Accessibility and Centrality Based Estimation of Pedestrian Activity and Safety in Urban Areas

A THESIS
SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL
OF THE UNIVERSITY OF MINNESOTA
BY

Brendan Alan Murphy

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
MASTER OF SCIENCE

DAVID LEVINSON

October, 2015

Acknowledgements

I first would like to thank my advisor, David Levinson, for continually providing thoughtful critique and guidance regarding this work. His expertise in the fields of transportation and economic analysis research have only improved my own investigation skills. I would also like to specifically thank Andrew Owen, without whose knowledge in spatial data analysis, programming, and database management this project would have been made exceedingly difficult. To the rest of the University of Minnesota Accessibility Observatory staff and Research Assistants I also give my sincere gratitude, for wrestling and wrangling data alongside me, and asking good questions of me.

I never would have proceeded down this particular path of transportation research had my interests not been piqued by conversations with David Levinson, Greg Lindsey, Frank Douma, Simon Blenski, Tony Hull, Prescott Morrill, and many others, well before I joined the University of Minnesota CEGE department. Their willingness to discuss urban transportation science with a curious cyclist, and then-prospective engineering graduate student, is greatly appreciated.

I would also like to acknowledge the support and contributions of the University of Minnesota Center for Transportation Studies and the Roadway Safety Institute, through which this research was made possible.

Finally, I give heartfelt thanks to those in my life who have always asked difficult and inspiring questions of me, challenged my assumptions and narratives with evidence and counterpoint, and never wavered in their support of my chosen paths, be them on- or off-road. Mom, Allison, Andrew, Esther, Chelsea, Miri, and so many others - thank you.

Dedication

I dedicate this thesis to those who have inspired my curiosity for the interconnected world around us, ignited my passion for critical inquiry and compassion, and to all who have bicycled by my side.

"Get a bicycle. You will certainly not regret it, if you live."
-Mark Twain

Abstract

The following thesis investigates the feasibility of using metrics of accessibility to jobs, betweenness centrality, and automobile traffic levels to estimate pedestrian behavior levels and automobile-pedestrian collision risks within an urban area. Multimodal count and crash report data from Minneapolis, Minnesota are used as a test of this scalable, translatable modeling framework; multiple stepwise linear regression is performed to compile a set of explanatory variables from which to construct a predictive model of pedestrian movement. The existence of the Safety In Numbers (SIN) phenomenon is investigated within both the raw and estimated pedestrian movement data; the SIN effect is the phenomenon where pedestrians are found to be safer from collisions, on average, when there are more pedestrians present in a given intersection, street, or area - that is, that the per-pedestrian risk of injury inflicted by drivers of automobiles decreases as a function of the increasing volume of pedestrian traffic. Economic accessibility, betweenness centrality, and Average Annual Daily Traffic (AADT) were found to be significant predictors of pedestrian traffic at intersections in Minneapolis, and the SIN effect was observed in both the raw and estimated pedestrian movement data when combined with the aggregated crash data. This investigation shows the potential utility of such a model that is both scalable to larger geographic areas, and translatable to varying jurisdictions due to its reliance on nationally-available datasets. Policy implications and concerns surrounding use of the Safety In Numbers effect in planning and engineering, and issues of data quality and availability in urban geographic science, are discussed.

Contents

List of Tables

List of Figures

Chapter 1

Background

Walking and bicycling are increasingly becoming important transportation modes in modern cities, and for a multitude of reasons, including individual and societal wellness, environmental externalities associated with motorized modes, and resource availability. Planning for biking and walking, and creating societal programs to increase their levels, has been cited as a targeted health need in urban planning going forward [2], [3], [4]. Resource limitations, particularly in high-population and developing third-world countries, impose constraints on the maximum level of personal motorized travel allowed, and as a result there is a greater need for viable alternatives. However, safety levels associated with these modes continue to be a problem, with 1.24 million vulnerable road users (VRUs) being killed in on-road accidents in 2010, with another 20-50 million injured globally. Further, a full 22% of traffic deaths worldwide are pedestrians, which is quite a high figure considering the transportation mode of walking harbors little danger unto itself [5].

Non-motorized transportation tends to be some degree of unsafe in most average developed urban areas where cars are ubiquitous, except where specific programs and treatments have been employed to address the safety concerns, such as in Copenhagen, Denmark [6]. Rates of walking and bicycling to work in the United States hover around 2.8% and 0.6%, respectively, with public transit use barely higher at 5% nationally [7]. Proper placement of pedestrian treatments and improvements has implications to both safety [8] and accessibility and mode choice [9], but proper information regarding estimated non-motorized traffic levels is needed to locate areas in need of improvement. In determining salient locations for non-motorized improvements, it is important to have accurate records of both existent and potential travel demand (e.g. current levels of walking in a neighborhood, as well as good models of increased demand due to potential treatments); however good quality, high-granularity datasets for non-motorized travel can be difficult to obtain, especially standardized for national spatial inventories [10]. For this reason, practitioners and researchers must frequently rely on estimation models for non-motorized traffic, and various methods can suffer from issues of data quality, granularity, and the presence of location-specific variables [11].

Many of the issues with the collection of standardized non-motorized transportation data have to do with the factors that influence pedestrian and bicycle behavior. A model of active transport risk assessment is uninformative if the pedestrian and vehicular flows do not accurately represent corresponding levels *in situ*, and many cities do not have dense data sets of active transport flow levels, instead favoring counts of vehicle traffic. As such, active transport flow levels must be extrapolated from sparse data sets using comprehensive methodologies. Land use data are well-documented by the U.S. Census Bureau to the Census Block level of resolution, and general socioeconomic characteristics are maintained as well, and can have significant influence [12]. However, more specific socioeconomic characteristics are salient in non-motorized travel beyond just adjusted income levels. Weather variables [13], [14] and latent, subjective variables such as visibility and perceptions of lighting may be relevant, which can be more difficult to obtain at high spatial resolution [15], and can complicate inter-city comparisons. For these reasons, as well as the overall lack in non-motorized travel counts for many communities, methods of estimating pedestrian and bicycle behavior that do not rely heavily on high-resolution count data are applied in this study.

The aims of this thesis are twofold: to investigate the viability of economic accessibility and network betweenness centrality metrics in estimating urban pedestrian movements, and to further the understanding of the Safety In Numbers phenomenon and its dependence on both pedestrian and vehicular flow levels. Safety in Numbers (SIN) refers to the phenomenon that pedestrians as road users become safer, on average, when there are more pedestrians present in a given locale or area, e.g. that the per-pedestrian risk of injurious interaction with motorized vehicles decreases as a function of the increasing volume of pedestrian traffic. The emergence of the SIN phenomenon is well-supported by pedestrian crash data across a number of studies in various urban environments and reviews [16], [17], [18], [19]. The concept has seen relatively widespread adoption in urban planning schools of thought, though its temporal causality is not clear-cut [18], and it is commonly discussed only in the context of pedestrian risk depending on pedestrian flow levels. Additionally, a critical mass of research regarding psychological aspects of driver behavior pursuant to the SIN effect has not yet been reached, and the potential behavioral and social factors behind SIN do warrant more attention [19]. The USDOT Strategic Plan for Fiscal Years 2012-2016 aims to reduce non-vehicle-occupant fatalities to 0.15 per 100 million vehicle-miles-traveled (VMT) by 2016. However, such a goal does not account for risk dependence on pedestrian flow levels, i.e. the guidelines ignore the SIN effect.

Aggregate travel behavior studies typically involve analysis at the level of Transportation Analysis Zones (TAZs), which are too coarse to allow robust analysis of non-motorized travel [12], [9]; Regional Travel Surveys consider many trip purposes, but are similarly coarse, and typically have too small of sample sizes to allow for robust city-to-city comparison. There is strong support for significant correlations between environmental factors such as intersection density, network connectivity, land use density, and other urban form variables, and local mode-share of walking [20], [21]. [22] reviewed 1990 travel diary and land use data from the U.S. Census, and found independent and significant correlations between each of density, diversity, and design urban form metrics and Vehicle-Miles Traveled (VMT), trip-making, and mode-choice; in combination, these three factors were found to positively influence non-motorized travel. In an analysis of an array of neighborhoods in Europe and Asia, [23] reported that more walkable neighborhoods led to an average of 766 additional steps taken per day for adults wearing step monitors, with walkability defined as a meta-factor influenced by block size, intersection density, land use density, and other variables. This meta-definition of walkability estimation is also supported by [24], where network connectivity, density, and land use metrics were grouped.

Census block-level information regarding economic accessibility (access to jobs) via both strictly walking, and via the net accessibility benefit of public transportation, will first be used to explain observed pedestrian traffic at a subset of intersections in the city of Minneapolis, Minnesota, using Ordinary Least Squares (OLS) regression. Road network betweenness centrality will also be used as an explanatory variable, as a proxy of the underlying network structure. A framework for comprehensive pedestrian risk assessment modeling, using pedestrian volume, vehicle volume, and an environmental factor (crosswalk length) on a university campus is provided by [8]. The motivation for constructing models of pedestrian and vehicular traffic is in supplementing the sparse data currently available, in order to estimate pedestrian risk-burdens of collisions at every intersection in Minneapolis. Pedestrian risk-burdens - the risk of an individual pedestrian being struck by a vehicle - are calculated and compared for both the raw available data, and the estimated pedestrian and vehicular activity levels. This process allows us to construct a more complete spatial picture of how pedestrian collision risk varies throughout an urban area at the level of individual intersections, based on non-location-specific available data, and to assess the existence of the SIN effect using nationally-available data in conjunction with local count and crash data.

Chapter 2

Methodology

2.1 Data

This section briefly describes the data sources used in the pedestrian activity estimation models, and the data preparation process.

- **Data Sources**

    1. U.S. Census TIGER 2010 datasets: blocks, core-based statistical area (CBSA) for Minneapolis-St. Paul
    2. U.S. Census Longitudinal Employer-Household Dynamics (LEHD) 2011 Origin-Destination Employment Statistics (LODES)
    3. OpenStreetMap (OSM) North America extract, retrieved April 2014
    4. Turning movement counts (TMC) 2000-2013, City of Minneapolis
    5. Average Annual Daily Traffic (AADT) measurements 2000-2013, City of Minneapolis
    6. Traffic crash records 2000-2013, City of Minneapolis
    7. GTFS data from Metro Transit

- **Data Preparation**

    1. Construct pedestrian travel network graph for Minneapolis

2. Geocode pedestrian Turning Movement Count (TMC), Average Annual Daily Traffic (AADT), and crash data to spatial locations

- **Accessibility and Centrality Calculation**

  1. For each Census block in Minneapolis, calculate travel time to all other blocks within a 5 km radius for a single departure time
  2. Calculate cumulative opportunity accessibility to jobs for each census block, using thresholds of 5, 10, . . . , 30 minutes
  3. Calculate net transit accessibility benefit using a threshold of 30 minutes
  4. Calculate betweenness centrality for the Minneapolis OSM road network

- **Model estimation**

  1. Construct linear regression of pedestrian behavior on walking accessibility, AADT, net transit accessibility, network centrality, and accessibility to job opportunities by sector
  2. Assess and validate model on sample of other intersections in Minneapolis
  3. Calculate estimated pedestrian risk-burden

2.2 Software

Here the various software used is enumerated.

- **Software**

  1. QGIS, PostGIS, and PostgreSQL for GIS database work
  2. OpenTripPlanner open-source routing software for accessibility calculations (using Dijkstra's Algorithm)
  3. ArcMap GIS with Urban Network Analysis Tools toolbox for network centrality measures
  4. PostgreSQL, Python, R, and MATLAB for statistical work
  5. TileMill and ProjectMill for mapping and automation

6. Python for general scripting, automation, and data processing

Intersection locations were determined from OSM road centerline data for the Minneapolis-St. Paul CBSA (Core-Based Statistical Area). The subset of intersections for which count data were available is displayed in Figure 2.1; these intersections were used to construct the estimated models. Accessibility calculations were performed using OpenTripPlanner (OTP) open-source routing software; GIS work performed in QGIS, PostgreSQL, and PostGIS; network centrality measures computed in ArcMap GIS with the Urban Network Analysis Tools toolbox; statistical work done in SQL, Python, MATLAB, and R; mapping and imaging performed in TileMill. Figure 2.2 displays the locations of intersections in Minneapolis with pedestrian-automobile crashes used to estimate pedestrian activity and safety.

2.3 Accessibility

The first type of explanatory variable used in the model of Minneapolis pedestrian count data is cumulative opportunity accessibility. Using OTP, walking travel times along the network are calculated from each Census block centroid in Minneapolis, to each other block centroid within the travel-time thresholds of 10, 20, ..., 60 minutes. Job opportunities are summed from each block centroid reachable within a given time threshold, yielding an X-minute accessibility measure. Job opportunities are broken down by economic sector, as defined by the North American Industry Classification System. There are two accessibility calculations used in this study:

1. Accessibility to jobs from Census block centroids by walking
2. Accessibility to jobs from Census block centroids by transit and walking

Pedestrian counts are often taken at intersections in either gross counts, or divided by turning movement type. This study uses Turning Movement Count (TMC) data from approximately 750 intersections in Minneapolis; intersection counts were calculated by adding the various TMC types for each intersection in the analysis group, to yield a gross figure of pedestrian activity within an intersection. Two-hour counts for pedestrian activity were used for morning peak (7-9AM), midday (11am-1pm), and evening peak (4-6PM).

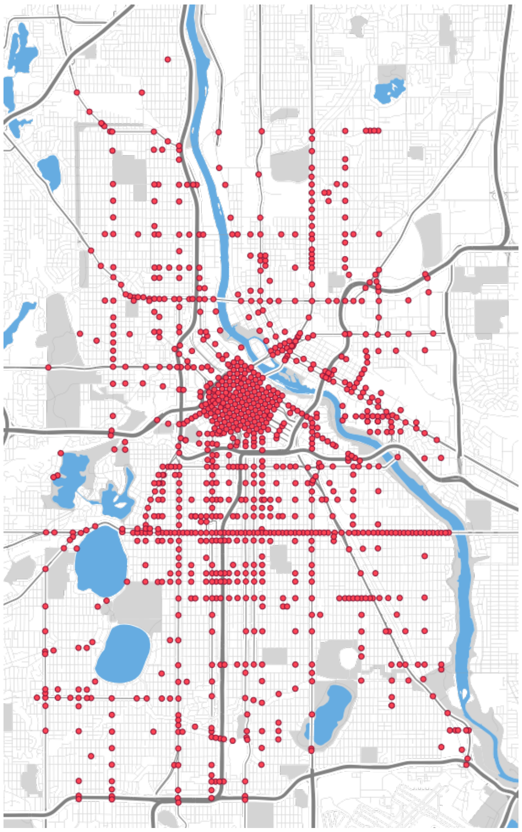

Figure 2.1: Locations of intersections in Minneapolis with raw pedestrian count data.

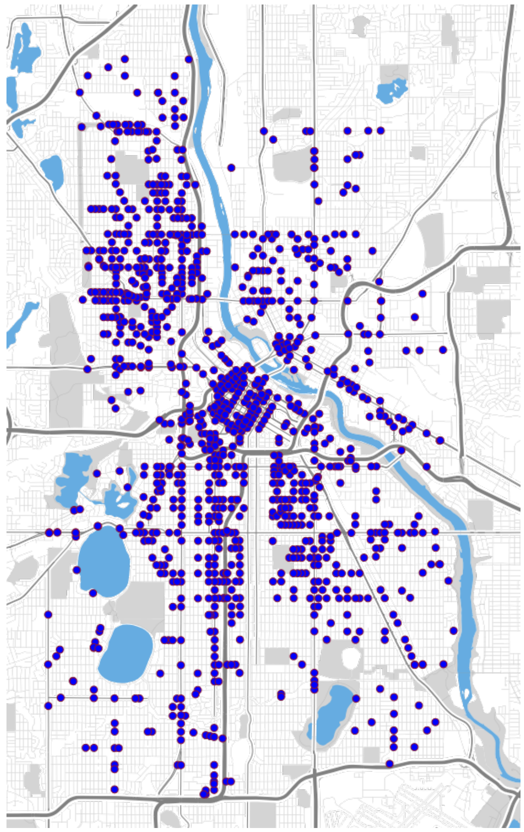

Figure 2.2: Locations of intersections in Minneapolis with pedestrian-automobile crashes; used in estimated pedestrian activity and safety analysis.

Accessibility calculations were performed using the following formulation of a cumulative opportunities model:

$$A_i = \sum_j O_j f(C_{ij}) \tag{2.1}$$

$$A_i = \text{accessibility for location } i \tag{2.2}$$

$$O_j = \text{number of opportunities at location } j \tag{2.3}$$

$$C_{ij} = \text{time cost of travel from } i \text{ to } j \tag{2.4}$$

$$f(C_{ij}) = \text{weighting function} \tag{2.5}$$

$$\tag{2.6}$$

The choice of weighting function has a large impact on the resulting Accessibility calculations; however, one of the simplest interpretations of cumulative opportunities is an integer count, using the following weighting function:

$$f(C_{ij}) = \begin{cases} 1 & \text{if } C_{ij} \leq t \\ 0 & \text{if } C_{ij} > t \end{cases} \tag{2.7}$$

$$t = \text{travel time threshold}$$

This intuitively makes sense when applied to opportunities such as jobs, number of restaurants, transit route departures, and other discrete integer variables in the surrounding environment. Cumulative opportunity models have been implemented historically since the 1970s for rote accessibility analysis [25] [26] [27] [28]. More current applications employ similar cumulative opportunity models, such as transportation and land use interactions in the Netherlands [29], residential land values [30] [31], regional accessibility [32], and analysis of accessibility on an individual-person, point-based network scale [33]. We hypothesize that origins exhibiting higher accessibility values would see greater pedestrian activity throughout the day. Accessibility for both walking, and walking + transit modes, are used in the estimation models; subtracting walking accessibility from the multimodal walking + transit accessibility yields the net transit benefit, and including walking and net transit separately in the regression models allows for explicit evaluation of how important transit benefits are to influencing pedestrian activity. Multiple regression in R statistical package was then performed to determine the explanatory power of the accessibility measures in estimating pedestrian and vehicular traffic in the AM, midday, PM peaks, as well as for a 6-hour summed count. These additional tables are reported in the Appendix. It was expected that origins exhibiting higher walking-accessibility values, and higher centrality values, would see greater pedestrian activity throughout the day.

2.4 Centrality

In an attempt to reflect pedestrian activity on the underlying topology of the transportation network, a centrality measure was computed in ArcGIS with the Urban Network Analysis Toolbox, and added to the regression models. Various types of network measures of centrality have been proposed in their applicability to estimation of non-motorized activity levels [10], [34], [35], and safety and collision rates [36], [37]. One of the most common measures of centrality is "betweenness" centrality, or how "between" other nodes or links a given node or link is. When considering route choice and estimating modal traffic flows, link betweenness centrality is often considered, and consists of the proportion of shortest paths between all node pairs that pass through a link or node [38]. Relatedly, stress centrality consists of counting the number of times each link in a given network is utilized among the set of shortest paths between all node pairs, and is given by:

$$C_s(k) = \sum_{i,j \in V} \sigma_{ij}(k) \tag{2.8}$$

where σ_{ij} is either 1 if link k is used in shortest path σ_{ij}, and 0 otherwise. This form of stress centrality has been used to spatially assess transportation systems [39]. This investigation does not examine the bicycle mode, but some modifications to centrality calculation are pertinent to both the bicycling and walking modes. In order to adapt stress centrality to the specific characteristics of non-motorized travel, [10] added the following modifications to the link betweenness schematics for the bicycle mode:

1. Restrict shortest paths to preferred bicycle routes

2. Restrict origin-destination (O/D) pairs to only locations reachable by bicycle

3. Modify O/D frequency with trip multipliers

However, for the walking mode, it is not reasonable to include the entire set of road network intersections as possible destinations for a given intersection-origin, due to the lower speed of the walking mode - an assumed 5 km/h. Thus, for the centrality calculations for the walking mode, an on-network radius of 5 kilometers, to represent an hour of walking at average speed, was implemented to increase the saliency and relevance of centrality to actual walking behavior. Additionally, similar modifications to the above for bicycle modes may be implemented for walking, in particular modifying O/D frequency to reflect that a certain subset of nodal origins and destinations exhibit much higher activity levels than others; for simplicity, such modifications were not attempted in this study. [1]

To reflect typical work trips, [10] chose O/D pairs such that origins were strictly residential parcels, and non-residential parcels were destinations in the morning, and the order was reversed in the evening. However, the authors speculated that allowing for non-residential destinations in the evening to reflect more complex after-work tours could increase model explanatory power [10]. Additionally, O/D pairs were limited by a network distance threshold of 5 km, per the *National Household Travel Survey* [41]. O/D multipliers specified relative magnitude of trip generation, since parcels are heterogeneous in their trip generation capacity; these included density of dwelling units within residential parcels, and square footage density for all other parcels.

These modifications constitute potentially salient areas for further investigation in our model of pedestrian traffic. O/D pairs can be tailored to favor walking trips from residential parcels to commercial destinations, as well as limited to reasonable walking distances attained within a 30-minute threshold (2.5 km). Rote stress centrality is first used to evaluate preliminary explanatory power, and feasibility of applying centrality metrics to this model.

[1] Preferred bicycle baths are chosen by [10] by defining an impedance value for each link, based on segment length, topographical slope, and segmental "friction" (i.e. ease of use, influenced by bike lanes, high traffic volume, attractions, etc.). The *Highway Capacity Manual* bicycle level-of-service defines friction to include: vehicle volumes, speeds, shoulder width, and other built environment factors [40]. Node impedance was defined as turn angle, the type of intersection control, and the hierarchical class of cross-street.

2.5 Activity Estimation and Safety

Multiple regression using Ordinary Least Squares (OLS) over the explanatory variables was performed in R for the walking mode. This estimator methodology was chosen for its ease of computation over other estimator methods, such as Generalized Linear Models (of which OLS is a special case), and Maximum Likelihood Estimation; while potentially vulnerable to autocorrelation and heteroscedasticity issues, OLS is the most straightforward process to investigate potential relationships between correlates, such as pedestrian movements and economic accessibility. The use of OLS is justified, and will yield best linear unbiased estimators (BLUE), so long as the three main axes of the Gauss-Markov assumptions upon the residuals are upheld: zero mean, uncorrelated, and constant variance.

Different time-thresholds of accessibility were compared for explanatory power of pedestrian activity, of which the strongest threshold was chosen for a final parsimonious model upon which to base the safety calculations. Iterative stepwise regression was performed using the economic sector accessibility variables, in increasing time-thresholds of travel, in an attempt to account for the possible differential walking trip generation levels of different job sectors. Pedestrian collision risk-burden is defined as the number of crashes occurring at an intersection per year (averaged over the 14-year analysis period), per pedestrian walking through that intersection on a given day. If the per-pedestrian rates of crashes are lower at intersections with more pedestrian activity, then a Safety in Numbers effect is observed. This effect is tested for both the raw data sets of pedestrian and vehicle activity, and for the estimated activity rates. In both safety models, the number of car-pedestrian crashes at intersections is not altered.

Estimated pedestrian activity and safety data are reported as 12-hour counts, extrapolated with a scaling factor from [14]. A scaling factor of 9.2 was calculated by weighted average of the hourly scaling factors for 4-5pm and 5-6pm, weighted by the percentage of daily count contributed by that hour block.

Chapter 3

Results

3.1 Pedestrian Behavior Model

Full tabulation of all bivariate regression models, to determine which time thresholds and peak-hour periods to use for greatest explanatory power in modeling pedestrian traffic levels, are included in the appendix. It was found that the 15-minute threshold of total accessibility, combined with the PM-peak period pedestrian counts and other variables, yielded the best explanatory power for walking accessibility. Table 3.1 lists summary statistics for the datasets used in the following analysis: automobile-pedestrian crashes between 2000 and 2013, and pedestrian turning movement counts between 2000 and 2013 for Minneapolis. A parsimonious model for walking activity, in terms of the strongest explanatory variables, is reported in Table 3.2. Net transit accessibility benefit was included as an explanatory variable in the pedestrian activity estimation model, to account for the effect of transit in urban cores of increasing pedestrian activity by attracting additional users beyond pure foot traffic.

The size of the sample of intersections where evening pedestrian counts were available was 741, and the number of total crashes involving pedestrians at this subset of intersections across the analysis period was 1064, of which 1052 were non-fatal and 12 were fatal. The total number of intersections included in the estimation modeling population was 1123, due to the relaxed restriction of not having pedestrian counts at intersections of estimation. It is important to note that for all three peak periods (morning 7-9am, midday 11am-1pm, and evening 4-6pm), the pedestrian count measurement distributions all showed higher standard deviation than their means, indicating a very high degree of dispersion and variability in the measurement data. The crash counts and pedestrian data at the intersection level are averaged ("intersection-average") in Table 3.1. The average number of crashes per year at intersections with evening pedestrian counts was calculated both with and without inclusion of intersections with zero crashes, to assess possible zero-inflation.

Table 3.1: Dataset summary statistics

Description	Value
Intersections with evening ped counts	741
Crashes at intersections with evening ped counts	1064 (1052 injuries, 12 deaths)
Crashes at intersections with evening ped counts per year	76
Intersections included in estimation modeling (number with crashes)	1123
Crashes at all intersections in estimation modeling	2513 (2478 injuries, 35 deaths)
Crashes at all intersections in estimation modeling per year	179.5
Intersection-average crashes/year with evening ped counts[1]	0.1518
Intersection-average crashes/year with evening ped counts[2]	0.2647
Intersection-average crashes/year in estimation modeling	0.1597
Intersection-average total ped activity per day	633.66, $\sigma = 2023.20$
Intersection-average morning ped activity per day	194.70, $\sigma = 570.34$
Intersection-average midday ped activity per day	270.74, $\sigma = 994.79$
Intersection-average evening ped activity per day	264.52, $\sigma = 733.49$

Note: Summary statistics for datasets used in pedestrian activity analysis: pedestrian turning movements between 2000 and 2013, and aggregate crash reports 2000-2013, for the City of Minneapolis.

[1] (including zero-crash intersections

[2] (without zero-crash intersections)

Table 3.2: Regression results for parsimonious models for walking activity, with and without AADT.

	Dependent variable:	
	Average PM pedestrians	
	(1)	(2)
Walking accessibility (15-minute)	0.410**	0.649***
	(0.173)	(0.112)
Net transit accessibility (30-minute)	0.320***	0.129**
	(0.093)	(0.053)
Betweenness	0.029	0.487***
	(0.371)	(0.186)
AADT	1.312*	
	(0.679)	
Management jobs 5min	−0.114***	−0.109***
	(0.033)	(0.017)
Education jobs 5min	0.922***	0.700***
	(0.086)	(0.058)
Finance jobs 10min	0.071***	0.054***
	(0.009)	(0.006)
Utilities jobs 15min	−0.968***	−0.729***
	(0.104)	(0.071)
Constant	−15.208	−1.698
	(9.874)	(4.795)
Observations	486	1,016
R^2	0.287	0.226
Adjusted R^2	0.275	0.221
Residual Std. Error	83.830 (df = 477)	72.773 (df = 1008)
F Statistic	23.970*** (df = 8; 477)	42.139*** (df = 7; 1008)
Note:	*p<0.1; **p<0.05; ***p<0.01; (standard error)	

Table 3.2 lists the predictive variables that were found to be significant correlates for pedestrian activity. Two models are listed, both with and without inclusion of the AADT metric, as a significant portion of intersections in the analysis population did not have AADT data available. Positive, significant correlates found were accessibility metrics for

walking and net transit, AADT (where present), accessibility to education jobs, and accessibility to finance jobs. Negative, significant correlates found were accessibility to management and utilities jobs. In the absence of AADT data (Table 3.2, right column), betweenness centrality became a significant, positive estimator of pedestrian activity. R^2 values for the two models were 0.29 and 0.23, respectively. The functional form of the regression equation employed is as follows:

$$P = c_1 * X_1 + c_2 * X_2 + \cdots + c_n * X_n \tag{3.1}$$

where constants c_i are regression coefficients, and X_n are the predictive correlates included in the analysis.

First, the pedestrian counts were modeled in terms of only walking accessibility, for different thresholds and times of day. From this, the strongest explanatory power was determined for PM peak period counts, at a 15-minute accessibility threshold. Results for the morning, midday, evening, and 6-hour periods are reported in the Appendix in Table 6.1, Table 6.2, Table 6.3, and Table 6.4, respectively.

Pedestrian counts were then modeled in terms of transit and walking accessibility (bimodal accessibility), for different times of day. A 30-minute transit threshold was used, in accordance with the reported data available in the Access Across America: Transit 2014 report [42]. Results for these regressions are reported in Table 6.5.

Net transit accessibility, a measure which looks at the contribution to accessibility from transit service, was also investigated as a potential explanatory variable for walking activity. A threshold of 30-minutes was again used, and models for the four time periods investigated are reported in Table 6.6 in the appendix. Betweenness centrality regression results for walking activity in the different time periods are reported in Table 6.7.

Regression results for the two parsimonious models for walking activity, with and without AADT (Average Annual Daily Traffic) included, are in Table 3.2. Figure 3.3 shows raw levels of daily pedestrian activity, aggregated from manual pedestrian counts between 2000 and 2013. The significance of the additional data sources - accessibility and betweenness centrality - is displayed in Figure 3.1 and Figure 3.2, respectively. Figure 3.1 shows the 30-minute walking accessibility for every census block within the city of Minneapolis, which given the walking mode's uniform nature, shows where economic activity is concentrated in the region. Figure 3.2 gives a sense of the most important nodes in the street network of Minneapolis - that is, the nodes that would affect the highest number of shortest paths, were they to be rendered impassible. Both of these calculations showed positive correlations with pedestrian activity, as shown in Table 3.2 for the parsimonious model, as well as in the bivariate models enumerated in the appendix. Additionally, spatial distributions of jobs in categories of Utilities, Finance, Management, and Education are shown in the appendix, in Figure 6.1, Figure 6.2, Figure 6.3, and Figure 6.4, respectively. Figure 3.4 displays the estimated PM pedestrian activity to be used in the subsequent safety analysis.

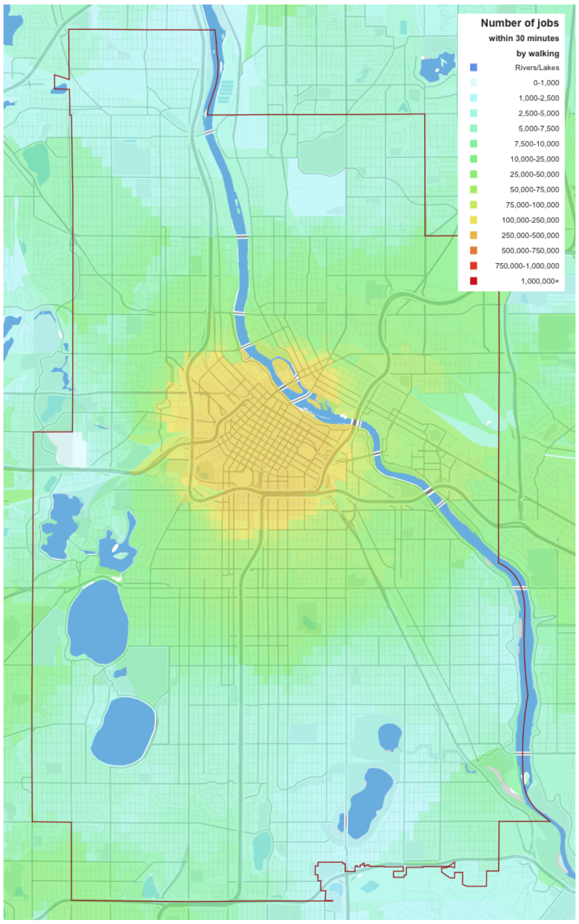

Figure 3.1: Accessibility to jobs within 30 minutes by walking in Minneapolis.

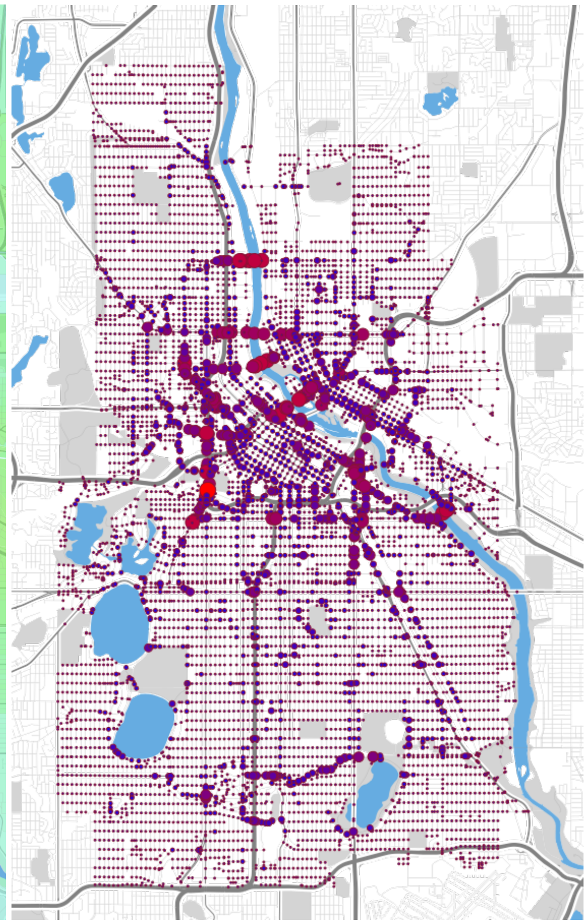

Figure 3.2: Betweenness centrality of all intersections in Minneapolis; radius of 5 km. Both size and color indicate scale of centrality metric.

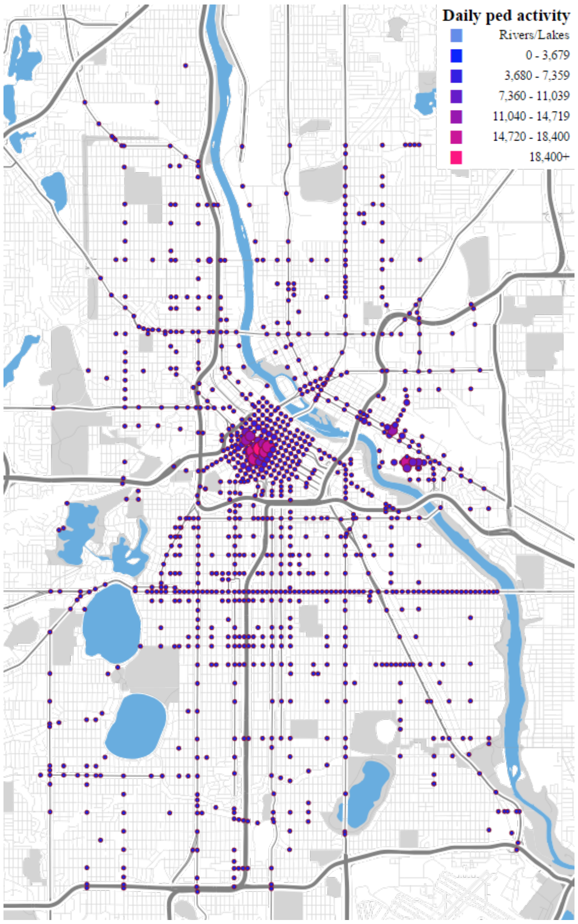

Figure 3.3: Raw levels of daily pedestrian activity in Minneapolis, 2000-2013.

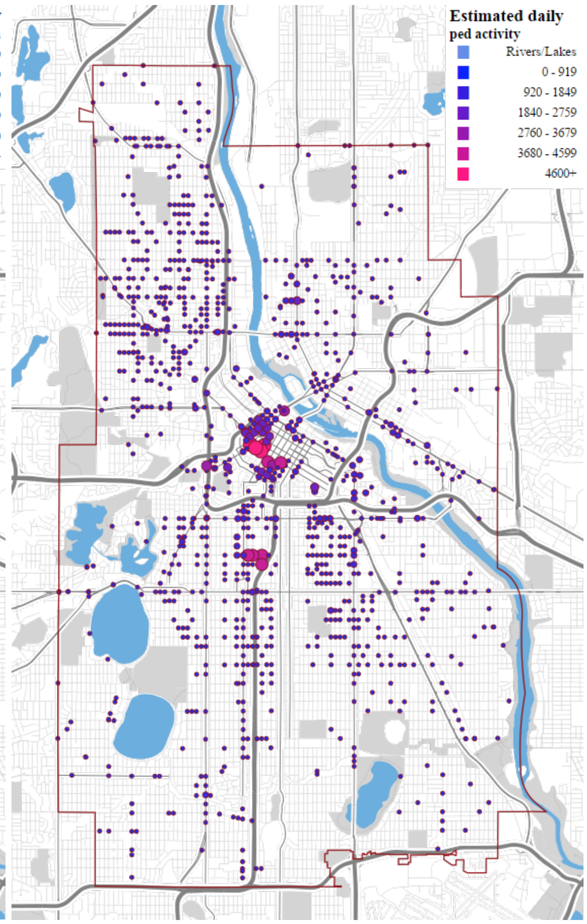

Figure 3.4: Estimated levels of evening peak pedestrian activity in Minneapolis.

3.2 Pedestrian Safety Model

A series of maps shows additional views of the data used in the safety modeling process; Figure 3.5 shows the raw count of pedestrian-auto crashes at intersections in the sample from 2000 to 2013, and Figure 3.6 shows the count of pedestrian-auto crashes, weighted by daily pedestrian activity, from 2000 to 2013. Weighting the crash data by pedestrian activity yields a visualization of the relative danger of given intersections. Safety analysis was also performed on the raw data, to attempt to verify the existence of the Safety in Numbers effect. The pedestrian-weighted data displayed in Figure 3.6 are plotted in Figure 3.9, which shows the relationship between per-pedestrian crash risk and the average daily pedestrian use level of an intersection. If Safety in Numbers is to be found within the data, then intersections characterized by greater daily levels of pedestrian activity should show lower per-pedestrian crash rates than less-active intersections; we find this to be the case, for both the raw and estimated datasets. Figure 3.10 shows the same relationship, but for estimated pedestrian count data based on the explanatory variables of accessibility and centrality (see Figure 3.1 and Figure 3.2). Exponential models are fitted to both the raw and estimated data, and both datasets appear to show significant trends towards exhibiting the SIN effect. Figure 3.7 shows a map of estimated pedestrian risk-burden of collisions, based on the estimated pedestrian activity data from the variable coefficients in Table 3.2. To validate the estimated model, the difference between actual and estimated pedestrian activity is mapped in Figure 3.8.

There are a few caveats to mention regarding the ability of simply accessibility and centrality to accurately estimate pedestrian behavior and collision risk. Figure 3.8 highlights sections of the urban area where the model differed significantly from the actual pedestrian counts. For 741 intersections, the number of daily pedestrians was over-estimated, and for 275 intersections the model under-estimated pedestrian activity. The distribution of differences has a mean $\mu = -8.10$ and standard deviation $\sigma = 72.50$; 91.11% of the sampled intersections had $actual - estimated$ differences within 1 standard deviation from the mean.

Figure 3.9 and Figure 3.10 show per-pedestrian risk vs. pedestrian activity for the raw data and estimated data, respectively. Intersections with zero crashes were excluded from the safety analysis. The units used are crashes per pedestrian per year, averaged across the 14-year analysis period (y-axis), and number of pedestrians traversing an intersection during the evening peak period per day (x-axis). Both plots show downward trends of pedestrian risk associated with increased levels of pedestrian activity at the intersection level. Negative exponential fits are included with their coefficients and relative standard errors.

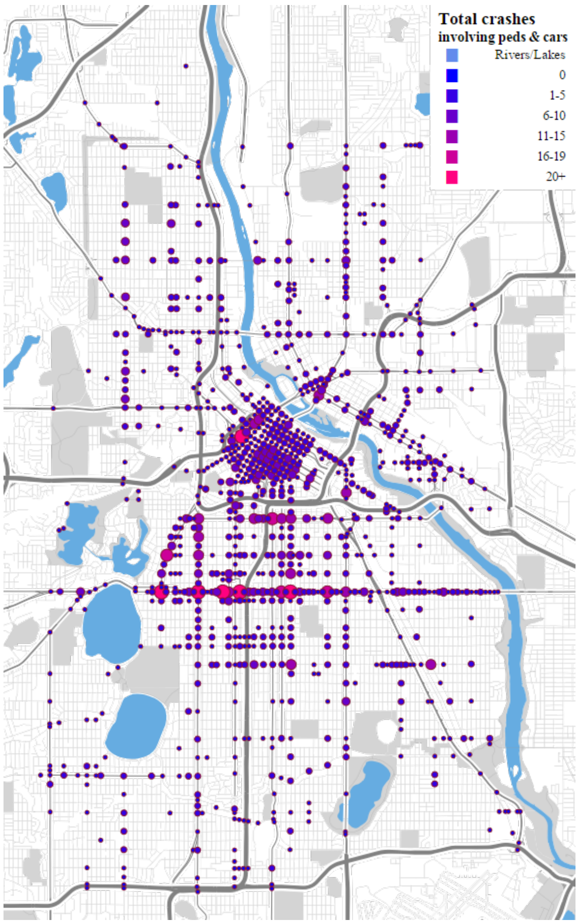

Figure 3.5: Raw levels of pedestrian-auto crashes in Minneapolis, 2000-2013.

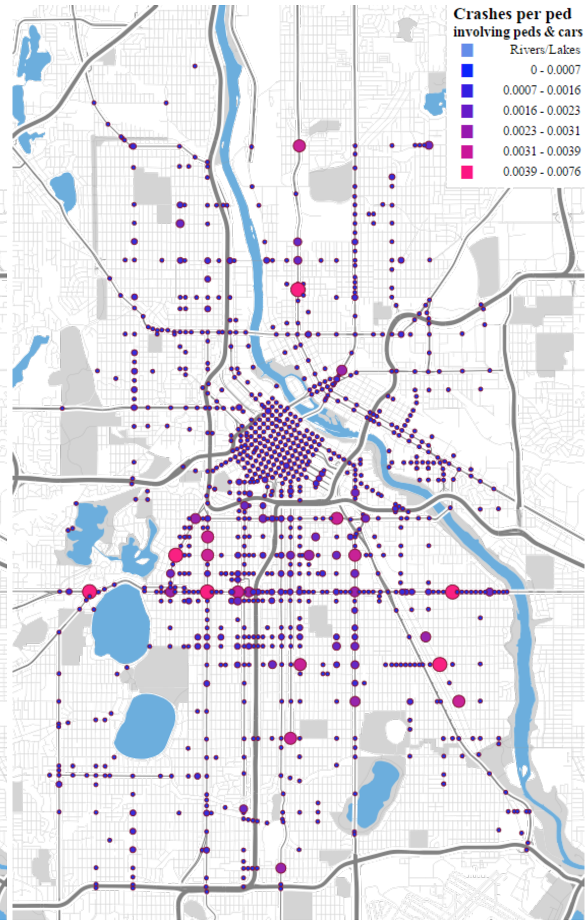

Figure 3.6: Pedestrian-weighted levels of ped-auto crashes in Minneapolis, 2000-2013. Pedestrian risk is defined as crashes per year per pedestrian per day (extrapolated 12-hour count).

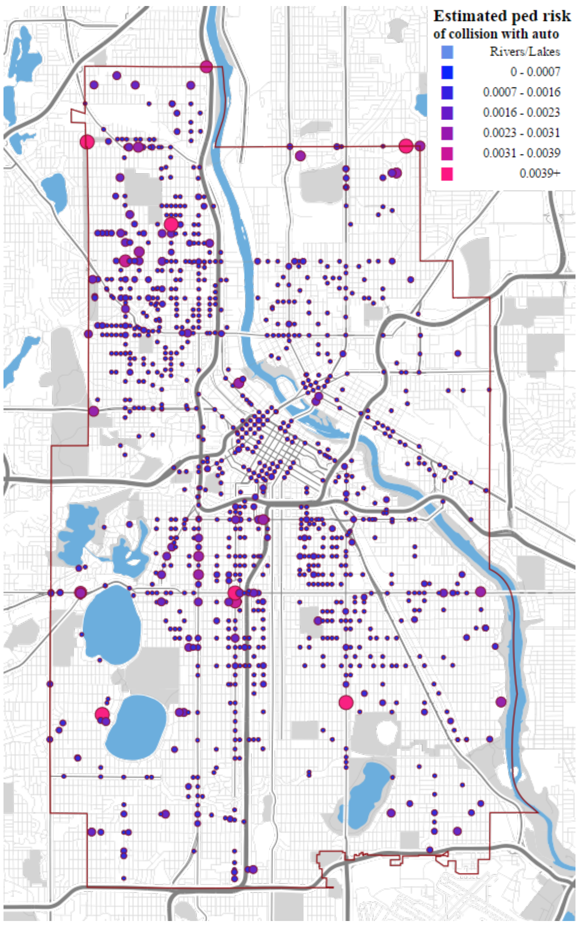

Figure 3.7: Estimated weighted pedestrian risk of crash. Pedestrian risk is defined as crashes per year per pedestrian per day (extrapolated 12-hour count).

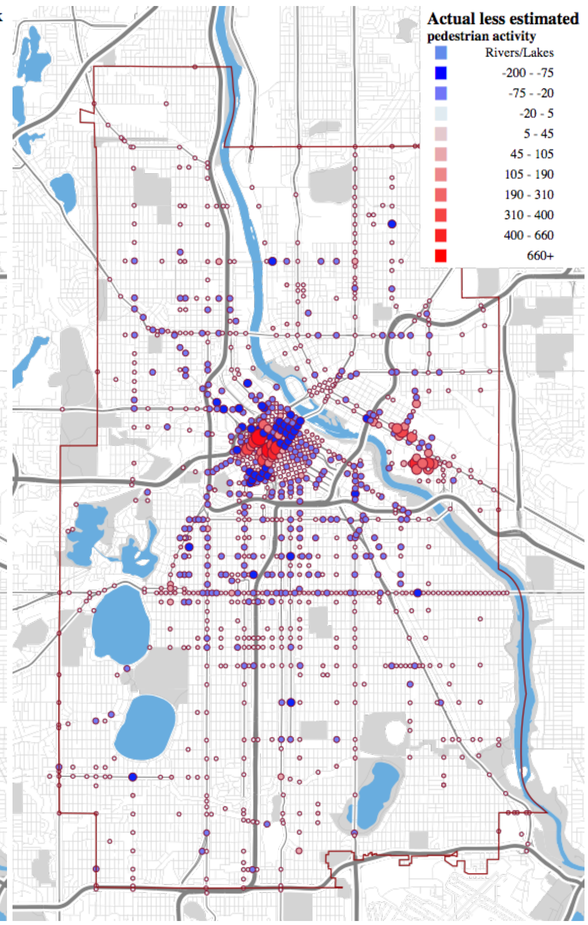

Figure 3.8: Actual minus estimated pedestrian activity, PM peak period. Reds are areas of underestimation; blues are areas of overestimation.

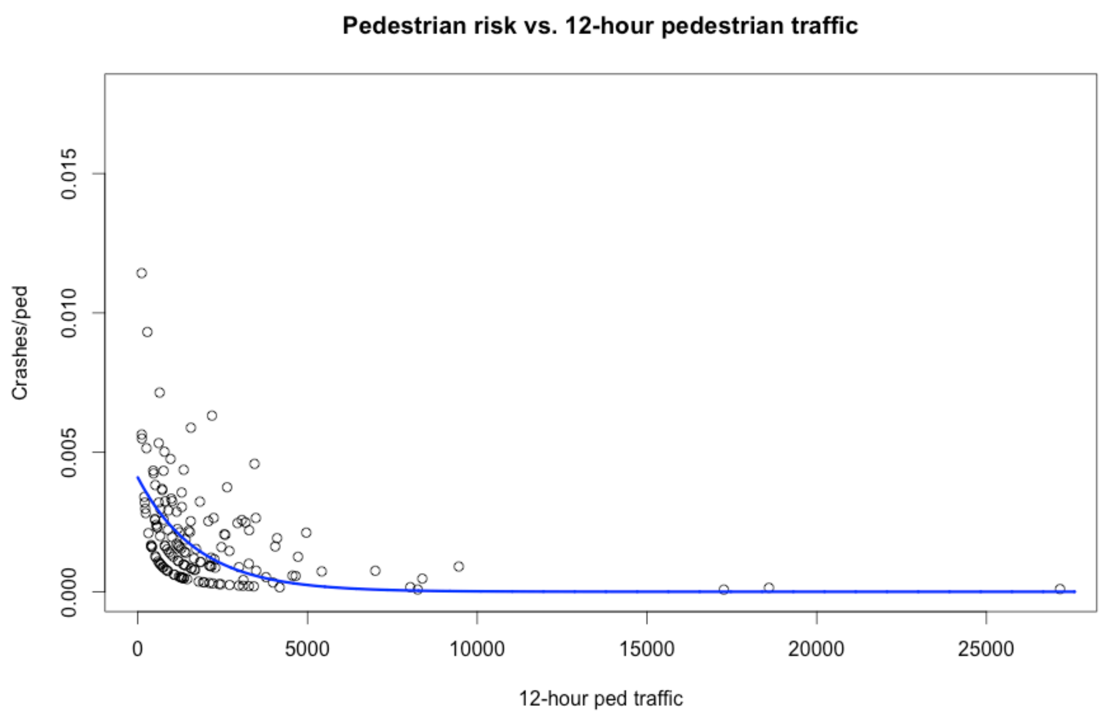

Figure 3.9: Pedestrian risk burden vs. pedestrian traffic levels, raw data; exponential fit, $b = -0.0506$, $RSE = 0.1018$, $p << 0.05$. Pedestrian risk is defined as crashes per year per pedestrian per day (extrapolated 12-hour count).

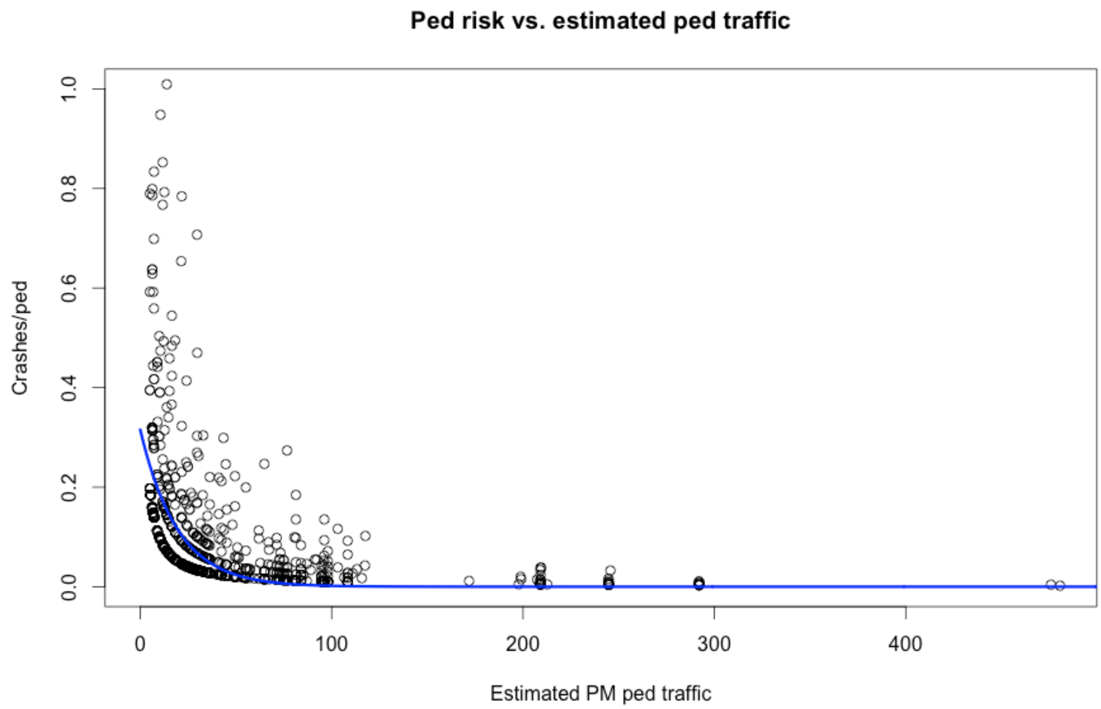

Figure 3.10: Pedestrian risk burden vs. pedestrian traffic levels, estimated data; exponential fit, $b = -0.0052$, $RSE = 0.0664$, $p << 0.05$. Pedestrian risk is defined as crashes per year per pedestrian per day (extrapolated 12-hour count).

3.3 Evaluation of the Model

To evaluate whether the modeling and regression framework was correct, the standard OLS assumptions are examined. Standard residuals of the combined (with and without AADT) model for pedestrian activity are shown in Figure 3.11, and the histogram distribution of residuals is shown in Figure 3.12. First, do the residuals have zero mean? Calculating the mean of the residuals (estimated minus actual) yields -0.1955, with a one-tailed t-statistic of 0.0585. The null hypothesis of zero-mean cannot be rejected, and so it is concluded that the model residuals do have a zero mean. Next, the serial correlation of the residuals is evaluated; the autocorrelation of residuals for a series of lags is displayed in Figure 3.13. And finally, heteroskedasticity of the model is assessed via a Breusch and Pagan test [43]; with Chi-square statistics of 843.32 and 1819.06, the null hypothesis of homoskedasticity is rejected. It would appear that only 1 of the 3 assumptions of the Gauss-Markov theorem is upheld.

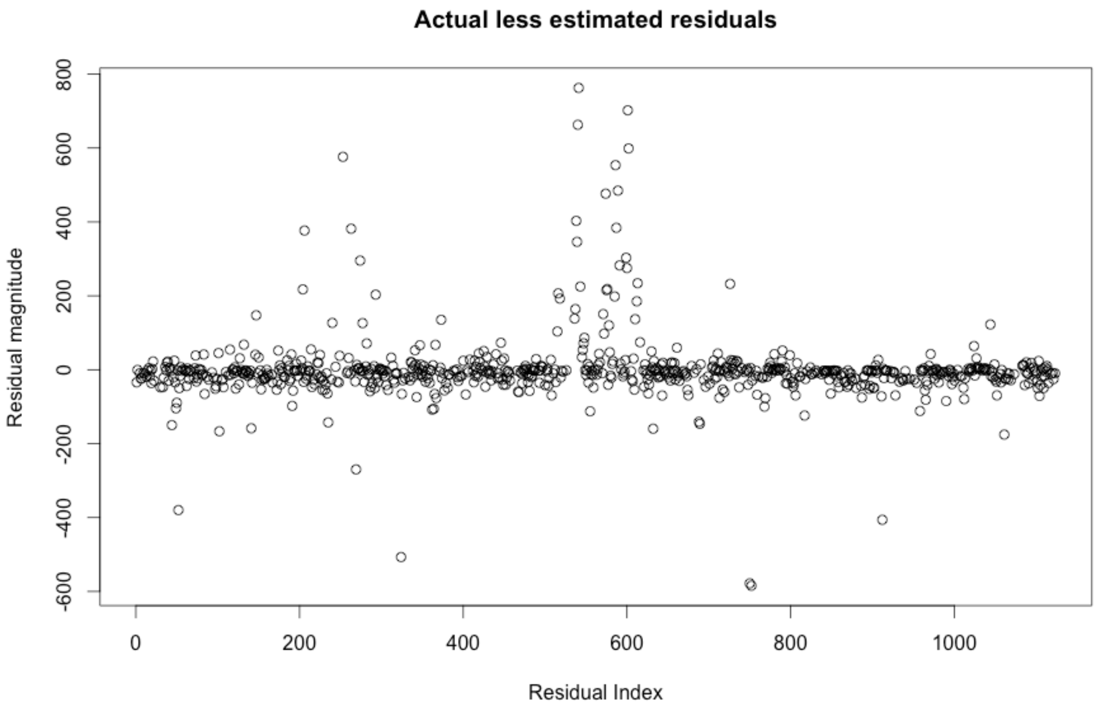

Figure 3.11: Plot of the actual less estimated residuals for the estimated model of pedestrian activity.

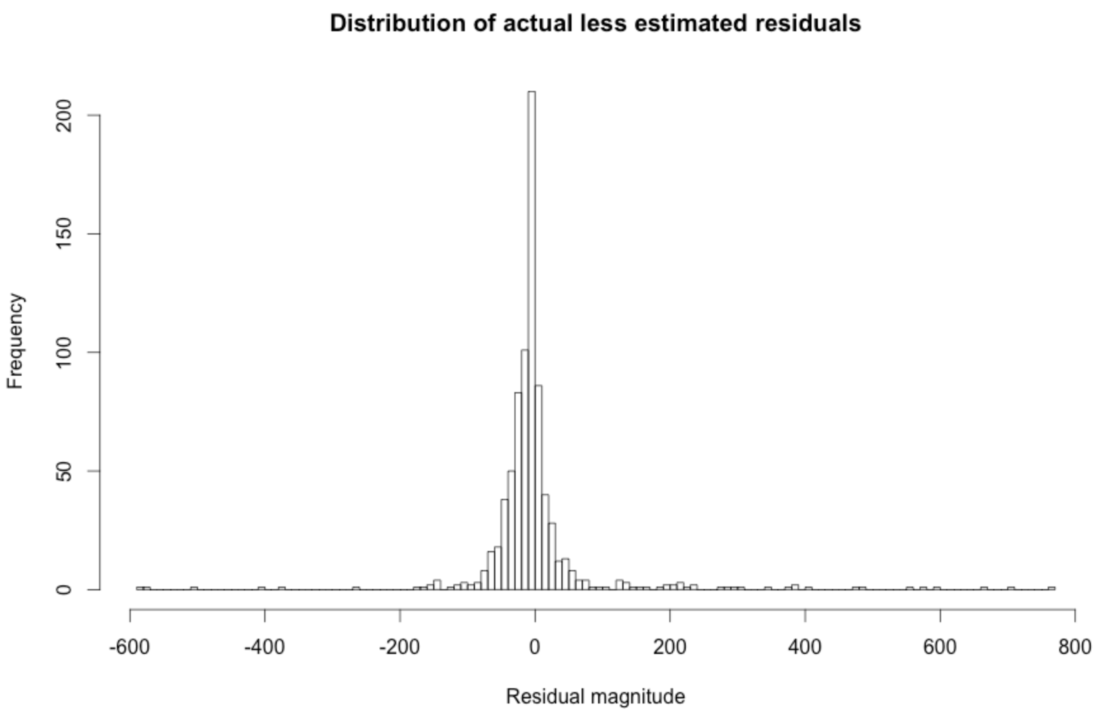

Figure 3.12: Histogram of actual less estimated residuals.

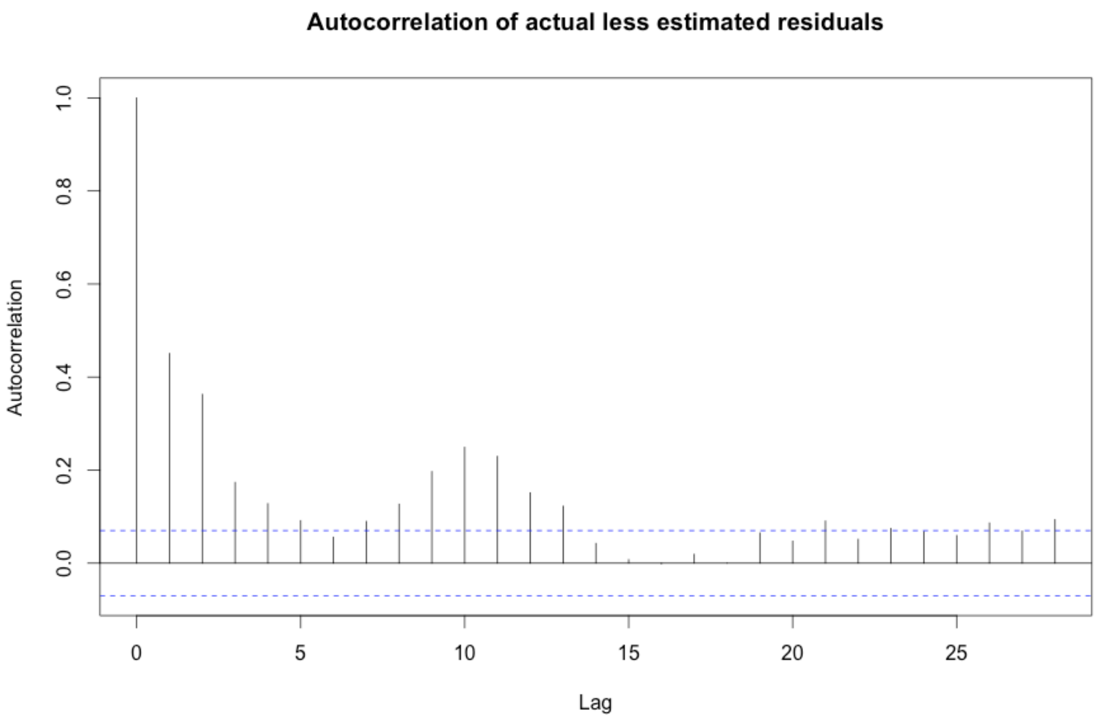

Figure 3.13: Autocorrelation of the actual less estimated residuals.

Chapter 4

Discussion

For the bivariate models of pedestrian activity in terms of census block centroid accessibility to jobs via walking, the evening peak period provided the best explanatory power. For all three time periods, as well as the 6-hour total count, R^2 values peaked near 15-minute thresholds, and dropped off in either direction. The correlation between walking accessibility and walking activity is positive. Walking is commonly thought of as a 15-minute-mode, in that the majority of people walking in urban areas will be on trips of duration 15 minutes or less. This is supported by both the 2009 National Household Travel Survey, in which mean walking trip time was found to be 14.9 minutes [44], and by the 2010 Travel Behavior Inventory survey performed by the Metropolitan Council [1]. In examination of the 2010 TBI data, an average trip time of 15.37 minutes was found for all walking trips with origin and destination contained within the city of Minneapolis; the distribution of these trips below 60-minutes in length is displayed in Figure 6.5 in the Appendix. Further, in dense urban areas, distance matters - a high-threshold measurement of walking accessibility will tend to blur the results and differences between origin points, thus potentially failing to reflect local variabilities in walking patterns. Additionally, accessibility data at the 5-minute threshold level were found to be a consistently less significant estimator of pedestrian activity than higher thresholds.

It was found that pedestrian counts in the evenings exhibited the strongest correlations with the accessibility variables tested, and midday counts exhibited the weakest correlation strengths. It is possible that midday pedestrian traffic is more dispersed in both nature of trip-making and timing, due to variable work schedules. Both the morning and evening periods exhibited stronger correlations with job-based accessibility metrics, in accordance with traditional work commute timings. The subtle difference between the two periods could be explained in part through analysis of individual trip diaries - specifically the distributions of departure and arrival times for morning and evening trips.

As was hypothesized, both the accessibility measures and betweenness centrality exhibited positive influences on pedestrian activity levels, with all the significant variables with strongest R^2 metrics having positive signs. However, centrality was only significant in the model which lacked the AADT variable, indicating some level of correlation between centrality and AADT. This gives a reasonable framework through which to estimate modal traffic levels at every intersection in Minneapolis and, by extension of the broader framework, in other cities as well. Figure 3.6 and Figure 3.7 suggest the level of policy and planning information that can be gathered from such a study. When multiple intersections with relatively high pedestrian injury risk-burden lie in the same corridor, such as Lake Street in Minneapolis, a discussion of pedestrian safety and the surrounding built environment should occur. Through the pedestrian risk-burden analysis, it is also possible to see intersections with a disproportionately high rate of crashes for its level of pedestrian activity. There are other factors at play in assessing risk (built environment characteristics, perceived risk, etc.), but direct measurement of traffic levels and crash reports offers the statistical likelihood of crashes as a useful metric.

However, betweenness centrality did not exhibit as strong a positive correlation as was hypothesized. This may have resulted from the specific methodology used - that is, a centrality calculation that takes into account heterogeneous trip generation within an urban area due to varying land use patterns may lead to higher estimating power of centrality measures toward actual pedestrian behavior patterns. Pedestrian behavior in urban areas does not exhibit uniform all-to-all trip generation distribution; rather, there are major sources and attractors, which would shift the distribution of route choices, and thus link and intersection centrality, to favor routes between those origin-destination pairs. Applying techniques analogous to those in [10] to the walking model may yield more accurate pedestrian behavior estimation based on the centrality metric.

The Safety in Numbers effect was indeed observed in both the raw Minneapolis pedestrian and crash data, as well as the modeled data at the broader sample of intersections (visible in Figure 3.3 and Figure 3.5). Intersections characterized by higher per-day pedestrian traffic exhibited lower per-pedestrian crash rates, a phenomenon that has been observed and described previously (see [16], [17], [18]). The precise reasons behind this effect are not definitively known; however, the aforementioned studies have hypothesized psychological effects on drivers; [45] determined salient factors in automobile-pedestrian crashes to include poor traffic management (speed), driving while impaired, inappropriate infrastructure, and lack of respect for vulnerable road users - the lattermost of which may be influenced by greater numbers of pedestrians. Additionally, spatial geometric probability of crashes for a given pedestrian necessarily varies with additional pedestrians present within an intersection. There are physical constraints imposed upon actors in an intersection by the built environment, such as upper limits on the level of vehicular flow through the intersection, number of vehicles within the intersection in a given moment, and number of pedestrians able to fit within its crosswalks. This necessarily limits the frequency and types of collisions that may occur. Continuing to add pedestrians, up to a point, to a built environment crossing where a fixed rate of cars are driven badly and would strike a pedestrian if one were there, should both increase the total number of collisions (more collision opportunities), and keep the rate of collisions per pedestrian approximately the same. It's possible there is a saturation point beyond which adding additional pedestrians does not create a significantly greater number of new collision opportunities, thus lowering the per-pedestrian collision risk. More research on the precise potential psychosocial mechanisms behind driver behavior at intersections, as well as more detailed statistical characterization of accident rates at varying types of intersections, is needed.

Additionally, there are policy implications to consider when observing and discussing any SIN-type effect among vulnerable road users and their exposure to deadly interactions with automobiles. The direction of temporal causality is not known between increased numbers of pedestrians and increased safety - does walking become safer because more people do it, or do more people walk places because it has been shown to be, or perceived to be, safer? Policies that promote the former, and simply promote walking as a viable transport mode without structural changes in the built environment, may be misleading and misguided at best, and may unduly expose pedestrians to places within a built environment that lack the proper treatments. Sufficient walking infrastructure and traffic calming measures should be a priority, and a precursor requirement to promotional campaigns for walking, rather than simply paradoxically promoting walking as itself a mechanism for increased walking safety.

Accessibility to Education and Finance jobs was found to be significantly correlated with increased pedestrian activity, while accessibility to Management and Utilities jobs was found to be significantly correlated with decreased pedestrian activity, relative to other categories; these spatial maps are visible in Figure 6.1, Figure 6.2, Figure 6.3, and Figure 6.4. Utility jobs tend to be concentrated in areas not immediately in the downtown core, as well as management jobs to a lesser degree; finance jobs are heavily concentrated in the downtown core area, and education jobs are concentrated on walkable campuses. Further, it is plausible that certain categories of jobs attract greater or lesser levels of walking among their workers, dependent on such factors as dress requirements, vehicle needs (e.g. construction and contract workers), and typical density of jobs within each category. Additional cross-comparison analysis among economic job categories is needed to investigate these effects, but initial analysis indicates these spatial distributions correlate to the regression coefficients in Table 3.2.

A significant and pervasive challenge with analysis dependent on pedestrian, bicycle, and vehicular count and crash data is the issue of data quality and format. Methodologies and data standards can vary from city to city and jurisdiction to jurisdiction; this study used a combination of national (Census, LEHD) datasets and local (Minneapolis traffic) data. Some cities, such as Boston, do not have robust pedestrian and bicycle counting programs throughout the city; others, such as Philadelphia, may have varying data release and non-disclosure agreements between MPOs, cities, and police departments; still other cities may have inconsistent data tracking and release practices, such as Washington, D.C. Such hurdles can make the collection and processing of pedestrian and bicycle spatial safety data on a national scale exceedingly difficult. Better standards of practice in data collection, management, and distribution are needed.

However, with pedestrian activity estimation based on sampling existing counts, accessibility analysis, and betweenness centrality of the underlying network, it becomes possible to estimate the individual risk-burden experienced by a single pedestrian at any intersection that has experienced a crash event. Such techniques may prove important in informing urban planning processes and decisions, pedestrian safety programs, and highlighting intersections characterized by excessive per-pedestrian risk that may be mitigated with more informed planning and engineering. An important extension of the identification of intersections with undue pedestrian risk burden is the visualization of such intersections - e.g. the obvious Lake Street corridor as demonstrated in Figure 3.1. The entire corridor stands out as an area with elevated pedestrian risk burdens given the number of pedestrians walking there. Further, if the sample data were to only contain a few intersections, the estimated models would enable planners and engineers to construct a more complete picture of pedestrian safety and activity throughout the entire corridor.

Regarding model accuracy, a few of the primary assumptions of the Gauss-Markov theorem for BLUE (best linear unbiased estimator) verification of OLS estimators were not upheld. Thus, in further investigation into this topic, relaxing the OLS assumptions and utilizing a GLM framework, in addition to testing the assumptions for negative binomial estimation of pedestrian activity correlates (as previously done by [14]), should be considered. Additionally, data quality and availability issues may contribute to the nullification of OLS first assumptions, and in cases of larger, better-maintained datasets, OLS may prove a justifiable estimator.

Qualitatively, the cases of underestimation and overestimation are geographically interesting to note; the two major areas of underestimation are the inner downtown core, and the East Bank Campus of the University of Minnesota, just east of the Mississippi River, while the major area of overestimation is located west of Hennepin Ave in downtown, near Dunwoody Boulevard and Olson Memorial Highway. The downtown core and the campus of the University are characterized by significant pedestrian activity and are considered walkable areas, whereas the areas just west of downtown are not as walkable; in fact, Dunwoody Boulevard, Olson Memorial Highway, and other roads in the area are multi-lane automobile thoroughfares. While the road network structure and proximity to downtown would suggest significant pedestrian activity, physical barriers exist within the built environment. These cases highlight the limitations of centrality and accessibility in capturing elements of the built environment relevant to pedestrian activity where local and hyper-local factors may play significant roles.

Chapter 5

Conclusion

Economic accessibility, AADT, and betweenness centrality were found to be significant positive correlates with pedestrian movements in the city of Minneapolis, including select economic categorical breakdowns of accessibility (Education and Finance jobs). Centrality was only significant in the model which did not include the AADT variable. A model to estimate gross pedestrian behavior at the intersection level was constructed from these correlates, in an attempt to create a more complete spatial picture of pedestrian activity throughout the city. Differential levels of pedestrian behavior correlation with categorical economic accessibility were discussed, as well as issues pertaining to local built environment characteristics and over- and under-prediction. More complete datasets with higher quality of pedestrian count data are desired, and additional modifications to betweenness centrality calculations, such as modeling O/D pair and trip-making frequencies after LEHD commute data, may lead to higher predictive power of pedestrian models.

The Safety In Numbers effect, wherein increased numbers of pedestrians correlate with a decrease in per-pedestrian collision risk, was observed in both the raw collision and count data in Minneapolis, as well as for the modeled pedestrian movement data. The potential driving forces and causality behind this phenomenon were discussed, along with policy implications of various types of integration of this phenomenon into built environment design and traffic governance. The ever-present issue of data quality and availability is germane to this research; good, complete data were not available for many other cities, despite repeated efforts in pursuit. Data gathering practices in traffic monitoring tend not to be multi-modal in scope, and the holy grail of non-motorized transportation research data availability is complete-coverage real-time counts from automated sensors. This is a large hurdle to clear, but as sensor technology improves and becomes more economically attainable, cities will be further able to monitor their transportation mode-shares in real-time, giving planners and practitioners a very dense and complete picture of non-motorized transportation use and demand. Until such systems become ubiquitously implemented, estimation and modeling techniques will be required to inform planners and engineers of potential pedestrian demand. A scalable and translatable framework of modeling with betweenness network centrality and economic accessibility has been given, and its utility shown in pedestrian safety analysis.

References

[1] Metropolitan Council. Metropolitan Council Travel Behavior Inventory. Technical Report December, Metropolitan Council, 2014.

[2] Les Lumsdon and Jayne Mitchell. Walking, transport and health: Do we have the right prescription? *Health Promotion International*, 14(3):271–279, 1999.

[3] Noah Raford and David Ragland. Space Syntax: Innovative Pedestrian Volume Modeling Tool for Pedestrian Safety, 2004.

[4] David Brownstone. Key Relationships Between the Built Environment and VMT. *Transportation Research*, (Special report 298: Driving and the Built Environment: The Effects of Compact Development on Motorized Travel, Energy Use and CO2 emissions):2–4, 2008.

[5] World Health Organization. Global status report on road safety. *Injury prevention*, page 318, 2013.

[6] Søren Underlien Jensen. Safety effects of blue cycle crossings: A before-after study. *Accident Analysis and Prevention*, 40(2):742–750, 2008.

[7] U S Census Bureau. Modes Less Traveled Bicycling and Walking to Work in the United States : 2008 2012. (May):2008–2012, 2014.

[8] Robert J. Schneider, Rhonda M. Ryznar, and Asad J. Khattak. An accident waiting to happen: A spatial approach to proactive pedestrian planning. *Accident Analysis and Prevention*, 36(2):193–211, 2003.

[9] Michael Iacono, Kevin J. Krizek, and Ahmed El-Geneidy. Measuring non-motorized accessibility: issues, alternatives, and execution. *Journal of Transport Geography*, 18(1):133–140, 2010.

[10] Stephen McDaniel, Michael B Lowry, and Michael Dixon. Using Origin-Destination Centrality to Estimate Directional Bicycle Volumes. page 15, 2013.

[11] Michael Lowry. Spatial interpolation of traffic counts based on origin-destination centrality. *Journal of Transport Geography*, 36:98–105, 2014.

[12] Robert J. Schneider, Lindsay S. Arnold, and David R. Ragland. Pilot Model for Estimating Pedestrian Intersection Crossing Volumes, 2010.

[13] Luis F. Miranda-Moreno, Patrick Morency, and Ahmed M. El-Geneidy. The link between built environment, pedestrian activity and pedestrian-vehicle collision occurrence at signalized intersections. *Accident Analysis and Prevention*, 43(5):1624–1634, 2011.

[14] Steve Hankey, Greg Lindsey, Xize Wang, Jason Borah, Kristopher Hoff, Brad Utecht, and Zhiyi Xu. Estimating use of non-motorized infrastructure: Models of bicycle and pedestrian traffic in Minneapolis, MN. *Landscape and Urban Planning*, 107(3):307–316, 2012.

[15] Maria Kamargianni. Generation s Y Travel Behavior and Perceptions Towards Walkability Constraints. *Transportation Research Board 93rd Annual Meeting. January 12-16, Washington, D.C.*, 2014.

[16] P Jacobsen. Safety in numbers: more walkers and bicyclists, safer walking and bicycling. *Injury Prevention*, 9(3):205–209, 2003.

[17] Lars Leden. Pedestrian risk decrease with pedestrian flow. A case study based on data from signalized intersections in Hamilton, Ontario. *Accident Analysis and Prevention*, 34(4):457–464, 2002.

[18] Rajiv Bhatia and Megan Wier. "Safety in Numbers" re-examined: Can we make valid or practical inferences from available evidence? *Accident Analysis and Prevention*, 43(1):235–240, 2011.

[19] Bushra Bader Obeidat and Imad Akram Al-Hashimi. Pedestrian risk index for Irbid city , Jordan. pages 1–12. International Space Syntax Symposium, 2015.

[20] J F Sallis, a Bauman, and M Pratt. Environmental and policy interventions to promote physical activity. *American journal of preventive medicine*, 15(4):379–397, 1998.

[21] N Humpel. Environmental factors associated with adults' participation in physical activity A review. *American Journal of Preventive Medicine*, 22(3):188–199, 2002.

[22] Robert Cervero and Kara Kockelman. Travel Demand And The 3ds : Density , Design And Diversity. *Transportation Research Part D: Transport and Environment*, 2(3):199–219, 1997.

[23] Samantha Hajna, Nancy a. Ross, Anne-Sophie Brazeau, Patrick Bélisle, Lawrence Joseph, and Kaberi Dasgupta. Associations between neighbourhood walkability and daily steps in adults: a systematic review and meta-analysis. *BMC Public Health*, 15(1):768, 2015.

[24] Eva Leslie, Neil Coffee, Lawrence Frank, Neville Owen, Adrian Bauman, and Graeme Hugo. Walkability of local communities: Using geographic information systems to objectively assess relevant environmental attributes. *Health and Place*, 13(1):111–122, 2007.

[25] S. L. Handy and D. a. Niemeier. Measuring accessibility: An exploration of issues and alternatives. *Environment and Planning A*, 29(7):1175–1194, 1997.

[26] R P McKenzie. The Measurement of Accessibility to Employment. *Transport Studies Unit, University of Oxford*, page 38, 1985.

[27] Walter Kondo, Brian Barber, and Len Sherman. Method for evaluating metropolitan accessibility. *Transportation research record*, 499:70–82, 1974.

[28] G V Wickstrom. Defining balanced transportation: a question of opportunity. *Traffic Quarterly*, 25(3):337–349, 1971.

[29] L. Bertolini, F. le Clercq, and L. Kapoen. Sustainable accessibility: A conceptual framework to integrate transport and land use plan-making. Two test-applications in

the Netherlands and a reflection on the way forward. *Transport Policy*, 12(3):207–220, 2005.

[30] Michael Iacono and David Levinson. Location, Regional Accessibility, and Price Effects. *Transportation Research Record: Journal of the Transportation Research Board*, 2245(-1):87–94, 2012.

[31] Issam Srour, Kara Kockelman, and Travis Dunn. Accessibility Indices: Connection to Residential Land Prices and Location Choices. *Transportation Research Record*, 1805(1):25–34, 2002.

[32] Thomas Straatemeier. How to plan for regional accessibility? *Transport Policy*, 15(2):127–137, 2008.

[33] Mei-Po Kwan. Space-Time and Integral Measures of Individual Accessibility: A Comparative Analysis Using a Point-based Framework. *Geographical Analysis*, 30(3):191–216, 1998.

[34] Paulo Rui Anciães. Urban transport , pedestrian mobility and social justice A GIS analysis of the case of the Lisbon Metropolitan Area. (July), 2011.

[35] Minh Tan Do, Offer Grembek, David Ragland, and Ching-Yao Chan. Weighting integration by block heterogeneity to evaluate pedestrian activity. pages 14p, schmas,\rtabl., ill., bibliogr., 2013.

[36] Yuanyuan Zhang, John Bigham, David Ragland, and Xiaohong Chen. Investigating the associations between road network structure and non-motorist accidents. *Journal of Transport Geography*, 42:34–47, 2015.

[37] Dajun Dai, Emily Taquechel, John Steward, and Sheryl Strasser. The impact of built environment on pedestrian crashes and the identification of crash clusters on an urban university campus. *The western journal of emergency medicine*, 11(3):294–301, 2010.

[38] Chris McCahil and Norman W. Garrick. The Applicability of Space Syntax to Bicycle Facility Planning. *Transportation Research Record: Journal of the Transportation Research Board*, 2074(-1):46–51, 2008.

[39] Sybil Derrible. No Title. In *Using Network Centrality to Determine Key Transfer Stations in Public Transportation Systems*, 2012.

[40] Richard G Dowling and Barbara K Ostrom. Highway Capacity Manual 2010. Technical report, Transportation Research Board, Washington, D.C., 2010.

[41] Federal Highway Administration. 2009 National Household Travel Survey. page 82, 2010.

[42] Andrew Owen and David Levinson. Access Across America: Transit 2014. (September), 2014.

[43] T.S. Breusch, T.S. Breusch, a.R. Pagan, and a.R. Pagan. A Simple Test for Heteroscedasticity and Random Coefficient Variation. *Econometrica*, 47(5):1287–1294, 1979.

[44] Author Manuscript. NIH Public Access. *American journal of preventive medicine*, 43(1):11–19, 2012, NIHMS150003.

[45] K. Goniewicz, M. Goniewicz, W. Pawowski, and P. Fiedor. Road accident rates: strategies and programmes for improving road traffic safety. *European Journal of Trauma and Emergency Surgery*, 2015.

Chapter 6

Appendix

6.1 Additional Figures

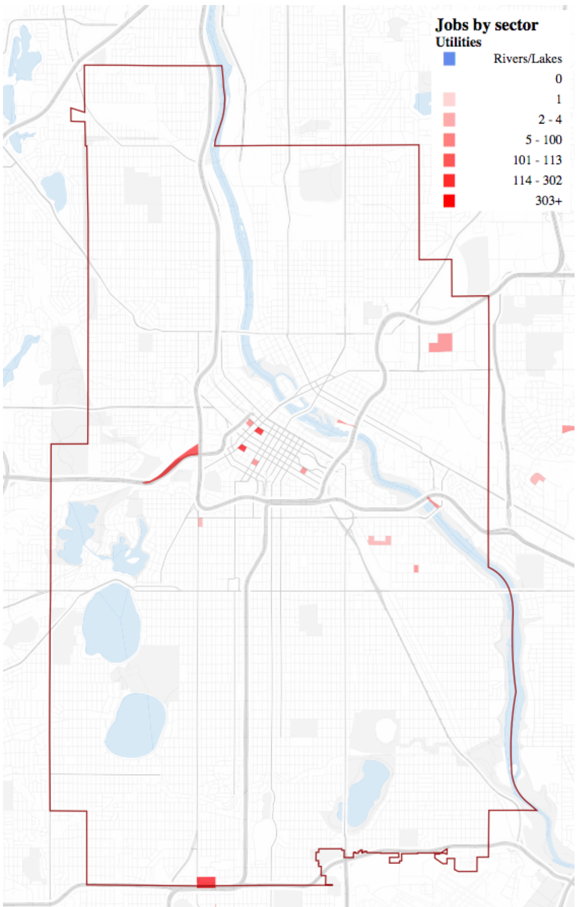

Figure 6.1: Spatial distribution of utility jobs in Minneapolis, based on LEHD data.

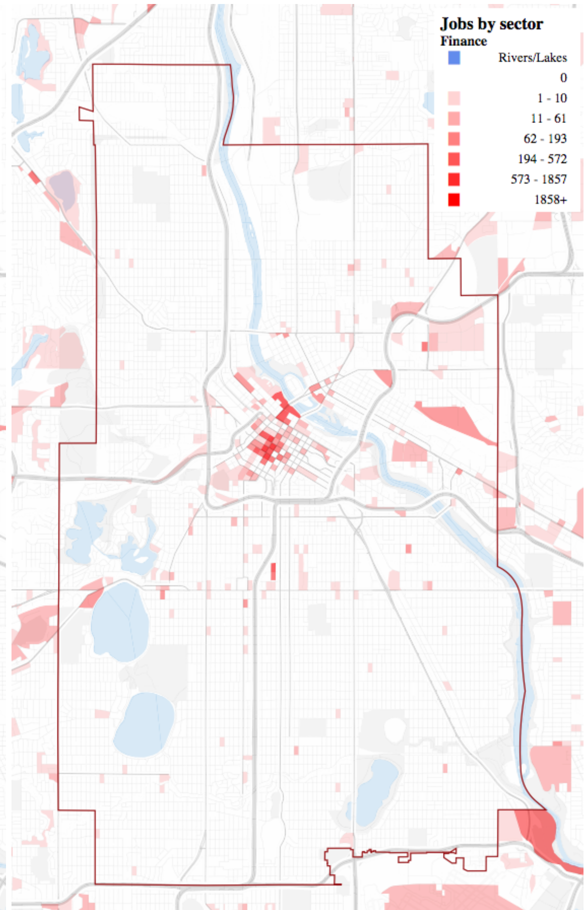

Figure 6.2: Spatial distribution of finance jobs in Minneapolis, based on LEHD data.

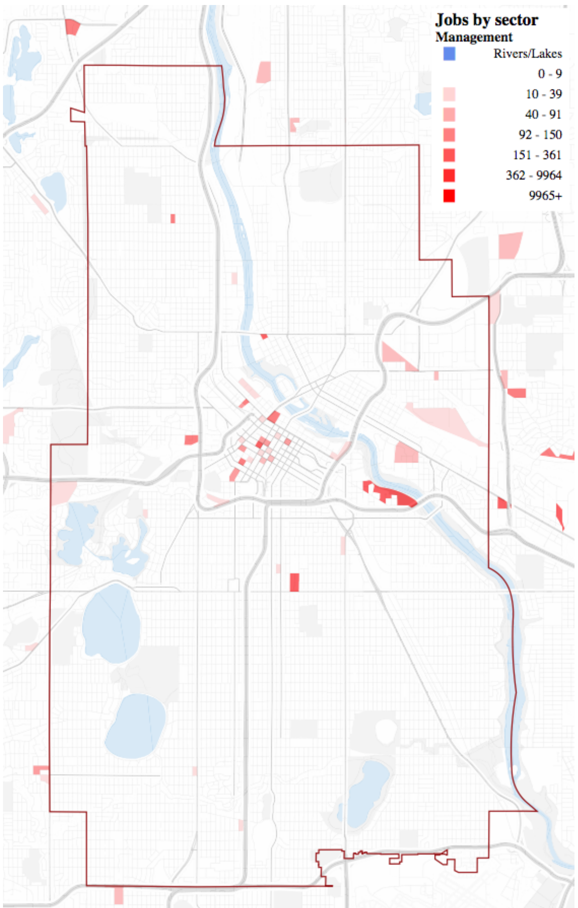

Figure 6.3: Spatial distribution of management jobs in Minneapolis, based on LEHD data.

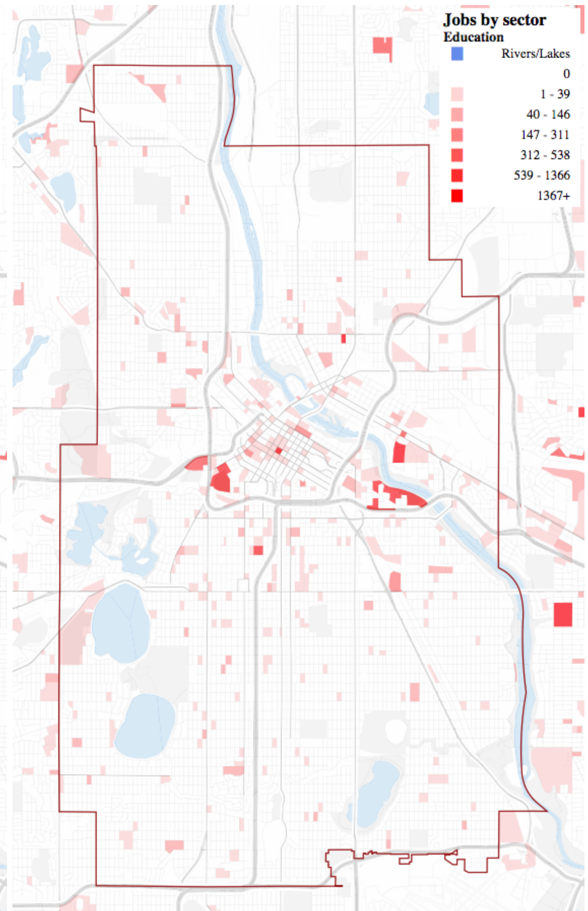

Figure 6.4: Spatial distribution of education jobs in Minneapolis, based on LEHD data.

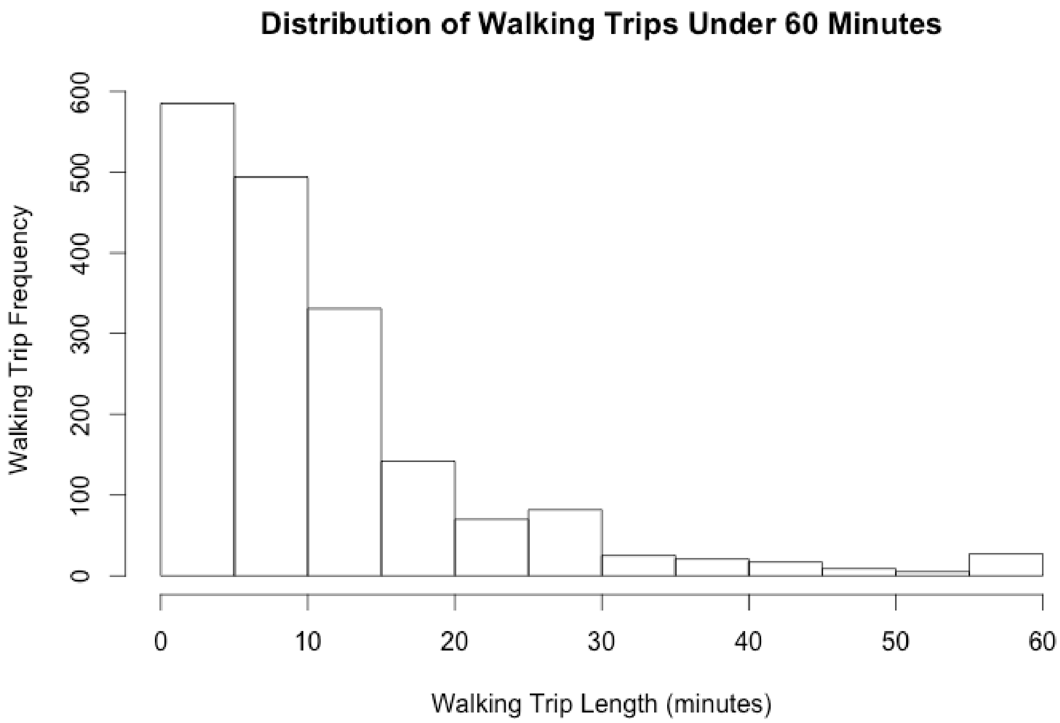

Figure 6.5: Distribution of walking trips with origin and destination contained within Minneapolis, and duration under 60 minutes. Data were extracted from the 2010 Travel Behavior Inventory report for the Twin Cities, Minnesota area [1].

6.2 Additional Tables

Table 6.1: Regression results explaining walking counts with walking accessibility, AM period.

	Dependent variable:					
	ped_am_avg					
	(1)	(2)	(3)	(4)	(5)	(6)
Walking accessibility (5-minute)	1.621*					
	(0.889)					
Walking accessibility (10-minute)		0.735***				
		(0.134)				
Walking accessibility (15-minute)			0.717***			
			(0.058)			
Walking accessibility (20-minute)				0.495***		
				(0.038)		
Walking accessibility (25-minute)					0.401***	
					(0.032)	
Walking accessibility (30-minute)						0.315***
						(0.028)
Constant	18.580***	16.185***	9.689***	7.108***	5.553***	4.820**
	(1.718)	(1.721)	(1.717)	(1.790)	(1.875)	(2.010)
Observations	1,056	1,056	1,056	1,056	1,056	1,056
R^2	0.003	0.028	0.127	0.138	0.133	0.110
Adjusted R^2	0.002	0.027	0.126	0.137	0.132	0.109
Residual Std. Error (df = 1054)	52.771	52.115	49.378	49.068	49.217	49.871
F Statistic (df = 1; 1054)	3.328*	30.107***	153.651***	168.915***	161.527***	129.860***

Note: *p<0.1; **p<0.05; ***p<0.01; (standard error)

Table 6.2: Regression results explaining walking counts with walking accessibility, midday period.

	Dependent variable:					
	ped_md_avg					
	(1)	(2)	(3)	(4)	(5)	(6)
Walking accessibility (5-minute)	1.966					
	(1.398)					
Walking accessibility (10-minute)		1.045***				
		(0.211)				
Walking accessibility (15-minute)			1.077***			
			(0.091)			
Walking accessibility (20-minute)				0.703***		
				(0.061)		
Walking accessibility (25-minute)					0.566***	
					(0.050)	
Walking accessibility (30-minute)						0.445***
						(0.044)
Constant	25.255***	21.632***	11.593***	8.742***	6.643**	5.589*
	(2.704)	(2.713)	(2.717)	(2.856)	(2.992)	(3.197)
Observations	1,056	1,056	1,056	1,056	1,056	1,056
R^2	0.002	0.023	0.116	0.113	0.107	0.089
Adjusted R^2	0.001	0.022	0.115	0.112	0.106	0.088
Residual Std. Error (df = 1054)	83.033	82.161	78.132	78.285	78.526	79.339
F Statistic (df = 1; 1054)	1.976	24.504***	138.616***	133.957***	126.660***	102.611***

Note: *p<0.1; **p<0.05; ***p<0.01; (standard error)

Table 6.3: Regression results explaining walking counts with walking accessibility, PM period.

	Dependent variable:					
	ped_pm_avg					
	(1)	(2)	(3)	(4)	(5)	(6)
Walking accessibility (5-minute)	2.869**					
	(1.363)					
Walking accessibility (10-minute)		1.227***				
		(0.205)				
Walking accessibility (15-minute)			1.164***			
			(0.088)			
Walking accessibility (20-minute)				0.759***		
				(0.058)		
Walking accessibility (25-minute)					0.611***	
					(0.048)	
Walking accessibility (30-minute)						0.484***
						(0.042)
Constant	27.681***	23.786***	13.379***	10.332***	8.062***	6.755**
	(2.636)	(2.634)	(2.612)	(2.748)	(2.880)	(3.083)
Observations	1,056	1,056	1,056	1,056	1,056	1,056
R^2	0.004	0.033	0.143	0.138	0.131	0.110
Adjusted R^2	0.003	0.032	0.142	0.137	0.130	0.109
Residual Std. Error (df = 1054)	80.941	79.768	75.102	75.312	75.601	76.512
F Statistic (df = 1; 1054)	4.429**	35.791***	175.397***	168.572***	159.230***	130.495***

Note: *p<0.1; **p<0.05; ***p<0.01; (standard error)

Table 6.4: Regression results explaining walking counts with walking accessibility, 6-hour count totals.

	Dependent variable:					
	six_hour_count					
	(1)	(2)	(3)	(4)	(5)	(6)
Walking accessibility (5-minute)	6.455* (3.563)					
Walking accessibility (10-minute)		3.007*** (0.537)				
Walking accessibility (15-minute)			2.958*** (0.231)			
Walking accessibility (20-minute)				1.958*** (0.153)		
Walking accessibility (25-minute)					1.577*** (0.127)	
Walking accessibility (30-minute)						1.244*** (0.111)
Constant	71.516*** (6.890)	61.603*** (6.895)	34.661*** (6.855)	26.182*** (7.194)	20.258*** (7.539)	17.165** (8.072)
Observations	1,056	1,056	1,056	1,056	1,056	1,056
R^2	0.003	0.029	0.135	0.134	0.128	0.107
Adjusted R^2	0.002	0.028	0.134	0.133	0.127	0.106
Residual Std. Error (df = 1054)	211.599	208.842	197.122	197.182	197.881	200.318
F Statistic (df = 1; 1054)	3.282*	31.379***	164.286***	163.538***	154.959***	125.716***
Note:	*p<0.1; **p<0.05; ***p<0.01; (standard error)					

Table 6.5: Regression results explaining walking counts with accessibility to jobs via transit & walking, for a 30-minute threshold, during different time periods.

	Dependent variable:			
	ped_am_avg	ped_md_avg	ped_pm_avg	six_hour_count
	(1)	(2)	(3)	(4)
Transit & walking accessibility (30-minute)	0.203***	0.289***	0.310***	0.802***
	(0.020)	(0.031)	(0.030)	(0.079)
Constant	−5.529*	−9.540**	−9.129**	−24.198**
	(2.949)	(4.671)	(4.523)	(11.830)
Observations	1,021	1,021	1,021	1,021
R^2	0.093	0.077	0.093	0.091
Adjusted R^2	0.093	0.076	0.092	0.090
Residual Std. Error (df = 1019)	51.105	80.958	78.380	205.021
F Statistic (df = 1; 1019)	105.068***	85.348***	104.647***	102.291***
Note:	*p<0.1; **p<0.05; ***p<0.01; (standard error)			

Table 6.6: Regression results explaining walking counts with accessibility to jobs via transit only (net transit), for a 30-minute threshold, during different time periods.

	Dependent variable:			
	ped_am_avg	ped_md_avg	ped_pm_avg	six_hour_count
	(1)	(2)	(3)	(4)
Net transit accessibility (30-minute)	0.161***	0.242***	0.247***	0.649***
	(0.037)	(0.058)	(0.056)	(0.147)
Constant	7.110**	7.554	10.164**	24.828*
	(3.351)	(5.266)	(5.138)	(13.425)
Observations	1,021	1,021	1,021	1,021
R^2	0.019	0.017	0.019	0.019
Adjusted R^2	0.018	0.016	0.018	0.018
Residual Std. Error (df = 1019)	53.175	83.562	81.539	213.034
F Statistic (df = 1; 1019)	19.232***	17.579***	19.268***	19.531***

Note: *p<0.1; **p<0.05; ***p<0.01; (standard error)

Table 6.7: Regression results explaining walking activity in terms of betweenness centrality, for different time periods.

	Dependent variable:			
	ped_am_avg	ped_md_avg	ped_pm_avg	six_hour_count
	(1)	(2)	(3)	(4)
Betweenness	0.512***	0.738***	1.002***	2.253***
	(0.124)	(0.195)	(0.189)	(0.495)
Constant	14.504***	19.147***	19.517***	53.169***
	(2.029)	(3.194)	(3.097)	(8.121)
Observations	1,056	1,056	1,056	1,056
R^2	0.016	0.013	0.026	0.019
Adjusted R^2	0.015	0.013	0.025	0.018
Residual Std. Error (df = 1054)	52.428	82.549	80.046	209.872
F Statistic (df = 1; 1054)	17.202***	14.409***	28.233***	20.755***
Note:	*p<0.1; **p<0.05; ***p<0.01; (standard error)			

Table 6.8: Spatial accessibility cross-correlation, 5-minute threshold

Economic sector 1	Economic sector 2	Correlation
Education 5min	Health care 5min	0.944
Construction 5min	Education 5min	0.904
Construction 5min	Health care 5min	0.901
Education 5min	Other 5min	0.864
Health care 5min	Other 5min	0.856
Construction 5min	Other 5min	0.850
Administrative 5min	Other 5min	0.714
Professional scientific technical 5min	Education 5min	0.675
Professional scientific technical 5min	Other 5min	0.668
Professional scientific technical 5min	Health care 5min	0.668
Construction 5min	Professional scientific technical 5min	0.619
Construction 5min	Administrative 5min	0.601
Administrative 5min	Health care 5min	0.566
Administrative 5min	Education 5min	0.550
Finance 5min	Arts entertainment 5min	0.495
Professional scientific technical 5min	Administrative 5min	0.455
Manufacturing 5min	Information 5min	0.402
Real estate 5min	Professional scientific technical 5min	0.356
Information 5min	Education 5min	0.333
Hospitality food 5min	Other 5min	0.315
Information 5min	Other 5min	0.289
Information 5min	Health care 5min	0.288
Information 5min	Professional scientific technical 5min	0.271
Information 5min	Administrative 5min	0.238
Retail trade 5min	Finance 5min	0.231
Construction 5min	Information 5min	0.223
Manufacturing 5min	Administrative 5min	0.215
Manufacturing 5min	Wholesale trade 5min	0.203
Wholesale trade 5min	Professional scientific technical 5min	0.190
Wholesale trade 5min	Administrative 5min	0.186
Construction 5min	Manufacturing 5min	0.175
Retail trade 5min	Arts entertainment 5min	0.173
Manufacturing 5min	Health care 5min	0.172
Finance 5min	Professional scientific technical 5min	0.172

Manufacturing 5min	Professional scientific technical 5min	0.166
Information 5min	Finance 5min	0.159
Manufacturing 5min	Other 5min	0.157
Real estate 5min	Other 5min	0.156
Real estate 5min	Administrative 5min	0.151
Manufacturing 5min	Education 5min	0.138
Professional scientific technical 5min	Arts entertainment 5min	0.138
Wholesale trade 5min	Retail trade 5min	0.137
Construction 5min	Transportation 5min	0.131
Wholesale trade 5min	Real estate 5min	0.130
Information 5min	Real estate 5min	0.127
Wholesale trade 5min	Transportation 5min	0.116
Construction 5min	Real estate 5min	0.105

Table 6.9: Spatial accessibility cross-correlation, 10-minute threshold

Economic sector 1	Economic sector 2	Correlation
Education 10min	Public administration 10min	0.925
Education 10min	Hospitality food 10min	0.846
Hospitality food 10min	Public administration 10min	0.801
Construction 10min	Professional scientific technical 10min	0.714
Management 10min	Other 10min	0.623
Wholesale trade 10min	Administrative 10min	0.613
Construction 10min	Other 10min	0.603
Manufacturing 10min	Wholesale trade 10min	0.576
Construction 10min	Management 10min	0.565
Health care 10min	Other 10min	0.551
Construction 10min	Manufacturing 10min	0.547
Finance 10min	Professional scientific technical 10min	0.534
Professional scientific technical 10min	Other 10min	0.468
Construction 10min	Wholesale trade 10min	0.468
Construction 10min	Administrative 10min	0.462
Manufacturing 10min	Professional scientific technical 10min	0.461
Professional scientific technical 10min	Administrative 10min	0.459
Wholesale trade 10min	Professional scientific technical 10min	0.449
Finance 10min	Real estate 10min	0.442

Management 10min	Health care 10min	0.438
Wholesale trade 10min	Real estate 10min	0.436
Real estate 10min	Hospitality food 10min	0.425
Finance 10min	Hospitality food 10min	0.422
Retail trade 10min	Arts entertainment 10min	0.420
Finance 10min	Administrative 10min	0.412
Professional scientific technical 10min	Management 10min	0.388
Manufacturing 10min	Information 10min	0.376
Real estate 10min	Public administration 10min	0.363
Professional scientific technical 10min	Health care 10min	0.345
Real estate 10min	Arts entertainment 10min	0.337
Real estate 10min	Professional scientific technical 10min	0.326
Hospitality food 10min	Other 10min	0.314
Manufacturing 10min	Administrative 10min	0.305
Retail trade 10min	Other 10min	0.305
Real estate 10min	Administrative 10min	0.305
Construction 10min	Retail trade 10min	0.301
Retail trade 10min	Real estate 10min	0.271
Finance 10min	Arts entertainment 10min	0.269
Construction 10min	Health care 10min	0.268
Arts entertainment 10min	Hospitality food 10min	0.262
Education 10min	Other 10min	0.257
Retail trade 10min	Hospitality food 10min	0.248
Manufacturing 10min	Other 10min	0.248
Transportation 10min	Information 10min	0.247
Retail trade 10min	Professional scientific technical 10min	0.228
Real estate 10min	Other 10min	0.225
Real estate 10min	Education 10min	0.222
Wholesale trade 10min	Finance 10min	0.220
Administrative 10min	Other 10min	0.219
Manufacturing 10min	Retail trade 10min	0.212
Utilities 10min	Administrative 10min	0.207
Other 10min	Public administration 10min	0.203
Retail trade 10min	Finance 10min	0.197
Agriculture 10min	Utilities 10min	0.194
Wholesale trade 10min	Retail trade 10min	0.188
Construction 10min	Transportation 10min	0.178

Wholesale trade 10min	Transportation 10min	0.176
Finance 10min	Education 10min	0.169
Construction 10min	Real estate 10min	0.168
Construction 10min	Finance 10min	0.164
Manufacturing 10min	Real estate 10min	0.162
Finance 10min	Other 10min	0.161
Wholesale trade 10min	Public administration 10min	0.157
Transportation 10min	Professional scientific technical 10min	0.157
Administrative 10min	Health care 10min	0.157
Utilities 10min	Wholesale trade 10min	0.153
Retail trade 10min	Administrative 10min	0.148
Information 10min	Finance 10min	0.141
Manufacturing 10min	Health care 10min	0.141
Finance 10min	Public administration 10min	0.139
Professional scientific technical 10min	Arts entertainment 10min	0.139
Professional scientific technical 10min	Hospitality food 10min	0.122
Manufacturing 10min	Transportation 10min	0.116
Information 10min	Real estate 10min	0.115
Manufacturing 10min	Management 10min	0.111
Wholesale trade 10min	Other 10min	0.104

Table 6.10: Spatial accessibility cross-correlation, 15-minute threshold

Economic sector 1	Economic sector 2	Correlation
Finance 15min	Management 15min	0.997
Finance 15min	Professional scientific technical 15min	0.979
Professional scientific technical 15min	Management 15min	0.972
Professional scientific technical 15min	Administrative 15min	0.953
Finance 15min	Arts entertainment 15min	0.920
Management 15min	Arts entertainment 15min	0.918
Finance 15min	Administrative 15min	0.914
Management 15min	Administrative 15min	0.902
Professional scientific technical 15min	Arts entertainment 15min	0.901
Administrative 15min	Arts entertainment 15min	0.816
Manufacturing 15min	Wholesale trade 15min	0.707
Construction 15min	Administrative 15min	0.697

Mining 15min	Wholesale trade 15min	0.691
Hospitality food 15min	Public administration 15min	0.669
Construction 15min	Professional scientific technical 15min	0.637
Information 15min	Administrative 15min	0.622
Administrative 15min	Other 15min	0.607
Real estate 15min	Other 15min	0.607
Information 15min	Professional scientific technical 15min	0.603
Professional scientific technical 15min	Other 15min	0.597
Information 15min	Finance 15min	0.596
Information 15min	Management 15min	0.593
Construction 15min	Wholesale trade 15min	0.582
Construction 15min	Finance 15min	0.579
Management 15min	Other 15min	0.575
Construction 15min	Other 15min	0.573
Mining 15min	Manufacturing 15min	0.569
Finance 15min	Other 15min	0.566
Education 15min	Public administration 15min	0.565
Construction 15min	Arts entertainment 15min	0.561
Construction 15min	Management 15min	0.560
Hospitality food 15min	Other 15min	0.556
Arts entertainment 15min	Other 15min	0.544
Information 15min	Arts entertainment 15min	0.525
Finance 15min	Hospitality food 15min	0.515
Professional scientific technical 15min	Hospitality food 15min	0.511
Arts entertainment 15min	Hospitality food 15min	0.510
Management 15min	Hospitality food 15min	0.503
Information 15min	Other 15min	0.486
Construction 15min	Manufacturing 15min	0.475
Education 15min	Hospitality food 15min	0.443
Administrative 15min	Hospitality food 15min	0.440
Wholesale trade 15min	Transportation 15min	0.430
Construction 15min	Information 15min	0.423
Construction 15min	Real estate 15min	0.419
Health care 15min	Other 15min	0.411
Construction 15min	Retail trade 15min	0.369
Retail trade 15min	Other 15min	0.358
Manufacturing 15min	Transportation 15min	0.348

Retail trade 15min	Real estate 15min	0.347
Wholesale trade 15min	Real estate 15min	0.340
Real estate 15min	Hospitality food 15min	0.337
Transportation 15min	Real estate 15min	0.325
Mining 15min	Transportation 15min	0.313
Mining 15min	Construction 15min	0.299
Manufacturing 15min	Retail trade 15min	0.298
Real estate 15min	Health care 15min	0.296
Wholesale trade 15min	Retail trade 15min	0.293
Professional scientific technical 15min	Health care 15min	0.292
Construction 15min	Hospitality food 15min	0.289
Management 15min	Health care 15min	0.286
Administrative 15min	Health care 15min	0.276
Construction 15min	Transportation 15min	0.275
Manufacturing 15min	Administrative 15min	0.271
Manufacturing 15min	Other 15min	0.270
Retail trade 15min	Transportation 15min	0.262
Health care 15min	Arts entertainment 15min	0.261
Information 15min	Hospitality food 15min	0.258
Finance 15min	Health care 15min	0.255
Mining 15min	Administrative 15min	0.252
Wholesale trade 15min	Other 15min	0.252
Real estate 15min	Administrative 15min	0.241
Wholesale trade 15min	Administrative 15min	0.237
Mining 15min	Utilities 15min	0.236
Manufacturing 15min	Information 15min	0.235
Utilities 15min	Wholesale trade 15min	0.217
Manufacturing 15min	Real estate 15min	0.210
Real estate 15min	Professional scientific technical 15min	0.200
Mining 15min	Real estate 15min	0.193
Real estate 15min	Arts entertainment 15min	0.190
Information 15min	Real estate 15min	0.189
Transportation 15min	Other 15min	0.188
Information 15min	Health care 15min	0.184
Utilities 15min	Manufacturing 15min	0.178
Other 15min	Public administration 15min	0.169
Real estate 15min	Public administration 15min	0.156

Manufacturing 15min	Professional scientific technical 15min	0.149
Health care 15min	Hospitality food 15min	0.147
Mining 15min	Other 15min	0.136
Real estate 15min	Management 15min	0.129
Education 15min	Other 15min	0.127
Utilities 15min	Administrative 15min	0.118
Transportation 15min	Hospitality food 15min	-0.117
Transportation 15min	Arts entertainment 15min	-0.117
Finance 15min	Real estate 15min	0.115
Retail trade 15min	Hospitality food 15min	0.112
Construction 15min	Health care 15min	0.109
Mining 15min	Hospitality food 15min	-0.108
Retail trade 15min	Arts entertainment 15min	0.104

6.3 Supplemental Code

Following is an example of the Python tools that were built to manage the various moving parts of this investigation - data, PostgreSQL instances, GIS shapefiles, and data analysis.

```
from time import sleep
import os
import psycopg2
from ConfigParser import SafeConfigParser
from aodb import AODB
import shapefile
from boto.s3.connection import S3Connection
import subprocess
import psutil
import time
import signal
import shlex
import glob
import numpy as np
import csv
import zipfile
from datetime import datetime
```

```
'''
Inputs: cbsaid

General outline of tasks for this script:
1. ssh into AODB machine
2. create postgres connection
3. for each cbsa, get a bounding box & feed that to the OSM fetcher
4. Insert osm.pbf location into XML file
5. build the CBSA's graph
6. run accessibility calculations
7. calculate walk accessibility for departure times 7-9am
7.1 dump results into database
***8. use worker_weighted_for_cbsa function in aodb.py to calculated
    ↪ worker-weighted accessibility for each 5-minute threshold

'''
config = SafeConfigParser()
config.read(os.path.expanduser("~/.aoconfig"))



def kill(proc_pid):
    process = psutil.Process(proc_pid)
    for proc in process.get_children(recursive=True):
        proc.kill()
    process.kill()

def create_tunnel(tunnel_cmd):
    ssh_process = subprocess.Popen(shlex.split(tunnel_cmd),
        ↪ universal_newlines=True,
                                                shell=False,
                                                stdout=subprocess.PIPE,
                                                stderr=subprocess.
                                                    ↪ STDOUT,
                                                stdin=subprocess.PIPE)

    # Assuming that the tunnel command has "-f" and "
        ↪ ExitOnForwardFailure=yes", then the
```

```
    # command will return immediately so we can check the return status
      ↪ with a poll().

    while True:
        p = ssh_process.poll()
        if p is not None: break
        time.sleep(1)

    if p == 0:
        # Unfortunately there is no direct way to get the pid of the
          ↪ spawned ssh process, so we'll find it
        # by finding a matching process using psutil.

        current_username = psutil.Process(os.getpid()).username()
        ssh_processes =[]
        i=0
        for proc in psutil.get_process_list():
            try:
                if proc.cmdline() == tunnel_cmd.split() and proc.
                  ↪ username() == current_username:
                    ssh_processes.append(proc.pid)
                    i=i+1
            except:
                pass

        if len(ssh_processes) == 1:
            return ssh_processes[0], ssh_process
        else:
            raise RuntimeError, 'multiple (or zero?) tunnel ssh
              ↪ processes found: ' + str(ssh_processes)
    else:
        raise RuntimeError, 'Error creating tunnel: ' + str(p) + ' :: '
          ↪ + str(ssh_process.stdout.readlines())


def match_cbsaid(name, cbsaid):
    if cbsaid is None:
```

```
        return True
    else:
        return name[0:5] == cbsaid

def get_cbsas():
    cbsaid=None
    os.chdir('/Users/brendanmurphy/Documents/North America db/
        ↪ origin_zones')
    ozones = shapefile.Reader("origin_zones")
    oids = [x[0] for x in ozones.records()]

    c = S3Connection(config.get("boto","aws_access_key_id"),config.get("
        ↪ boto","aws_secret_access_key"))
    bucket = c.get_bucket('ao-results')
    S3_files = [x.name.encode('utf-8') for x in bucket.list() if
        ↪ match_cbsaid(x.key,cbsaid)]

    cbsas=[]
    for oid in oids:
        matches = [x for x in S3_files if x.startswith(oid)]
        if len(matches) > 0:
            cbsa = oid[:5]
            cbsas.append(cbsa)
    cbsalist = list(set(cbsas))
    return cbsalist, oids

def get_OSM(self,cbsa,cur):
    #os.chdir('/Users/brendanmurphy/Dropbox/accessibility Observatory
        ↪ research/OpenTripPlanner/OSMs')
    os.chdir('/Users/brendanmurphy/Dropbox/accessibility Observatory
        ↪ research/OpenTripPlanner/OSMtesting')

    s = 5000 #average human walking speed in meters/h
    d = s * 1 #meters of buffer for 1 hour radial walking

    query = '''
```

```
  SELECT ST_XMin(ST_Buffer_meters(geom,{0})), ST_YMin(
     ↪ ST_Buffer_meters(geom,{0})), ST_XMax(ST_Buffer_meters(geom
     ↪ ,{0})), ST_YMax(ST_Buffer_meters(geom,{0}))
  FROM census.cbsas
  WHERE cbsafp10 = '%s'
  '''.format(d) % cbsa

cur.execute(query)
left, bottom, right, top = cur.fetchone()

print left, bottom, right, top

#shrink by 75% for testing
top = top - .75*(top-bottom)/2
bottom = bottom - .75*(bottom-top)/2
left = left - .75*(left-right)/2
right = right - .75*(right-left)/2

print left, bottom, right, top

#command = 'osmosis --read-pgsql host={}:{} database={} user={}
    ↪ password={} --dataset-bounding-box left={} right={} top={}
    ↪ bottom={} completeWays=yes --write-pbf omitmetadata=true file
    ↪ ={}'.format(self.config.get("aodb","host"), 5900, self.config.
    ↪ get("aodb","dbname"), self.config.get("aodb","user"), self.
    ↪ config.get("aodb","password"), left, right, top, bottom, '{0}.
    ↪ osm.pbf'.format(cbsa))
command = 'osmosis --read-pgsql host={}:{} database={} user={}
    ↪ password={} --dataset-bounding-box left={} right={} top={}
    ↪ bottom={} --write-xml file={}'.format(self.config.get("aodb","
    ↪ host"), 5900, self.config.get("aodb","dbname"), self.config.get
    ↪ ("aodb","user"), self.config.get("aodb","password"), left,
    ↪ right, top, bottom, '{0}.osm'.format(cbsa))
print command
#wait=input('test')

p = subprocess.Popen(command, shell=True)
```

```
    p.communicate()

def merge_OSM(cbsa,zones):
    command = 'osmosis '
    for zone in zones:
        zonefile = '{0}.osm.pbf'.format(cbsa)
        command = command.append('--rx {0}'.zonefile)
    command = command.append(' --m --wx {0}_merged.osm'.format(cbsa))

    #run the bash command
    print command
    p1 = subprocess.Popen(command,stdout=subprocess.PIPE, stderr=
        ↪ subprocess.PIPE, shell=True)
    for item in iter(p1.stdout.readline, ''):
        print item

def build_graph(cbsa):
    os.chdir('/Users/brendanmurphy/Dropbox/accessibility Observatory
        ↪ research/OpenTripPlanner')
    graphconfig = open('walk-graph-config.xml', "r")
    lines = graphconfig.readlines()
    lines[23]='          <property name="path" value="OSMs/{0}.osm.pbf" />
        ↪ <!--input your OSM file here mpls-stpaul-20140416.osm.pbf-->\n'
        ↪ .format(cbsa)
    lines[41]='    <property name="path" value="graphs/graph_{0}" /> <!--
        ↪ specify output location of graph OBJ -->\n'.format(cbsa)

    newgraphconfig = open("{}_walk-graph-config.xml".format(cbsa),"w")
    newgraphconfig.writelines(lines)
    newgraphconfig.close()
    graphconfig.close()

    command = 'java -Xmx8G -cp otp-ao-0.11.x.jar org.opentripplanner.
        ↪ graph_builder.GraphBuilderMain {}_walk-graph-config.xml'.format
        ↪ (cbsa)
    p2 = subprocess.Popen(command,stdout=subprocess.PIPE, stderr=
        ↪ subprocess.PIPE,shell=True)
```

```
  for item in iter(p2.stdout.readline, ''):
    print item

def fetch_shapefiles(self,cbsa):
  os.chdir('/Users/brendanmurphy/Dropbox/accessibility Observatory
    ↪ research/OpenTripPlanner/cbsa_jobs_shapefiles')
  #get the jobs data for a given CBSA, and write it to a Shapefile
  #query for aodb on Andrew's machine
  """query = '''
    CREATE VIEW temp AS SELECT id, COALESCE( alljobs, 0) as jobs,
      ↪ COALESCE(lowjobs, 0) as lowjobs, COALESCE(medjobs,0) as
      ↪ medjobs, COALESCE(highjobs,0) as highjobs, centroid FROM
                    (   (   SELECT id, centroid
                            FROM zones.blocks
                            WHERE ozid LIKE '{}%'
                            ) b
                        LEFT JOIN
                        (   SELECT geocode, c000 AS alljobs, CE01 AS
                          ↪ lowjobs, CE02 AS medjobs, CE03 AS
                          ↪ highjobs
                            FROM lehd.wac2011 ) l
                        ON b.id = l.geocode );
  '''.format(cbsa)
   query for Boston"""
  query = '''
    CREATE VIEW temp AS SELECT geoid10, COALESCE(jobs, 0) as jobs,
      ↪ centroid FROM
                    (   (   SELECT geoid10, centroid
                            FROM census.blocks
                            WHERE cbsa='{}'
                            ) b
                        LEFT JOIN
                        (   SELECT geoid, jobs
                            FROM smartlocation.ma ) l
                        ON b.geoid10 = l.geoid );

  '''.format(cbsa)
```

```
#get the origins
"""query = '''
  CREATE or replace VIEW temp AS SELECT geoid10, COALESCE(alljobs, 0)
     ↪ as jobs, COALESCE(lowjobs, 0) as lowjobs, COALESCE(medjobs
     ↪ ,0) as medjobs, COALESCE(highjobs,0) as highjobs, centroid
     ↪ FROM
                    (    (   SELECT geoid10, centroid
                             FROM census.blocks
                             WHERE cbsa='{}'
                             ) b
                         LEFT JOIN
                         (   SELECT geocode, c000 AS alljobs, CE01 AS
                            ↪ lowjobs, CE02 AS medjobs, CE03 AS
                            ↪ highjobs
                             FROM lehd.wac2011 ) l
                         ON b.geoid10 = l.geocode );'''.format(cbsa)"""

try:
  cur.execute(query)
  conn.commit()
  print 'created view'
  #wait=input('test)')
except psycopg2.Error as e:
  print str(e)
  pass

#command = 'pgsql2shp -f {}_jobs -h {} -p {} -u {} -P {} {} "{}"'.
   ↪ format(cbsa,self.config.get("aodb","host"),5900,self.config.get
   ↪ ("aodb","user"),self.config.get("aodb","password"),self.config.
   ↪ get("aodb","dbname"),'temp')
command = 'pgsql2shp -f {}_origins -h {} -p {} -u {} -P {} {} "{}"'.
   ↪ format(cbsa,self.config.get("aodb","host"),5900,self.config.get
   ↪ ("aodb","user"),self.config.get("aodb","password"),self.config.
   ↪ get("aodb","dbname"),'temp')
print command
p = subprocess.Popen(command, shell=True)
```

```
p.communicate()

try:
  query = 'DROP VIEW temp'
  cur.execute(query)
  conn.commit()
  print 'dropped view'
except psycopg2.Error, e:
  print str(e)
  pass



#now create a shapefile for destinations, with a 5km buffer
"""query = '''
  CREATE or replace VIEW temp2 AS SELECT geoid10, COALESCE(alljobs,
    ↪ 0) as jobs, COALESCE(lowjobs, 0) as lowjobs, COALESCE(medjobs
    ↪ ,0) as medjobs, COALESCE(highjobs,0) as highjobs, centroid
    ↪ FROM
                  (   (   SELECT geoid10, centroid
                          FROM census.blocks
                          WHERE cbsaid_destination='{}'
                          ) b
                      LEFT JOIN
                      (   SELECT geocode, c000 AS alljobs, CE01 AS
                          ↪ lowjobs, CE02 AS medjobs, CE03 AS
                          ↪ highjobs
                          FROM lehd.wac2011 ) l
                      ON b.geoid10 = l.geocode );
'''.format(cbsa)"""



#destinations for Boston
query = '''
  CREATE VIEW temp2 AS SELECT geoid10, COALESCE(jobs, 0) as jobs,
    ↪ centroid FROM
                  (   (   SELECT geoid10, centroid
                          FROM census.blocks
```

```
                            WHERE cbsaid_destination='{}'
                            ) b
                        LEFT JOIN
                        (   SELECT geoid, jobs
                            FROM smartlocation.ma ) l
                        ON b.geoid10 = l.geoid );'''.format(cbsa)



    try:
        cur.execute(query)
        conn.commit()
        print 'created view'
    except psycopg2.Error as e:
        print str(e)
        pass

    command = 'pgsql2shp -f {}_destinations -h {} -p {} -u {} -P {} {}
        ↪ "{}"'.format(cbsa, self.config.get("aodb","host"),5432,self.
        ↪ config.get("aodb","user"),self.config.get("aodb","password"),
        ↪ self.config.get("aodb","dbname"),'temp2')
    print command
    p = subprocess.Popen(command, shell=True)
    p.communicate()

    try:
        query = 'DROP VIEW temp2'
        cur.execute(query)
        conn.commit()
        print 'dropped view'
    except psycopg2.Error, e:
        print str(e)
        pass


def calc_accessibility(cbsa):
```

```
    os.chdir('/Users/brendanmurphy/Dropbox/accessibility Observatory
        ↪ research/OpenTripPlanner')
    analystconfig = open('analyst_config_walk.xml',"r") #
        ↪ analyst_config_walk.xml2
    lines = analystconfig.readlines()
    lines[14] = '      <BEANS:property name="defaultRouterId" value="graph_
        ↪ {}"/>\n'.format(cbsa)
    lines[19] = '      <BEANS:property name="sourceFilename" value="
        ↪ cbsa_jobs_shapefiles/{}_origins.shp"/> <!--shapefile loading
        ↪ goes here -->\n'.format(cbsa) #origins
    lines[24] = '      <BEANS:property name="sourceFilename" value="
        ↪ cbsa_jobs_shapefiles/{}_destinations.shp"/> <!--shapefile
        ↪ loading goes here -->\n'.format(cbsa) #destinations
    lines[79] = '      <BEANS:property name="outputPath" value="
        ↪ cbsa_results/{}_wa_2014_0700-0700.csv"/> <!-- update output
        ↪ -->\n'.format(cbsa) #79

    newanalystconfig = open("{}_analyst_config_walk.xml".format(cbsa),"w"
        ↪ )
    newanalystconfig.writelines(lines)
    newanalystconfig.close()
    analystconfig.close()

    command = 'java -Xmx12G -jar ao-otp-analyst.jar {}
        ↪ _analyst_config_walk.xml'.format(cbsa)
    print 'Calculating accessibility results for cbsa {}'.format(cbsa)
    p = subprocess.Popen(command, shell=True)
    p.communicate()

def copy_to_db(cbsa):
    os.chdir('/Users/brendanmurphy/Dropbox/accessibility Observatory
        ↪ research/OpenTripPlanner/cbsa_results')
    datastring = os.path.join(os.getcwd(),'%s*.csv' % cbsa)
    numfiles = len(glob.glob(datastring))
    print ("The total number of files to process is %s" % numfiles)

    for item in glob.glob(datastring):
```

```
    try:
      #os.system('scp %s murph677@134.84.148.4:/Users/murph677/cbsa_results/' % item.replace(' ','\ '))
      #query="COPY results.walk FROM '/Users/murph677/cbsa_results/%s_wa_2014_0700-0700.csv' DELIMITER ',' CSV HEADER" % cbsa
      query = "COPY results.walk FROM '/Users/brendanmurphy/Dropbox/Accessibility Observatory research/OpenTripPlanner/cbsa_results/%s_wa_2014_0700-0700.csv' DELIMITER ',' CSV HEADER" % cbsa
      #print query
      cur.execute(query)
      #cur.copy_from(fopen,'results.walk',columns=('label','deptime','threshold','JOBS'))
      #conn.commit()
      conn.commit()

    except psycopg2.Error, e:
      pass
      print str(e)

def worker_weighted_for_cbsa(cbsa, thresholds=[5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60], firsthour=700, lasthour=700):
  os.chdir('/Users/brendanmurphy/Dropbox/accessibility Observatory research/OpenTripPlanner')
  weighted_avgs=[]
  for threshold in [30]:
    print("Calculating {}-minute worker-weighted average accessibility for CBSA {}...".format(threshold, cbsa))
    query = """ SELECT COALESCE(workers, 0) as workers, COALESCE(jobs, 0) as jobs FROM
          ( ( SELECT geoid10
          FROM census.blocks
          WHERE cbsa='{}' ) b
          LEFT JOIN
          ( SELECT geocode, c000 AS workers
          FROM lehd.rac2011 ) l
          ON b.geoid10 = l.geocode ) j
```

```
            LEFT JOIN
            ( SELECT blockid , jobs
            FROM results._12580test
            WHERE threshold = {} ) r
            ON j.geoid10 = r.blockid; """.format(cbsa, threshold*60)


      print "running the {}-minute query for CBSA {}".format(threshold,
         ↪ cbsa)
      print query
      #wait=input('test')
      cur.execute(query)
      results = cur.fetchall()
      try:
        weights, values = zip(*results)
        w = list(weights)
        v = list(values)
        weighted_avg = np.average(v, weights=w)
        print weighted_avg
        weighted_avgs.append(weighted_avg)
      except TypeError, e:
        print str(e)
        #return None
    results = [cbsa] + weighted_avgs
    fopen=open('cbsa_12580_test.csv','a')
    writer = csv.writer(fopen,delimiter=',')
    writer.writerow(results)
    fopen.close()
    print results
    return results

def get_blockgroups(cbsa):
    query = """SELECT left(geoid10,12) FROM census.blocks b
          where b.cbsa='{}'""".format(cbsa)
    try:
      begin=datetime.now()
      cur.execute(query)
```

```
    results=cur.fetchall()
    elapsed = datetime.now() - begin
    print("Fetched blockgroups for CBSA {} in {}".format(cbsa,elapsed))
  except psycopg2.Error,e:
    print str(e)
  return results

def worker_weighted_for_bgs(cbsa, thresholds=[5, 10, 15, 20, 25, 30,
    ↪ 35, 40, 45, 50, 55, 60]):
  os.chdir('/Users/brendanmurphy/Dropbox/accessibility Observatory
      ↪ research/OpenTripPlanner')
  start=datetime.now()
  agg_jobs=[]
  for threshold in thresholds:
    begin = datetime.now()
    print("Calculating {}-minute worker-weighted average
        ↪ accessibilities at the blockgroup level for CBSA {}...".
        ↪ format(threshold, cbsa))
    query = """SELECT left (geoid10,12), CASE when sum(COALESCE(workers
        ↪ ,0))=0 THEN 0 ELSE sum(COALESCE(workers, 0)*COALESCE(walkjobs
        ↪ ,0))/sum(COALESCE(workers,0)) END FROM
           (( ( SELECT geoid10
             FROM census.blocks
             WHERE cbsa = '{0}' ) b
             LEFT JOIN
             ( SELECT geocode, c000 AS workers
             FROM lehd.rac2011 ) l
             ON b.geoid10 = l.geocode ) j
             LEFT JOIN
             ( SELECT blockid, jobs as walkjobs
             FROM results.walk
             WHERE threshold = {1} ) r
             ON j.geoid10 = r.blockid) as foo
             GROUP BY left(geoid10,12)
             ORDER BY left(geoid10,12); """.format(cbsa, threshold*60)
    print query
    cur.execute(query)
```

```
    results = cur.fetchall()
    try:
      blockgroups,avgjobs = zip(*results)
      #print blockgroups, avgjobs
      #wait=input('test')
      agg_jobs.append(avgjobs)
      elapsed=datetime.now() - begin
      print("Calculated blockgroup weighted averages in {} for
        ↪ threshold {} minutes".format(elapsed,threshold))
    except TypeError,e:
      print str(e)

  results = map(list,zip(*[blockgroups]+agg_jobs))
  #print results
  fopen=open('blockgroup_walk_accessibility_{}.csv'.format(cbsa),'a')
  writer = csv.writer(fopen,delimiter = ',')
  writer.writerow(['Blockgroup ID','5min','10min','15min','20min','25
    ↪ min','30min','35min','40min','45min','50min','55min','60min'])
  writer.writerows(results)
  fopen.close()

def total_jobs_for_cbsa(cbsa, thresholds=[5, 10, 15, 20, 25, 30, 35,
  ↪ 40, 45, 50, 55, 60]):
  os.chdir('/Users/brendanmurphy/Dropbox/accessibility Observatory
    ↪ research/OpenTripPlanner')
  total_jobs = []
  for threshold in thresholds:
    print('Fetching {}-minute total accessibility for CBSA {}...'.
      ↪ format(threshold,cbsa))
    query = """ SELECT SUM(COALESCE(jobs,0)) as jobs FROM
          (  (   SELECT id
           FROM zones.blocks
           WHERE ozid like '{}%' ) b
           LEFT JOIN
           ( SELECT blockid, jobs
             FROM results.walk
             WHERE threshold = {} ) r
```

```
            ON b.id = r.blockid);""".format(cbsa,threshold*60)
        cur.execute(query)
        result = cur.fetchall()
        jobs = result[0][0]
        print jobs
        total_jobs.append(jobs)
    results = [cbsa] + total_jobs
    fopen = open('cbsa_total_jobs.csv','a')
    writer = csv.writer(fopen,delimiter=',')
    writer.writerow(results)
    fopen.close()
    print results
    return results


def copy_results_table(cbsa):
    command = 'pg_dump -t accessibility_results_{0} usadb | psql aodb'.
        ↪ format(cbsa) #accessibility_avgs_{0}
    p = subprocess.Popen(command, shell=True)
    p.communicate()

def rectangularize(cbsa):
    os.chdir('/Users/brendanmurphy/Dropbox/Accessibility Observatory
        ↪ research/OpenTripPlanner/cbsa_results/rectangularized_walk')
    #get the state FIPS codes
    query = "select distinct on (left(geoid10,2)) left(geoid10,2) from
        ↪ census.blocks where cbsa='{0}'".format(cbsa)
    cur.execute(query)
    results = cur.fetchall()
    statecodes = [i[0] for i in results]
    statecodes = str(map(str,statecodes))
    statecodes = statecodes.replace('[','(').replace(']',')')

    #create view
    query = """create view rectangularize as
            with data as (select blockid,threshold,jobs from results.walk
            where left(blockid,2) IN {0})
```

```
        select b.geoid10, r.jobs as jobs_5min, r1.jobs as jobs_10min,
           ↪ r2.jobs as jobs_15min, r3.jobs as jobs_20min,
            r4.jobs as jobs_25min, r5.jobs as jobs_30min, b.geom from
              ↪ census.blocks b
        left join (select*from data
             where threshold=300) r
        on b.geoid10=r.blockid
        left join (select*from data
             where threshold=600) r1
        on b.geoid10=r1.blockid
        left join (select*from data
             where threshold=900) r2
        on b.geoid10=r2.blockid
        left join (select*from data
             where threshold=1200) r3
        on b.geoid10=r3.blockid
        left join (select*from data
             where threshold=1500) r4
        on b.geoid10=r4.blockid
        left join (select*from data
             where threshold=1800) r5
        on b.geoid10=r5.blockid
        where b.cbsa='{1}'
""".format(statecodes,cbsa)

query = """create view rectangularize as
        with data as (select blockid,threshold,jobs from results.walk
        where left(blockid,2) IN {0})
        select b.geoid10, r.jobs as jobs_tot, b.geom from census.blocks
           ↪ b
        left join (select*from data
             where threshold=1800) r
        on b.geoid10=r.blockid
        where b.cbsa='{1}'
""".format(statecodes,cbsa)
cur.execute(query)
conn.commit()
```

```
#export the csv
query = """
    copy (select geoid10, jobs_5min, jobs_10min, jobs_15min,
      ↪ jobs_20min, jobs_25min, jobs_30min from rectangularize)
    to '/Users/brendanmurphy/Dropbox/Accessibility Observatory
      ↪ research/OpenTripPlanner/cbsa_results/rectangularized_walk
      ↪ /{0}_wa_2014_0700-0700.csv' delimiter ',' CSV HEADER
    """.format(cbsa)

query = """
    copy (select geoid10, jobs_tot from rectangularize)
    to '/Users/brendanmurphy/Dropbox/Accessibility Observatory
      ↪ research/OpenTripPlanner/cbsa_results/rectangularized_walk
      ↪ /{0}_wa_2014_0700-0700.csv' delimiter ',' CSV HEADER
    """.format(cbsa)
cur.execute(query)
print 'Exported csv for CBSA {0}'.format(cbsa)

#export the shapefile
command = 'pgsql2shp -f {}_wa_2014_0700-0700 -h {} -p {} -u {} {}
   ↪ "{}"'.format(cbsa,'/tmp/',5432,'brendanmurphy','aodb','
   ↪ rectangularize')
print command

p = subprocess.Popen(command, shell=True)
p.communicate()
print 'Exported shapefile for CBSA {0}'.format(cbsa)

#zip up the shapefile
zfile = zipfile.ZipFile("{}_wa_2014_0700-0700.zip".format(cbsa),"w")
for i in glob.glob("{}_wa_2014_0700-0700.*".format(cbsa)):
  print i[-3:]
  if i[-3:]!='csv' and i[-3:]!='zip':
    zfile.write(i)
zfile.close()
print 'Shapefile zipped for CBSA {0}'.format(cbsa)
```

```
    #dump the view
    query = 'drop view rectangularize'
    cur.execute(query)
    conn.commit()

if __name__ == '__main__':
    tunnel_cmd = 'ssh -o BatchMode=yes -o ServerAliveInterval=1 -o
        ↪ ServerAliveCountMax=5 -f -o ExitOnForwardFailure=yes -N -L
        ↪ 5900:localhost:5432 murph677@134.84.148.4'

    tunnel=0 #set flag to open SSH tunnel or not
    if tunnel == 1:
        try:
            ssh_tunnel_process_pid, ssh_process = create_tunnel(tunnel_cmd)
            print 'made the tunnel at process ID {0}...'.format(
                ↪ ssh_tunnel_process_pid)

            while True:
                try:
                    aodb=AODB()
                    conn = AODB._con(aodb)
                    break
                except psycopg2.OperationalError as e:
                    print e
                    sleep(3)

            cur = conn.cursor()
            print 'made the AODB cursor and connection'

        except Exception,e:
            print str(e)

            #con = psycopg2.connect(database='usadb',user='brendanmurphy',
                ↪ host='/tmp/')
            #cur = conn.cursor()
```

```
conn = psycopg2.connect(database='aodb',user='brendanmurphy',host='/
    ↪ tmp/')
cur = conn.cursor()
cbsas, oids = get_cbsas()
#aodb=AODB()
#results=[]
print cbsas

for cbsa in cbsas:
  try:
    print cbsa
    #get_OSM(aodb,cbsa,cur)
    #print 'Downloaded OSM data for cbsa {}'.format(cbsa)
    #merge_OSM(cbsa,zones)
    #build_graph(cbsa)
    #fetch_shapefiles(aodb,cbsa)
    #calc_accessibility(cbsa)
    #copy_to_db(cbsa)
    #copy_results_table(cbsa)
    #worker_weighted_for_cbsa(cbsa)
    #worker_weighted_for_bgs(cbsa)
    #total_jobs_for_cbsa(cbsa)
    rectangularize(cbsa)
  except Exception,e:
    print str(e)
if tunnel==1:
  kill(ssh_tunnel_process_pid)
  print 'terminated the tunnel'
if conn:
  conn.close()
```