Detecting weak public transport connections from cellphone and public transport data

Thomas Holleczek
SingTel
Singapore
thomasas@singtel.com

Liang Yu
Institute for Infocomm
Research
Singapore
yul@i2r.a-star.edu.sg

Joseph Kang Lee
University of British Columbia
Vancouver, Canada
josephkanglee@gmail.com

Oliver Senn
Singapore-MIT Alliance for
Research and Technology
Singapore
oliver.senn@smart.mit.edu

Carlo Ratti
Massachusetts Institute of
Technology
Cambridge, MA
ratti@mit.edu

Patrick Jaillet
Massachusetts Institute of
Technology
Cambridge, MA
jaillet@mit.edu

ABSTRACT

Many modern and growing cities are facing declines in public transport usage, with few efficient methods to explain why. In this article, we show that urban mobility patterns and transport mode choices can be derived from cellphone call detail records coupled with public transport data recorded from smart cards. Specifically, we present new data mining approaches to determine the spatial and temporal variability of public and private transportation usage and transport mode preferences across Singapore. Our results, which were validated by Singapore's quadriennial Household Interview Travel Survey (HITS), revealed that there are 3.5 million public and 4.3 million private inter-district trips (HITS: 3.5 million and 4.4 million, respectively). Along with classifying which transportation connections are weak, the analysis shows that the mode share of public transport use increases from 38 % in the morning to 44 % around mid-day and 52 % in the evening.

Categories and Subject Descriptors

H.4 [**Information Systems Applications**]: Miscellaneous

General Terms

Public transport, private transport, clustering, cellphone call detail records

1. INTRODUCTION

Securing public transportation ridership is critical for developing a sustainable urban future. However, existing systems for analyzing and identifying weaknesses in public transport connections face major limitations. In cities, origin-

destination (OD) matrices—which measure the flow of people between different geographical regions—are often generated using household surveys and roadside monitoring, approaches which are time consuming, expensive, and lack spatial and temporal accuracy [9]. Focus in more recent research has been drawn towards using cellphones to overcome the limitations mentioned above—using the cellphone fluctuations in GSM signal strength [18, 7, 8] and location, speed, and acceleration estimates obtained through GPS [16, 21, 19] to detect individual trips and categorize them into classes such as walking, cycling, public transport and driving a car. These approaches are indeed innovative and capture in detail individual travel behavior, but are limited by their sample sizes (e.g. number of volunteers) and currently face difficulties scaling up. While these methods will continue to be valid sources of data and insight, there are emerging opportunities to analyze urban transportation networks using data mining approaches, specifically in using cellphone call detail records to detect spatio-temporal patterns in urban mobility and transport mode preferences.

As cities not only become denser but also more extensive, there is an increasing pressure to maintain the connectivity and accessibility of goods and services through public transportation networks and integrated urban design. However, despite operating highly effective public transportation systems, modern and growing cities such as Singapore are facing declines in public transport usage; local governments and city planners are often equipped with few tools to explain where and why such declines are occurring [11]. Given that cities are becoming increasingly digitized and with the pervasiveness of cellphones, new data-driven methods are surfacing to help understand and explain characteristics of urban mobility [20, 17, 9, 10, 13, 15]. Unlike previous work, we derive the mode share of public and private transportation (including taxis) by examining multiple datasets, namely cellphone call detail records and a public transportation smart card system.

In this article, we use 3.4 million cellphone users' call detail records to derive the inter-district (55 districts) mobility of people in the dense, metropolitan city-state of Singapore. We then subtract the number of Singapore's public transportation smart card system users from our calculated value

of urban mobility to estimate the share of private transportation users between districts. By examining the spatial and temporal patterns of public and private transportation, we can determine to and from where people prefer to take public or private transit and furthermore classify the time periods and the districts that may be considered weakly connected by public transportation. Ultimately, we present a novel method of using cellphone data to map the mobility of people in cities and posit that its application can not only benefit urban transportation and planning efforts but also other domains focused on or affected by the movement of people across space.

2. DATASETS

We use cellphone call detail records and a public transport dataset to estimate the flows of people using public and private transport between the 55 administrative districts of Singapore.

2.1 Cellphone dataset

The cellphone dataset consists of location data of 3.4 million anonymized users of SingTel, Singapore's largest telecommunications company with a market share of 45.3 %. The data were recorded during a two-month period from mid-March to mid-May 2011. A call detail record (CDR) includes the location of the cell tower each cellphone connects to and was created by the billing system of SingTel in the case of following network events:

- a phone call was initiated or received (at the beginning and at the end of the call).

- a short message was sent or received.

- the cellphone user accessed the data network (for example, to open a website or retrieve emails).

There is no information about handovers between base stations during long phone calls. Logging in and out of the cellphone network was also not recorded.

We use the location of the base station a cellphone connects to as the location of a cellphone user and estimate the localization error as follows. The average service radius of a base station is approximately $r = 212$ m. Due to confidentiality of the data, the locations of the base stations were offset by a maximum of 200 m. This leads to an average uncertainty radius of 412 m for the location estimation of cellphones. However, the cellphone might not always connect to the closest base station (for example, if the closest base station is busy), which can further increase the localization error. A study by Ferris et al. [12] shows that the median localization error can be reduced to 128 m with the help of Gaussian processes when considering the received GSM signal strength as well. The received signal strength is, however, not included in our dataset.

The location of a cellphone is only recorded in the case of the network events listed above, which means that subscribers using their cellphone frequently can be tracked more precisely. To measure how frequently cellphone subscribers access the cellular network, we introduce the network inter-event time. Let $\mathbf{t} = [t_1, t_2, \ldots t_n]$ be the timestamps of the network events related to the cellphone subscriber α. We
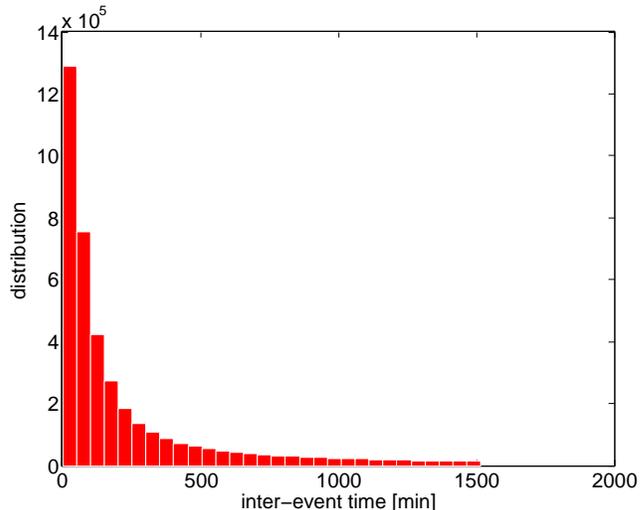

Figure 1: **To measure how frequently cellphone subscribers access the cellular network, we introduce the network inter-event time. The average inter-event time for the entire dataset is** 320 min**. The first quartile of the inter-event time is** $\tau_{25} = 41$ min**, the median is** $\tau_{50} = 114$ min **and third quartile is** $\tau_{75} = 406$ min**.**

then define the network inter-event time τ_α of α as

$$\tau_\alpha = \frac{1}{n-1} \sum_{i=2}^{n} (t_i - t_{i-1}). \qquad (1)$$

We determined the distribution of the inter-event time. The average inter-event time for the entire dataset is 320 min. The first quartile of the inter-event time is $\tau_{25} = 41$ min, the median is $\tau_{50} = 114$ min and third quartile is $\tau_{75} = 406$ min. The distribution of the inter-event time is shown in Figure 1.

To show that the cellphone dataset is representative of Singapore, we compare the distribution of the home locations of SingTel customers with official census data from 2010 [1]. Call detail records accumulate at home and work locations as these are the places where people spend most of their time [14]. To determine these clusters of call detail record locations, we apply the K-means clustering algorithm to the recorded locations of each subscriber, searching for at most $K = 10$ clusters. Next, we remove those location points from each cluster that are farther than 1.0 km away from the respective centroid to make sure that no outliers are assigned to clusters. Finally, we have to find home and work locations among the identified clusters. If the majority of call detail records in a particular cluster occurs between 6 pm and 6 am, we consider that cluster to be a potential home location. Otherwise, that cluster is assumed to be a potential work location. Among all potential home (work) locations of a cellphone subscriber, we consider the cluster with the most call detail records to be the real home (work) location.

Singapore is divided into 55 administrative districts [4]. We determine the population of SingTel customers in these districts based on the identified home locations (see Figure 2) and compare the results with the 2010 census data
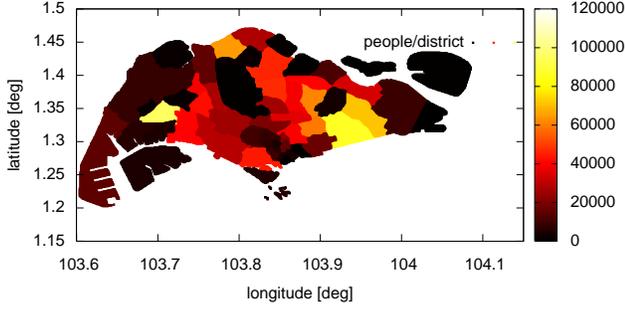
Figure 2: Population (number of home locations) per district determined from the SingTel dataset.
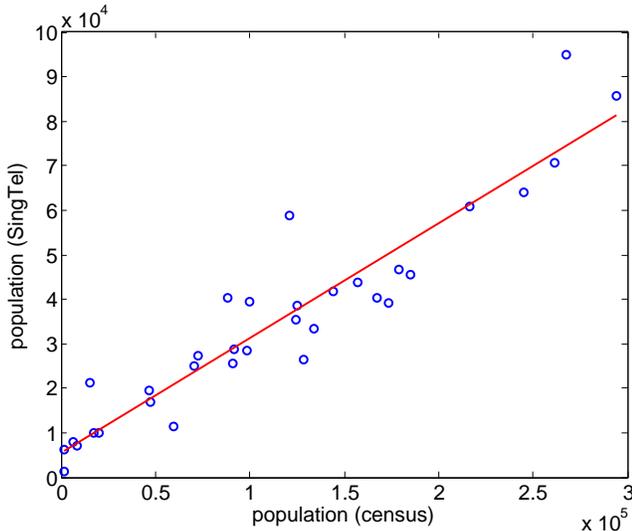

Figure 3: Population in Singapore's districts according to the 2010 census versus the population determined from the SingTel dataset. With a correlation coefficient of 0.94, the two distributions are highly linearly correlated.

[1]. With a correlation coefficient of 0.94, the two distributions are highly linearly correlated (see Figure 3).

2.2 Public transport dataset

The public transport dataset consists of trips made by 4.4 million anonymized users of Singapore's public transport system during a two-week period in April 2011, which overlaps with the recording period of the cellphone dataset. As mentioned earlier on, we consider taxi trips as private rather than public transport in this study. In Singapore, passengers use smart cards when getting on and off trains and buses [5]. The smart card system records the station and the time of departure and arrival for each trip. Single trips from the same passenger within a certain short period are merged as a combined trip which reflects the actual origin-destination pair.

3. APPROACH

The flow of people between different geographical regions $R = \{r_1, r_2, \ldots\}$ in a given time interval is commonly represented through a quadratic origin-destination (OD) matrix $A(t_s, t_e)$. $A_{i,k}(t_s, t_e)$ counts the number of trips from region r_i to region r_k that arrive in the time interval $[t_s, t_e]$. OD matrices can be aggregated and normalized over time, for example to get hold of the average number of trips on Mondays between 6 and 9 in the morning. We determine origin-destination matrices of Singapore on an hourly basis and average them over all workdays covered by the two datasets.

The flows of people can be partitioned into flows by public and by private transport, as people travel either by a private car (including taxis) or by the public transport system. That means, the overall mobility $A_{\mathrm{singapore}}$ in Singapore can be decomposed into the OD matrix A_{public} containing trips by public transport and the OD matrix A_{private} containing trips by private transport:

$$A_{\mathrm{singapore}} = A_{\mathrm{public}} + A_{\mathrm{private}}. \tag{2}$$

As geographical regions, we use the 55 administrative districts of Singapore. We estimate the overall mobility $A^*_{\mathrm{singapore}}$ in Singapore by upscaling the OD matrix A_{singtel} derived from the cellphone dataset (see 3.1) and determine the accurate public transport OD matrix A_{public} from the public transport dataset (see Section 3.2). To do this, we extract individual trips from the cellphone and the public transport datasets. We then map the start and end points of all determined trips to the 55 administrative districts and group them into hourly OD matrices. Finally, subtracting A_{public} from the estimated overall mobility $A^*_{\mathrm{singapore}}$ yields an estimate of the private transport OD matrix A^*_{private} (see Section 3.3).

As every public trip is captured by the public transport smart card system, the public transport OD matrix represents actual numbers. This means that the accuracy of A^*_{private} only depends on the accuracy of the overall mobility $A^*_{\mathrm{singapore}}$ estimated from the cellphone dataset. More specifically, every trip that is not detected in the cellphone dataset directly contributes to the underestimation of the number of private trips. At first sight, it might seem that private trips are more difficult to detect from the cellphone data as people driving a car are much less likely to use their cellphones than while traveling on a train or a bus. However, our trip extraction algorithm is based on the recognition of origins and destinations of trips and only requires cellphones be used before and after, but not during trips.

3.1 Overall OD matrices

We determine the mobility of SingTel customers A_{singtel} from the cellphone call detail records with a trip detection algorithm (Section 3.1.1) and estimate the overall mobility $A^*_{\mathrm{singapore}}$ by upscaling A_{singtel} to the entire population (Section 3.1.2).

3.1.1 Trip extraction

The idea behind our trip detection algorithm is that call detail records of a particular subscriber accumulate at origins and destinations of trips in the form of clusters. We use an approach inspired by Calabrese et al. [10] to detect these clusters:

- Let $\mathbf{x}_\alpha = [\mathbf{x}_\alpha^1, \mathbf{x}_\alpha^2, \ldots, \mathbf{x}_\alpha^n]$ denote the sequence of recorded

locations of the cellphone user α.

- Then, the consecutive locations from $\mathbf{x}_\alpha^k$ through to $\mathbf{x}_\alpha^{k+s}$ (with $1 < k < k+s < n$) can be combined if $\|\mathbf{x}_\alpha^k - \mathbf{x}_\alpha^i\| < \Delta d \, \forall i \in [k+1, k+s]$. We use the threshold value $\Delta d = 2\,\mathrm{km}$, which is sufficiently high to filter out jumps among adjacent cell towers that do not happen due to trips of the cellphone user in Singapore. We define the virtual location $\mathbf{y}_\alpha^p$ as the centroid of the combined locations:

$$\mathbf{y}_\alpha^p = \frac{1}{s+1} \sum_{i=k}^{s} \mathbf{x}_\alpha^i. \tag{3}$$

Let $\mathbf{y}_\alpha = [\mathbf{y}_\alpha^1, \mathbf{y}_\alpha^2, \ldots, \mathbf{y}_\alpha^m]$ denote the sequence of virtual locations determined for α.

- These virtual location do not necessarily have to be origins or destinations of trips as call detail records may also occur cumulatively during a trip (for example, when writing texts on a train) or a short interruption of a trip (for example when refueling at a gas station or when waiting for a bus). We therefore define clusters as those virtual locations that can be considered as origins and destinations of trips. A virtual location $\mathbf{y}_\alpha^k$ is referred to as a cluster if and only if there are at least two call detail records associated with $\mathbf{y}_\alpha^k$ and the time α spends at $\mathbf{y}_\alpha^k$ exceeds the threshold value Δt, which we select as $20\,\mathrm{min}$. Let $\mathbf{z}_\alpha = [\mathbf{z}_\alpha^1, \mathbf{z}_\alpha^2, \ldots, \mathbf{z}_\alpha^p]$ denote the sequence of clusters of α.

- We then define a trip $\mathbf{s}_\alpha^k$ as the path between the two consecutive clusters $\mathbf{z}_\alpha^k$ and $\mathbf{z}_\alpha^{k+1}$. $\mathbf{z}_\alpha^1$ is the origin of the first trip of α in the recorded period, and $\mathbf{z}_\alpha^p$ is the destination of the last trip. All other clusters $\mathbf{z}_\alpha^k$ serve as both origins and destinations of trips.

To determine the origin-destination matrix $A_{\mathrm{singtel}}(t_s, t_e)$, we map the start and end points of all detected trips to the 55 administrative districts of Singapore and add only these trips to $A_{\mathrm{singtel}}(t_s, t_e)$ whose *end time* is $\in [t_s, t_e]$.

3.1.2 Upscaling

Next, the OD matrix A_{singtel} must be upscaled to represent the overall mobility of Singapore. Taking into account the market share of SingTel, we estimate there were about 7.5 million cellphone subscribers in Singapore in 2011:

$$\begin{aligned} &\text{number of cellphone subscribers} = \\ &\frac{\text{number of SingTel subscribers}}{\text{SingTel market share}} = 7.5 \times 10^6. \end{aligned} \tag{4}$$

A population of around 5.2 million in 2011 [3] yields a cellphone penetration of about $144\,\%$ when not considering pre-paid cellphones. To estimate the overall mobility $A_{\mathrm{singapore}}^*$, we upscale A_{singtel} by dividing it by the SingTel market share and the cellphone penetration in Singapore:

$$A_{\mathrm{singapore}}^* = \frac{A_{\mathrm{singtel}}}{\text{market share} \times \text{cellphone penetration}}. \tag{5}$$

This scaling assumes that the travel behavior of SingTel customers represents the travel behavior of all people in Singapore.

3.2 Public transport OD matrices

Trips can be extracted fairly easily from the public transport dataset as each entry represents a single trip. Should the time between two consecutive trips of a particular person be less than $45\,\mathrm{min}$—which is the time allowed by the smart card system to transfer [6]—we consider this as transit time and combine these trips as we are only interested in the final destination of trips. Finally, the origins and the destinations of the trips are mapped to the 55 administrative districts of Singapore, and hourly public OD matrices can be determined. As each trip is captured by the smart card system, the resulting OD matrices are not an estimate but represent actual numbers of public transport trips.

3.3 Private transport OD matrices

Subtracting the public transport OD matrix A_{public} from the estimated overall mobility $A_{\mathrm{singapore}}^*$ yields an estimate of the number of people that do not take public transport:

$$A_{\mathrm{private}}^* = A_{\mathrm{singapore}}^* - A_{\mathrm{public}}. \tag{6}$$

As a trip is either public or private, the resulting OD matrix describes the flows of passengers using private transport.

4. EVALUATION

We use the results of the latest Household Interview Travel Survey (HITS) from 2008 to evaluate the performance of our trip extraction algorithm and the precision of the corresponding OD matrices. According to the HITS, $20\,\%$ of all trips remain in the same district. On the contrary, only $4\,\%$ of all trips remain in the same district according to our study. Short intra-district trips cannot be recognized from the cellphone dataset for the following two reasons:

1. In our study, we estimate the location of a cellphone as the location of the connecting base station. To filter out jumps between adjacent base stations, our trip recognition algorithm is based on a clustering algorithm. As a consequence, trips shorter than $2.0\,\mathrm{km}$ cannot be detected with our dataset.

2. If two consecutive trips are too short, it is unlikely even for very frequent users with an inter-event time of $60\,\mathrm{min}$ that a network event occurs, resulting in only one detected trip rather than two.

We therefore do not consider intra-district trips in our analyses. The number of samples of the latest HITS is not sufficient to compare the number of people traveling on specific connections (only one percent of all households were interviewed), which is why we focus on the overall number of trips and the mode share of public transport. Our mobility study shows a good correspondence with the latest HITS and a more recent study by the Land Transport Authority:

- We estimate the number of inter-district trips (both public and private) as 7.8 million per day. According to the HITS from 2008, there were 9.9 million trips per day in Singapore, 7.9 million of which were inter-district [11].

- According to our study, $45\,\%$ of the trips in Singapore are produced by public transport and $55\,\%$ by private transport (including taxi trips) on average. These results are backed up by a travel survey conducted by

the Land Transport Authority in 2011 (44 % by public transport and 56 % by private transport) [2].

These figures suggest that our trip detection algorithm for call detail records as well as the presented correction and up-scaling methods produce good estimates of the overall mobility in Singapore as well as the flows of private transport.

5. RESULTS

To identify weak public transport connections depending on the time of the day, we determine OD matrices for the morning (6 am to 10 am), mid-day (10 am to 5 pm) and the evening (5 pm to 10 pm) based on hourly OD matrices that were averaged over all workdays. We investigate the evolution of the public transport mode share over the day and identify weak public transport connections.

5.1 Mode share of public transport

The mode share of public transport in Singapore increases with time of the day. It is 38 % in the morning, reaches 44 % around mid-day and peaks at 52 % in the evening. One reason for this observation could be people feeling tempted to go to work by taxi when being in a hurry in the morning and using public transport back home in the evening, people car-pooling in the morning and returning back home separately, or parents dropping their kids at school on the way to work.

5.2 Weak public transport connections

We define a connection to be weak by public transport when more people use private than public transport on this connection in a given period of time. Being particularly interested in the major connections, we present only the 50 busiest out of the 2970 inter-district connections in Figure 4. The plots show the number of passengers taking public as well as private transport. Almost all trips are directed towards the city center in the morning. In the evening, the situation is reverse with most trips departing from the city center. The highest mode share of private traffic (around 80 %) can be observed between Bukit Timah, a large residential area without access to a subway line, and the city center in the morning and in the evening. The Downtown Line, a new subway line in Singapore opening from 2013 to 2017, will cover this connection. Moreover, many weak public transport connections can be found around Bedok and Tampines, the most and third-most populated districts, which are located in the east of Singapore.

6. OUTLOOK

Traditional approaches to understanding public and private transportation flows through a city use travel surveys that are not only expensive, but also time consuming, inaccurate, and only sample a small percentage of a city's total population. Using a data mining approach, we present methods and analyses that show that mobility and connectivity in cities can be accurately described by cellphone call detail records quickly, efficiently, in real-time, and with district-level spatial resolution or better. Coupled with public transport smart card records, we show that we can further identify the spatio-temporal variability of public and private transport use and begin to examine how and why such patterns exist. However, the value and reliability of such an approach is not limited to only this use-case. In fact, the implications of harnessing these types of datasets, specifically those being created from pervasive urban sensor networks and smartphones, are immense for planning and designing more livable and sustainable cities.
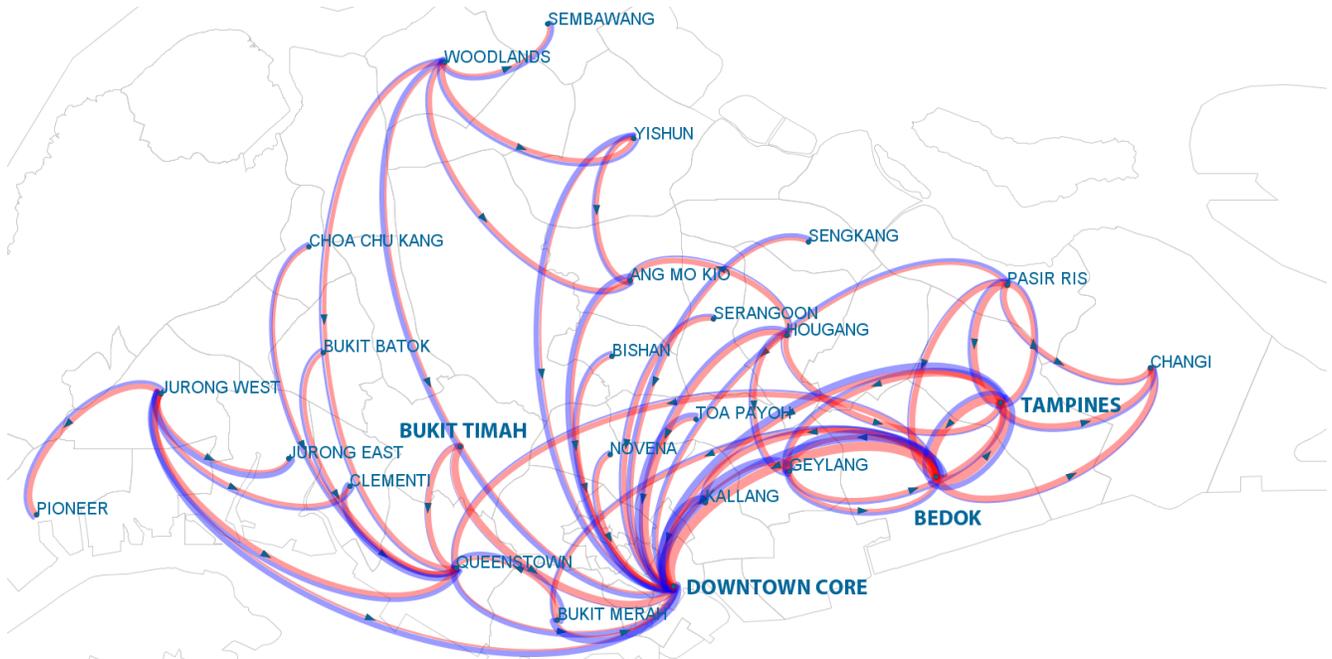
By analyzing these highly granular datasets, produced in real-time, there are emerging opportunities to address a wide range of environmental, epidemiological, and socio-cultural questions that result from urban living and vis-a-vis human mobility in cities. With respect to transportation, using cellphone based mobility patterns opens up new ways to plan for the ever growing population of aging citizens—a particular issue for countries in Europe and in Japan, the United States, Singapore, among others. In order to guarantee accessibility to goods and services for populations with limited mobility options, big, urban datasets such as cellphone call detail records can help inform short-, medium-, and long-term decisions to plan, locate, and design, for example, mixed-use settlements, walkable connections and corridors, and mobility on-demand services.

Lastly, a role for cellphone based urban mobility detection has significant implications for disease control in cities. As humans serve as the primary and secondary vectors of many infectious diseases, understanding from where people arrive and depart and by which transportation modes people are traveling, we have the potential to model how and where diseases might be spreading and from where they might originate. Big data analytics such as the analysis presented here, may introduce a distinctly new and interdisciplinary approach to modern epidemiological studies in both the developed and developing world.
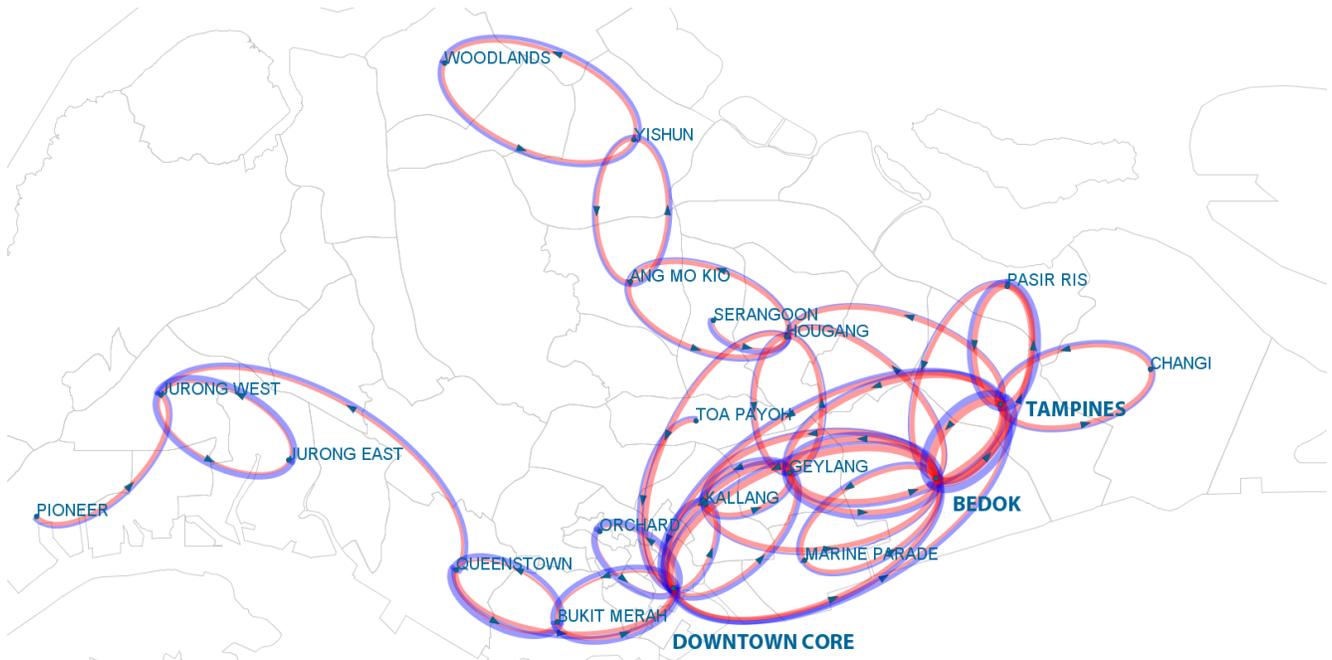
Examining the digital breadcrumbs left behind by people in cities introduces new methods for examining urban mobility. Here, we emphasize the promising role for cellphone network analysis to generate meaningful descriptions of city scale transportation use and to comment on future avenues of urban research and planning.

7. REFERENCES

[1] Singapore census of population 2010. Technical report, Singapore Department of Statistics, 2010.

[2] *JOURNEYS – Sharing Urban Transport Solutions*, volume 7. Land Transport Authority, Singapore, 2011.

[3] Statistics Singapore – Population (mid-year estimates). `http://www.singstat.gov.sg/stats/latestdata.html#12`, retrieved on 2012-11-29.

[4] Administrative divisions of Singapore. `http://en.wikipedia.org/wiki/Administrative_divisions_of_Singapore`, retrieved on 2012-12-03.

[5] Ez-link card. `http://www.ezlink.com.sg/index.php`, retrieved on 2013-04-04.

[6] Transport fares. `http://www.sbstransit.com.sg/transport/trpt_lrt_fares.aspx`, retrieved on 2013-04-04.

[7] I. Anderson and H. Muller. Practical context awareness for GSM cell phones. In *Proceedings of the 10th IEEE International Symposium on Wearable Computers (ISWC 2006)*, pages 127–128, 2006.

[8] I. Anderson and H. Muller. Exploring GSM data in pervasive environments. *International Journal of Pervasive Computing and Communications*, 4(1):8–25, 2008.

[9] N. Caceres, J. Wideberg, and F. Benitez. Deriving

(a) Morning (6 am to 10 am).

(b) Mid-day (10 am to 5 pm).

Figure 4: The 50 busiest inter-district connections, depending on the time of the day. The width of the lines represents the number of people traveling between the connected districts. Blue lines indicate the number of people taking public transport, and red lines the number of people taking private transport. The highest mode share of private traffic can be observed between Bukit Timah and the city center. Moreover, many weak public transport connections can be found around Bedok and Tampines.
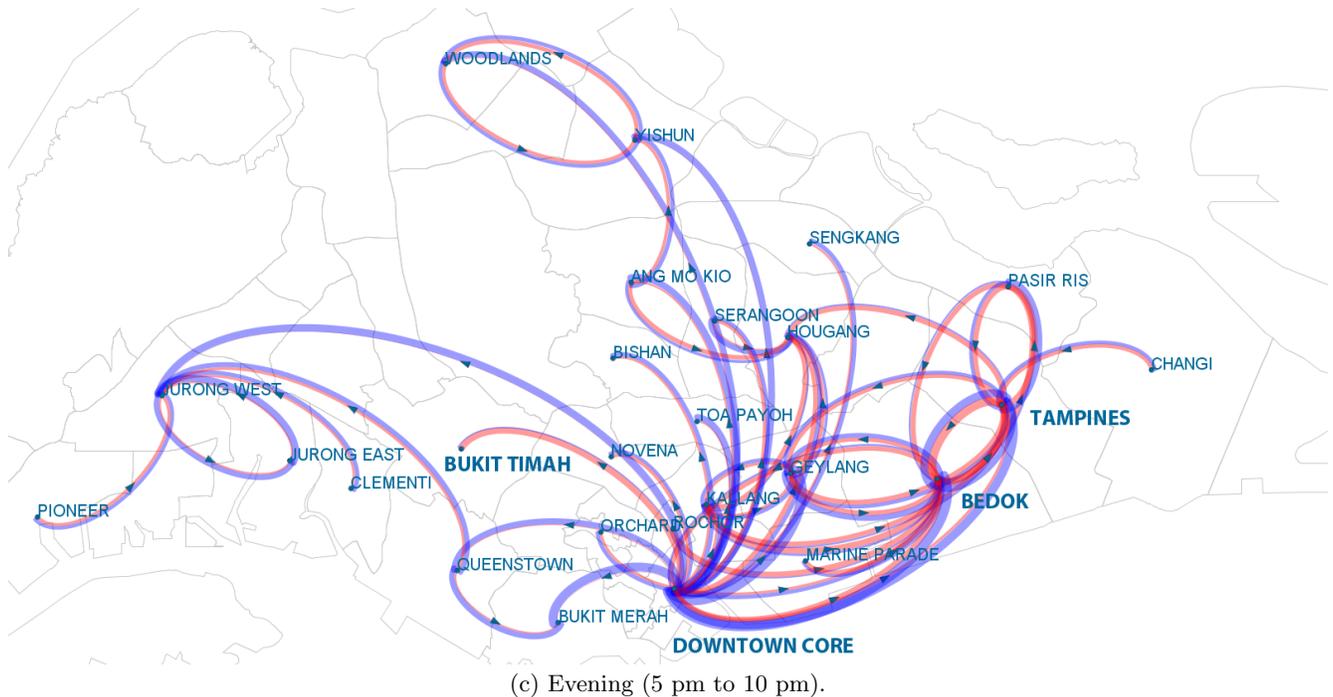
(c) Evening (5 pm to 10 pm).

Figure 4: The 50 busiest inter-district connections, depending on the time of the day. The width of the lines represents the number of people traveling between the connected districts. Blue lines indicate the number of people taking public transport, and red lines the number of people taking private transport. The highest mode share of private traffic can be observed between Bukit Timah and the city center. Moreover, many weak public transport connections can be found around Bedok and Tampines.

origin destination data from a mobile phone network. *Intelligent Transport Systems*, 1(1):15–26, 2007.

[10] F. Calabrese, G. Di Lorenzo, L. Liu, and C. Ratti. Estimating origin-destination flows using opportunistically collected mobile phone location data. *Pervasive Computing*, 10(4):36–44, 2011.

[11] C. Choi and R. Toh. Household interview surveys from 1998 to 2008 – a decade of changing travel behaviors. Technical report, Land Transport Authority, Singapore Government, 2008.

[12] B. Ferris, D. Hähnel, and D. Fox. Gaussian processes for signal strength-based location estimation. In *Proceedings of Robotics: Science and Systems II*, Cambridge, Massachusetts, USA, 2007. MIT Press.

[13] V. Frias-Martinez, C. Soguero, and E. Frias-Martinez. Estimation of urban commuting patterns using cellphone network data. In *Proceedings of the ACM SIGKDD International Workshop on Urban Computing (UrbComp 2012)*, pages 9–16. ACM, 2012.

[14] S. Isaacman, R. Becker, R. Cáceres, S. Kobourov, M. Martonosi, J. Rowland, and A. Varshavsky. Identifying important places in people's lives from cellular network data. In *Pervasive Computing*, pages 133–151. Springer, 2011.

[15] C. Kang, S. Sobolevsky, Y. Liu, and C. Ratti. Exploring human movements in Singapore: a comparative analysis based on mobile phone and taxicab usages. In *Proceedings of the 2nd ACM SIGKDD International Workshop on Urban*

Computing (UrbComp 2013), 2013.

[16] L. Liao, D. Patterson, D. Fox, and H. Kautz. Learning and inferring transportation routines. *Artificial Intelligence*, 171(5-6):311–331, 2007.

[17] C. Pan, J. Lu, S. Di, and B. Ran. Cellular-based data-extracting method for trip distribution. In *Transportation Research Record: Journal of the Transportation Research Board*, volume 1945, pages 33–39, 2006.

[18] T. Sohn, A. Varshavsky, A. LaMarca, M. Chen, T. Choudhury, I. Smith, S. Consolvo, J. Hightower, W. Griswold, and E. De Lara. Mobility detection using everyday GSM traces. In *Proceedings of the 8th International Conference on Ubiquitous Computing (Ubicomp 2006)*, pages 212–224, Orange County, CA, USA, 2006. Springer.

[19] L. Stenneth, O. Wolfson, P. Yu, and B. Xu. Transportation mode detection using mobile phones and GIS information. In *Proceedings of the 19th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems (GIS 2011)*, pages 54–63. ACM, 2011.

[20] J. White and I. Wells. Extracting origin destination information from mobile phone data. In *Proceedings of the 11th International Conference on Road Transport Information and Control*, pages 30–34, 2002.

[21] Y. Zheng, Y. Chen, Q. Li, X. Xie, and W. Ma. Understanding transportation modes based on GPS data for web applications. *ACM Transactions on the*

Web (TWEB), 4(1):1–36, 2010.