The unmatched HERE nodes still need to be addressed. Remember that BikewaySim links are defined by reference nodes in order for a link to have attribute data, it must have a pair of node IDs that match to a link in the ABM or HERE network. In other words, if a link has 0 HERE IDs, 1 HERE ID, or 2 HERE IDs that do not match to a HERE link, that link won't carry HERE attribute data.

This problem is shown in Figure 24, which shows several unmatched HERE nodes in red overlapping with a single BikewaySim link. Even though the one BikewaySim link shown in Figure 24 in white has two HERE nodes as reference points (see the red HERE nodes that are overlapping with the large pink BikewaySim nodes), there is not a HERE link with these nodes as reference points. This is because there are several HERE links that overlap with the single BikewaySim link.

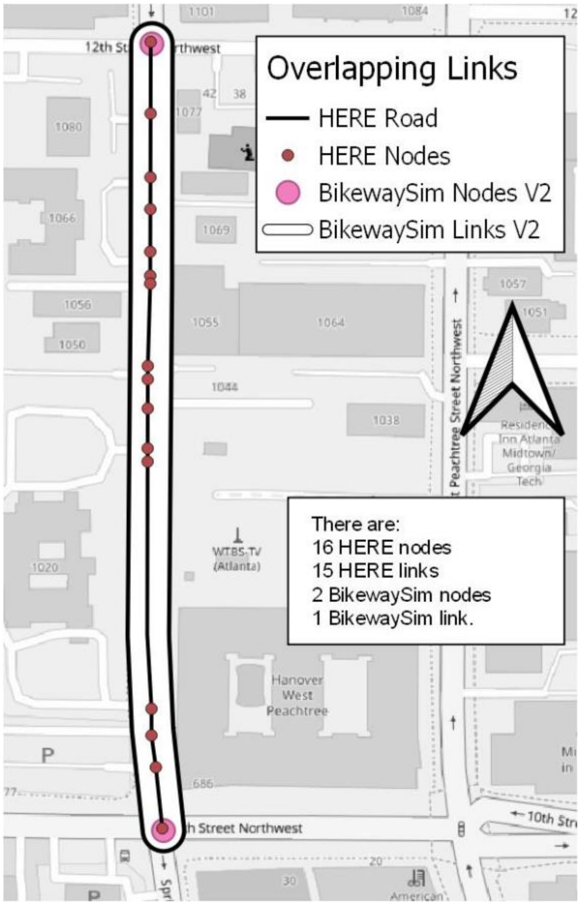


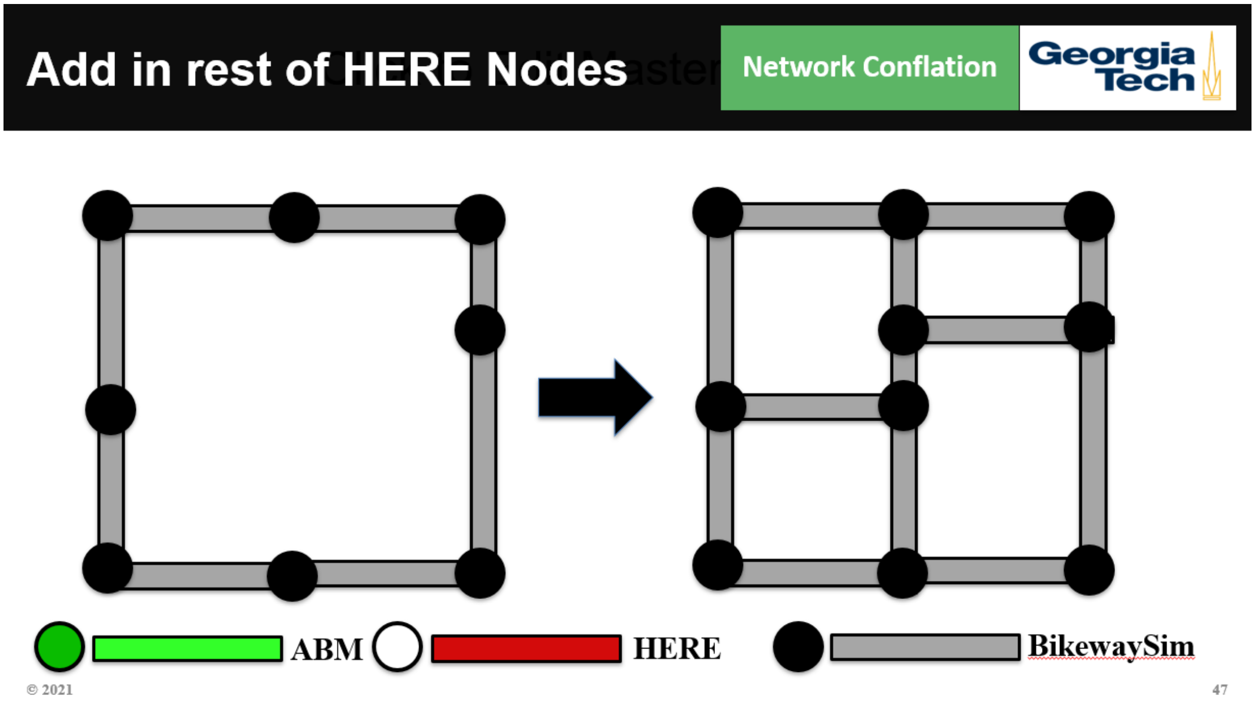
Figure 24. Overlapping Links between BikewaySim and HERE

There were two possible approaches to solving this issue. The first and chosen path was to just associate HERE link information to the BikewaySim network by finding the HERE link that had the greatest length of overlap on a BikewaySim link (shown in orange in Figure 24). This was done by buffering the BikewaySim links by 30 feet and performing an intersection with the HERE links. HERE links that were completely covered in this buffer were filtered so that only HERE links with no known ABM analog remained. These links can then be added in without the need to conflate to anything.

In the set of HERE links that overlap BikewaySim links, the link that had the greatest amount of overlap with its associated BikewaySim link was paired with that BikewaySim link. This association was recorded in a separate column in the BikewaySim links, so that a record would be retained that the matched HERE link should only be used for importing HERE link attributes into BikewaySim link attributes. These additional HERE nodes were removed because they were not necessary for routing (because these are not intersections). However, these nodes may represent locations where the number of lanes changes, the speed limit changes, or there is some other factor that a navigation app tracks and uses in estimating congestion or displaying messages to a driver.

As the BikewaySim network is further developed, the team may elect to split the current set of BikewaySim links into smaller links by incorporating more HERE nodes. Splitting the links and assigning HERE attributes to the different links, may allow a shortest path algorithm to account for these differences in link impedance calculations (this will come at a cost of computational efficiency). This approach was not taken for this report because the nodes themselves were not necessary as none of the HERE service links were incorporated into this network. Both the service links and road attribute changes will be considered for inclusion as BikewaySim develops.

Now that these links have been addressed, the rest of the HERE links can just be added to the BikewaySim network. The result of this is shown in Figure 25 below. It is the 2nd version of the BikewaySim links and 3rd version of the BikewaySim Nodes. There are now 4,484 links and 5,296 nodes.



Add in rest of HERE Nodes

Network Conflation

Georgia Tech

Figure 25. Bringing in the rest of the HERE links.

Figure 26. BikewaySim Network with HERE links added.

Adding in Bike Links

The last step in the BikewaySim network generation process is to add in the OSM bicycle subnetwork. This subnetwork is linked to the current BikewaySim network by finding the nearest BikewaySim node from each OSM bicycle node. Node pairs that were within 25 ft apart were considered connected. Note that these node pairs need to be ground-truthed through a QA/QC. The network with the added bicycle paths is shown in Figure 27 below. This is the 4th version of BikewaySim nodes and the 3rd version of BikewaySim links. There are 7,658 links and 8,228 nodes.

Figure 27. BikewaySim links with OSM bicycle subnetwork added.

Adding in On-street Bicycle Infrastructure

Because neither the ABM or HERE carries information about bicycle facilities, and because OSM's on-road bicycle facility tagging lacks the attributes contained in the ARC's bicycle facility inventory, on-road bicycle lane information was coded in using a simple buffer and filter method. The ARC bicycle facility inventory shown earlier in Figure 3, was buffered by 30 feet. Then this buffered layer and Step 3 of the BikewaySim links were intersected using GeoPandas' overlay function. To prevent a mistaken association with a bicycle facility on an intersected link, only intersections in which at least 90% of a BikewaySim link was covered were accepted. As BikewaySim develops further, more accurate results might be achieved by snapping the ARC bicycle facility inventory lines to the bicycle network. The number of links and nodes in the BikewaySim network remained the same after this process.

BikewaySim Network Graph

The finalized BikewaySim network graph attribute table contains a column for each network indicating what node ID it has with that network. If a node does not have an association to one network, then it will be filled with a null value. For BikewaySim nodes, the BikewaySim node ID column can be used to tell which network a point's geometry originated from. For BikewaySim links, the BikewaySim reference node columns can be used to tell which network a link's geometry originated from. The BikewaySim ID will be filled based on what network IDs are available. If an ABM ID is listed, then the BikewaySim ID field will be populated with the ABM ID. If no ABM ID is listed but a HERE ID is listed, then the HERE ID is used. If neither an ABM ID or HERE ID is available then the OSM ID will be used. Every point and link have at least one network associated with it. An example of how this hierarchy works is shown in Table 8.

Table 8. Example of BikewaySim ID hierarchy.

ABM ID	HERE ID	OSM ID	BikewaySim ID
110XX	210XX	320XX	110XX
-	210XX	320XX	210XX
-	-	320XX	320XX

XX – Represents the specific node ID following the three identifying numbers

The final BikewaySim network information is shown in Table 9 below. While all attributes can be carried on the BikewaySim network, there is some processing that is needed for the network to be in a usable state for shortest path routing. For instance, the way speed limit is defined across the networks differs. The HERE network groups speeds into categories while the OSM network just lists the speed. Even when the data format is the same, there may be disagreement between the networks where one network reports the speed as 25 miles per hour when the other reports it as 30 miles per hour. This is true for data on road classifications as well. All three networks indicate a functional classification for its links, but these functional classifications are usually not the same across networks. As such, BikewaySim's link attributes will need to be further examined and reconciled in the future. In the meantime, this initial BikewaySim network can be used to calculate the shortest path from any origin to any destination using BikewaySim.

Table 9. BikewaySim Network information.

Network Name	Attributes	Number of Links	Number of Nodes	Total Length (miles)	Average Links Length (feet)
BikewaySim	477	7762	8228	306	208

New networks or updated versions of the networks used in this study will become available in the future. Updated versions of existing network data can be brought in as long as the updated network uses the same reference IDs. In the case where new links are added to one of the existing networks, it may make more sense to manually add in the new links depending on how many there are.

In the case of a new network, the filtering and conflation process described in the above sections of this report can be used to reconcile the new network data with the existing BikewaySim network.

BikewaySim Shortest Path Calculator

In this section, the BikewaySim shortest path calculator is described. The BikewaySim shortest path application reads a user-supplied set of coded links and nodes, calculated impedance values for each link, and a set of origin-destination (OD) pairs and calculates the shortest path between each OD pair using Dijkstra's algorithm (*1*). The Python-based code allows users the flexibility to employ any set of equations to calculate impedance values for each link, prior to performing the shortest path optimization run. By default, this impedance method assigns the impedance in the form of travel time to each link. Hence, user-defined impedance functions need to convert other costs, such as tolls or preference costs, to a travel time increment that is added to distance/speed based travel times. BikewaySim utilizes the Python NetworkX package to construct each network graph and perform the Dijkstra's shortest path routines (*1*, *17*).

BikewaySim uses the set of user provided links and nodes to construct a network graph. Hence, before the user-provided network can be converted to the BikewaySim network graph, the user needs to ensure that the network they are providing conforms to BikewaySim structure requirements. All links need to be defined by corresponding from-to node IDs, and all node IDs must be present in the nodes file. In the BikewaySim network, these reference IDs can be found in the 'bikewaysimNodeA' and 'bikewaysimNodeB' columns.

Once the network graph is created, the program reads the trip input array, where every trip is reported as a row and the origin and destination locations are specified with both latitude and longitude position values and in UTM coordinates (NAD 83 Georgia West). For each trip, the

BikewaySim finds the closest node in the network to the origin and to the destination. The program adds a travel time increment to the start and end of each trip to represent the time it takes to get from the trip origin location to its closest network link, and to the trip destination location from its closest network link. These travel time increments are currently based upon an assumed walking speed of 2.0 miles per hour, but this default value can be modified by the user. The research team is considering using a higher network access speed in the next BikewaySim version, to represent the use of a bicycle to get from the trip origin location to the network (some constrained bicycle speed that would account for mobility limitations getting to the roadway network). The BikewaySim network is very dense in more urbanized areas, so most trips begin and end close to a network node. However, this is not the case in rural areas, where the roadway network is more dispersed and a trip might start within a large parcel of land some distance from the nearest network node. The research team will need to address this limitation in future versions of BikewaySim. In an area where a TAZ is large and the road network is sparse, it might be more desirable to have more control over where an origin or destination matches to the network. In addition, structuring the all-streets network into sets of microanalysis zones (MAZs) (typically by census block) will be important in enhanced travel demand modeling (*12*).

Once the network graph is constructed and the OD pairs are matched to the network graph, shortest path calculations can begin. NetworkX uses Dijkstra's' algorithm to calculate the shortest path, but this routine can be modified to report back k-shortest paths (*1*, *17*). The shortest path (series of links from origin to destination along the lowest impedance cost path) is retained in a CSV file that can be manipulated to plot or analyze the shortest path.

Shortest Path Routing between TAZ Pairs

To test the performance of BikewaySim, shortest path calculations were performed for trips between each TAZ pair, with the start point set as the coordinates of the origin travel demand model TAZ centroid and the destination as the coordinates of the destination travel demand model TAZ centroid. Even for this relatively small study area, there the 187 TAZs yield 34,782 TAZ origin destination pairs (duplicate TAZ pairs were removed). The 34,782 trips were run using BikewaySim on both the original low-density ABM network as well as the new high-density BikewaySim network. This code was run on a desktop with 32 Gbs of RAM and a 10 core 3.70 GHz processor. The results of these shortest path calculator runs are shown in Table 10. This table shows that the runtime for the BikewaySim network was over twice as long as the run on the original ABM network. However, most of this increase in run time was not due to performing the shortest path routine but instead matching the OD coordinates to the network (Time to prepare trips). Because bicycle trips are assumed to be unaffected by roadway congestion levels, shortest path calculations for bikes do not need to run in an iterative fashion as is in a travel demand model. BikewaySim is run once, and the shortest paths it generates can be referenced in future work without having to run BikewaySim again. The average trip distance was very similar between the two networks, although slightly less for the larger network which allowed more pathways. The distribution of trip distance across the runs was very similar too. Still, the BikewaySim links were utilized since only 70% of trip miles were routed on the ABM links in the BikewaySim network, meaning that a significant percentage of trips diverted from the ABM network onto links in the more detailed network (which should be more realistic). In addition, the number of trips per link was plotted for each network.

Table 10. Results of shortest path calculator runs.

Network Used	Average Trip Distance (miles)	% of total trip miles on ABM	Time to Build Network Graph (minutes)	Time to Prepare Trips (minutes)	Time to Find Shortest Path (minute)	Total Runtime (minutes)
ABM	2.46	100%	< 1	52	66	118
BikewaySim	2.43	70%	< 1	213	63	276

Figure 28 below shows the trips by link for the ABM network and Figure 29 shows them for the BikewaySim network. The links in the network are colored according to the number of trips that utilize them. From a visual inspection of these two maps, it is clear that the added in links in BikewaySim are being utilized.

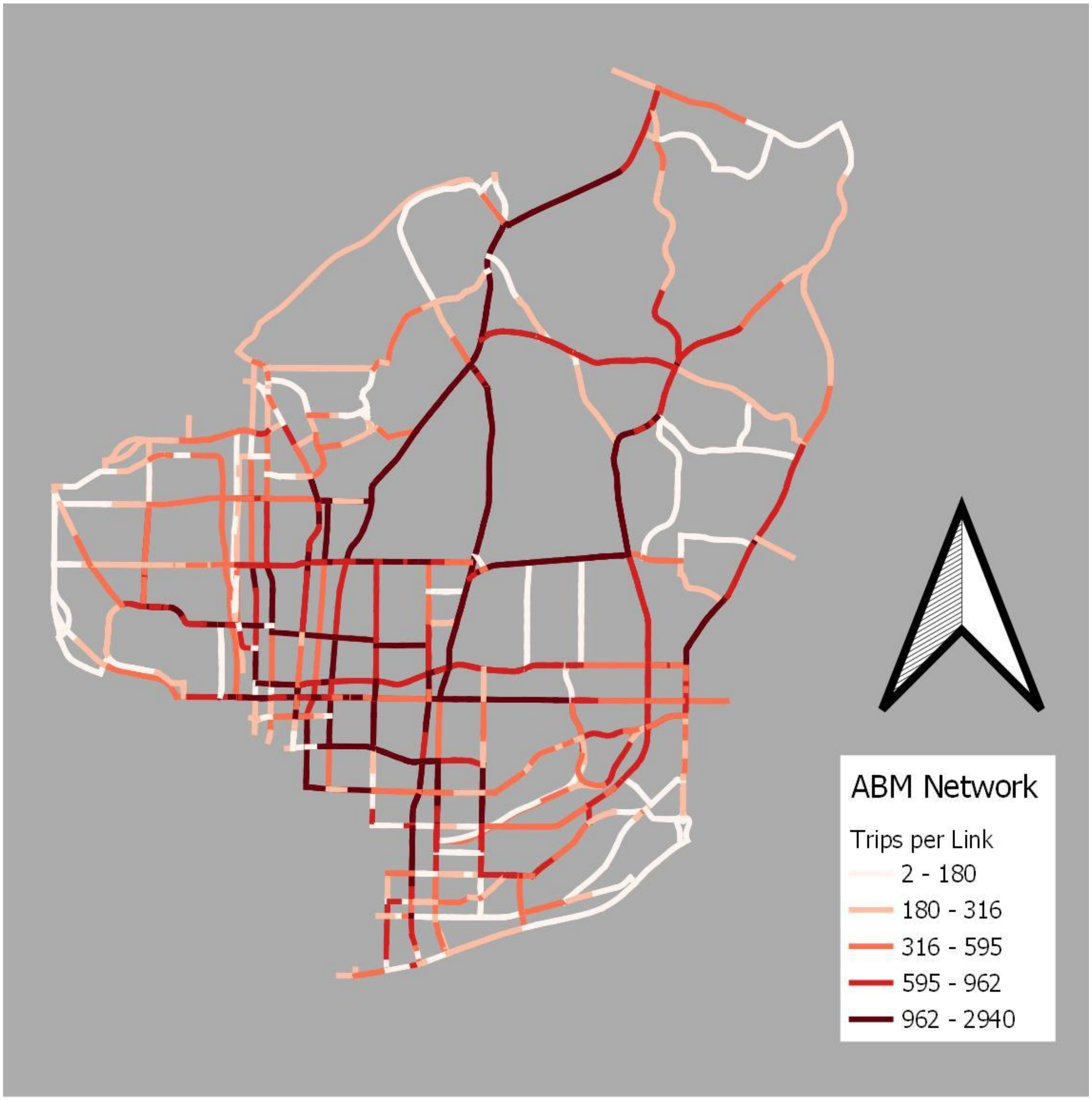


Figure 28. Shortest path routing results for ABM network.

Figure 29. Shortest path routing results for BikewaySim Network

Cost Functions for Bicycle Travel

BikewaySim currently finds shortest paths using link travel times assigned to each network link as the impedance factor. However, the shortest path routing can be performed using any impedance factor (travel time, monetary cost, utility, etc.). One objective of BikewaySim is to be able to account for a cyclist's preferences for certain link attributes. In the literature, most bicycle route choice models use a multinomial logit choice model to select the optimal route from a choice set of routes (*8*, *10*, *11*). The coefficients estimated in these models give insight as to how road attributes are preferred relative to another. Several of these coefficients are listed in Table 11 below. As negative coefficients increase, the route is less attractive to cyclists. A cyclist may consider riding on an ordinary street as four time worse than taking a multi-use path of the same length. Based on these interpretations, a distance modification factor can serve as a proxy to integrate cyclist preferences from other models. These modification factors can then be used to calculate a perceived time cost associated with each difference.

These values shown in Table 11 were calculated by treating the ordinary road coefficient as the base value and dividing all other coefficients by this base value. Distance modification values that are between 0 and 1 are considered preferable as they reduce the distance of a link while values above 1 are considered undesirable as they increase the distance of a link.

Table 11. Table of cycling utility parameters recreated from NCHRP report (*12*).

Variable	Coefficients from NCHRP report	Distance Modification
Distance on ordinary streets (miles)	-0.858	1
Distance on multi-use paths (miles)	-0.248	0.289
Distance on bicycle lanes (miles)	-0.544	0.634
Distance on arterials without bicycle lanes (miles)	-1.908	2.224
Distance on protected bicycle lanes (miles)	-0.424	0.494

The shortest path calculator was run again after including the costs coefficients from the NCHRP report, routing changes were noted in certain areas, as seen in Figure 30. Visually, the new impedance factors yield a greater concentration of trips using the paths through Piedmont Park and the BeltLine. While this may not be representative of cyclist behavior in Atlanta, this serves as a demonstration for how link costs can be modified in BikewaySim to reflect the preferences demonstrated in revealed preference studies. The next stage of research will be to further explore the literature and create a more comprehensive set of impedance factors to represent cyclist preferences.

In this application example, the mean trip distance increased to 2.7 miles when the distance modification factors from the literature were applied. This corresponds to a 12% increase in distance on average over the shortest path calculated using the time impedance, a ratio that is within the findings from the literature (*7*, *8*, *9*). One interesting finding was that only 40% of trip miles utilized ABM links when using the perceived time cost method, indicating that the denser all-paths network further increases the percentage of routes that shift from the ABM network when other factors affecting cyclist preferences are brought into the modeling approach.

Figure 30. Number of trips on links with preference factors.

Conclusion and Future Work

This project presented a methodology for reconciling multiple GIS transportation network datasets to create an efficient all-path network for bicycle trips routing. The three networks used were ABM, HERE, and OSM. These networks were cleaned and filtered into three subnetworks: road, bike, and service. Then a combination of these networks was conflated to construct a routable network graph that could be used with Dijkstra’s shortest path algorithm (*1*). Even though there are some additional modifications that need to be made to the BikewaySim network after this automated conflation process, this process still has the potential to significantly reduce time and labor costs associated with developing all-paths networks. This network is fundamental to being able to model bicycle trips based on preferences or any other impedance.

During the filtering process, these networks were compared based on their relative network density and coverage relative to one another. It was found that OSM had more total miles of both bicycle and service links, and it was also found that HERE had denser and finer road links than OSM. The filtering methods used in this project for HERE and OSM should also be transferable.

For this project, only the ABM and HERE networks were conflated in the stepwise conflation process given that the node density in HERE was higher, we were confident that the HERE network was not missing any links that OSM would have, and it would have added more time to developing the initial BikewaySim network. In the future, OSM will be added to the stepwise conflation process in addition to HERE, and the performance of OSM conflating to ABM vs HERE conflating to ABM will be compared. The BikewaySim network development is an ongoing process, and in addition to OSM roads, the service links from both the OSM and HERE subnetworks will be considered for addition if they become necessary.

There were four major steps in the conflation process: intersection matching, link splitting, adding in non-overlapping links, and adding in bicycle paths and on-street bicycle facility data. In the intersection matching process, intersections between ABM and HERE road subnetworks were matched according to a given tolerance. As this tolerance varied, it was found that the number of nodes matched plateaued around 25 feet. Because HERE is intended for navigation purposes, it often draws intersections differently than ABM. This results in there being circumstances in which network nodes do not match correctly or at all.

The link splitting process took the remaining unmatched nodes from HERE and used them to split up the ABM network into finer links. One objective of this project was to add local roads into ARC’s ABM, and this step helped accomplished this as the ABM links were not split at every road intersection.

In the adding in non-overlapping links step, network attributes from overlapping HERE links were added to ABM. With the current approach, network attributes from the longest overlapping link with ABM are added to the BikewaySim network. In the future, we would like to examine further splitting links so that attributes from one network can be better transferred to the BikewaySIm network.

The last step was adding in bike paths and on-street bike infrastructure. Adding in the bike links did not require conflation since none of the links thus far in the BikewaySIm network had a similar topology to a bike link. However, the bike network did need to be connected to the other BIkewaySim links to ensure proper network routing. These connections will need to be verified in the QA/QC process. Additionally, it is likely that there are missing connections that either were not added because the matching distance was not great enough or because the connection occurred through a service link like a parking lot road.

Shortest path routing occurred with both time impedance and bicycle facility preference impedance. It was found that when using time impedance between the ABM and BikewaySim network, the average mileage of the trips routing was only slightly less with the BikewaySim network. It also took almost twice as long to complete shortest path routing calculations using the BikewaySim network because matching OD coordinates to the nearest network node took substantially longer for the BikewaySim network. It was also found that there was a large shift of trip miles that diverted from the ABM links, which demonstrated that the additional paths added in the BikewaySim network were being utilized.

Lastly, using a preferences impedance, we further able to show a shift in routing patterns in that more bike specific links were utilized and the average detour rate from the shortest path was about 12%.

As stated in the network conflation section, there will need to be a manual QA/QC process on the conflated network to reconcile difference between network attributes and verify the validity of each step in the conflation process. There are differences in how attributes are enumerated in each network and the difference between these attribute values across networks will need to be reconciled.

However, before manual QA/QC modifications are made to this generated BikewaySim network, there needs to be a way of making sure that any changes are accessible to the public. One of the future objectives of this project will be establishing a procedure for pushing network updates to OSM. This network can also be used for asset management purposes since it contains a geospatial record of infrastructure that can easily modified. In addition to the QA/QC process, cycling ridership data need to be brought in to detect any missing links that cyclists frequent that are not included in the current network.

References

1. Edsger, D. A note on two problems in connexion with graphs. *Numerische mathematik*, Vol. 1, no. 1, 1959, pp. 269-271.

2. Federal Highway Administration. Explore Vehicle Trips Data. *NHTS*, https://nhts.ornl.gov/vehicle-trips. Accessed June 23, 2021.

3. Buehler, R., and J. Pucher. *City Cycling*. MIT Press, Cambridge, Massachusetts, 2012.

4. The League of American Bicyclists. Rates of Active Commuting. https://data.bikeleague.org/show-your-data/city-data/topic-iii-rates-of-active-commuting/. Accessed June 22, 2021.

5. The League of American Bicyclists. Infrastructure for People Biking & Walking. https://data.bikeleague.org/show-your-data/city-data/3-10-cities-infrastructure-for-people-biking-walking/. Accessed June 23, 2021.

6. Buehler, R., and P. John. *Cycling for Sustainable Cities*. The MIT Press, 2021.

7. Broach, J., and J. Dill. Using Predicted Bicyclist and Pedestrian Route Choice to Enhance Mode Choice Models. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 2564, no. 1, 2016, pp. 55-59.

8. Bernardi, S., K. Geurs, and L. L.P. Puello. Modelling route choice of Dutch cyclists using smartphone data. *The Journal of Transport and Land Use*, Vol. 11, no. 1, 2018, pp. 883-900.

9. Misra, A. Mapping Bicyclist Route Choice Using Smartphone Based Crowdsourced Data. Georgia Institute of Technology, Atlanta, PhD Thesis 2016.

10. Hood, J., E. Sall, and B. Charlton. A GPS-based bicycle route choice model for San Francisco, California. *Transportation Letters*, Vol. 3, no. 1, 2011, pp. 63-75.

11. Fitch, D., and S. Handy. Road environments and bicyclist route choice: The cases of Davis adn San Francisco, CA. *Journal of Transport Geography*, Vol. 85, 2020.

12. RSG The Rand Corporation. NCHRP 08-36, Task 141 Evaluation of Walk and Bicycle Demand Modelling Practice. AASHTO, 2019.

13. Li, H., Y. Wang, X. Xu, H. Liu, A. Guin, M. Rodgers, M. Hunter, J. Laval, and R. Guensler. Assessing the Time, Monetary, and Energy Costs of Alternative Modes. in *97th Annual Meeting of the Transportation Research Board*, Washington D.C., 2018.

14. GEOFABRIK. *geofabrik*, https://www.geofabrik.de/. Accessed August 2, 2021.

15. Documentation. *Overpass API*, http://overpass-api.de/.

16. Ferster, C., J. Fischer, K. Manaugh, T. Nelson, and M. Winters. Using OpenStreetMap to inventory bicycle infrastructure: A comparison with open data from cities. *International Journal of Sustainable Transportation*, Vol. 14, no. 1, February 2019, pp. 64-73.
17. Hagberg, A., and D. S.P. Schult. Exploring Network Structure, Dynamics, and Function using NetworkX. in *7th Python in Science*, 2008, pp. 11-15.

Data Summary

Products of Research

Three GIS datasets containing roads and road attribute information were acquired for this study: the Atlanta Regional Commission activity-based mode network, HERE streets data, and OSM data.

The other product of this research are Python scripts used for filtering and conflating the GIS networks.

Data Format and Content

The three GIS datasets used are in standard GIS data formats. The Atlanta Regional Commission activity-based model network data is in Geodatabase format and contains 75,289 roadway links and 27,524 nodes for the Atlanta metropolitan region. The HERE streets data are in shapefile format and contains 1,673,345 roadway links for the entire state of Georgia. The OpenStreetMap data are in GeoJSON format and contains 38,508 roadway links containing OSM street data for the study area.

The Python scripts are stored as PY files. The scripts are for filtering GIS roadway datasets and conflating them.

Data Access and Sharing

For the Atlanta Regional Commission activity-based model data, the public should contact the Atlanta Regional Commission, who will provide the data upon completion of a data user agreement. HERE street data must be licensed directly from HERE. OpenStreetMap data can be acquired using the Overpass API.

The Python scripts are available from the following places:

- Version used for this project: https://doi.org/10.5281/zenodo.5750140
- Latest Version: https://github.com/gti-gatech/BikewaySim

Reuse and Redistribution

For the Atlanta Regional Commission activity-based model data, the general public should review the data use agreement for restrictions. The same is true for the HERE data. OpenStreetMap data have an Open Data Commons Open Database License and can be reused and redistributed so long as OpenStreetMap and its contributors are credited.

The Python scripts are considered open data and can be reused and redistributed so long as the researchers in this report are credited. The following citation is recommended:

> Reid Passmore, Kari Watkins, & Randall Guensler. (2021). BikewaySim Technology Transfer: City of Atlanta, Georgia. Zenodo. https://doi.org/10.5281/zenodo.5750140