# Quantifying and Visualizing Jobs-Housing Balance with Big Data:

## A Case Study of Shanghai

**Abstract**

Existing jobs-housing balance studies have relied heavily if not solely on small data. Via a case study of Shanghai, this study shows how cellular network data can be processed to derive useful information, job and housing locations of commuters in particular, for those studies. Based on cellular network data, this article quantifies and visualizes Shanghai's jobs-housing balance with a much larger sample (n=6.3 million), finer spatial resolution and greater geographic coverage than before. It identifies and geocodes the local commuters by Base Transceiver Station (BTS), which has on average a service area of 0.16 square kilometers. After detecting jobs and housing by BTS, it aggregates them by subareas of particular interest (e.g., traffic analysis zones, inner city, suburbs and exurbs) to local planners and decision-makers. It also visualizes the traffic flows associated with the actual ($T_{act}$), theoretical minimum ($T_{min}$) and maximum ($T_{max}$) commutes. It shows that Shanghai's commuting pattern is far from the extremes (indicated by $T_{max}$ and $T_{min}$ traffic flows) and Shanghai's relative balance of jobs with respect to housing is decent (3.2 km) despite of its huge population (24 million) and land area sizes (6,800 square kilometers). The distance distribution of the $T_{min}$ and $T_{act}$ flows in Shanghai is similar when the distance is larger than 12.5 km, which means that if Shanghai hopes to optimize its commuting pattern, it should focus more on commuting trips that are shorter than 12.5 km.

**Key Words**

Cellular Network Data; Jobs-Housing Balance; Excess Commuting; Visualization

## INTRODUCTION

Car dependence, traffic congestion, long commute and associated air pollution and Greenhouse Gas (GHG) emissions are torturing many metropolises across the world. They are thus some of the most significant challenges faced daily by millions of people (Litman and Burwell 2006). To deal with these challenges, many planners, policy analysts and public agencies have proposed different countermeasures. Among them, the jobs-housing balance has been considered or even advocated as one of the most effective (California Planning Round Table 2008; Cervero 1991; Weitz 2003). Despite that, the jobs-housing balance, in academia, however, has not been understood and defined unanimously (e.g., Giuliano 1991; Ma and Banister 2006; Peng 1997). There have also been a variety of input data for one to characterize and quantify the "jobs-housing balance" that has been defined differently. Among the existing input data, the most typical and dominant include household travel surveys and interviews, which can be called "small data" as compared to the emerging "big data" such as cellular network data. As a whole, there have been relatively mature and systematic ways for us to process, validate and calibrate small data. Based on validated and calibrated small data, most authors/scholars implicitly believe that their derived information, conclusions and findings would be reliable and even transferable. The presence and availability of big data, in particular, cellular network data, has provided new opportunities for authors/scholars to quantify the "jobs-housing balance", regardless of its exact definition.

The big data, nevertheless, are often not purposefully designed for scholarly studies; rather, they are designed to serve particular business function(s), e.g., validating and collecting bus fares (Pelletier et al. 2011). How can one then derive useful information from big data for scholarly studies of the jobs-housing balance? How can the derived information complement small data for such studies? Would big data shed new/more lights on pressing urban issues such as the jobs-housing imbalance and long commutes? These are some interesting and important questions that scholars and decision-makers need to well address in the era of big data.

This article argues that cellular network data are a kind of big data that can be used to effectively facilitate the jobs-housing balance studies, making them transcend the constraints such as detection of latency and limited geographic/temporal coverage posed by small data (Pucci and Tagliolato 2015). Via a case study of Shanghai, it shows how cellular network data can be processed to derive useful information for the aforementioned studies. It characterizes, quantifies and visualizes the jobs-housing balance based on existing analytical frameworks ("excess commuting" in particular) in a metropolis, attempting to shed more lights on the issue than existing studies.

The article is organized as follows. The next section (Section 2) reviews relevant literature, which helps place this manuscript into a bigger picture of the jobs-housing balance studies, in particular, what big data are and how big data may improve and even revolutionize the jobs-housing balance studies. Section 3 provides a case study of Shanghai, demonstrating how cellular network data could be processed to facilitate and improve the jobs-housing balance studies. Section 4 concludes and discusses how cellular network data could further enhance and improve the jobs-housing balance studies in the future.

## RELATED LITERATURE

### Defining the Jobs-Housing Balance

In academia, there have been quite a few definitions of “jobs-housing balance”. It is “the spatial relation between the number of jobs and housing units within a given geographical area” (Peng 1997: p.1216).It can also be as a ratio of jobs and housing units at the level of spatial units such as census tract, Zipcode area or Traffic Analysis Zone (TAZ) (e.g.,Margolis 1973; Cervero 1989). If a spatial unit achieves certain ratio of jobs and housing units, it is in the “quantitative balance” and otherwise is in the “quantitative imbalance” (Ma and Banister 2006: p. 2104). The jobs-housing balance, per other authors; however, cannot simply be defined as a ratio of jobs to dwelling units. Differences in the household size, workforce participation rate and dwelling unit, for instance, can make the ratio biased and even problematic. True jobs-housing balance thus involves “perfectly complementary housing and job characteristics” (Giuliano 1991: p.305). When housing and job characteristics in a spatial unit do not complement one another, the qualitative jobs-housing imbalance arises (Ma and Banister 2006). More generally, the jobs-housing balance is a dynamic process of adjustments of jobs and/or housing in urbanization or suburbanization. In this process, commuting time can be a proxy for the jobs-housing balance (e.g., Dubin 1991; Gordon et al. 1989). A free market automatically generates some degree of the jobs-housing balance (i.e., co-location effects) so long as firms and resident workers can choose their locations at will. Inappropriate planning and policy interventions sometimes distort the market and thus contribute to the jobs-housing imbalance, lengthening the average commute (Cervero 1989; Cervero and Landis 1995).

In practices, the jobs-housing balance has been treated as a planning tool or a recommended policy target. Thus it has been defined with considerations such as data requirements, selection of indicators, application of the indicators and how the indicators would affect the attainment of goals. California Planning Roundtable (CPR) (2008), an organization of experienced planning professionals who are members of the American Planning Association (APA), has proposed that communities use the same input to define and measure the jobs-housing balance and there are three commonly-used quantitative measures to define the jobs-housing balance: jobs-households ratio, jobs-housing units ratio and jobs-employed residents ratio. When applying those measures, communities should account for the relationship among different types of jobs and housing characteristics, which can affect how good a specific measure is in terms of defining the jobs-housing balance for a particular community. What is more, communities should be aware of the fact that workers' price elasticity for commuting/housing costs, mode choice, gender and family concerns all could influence how they could achieve the jobs-housing balance. Similar to CPR, Weitz (2003), in a document published by APA, argues that the jobs-housing balance can be expressed as a ratio of jobs to housing. But when applying this ratio as a planning tool, planners should ensure that job and housing characteristics match each other. For practitioners, he contends that two jobs-housing ratios: jobs to housing units ratio and jobs to employed residents ratio can be used to pursue the same policy targets related to the jobs-housing balance. These targets are applicable when local data on the number of workers per household are well considered. To a geographical location of a region, there are four types of the jobs-housing imbalance, depending how well jobs complement housing units (Table 1):

Tab. 1: Types of Jobs-Housing Imbalance

| Type # | Jobs | Housing Units |
|---|---|---|
| 1 | Too many low-wage | Too few low-end |
| 2 | Too many high-wage | Too few high-end |
| 3 | Too few low-wage | Too much low-end |
| 4 | Too few high-wage | Too much high-end |

Source: Adapted from Weitz (2003).

Communities should formulate different plans and strategies to reduce different types of the jobs-housing imbalance that they encounter.

## Characterizing and Quantifying the Jobs-housing Balance

Other than the ratios mentioned above, scholars have developed more sophisticated analytical frameworks to characterize and quantify the jobs-housing balance and associated issues such as commuting, spatial mismatch and job/housing accessibility. At some risk of oversimplifying, these analytical frameworks can be categorized into three groups: excess commuting, gravity-based accessibility and commuting spectrum.

## Excess commuting

This framework uses indicators such as the theoretical minimum commute ($T_{min}$), theoretical maximum commute ($T_{max}$), random commute ($T_{ran}$), actual commute ($T_{act}$), excess commuting (EC), commuting potential used ($C_u$), commuting economy ($C_e$) and normalized commuting economy ($NC_e$) to quantify and connect the jobs-housing balance with commuting efficiency, which measures how efficient a commuting pattern of a city/region is. In this framework, it is assumed that:

- All workers, employment and housing opportunities are homogeneous and thus they can be enticed to any employment and/or housing opportunities without losing any utilities
- The travel cost or impedance between any two spatial units remains the same, e.g., the cost is always the linear distance between centroids of the two spatial units, regardless of how many trips there are.

$T_{min}$ in a region is achieved where workers travel to the closest possible workplace on average in terms of some measure of zonal separation (e.g. time, distance). $T_{min}$ indicates the relative balance of jobs with respect to housing in a region (Small and Song 1992). $T_{max}$ occurs in a region "when workers are assigned, on average, to their most distant workplaces" (Horner 2002: p.550). It reflects the worst commuting pattern for a given distribution of jobs and housing of a region.

$T_{act}$ can be directly calculated if the existing commuting pattern is known. For instance, most household travel surveys report on average how long a commuter travels for his/her journey to work. This figure based on such surveys can be regarded as $T_{act}$. EC is “the nonoptimal or surplus work travel occurring in cities because people do not minimize their journeys to work” (Horner 2002: p.543), that is,

$$EC= (1-T_{min}/T_{act})*100 \tag{1}$$

$C_u$ quantifies how much of the available commuting range, which is the difference between $T_{max}$ and $T_{min}$, has been consumed, that is,

$$C_u=(T_{act}-T_{min})/(T_{max}-T_{min})*100 \tag{2}$$

A region will have $T_{ran}$ for its commuters if all these commuters make no efforts to minimize their commutes and randomly choose their respective residences and workplaces. Thus, $T_{ran}$ should on average always have a value that is greater than $T_{min}$. Charron (2007) uses the following equation to get an approximate value of $T_{ran}$:

$$T_{ran} = \frac{1}{N^2}\sum_{i=1}^{m}\sum_{j=1}^{n} O_i D_j C_{ij} \qquad (3),$$

where

N is the total number of commuters in the study area;

$O_i$ is the number of origins where commuters start;

$D_j$ is the number of destinations where commuters end;

$C_{ij}$ is the cost of travel between i and j and the cost can be time, distance or monetary value.

More recently, Murphy and Killen (2011) have proposed a feasible but more sophisticated method than the above to calculate $T_{ran}$. In a nutshell, their method has three steps.

Step 1 is to simulate as many as possible commuting trip distributions given the fixed/known numbers/distribution of jobs and housing in a city or a region, where jobs and housing are aggregated by some spatial divisions such as TAZ.

Step 2 is to calculate the respective total commutes of the simulated distributions.

Step 3 is to get the average of a very large number (say n=10,000) of total commutes resulting from Steps 1 and 2. This average is then treated as an approximate $T_{ran}$.

With $T_{ran}$ and $T_{max}$, $C_e$ can be derived and it “demonstrates the extent to which actual behaviour is reacting to the cost of consuming the separation that exists between residences and workplaces in the urban region” (Murphy and Killen 2011: p. 1261).

Specifically, $C_e$ is calculated using the following equation:

$$C_e=(1-T_{act}/T_{ran})*100 \tag{4}$$

With $T_{ran}$ and $T_{min}$, $NC_e$ is a better alternative to $C_e$ and allows one to “determine the extent to which collective behaviour is tending towards commuting economy while talking account of the theoretical extent to which it is possible within the constraints set by land use geography (Murphy and Killen 2011, p. 1261). Specifically, $NC_e$ can be expressed in the following equation:

$$NC_e=(T_{ran}-T_{act})/(T_{ran}-T_{min})*100 \quad (5).$$

Given that a city or region have always to be divided into smaller units of analysis before one can estimate the above indicators such as $T_{min}$, $T_{max}$ and $C_u$, how would different units of analysis would affect the values of these indicators and their stability? That is, would those indicators be subject to the modifiable areal unit problem (MAUP)? Several authors have looked into this issue (e.g., Horner and Murray 2002; Niedzielski et al. 2013). The overall finding is that metrics such as $T_{min}$, $T_{max}$, EC and $C_u$ tend to suffer more from the issue while metrics such as $C_e$ and $NC_e$ are largely immune from the issue.

Assuming that the MAUP issue is now addressed, there still is an issue of when to use which indicator(s)? Kanaroglou et al. (2015) review a large amount of literature applying or quantifying the above indicators, concluding that none of these indicators can adequately measure the commuting performance of a city but each indicator can still be used to address a specific policy question. When they are combined, they can provide "a reasonably good understanding of urban form and commuting behaviors" (p.13).

## Gravity-based accessibility

Advocators of gravity-based accessibility argue that the jobs-housing balance should consider jobs or employment opportunities not only within a predefined area but also around it. Levinson (1998), one of such advocators, develops an accessibility measure for the jobs-housing balance to account for jobs or housing units in and around a subarea/zone according to some spatial distance decay functions. He contends that this measure is more powerful than zone-based jobs-housing ratios in terms of explaining the variations in commuting. His case studies show that the accessibility to jobs and housing has a negative relationship with the commuting distance, and that transit commuters appear to have higher accessibility to jobs and housing than their automobile counterparts. In a similar vein, Horn and Mefford (2007) use the minimum and maximum commutes and the ranges between the minimum and maximum commutes by different social groups to show how the spatial mismatch and the jobs-housing balance vary across different social groups. More specifically, if we assume that different social groups could only swap jobs and housing within groups, they could have different degrees of accessibility to jobs and housing and of the jobs-housing balance and imbalance. There could be cases that there are jobs around/near some residences or residences around/near some workplaces but some workers are simply excluded from those employment or housing opportunities because of implicit discriminations in local job and housing markets. In other words, the proximity to jobs or housing sometimes does not necessarily means accessibility to, and availability of them among all workers. The jobs-housing balance thus should well account for job and/or housing accessibility and availability across worker groups.

Commuting spectrum

The commuting-spectrum scholars view the existing commuting pattern (which generates $T_{act}$) in a city/region as one of many possible commuting patterns, that is, commuting trip distributions given that jobs and workplaces are fixed by some spatial disaggregation (e.g., TAZ), in a city/region (Yang and Ferreira 2008). If we assume now travel cost is the only factor that influences commuters' job and housing decisions, then a gravity model can be calibrated to derive the value of $T_{min}$, that is, the relative balance of jobs with respect to housing in a city/region, as well. That is, the relative jobs-housing balance for a given distribution of jobs and housing can be derived based on a gravity model. When workers from a unit of analysis i are allocated to other units of analysis (j's) per j's share of the entire region's workers, proportionally matched commuting (PMC) is generated. PMC means a scenario where workers are insensitive to travel cost, that is, "every worker in the region competes for every job in that region, regardless of the commuting cost" (Yang and Ferreira 2008: p. 367). Like $T_{min}$, PMC represents another extreme commuting pattern. While the former is determined mainly by the local-level jobs-housing distribution the latter is more dependent on the regional-level one. Based on a case study of Boston, Yang and Ferreira (2008) find that the average of PMC at the census tract level better explains the spatial variation of commuting. In other words, the regional jobs-housing balance or the jobs-housing balance within a region's commuting shed should carry more weight if a city or region hope to reduce the average commuting cost by adopting jobs-housing balance policies.

**Jobs-housing Balance Studies with Small Data**

Regardless of their respective analytical frameworks, most existing studies of the jobs-housing balance, including the above-mentioned ones, have relied heavily if not solely on small data as input. Table 2 provides a snapshot for some representatives of existing studies.

Tab. 2: Jobs-Housing Balance Studies with Small Data

| Study | Analytical Framework(s) (Or Indicators) | Sample Size | Data Type/Source |
|---|---|---|---|
| Giuliano (1991) | Resident workers/jobs ratio | Not mentioned (NM) | Census/Employment data of the government for two years |
| Wachs, et al. (1993) | Mode choice and commuting distance | 1,500 | Ad-hoc surveys over 6 years |
| Peng (1997) | Jobs-housing ratio and vehicle miles traveled | NM | Travel model data |
| Sultana (2002) | Jobs-housing ratio and commuting time | NM | Census data (Census Transportation Planning Products [CTPP]) |
| Morrison and Monk (2006) | Numbers of jobs and housing units | NM | Employers' surveys; Housing surveys |

| | | | |
|---|---|---|---|
| Horner and Mefford (2007) | Range between the minimum and maximum commute | NM | CTPP |
| Yang and Ferreira (2008) | Commuting spectrum | NM | CTPP |
| Liu et al. (2008) | Excess commuting | 1,500 | Household interviews |
| Wang and Chai (2009) | Average commuting time, physical relation of job and housing locations | 736 | Household interviews |
| Horner (2010) | Excess commuting | NM | CTPP |
| Loo and Chow (2011) | Excess commuting; Commuting time | NM | Census data |
| Suzuki and Lee (2012) | Excess commuting; Spatial correlation of jobs and housing (Vaughan's model) | NM | Census data |
| Zhou, et al. (2013) | Excess commuting | 59,967 | Household interviews |
| Zhou and Long (2014) | Excess commuting | 216,844 | Smart-card data; travel survey data |

Based on a careful scan of the literature listed in Table 2, one can notice that most if not all the selected existing studies implicitly assume that their input data have been validated and thus can be directly fed into related studies. Many of these existing studies even do not mention how big their respective sample sizes are and how the samples are selected. Of the existing studies reviewed, the biggest sample size is 216,844. But this may still be small if one takes into account the fact that the study area is Beijing, which cover a land area of over 16,410 square kilometers, contains 1,119 TAZs and has over 20 million residents. It is also unclear that how random or representative the samples are as compared to the whole population and how well the samples can cover all the TAZs.

If one assumes, of course, that the samples in existing studies are all randomly drawn and well represent the population, then there is little to worry about per classic statistical and sampling theories. But there remain some interesting questions in the era of big data, for instance, would input data of a much bigger sample size and of even the whole population, in particular, big data such as cellular network data challenge the existing knowledge and findings about the jobs-housing balance, which are based primarily on small data? Would big data generate new ways and visuals to study the jobs-housing balance, resulting in new knowledge about it? These are what this article hopes to address via a case study of Shanghai. In the context of Shanghai, household travel survey data were the primary source of data for the jobs-housing and commuting studies prior to the emergence of big data. On the one hand, the former (small data) can only cover 0.75% of the population; on the other hand, they only record travel behaviors of the respondent on a weekday (Ding et al. 2015). These characteristics mean that scholars have to find reliable ways to extrapolate the samples so long as to get a fuller and longer (multi-day) picture of the population. Based on survey data of selected subareas, Feng et al. (2011) and Sun et al. (2013) have examined impacts of polycentrism on Shanghai's commuting efficiency. They argue that multiple employment centers can improve the road traffic condition and shorten the average commuting time in Shanghai. Using samples from large planned communities in Shanghai, Chen et al. (2014) quantity the commuting time and mode choice of resident-workers therein. They find that the resident-workers therein have an earlier departure time for their daily commute and higher dependence on public transit and scooters.

**Jobs-housing Balance Studies with Big Data**

Big data "is data that exceeds the processing capacity of conventional database systems. The data is too big, moves too fast, or doesn't fit the strictures of your database architectures. To gain value from this data, you must choose an alternative way to process it" (Dumbill 2012: p.3). As compared to small data, big data have seven features:

- Huge in volume--big data consist of a much larger size of data than small data, usually in magnitudes of terabytes or petabytes;
- High in velocity--unlike the small data, big data can be generated in or near real-time;
- Diverse in variety--big data can be either structured or unstructured in nature and can contain both temporal and spatial information;
- Exclusive in scope--big data can capture even the whole population or at least in a sample size that is much larger than small data;
- Fine-grained in resolution--big data can hold much more details about the subjects that scholars or administrators want to have than small data;
- Relational--big data can have much more common fields that a large number of diverse datasets can be joined together;
- Flexible and scalable--big data can add new fields and expand in size efficiently (Boyd and Crawford 2012; Dodge and Kitch 2005; Mart and Warren 2012; Kitch 2013; Mayer-Schonberger and Cukier 2013; McKinsey Global Institute 2015).

Based on two popular academic databases Web of Science and Web of Social Sciences, there have been few specific jobs-housing balance studies with big data, cellular network data in particular, as input. But commuting and dense locations of cell phone users (e.g., their homes and workplaces) seem to be a topic of interest to many authors if we expanded the search using other tools such as Google Scholars. Ahas et al. (2007, 2010a, b) were some of the pioneers in this topic. Using cellular network data from Estonia, they characterize the daily rhythms of a subgroup of commuters' movement and identify meaningful locations of mobile phone users. Similarly, Vieira et al. (2010) have used mobile phone-call data to detect dense urban areas. Utilizing cell-phone call data across four countries, Kung et al. (2014) study home-work commuting patterns' regularity in terms of home-work time distribution. They find that when all modes of travel are considered, people across countries on average tend to spend a similar amount of time on commuting. Chen (2014), based on the US data, is able to determine 90 percent of homes and workplaces within a certain area. But she still argues that there remain challenges for the usage of cellular data in transport, in particular, what are the market penetration rates of different mobile phone companies and what is the actual sample size of the cellular network data (e.g., some users can have two SIM cards or two cell phones). These challenges engender uncertainties to researchers when they try to extrapolate the cellular network data to the whole population.

In the case of Shanghai, Ding et al. (2015) have used two-week-worth cellular network data of two years to estimate the commuting shed of the inner city, which had long been an undetermined issue among local planners and decision-makers. Using the same data as Ding et al. (2015), Niu and Ding (2015) further examine commuting patterns of different subareas in Shanghai: the subarea within the inner ring road ("inner city") and seven new satellite cities outside the external ring road. They find that 97 percent of the inner-city workers have their residence within the commuting shed that Ding et al. (2015) identify and only 5 percent of the workers of the satellite cities have their workplace in the inner city. Zhang (2016) develop methodologies to use cellular network data as input to derive homes and workplaces of cell phone users in Shanghai. Based on the derived information, they quantify the commuting distance of workers between TAZ and compare them with those based on local household travel surveys. Their comparison indicates that cellular network data can be used to derive jobs-housing locations and separations at least as accurately as household travel survey data.

Given the above examples and features of big data, not only specific studies of the jobs-housing balance but also several related academic fields such as Urban Planning, Geography and Transportation would have to adapt and change (e.g., see McKinsey Global Institute 2015; Schweitzer 2014; Batty 2012, 2013). This also necessitates this article, which uses a case study to show how big data, cellular network data in particular, can facilitate and improve the jobs-housing balance studies. Compared to small data, big data can have the following advantages when they are used to study the jobs-housing balance and commuting issues:

- They provide a much larger sample of the population;
- They provide continuous and timely information about commuting, jobs and housing at finer spatial and temporal resolutions;
- They are automatically generated and are more cost-effective;
- They are less likely to subject to respondents' reporting errors or hoarding of information as respondents (samples) do not have to answer any survey questions and their information is passively collected (c.f., Pucci and Tagliolato 2015; Ding et al. 2015).

**Gaps in Existing Studies**

In light of the literature reviewed above, the following gaps can be identified regarding the characteristics of, and gaps in current research on the jobs-housing balance:

First, there have been many studies that have tried to define the “jobs-housing balance” but there has not been one universally accepted definition of it;

Second, regardless of how the jobs-housing balance is defined, few have considered the differences between related studies based on small data and big data and how the introduction of the latter can affect existing studies of the jobs-housing balance: their input data, methodologies, findings, conclusions and visualizations.

Third, little has been done on using cellular network data to characterize, quantify and visualize the jobs-housing balance and related commuting pattern at the metropolis level.

Fourth, the excess commuting framework has been utilized to study the jobs-housing balance issue but in the past the input data for related studies are mostly if not solely based on small data.

In light of the above gaps, this manuscript will conduct a case study of Shanghai, trying to show how cellular network data can be used to improve and enhance the existing studies.

## A CASE STUDY

### The Site

Shanghai was chosen as the site for the case study. Shanghai is the most populous metropolis in China. It has over 24 million registered residents and covers a land area of

6,800 square kilometers. As of 2015, millions of Shanghai residents have at least one active mobile phone. To effectively serve, manage and charge millions of users, three mobile-phone companies collect and process cellular network data constantly. The cellular network data contain records such as anonymous and unique ID for each user, time and duration the mobile phone was in the local service area, which Base Transceiver Station (BTS) the mobile phone had been connected to, when, how long and/or whether the mobile phone has sent or received information (voice, message and data). As each BTS has a service area, typically a triangle, which on average is about 0.16 square kilometers in size, one can usually detect the location of each mobile phone (that is, a mobile-phone user) by that scope so long as the phone is not continuously shut off, does not malfunction and has communicated with at least one BTS. This study utilizes these detected locations to derive mobile-phone users' home and workplace by BTS, which can then be aggregated by other larger spatial units such as TAZ. More technical details about the processes are given below.

**The Jobs-Housing Balance Definition and Indicators**

In this case study, we adapted two existing jobs-housing balance definitions and use corresponding indicators to quantify and visualize jobs-housing balance and commuting efficiency in Shanghai. The first definition considers jobs-housing balance as a ratio of "commuter residents" and "commuter workers" by subareas that have been predefined and used by local planners and decision-makers. If a commuter lives in subarea A, then s/he is categorized as commuter resident of A, regardless of where is her/his workplace's location. Similarly, a commuter worker of A is a commuter whose workplace is in A, regardless of where is the location of her/his residence. In Shanghai, the three commonly used subareas are inner city, suburb and exurb and their boundaries have been clearly defined by local planners and decision-makers. Thus, once we have identified “commuter residents” and “commuter workers” by these three subareas, we can easily map out the corresponding jobs-housing distribution and calculate different commuting distances by subarea and use them to inform local planners and decision-makers.

The second definition was based on the excess commuting framework, which uses $T_{min}$ and corresponding commuting pattern to represent the relative balance of jobs with respect to housing at the city level. It also uses extra indicators such as $T_{max}$, EC, $C_e$ $C_u$, and $NC_e$ to show how efficient actual commuting pattern is or how imbalanced the actual jobs-housing is as compared to that producing $T_{min}$ for the city. Based on the above, we can the further visualize commuting patterns at the city level when $T_{min}$, $T_{act}$ and $T_{max}$ are in presence, respectively. These visuals can more or less inform local planners how actual commuting flows look like and how much more efficient or inefficient they can possibly be. According to our knowledge, the above indicators or visuals have not been presented in any existing studies of Shanghai’s jobs-housing balance.

**Deriving and Verifying Job and Housing Locations**

All the local cellular network data (about 1.5 billion records every day) in Shanghai were used as the initial input to derive the job and housing locations of people who

(a) had at least one active mobile phone;

(b) who stayed in Shanghai at least 60% of the time between January 2013 and June 2013, that is, their mobile phone has been detected 60% or more the time during the study period.

To ensure high accuracy of the derived job and housing locations, in addition to the above filtering criteria, only the users/data that meet the following criteria were further processed and used to derive job and housing locations:

- The users who had their mobile phone frequently detected in a BTS service area between 8 pm of a day and 8 am of the next day (for housing locations) and between 9 am and 6 pm of a day (for job locations)----at least four times per week.
- Assume a user's mobile phone had been detected n times during the above two periods, a derived home or job location would have to be in the BTS service area that had been detected at least n*60% times.
- For all users whose derived home and job locations from cellular network data were within 400 meters, they were either treated as telecommuters, unemployed, on vacation or retirees.

Implementing and/or applying the above processes/criteria generated 12.7 million housing locations and 6.3 million job locations of the local workers by BTS. If one recalls Table 2, the numbers of generated locations of housing and jobs are much larger than any existing studies listed therein. Given that Shanghai has a population of 24 million, one can say that more than half of the local residents' housing locations and most workers' workplaces and residences have been detected using cellular network data as input. The above fact indicates that big data can capture a much larger sample than small data, which usually draw five percent or even a smaller portion of the population. In the case of Shanghai, only 0.75% of the residents are selected into the local household travel survey (Ding et al. 2015).

Given that BTS' service area is not a typical spatial unit for local planners and decision-makers, the above derived information about jobs and housing would still need to be aggregated by larger spatial units such as TAZ. In the case of Shanghai, there are far more BTS service areas (n=420,000) than TAZ (n=4,518) and thus aggregating BTS-level data to TAZ-level data is generally straightforward. Most BTS service areas are fully inside a TAZ and therefore we can conveniently relate them together. For those BTS service areas intersect with two or more TAZs, we assume that the detected mobile phone users' residences or workplaces are evenly distributed and therefore they can be allocated into related TAZs in light of the portion of a BTS service area that falls into different TAZs:

$$b_i = \frac{Z_i}{\sum_{i=1}^{n} Z_i} U_j \qquad (6),$$

where

$b_i$ the estimated number of residences or workplaces that is in TAZ i;

$Z_i$ is the subarea of a BTS service area falls into TAZ i;

n is the total number of TAZs that has a portion of BTS service area j;

$U_j$ is the estimated number of residences or workplaces for BTS service area j.

How reliable are the derived housing and workplace locations? To address this, we compared them with the local household travel survey data (n=15,000) collected in the same year (2013). The latter only cover four new towns in Shanghai: Jiading, Qingpu, Songjiang and Jinshan and so we aggregated our BTS-level data into these towns, following similar procedures and methods described above regarding how we assembly BTS-level data into TAZ-level ones. With the housing and workplace locations from the two sources for the same spatial unit, we compared them in two dimensions: workplace distribution and commuting-distance distribution. Table 3 presents the workplace distribution of the four towns by the two sources.

Tab. 3: Workplace Distribution Based on Two Sources

| New town | Jiangding | | | Qingpu | | | Songjiang | | | Jinshan | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Data source/Location | a* | b | c | a | b | c | a | b | c | a | b | c |
| Cellular network (%) | 70 | 6 | 24 | 78 | 3 | 19 | 86 | 4 | 10 | 69 | 2 | 29 |
| Survey (%) | 75 | 4 | 21 | 83 | 2 | 15 | 82 | 5 | 13 | 79 | 2 | 19 |

*a=inside the new town; b=inner city; c=other.

If we treat all the percentage based on cellular network data as one population and all the percentage based on the survey on the other, their correlation coefficient is 0.99. This indicates that both sources would generate very similar workplace distribution for us.

Figure 1 shows the commuting-distance distribution of all the four towns by the two sources.

Fig.1: Commuting Distance Distribution Based on Two Sources

Figure 1 indicates that the two sources' distributions notably diverge when the commuting distance is less than one km but largely converge when the commuting distance is more than one km. The divergence can be a result of the assumption we made about those mobile users whose detected "workplaces" and "residences" are consistently 400 meters or shorter. As a whole, however, if we treat all the percentage based on cellular network data as one population and all the percentage based on the survey on the other, their correlation coefficient is 0.95. This indicates that both sources would generate very similar commuting-distance distribution for us.

In light of the above comparisons and correlation coefficients, we conclude that cellular network data can be used to detect housing and workplace locations of local workers quite accurately as compared to the household travel survey, should we assume that the latter is the most reliable and accurate way to obtain the locations.

**Distribution of Jobs and Housing**

With the derived and somewhat verified job and housing locations mentioned above, one can map out their distribution with a much larger sample and in a finer spatial resolution than ever before. In other words, at least two features of the big data mentioned above have been "materialized" in this case of Shanghai. Panels (a) to (e) of Figure 2 show the distribution of commuters' jobs and housing by BTS service area in Shanghai. One thing should be emphasized again is that those jobs and housing are in a magnitude of million and should well represent 50% of their respective population.

To be consistent with local planners’ conventions, we divided Shanghai into three large subareas: the inner city, the suburb and the exurb. The inner city is all the area within the inner ring roads of Shanghai. The suburb is all the area outside the inner city but are within some irregular boundaries, which is a buffer area of the external ring roads of Shanghai (See “suburb boundaries” in Figure 2) that are 20 to 35 kilometers from the 5the subareas defined, we further categorized commuters into six groups. Table 4 highlights the characteristics of these groups.

Tab. 4: Commuters by Subarea

| **Group** | **Characteristics** |
|---|---|
| Inner-city workers | Commuters whose workplace is in the inner city |
| Inner-city residents | Commuters whose residence is within the inner city |
| Suburb workers | Commuters whose workplace is outside the inner city but within the suburb |
| Suburb residents | Commuters whose residence is outside the inner city but within the suburb |
| Exurb workers | Commuters whose workplace is outside the suburb |
| Exurb residents | Commuters whose residence is outside the suburb |

Panel (a) of Figure 2 indicates that most commuters’ residences and workplaces cluster in and around the inner city. Overall, the spatial correlation of workplaces and residences is strong across Shanghai. Not surprisingly, the job and residential density in the inner city is among the highest in the city. Some locales in suburbs, in particular, some spots in the east and northeast suburbs, also have some of the highest concentration of residences and workplaces in the city. Exurbs have gained some concentration of residences and workplaces but this is not evenly distributed across the space. Panels (b) to (d) of Figure

2 show the distribution of commuters' residences and workplaces by different groups defined in Table 3. Based on these panels, we can see that the inner-city residents tend to have the best jobs-housing balance, i.e., most of them are able to work in or around the inner city. Most suburb and exurb workers cannot afford a residence or are not willing to live in the inner city. More suburb or exurb residents have their workplace outside suburbs or exurbs. In other words, the inner city has more workplaces and many suburb and exurb residents have to commute to the inner city to be employed.

(a) Overall Distribution

(b) Inner-City Workers and Residents

(c) Suburb Workers and Residents

(d) Exurb Workers and Residents

Fig.2: Distribution of Commuters' Workplaces and Residences in Shanghai

Table 5 quantifies the distribution of commuters by the groups defined in Table 4 in a commuting-flow matrix format.

Tab.5: Commuting Flows by Subarea in Shanghai

| **Resident Groups** | **Workplace Location** | | | |
|---|---|---|---|---|
| | Inner City | Suburbs | Exurbs | Total |
| Inner-city | 71% | 22% | 7% | 100% |
| | 784,573 | 243,107 | 77,352 | 1,105,033 |
| Suburb | 32% | 57% | 11% | 100% |

| | | | | |
|---|---|---|---|---|
| | 759,679 | 1,353,178 | 261,140 | 2,373,997 |
| Exurb | 12% | 21% | 67% | 100% |
| | 154,824 | 270,941 | 864,431 | 1,290,196 |
| **Worker group** | **Residential Location** | | | |
| | Inner City | Suburbs | Exurbs | Total |
| Inner-city | 45% | 44% | 11% | 100% |
| | 783,568 | 766,155 | 191,539 | 1,741,262 |
| Suburb | 13% | 70% | 17% | 100% |
| | 247,491 | 1,332,643 | 323,642 | 1,903,775 |
| Exurb | 4% | 17% | 80% | 100% |
| | 42,286 | 179,717 | 845,728 | 1,057,160 |

Based on Figure 2 and Table 5, the inner-city residents have the best jobs-housing balance. Less than 30% of these workers need to commute outside the inner city. Comparatively speaking, suburb residents are most likely to commute outside their communities, i.e., suburbs, to work. So if we consider the jobs-housing balance by subarea, suburb residents have the worst jobs-housing imbalance. Forty-three percent of them would have to work outside the suburb. But as a whole, most of residents in Shanghai are able to work within their respective subareas. At least 57% of them are able to find a job within their respective subareas.

From the perspectives of workers by subarea, most exurb workers choose a residence in exurbs. Only one out of five such workers choose to live in suburbs or the inner city. Most inner-city workers cannot or are unwilling to live in the inner city----55% of them reside outside this subarea. The suburb workers are fortune in this regards----70% of them are able to live in suburbs.

If we turn to the other two popular indicators of jobs-housing balance: commuting distance and jobs-housing ratio, the three subareas also have different patterns (See Table 6). The inner-city residents enjoy the shortest commuting distance (6.77 km) and the highest jobs-housing ratio (1.58). The inner-city workers suffer from the longest commuting distance (8.63 km). For suburb and exurb workers and residents, they have similar average commuting distances and comparable jobs-housing ratios.

Tab.6: Commuting Distances and Jobs-housing Ratios by Subarea

| **Subareas** | **Average Commuting Distance (km)*** | | **Jobs-housing Ratio** |
|---|---|---|---|
| | Residents (Origin-based) | Workers (Destination-based) | |
| Inner city | 6.8 (n=1,105,033) | 8.6 (n=1,741,262) | 1.58 |
| Suburb | 9.1 (n=2,373,997) | 7.9 (n=1,903,775) | 0.80 |
| Exurb | 9.0 (n=1,290,196) | 7.7 (n=1,057,160) | 0.82 |

*It is assumed that commuters travel along the straight line between two centroids of two BTS service areas.

**Jobs-Housing Balance in the Excess-Commuting Framework**

Adopting the existing excess-commuting framework mentioned above, several more indicators other than the commuting distance and jobs-housing ratio were calculated, using the derived numbers of jobs and housing by the local TAZs. More technical details about how to calculate those indicators and how to deal with changed analysis zone boundaries can be found in (Horner 2002; Murphy and Killen 2011; Zhou et al., 2014a).

Table 7 presents the values of those indicators in Shanghai we obtained and their counterparts in several other metropolises based on existing studies.

Tab.7: Excess-Commuting Indicators across Cities

| Indicator | Unit | Shanghai | Beijing | Guang -zhou* | Los Angeles | Tokyo** | Dublin |
|---|---|---|---|---|---|---|---|
| Year | | 2013 | 2008/2010 | 2005 | 1991 | 2000 | 2001 |
| Sources | | This study | Zhou and Long (2014) | Liu et al.(2008) | Kim (2005) | Lee et al. (2006) | Murphy & Killen (2011) |
| $T_{min}$ | km | 3.2 | 2.5(bus) 3.5(car) | 2.7 | 16.5 | 6.7 | 2.7 |
| $T_{max}$ | | 49.4 | 24.7(bus) 35.6(car) | - | - | 50.5 | 21.7 |
| $T_{ran}$ | | 16.6 | - | - | - | - | 11.0 |
| $T_{act}$ | | 8.2 | 8.2(bus) 11.2(car) | 5.0 | 24.6 | 11.0 | 9.9 |
| EC | % | 61.6 | 69.5(bus) 68.8(car) | 44 | 33.0 | 39 | 73 |
| $C_u$ | | 11.0 | 25.7(bus) 24.3(car) | - | - | 10 | 38 |
| $C_e$ | | 50.1 | - | - | - | - | 34.5 |
| $NC_e$ | | 61.9 | - | - | - | - | 43.2 |

*Unit of analysis is zip code area.
**Unit of analysis is shikuuchoson.

Niedzielski et al. (2013) show that $T_{ran}$, $T_{max}$, $C_u$ and $C_e$ are more likely to be scale independent, that is, their values are relatively stable regardless of the sizes of unit of analysis; thus, when making comparisons between Shanghai and other metropolises, this article focuses on the former rather than the indicators that are scale dependent when the units of analysis are different. The comparisons between Shanghai and the other metropolises indicate that:

First, Shanghai, Beijing and Dublin have comparable $T_{min}$. This means that despite that the three cities vary in their urban form, land use, population size, etc., the spatial correlation and separation of jobs and housing therein are somehow similar. In the ideal scenario that all jobs and housing are homogeneous and every worker can be enticed to any job or housing and s/he minimizes her/his commute, that is, when the relative balance of jobs with respect to housing is achieved, the three cities' commuters would have comparable average commuting distance. Or in other words, if commute costs are the only utility that we care about, the initial Pareto optimality in the three cities, measured by $T_{min}$, is comparable (Zhou and Long 2015).

Second, in terms of $T_{max}$, which measures the worst imbalance of jobs with respect to housing, Shanghai and Tokyo have almost identical values. This means that the scale and degree of jobs-housing separation of the two cities, in the worst scenario, are comparable.

Third, for $T_{ran}$, Shanghai has a value that is almost 50% more than that of Dublin. This may simply result from the facts that jobs and housing in Shanghai are distributed across a larger piece of land and if commuters therein no longer care about the travel costs, they would on average have a longer commuting distance that is larger than their counterparts in smaller cities such as Dublin.

Fourth, Shanghai's actual jobs-housing balance, if measured by average commuting distance, is decent despite it is the most populous city in China. For all the four studies/cities (Beijing, Los Angeles, Shanghai and Dublin) use TAZ as the unit of analysis, Shanghai's average commuting distance is the lowest. Based on Figure 2, this could be due to the fact that workplaces and residences in Shanghai have a strong spatial correlation.

Fifth, Shanghai's jobs-housing imbalance, if measured by EC, is better than that of Beijing. This means that Shanghai's commuters as a whole are able to minimize their commutes to a larger degree than their counterparts in Beijing.

Sixth, with respect to $C_u$, which measures the degree to which the commuting range afforded by the existing distribution of jobs and housing has been consumed, Shanghai also performs better than Beijing.

Last but not least, based on $C_e$ and $NC_e$, which measure how collective behaviors of commuters depart from random behaviors constrained by the existing distribution of jobs and housing, Shanghai's commuters tend to depart more from random behaviors, as compared to Dublin's commuters. That is, commuting behaviors in Shanghai are probably not as random as those in Dublin. This may be due to two facts. First, compared to Dublin, Shanghai has a very high concentration of employment, that is, a dominant employment center within the inner city. This concentration has greatly shaped or constrained the local commuting behaviors. Based on our derived information, Shanghai's inner city has about 1.7 million jobs, which accounts for nearly a third of all jobs in Shanghai. But the inner city has only 1.1 million residences and the average price of these residences is the highest in Shanghai. This forces the inner-city workers to find other residences outside the inner city. Second, as shown in Figure 2, Shanghai's workplaces and residences have a strong spatial correlation and this enables workers therein to enjoy some co-location effects.

**Visualizations of Jobs-Housing Balance**

The above quantitative indicators have already shed some new lights on the jobs-housing balance in Shanghai. Taking advantage of the very large sample size, one can further map out the different commuting flows when different indicators such as $T_{min}$, $T_{act}$ and $T_{max}$ are achieved, assuming that all commuters take the shortest path regardless of the commuting costs. Similar figures have been drawn by Zhou et al. (2014b) in the case of Beijing. Thus some comparisons can be made between Shanghai and Beijing as well.

Different panels of Figure 3 visualize the Shanghai's commuting flows associated with $T_{min}$, $T_{act}$ and $T_{max}$, respectively.

(a) $T_{min}$

(b) $T_{act}$

(c) $T_{max}$

Fig.3: $T_{min}$, $T_{act}$ and $T_{max}$ Commuting Flows in Shanghai

When one only looks at the flows of Shanghai, one can see that $T_{max}$ would require most workers to commute along the corridors originating from the inner city ($T_{max}$-Panel of Figure 3). On average, these corridors could have a volume in the magnitude of at least 125,000. $T_{min}$, on the contrary, evenly distribute the commuters across different roads in the city ($T_{min}$-Panel of Figure 3). $T_{act}$ generates a commuting pattern that is different from those produced by $T_{max}$ and $T_{min}$. Most notably, most commutes occur on roads that are closer to the inner city and there are several dominant commuting corridors, for instance, the ones originating from the inner city (area within the small red circle) and penetrating into suburbs or exurbs southeast, northeast and east to the inner city.

By comparing the flows between Shanghai and Beijing (See Figures 3 and 4), we can know better the characteristics of those in Shanghai. Based on the $T_{act}$ panels, in Shanghai, the most notable commuting flows are within the inner city (areas within the inner ring roads) and in the southwest while in Beijing, the inner city (areas within the third ring roads) and the north tend to have a much higher concentration of the commuting flows. Based on the $T_{min}$ panels, there tend to more commuting flows outside the inner city of Shanghai as compared to Beijing, especially in the west and in the southwest. In Beijing, there are significantly more traffic flows in the north of the center (Tian'anmen). Based on the $T_{max}$ panels, there are much more commuting flows from the south into the inner city of Shanghai while there are more commuting flows from the north into the inner city of Beijing. This may reflect that the two cities have significantly different jobs-housing distribution/separation. But one thing should be noticed is that in the case of Beijing, only commuting flows by bus are considered while all commuting flows are considered in Shanghai.

Fig.4: $T_{min}$, $T_{act}$ and $T_{max}$ Bus-Commuting Flows in Beijing

Except the above figures, the other way to visualize commuting flows associated with $T_{min}$, $T_{max}$ and $T_{act}$ is to map out the percentage of commuters across different distance ranges. Figure 5 presents the case of Shanghai, which represents a sample of 6.3 million local commuters.

Fig.5: Commuting Distance Distribution of Different Flows: $T_{min}$, $T_{act}$ and $T_{max}$

It can be seen from Figure 5 that the distributions of the $T_{act}$ and $T_{min}$ flows are almost identical when commuting distance is larger than 12.5 km. This can mean that for most commuters in the city, when they are willing to travel a distance of more than 12.5 km, there is always at least one acceptable job available to them. But if they are unwilling to do so, their odds of finding an acceptable job are low. Given that the above is only about Shanghai, it is unclear whether the 12.5 km or another cutting-off point exist for other metropolises. But it should be interesting to expand related work in the future. If one further compares Figures 3 and 5, s/he can also realize that Figure 3 could be somewhat misleading as it shows vast differences between the $T_{min}$ and $T_{act}$ flows. But Figure 5 indicates that some flows are very similar to one another, at least distance-wise.

## DISCUSSION AND CONCLUSIONS

The contributions of this study can be seen by comparing it to similar studies done before. Compared to existing studies focusing on Shanghai such as Feng et al. (2011), Sun et al.

(2013) and Chen et al. (2014) that are based on small data, this study has produced some new findings/visuals that are not possible before such as:

- The spatial distribution of workplaces and residences of a much higher percentage (about one fourth) of all the local residents in ShanghaiJOurn;
- Commuting flows to and from different types subareas of interest to local planners and decision-makers;
- Where there could be potential for better jobs-housing balance or shorter commutes based on the comparisons/visuals of the $T_{min}$ and $T_{act}$ commuting flows (Figures 3 and 5).

Compared to existing studies focusing on Shanghai such as Zhang (2016), Ding et al. (2015) and Niu and Ding (2015) that are also based on cellular network data, this study has completed extra tasks and generated many more new findings such as:

- It quantified several extra jobs-housing balance indicators for Shanghai and found that (a) inner-city resident-workers (those workers whose residence is in the inner city) in Shanghai enjoy the highest jobs/housing ratio and have the shortest average commuting distance; (b) inner-city workers (those workers whose workplace is within the inner city) on average have the longest commuting distance.
- It considered commuting patterns of Shanghai workers in two extreme situations according to the excess commuting framework, compared them with those for other cities whenever possible and found that: (a) $T_{min}$ value of Shanghai is comparable to those in Guangzhou, Dublin and Beijing, meaning all these cities have similar levels of spatial correlation and separation of jobs and housing; (b) $T_{max}$ of Shanghai is

comparable to that of Tokyo (both are nearly 50km), meaning that in the worst scenario, workers in both cities can suffer from a long commute; (c) $C_e$ and $NC_e$ values (which measure how corresponding community patterns depart from the random one) of Shanghai are much higher than those in Dublin, meaning that Shanghai's commuting pattern is not as random as that in Dublin; (d) Shanghai's actual commuting pattern is notably different from the two extreme ones ($T_{min}$ and $T_{max}$), meaning that the commuting pattern of Shanghai is probably right in the middle of two extremes; (e) the distance distribution of the $T_{min}$ and $T_{act}$ flows in Shanghai is quite similar when the distance is larger than 12.5 km, meaning that if we want to optimize the commuting pattern of Shanghai, more attention should be paid to commuters whose commuting distance is less than 12.5 km.

More generally, most existing studies of the jobs-housing balance have relied on small data. The emergence of big data has provided new opportunities for, and challenges to these studies. This study, via the case of Shanghai, shows that big data could at least change the existing studies in two aspects. One, it can provide researchers with a much larger sample than even before; Second, it can supply researchers with samples in a much higher resolution than before. In the case of Shanghai, millions of commuters who use a mobile phone were detected, which account for at least one fourth of the metropolis' long-term residents, and their jobs and housing locations were geocoded at the BTS service area level, which on average is 0.16 square kilometers.

With a larger sample in a higher resolution, researchers can do much more than what they can do in the small-data era. This study, for instance, maps out the job and housing locations of millions of commuters by subareas of interest to local planners and decision-makers. One can quantify the commuting flows and average commuting distances between and within those subareas as well. With a larger sample in a higher resolution for a longer period of time, there is also much more what could be done about the jobs-housing balance studies. In particular, the daily, weekly, monthly and yearly changes in the local jobs-housing balance and associated commuting patterns. This is simply difficult and costly if we only rely on small data. But understandings of those changes and their underlying dynamics could help us better manage our land use-transportation systems and increase the overall social welfare of commuters and/or travelers. For instance, systematic and comprehensive land-use adjustments in light of the $T_{min}$ and $T_{act}$ flows (e.g., Figures 3 and 5) could help us reduce the average commuting distance among workers. What's more, with some add-on surveys of local mobile-phone users, one could also segment local workers/mobile-phone users into more meaningful subgroups, for example, the low-income and the migrants, and better study and serve them. Related insights that are routinely updated would also help us make informed housing and employment policies and better keep track of the social welfare of various subgroups that are of policy relevance.

## REFERENCES

Ahas, R. Aasa, A. Mark, U., Pae, T. And Kull, A. (2007). "Seasonal tourism spaces in Estonia: Case study with mobile positioning data." Tourism Management, 28, 898-910.

Ahas R, Aasa A, Silm S, Tiru M. (2010a). "Daily rhythms of suburban commuters' movements in the Tallinn metropolitan area: Case study with mobile positioning data." Transportation Research C, 18(1), 45–54

Ahas R, Silm S, Jarv O, Saluveer E, Tiru M. (2010b). "Using mobile positioning data to model locations meaningful to users of mobile phones." Journal of Urban Technology, 17(1), 3–27

Batty, M. (2012). “Smart cities, big data.” Environment and Planning B, 39(2), 191-193.

Batty. M. (2013). “Big Data, Smart Cities and City Planning.” Dialogues in Human Geography, 3(3), 274-279.

Boyd, D. and Crawford, K. (2012). “Critical questions for big data”. Information, Communication and Society 15(5), 662–79.

California Planning Roundtable. (2008). “Deconstructing Jobs-Housing Balance. ” California Planning Roundtable, Sacramento, CA. Available at: http://goo.gl/Q5Z4m4, Accessed November 20, 2015.

Cervero, R. (1989). “Jobs–housing balancing and regional mobility.” Journal of the American Planning Association, 55(2), 136–150.

Cervero, R. (1991). “Jobs-housing balance as public policy.” Urban Land, 10, 14.

Cervero, R. and Landis, J. (1995). “The transportation-land use connection still matters.” ACCESS Magazine, 1(7), 2-10.

Charron, M., (2007). “From excess commuting to commuting possibilities: More extension to the concept of excess commuting.” Environment and Planning A, 39(5), 1238–1254.

Chen, C. (2014).” Agency and academic experience using cell data.” In: Federal Highway Administration (FHWA), 2014. Cell phone data and travel behavior research: symposium summary report. Washington DC:  FHWA. US Department of Transportation.

Chen, Z., Yang, D. and Guo, G. (2014). “Research on separation between workplace and residence in large community in Shanghai” (In Chinese). Transportation Standardization, 42(15), 19-24.

Ding, L., Niu,X., and Song. X. (2015). “Identifying the commuting area of Shanghai Central City using mobile phone data” (In Chinese). City Planning Review, 39(9), 100-106.

Dodge, M. and Kitchin, R. (2005). “Codes of life: Identification codes and the machine-readable world.” Environment and Planning D, 23(6), 851–81.

Dubin, R. (1991). “Commuting patterns and firm decentralization”. Land Economy, 67(1), 15–29.

Dumbill, E. (2012). “Getting up to Speed with Big data”. In O’Reilly. 2012. Big Data Now: Current perspectives from O’Reilly Media. Beijing, Cambridge, Farnham, Koln, Sebastopol, Tokyo.

Feng, X., Yu, Y., Sun, B. and Guo, Y. (2011). “Commuting performance of polycentric urban spatial structure” (In Chinese). Urban Economy of China, no. 11. 20-21.

Giuliano, G., 1991. “Is jobs-housing balance a transportation issue?” Transportation Research Record, 1935, 305–312.

Gordon, P., Kumar, A. and Richardson, H.W. (1989). “The influence of metropolitan spatial structure on commuting time.” Journal of Urban Economy, 26(2), 138–151.

Horner, M.W. (2002). “Extensions to the concept of excess commuting.” Environment and Planning A, 34(3), 543–566.

Horner, M.W. (2010). “How does ignoring worker class affect measuring jobs-housing balance? Exploratory spatial data analysis.” Transportation Research Record, 2163, 57-64.

Horner, M.W. and Mefford, J. (2007). “Investigating urban spatial mismatch using jobs-housing indicators to model home-work separation.” Environment and Planning A, 39(6), 1420–1440.

Horner, M.W. And Murray, A.T.(2002). “Excess commuting and the modifiable areal unit problem.” Urban Studies, 39(1), 131-139.

Kanaroglou, P.S., Higgins, C.D. and Chowdhury, T.A. (2015). “Excess commuting: a critical review and comparative analysis of concepts, indices, and policy implications.” Journal of Transport Geography, 44, 13-23.

Kim, S. (2005). “Excess commuting for two-worker households in the Los Angeles Metropolitan Area.” Journal of Urban Economics, 38(2), 166-182.

Kitch, R. (2013). “The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences.” London, UK, Sage.

Kung, K. S., Greco, K., Sobolevsky, S. and Ratti, C. (2014). “Exploring universal patterns in human home-work commuting from mobile phone data.” PLoS ONE, 9, e96180.

Lee, S.Suzuki, T. and Lee, M. (2006). “A study on the change of urban structure and commuting based on optimal commuting assignment problem in Korean and Japanese Metropolitan Areas.” Journal of the Korea Planners Association, 41(2),57-65.

Levinson, D.M., (1998). “Accessibility and the Journey to Work. ” Journal of Transport Geography, 6(1), 11–21.

Litman, T. and Burwell, D. (2006). “Issues in Sustainable Transportation.” International Journal of Environmental Issue, 6(4), 331-347.

Liu, W., Yan, X., Fang, Y. and Cao, X. (2008), “Related characteristics and mechanisms for excess commuting in Guangzhou (in Chinese).” Acta Geographica Sinica, 63, 1085–1096.

Loo, B.P.Y. and Chow, A.S.Y. (2011). Jobs-housing balance in an era of population decentralization: An analytical framework and a case study.” Journal of Transport Geography, 19(4), 552–562.

Ma, K.R. and Banister, D. (2006). “Extended excess commuting: A measure of the jobs-housing imbalance in Seoul.” Urban Studies, 43(11), 2099–2113.

Margolis, J. (1973). “Municipal fiscal Structure in a metropolitan region.” In: Grienson, R.E. (ed.), Urban Economics: Readings and Analysis. Little Brown, Boston.

Marz, N. and Warren, J. (2012). “Big Data: Principles and Best Practices of Scalable Realtime Data Systems.” MEAP edition. Manning, Shelter Island, New York.

Mayer-Schonberger, V. and Cukier, K. (2013). “Big data: A Revolution that Will Transform How We Live, Work, and Think.” New York: Houghton Mifflin Publishing House.

McKinsey Global Institute. (2015). “Open Data: Unlocking Innovation and Performance with Liquid Information”, Available at: http://goo.gl/IavBty, Accessed Feb 13, 2015.

Morrison, N. and Monk, S. (2006). “Job housing mismatch: Affordability crisis in Surrey, South East England.” Environment and Planning A, 38(6), 1115-1130.

Murphy, E. and Killen, J.E. (2011). “Commuting economy: An alternative approach for assessing regional commuting efficiency.” Urban Studies, 48(6), 1255–1272.

Niedzielski, M. A., Horner, M. W. and Xiao, N. C. (2013). “Analyzing scale independence in jobs-housing and commute efficiency metrics.” Transportation Research A, 58,129-143.

Niu, X. and Ding. L. (2015). “Analyzing job-housing spatial relationship in Shanghai using Mobile phone data: Some conclusions and discussions” (In Chinese). Shanghai Urban Planning Review, No.2. 39-43.

Peng, Z. R. (1997). The Jobs–Housing Balance and Urban Commuting. Urban Studies 34(8), 1215-1235.

Pelletier, M.P., Trepanier, M. and Morency, C. (2011). “Smart card data use in public transit: A literature review.” Transportation Research Part C, 19(4), 557-568

Pucci, P. and Tagliolato, P. (2015). “Mapping urban practices through mobile phone data.” PoliMI SpringerBriefs, DOI 10.1007/978-3-319-14833-52

Schweitzer, L. (2014). “Planning and social media: A case study of public transit and stigma on Twitter.” Journal of the American Planning Association, 80(3), 218-238.

Small, K.A. and Song, S. (1992). “Wasteful commuting – A resolution.” Journal of Political Economy, 100(4), 888–898.

Sultana, S. (2002). “Job/Housing imbalance and commuting time in the Atlanta Metropolitan Area: Exploration of causes of longer commuting time.” Urban Geography, 23(8), 728–749.

Sun, B., Tu, T., Shi, W. and Guo, Y. (2013). “Test on the performance of polycentric spatial structure as a measure of congestion reduction in megacities: The case study of Shanghai” (In Chinese). Urban Planning Form, no.2, 63-69.

Suzuki, T. and Lee, S. (2012). “Jobs–housing imbalance, spatial correlation, and excess commuting.” Transportation Research A, 46(2), 322–336.

Vieira, Marcos R. , Frias-Martinez, Vanessa , Oliver, Nuria. (2010). “Characterizing dens e urban areas from mobile phone-call data: Discovery and social dynamics.” Proceeding of IEEE Second International Conference on Social Computing, Minneapolis, MN. (DOI: 10.1109/SocialCom.2010.41)

Wachs, M., Taylor, B.D., Levine, L. and Ong, P. (1993). “The changing commute: A case study of the jobs-housing relationship over time.” Urban Studies, 30, 1711–1729.

Wang, D., Chai, Y.W. (2009). “The jobs–housing relationship and commuting in Beijing, China: The legacy of Danwei.” Journal of Transport Geography, 17, 30–38.

Weitz, J. (2003). “Jobs-housing Balance.” Chicago, IL: American Planning Association.

Yang, J. W. and Ferreira, J. R. (2008). “Choices vs. Choice sets: A commuting spectrum method for representing job–housing possibilities.” Environment and Planning B, 35(2), 364-378.

Zhang, T. (2016). “Shanghai job-housing spatial analysis based on cell phone big data” (In Chinese). Urban Transport of China, no. 1, 15-23.

Zhou, J., Chen, X., Huang, W., Yu, P. and Zhang, C. (2013). “Jobs-housing balance and commute efficiency in cities of central and western China: A case study of Xi’an” (In Chinese). Acta Geographica Sinica, 68, 1316–1330.

Zhou, J. and Long, Y. (2014). “Jobs-housing balance of bus commuters in Beijing.” Transportation Research Record, 2418, 1-10.

Zhou, J. and Long, Y. (2015). “Losers and Pareto Optimality in optimizing commuting patterns.” Urban Studies. DOI: 10.1177/0042098015594072.

Zhou, J., Murphy, E. and Long, Y. (2014a). “Commuting efficiency in the Beijing Metropolitan Area: An exploration combing smartcard and travel survey data.” Journal of Transport Geography, 41, 175-183.

Zhou, J., Murphy, E. and Long, Y. (2014b). “Visualizing the minimum and maximum solutions of the transportation problem of linear programming for Beijing’s Bus Commuters.” Environment and Planning A, 46, 2051-2054.