METROSCORE: A RELATIVE MEASURE OF TRANSIT QUALITY

Arunav Gupta
Halıcıoğlu Data Science Institute
UCSD
arg002@ucsd.edu

ABSTRACT

In recent years, rideshare alternatives have significantly impacted public transit ridership in large American cities by offering speed and convenience. To address this issue, transit agencies require a measure of transit mode preference to determine when and where riders choose public transit over car-based options. This paper introduces Metroscore, an arcpy-powered API that computes a multidimensional preference statistic for transit agencies to evaluate their services in comparison to car-based transit. Metroscore allows transit planners to input their own networks or build one for their city and analyze the effects of potential changes on the transit system under various spatiotemporal constraints. The methodology is illustrated through case studies on three major cities: Cincinnati, San Diego, and New York. With Metroscore, transit agencies can make informed decisions regarding the development of transit services, respond to the growing popularity of rideshare options, and provide strong evidence to support transit expansion proposals. Metroscore is available as an open-source, pip-installable package, with opportunities for feedback and collaboration on future developments. This paper not only introduces the innovative tool, but also highlights its potential in improving transit quality and fostering sustainable urban mobility.

Contents

1 Introduction

Recent developments in rideshare alternatives and systems take valuable ridership away from public transit systems. Because these solutions utilize existing road networks, they can take advantage of high levels of road connectivity in most American cities, thereby offering riders a tantalizing advantage over public transit options: speed and convenience.

The result has been a precipitous drop in the use of both rail and bus-style public transit options in several large American cities since the introduction of rideshare options. When this happens, cities get caught in a vicious cycle — car-based options appear to serve the public more effectively, so road networks are prioritized while public transit systems are neglected. This results in an even larger convenience disbalance between the two modes, which results in even more funding inequality.

Given this drop in ridership, how do transit agencies respond? Agencies need a measure of transit mode *preference*, i.e when, where, and under what circumstances do riders prefer to take public transit over car-based modes of transit. Metroscore provides this in an easy-to-use arcpy-powered API. Transit planners can bring their own transit networks or build one for their city using Metroscore and then compute the metroscore statistic under various spatiotemporal constraints. The score is designed to answer the preference question stated above. In addition, the system supports running hypotheticals, in which transit networks are altered according to various proposals (new lines, new route schedules, etc.). When metroscore is rerun on the altered networks, transit planners can quickly estimate the effects of the planned changes and present them to city leadership in support of transit expansion projects.

2 Motivation & Background

Metrics exist today to quantify the quality of transit around a region [1]. However, these metrics fail to consider the decision making process that is behind ridership [2, 3]. With the advent of rideshare services like Uber and Lyft that pose a real threat to transit ridership, a new metric that evaluates transit quality relative to car-based connectivity is required.

Namely, I seek to answer the following questions:

1. Can we build a better scoring system for transit that takes into account how good the system is relative to driving?
2. Can the score calibrate itself well to known "native" measures of transit quality, such as revenue and ridership?
3. Can we "play around" with the designed score to see the expected effects of transit changes, such as decreased headways, new lines, etc?

The business case for a new transit score is overwhelming: real estate buyers and city planning commissions are the two primary target audiences. Those buying real estate already consider transit quality as a key buying decision, and making one that is up-to-date with the current state of "getting from A to B" is valuable. City planning commissions would also find this statistic useful as it does a much better job of predicting key success factors than existing metrics.

3 Methodology

With the goal of the score being to measure preference, it is loosely based on a probability — how likely is a trip at a given origin to a random destination to be feasibly traversed using public transit?

In order to estimate this, the calculation of the metroscore can be broken down into the algorithm that is shown in Algorithm 1. Note that metroscore appears to be a parametric function that depends on a point location, a time-of-day, and a trip duration. More information on this can be found in section 3.3. The ArcGIS API calls for multiple point locations and trip durations can be parallized; however the time-of-days must be computed in series. For the purposes of this methodology section, we will focus on a single metroscore measurement taken at a single point, time, and trip duration.

3.1 Service Area Generation

The first part of the algorithm requires that two service areas be generated, one using only public transit and walking, and one using driving. The driving service area is readily generated using the `generate_service_areas` function in the arcgis network analysis module. The public transit version, however is more complicated, and requires the building of a custom network dataset. By taking in the city's GTFS (General Transit Feed Specification) data, which contains

Algorithm 1 metroscore algorithm

Require: point p, time t, duration d
  function METROSCORE(p, t, d)
    $w = 2$ ▷ Weightage to give to transit bonus
    $D \leftarrow$ DRIVETIMESERVICEAREA(P, T, D) ▷ D is a polygon
    $T \leftarrow$ TRANSITTIMESERVICEAREA(P, T, D) ▷ T is a polygon
    $TDTC \leftarrow T \cap D$
    $TB \leftarrow T - D$
  return $(\text{AREA}(\text{TCTD}) + w \times \text{AREA}(\text{TB})) / \text{AREA}(\text{D})$
  end function

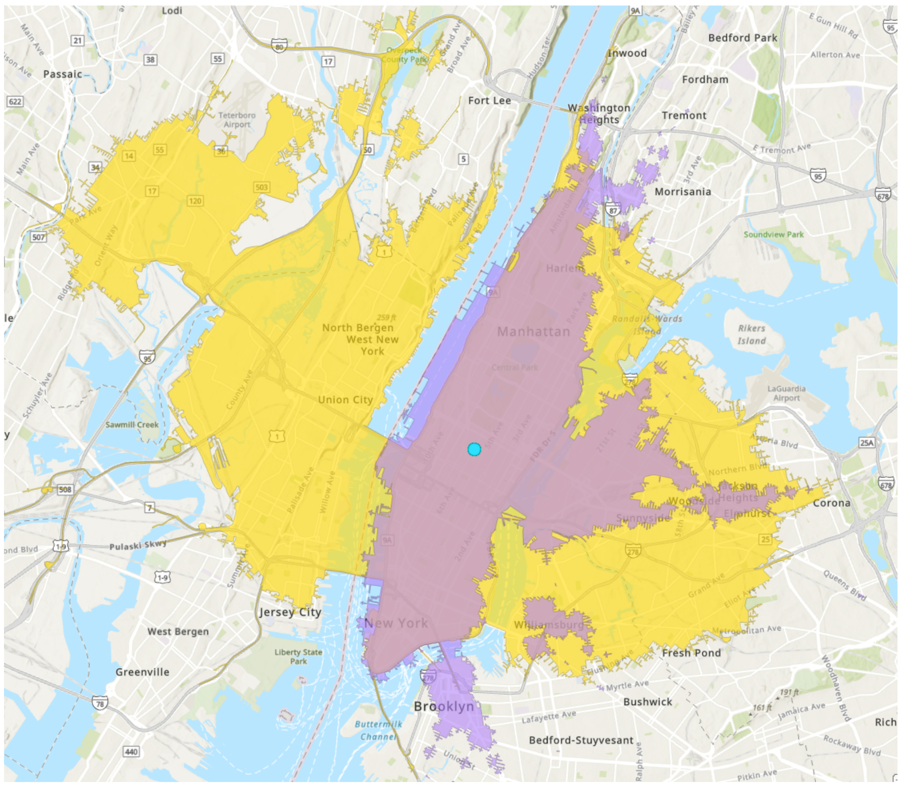

Figure 1: Comparison of 30-minute service areas for transit (purple) and driving (yellow) from central Manhattan at 4PM on Tuesday. Note the region in Brooklyn which is accessible by transit within 30 minutes, but not by driving.

information on transit routes, schedules, and stops in a standardized fashion, and the city's street data, we can build a public transit network dataset by following the instructions in [4]. Then, we can use the ServiceArea module in arcpy's network analyst module to compute the service areas. Figure 1 shows the computation of service areas in New York, NY.

3.2 Score calculation

Before comnputing the metroscore, we must create two new polygons from the transit and drive service area polygons:

1. **Transit Drive-Time Coverage (TDTC):** This is the intersection of the transit and drive-time polygons, and it represents the area in which transit can service a trip within the same time as driving. The TDTC can be reduced to the probability calculation:

$$P_{TDTC} = \frac{\text{AREA}(TDTC)}{\text{AREA}(D)} \tag{1}$$

Which approximates the probability that a random trip from the origin, given the utilized trip duration and trip time, would take the *same or less time* using driving and public transit. Note that $P_{TDTC} \in [0, 1]$.

2. **Transit Bonus (TB):** This is the difference between the transit and drive-time polygons (T - D), and it represents the area in which transit can service a trip faster than driving. Transit bonuses are valuable to transit networks because the represent pockets of unique accessibility, destinations to which riders are significantly more likely to use transit since it will get them there faster than driving/rideshare. The TB reduces to the probability:

$$P_{TCTD} = \frac{\text{AREA}(TB)}{\text{AREA}(D)} \tag{2}$$

   Which approximates the probability that a random trip from the origin would take *less time* via transit than via driving/rideshare. Note however, that this is not truly a probability metric, since TB can be larger than D (although this is rare. A true probability metric would divide by D', the minimum drive-time service area that encompasses TB. However calculating this is difficult, and it can be proven (via inequality) that $P_{TB}(T, D) > P_{TB}(T, D')$. Since we know that riders will prefer transit in cases where it is faster than driving much more frequently than in cases where transit and driving are the same speed, we are okay with this inequality (and will adjust for it accordingly below).

Combining the components computed in equations 1 and 2 allows us to generate the final metroscore:

$$\begin{aligned}\text{METROSCORE}(T, D) &= P_{TDTC}(T, D) + P_{TB}(T, D) \\ &= \frac{\text{AREA}(TDTC) + w \times \text{AREA}(TB)}{\text{AREA}(D)}\end{aligned}$$

Where w is a weightage given to the Transit bonus to amplify the difference in preference of taking transit when one's destination is inside a transit bonus. In the metroscore package, $w = 2$.

3.3 Multi-resolution analysis

Any given metroscore depends on three variables: location, time of day, and trip duration. Naively, users can fix all three variables when running calculations. However, the value of metroscore comes when users "slice", or take multiple measurements across a single dimension while fixing or averaging across the remaining two.

1. **Slice by location:** Agencies can test multiple locations within a city or county's limits in order to understand in which regions of a city is transit convenient and a legitimate option to take versus car-based options. This is especially valuable to agencies looking to expand their transit network, as it encourages them to surgically target regions in which rideshare/driving is predominately more convenient than existing transit options. All three case studies in section 4 contain this type of analysis, and users can generate testing points using either the `make_random_points` or `make_grid_points` functions in metroscore. The former produces randomly generated points within a polygon, and the latter produces a grid of evenly-spaced points within a polygon, which is useful when generating heatmaps.

2. **Slice by time of day:** By testing multiple times of day, users can learn about how time affects their transit quality. For example, to understand how rush hour affects how much one benefits from a bus stop, users can compare the metroscore during and outside of rush hour time. To test how transit hours affect preference, users can test times when transit is unavailable (2AM, for example), against times when they are available. To understand the effects of frequency/headways, users can test the metroscore across an hour (an example of this type of analysis can be found in section 4.2).

3. **Slice by trip duration:** Users can fix or select location and time of day, but test different trip durations to understand how the metroscore changes based on trip duration. Do riders prefer to take shorter or longer trips using transit? Answering this question can provide insight into the extent of a transit network, as well as its stop density. Examples of this analysis can be seen in sections 4.1, 4.2, and 4.3.

In addition to the single-variable slices described above, one could also slice by multiple dimensions. One example would be to slice across both location and time-of-day, resulting an a heatmap animation that shows how the metroscore across the entire region changes as the day progresses. However, such multidimensional slices require far more computational power than was available for this project, but represent a potential future avenue of exploration.

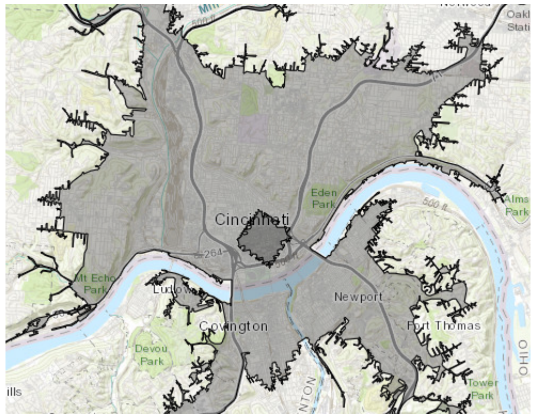

Figure 2: 10-minute drive time service area (larger polygon) vs. 10-minute transit time service area (smaller polygon) in Cincinnati, OH

4 Results

Metroscore was tested on three candidate cities: Cincinnati, OH, San Diego, CA, and New York, NY to validate its effectiveness. The cities were selected to represent a cross-section of transit quality in the United States, as well as due to their readily available and clean transit and road network datasets.

4.1 Case Study A: Cincinnati, OH

The city of Cincinnati (Data: [4]) supports two public transit agencies, which primarily run buses through downtown Cincinnati and its suburbs. SORTA (Southwestern Ohio Retional Transit Authority) and TANK (Transity Authority of Northern Kentucky) run the bus systems. Cincinnati also exists at the center of various interstates, including I-75, I-71, and I-471. These three interstates form an "H" pattern around downtown Cininnati, which gives them high levels of connectivity to Cininnati's suburbs. Figure 2 shows the 10-minute service areas from downtown Cincinnati — one can clearly see the advantage driving networks have over transit networks due to the interstate connectivity.

For this analysis, two slices were created: first along trip duration, and another by location. Figure 3 shows the metroscore as a function of trip duration — while transit preference is (relatively) high for short, 10-minute-long journeys, it drops sharply as the interstate advantage of driving begins to take over and make transit highly inconvenient for longer journeys. Figure 4 shows a heatmap of metroscores across the Cincinnati metro — one can clearly see how the metroscore is higher in the downtown, where more bus lines frequent and smaller in the suburbs.

4.2 Case Study B: San Diego, CA

San Diego (Data: [5]) is served by two transit agencies: the MTS (Metropolitan Transit System) and NCTD (North County Transit District). With slightly more transit support than Cincinnati, San Diego has not only buses (both MTS and NCTD), but also commuter rail (NCTD Coaster) and a 4-line trolley system (MTS). However, San Diego's driving environment relies heavily on a network of interstate, state, and local highways that connect the city.

For this study, 4 "slices" were created: two by time of day, one by trip duration, and fourth by location. Figure 5 shows the metroscore as a function of trip duration — this curve is remarkably different from the one in Cincinnati. The trip duration slice was produced from a single location on the UCSD campus in La Jolla, at 4PM on a Tuesday. At this time, rush hour traffic hampers transportation along I-5, the largest interstate highway in San Diego that forms the backbone of the interstate region. In this setting, the MTS trolley shines: relatively long (30-60) minute trips via light rail that bypass traffic become remarkably faster than driving by comparison. Still, shorter trips (which are mostly serviced by buses) appear to fall victim to rush hour congestion. It is estimated that when traffic subsides from the freeways, the curve will invert itself, as the the trolley no longer travels faster than the car and therefore loses its primary competitive advantage.

The time-of-day analysis supports this hypothesis, as shown in figure 6. 4PM appears to have the highest metroscore, when compared with 4AM, 10AM, and 10PM. However, the largest complaint against the trolley is its low frequency (every 15 minutes). How does this low frequency affect the metroscore? We can run another slice, this time running metroscore every 2 minutes from 4PM to 5PM. The results are shown in figure 7: the large spikes make it clear that the

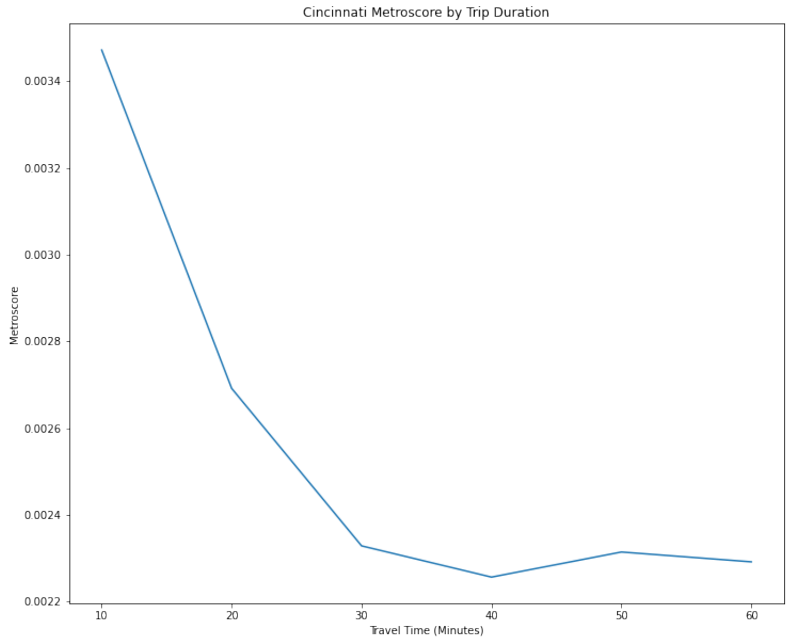

Figure 3: Cincinnati metroscore as a function of trip duration

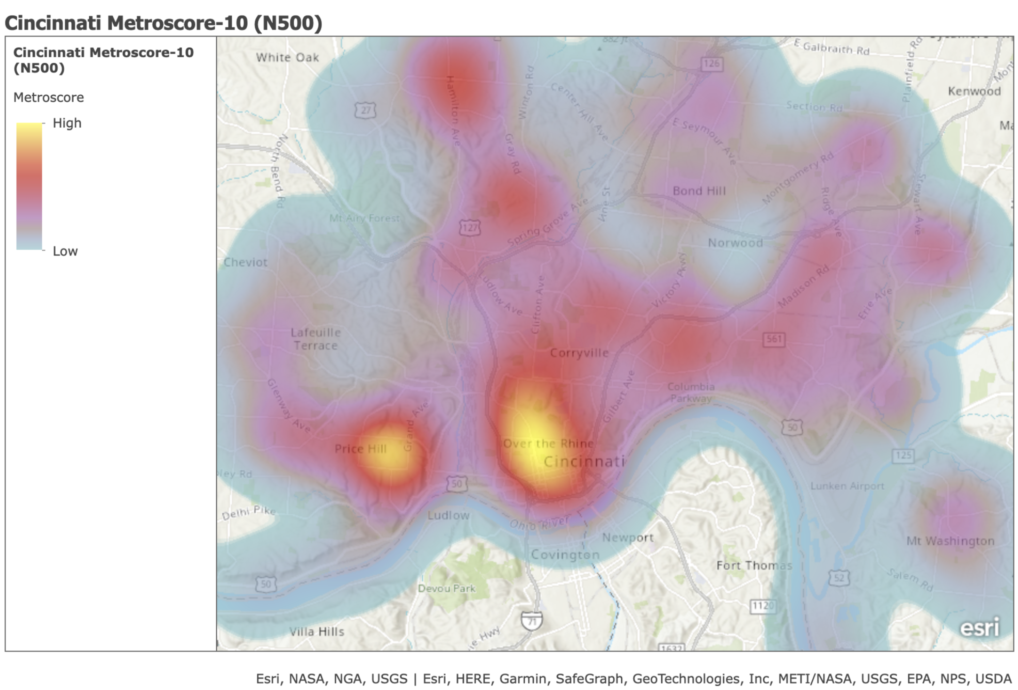

Figure 4: Cincinnati metroscore as a function of location. 60-minute trips, time agnostic, N=500.

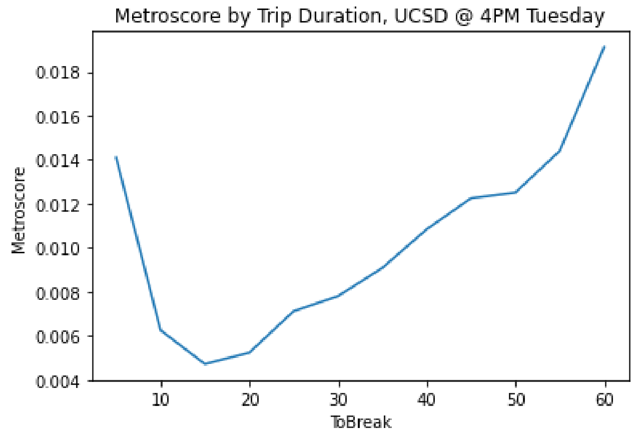

Figure 5: San Diego metroscore as a function of trip duration. Location: UCSD Campus, La Jolla, 4PM Tuesday.

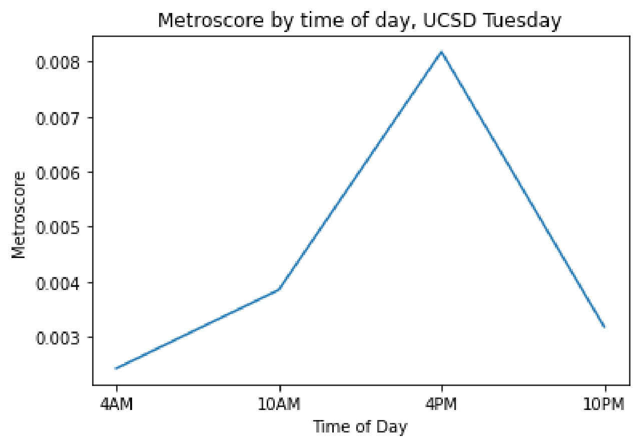

Figure 6: San Diego metroscore as a function of time of day, across a 24-hour period.

metroscore varies widely with whether or not the trolley is about to arrive. If one were to start their journey at 4:04PM, for example, it would take them roughly 6 minutes to reach the nearest trolley stop at UCSD. Since this stop only sees a southbound trolley 4 times every hour [6] (at 4:09PM, 4:24PM, 4:39PM, and 4:54PM) They would reach the station at 4:10PM and then have to wait 14 minutes until the next trolley comes at 4:24PM. Meanwhile someone who made it on the 4:09PM trolley would be able to travel an additional 14 minutes along the trolley, making significant headway into San Diego. This can be presented as evidence that the headways along all trolley lines should be dramatically improved.

Finally, a spatial analysis was conducted by slicing by location. Figure 8 shows these results, which indicate a hotspot near MCAS Miramar where a large bus depot exists. Aside from the Miramar outlier, metroscore improves slightly as one gets closer to downtown.

4.3 Case Study C: New York, NY

A final case study was conducted on New York, NY (Data: [7, 8, 9, 10]), due to its high levels of metro ridership and a transit system widely regarded as successful. Like San Diego, slices were taken along each of the three metroscore dimensions: time of day, trip duration, and location. Unless otherwise specified, tests were conducted for 30-minute trips at 4PM on Tuesday from the corner of 7th Ave and 57th St in Manhattan (Carnegie Hall). NYC is served primarily by the MTA, which manages buses and the subway system, as well as the Metro-North Rail Road (MNRR) and Long Island Rail Road (LIRR), which both operate commmuter rail around the tri-state area and Long Island. NJ Transit also runs commuter rail (PATH trains) through parts of NYC, but clean GTFS data was difficult to find for this agency and was thus ommitted from this analysis.

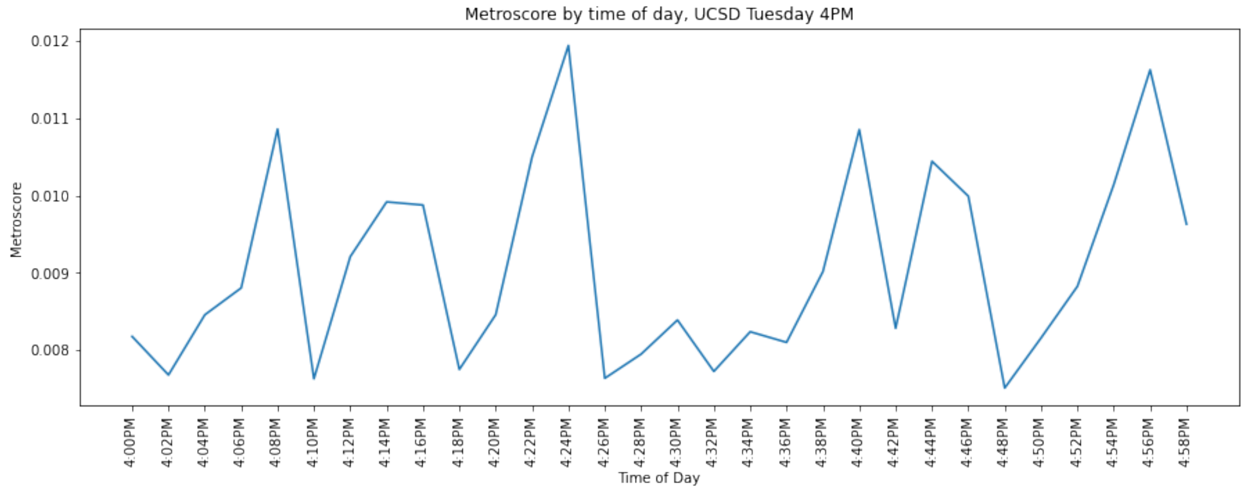

Figure 7: San Diego metroscore as a function of start time, from 4PM to 5PM (every two minutes).

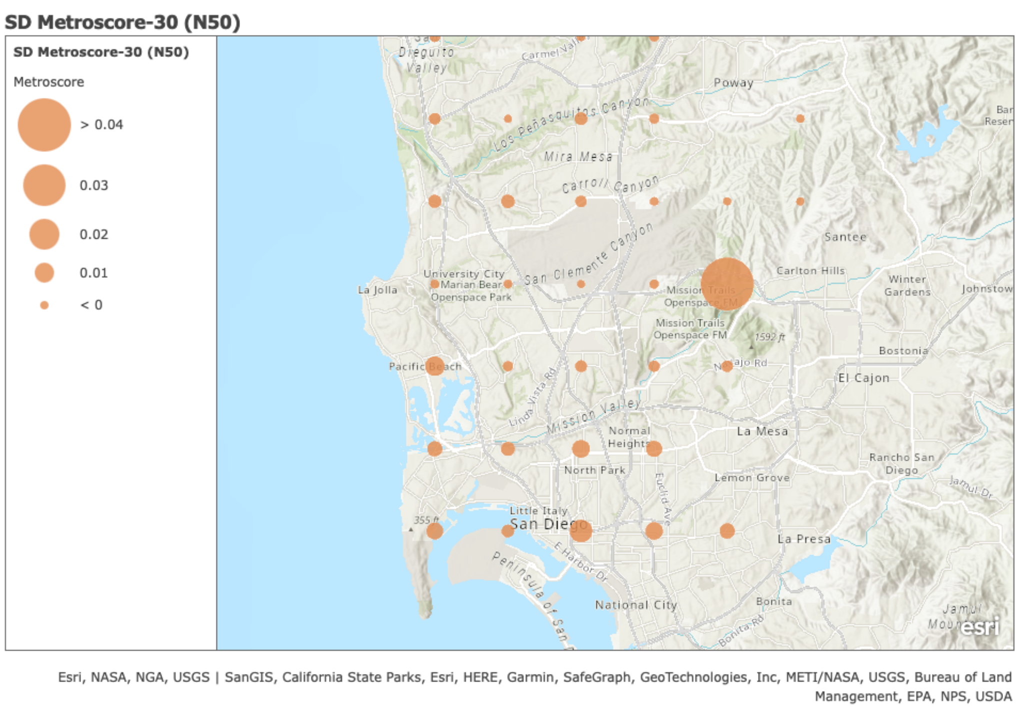

Figure 8: San Diego metroscore as a function of location. 30-minute trips, 4PM Tuesday, N=50.

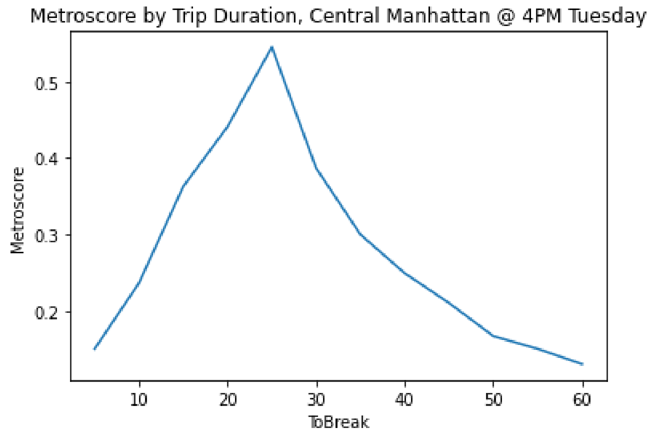

Figure 9: NYC metroscore as a function of trip duration.

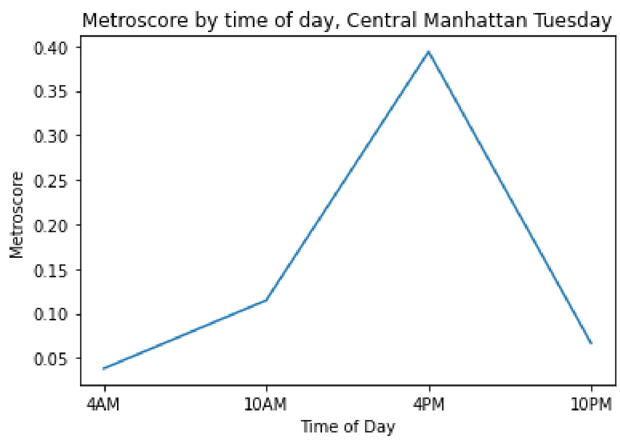

Figure 10: NYC metroscore as a function of time of day across a 24-hour period.

Figure 9 shows the result of the trip duration analysis. One can see a sharp rise in metroscore up to 30-40 minutes, and then a sharp decline afterwards. A possible explanation could be that it takes roughly 30 minutes to make it out of the Lincoln tunnel to New Jersey, after which one can access the plethora of interstate available in the state to reach land that cannot be accessed by any of NYC's transit systems. Before this tipping point, car users are confined to Manhattan's dense, slow street network, which makes the subway a significantly quicker option.

Figure 10 shows the time-of-day analysis, and there is little remarkable here: just like we expected in the case of San Diego, commute times around 4PM and 10AM have higher metroscores than off-peak hours of 4AM and 10PM. More interesting, however is the location analysis, seen in 11, which shows how central Manhattan and Upper East Side, which are all served by the subway, appear to have higher metroscores than Hell's Kitchen and Hudson Yards, which are not served by any subway lines.

5 Limitations

Given the computational and algorithmic requirements of the score's design, there are several limitations that represent future avenues of research and development:

1. **Computational Complexity:** the advantage of metroscore being multi-resolutional also means that calculating a city-wide metroscore comes at a high computational cost. For every sampled location (of which in a large MSA could be in the thousands), two service area polygons must be generated for each trip duration and time. For example, a high-resolution metroscore of a MSA might have the following parameters: 1,000 sampled points, trip durations of 5-60 mins (every 5 minutes), run at 6AM, 12PM, 6PM, and 12AM. Across these four

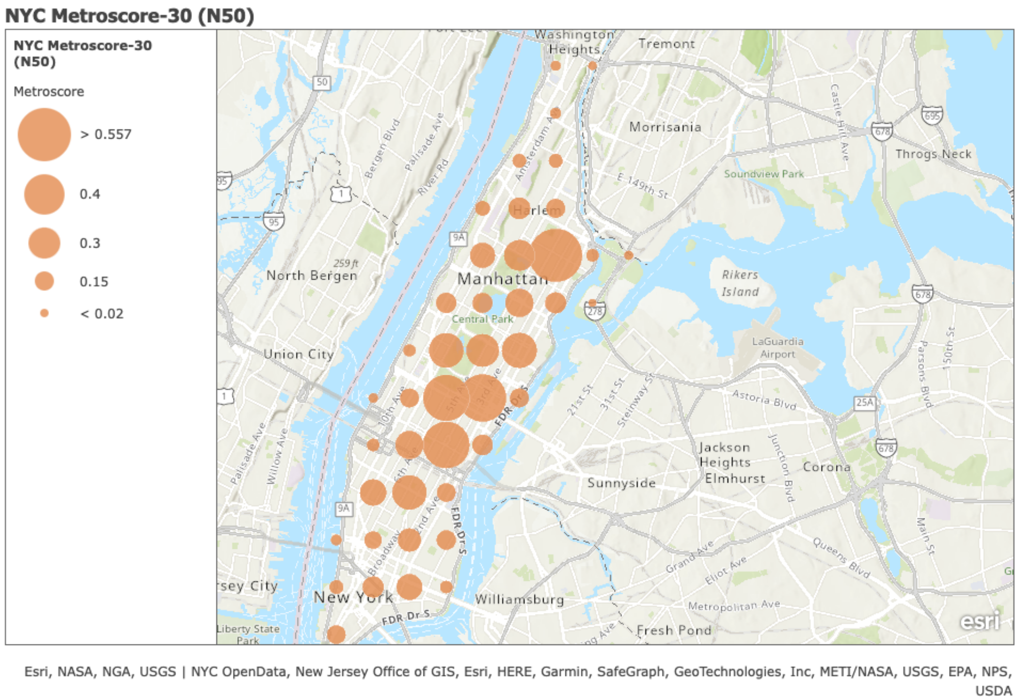

Figure 11: NYC metroscore as a function of location.

variables, there would be 96,000 service areas generated! This imposes a significant time and resource (since drive-time areas use ArcGIS credits in metroscore v0.1) cost which should be alleviated.

2. **Boundary Conditions:** While the metro-time network is provided (or created) by the user, the drive-time network is pulled from ArcGIS, as the public model has access to updates streets and granuar traffic data. Creating this on the user's end would be exceedingly complicated. However, this difference in datasets means that MSAs which end at network boundaries (i.e. New York, Cincinnati) can have drive-time service areas which stretch beyond the extent of the transit network. This is the case in New York, for example, since the NYC MTA, LIRR, and Metro-North Railroad do not contain data about NJ Transit systems. Thus drive-time service areas extend into New Jersey, while metro-time service areas are restricted at the NJ/NY border. This issue can be alleviated through careful network design that considers all transit providers which surround an MSA.

3. **Multi-leg Journeys:** The ArcGIS Pro Network Analyst module currently does not support computing transit service areas that rely on multiple transit connections (i.e. bus, then tram, then bus). Thus, metroscore (as of v.0.1) does not support this mode of transportation. A fix can be implemented however, since the transit networks are graph networks; by squaring or cubing the OD-cost matrix we can approximate the cost of multi-leg journeys and add those results to our solver.

4. **Non-traversable Areas:** When testing metroscore I noticed that several drive-time service areas encompassed bodies of water, parks, and non-traversable geographic features. This could cause scores to be poorly calibrated, as it would be impossible for these areas to be the destination of a potential journey. A quick fix to this problem (that will be released in future metroscore versions) would be to filter the service areas by the intersection of the 10m buffers of the street network.

5. **Calibration:** At the moment, raw metroscores are not readable: values like 0.003 are meaningless to the average consumer. Given that real estate consumers are a viable target audience for metroscores (i.e. home listings include metroscore information alongside walkability, schoools, etc.). Once we have a dataset of metroscores for a variety of locations/transit environments, the "raw" metroscores can be scaled and adjusted into a 0-100 scale, making them readable to the average consumer.

6 Conclusion

Metroscore represents a viable option for transit agencies to evaluate the health of their transit services in comparison to car-based modes of transit. Given that car-based modes of transit are the biggest competition for most public transit systems, it follows that such a relative metric is valuable to these agencies. While there are several existing transit evaluation metrics, metroscore is grounded in principles of probability and is multidimensional by design, giving agencies and planners a way to analyze each metroscore result from a variety of different angles.

This technical report introduces the metroscore methodology and provides case studies on three major cities in the United States: Cincinnati, San Diego, and New York. It also outlines avenues of further research into service-area-based measures of transit quality for the future. Armed with insights from metroscore and future related metrics, transit agencies and planners have the ability to make informed decisions about where, when, and how to develop transit, respond to the incursion of car-based transit options such as rideshare, and provide city councils with strong evidence to support their proposals for transit expansion.

Metroscore is available as a pip-installable package for others to use, and is available open-source on GitHub [11] for others to leave feedback, feature requests, and suggestions.

Acknowledgments

Experiments described in this technical report were done within the DSC 170: Spatial Data Science and Applications course The author is immensely grateful to Professor Ilya Zaslavsky and Harshil Jain for their guidance during the project.

References

[1] Transit score methodology.
URL `https://www.walkscore.com/transit-score-methodology.shtml`

[2] Case study: Analyzing light rail station area performance in phoenix.
URL `https://cdn2.walk.sc/2/pdf/WS-Phoenix-TOD.pdf`

[3] S. Bokhari, How much is a point of walk score worth?, Redfin.
URL `https://www.redfin.com/news/how-much-is-a-point-of-walk-score-worth/`

[4] Create and use a network dataset with public transit data, ArcGIS Pro.
URL `https://pro.arcgis.com/en/pro-app/latest/help/analysis/networks/create-and-use-a-network-dataset-with-public-transit-data.htm`

[5] Sangis data warehouse.
URL `https://gis-sangis1.hub.arcgis.com/pages/download-data`

[6] San diego trolley timetable.
URL `https://www.sdmts.com/sites/default/files/attachments/trolley-timetable-sept2022.pdf`

[7] Nyc mta subway gtfs.
URL `https://transitfeeds.com/p/mta/79`

[8] Nyc mta bus gtfs.
URL `https://transitfeeds.com/p/mta/85`

[9] Metro-north railroad gtfs.
URL `https://transitfeeds.com/p/mta/87`

[10] Long island rail road gtfs.
URL `https://transitfeeds.com/p/mta/86`

[11] A. Gupta, Metroscore.
URL `https://github.com/agupta01/metroscore`