Bikeshare as a Substitute and Complement to Transit in Washington, D.C.

Elisabeth Ericson

CPLN 550: Introduction to Transportation Planning
Fall 2021

Introduction

In not much more than a decade, the idea of unlocking a bicycle on a street corner, riding it to one's destination, and leaving it for someone else to use has transformed from a notion that might raise eyebrows to an integral part of major cities' transportation systems. As bikeshare systems have thrived, so has curiosity about their precise role within those transportation systems, including their relationship to public transit.

A growing body of work has sought to address the extent to which bikeshare use tends to serve as a transit substitute, complement, or some combination of the two. This paper seeks to contribute to that exploration through the lens of Washington, D.C.'s Capital Bikeshare ("CaBi"), which was founded in 2010 and now comprises 678 docking stations and more than 5,000 bicycles across the District of Columbia, Maryland, and Virginia.

In this analysis, I apply methods developed by Blanchard and Waddell (2017) and Foti et al. (2012) to estimate travel times between the start and end coordinates of bikeshare trips along an integrated transit and pedestrian travel network that combines transit schedule and street network data. In a sample of 22,856 Capital Bikeshare trips taken on weekday mornings in November 2021, I find that up to 17% can be considered substitutes for transit – or potentially walking, in the case of some very short trips – based on the network analysis estimates.

I also attempt to measure transit complementarity, by identifying bikeshare trips that connect heavy rail stations to points outside their catchment areas. Although trips meeting the criteria I set out represent more than 7% of the sample, I find that their destinations tend to be downtown employment centers in their own right, and caution against relying on transit station adjacency alone as a determinant for what makes a bikeshare trip complementary to transit.

Only 271 trips, or 1.2% of the total, met the criteria to be classified as both substitutes and complements. The remaining 74.8% of trips neither had a transit alternative competitive on time, nor connected Metrorail stations to areas distant from them.

Methods

Below, I describe the methods used first for data collection and preparation, and then for the analysis itself. I performed the initial data processing and network analysis in Python, and performed additional analysis and data visualization in R.

Data collection and preparation

This paper relied on two categories of data: first, and primarily, the trip-level bikeshare usage data that formed the basis of the analysis; and second, the transit system and schedule data used to map the relationship between bikeshare trips and public transit.

Bikeshare trip data

Capital Bikeshare has published trip history data in General Bikeshare Feed Specification (GBFS) format on its website since the system's inception in 2010. For this project, I chose to limit the scope to data for November, 2021 – the most recent available – to ensure that analysis based on current transit schedules would be reasonably accurate. To simplify the analysis of bikeshare interactions with transit in a system operating across multiple jurisdictions, I restricted the scope to trips whose start or end coordinates placed them within the boundaries of the District of Columbia.

I excluded trips shorter than two minutes or longer than two hours, along with trips that started and ended at the same docking station. These cleaning steps may have excluded some legitimate trips, but were intended to screen out instances where a bike was checked out and immediately returned, or where a station did not correctly register a return.

The original intent for this project included comparing travel patterns between docked and dockless bicycle trips, as the trip history data nominally included start and end coordinates for both. Closer examination revealed that the latitude and longitude data for dockless trips was reported at a much lower spatial resolution than for the docked trips, presumably to protect rider privacy. I ultimately excluded dockless trips from the analysis due to the difficulty of comparing trip characteristics across different spatial scales.

Lastly, as described under "Analysis" below, the proof-of-concept transit-and-pedestrian network developed to estimate equivalent point-to-point travel times for this paper relied on transit service and headways for weekday mornings between 07:00 and 10:00 a.m., a time window covering a subset of 22,856 individual bikeshare trips. (Veteran's Day, Thanksgiving, and the day after Thanksgiving were excluded from the sample, due to modified transit operations on those dates.) For the sake of consistency, the rest of the analysis was limited to the same subset – acknowledging that a sample limited to peak commuting hours is unlikely to be wholly representative, and that future work would ideally represent a much more comprehensive range of mid-day, evening, overnight, and weekend trips.

Transit system and schedule data

To compute travel time estimates in order to identify trips with and without a reasonable transit equivalent, I relied on standardized schedule data published by District of Columbia transit authorities. I obtained current transit schedule data through API access to General Transit Feed Specification (GTFS) static feeds from the Washington Metropolitan Area Transit Authority (WMATA), which operates the Metrorail and Metrobus systems, along with GTFS feeds for the D.C. Circulator bus and D.C. Streetcar, which operate independently.

To classify bikeshare stations by their distance from transit, in order to identify potentially complementary trips, I relied on a DC Geographic Information System (DC GIS) point layer of regional Metrorail station entrances available through the District of Columbia open data portal.

Analysis

Background

To inform my analysis, I reviewed relevant literature from the last decade addressing the relationship between bikeshare use and public transit. The discussion in this paper focuses on two studies I found particularly interesting.

Song and Huang (2020) propose one of the more comprehensive frameworks for classifying Minneapolis bikeshare trips as complementary, competitive, or both, observing that these attributes are not necessarily exclusive. Their work draws on a combination of bikeshare trip records and detailed transit ridership records to identify competitive relationships by matching bikeshare trips with comparable transit trips, and complementary relationships by examining boarding and alighting patterns at nearby transit stops for each bikeshare docking station. The authors ultimately classify 3.55% of trips as both competitive and complementary, 2.56% as competitive alone, and 29.66% as complementary alone.

The authors state that the complementary designation is derived from passenger flow patterns within “subgraphs” of bikeshare stations connected to nearby transit stops. However, because the discussion also refers to complementary trips as those that provide first- or last-mile transit *access*, it is not entirely clear whether all trips with the *potential* to create a first- or last-mile transit connection are classified as complementary, or whether the high share of complementary trips reflects actual rider behavior.

Not all approaches to this question rely on trip-level data. Modeling the impact of Capital Bikeshare on Metrorail ridership in the Washington, D.C. metropolitan area between 2010 and 2015 using a combination of detailed rail ridership data and bikeshare station locations, Ma and Knaap (2019) found that the effect varied depending on the built environment surrounding stations, which they classified as either “core” or “peripheral.” The authors found that CaBi docking stations within a quarter-mile of a Metro station reduced rail ridership for “core” stations, concentrated in downtown D.C., but increased it for “peripheral” stations outside of downtown: in other words, bikeshare served as a substitute for rail in the core, but as a complement in the periphery. If this holds true, bikeshare may extend the reach of transit by serving as a “last-mile” option for areas below the density threshold required to support rail, but may draw riders away from rail in better-connected areas.

Identifying “transit substitute” trips

Like Song and Huang’s (2020), my approach to identifying potential transit-substitute trips relied on comparing a bikeshare trip’s recorded duration to the duration of the equivalent transit trip; however, the implementation details differed. In brief, their method identified likely candidate trips that started and ended near transit stops, matched those trips to real recorded transit trips with the same profile, and compared the durations of the matched trips. By contrast, I compared the recorded duration of each Capital Bikeshare trip to the shortest estimated travel time between the same start and end coordinates along a combined transit-pedestrian network that modeled the District of Columbia transit system as a set of nodes and edges connected to the city’s street network.

This analysis took advantage of the Python packages UrbanAccess and Pandana, whose theoretical and methodological underpinnings are described by Blanchard and Waddell (2017) and Foti et al. (2012), respectively. Though both packages were developed to study aggregate accessibility at a regional scale, they also include the capability to compute the shortest path between individual points. I used UrbanAccess to integrate the heavy rail, bus, and streetcar schedule data cited above with a pedestrian network drawn from OpenStreetMap, and used Pandana’s shortest path computation functionality to estimate travel times along the combined pedestrian-transit network. Travel time estimates were derived from a combination of transit schedules for travel times along transit segments, mean headways to represent wait time, and a 3 m.p.h. estimated pedestrian walking pace.

Fig. 1. Transit network (left) and pedestrian network (right).

Fig. 2. Integrated transit and pedestrian network.

The travel network used in this preliminary analysis was based on transit service and average headways for the weekday morning service peak between 07:00 and 10:00 a.m., a time window covering a subset of 22,856 individual bikeshare trips. In the future, the same method could be adapted to create multiple networks covering a more comprehensive service calendar and a more diverse selection of trips than in this subset, which likely oversampled commuting trips and undersampled trips taken for leisure, tourism, or other reasons.

Having generated the travel time estimates, I compared the results to the recorded bikeshare ride time to identify trips with a reasonable substitute. Determining where to set the threshold involved certain subjectivity and experimentation, discussed under “Findings” below.

To verify that the travel time estimates generated by the network were roughly accurate, I manually spot-checked a random sample of origin-destination pairs against Google Maps transit time estimates for the same trip at the same time of day. I found that UrbanAccess/Pandana estimates generally fell at the lower end of Google's transit estimates for the same trip, and more often underestimated than overestimated travel time compared to Google Maps. Though these schedule-based travel time estimates by definition do not account for traffic conditions, weather, mechanical breakdowns, and other causes of transit delays, I considered the network's accuracy sufficient to model a traveler's decision to use bikeshare over transit in the *best-case* scenario that the transit option is on time.

One caveat, however: during subsequent data analysis, I identified a handful of origin-destination clusters where network trip time estimates were several times larger than what they should be, for no immediately apparent reason. These errors were few enough in number not to invalidate the rest of the analysis, but serious enough in magnitude to require resolution before this method can be relied on further. My first guess is that a few missing or disconnected nodes or edges might be causing routing errors in specific parts of the network, but ruling this in or out will require closer examination of the paths in question than is feasible at this time.

Identifying "transit complement" trips

Without travel diaries, GPS tracking, integrated payment data, or any other way to follow traveler movements across modes, a definitive test for transit-complementary bikeshare trips is harder to define. As such, the criteria laid out here are by necessity circumstantial.

With that caveat, the clearest indication of a potentially complementary trip might be one connecting a transit station to a location poorly served by transit, or vice versa. To identify these trips, I computed the k-nearest neighbor distance from each Capital Bikeshare station to the nearest Metro station entrance in R, and classified each bikeshare station as adjacent to the Metro, far from the Metro, or neither.

At the far end, conventional planning wisdom appears to suggest half a mile as the distance most Americans are willing to walk to transit, and Guerra et al. (2012) found support for the "half-mile circle" as an appropriate population catchment area for transit ridership models. Accordingly, I counted bikeshare stations more than half a mile from a Metro station as "far" from transit for the purposes of this paper.

As a measure of proximity, Ma and Knaap (2019) used a quarter-mile threshold for their study of the effect of bikeshare stations on metropolitan D.C. transit ridership. However, I found that nearly 40% of D.C. bikeshare stations – and nearly all downtown stations – fell within a quarter-mile radius of at least one Metro entrance, and the closest bikeshare station to a given Metro entrance was usually no more than a tenth of a mile away. Likewise, Song and Huang (2020) set their threshold for spatial proximity at 100 meters, or just over 0.6 of a mile.

With this in mind, I set a relatively conservative maximum distance of one-tenth of a mile to designate a station as Metro-adjacent, in an attempt to distinguish between bikeshare stations simply convenient to transit and those most likely to actually facilitate a transfer between modes.

Distance from Metro	Stations
0 to 0.1 mile	56
0.1 to 0.25 mile	75
0.25 to 0.5 mile	91
More than 0.5 mile	113
Total	*335*

Table 1. Capital Bikeshare stations by distance from Metrorail station entrance.

Bikeshare stations by distance from Metro

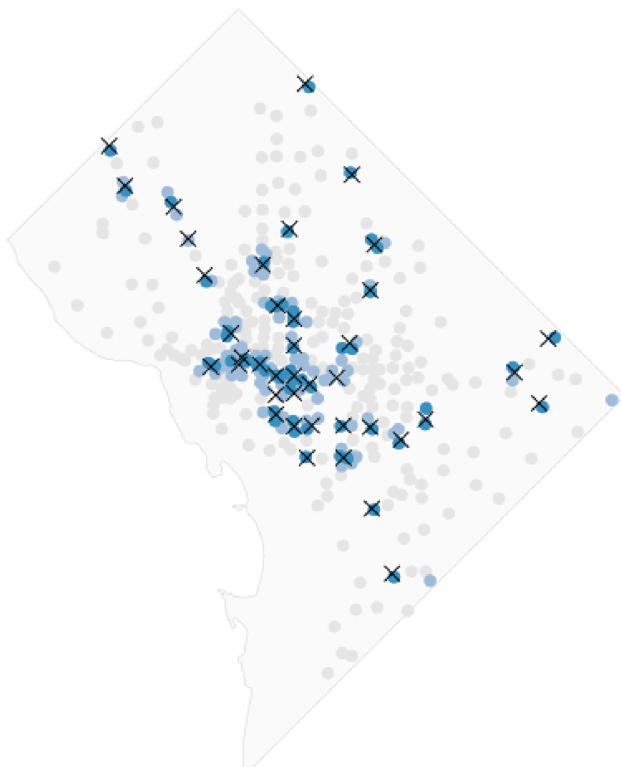

Fig. 3. Bikeshare stations by distance from Metro. Darker blue indicates tenth-mile distance; lighter blue indicates quarter-mile distance; crosses indicate Metro stations.

Alongside proximity to transit, Song and Huang (2020) included a short travel distance of no more than one mile, along with a short duration of no more than 10 minutes, as criteria for identifying a trip as potentially complementary. This is analytically sound if we limit our definition of complementary trips to those that solve a literal first- or last-mile problem, reasoning that a longer ride qualifies as an independent trip and not simply a connection to transit. However, given that 18 of D.C.'s bikeshare stations are located more than one mile from the nearest Metro station, and given that the nearest Metro station may not be on the relevant line, I chose not to use trip distance or duration as criteria for complementarity.

Findings

Bikeshare as transit substitute

To better understand the relationship between the bikeshare ride times and the travel time estimates generated by the transit-pedestrian network, I first examined distributions and summary statistics for both sets of values.

As the paired histograms below show, the estimated transit-walking times followed a roughly normal distribution, while the bikeshare trips skewed toward shorter average trip times with a substantially longer tail. The histogram of bikeshare ride times omits 117 trips longer than an hour, representing 0.51% of the total sample, which were not visible on the chart.

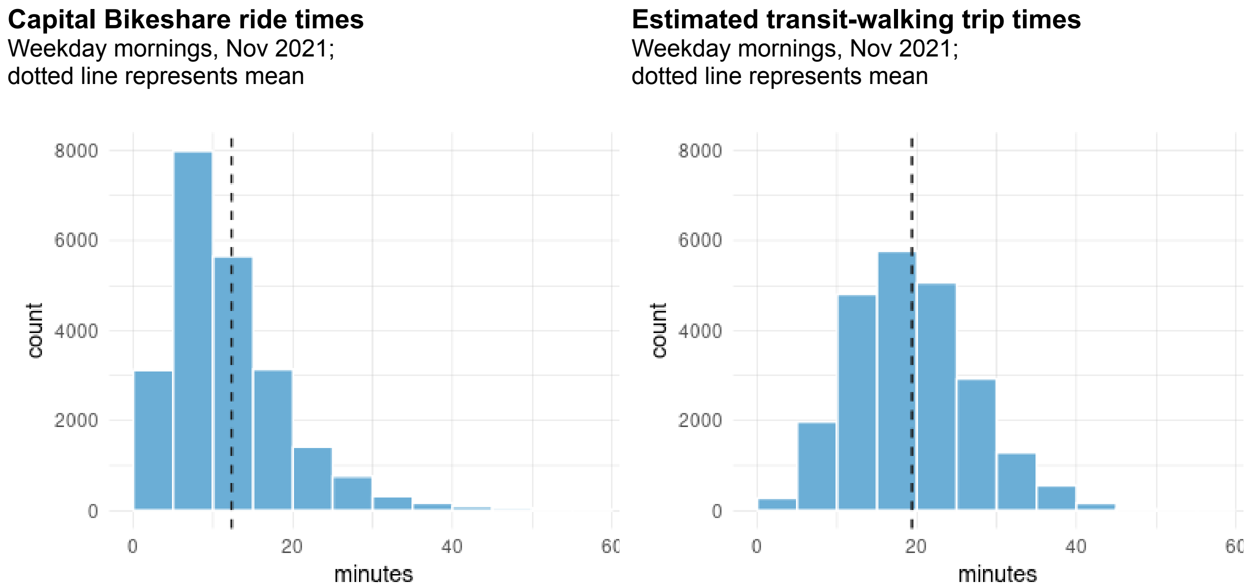

Fig. 4. Distributions of recorded ride times and estimated transit-walking trip times. Outlier ride times omitted for legibility.

The table below provides more detailed summary statistics. As the table shows, both actual ride times and estimated travel times varied widely, suggesting that an absolute measure of equivalency – for instance, defining a bikeshare trip as substituting for transit if the ride length differed from the transit estimate by less than five minutes – might not be appropriate.

	Bikeshare ride times HH:MM:SS	Estimated trip times HH:MM:SS
Min	00:02:01	00:00:00
Max	01:58:21	00:52:55
Mean	00:12:19	00:19:22
Median	00:10:13	00:18:57
Std. Dev.	00:09:04	00:07:38

Table 2. Summary statistics, recorded ride times and estimated transit-walking trip times.

The longest bikeshare trip was more than twice as long as the longest travel estimate, which may be indicative of biking for leisure and not just as transportation, or may simply be an artifact of a returned bike not immediately being recognized as such in the system. Simultaneously, mean and median bikeshare ride times were substantially shorter than the mean and median travel estimates. Though tentative, this might suggest that, rather than viewing bikeshare and transit as equivalent alternatives, Capital Bikeshare riders – or at least weekday morning commuters pressed for time – strategically choose bikeshare when they know it to be faster than their best transit or walking option.

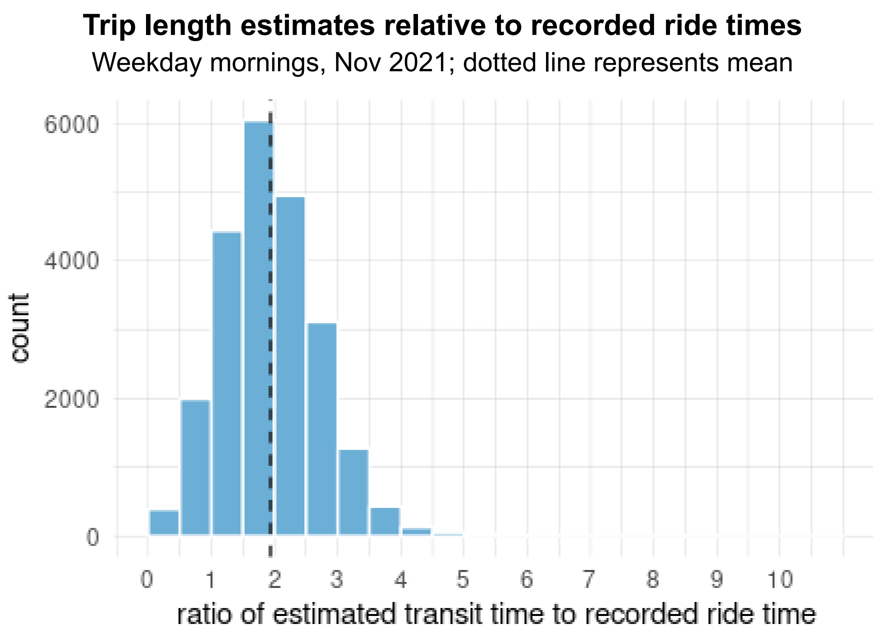

Fig. 5. Distribution of transit-pedestrian travel estimates relative to bikeshare ride lengths.

Above, the histogram sheds additional light on the relationship between recorded bikeshare trip times and estimated travel times between the same origin and destination along the transit-pedestrian network. The mean and median travel time estimates corresponded to 1.93 and 1.88 times the recorded ride times, respectively.

As a heuristic to determine whether a given bikeshare trip could be considered to have a viable transit substitute, I settled on a margin of error from 50% below to 25% above the recorded ride time. I included the lower bound as a simple screen for leisure rides where a faster transit trip to the same place would not be considered a substitute or equivalent.

Bikeshare ride time	Transit estimate considered reasonable substitute
10 min	5 min - 12.5 min
20 min	10 min - 25 min
30 min	15 min - 37.5 min
45 min	22.5 min - 56.25
1 h	30 min - 1 h 15 min

Table 3. Comparison of travel time ranges used to identify bikeshare trips with viable transit substitutes.

By this metric, 17.05% of trips in the weekday morning sample – representing 3,897 of 22,856 trips – could be considered substitutes for equivalent transit trips. This is considerably higher than the 6.12% identified by Song and Huang (2020), which could reflect methodological differences, D.C.'s extensive transit network by U.S. standards, or simply the fact that my analysis was limited to peak commuting hours when transit service is likewise at its peak.

Reviewing the results, at least some of the trips shorter than 10 minutes appear to involve traveling distances of a block or two, and the estimated travel times are more likely to represent walking than transit. Future work might involve more formal protocols to exclude these trips from the sample, but as a rough estimate, omitting them might bring the transit-substitute share of the total closer to 16%.

Bikeshare as transit complement

Of the 22,856 trips in the sample, only 1,695 – or 7.4% – met the criteria I established for a trip potentially complementary to the Metro. Of those 1,695, 69.8% (1,183 trips) took the rider from a bikeshare station far from Metro to a Metro-adjacent one, while only 30.2% (512 trips) were from a Metro-adjacent bikeshare station to one more than half a mile from the Metro.

Even before disaggregating further, this asymmetry – combined with what we know about weekday morning travel trends and the spatial distribution of Metro-adjacent bikeshare stations – suggests potential limitations of relying on the spatial relationship to transit as the sole determinant of whether a trip is likely to be complementary. Tabulating the results confirmed that the top 10 destinations represented in the “potentially complementary” subset, responsible for a combined 600 trips, were all in employment-dense downtown areas that also coincided with Metro station locations; most other bikeshare stations high on the list were also located downtown or close to it.

Of course, the fact that a transit-connected part of town is an employment destination in its own right does not inherently mean that *no* bikeshare rides there resulted in a transit connection. However, Occam’s razor suggests that many, if not most, of the weekday morning downtown-bound bikeshare riders probably do in fact work there, and that the false positive rate for this initial test is likely quite high.

Interestingly, even with a very high suspected false positive rate, a 7.4% rate of potentially complementary trips is far lower than the 29.66% identified by Song and Huang (2020). This is probably unsurprising: in addition to addressing a different city and transit system, their definition of a complementary trip included connections to all transit stops, including bus stops, and did not apply a minimum distance from transit on the non-transit-connected end. Still, the difference between the two results serves to underscore the difficulty of attempting to devise an appropriate transit complementarity measure absent any data on non-bikeshare travel behavior, and points to the need for additional research and validation.

It is also entirely possible that requiring complementary trips to connect to an area more than half a mile from transit is too restrictive, and that eliminating this standard would capture genuine complementary trips that are otherwise screened out.

Bikeshare as both substitute and complement

Only 271 trips, or 1.2% of the total, met the criteria set out above to be classified as both substitutes and complements. Consistent with the rest of the sample, 69.4% of these rides represented trips to Metro-adjacent bikeshare stations from areas not served by Metro, and only 30.6% represented the reverse.

Discussion

This analysis explored the use of open data and transportation network analysis to model the relationship between Washington, D.C.’s Capital Bikeshare system and the city’s transit network, with an emphasis on substitute and complementary trips. The findings built on previous literature addressing similar questions in D.C. and elsewhere, but may have raised more questions than answers regarding the analytical methods best suited to a comprehensive investigation of the role of bikeshare within a city’s transportation network. Below, I address limitations of the methods described here and suggest some opportunities for future work.

Model limitations

This paper served as a proof of concept for the network analysis methodology described in detail above, and as such, the scope of that analysis was necessarily limited. With the resolution of the sporadic network errors that resulted in inaccurate travel time estimates for a small subset of trips, the same method could be adapted to expand the analysis to daytime, evening, night, and weekend travel, as well as to the suburban jurisdictions that comprise the remainder of the Capital Bikeshare network.

In its identification of trips with a possible complementary relationship to transit, this paper relied wholly on proximity to heavy rail. To avoid unduly privileging rail over high-quality bus service, future work could also include bus stops on high-frequency lines; however, manually extracting such stops from the larger dataset of all D.C. bus stops was beyond the scope of this analysis.

Due to the density of Metro-adjacent bikeshare stations near downtown employment centers, the identification of complementary trips in this paper also suffered from a high false positive rate. One possible improvement may be to implement Ma and Knaap's (2019) distinction between core and peripheral transit stations, so as to focus on identifying first- and last-mile connections and reduce the outsize influence of downtown stations that serve as destinations at least as much as connecting points. If so, it may also be interesting to expand the study area to include the Maryland and Virginia suburbs within the Capital Bikeshare network, even if building a complete picture of transit service across the full set of jurisdictions would require a more complex data collection effort.

A more complete analysis of complementary trips would also account for transit hours of operation, and potentially also frequency of service. These temporal aspects were of lower concern for this analysis, which was limited to morning commuting hours when transit service is at its peak.

An aspect of the bikeshare-transit relationship not fully addressed by the substitute-complement framework, but which the analysis of estimated equivalent trips has the opportunity to address, is whether and to what extent bikeshare fills gaps within the existing transit network. A longer-term, more comprehensive study would also seek to obtain a data sharing agreement with Capital Bikeshare in order to also apply this analysis to dockless trips. This would be particularly interesting given the contentious politics surrounding dockless electric transportation, as well as the ever-vexing last-mile problem for neighborhoods poorly served by conventional transit.

Future work might also investigate more deeply what spatial patterns characterize bikeshare trips with varying transit connections. One such question might be whether trips along D.C.'s north-south axis differ from those along its east-west axis, which has tended to be far more poorly served by transit.

Conclusion

In this paper, I have demonstrated the utility of applying transit-pedestrian network analysis to explore the relationship of bikeshare to transit, as well as some of the pitfalls of attempting to measure transit complementarity based on proximity alone. Using the methodology described, I identified 17% of weekday morning Capital Bikeshare trips during November, 2021 as substitutes for transit, 7% as complementary to transit, and 1.2% as both. However, the figure for complementary trips is likely a substantial overestimate caused by the density of employment in the most transit-dense parts of downtown D.C.

Future work might build on the methods developed here to further systematize and improve the equivalent-trip estimation component; investigate other data sources and research methods to more rigorously examine transfers between modes; and further explore the spatial distribution of trips with differing relationships to transit. In addition, while the need to protect rider privacy presents challenges from a data collection standpoint, the expansion of dockless bikeshare also presents new opportunities to study the possible role of bikeshare in extending first- and last-mile transit connections to communities without the density to support traditional bikeshare docks.

Citations

Blanchard, Samuel D., and Paul Waddell. 2017. "UrbanAccess: Generalized Methodology for Measuring Regional Accessibility with an Integrated Pedestrian and Transit Network." *Transportation Research Record* 2653 (1): 35–44. https://doi.org/10.3141/2653-05. https://github.com/UDST/urbanaccess

Foti, Fletcher, Paul Waddell, and Dennis Luxen. 2012. "A Generalized Computational Framework for Accessibility: From the Pedestrian to the Metropolitan Scale." In *Proceedings of the 4th TRB Conference on Innovations in Travel Modeling. Transportation Research Board.* https://onlinepubs.trb.org/onlinepubs/conferences/2012/4thITM/Papers-A/0117-000062.pdf. https://github.com/UDST/pandana

Guerra, Erick, Robert Cervero, and Daniel Tischler. 2012. "Half-Mile Circle: Does It Best Represent Transit Station Catchments?" *Transportation Research Record* 2276 (1): 101–9. https://doi.org/10.3141/2276-12.

Ma, Ting, and Gerrit-Jan Knaap. 2019. "Estimating the Impacts of Capital Bikeshare on Metrorail Ridership in the Washington Metropolitan Area." *Transportation Research Record* 2673 (7): 371–79. https://doi.org/10.1177/0361198119849407.

Song, Ying, and Yuchuan Huang. 2020. "Investigating Complementary and Competitive Relationships between Bikeshare Service and Public Transit: A Spatial-Temporal Framework." *Transportation Research Record* 2674 (1): 260–71. https://doi.org/10.1177/0361198119899389.

Data sources

Capital Bikeshare. “Trip History Data.” Accessed December 8, 2021. https://www.capitalbikeshare.com/system-data

D.C. Circulator. “General Transit Feed Specification (GTFS).” Accessed December 18, 2021. https://www.dccirculator.com/developer-resources/

D.C. Streetcar. “General Transit Feed Specification (GTFS).” Accessed December 29, 2021. https://dcstreetcar.com/developer-resources/

Open Data D.C., Office of the Chief Technology Officer. “Metro Station Entrances (Regional).” Accessed December 29, 2021. https://opendata.dc.gov/datasets/DCGIS::metro-station-entrances-regional/about

Washington Metropolitan Area Transit Authority (WMATA). “Bus GTFS Static.” Accessed December 18, 2021. https://developer.wmata.com/docs/services/gtfs/operations/bus-gtfs-static

Washington Metropolitan Area Transit Authority (WMATA). “Bus GTFS Static.” Accessed December 18, 2021. https://developer.wmata.com/docs/services/gtfs/operations/5cdc5367acb52c9350f69753