Journey Estimation with Smartcard Data for Land Use Planning

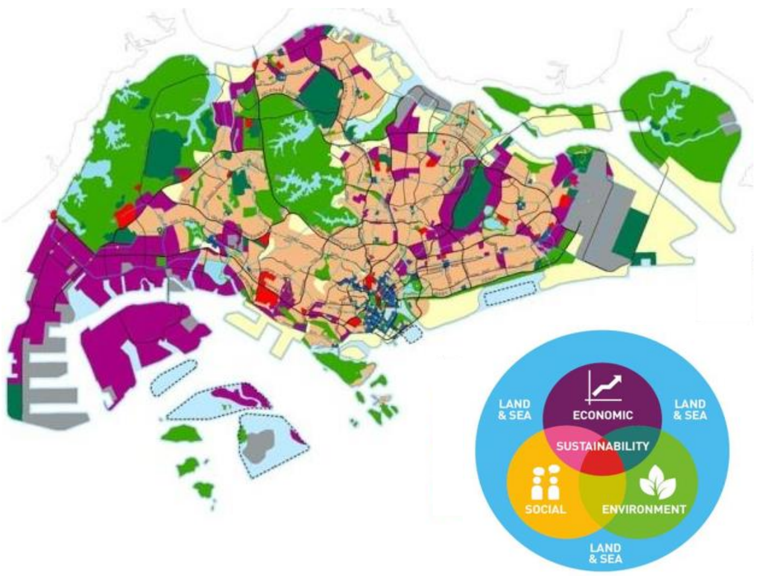
Urban Redevelopment Authority
Digital Planning Lab

Alvin CHUA (Corresponding Author)
WANG Songyu

Land Use Planning in Singapore

Planning for a Smart and Sustainable City

Concept Plan



Strategic land use and transportation plan guiding development over the next 40-50 years

Project future land demand for homes, employment centres, ports & key industries

Master Plan

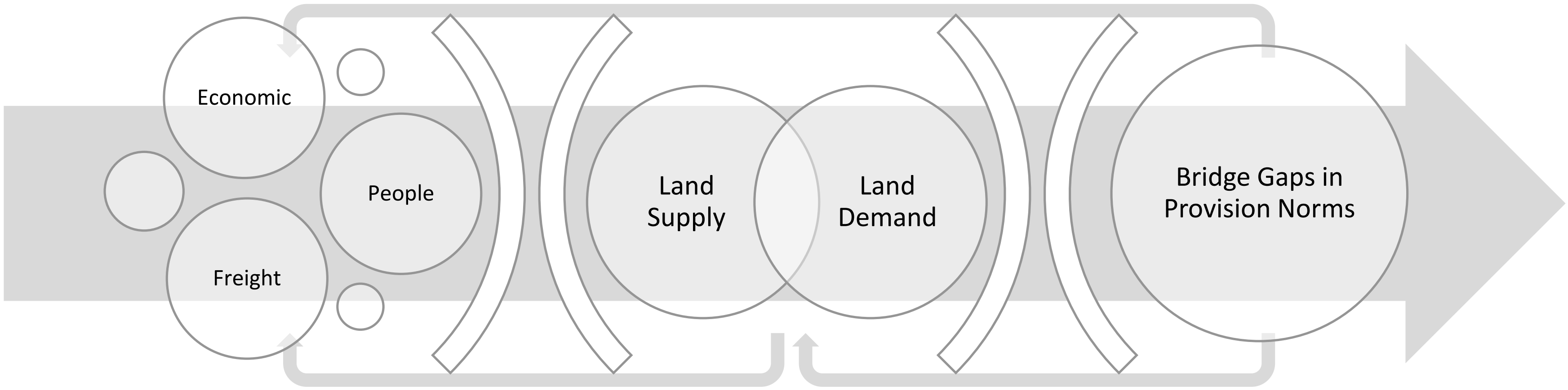
Translates concept plan into detailed plans for near term implementation over 4-5 years

Ensure adequate facilities & infrastructure to provide a comfortable living environment

Urban Redevelopment Authority

Land Use Planning in Singapore

How Data Analytics Inform Land Use Planning



Build a comprehensive understanding of past & present activity patterns

Model future demand to ensure sufficient supply of land

Refine planning policies

Urban Flow Analysis for Land Use Planning [4]

- Identify location of activities, induced by different facilities island wide.
- Inform near term facility siting and sizing as well as the alignment of mobility infrastructure connecting them.

Using Smartcard Data to Study People Movement in Singapore

- Bulk of population (~60%) commute by public transport [9].
- Smart card data provides granular information of public transport trips.

Journey Estimation with Smartcard Data

Deriving Journeys from Smartcard Trip Data for Urban Flow Analysis

- ***Journeys*** refer to one-way travel between two locations for specific purposes besides transfer.
- Smart card data contains information of every public transport trip in Singapore (i.e. unique card ID, origin, destination, time start & stop) but does not contain ready information on ***journeys***.
- ***Journeys*** may consists of more than one trip if it involves transfers between services (e.g. bus to rail, rail to bus, bus to bus).
- Hence there is a need to distinguish consecutive trips that involve transfer(s) between services, from those that stop for other activities.
- To derive an accurate estimate of journeys, trip chaining [1] is required for trips involving transfers.
- Yet, journey estimation is affected by the assumptions of what constitutes a transfer, based on the time interval between consecutive trips.

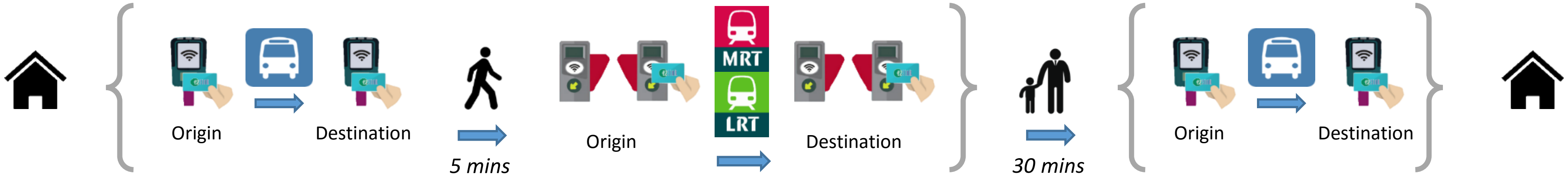
Transfer Time Intervals

- Singapore adopts a 45 minute transfer time interval for ticketing purposes, yet many types of activities are possible within the time span.
- Past studies have applied a range of timings (30 – 90 minutes) for transfer [1][2][8] but did not study in detail whether these intervals reasonably reflect the actual transfer time.
- Amount of time needed for transfer is influenced by many factors like the frequency of services as well as the distance between alighting and boarding locations.
- A recent study in Singapore reported that transfer time varies by location, time of day, commuter demographics, crowdedness at train stations & bus stops, as well as the pace of walking [7].

Effect of Journey Estimation with 45 Minute Transfer Time Interval

Undercounting Journeys

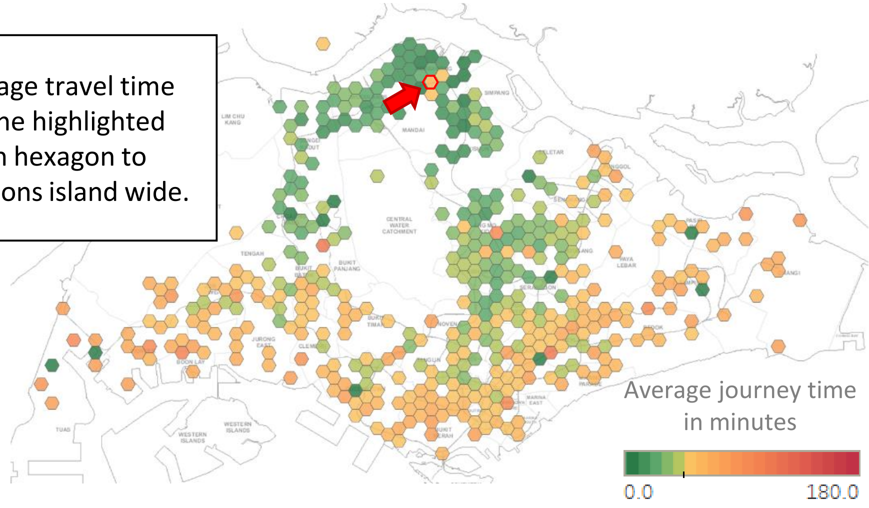
Two distinct journeys are counted only once.



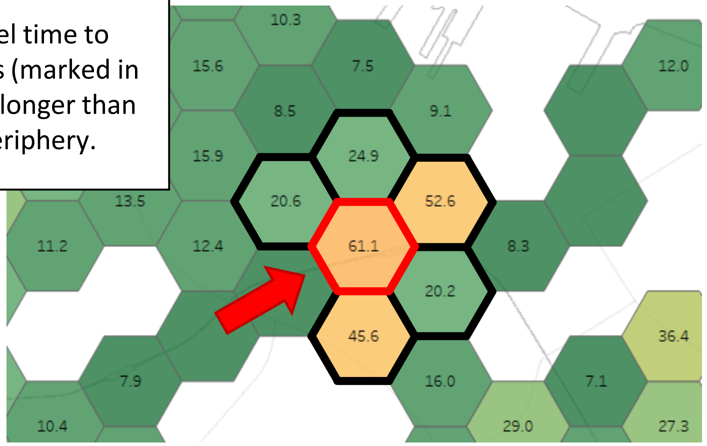
Over-estimating Average Travel Time

Irregularly long trip chains will skew journey travel time statistics.

a) Average travel time from the highlighted origin hexagon to destinations island wide.



b) Average travel time to adjacent hexagons (marked in black) observably longer than those in the periphery.



We contribute a practical approach to identify more representative transfer time intervals to address the gaps in journey estimates.

Methodology

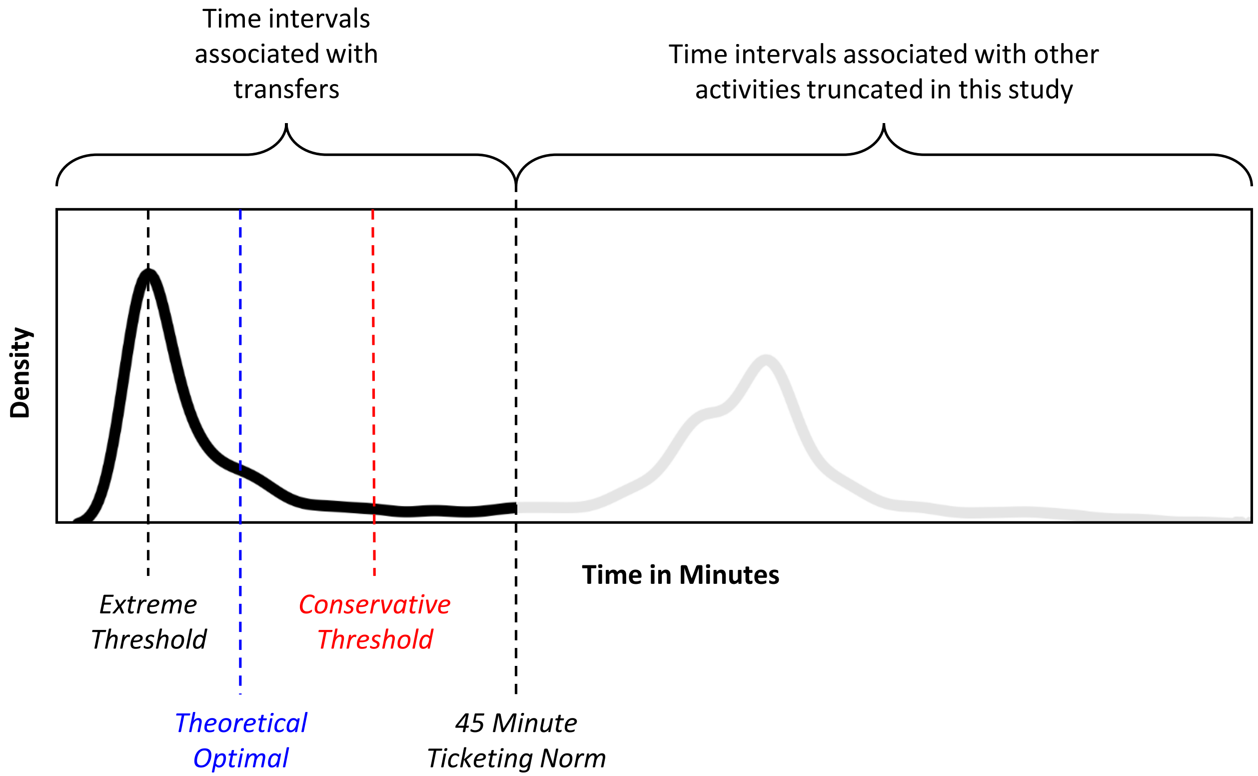
Using time between consecutive trips at regional transport hubs to identify transfer time between services.

- Singapore's transport infrastructure is based on a hub & spoke model.
- Most transfers occur at the regional transport hubs.
- Transport hubs are also major activity centres where commuters may stop for other purposes (e.g. retail, recreation, etc.).
- Challenge is to identify the threshold time interval that best represents only transfer activities.
- Thus, plausible threshold time intervals should be identified based on the distribution of commuter time interval between consecutive trips.
- This can be analysed as a transfer time interval Probability Density Function (PDF).

Thresholds Evaluated

- ***45 Minute Ticketing Norm (TN)*** *which is the current criteria* will be the baseline for comparison to other thresholds.
- 95th Percentile is a ***Conservative Threshold (CT)*** expected to exclude irregularly long time intervals, which are unlikely to be transfers.
- Inflection point, the ***Theoretical Optimal (TO)*** threshold we believe transfers would not exceed.
- Mode, an ***Extreme Threshold (ET)*** we observe most transfers occurring at.

Diagrammatic representation of transfer time interval PDF & the thresholds evaluated in this study



Data Processing

Smart card data fields & calculations to derive the time interval between consecutive trips.

Let $t\{C, En, Ex, In\}$ represent a trip in sequence $T\{t_1, t_2, ..., t_n\}$ sorted in ascending order by En & Ex respectively.

C	Unique Card ID *(actual ID truncated)*
En	Entry *(trip start time)*
Ex	Exit *(trip stop time)*
In	Time interval between consecutive trips

where

$$In_n = \begin{cases} En_{n+1} - Ex_n & C_n = C_{n+1} \\ NA & \textit{otherwise} \end{cases}$$

Tabular extract of smart card data fields & time intervals derived from calculations

Card Id	Entry	Exit	*Time Interval*	Transport Mode	Origin	Destination
A93EE	1/7/2016 09:38	1/7/2016 09:44	*3.95*	BUS	OPP TAMPINES JC	TAMPINES INT
A93EE	1/7/2016 09:48	1/7/2016 10:31	*695.1333333*	RTS	TAMPINES MRT STATION	HARBOURFRONT MRT STATION
A93EE	1/7/2016 22:06	1/7/2016 22:54	*6.516666667*	RTS	HARBOURFRONT MRT STATION	TAMPINES MRT STATION
A93EE	1/7/2016 23:00	1/7/2016 23:12	*NA*	BUS	TAMPINES INT	NGEE ANN SEC SCH
FEDE9	1/7/2016 19:06	1/7/2016 19:16	*1.85*	BUS	BLK 109B	QUEENSTOWN STN
FEDE9	1/7/2016 19:18	1/7/2016 19:45	*21.46666667*	RTS	QUEENSTOWN MRT STATION	PIONEER MRT STATION
FEDE9	1/7/2016 20:06	1/7/2016 20:38	*9.766666667*	RTS	PIONEER MRT STATION	QUEENSTOWN MRT STATION
FEDE9	1/7/2016 20:48	1/7/2016 20:58	*NA*	BUS	QUEENSTOWN STN	BLK 118

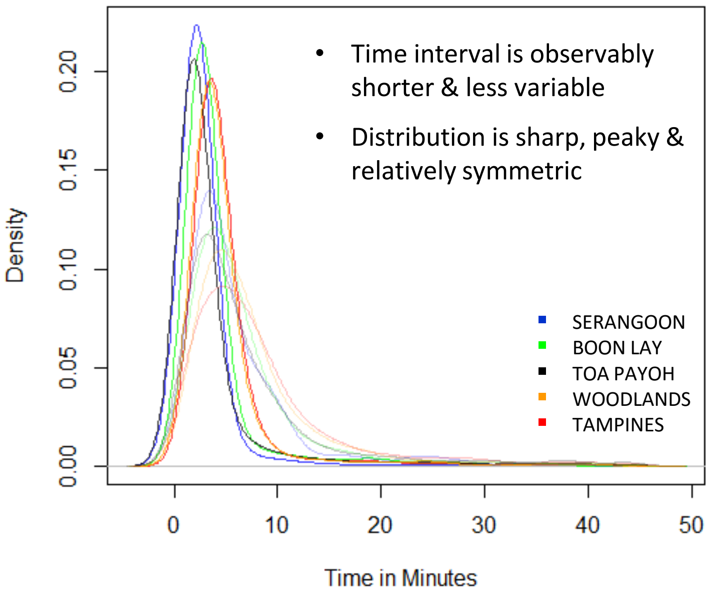
Data fields to estimate transfer time interval PDFs

Data fields for time interval calculation

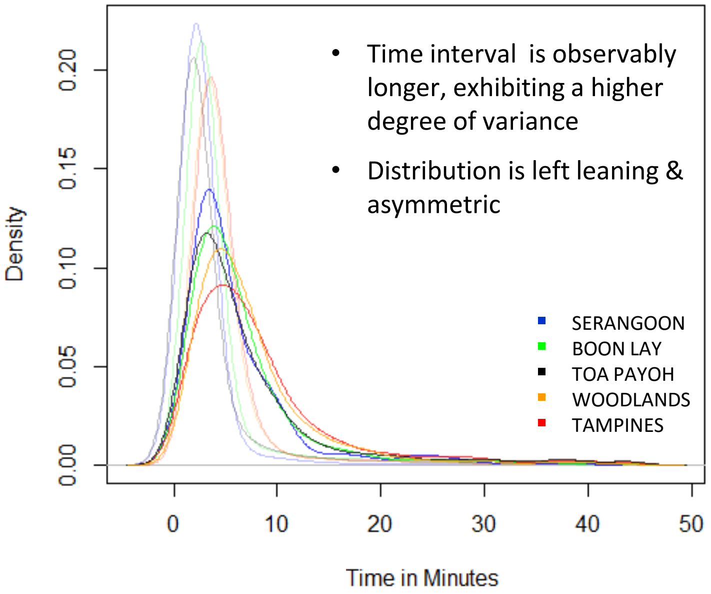
Transfer Time Modelling & Threshold Estimation

Transfer Time interval PDFs at regional transport hubs. We model with observations from 8 - 9 am as commuters tend to stop for transfers at transport hubs for journeys to work [3].

Towards Rail

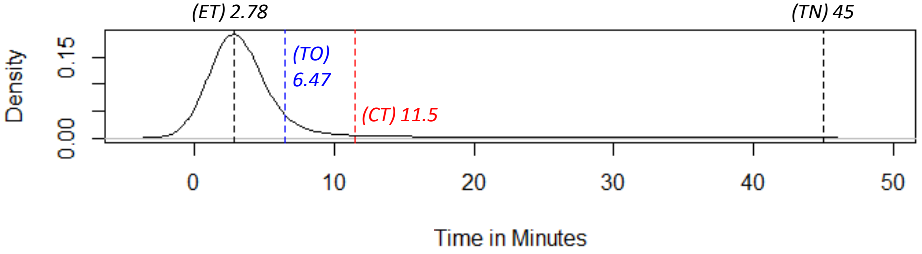


Toward Bus

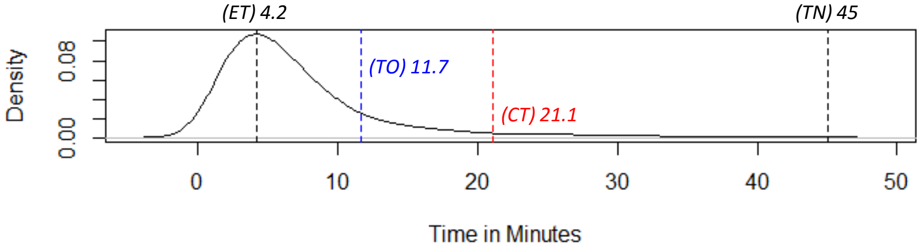


Joint probability density functions of dwell times at regional transport hubs generated with kernel density estimation, and different parameters for evaluation.

Towards Rail



Toward Bus



Urban Redevelopment Authority

Journey Estimates Comparison with Different Threshold

	Unchained Trips	**Ticketing Norm (TN)** *45 mins*	**Conservative Threshold (CT)** *to train = 11.5 mins,* *to bus = 21.1 mins*	**Theoretical Optimal (TO)** *to train = 6.47 mins,* *to bus = 11.7 mins*	**Extreme Threshold (ET)** *to train = 2.78 mins,* *to bus = 4.2 mins*
N. Origins	4,439	4,331	4,353	4,396	4,429
N. Destinations	4,612	4,509	4,537	4,572	4,604
N. OD Pairs	193,526	235,441	243,896	250,889	244,674
N. Journeys	-	340,032	354,532	371,628	455,024
N. Trips	556,588	-	-	-	-

Key Observations

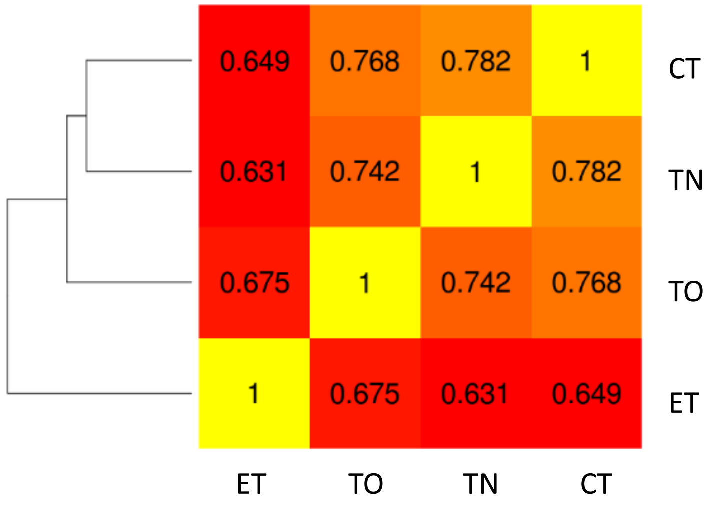
- Broadly, the number of origins, destinations, journeys and OD pairs increase as the thresholds become lower.
- The number of origins and destinations in ET is close to Unchained Trips. While some journeys in ET are valid, the transfer time threshold is too short for a good number of trips resulting in a sizable overestimation of journeys, and exclusion of plausible unique OD pairs.
- CT & TO provides more representative estimates which includes journeys that were previously undercounted in TN.
- TO exhibits the most desirable outcomes in general because we were able to recover more journeys while retaining a higher number of unique OD pairs.

Quantitative Evaluation

Journey estimates with different threshold benchmarked quantitatively to verify findings. Both metrics evaluate the connectivity between origins and destinations based on the number journeys occurring between them. Results of evaluation agree with earlier comparisons in that (1) ET differs from CT, TN & TO, alluding to substantial over counting in the former; (2) TO appears further from TN than CT indicating some degree of under counting in the latter.

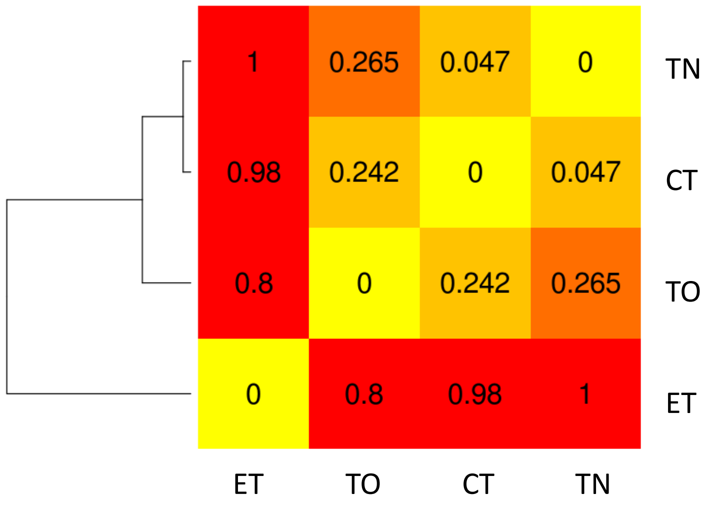
Geoffrey E. Havers (GEH) Distance

- Metric from transportation literature to assess the similarity of estimated results [8].
- GEH score is computed for every OD pair in each set of estimates.
- The percentage of ODs with GEH equal or less than 5 is calculated to indicate the closeness between two sets.
- A score of 1 suggest that two sets of journey estimates are similar while 0 suggests that both are different.

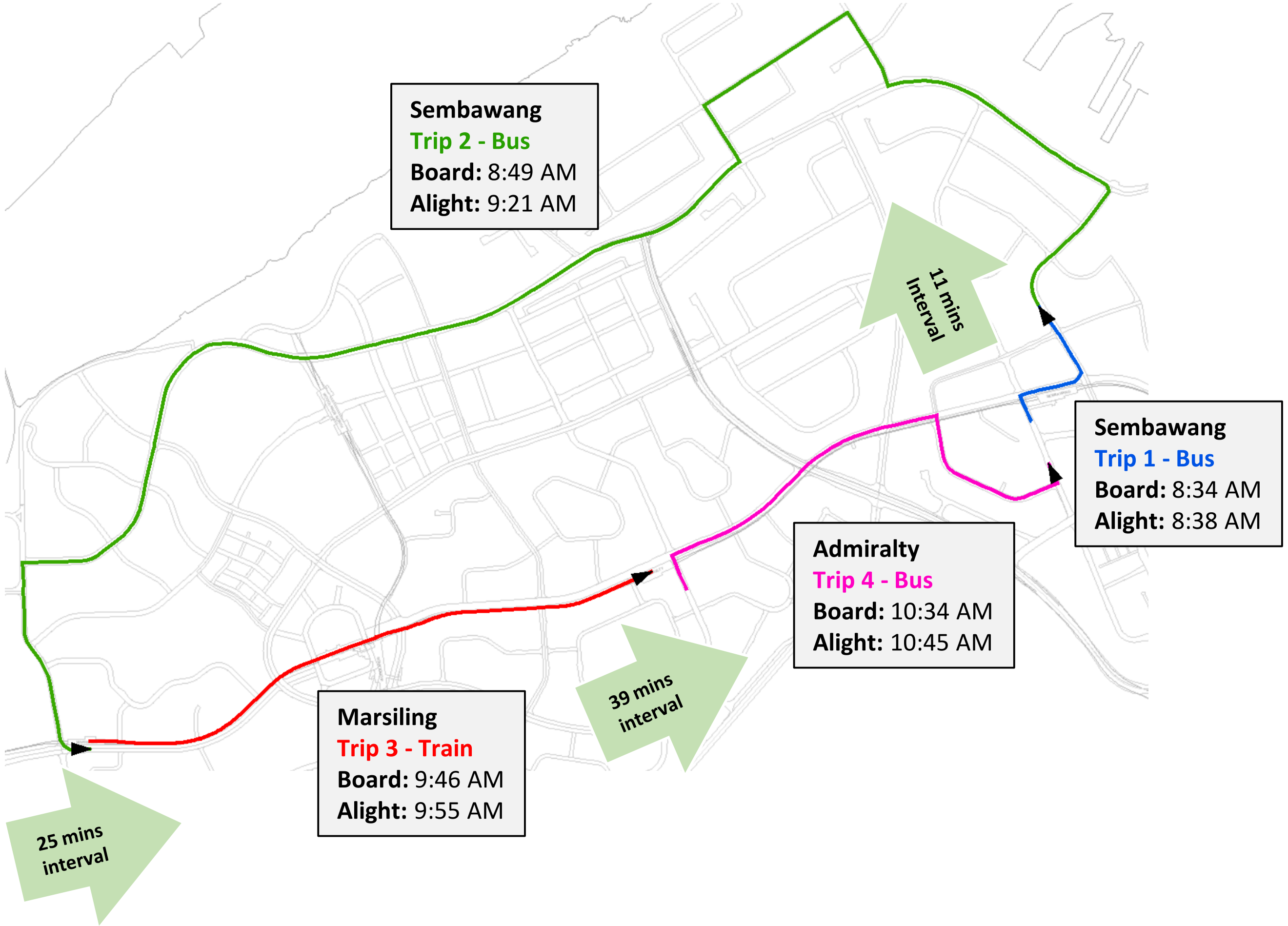


Hamming-Ipsen-Mikhailov (HIM) Distance

- Metric with general application in many scientific fields [5] to assess the similarity of estimated results.
- Derived from a linear combination of edit and spectral distances.
- A score of 1 suggests two matrices are different while 0 suggests that both matrices are the same.



Example: A commuter's Trips Around Woodlands



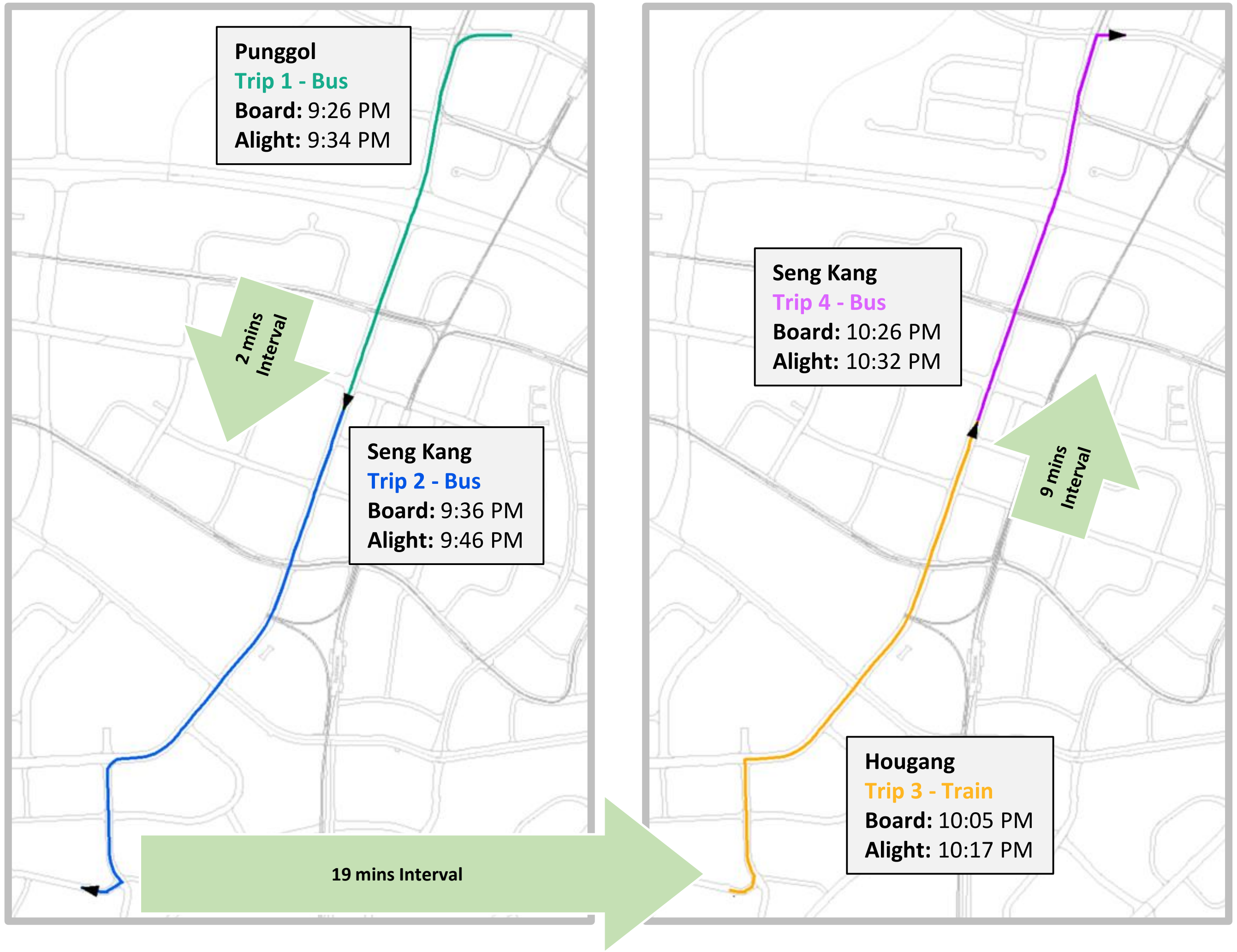
Time Interval Between Trips

Trip Sequence	Time Interval	Ticketing Norm (TN)	Theoretical Optimal (TO)	Extreme Threshold (ET)
1 → 2	11 mins (→ Bus)	✓ (≤ 45mins)	✓ (≤ 11.7mins)	✗ (> 4.2mins)
2 → 3	25 mins (→ Train)	✓ (≤ 45mins)	✗ (> 6.47mins)	✗ (> 2.78mins)
3 → 4	39 mins (→ Bus)	✓ (≤ 45mins)	✗ (> 11.7mins)	✗ (> 4.2mins)

Derived Journeys

	Trip Chaining		
Derived Journey ID	Ticketing Norm (TN)	Theoretical Optimal (TO)	Extreme Threshold (ET)
Journey A	1 → 2 → 3 → 4	1 → 2	1
Journey B	-	3	2
Journey C	-	4	3
Journey D	-	-	4

Example: A commuter's Trips Through Seng Kang



Time Interval Between Trips

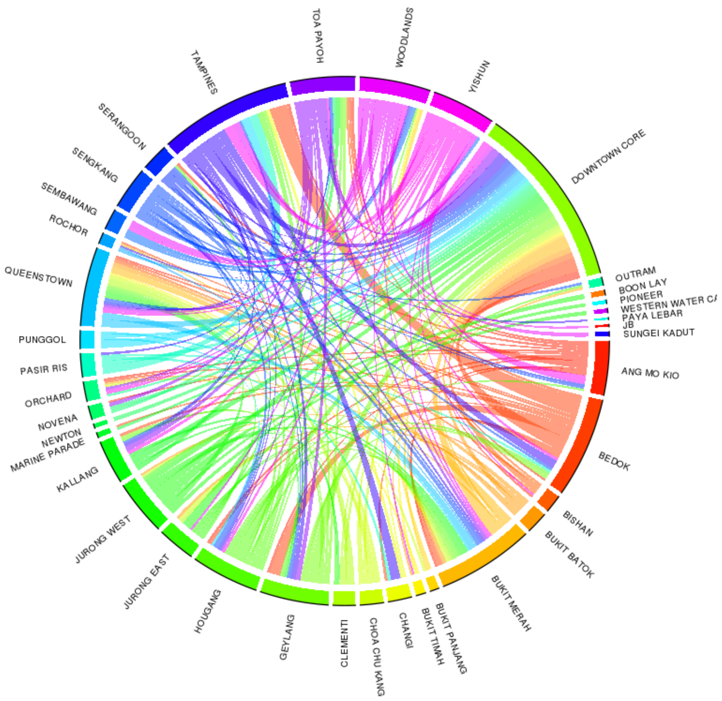
Trip Sequence	Transfer Time Interval	Ticketing Norm (TN)	Extreme Threshold (ET)	Theoretical Optimal (TO)
1 → 2	2 mins (→ Bus)	✓ (≤ 45mins)	✓ (< 4.2mins)	✓ (≤ 11.7mins)
2 → 3	19 mins (→ Bus)	✓ (≤ 45mins)	× (> 4.2mins)	× (> 11.7mins)
3 → 4	9 mins (→ Bus)	✓ (≤ 45mins)	× (> 4.2mins)	✓ (≤ 11.7mins)

Derived Journeys

Derived Journey ID	Ticketing Norm (TN)	Extreme Threshold (ET)	Theoretical Optimal (TO)
Journey A	1 → 2 → 3 → 4	1 → 2	1 → 2
Journey B	-	3	3 → 4
Journey C	-	4	-

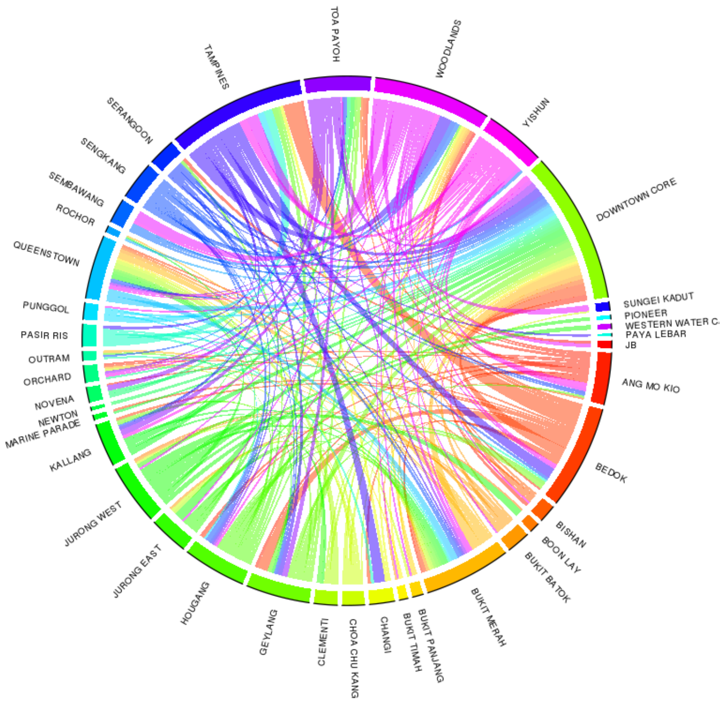
Journey Comparison by Origin & Destination

Journeys aggregated to coarse grain zonal boundaries for visual comparison. Here, we Illustrate the gain from journey estimation with the TO transfer time interval.

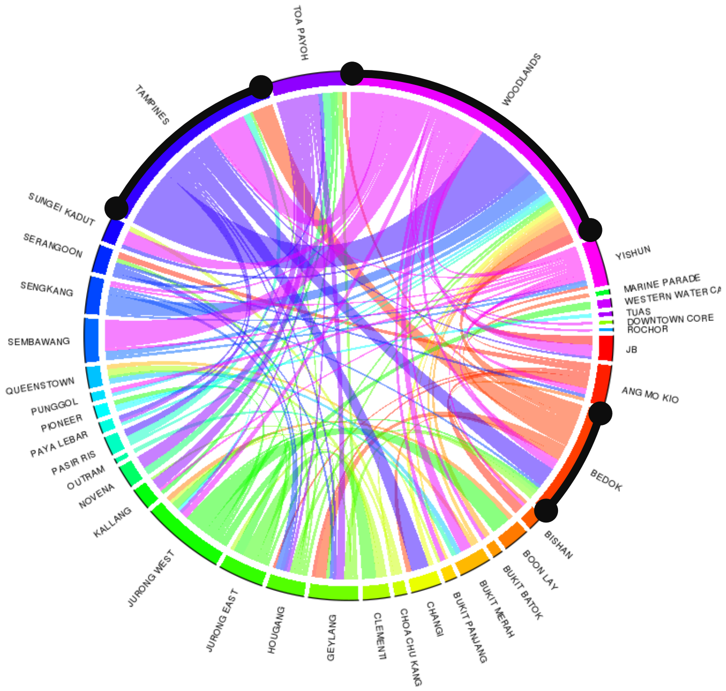


Ticketing Norm (TN)

Baseline for comparison



Theoretical Optimal (TO)



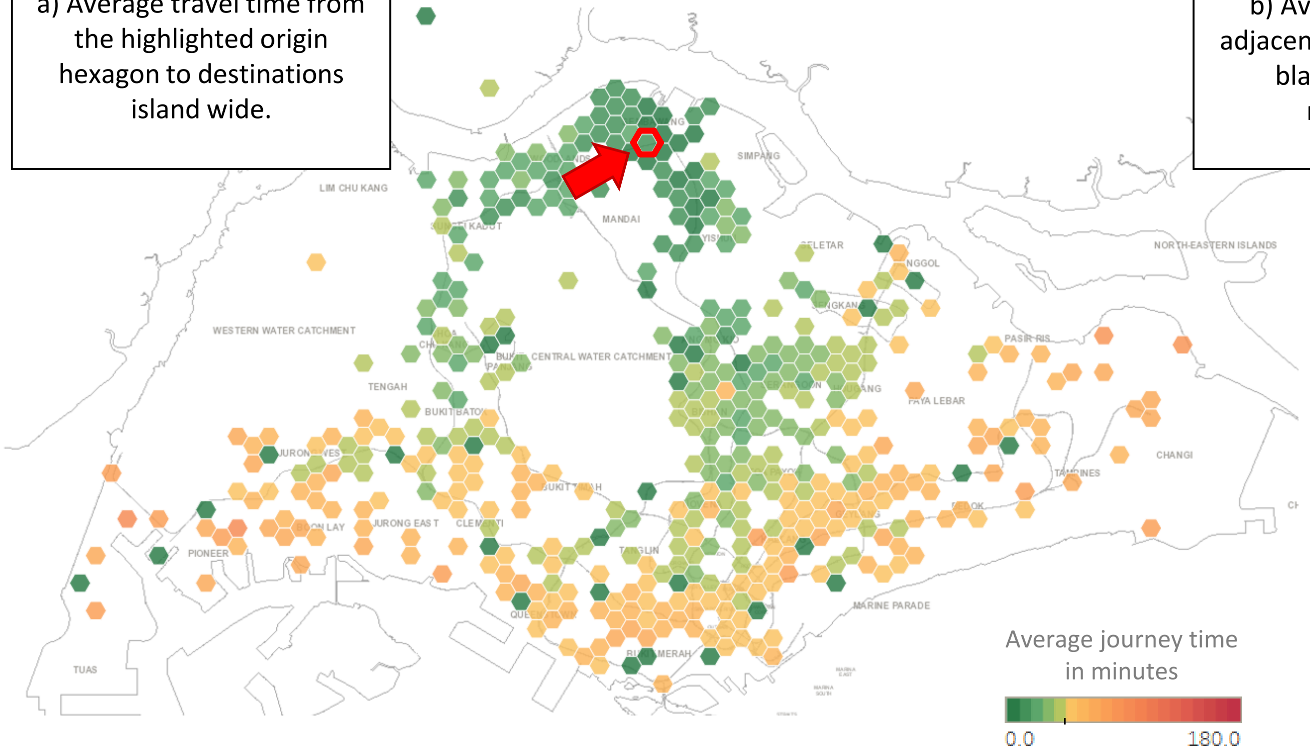
Recovered Journeys in TO that were not in TN *(TO – TN)*

Marked regions indicate where majority of OD pairs were retained

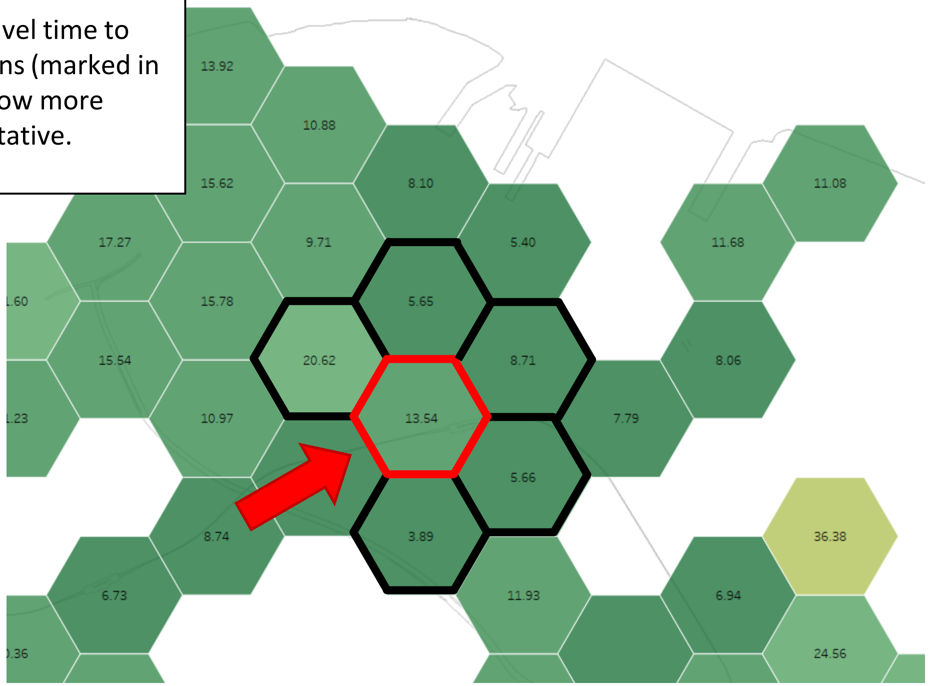
Urban Redevelopment Authority

Travel Time Statistics based on Journeys Estimated with the Theoretical Optimum Threshold

a) Average travel time from the highlighted origin hexagon to destinations island wide.



b) Average travel time to adjacent hexagons (marked in black) are now more representative.

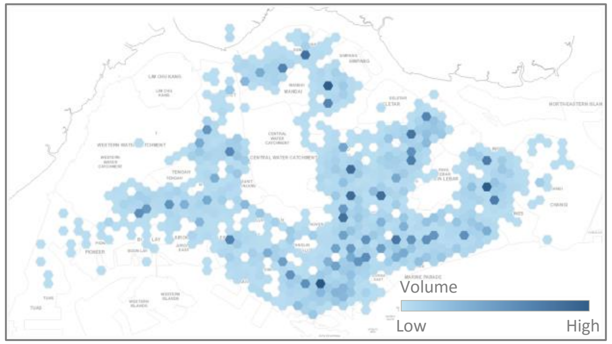


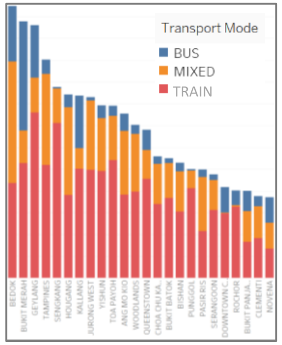
Analyse Catchment of Key Employment Nodes

Towards Destination Planning Areas: Outram, Downtown Core, Straits View & Marina South

Macro Analysis → Micro Analysis

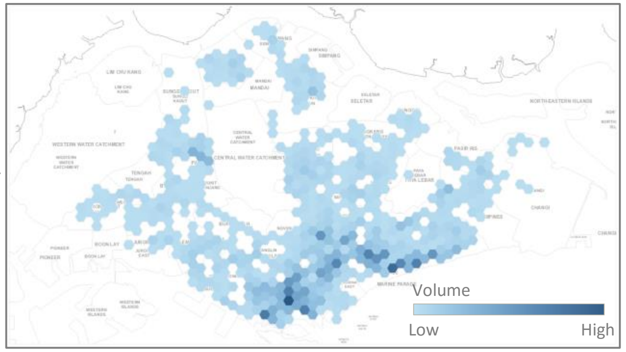
Catchment of Destination by various public transport modes



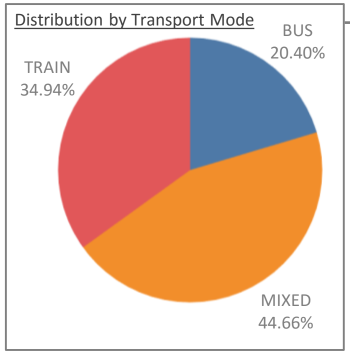


Volume by Planning Area

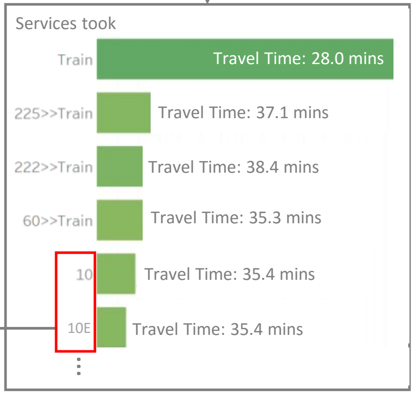
Filter
Bus
Only



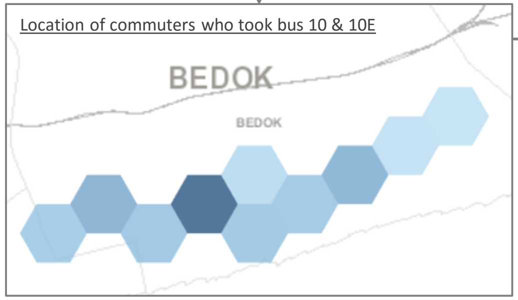
Local Scale Studies: Bedok to Destination



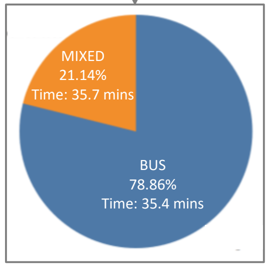
Breakdown by services took by commuters



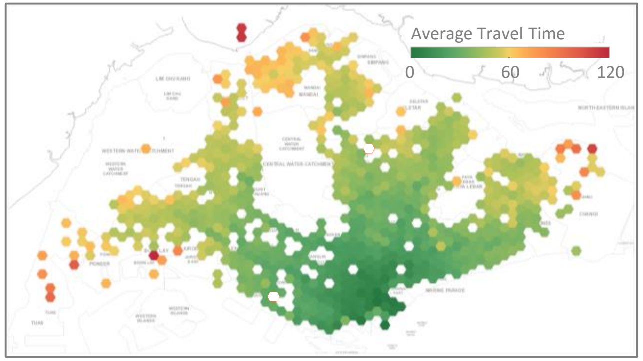
Locate areas served by buses



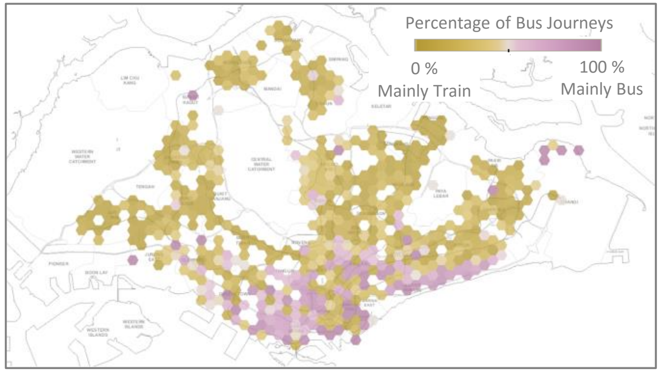
Understanding PT Mode Choice Preference of Commuters in these areas



Average zonal Origin Destination travel time



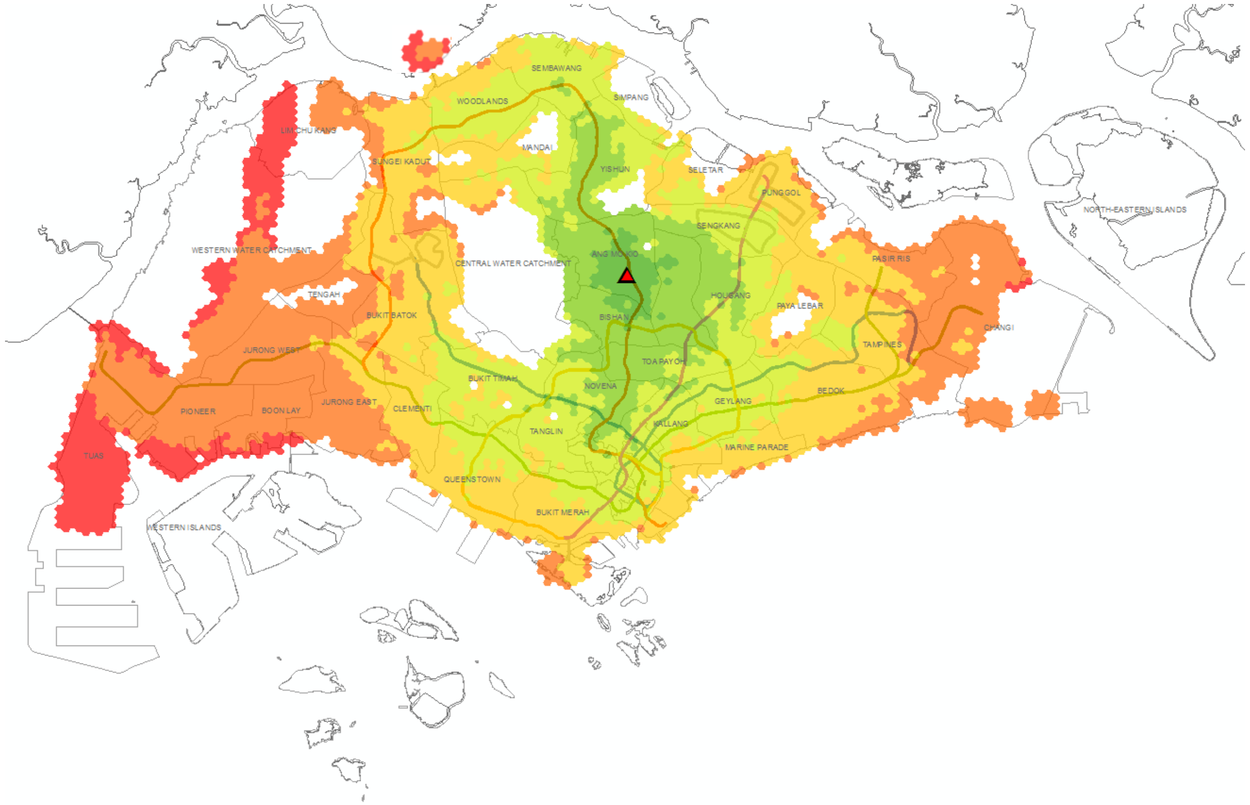
Public transport mode share



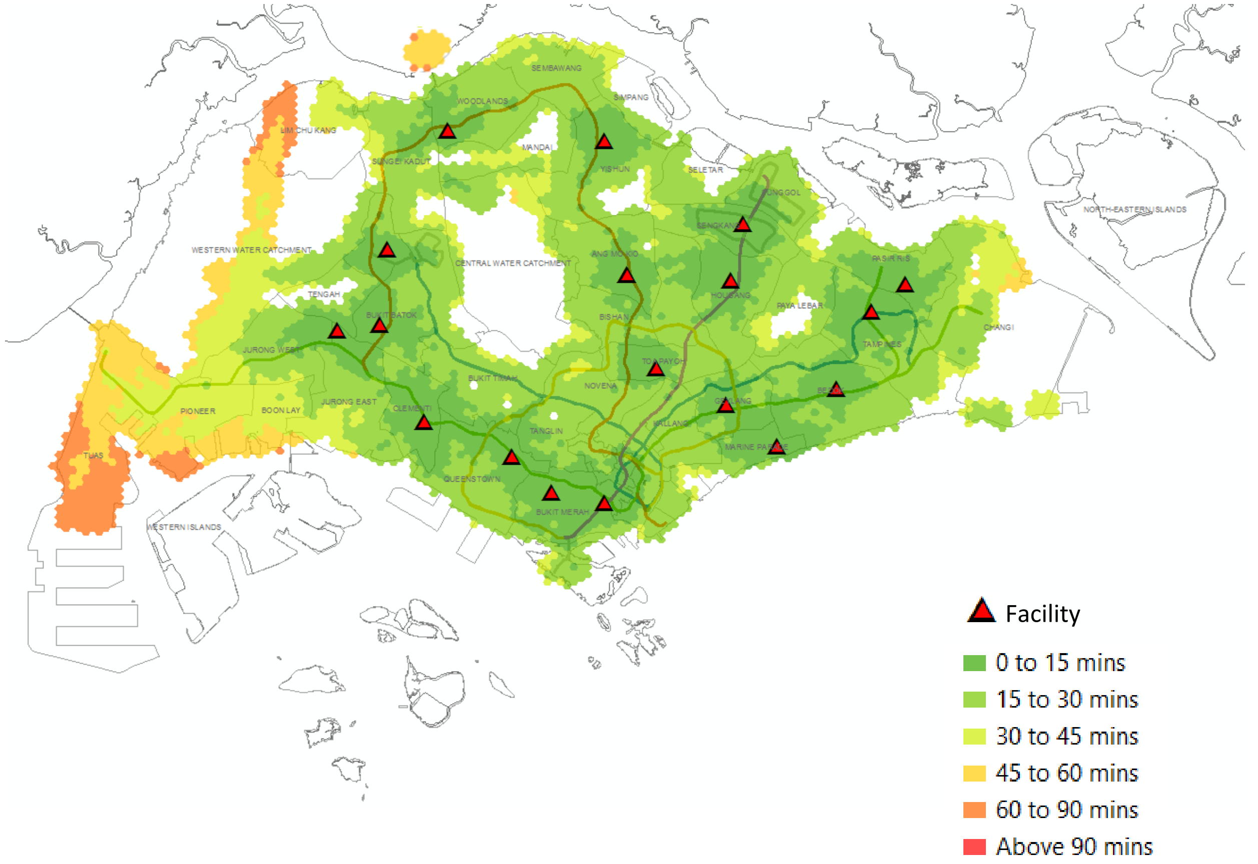
Analyse Population Access to Opportunities & Services

Evaluate facility service area with estimated public transport travel time as criteria

Single Facility



Multiple Polyclinics



Travel time estimates do not take into account the possibility of missing trains or buses.
However, the estimated travel time does consider traffic conditions and average waiting time during transfers.

Summary

- Journey estimation with 45 Minute Transfer time threshold result in inaccurate journey counts and average travel time statistics.
- We contribute a practical approach to identify more representative transfer time intervals that address the gaps in journey estimates. The approach is novel for the purpose of land use planning in Singapore.
- Our approach identifies a theoretically optimal transfer time interval threshold that best represents only transfer activities.
- We compare the journeys estimated with our theoretically optimal threshold to three other thresholds and show that our approach recovered more journeys and retained a higher number of unique OD pairs.
- While our theoretically optimal threshold provides better journey estimates for land use planning, the 45 Minute Transfer time threshold is still necessary for ticketing and fare revenue estimation.

Future Work

- Our approach to identify representative transfer time interval is based on observations at regional transport hubs. To further improve journey estimation, the approach maybe extended to every bus stop and train station island wide, for location specific transfer time intervals. While theoretically feasible, this will require substantial computational resources.
- The proposed approach is also generalizable to other domains like freight & facilities management with similar data for deeper insights into movement patterns.

References

1. Alsger, A. A., Mesbah, M., Ferreira, L. & Safi H. **Public Transport Origin-Destination Estimation Using Smart Card Fare Data.** In: Transportation Research Board 94th Annual Meeting, Washington DC, USA, 2015.

2. Alsger, A.A., Mesbah, M., Ferreira, L. and Safi, H., 2015. **Use of smart card fare data to estimate public transport origin–destination matrix**. Transportation Research Record: Journal of the Transportation Research Board, (2535), pp.88-96.

3. Andris, C. and Ferreira, J., 2014. **Featured graphic. Visualizing commuting in Singapore.** Environment and Planning A, 46(11), pp.2543-2545.

4. Chua, A., Marcheggianni, E., Servillo, L., and Vande Moere, A. V. 2014. **Flow Sampler: Visual analysis of urban flows in geolocated social media data.** In: International Conference on Social Informatics, pp. 5-17.

5. Jurman, G., Visintainer, R., Filosi, M., Riccadonna, S. and Furlanello, C., 2015. **The HIM glocal metric and kernel for network comparison and classification.** In: IEEE International Conference on Data Science and Advanced Analytics (DSAA) 2015, pp. 1-10.

6. Rao, L. **Backed by Google Ventures and Eric Schmidt, Urban Engines wants to solve urban congestion using data intelligence**, 2014. Retrieved May 12, 2018, from Techcrunch: *https://techcrunch.com/2014/05/15/backed-by-google-ventures-and-eric-Schmidt-urban-engines-wants-to-solve-urban-congestion-using-data-intelligence/*

7. Sun. L., Jin J. G., Lee D. H., Axhuasen, K. W., 2015. **Characterizing travel time reliability and passenger path choice in a metro network.** In: Transportation Research Board 94th Annual Meeting, Washington DC, USA, 2015.

8. Tavassoli, A., Alsger, A., Hickman, M. and Mesbah, M., 2016. **How Close the Models are to the Reality? Comparison of Transit Origin-destination Estimates with Automatic Fare Collection Data.** In: Australasian Transport Research Forum (ATRF), 38th, 2016, Melbourne, Victoria, Australia.

9. **Making Public Transport the Choice Mode**, Retrieved May 12, 2018, from Ministry of Transport: *https://www.mot.gov.sg/about-mot/land-transport/public-transport*