Forecasting district-wide pedestrian volumes in multi-level networks in high-density mixed-use areas

Michael A.B. van Eggermond[1,2], Panos Mavros[2], Alex Erath[1,2], Veera Helle[3,2], Pablo Acebillo[4,2], and Shuchen Xu[5,2]

[1]University of Applied Sciences Northwestern Switzerland (FHNW), Muttenz, Switzerland
[2]Future Cities Laboratory, Singapore ETH Centre, Singapore
[3]verkehrsteiner AG, Bern, Switzerland
[4]enCity, Singapore
[5]To be confirmed

March 2022

Abstract

This paper is concerned with improvements in the forecasting of pedestrian flows in multilevel pedestrian networks in high-density urban environments. 3D network topology measures are combined with land-use data, and validated against extensive pedestrian counts, to provide both evidence for the applicability of network analysis in tropical metropolises, as well as a calibrated tool for urban planners.

The research focuses on four area in Singapore. These areas have in common that they all are prominent transport hubs, but differ in surrounding land-use types and dominant network topology (e.g. indoor, outdoor, above ground, below ground, at grade). Multi-level pedestrian networks were drawn based on OpenStreetMap, include sidewalks on both sides of major roads for a radius up to 2 kilometres from the site centroids. Spatial network analysis was performed using sDNA which allows vertical networks to generate measures describing the spatial configuration of the network.

Subsequently, pedestrian counts were conducted during three consecutive days. In total, counts were conducted at more than 250 locations in 2018 and 2019, well before the global COVID19 pandemic.

Pedestrian flows are set against a series of variables, including pedestrian attractors and generators (e.g. shops, offices, hotels, dwellings), and variables describing the spatial configuration of the network, using advanced regression models.

Our results show that betweenness metrics (i.e. space syntax *choice*) combined with land-use yield high predictive power. Dependent on the study site, network metrics based on angular distance outperform those based on metric distance or perceived link distance.

This research demonstrates that is necessary to account for the multi-level nature of networks, and that indoor flows through private developments cannot be neglected, in particular when planning for integrated transport developments. The paper concludes with recommendations and implications for practice.

1 Introduction

Background Understanding what are the determinants of pedestrian flows can support urban and transport planners and inform the planning of future urban developments,. On one hand, the configuration of the urban network itself is the main generator of movement patterns in the city by and influences the routes that are potentially more frequently used for various trips (Hillier *et al.*, 1993). On other hand, the size, density, type and distribution of origins, destinations and their spatial configuration also influence pedestrian flows. Different methods and covariates have been used to determine how many pedestrians might walk along a walkway (link-based pedestrian volume), varying from methods including proximity to population and employment (e.g. Griswold *et al.*, 2019), Space Syntax measures (e.g. Cambra *et al.*, 2017) and network topology measures that take into account land-use (e.g. Cooper *et al.*, 2019).

New district-level urban developments in South East Asia are increasingly characterised by the fusion of indoor and outdoor, underground, at-grade, elevated, private and public pedestrian networks. As compared to outdoor environments, the configuration and dimension of indoor environments are difficult to change after construction, partially because of constraints resulting from the built environment and partially because of constraints in ownership. At the same time, measures of pedestrian flow typically include only the (outdoor) urban space (e.g. streets and sometimes pavement). However, in regions characterized by high temperatures and high humidity, traversing indoor walkways is an attractive alternative to using outdoor walkways, and such walkways have indeed been constructed, and are often encouraged by planning authorities. Therefore, it is necessary to better understand the determinants of route choice —and resulting pedestrian volumes —in both indoor, privatized, environments and the outdoor public pedestrian network.

This paper sets out to assess different metrics to predict pedestrian volumes in complex multilevel environments, starting with 'classic' space syntax metrics (e.g. angular choice), and moving further enhancing the predictive capabilities of such measures with additional information about land-use, gross floor area (GFA) and population density.

The study is validated with four case-study sites from Singapore. Singapore is located just one degree north of the equator in Southeast Asia with a land area of 712 km^2 and a total population of 5.9 million in 2021. GNI per capita amounts to US$ 86,480 (2020). Singapore has witnessed tremendous growth since its independence in 1965; a key to economic growth in Singapore has been the formulation and execution of land-use and transportation policies in an integrated manner, with a long-term vision (e.g. Huat, 2011; Richmond, 2008). One characteristic of integrated transport land-use planning in Singapore is the widespread development of transport-oriented neighbourhoods, in which stations of the Mass Rapid Transit, or MRT (Singapore's rapid transit system) , are surrounded and integrated with high-density commercial and residential developments. Often, these developments are characterised the vertical separation of traffic, where pedestrian connections between buildings are made through elevated or underground pedestrian walkway (bridges and/or tunnels) that can span entire precincts as described later in this paper. However, it is unclear whether climatic (hot and humid weather conditions throughout the year) and urban design (fusion of indoor and outdoor, multilevel pedestrian networks) skew the distribution of pedestrian flows towards certain walkways, or if space syntax analysis has similar forecasting abilities.

Contributions The study at hand seeks to overcome several issues that have not been taken into by previous studies. First, this study has set out to map pedestrian networks in the public realm and in private realm, perform pedestrian counts indoor and outdoor and quantify pedestrian demand models. Second, the research is situated in environments with multi-level pedestrian networks; networks have been mapped for all publicly accessible floor levels and pedestrian counts have been conducted below grade, at grade and above grade. Finally, analysis across four distinct sites allows for an elaborate comparison of the different methods used.

2 Related work

In this section we briefly review related work from the field of urban planning, transport planning, space syntax and spatial network analysis. We focus our attention on the progression from direct demand models to graph-theoretic approaches , in especially space syntax, and the increasing predictive ability afforded by taking into account behavioural constraints such as distance, and spatial characteristics, such as land-use.

Direct demand models Traditional approaches to forecast demand for motorized and public transport can be described as *four-step* transport demand models; these consist of four steps: (1) trip generation, (2) trip distribution (destination choice), (3) mode choice and (4) trip assignment (route choice). This type of models are constrained by the need to have exhaustive data to capture demand (1) an route choice (4) – typically using multi-attribute discrete-choice approaches. By contrast, instead of estimating pedestrian demand through an exhaustive demand generation process, direct demand models aim to estimate pedestrian flows as a function of socio-demographic and built environment variables (Cooper *et al.*, 2019). These flows are either estimated for intersections and/or segments. Kuzmyak *et al.* (2014) lists an exhaustive set of variables typically employed in direct demand models: population and employment density, possible differentiated by age or employment type), population or employment activity levels within a buffer distance from the pedestrian link, land-use mix, link characteristics, vehicular traffic (e.g. speed, composition), transit availability and proximity to points of interest. These variables largely coincide with the 5D's that have been used to analyse the relationship between the built environment and travel behaviour: density, diversity, design, destination accessibility and distance to transit (Ewing & Cervero, 2010). These built environment characteristics are spatially aggregated using Euclidean or network buffers and subsequently related to pedestrian flows. Indeed, when examining the variables used to explain pedestrian flows in studies reviewing direct demand models (Kuzmyak *et al.*, 2014; Munira & Sener, 2017), many studies incorporate aggregate spatial measures that describe the pedestrian's environment rather than describing a link's properties and a segments role in the larger network. By doing so, for links, and to a lesser extent intersections, that are not relevant in pedestrian's route choices, a higher flow will be predicted than might be actually the case. This happens as less-frequented links adjacent to links with a high footfall will exhibit similar characteristics.

Graph theoretic approaches Recognizing that spatial measures describing network topology can aid in understanding pedestrian flows, research has set out to describe a pedestrian network's properties. A distinction can be made between approaches solely considering the pedestrian network, and approaches accounting for network topology as well as population, employment and land-use.

One common metric is *betweenness centrality*. This is a measure that, for every link in a network, sums up the number of shortest paths that from everywhere to everywhere (Freeman, 1977): this definition solely relies on the characteristics of the network and the definition of distance to determine the shortest path.

During the last decades, an extensive set of network topology measures have been developed under the theory of *Space Syntax*. As a starting point, space syntax uses the notion *Natural Movement*: individuals prefer to move along visibility axes (Hillier *et al.*, 1993), that can be approximated as *axial lines*. An implication of this starting point is that, at least in earlier applications, the notion of metric distance was discarded and, rather, the number of turns, referred to as steps, needed to reach a destination was used as a measure of distance between any two locations. The number of links that an individual can reach within a number of turns, or steps, is referred to as *depth*.

Integration measures the amount of turns needed from a street segment, to reach all other street segments in the network. *Choice* is similar to the aforementioned betweenness centrality. Instead of counting the number of links, it is possible to use custom link weights or link length as a proxy thereof. Initially designated networks were drawn for Space Syntax analysis, called axial maps. Axial maps described visibility axes and/or axes of potential lines of movement. These maps were criticized for a lack of transparency and a certain amount of arbitrariness (Ratti, 2004). In response, new methods were developed that take into account both metric, angular and topological distance; furthermore, these were applied on street centre-lines that are more readily available in transport planning (Turner, 2007). Space syntax measures have been used in direct demand models to predict pedestrian flow (Desyllas *et al.*, 2003; Lerman *et al.*, 2014; Porta *et al.*, 2009).

Distance metrics Implicit in all routing applications for transport planning is the concept of distance, which is used to compute the shortest path between any two points. As discussed above *distance* can refer to metric, angular or topological measures. Typically in space syntax analysis one type of distance is used to analyse spatial networks, e.g. using angular integration. Researchers have previously combined analyses that rely on different types of distance and/or radii, by merging the results (i.e. the network measures). More recently, Cooper *et al.* (2019) introduced the concept of hybrid distance: the distance between an origin and a destination can be based on any combination of metric distance and euclidean distance. Likewise, *distance* can incorporate incorporate slope, the presence of vehicular traffic or the availability of certain infrastructure (Cooper, 2017) as additional weights to be consider when estimating the shortest path. In these cases, the costs to traverse a link are translated to generalized costs, a concept well-known in transport planning, where travel time can include components that include congestion, crowding, in-vehicle time, waiting time and number of transfers. All these components are weighted differently and can depend on trip purpose and socio-demographic characteristics. For pedestrians, generalized costs consist of distance (Guo & Loo, 2013; Muraleetharan & Hagiwara, 2007; Rodriguez *et al.*, 2014), turns (Lue & Miller, 2019), greenery (Rodriguez *et al.*, 2014) as well as commercial activity (Rodriguez *et al.*, 2014).

Improving models for TODs in the tropics Tropical metropolises like Singapore and Hong Kong are characterised climatic and spatial conditions that have encouraged the development of multilevel pedestrian networks that effectively fuse outdoor and indoor, publicly- and privately-owned public spaces (see also **?**). For instance, many MRT stations in Singapore are located within a pedestrian-oriented neighbourhood core, comprised of several mixed-use buildings. Retail activity can extend in multiple above ground as well as underground levels, while buildings are often interconnected at grade, but also with elevated walkways (bridges) and underground passages. It is unclear how well the current generation of spatial network measures, including space syntax, capture the pedestrian flows that occur in such networks. For instance, it is possible that crossing through an elevated walkway appears as having a higher-cost (in terms of metric or angular distance) although pedestrians opt for a longer route to avoid at-grade crossings; similarly for walkways through a climate-controlled (i.e. cooler) publicly accessible building. Despite these differences there has been limited research on the applicability and predictive ability of spatial network analysis for such urban environments. Moreover, if such measures are to be used by urban and transport planning authorities in these cities, it is important to know how well spatial network analysis captures pedestrian dynamics and if they can be enhanced with additional data such land-use, GFA and population. Thus, it is essential to both (a) validate and demonstrate their predictive potential in these environments and (b) calibrate model parameters to the social and spatial environment.

3 Methodology & data

To understand pedestrians' preference and predict link-based (i.e. segment analysis) pedestrian volumes, this study consisted of four steps: (i) map pedestrian network in four sites, (ii) compute hybrid network measures augmented with GFA, land-use and population data, (iii) perform pedestrian counts in all four sites, and (iv) estimate pedestrian flow regression models to assess the overall predictive ability of the network measures (variance explained) and the relative influence of different parameters. In the following section we describe the methodology followed, starting from the case-study sites.

Sites The research focuses on four areas in Singapore: Orchard Road, Raffles Place, Jurong East and Tampines. These areas have in common that they all are prominent transport hubs, but differ in surrounding land-use types and dominant network topology. Orchard Road is located in central Singapore and is flanked by shopping malls and hotels. In the immediate surrounding area a large number of condominiums are located and several private hospitals. Orchard Road itself has five lanes for vehicular traffic and can only be crossed at grade at a limited amount of junctions by pedestrians. We focus on the area surrounding Orchard MRT, which is located at the intersection of Orchard Road and Scotts Road. This intersection can only be traversed underground by pedestrians. Orchard Road MRT is located directly under the mall ION Orchard and connected to several other malls through a network of underground walkways. Raffles Place is located at the southern tip of Singapore and holds the Central Business District. The train station is located directly under the square 'Raffles Place'. Raffles Place MRT station is along Singapore's main (and oldest) train lines (Green/East-West and Red/ North-South). The MRT station is connected to all surrounding buildings and roads by a series of underground walkways. The precinct surrounding the station is characterised by retail and office land-uses. Jurong East is a suburb located in western Singapore. Although initially developed as a housing and industry estate, it has increasingly witnessed commercial developments. The area surrounding Jurong East MRT is characterized by a series of malls, high-rises, a business park and a hospital. As opposed to the two aforementioned sites, the developments are connected through a network of elevated (i.e. above grade) walkways, called the J-Walk. Tampines is a primarily residential precinct located in eastern Singapore. Tampines MRT station is surrounded by several shopping malls. In contrast to the other sites, the pedestrian network is mostly located at grade.

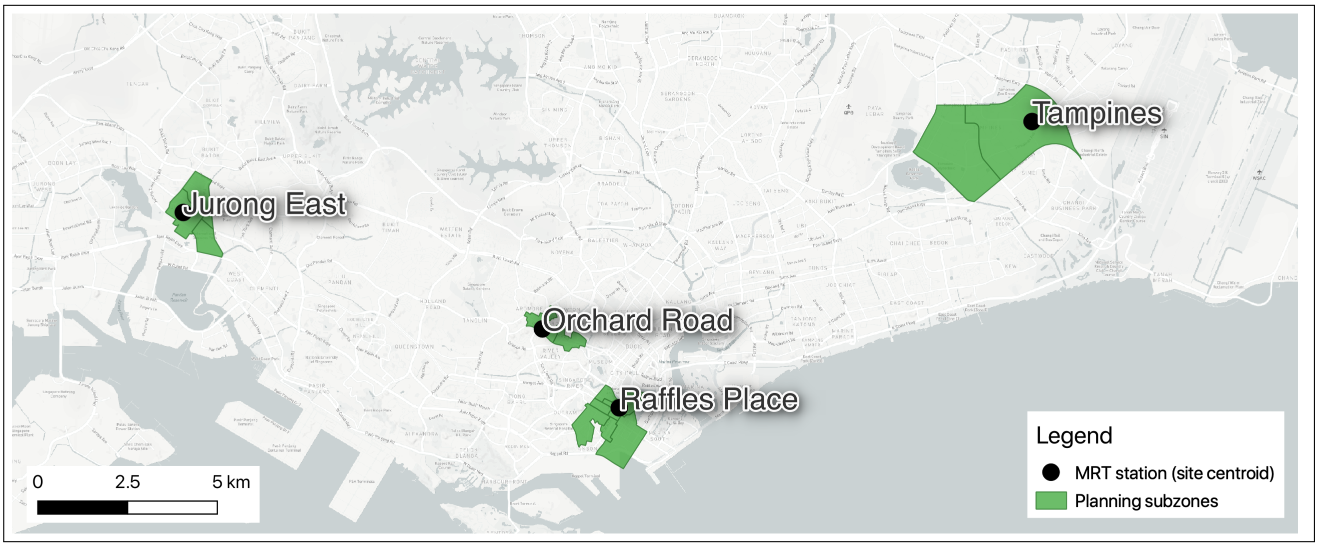


Figure 1: Location of four study areas in Singapore

Multi-level networks As a first step, for all four sites a multi-level pedestrian network was drawn. OpenStreetMap's road centrelines were used as a starting point for the network. Subsequently, a distinction was made between roads that can be crossed at any point and roads that require designated crossings. This distinction is important for eventual routing purposes. Sidewalks were drawn manually for roads that could not be crossed without a designated crossing (zebra crossing, signalized intersection) due to the number of lanes, traffic volumes or hindrances to crossings. Crossings, pedestrian underpasses and pedestrian overhead bridges were added to the network. Finally, observed informal crossings were added to the network. To account for permeable ground floors of buildings that are built on pilotis, a feature common in Singapore's public housing estates (referred as 'void-decks'), mid-building links were added for additional network connectivity.

In the next step the indoor pedestrian network was drawn. Publicly available building floor plans were obtained from Google Maps, fire evacuation plans and building's websites. These plans were digitized and every publicly accessible building level was included in the pedestrian network. Vertical connectors were added, including lifts, staircases, and escalators.

Finally, attributes were added to the network to distinguish between the different link types (e.g. indoor link, outdoor link), whether a link was a vertical connector, and the type of vertical link (stair, escalator, elevator; where applicable).

Table 1 shows the mapped network length per site. From the table it becomes apparent that the majority of publicly accessible network within close proximity to each transport hub (i.e. MRT station) is located indoors. For example, in Orchard Road, the majority of the network within 400 meter from Orchard Road MRT is located indoors (36 kilometres or 65%). Approximately of 13 km of the network in Orchard is located below grade (13 km), another 11 km grade (11 km) and 28 km is located above grade.

Land-use Detailed land-use data was received from Singapore's authorities. Per parcel a detailed breakdown on gross floor area (GFA) was given for the following land-use types: retail, office, residential, entertainment and food & beverage. For cases where GFA data was not available by the

Site Location	Jurong East			Orchard Road		
	<400 m	400 m - 800 m	800 m - 1200 m	<400 m	400 m - 800 m	800 m - 1200 m
Indoor	12	4	0	36	8	8
Outdoor	16	31	28	17	36	37
Vertical	2	0	0	4	4	1.1
Total	30	35	28	57	48	46.1
Site Location	**Raffles Places**			**Tampines**		
	<400 m	**400 m - 800 m**	**800 m - 1200 m**	**<400 m**	**400 m - 800 m**	**800 m - 1200 m**
Indoor	10	15	0	6	0	0
Outdoor	22	56	88	19.2	46	55
Vertical	0.5	0	0	0.5	0	0
Total	32.5	71	88	25.7	46	55

Table 1: Network length [km] per site and distance to MRT station

authorities, GFA was either obtained via online research (e.g. developer's yearly reports or websites) or Singapore's land-use Master Plan was used (URA, 2014). The number of residential units per private and public buildings were available as well from online sources. For public housing, the number of residential units were webscraped but are currently available as open data; the number of units in private flats are available in URA's real estate information system REALIS (URA, 2016).

Combining pedestrian networks and land-use In order to merge the land-use data to the pedestrian network, we used the approach of computing the centroid of each parcel. For parcels with a single building the building was used. For parcels with multiple buildings multiple centroids were calculated. In both cases, building centroids were move adjacent to the link were the entrance of a building was located. In specific cases, the GFA figures, which are available on a parcel level, were divided in multiple, 'virtual' buildings, to better reflect actual land-usage. For instance, in Raffles Place, where some buildings only have retail activities (food & beverage) in the below grade network, while all the above-ground network is allocated to office space. Finally, GFA data per building were allocated to each walkway segment (link) proportionally to its length.

Pedestrian counts Pedestrian counts were conducted at each of the four site in 2018 and 2019. At every site, approximately 80 locations were selected after multiple site visits. These locations were selected based on several criteria. These criteria included, but were not limited to the distance to the site centroid, possible coverage of parallel routes between important origin-destination pairs, variation in land-use, variation in walkway type (outdoor, indoor), variation in pedestrian flow and variation in floor levels. Table 2 lists the total number of count locations per site, and the dates and times counted.

Network analysis Several groups of network measures have been calculated. The first set of measures includes accessibility measures. Accessibility is defined as the number of opportunities

Site Device	Jurong East	Orchard Road	Raffles Place	Tampines
Counting mat	2	0	2	0
Sensmax (Infrared)	4	5	2	0
Pyrobox (Infrared)	3	3	3	3
Video cameras	17 locations (10 cameras)	13 locations (10 cameras)	24 locations (12 cameras)	16 locations (11 cameras)
Manual	59 locations (12 assistants, covering 4 to 5 locations per hourly round)	64 locations (12 assistants, covering 4 to 6 locations per hourly round)	67 locations (12 assistants, covering 5 to 6 locations per hourly round)	63 locations (12 assistants, covering 4 to 6 locations per hourly round)
Dates and times	Thursday, May 17th 2018 (Manual counting 7:00 - 19:30)	Thursday, May 10th 2018 (Manual counting 7:30 - 19:30)	Tuesday, October 30th 2018 (Manual counting 7:30 - 20:00)	Thursday, 25th of April 2019 (Manual counting 7:00 - 19:30)
	Friday, May 18th 2018	Friday, May 11th 2018	Wednesday, October 31st 2018	Friday, 26th April 2019
	Saturday, May 19th 2018 (Manual counting 9:30 - 20:30)	Saturday, May 12th 2018 (Manual counting 9:30 - 20:30)	Thursday, November 1st 2018 (Manual counting 8:30 - 21:00)	Saturday, 27th April 2019 (Manual counting 9:30 - 20:30)

Table 2: Count locations per site number of days counted and dates

that can be reached within a pre-defined distance. In this case specifically, land-use is attached as link weights and summed up.

The second set of measures concerns space syntax metrics. Two types of measures have been calculated: integration and different variants of choice (with and without length as weight).

The third and final set of measures include land-use and network topology. Following Cooper *et al.* (2019) two-phase betweenness metrics are calculated. Whereas betweenness metrics counts the number of paths that pass through a link, the origin weight is distributed over destination weights.

Accessibility measures and two-phase betweenness measures have been calculated using several types of distance metrics. The first distance metric uses link length to determine routes between origin-destination pairs. The second distance metric uses angular distance to determine routes between origin-destination pairs: a route with the least amount of angular change is considered the shortest route. The third distance metric uses perceived link distance to determine shortest routes. Based on the results of an earlier study on pedestrian route choice conducted in Singapore (Erath *et al.*, 2015), crossings, signalized intersections and vertical changes were considered as 20% longer, to account for waiting time and physical effort respectively. Based on the same study, indoor links and links with commercial land-usage were considered as 20% shorter. Finally, shortest routes were determined by using a combination of metric distance and angular distance. Each component was weighted as 50%: i.e. link length accounts for 50% of route costs while angular changes, expressed in degrees, accounts for 50% of the route costs.

To limit correlation between network measures, *banded*, or 'donut', measures were calculated. Banded measures only consider links that are within a certain distance from each other. We used distance-bins of 200 meters for a distance of up to 1200 meters from each link (e.g. 0 - 200 meters, 200-400 mmeters, 1000 - 1200 meters, etc). In addition to limiting correlations, banded measures allow for the estimation of a piece-wise linear distance decay function.

Model estimation and prediction Multivariate regression models with pedestrian counts as a dependent variable and betweenness calculations as independent variables have been estimated using lasso and ridge regression. Unlike ordinary least squares regression, lasso and ridge regression are able to cope with correlation between the independent variables. This correlation is likely to exist, as betweenness variables, calculated for different radii and land-uses, are likely to be correlated. Lasso and ridge regression make use of a penalty term λ that prevents overfitting of the regression coefficients. The parameter λ is determined by cross-validation and repetition. The model is fitted against on a random subset of the data and tested on the remainder of the data.

Software The pedestrian network was initially drawn in Autodesk AutoCAD using 3D geometries, and was stored in a spatially enabled database (PostgreSQL / Postgis). This database could be accessed through ESRI's ArcGIS, as well an external spatial database through ArcSDE and in AutoCAD through ESRI's ArcGIS for AutoCAD extension. Using AutoCAD was a necessity to be able to draw and edit 3D networks, a capability that was found to be lacking in GIS software. However, GIS software was necessary to keep track of attributes and pictures of links, and perform spatial analysis. Network analysis measures were calculated using the publicly available sDNA software (Cooper & Chiaradia, 2020). Finally, pedestrian demand models were estimated with R package *glmnet* (Friedman *et al.*, 2021).

Overall, this workflow enabled us to integrate ESRI's tools for visualization, network editing and in-field network auditing, with the power of modern databases like PostgreSQL, the visualization with QGIS and statistical analysis with the open source statistical software R.

4 Results

Multivariate ridge regression models were initially estimated for different aggregates of the pedestrian counts: weekday evening peak, weekday lunch peak and weekday peak. Finally, it was decided to use weekday average peak as a dependent variable in the final model estimations, as the highest counts for most sites were observed for this time period (with the exception of Raffles Place site, where lunchtime flows were highest). These flows are more relevant from a planning perspective as well, as they include regular commuter trips and this are thought to be consistently and constant high across multiple days.

In all tables with model estimates, standardized and unstandardized model estimates are presented. Standardized model estimates are presented to be able to compare model estimates from the same model with each other; parameters are standardized by using the standard deviation.

The goodness-of-fit for ridge regression model is calculated in a different fashion than for ordinary least squares. In this research, different goodness-of-fit indicators will be presented. While the commonly known rho-square will be presented, other indicators will also be mentioned. These include mean squared error, the cross-validated rho-square and the GEH statistic. The goodness-of-fit of the final model are calculated in a similar fashion as in the estimation: by using the estimated parameters the goodness-of-fit can be evaluated by using simulation: calculating the goodness of fit on a subsample rather than on the entire sample multiple times.

The values should be interpreted as following: the RMSE or root mean squared error provide an indication of the standard deviation. A lower value indicates a better model fit. The presented rho square is the commonly used rho-square: a higher value indicates a better model fit. The 85% GEH statistic indicates how many count locations are within 85% of the counted values, using the GEH statistic, which combines the absolute and the relative error in a single measure. A higher value is considered better.

Models incorporating accessibility The first set of models, presented here, only include accessibility metrics, i.e. how much land-use is accessible within a certain radius / band (nb. these models do not include space syntax measures). Parameter estimates per site are shown in Table 3. The results show the estimated coefficients ("Est. Coeff") and scaled coefficients ("Z-Coeff"). To compare parameter estimates between, sites, distance bins and GFA types the scaled coefficients should be used. Estimated coefficients cannot be compared directly as the amount of GFA per site varies between site and land-use type; only the estimated coefficients for the access to MRT stations can be interpreted easily.

Starting with access to MRT, it can observed that walkways in close proximity to MRT ($<$200 meters) receive the highest footfall. In the residential site of Tampines, the effect of the MRT extends up to 800 meters from the train station. Proximity to retail GFA is the second highest contributor to footfall in all sites except in Raffles Places, where proximity to office GFA is the second highest contributor.

Model goodness-of-fit varies considerably between sites and is highest for Tampines (Rho-square 0.63), followed by Jurong East (Rho-square 0.37) and Orchard (Rho-square 0.23).

	Jurong East		Orchard		Raffles		Tampines	
	Est. Coeff	Z-Coef	Est. Coeff	Z-Coeff	Est. Coeff	Z-Coeff	Est. Coeff	Z-Coeff
Intercept	-74.58	157.56	-28.31	172.56	68.41	209.01	-656.96	218.29
Access to MRT station								
0 - 200m	202.13	70.42	34.52	14.53	50.41	23.76	273.69	216.11
200 - 400m	25.19	12.38	0	0	0	0	141.93	146.2
400 - 600m	0	0	11.71	6.38	0	0	125.2	104.29
600 - 800m	0	0	0	0	0	0	102.54	81.21
800 - 1000m	0	0	0	0	4.24	4.59	63.13	25.36
Access to retail GFA								
0 - 200m	0	88.67	0	30.83	0	0	0.01	125.71
200 - 400m	0	9.18	0	11.66	0	6.68	0	3.37
400 - 600m	0	0	0	7.32	0	0.09	0	0
600 - 800m	0	13.54	0	0	0	10.4	0	16.45
800 - 1000m	0	10.03	0	0	0	0	0	28.94
1000 - 1200m	0	9.17	0	0	0	0	0	8.35
Access to office GFA								
0 - 200m	0	0	0	13.59	0	21.23	0	11.11
200 - 400m	0	17.85	0	20.59	0	16.68	0	16.85
400 - 600m	0	11.95	0	0	0	0	0	28.65
600 - 800m	0	0	0	0	0	0	0	20
800 - 1000m	0	0	0	0	0	2.31	0	0
1000 - 1200m	0	25.3	0	0	0	0	0	32.69
Access to residential units								
0 - 200m	0	0	0	0	0	0	0	0
200 - 400m	0	0	0	0	0	0	0	0
400 - 600m	0	0	0.18	24.12	0	0	0	0.26
600 - 800m	0	0	0	0	0	0	0.04	23.09
800 - 1000m	0	0	0	0	0	0	0	2.29
1000 - 1200m	0	0	0	0	0	0	0	0
Estimation								
Lambda	40.37		148.5		84.97		5.214	
Weight	0.7		0.7		0.7		0.7	
Goodness-of-fit								
Rho-square	0.37		0.24		0.16		0.63	
Mean RMSE	179.15		172.74		141.78		181.51	
85% GEH	31		23		42		34	
Observations	79		71		89		85	

Table 3: Parameter estimates using accessibility and metric distance

Model predictions using the estimated parameter estimates for Raffles Place and Tampines can be in seen in Figure 2. The highest predicted link volumes can be seen in close proximity to the MRT station in both sites. At the same time, it can be seen in both sites that the predicted link flows are similar for adjacent links in all directions. This shows that the calculated accessibility values extend homogeneously in all directions and highlights the aforementioned shortcoming

Models incorporating network topology The second set of models uses betweenness metrics as independent variables to describe pedestrian link flows (i.e. without land-use information). Parameter estimates are presented in Table 4. In three sites (Jurong East, Orchard Road and Tampines) angular betweenness between 0 and 200 meters and angular betweenness between 200 and 400 meters contribute most to pedestrian link flows, indicating that most links serve short trips. This is inline with other studies showing that the majority of pedestrian trips are under 400 meters long Erath *et al.* (2015). Only in Orchard Road angular betweenness for links located between 400 and 600 meters is also relevant to predict observed link flows.

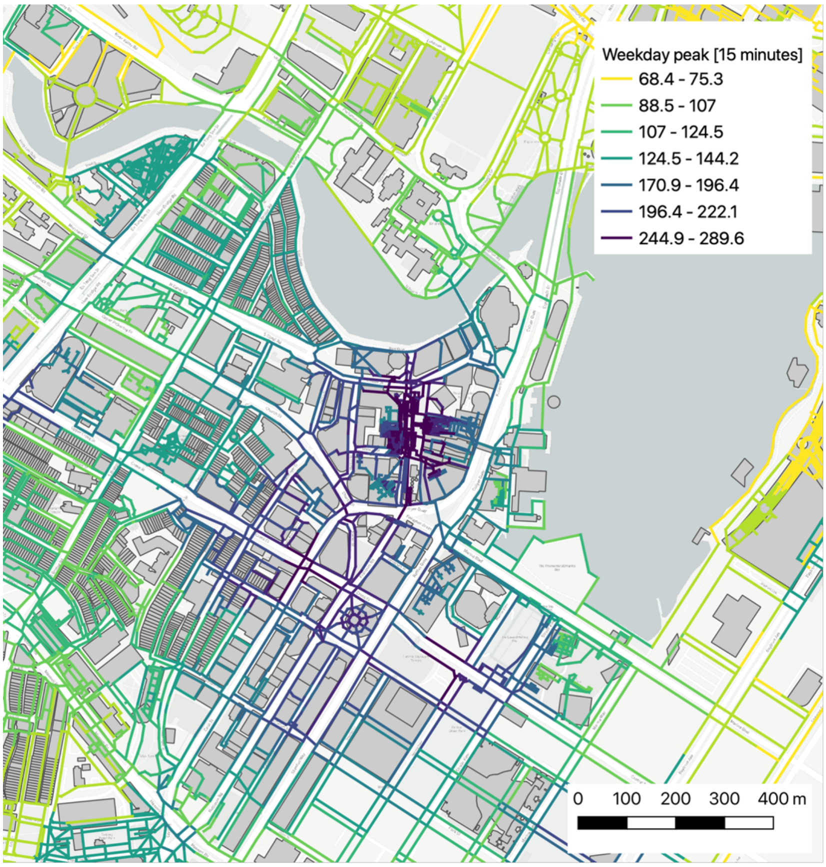
	Jurong East		**Orchard**		**Raffles Place**		**Tampines**	
	Est. Coeff	**Z-Coeff**	**Est. Coeff**	**Z-Coeff**	**Est. Coeff**	**Z-Coeff**	**Est. Coeff**	**Z-Coeff**
Intercept	75.69	157.46	31.66	194.73	210.17	214.74	102.37	202.71
Angular betweenness								
0 - 200m	0	88.93	0	48.19	0	1.79	0	83.05
200 - 400m	0	11.41	0	62.74	0	0.8	0	48.82
400 - 600m	0	0	0	35.94	0	0.74	0	0
600 - 800m	0	0	0	0	0	0.51	0	0
800 - 1000m	0	0	0	0	0	0.54	0	0
1000 - 1200m	0		0	0	0	0.64	0	0
Estimation								
Lambda	48.62		58.57		2420.12		70.54	
Weight	0.8		0.8		0.8		0.8	
Rho-square	0.17		0.46		0.01		0.24	
Mean RMSE	211.1		150		150.68		225.85	
85% GEH	18		32		36		23	
Observations	79		71		89		85	

Table 4: Parameter estimates with angular distance and link weights

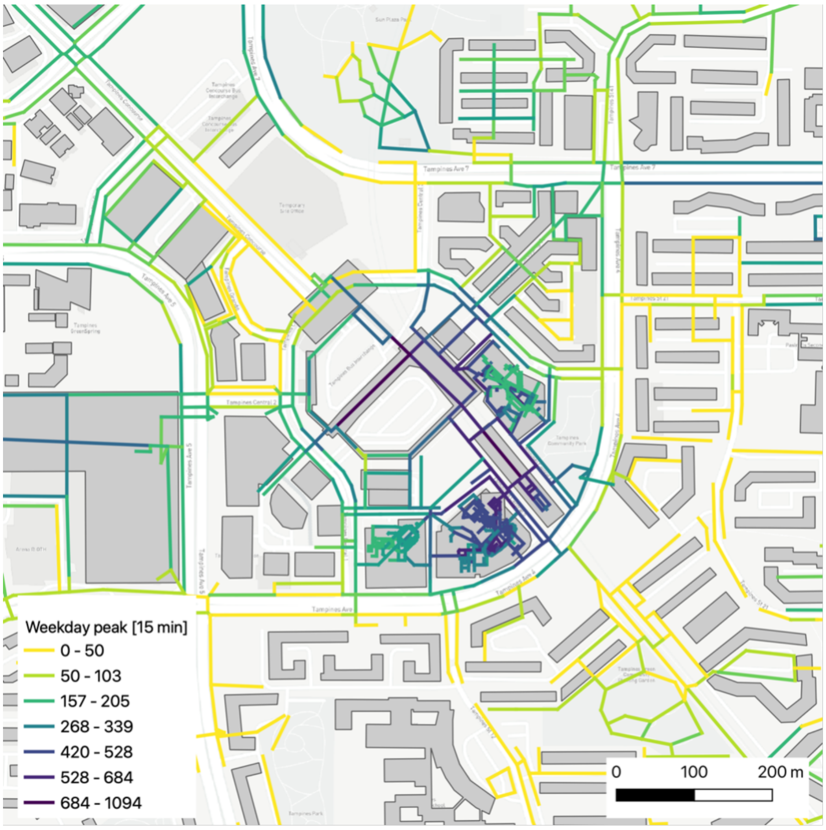
Model fit varies considerably between sites and is highest for Orchard (Rho-square 0.46), followed by Tampines (Rho-square 0.24) and Jurong East (Rho-square 0.17). For Raffles Place, model fit is extremely low (Rho-square 0.01).

Models combining land-use with network topology The final set of models includes betweenness metrics that include land-use. Two sets of models have been defined. First, models were estimated that included commercial, MRT and residential land-use. Second, models were estimated that included that contained a further break down of commercial GFA into retail and office GFA.

The presented models in Figure 3 and Figure 4 present results including two-phase betweenness



(a) Raffles Place



(b) Tampines

Figure 2: Model predictions with angular change as distance metric and link length as weights

metrics. The variable names indicate the origin and destination used in the betweenness computation. For instance, bus_com_tbt200 indicates 'bus to commercial 0 – 200 meters', 'mrt_resid_tbt1200' indicates betweenness between MRT and residences 1000 – 1200 meters. The colour indicates the relative importance of each variable; a darker colour (red) indicates a higher importance. The figures only present standardized coefficients for ease of comparison.

Starting with the model estimates including commercial land-usage, it can be seen that the links located between the MRT and commercial GFA receive the highest footfall. This effect is most pronounced in Jurong East, Orchard Road and Raffles Place. This effect extends up to 1000 meters in Jurong East and Tampines; in Orchard Road and Raffles Place up to 400 meters. This can be, as there is more overlap in catchment areas of MRT stations serving commercial developments in Orchard Road and Raffles Place. The second most important driver are links leading to commercial developments; this effect extends up to 600 meters in Orchard Road. In Raffles Place and Tampines, links between commercial GFA attract a high amount of pedestrians. In both Jurong East and Tampines, links between MRT and residential units attract pedestrians. Where in Jurong East this effect is limited to 600 meters from the MRT station, in Tampines this effect extends up to 1200 meters.

When a distinction is made between retail and office GFA (Figure 4), instead of solely entering commercial GFA in the model specification, it becomes apparent that solely relying on the aggregate commercial GFA may not be sufficient. In Jurong East, links between the MRT station and retail GFA attract the most footfall. This effect extends up to 800 meters from the MRT station. In Orchard Road, this effect is present as well, but more constant up to 800 meters. In Tampines the effect of retail is less strong; whereas in Raffles Place is almost not present for weekday peak hour flow. In Raffles Place, weekday peak hour pedestrian demand is highest on links between the MRT station and office GFA up to 400 meters.

Goodness-of-fit metrics are presented separately in Figure 5. A distinction is made between models including solely commercial GFA (commercial), or models that make a distinction between retail and office GFA (retail_office). Also, a distinction is made between different the routing metrics considered: (a) Ldp: Perceived Link Distance, (b) Ang: Angular and (c) Euc: Euclidean (metric distance). Judging by model fit, all models outperform the earlier presented models (solely network topology and accessibility): the root mean squared error drops and rho-square increases. Including more variables does not necessarily result in an improved model: in Tampines, model fit is better when solely including commercial GFA. In other sites, model fit is improved when a distinction is made between retail and office GFA. In different sites it can be seen that the usage of different routing metrics results in differences in goodness-of-fit. In models including commercial GFA as dependent variable, the usage of perceived link distance yields slightly better model estimations, expect for the case of Orchard Road. When a breakdown is made between retail and office, models using angular distance perform slightly better.

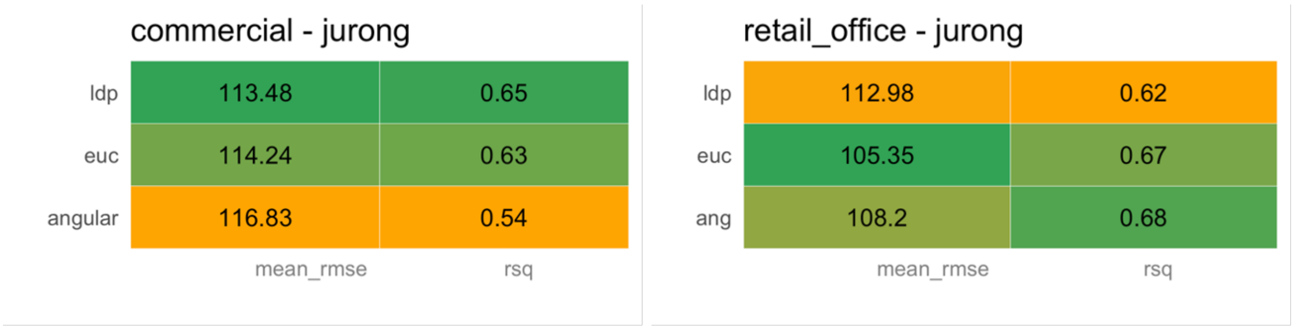
A prediction of pedestrian link-based flows to the entire network is visualized for Jurong East and Orchard Road is visualized in Figure 6. For Raffles Place and Tampines the prediction is visualized in Figure 7. In the visualizations, it becomes apparent, that certain links (most notably links between MRT and commercial land-use) receive most footfall.

	jurong	orchard	raffles	tampines
bus_com_tpbt200	4.1	0	7.1	4.8
bus_com_tpbt400	0	0	0	0
bus_none_tpbt200	2.1	0	0	0
bus_none_tpbt400	0	0	0	0
bus_resid_tpbt200	0.1	0	0	0
bus_resid_tpbt400	0	0	0	0
com_com_tpbt200	0	0	0	0
com_com_tpbt400	1.4	0	0	4.6
com_com_tpbt600	8.9	0	0	0.6
com_com_tpbt800	0	0	0	0
com_com_tpbt1000	0	0	19.2	16.2
com_com_tpbt1200	0	0	6.2	0
mrt_bus_tpbt200	0	0	0	14.7
mrt_bus_tpbt400	15	0	10.5	17.9
mrt_com_tpbt200	61	10	24.4	38.1
mrt_com_tpbt400	38.1	47.5	52.4	35.5
mrt_com_tpbt600	16.3	15	0	26.4
mrt_com_tpbt800	24.1	0	0	13.1
mrt_com_tpbt1000	13.8	0.4	0	0
mrt_com_tpbt1200	3.8	1.8	0	0
mrt_resid_tpbt200	0	0		0
mrt_resid_tpbt400	2.2	0		0.2
mrt_resid_tpbt600	2.7	0		13
mrt_resid_tpbt800	0	6.6		5.4
mrt_resid_tpbt1000	0	0.5		10.1
mrt_resid_tpbt1200	0	0		12.1
none_com_tpbt200	11.9	2	0	10.7
none_com_tpbt400	1.7	28.2	7.1	13.1
none_com_tpbt600	16.4	13.9	0	3.6
none_com_tpbt800	2.5	0	0	0
none_com_tpbt1000	0	11	0	0
none_com_tpbt1200	0	20.4	0	0
none_resid_tpbt200	6.2	3.1		0
none_resid_tpbt400	5.3	0		0
none_resid_tpbt600	0	0		0
none_resid_tpbt800	0	10.8		0
none_resid_tpbt1000	0	0		0
none_resid_tpbt1200	0	0		0
resid_com_tpbt200	0	0		0
resid_com_tpbt400	0	0		0
resid_com_tpbt600	0	0		0
resid_com_tpbt800	0	0		0
resid_com_tpbt1000	0	0		0
resid_com_tpbt1200	0	0		0

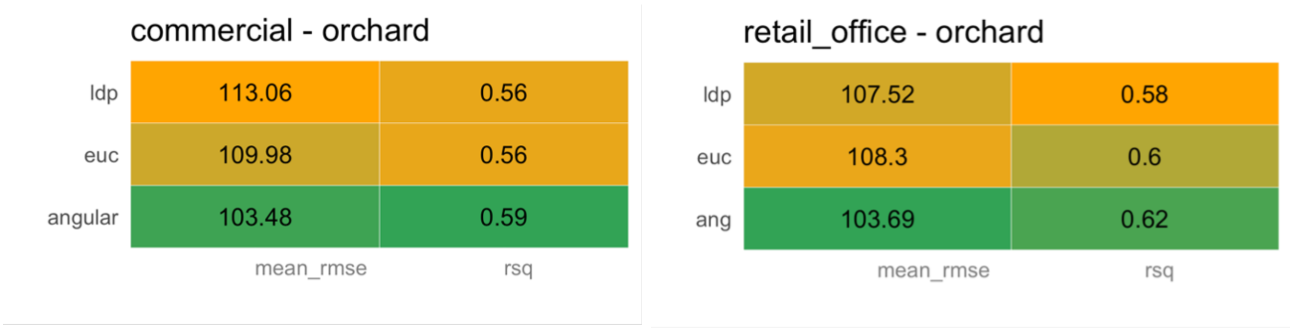
Figure 3: Two-phase betweenness / Commercial / Euclidean distance / Weekday peak

	jurong	orchard	raffles	tampines
bus_none_tpbt200	1.9	0	0	2.2
bus_none_tpbt400	0	0	1.6	3.7
bus_office_tpbt200	2.9	0	0	0
bus_office_tpbt400	0	0	0	0
bus_resid_tpbt200	0.8	0	0	0
bus_resid_tpbt400	0	0	0	0
mrt_bus_tpbt200	0	0	0	7.7
mrt_bus_tpbt400	4.2	0	4.3	8.2
mrt_office_tpbt200	0	0	30.9	2.3
mrt_office_tpbt400	6.5	23.1	25.5	8.6
mrt_office_tpbt600	11.9	2.3	3.7	9.2
mrt_office_tpbt800	12.9	0	0	12.2
mrt_office_tpbt1000	14.4	0	10.5	0
mrt_office_tpbt1200	5.6	0	2.1	0
mrt_resid_tpbt200	0	0	0	0
mrt_resid_tpbt400	5.1	0	9.1	2.2
mrt_resid_tpbt600	5.4	0	0	5.2
mrt_resid_tpbt800	1.5	2.2	0	4.2
mrt_resid_tpbt1000	2.9	0.1	0	6
mrt_resid_tpbt1200	0.9	0	0	5.6
mrt_retail_tpbt200	39.1	5.7	5.1	12.7
mrt_retail_tpbt400	38	16.1	6.7	12.9
mrt_retail_tpbt600	24.9	19.9	20.1	8.4
mrt_retail_tpbt800	8.4	12.8	0	3.9
mrt_retail_tpbt1000	2.9	0.5	0	0
mrt_retail_tpbt1200	0	0	1.4	0
none_office_tpbt200	0	0	0	0
none_office_tpbt400	0	0	0	2.3
none_office_tpbt600	3.5	3.4	10.6	3.5
none_office_tpbt800	0	0	0	1.2
none_office_tpbt1000	0	0.2	0.6	5.2
none_office_tpbt1200	0	0	0	0
none_resid_tpbt200	5.2	0.5	23.5	0
none_resid_tpbt400	1.3	0.8	0	1.6
none_resid_tpbt600	0	0	0	0.3
none_resid_tpbt800	0	5.9	0	0.9
none_resid_tpbt1000	0	0	0	0
none_resid_tpbt1200	0	0	0	0.4
none_retail_tpbt200	14.1	0	13.3	8.1
none_retail_tpbt400	0	13	0	7.2
none_retail_tpbt600	6.1	10.3	0	5.5
none_retail_tpbt800	0	8.2	0	1.1
none_retail_tpbt1000	0	7.5	0	0.7
none_retail_tpbt1200	0	5.8	0	0.7
resid_office_tpbt200	0	0	0	0
resid_office_tpbt400	0	0	0	0
resid_office_tpbt600	0	0	0	0
resid_office_tpbt800	0	0	0	0
resid_office_tpbt1000	0	9.6	0	0
resid_office_tpbt1200	0	0.7	0	0
resid_retail_tpbt200	0	0	0.2	0
resid_retail_tpbt400	0	0	0	0
resid_retail_tpbt600	0.4	0	0	0.4
resid_retail_tpbt800	0	0	0	0.4
resid_retail_tpbt1000	0	0	0	0
resid_retail_tpbt1200	0	1.9	0	0.2
retail_office_tpbt200	0	0	0	0
retail_office_tpbt400	0	0	0	1.9
retail_office_tpbt600	10.2	4.7	0	2.8
retail_office_tpbt800	0	0	0	2.4
retail_office_tpbt1000	0	0	0	12.8
retail_office_tpbt1200	1.7	0	0	0
retail_retail_tpbt200	9.2	0	0	3.4
retail_retail_tpbt400	0	2.8	0	6.6
retail_retail_tpbt600	12.7	3.7	0	5.1
retail_retail_tpbt800	0	4.2	0	0
retail_retail_tpbt1000	1.1	5.9	0	0
retail_retail_tpbt1200	3	2	0	0

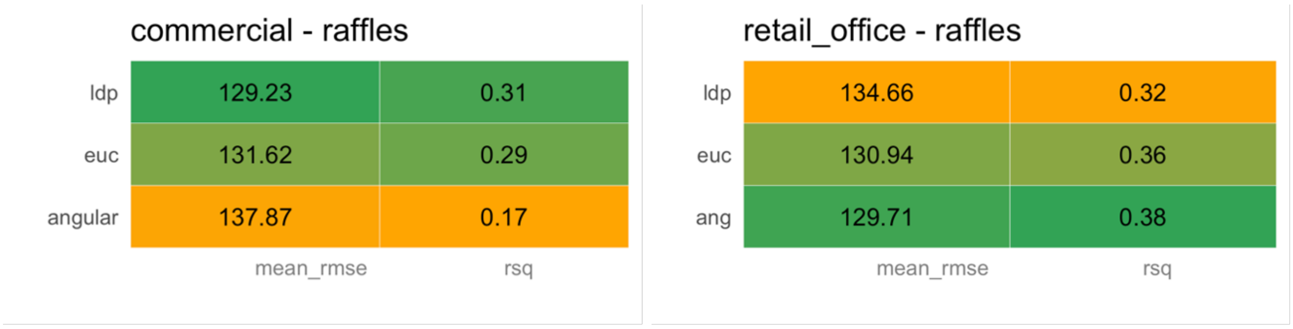
Figure 4: Two-phase betweenness / Retail Office / Euclidean distance / Weekday peak



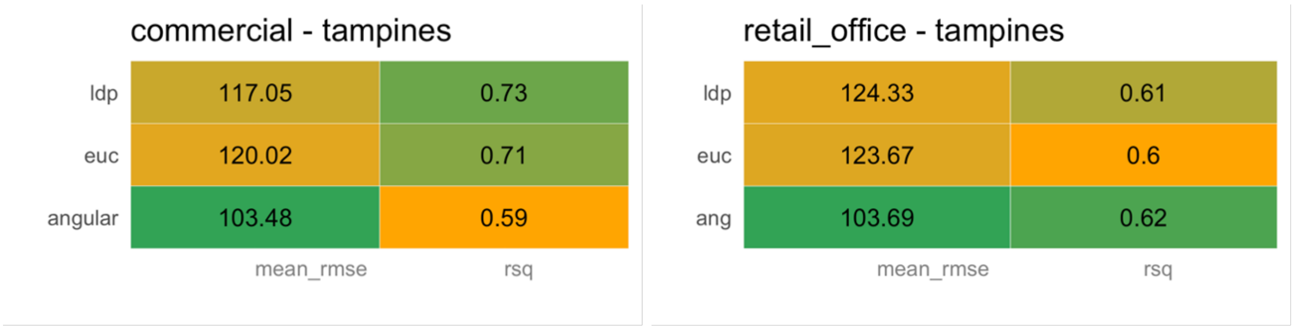
(a) Goodness-of-fit: Jurong East



(b) Goodness-of-fit: Orchard Road

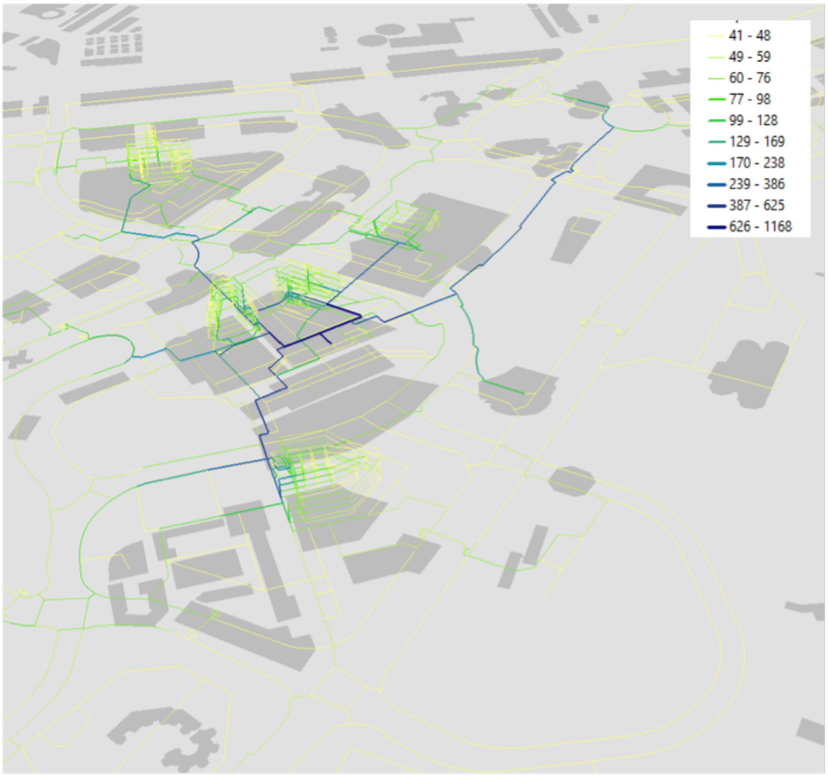


(c) Goodness-of-fit: Raffles Place

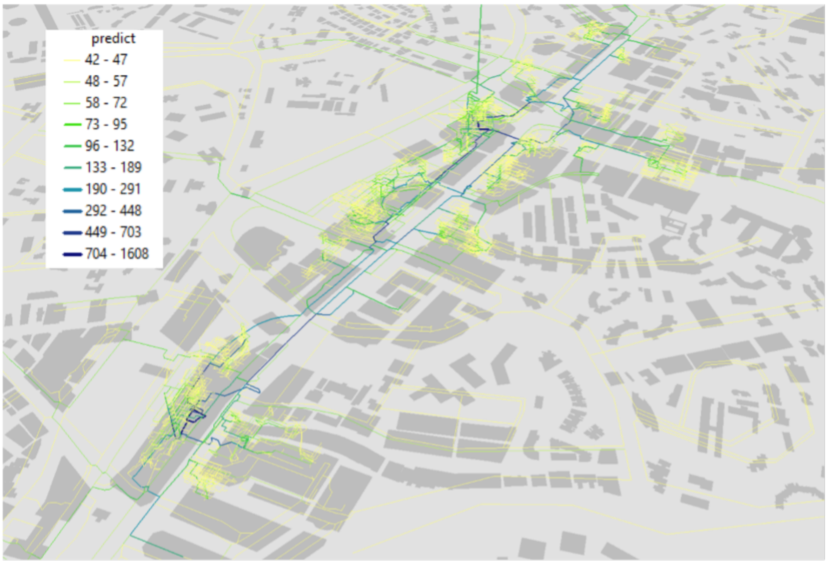


(d) Goodness-of-fit: Tampines

Figure 5: Goodness-of-fit using two-phase betweenness metrics for all 4 sites



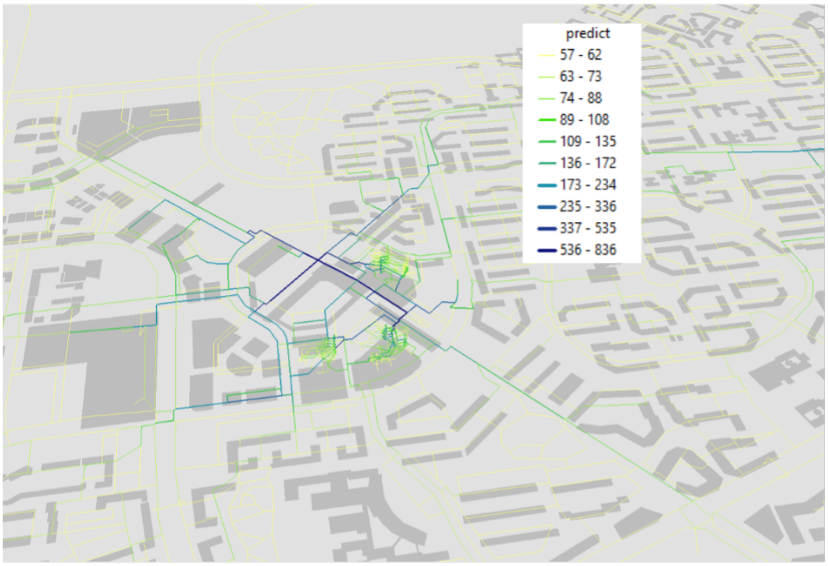
(a) Jurong East



(b) Orchard Road

Figure 6: Model prediction using parameter estimates for metric distance and office / retail covariates

(a) Jurong East



(b) Orchard Road

Figure 7: Model prediction using parameter estimates for metric distance and office / retail covariates

5 Conclusion & outlook

5.1 Conclusion

To the authors knowledge, this is the first time that models have been estimated for pedestrian demand in combined indoor / outdoor and public / private environments using highly detailed pedestrian networks combined with GFA and land-use data. Model results allow for the prediction of pedestrian link flows based on data that describes the GFA for different land-use types.

In all four sites, links leading from MRT stations to commercial developments attract most footfall. When commercial land-use is disaggregated, links leading from MRT stations to retail developments attract most footfall.

Network analysis Judging by model fit, models that include betweenness metrics that account for land-use outperform models that solely include network topology. Models that include betweenness measures outperform accessibility in Orchard, but not in the other sites. It is thought that this is the case as the pedestrian network in Orchard consists of many links in shopping malls connected to pedestrian through fares, whereas in the other sites the number of links is less related to pedestrian flows. It is concluded that solely accounting for network typology is not suited for the estimation of district-wide pedestrian demand models.

Route choice preferences The presented model results do not reveal a clear preference for routing metric. Dependent on the goodness-of-fit indicator chosen either angular, metric or perceived link distance perform similarly well. From pedestrian counts, it is clear that individuals prefer direct connections that provide same-level connections.

Multi-level pedestrian networks and land-use When using network measures that include land-use, choosing the correct link to attach weights is important for model fit and prediction. In complex, multi-level, networks, not only the link but the floor level is of importance as well. For instance, it is important to attach the MRT station to the links connecting the MRT to other developments. Initially building GFA was distributed proportionally across along links in a building. After inspecting residuals it become apparent that several buildings should be split into multiple, 'virtual' buildings, to account for differences in the amount of footfall that certain levels in buildings generate and attract. For instance, in a certain development along Orchard Road is characterized by shops, restaurants and an event-space with a low to medium price level below grade and high-end shops at grade and above grade. In Raffles Place, retail GFA has been attached to the basement links of buildings, at locations where shops are located; office GFA has been attached to links above grade, where companies are situated.

Pedestrian counts Ideally, pedestrian counts should be available for every distance bin for every land-use combination to estimate the effect of distance on pedestrian flow. This was not the case for the conducted pedestrian counts, including, but not limited to, permissions to place pedestrian counters or conduct counts and technical issues. Permanent pedestrian counting devices at key locations can allow for the monitoring of pedestrian flow over longer periods of time. These pedestrian counts should be available for indoor and outdoor environments.

5.2 Outlook

For further research, we recommend the evaluation of a distance metric that incorporates both metric and angular distance. Recent approaches include a computation of hybrid distance Cooper *et al.* (2019) and have shown promising results, but we don't know how well these correspond to individual navigation and spatial decision making. Also, we recommend in looking into usage of 'randomized' link costs. Instead of calculating betweenness metrics on a single, deterministic, distance, link costs will be drawn from a distribution, in other words varied slightly and averaged over a large number of calculations. This can capture the uncertainty around micro-level route choice at crossings. Finally, at the research-to-application level, we recommend increasing the interaction with planners and policy makers, to increased the adoption of these new tools and methods. As Yamu and van Nes discuss this also involves educating planners the underlying rationale, and theory behind these measures (**?**).

6 Acknowledgments

This research is supported by the Singapore Ministry of National Development and the National Research Foundation, Prime Minister's Office under the Land and Liveability National Innovation Challenge (L2 NIC) Research Programme (L2 NIC Award No. L2NICTDF1-2017-1). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not reflect the views of the Singapore Ministry of National Development and National Research Foundation, Prime Minister's Office, Singapore. All authors were affiliated with the Future Cities Laboratory, Singapore ETH Centre at the time of the research.

References

Cambra, Paulo, Moura, Filipe, & Gonçavales, Alexandre. 2017. On the Correlation of Pedestrian Flows to Urban Environment Measures: A Space Syntax and Walkability Analysis Comparison Case. *In: Proceedings of the 11 Th Space Syntax Symposium*.

Cooper, Crispin H. V., Harvey, Ian, Orford, Scott, & Chiaradia, Alain J. F. 2019. Using Multiple Hybrid Spatial Design Network Analysis to Predict Longitudinal Effect of a Major City Centre Redevelopment on Pedestrian Flows. *Transportation*, Dec.

Cooper, Crispin H.V. 2017. Using Spatial Network Analysis to Model Pedal Cycle Flows, Risk and Mode Choice. *Journal of Transport Geography*, **58**(Jan.), 157–165.

Cooper, Crispin H.V., & Chiaradia, Alain J.F. 2020. sDNA: 3-d Spatial Network Analysis for GIS, CAD, Command Line & Python. *SoftwareX*, **12**(July), 100525.

Desyllas, Jake, Duxbury, Elspeth, Ward, John, & Smith, Andrew. 2003. Pedestrian Demand Modelling of Large Cities: An Applied Example from London.

Erath, Alexander, Van Eggermond, Michael A.B, Ordóñez Medina, Sergio, & Axhausen, Kay W. 2015. Modelling for Walkability: Understanding Pedestrians' Preferences in Singapore. *In: 14th International Conference on Travel Behavior Research (IATBR 2015)*.

Ewing, Reid, & Cervero, Robert. 2010. Travel and the Built Environment: A Meta-Analysis. *Journal of the American Planning Association*, **76**(3), 265–294.

Freeman, Linton C. 1977. A Set of Measures of Centrality Based on Betweenness. *Sociometry*, **40**(1), 35.

Friedman, Jerome, Hastie, Trevor, Tibshirani, Rob, & Narasimhan, Balasubramanian. 2021. *Package 'Glmnet'*.

Griswold, Julia B., Medury, Aditya, Schneider, Robert J., Amos, Dave, Li, Ang, & Grembek, Offer. 2019. A Pedestrian Exposure Model for the California State Highway System. *Transportation Research Record: Journal of the Transportation Research Board*, **2673**(4), 941–950.

Guo, Zhan, & Loo, Becky P.Y. 2013. Pedestrian Environment and Route Choice: Evidence from New York City and Hong Kong. *Journal of Transport Geography*, **28**(Apr.), 124–136.

Hillier, B., Penn, A., Grajewski, T., & Xu, J. 1993. Natural Movement: Or, Configuration and Attraction in Urban Pedestrian Movement. *Environment and Planning B: Planning and Design*, **20**, 29–66.

Huat, Chua Beng. 2011. Singapore as Model: Planning Innovations, Knowledge Experts. *Worlding cities: Asian experiments and the art of being global*, **31**, 29–54.

Kuzmyak, J. Richard, Walters, Jerry, Bradley, Mark, Kockelman, Kara M., & Transportation Research Board. 2014. *Estimating Bicycling and Walking for Planning and Project Development: A Guidebook*. Washington, D.C.: Transportation Research Board.

Lerman, Yoav, Rofè, Yodan, & Omer, Itzhak. 2014. Using Space Syntax to Model Pedestrian Movement in Urban Transportation Planning: Using Space Syntax in Transportation Planning. *Geographical Analysis*, **46**(4), 392–410.

Lue, Gregory, & Miller, Eric J. 2019. Estimating a Toronto Pedestrian Route Choice Model Using Smartphone GPS Data. *Travel Behaviour and Society*, **14**(Jan.), 34–42.

Munira, Sirajum, & Sener, Ipek N. 2017. *Use of Direct-Demand Modeling in Estimating Nonmotorized Activity: A Meta-analysis*. Tech. rept. UTC Safe-D 01-003. Texas A&M Transportation Institute.

Muraleetharan, Thambiah, & Hagiwara, Toru. 2007. Overall Level of Service of Urban Walking Environment and Its Influence on Pedestrian Route Choice Behavior: Analysis of Pedestrian Travel in Sapporo, Japan. *Transportation Research Record: Journal of the Transportation Research Board*, **2002**(1), 7–17.

Porta, Sergio, Strano, Emanuele, Iacoviello, Valentino, Messora, Roberto, Latora, Vito, Cardillo, Alessio, Wang, Fahui, & Scellato, Salvatore. 2009. Street Centrality and Densities of Retail and Services in Bologna, Italy. *Environment and Planning B: Planning and Design*, **36**(3), 450–465.

Ratti, Carlo. 2004. Space Syntax: Some Inconsistencies. *Environment and Planning B: Planning and Design*, **31**(4), 487–499.

Richmond, Jonathan E. D. 2008. Transporting Singapore—The Air-Conditioned Nation. *Transport Reviews*, **28**(3), 357–390.

Rodriguez, Daniel A., Merlin, Louis, Prato, Carlo G., Conway, Terry L., Cohen, Deborah, Elder, John P., Evenson, Kelly R., McKenzie, Thomas L., Pickrel, Julie L., & Veblen-Mortenson, Sara. 2014. Influence of the Built Environment on Pedestrian Route Choices of Adolescent Girls. *Environment and Behavior*, Jan., 0013916513520004.

Turner, Alasdair. 2007. From Axial to Road-Centre Lines: A New Representation for Space Syntax and a New Model of Route Choice for Transport Network Analysis. *Environment and Planning B: Planning and Design*, **34**(3), 539–555.

URA. 2014. *Master Plan*. Tech. rept. Urban Redevelopment Authority, Singapore.

URA. 2016. *Real Estate Informations System* {*REALIS*}. http://www.ura.gov.sg/realis/.