Urban Science Methods for Characterizing Human Mobility: A Case Study of Mexico City

Pierre Melikov, Jeremy A. Kho, Vincent Fighiera, Fahad Alhasoun, Jorge Audiffred, José L. Mateos and Marta C. González

Abstract Throughout the world, people struggle with safety, and access to economic opportunities, and health. Seamless access to destinations they value, such as workplaces, schools, hospitals, and parks, influences their quality of life. One of the first steps to planning and improving accessibility is to estimate the number of trips being made across different parts of a city. A challenge is that the spatial distribution and availability of urban services, may come short in supplying the needs of their inhabitants. Relying on expensive and infrequently collected travel surveys for modeling trip distributions to their facilities slow down the decision making process. The growing abundance of data already collected, if analyzed with the right methods, can help us in planning and understanding cities. In this chapter, we examine the use of points of interests (POIs) registered on Google Places to approximate trip attraction in a city. First, we compare the result of trip distribution models that utilize only POIs with those utilizing conventional data sets, based on surveys. We show that an extended radiation model provides a high degree of accuracy when compared with the official Origin-Destination Matrices from the latest census in Mexico City. Finally, we analyze trip modes and travel time showing that a clustering method is suitable to extract groups by their mode and travel time.

P. Melikov (pierre_melikov@berkeley.edu) – J. A. Kho (jerkho@berkeley.edu) –

V. Fighiera (vincent.fighiera@berkeley.edu) – M. C. González (martag@berkeley.edu)

University of California, Berkeley, Berkeley, CA

F. Alhasoun (fha@mit.edu), Massachusetts Institute of Technology, Cambridge, MA,

Jorge Audiffred (ja@digitalstate.mx), Data Lab MX, Mexico City, México.

J. L. Mateos (mateos@fisica.unam.mx), Universidad Nacional Autónoma de México, Mexico City, México

Codes avialable: https://github.com/VincentFig/urban_computing_mexico

1 Introduction

As more people continue to migrate from rural to urban settings, the challenges of improving cities increase in pace and complexity. Planning for daily mobility, within metropolitan areas is one important topic of the coming years. Climate change, overpopulation, and the cost of living in highly urbanized areas are issues at the heart of their public policies. These are all in connection with meeting transportation needs and have economic and environmental impact. The estimates of the total daily trips specific to a metropolis is the first step to establish efficient strategies that inform the transportation planning process. However, the lack of reliable and accessible data sources of individual mobility greatly slow down the planning progress. Data on human mobility has thus far been collected through individual surveys that are prohibitively expensive to administer and are plagued by small and potentially biased sample sizes. This is because they require active participation and often rely on self-reporting [1]. While conventional travel surveys provide a wealth of valuable information, they are also very expensive and time-intensive. For most major cities, these are conducted about once a decade; for smaller cities and towns, it is more seldom than that or none at all. Between the publications of these surveys, a lot can happen that could change the dynamic of the city: new attractions, redevelopment of entire city blocks, changing economic trends, impact of a natural calamity, or just the gradual shift of a city's characteristics. These changes would not be captured until the next travel survey is issued, which could be anywhere from the following year to a decade. With the abundance of information and connectivity today, other sources of easily accessible data could prove to be useful as proxy to the data obtained in conventional surveys. One example of this is the use of triangulated mobile phone data to form mobility networks to extract individual trip chains [6]. Another such potential is points of interest (POIs) registered on Google Places, a feature of the mapping service developed by Google LLC (Google), which are extensive, updated frequently, and relatively accessible for most people. Google Places lists various types of establishments, such as restaurants, schools, offices, and hospitals, allowing it to serve as a good indicator of trip attraction.

As a complement to the development of statistical methods to carefully treat travel diaries [2–4], new, cheaper, and larger data sources are necessary to push our understanding of human mobility efforts further. The evolution of technology over the past decade has given rise to ubiquitous mobile computing, a revolution that allows billions of individuals to access people, information, and services through information technologies such as their cellular phones. Using today's large-scale computing infrastructure and data gathered from sensing technologies, one can combine methods from computer science with urban planning, transportation, and environmental science, to tackle specific problems with fined tuned methodologies in a data-centric computing framework.

In this chapter we focus in methods to analyze and model human mobility both aggregated and individually. We take advantage of novel data sources passively collected, to enrich the information on human mobility patterns. Namely, we parse an alternative source of geospatial data, apply trip distribution models to estimate aggregated trips, and implement unsupervised machine learning to characterize different types of commuters by their mode of transportation and travel time.

As a sample case, we focus on Mexico City, one of the largest cities in the world with over 21 million people in the greater metropolitan area. It is also one of the most important cultural and historical centers in the Americas. With such a large amount of people and a high level of vibrancy, mobility in the region can be quite a challenge. In 2017, a major household travel survey [5] was completed for the Metropolitan Zone of the Valley of Mexico. Conducted from January to March of 2017, the survey obtained information to facilitate better understanding of mobility of the inhabitants in the metropolitan region. This includes data on trip generation, trip attraction, mode choice, trip purpose, trip duration, socio-demographics, and more, which is representative of 34.56 million daily trips occurring in our study zone.

2 Data Collection of POIs

In order to obtain POIs from Google Places, programming scripts are written to utilize the Application Programming Interface (API) that Google provides [7]. However, Google sets limits on the number of POIs a single request can return and on the number of API requests an account is allowed to make in order to differentiate commercial and non-commercial applications. While the conduct of this undertaking is non-commercial, the data to be collected tends to exceed Google's limitations. Hence, an efficient algorithm needs to be implemented to collect the most information from a minimal number of API requests.

To achieve this, API requests are framed and constrained by geometries defined by the Hexagonal Hierarchical Geospatial Indexing System (H3) of Uber Technologies, Inc (Uber) [8]. Uber's H3 system is an application of the concept of fractals. Maps are divided into large hexagonal tiles, with each tile further divided into seven

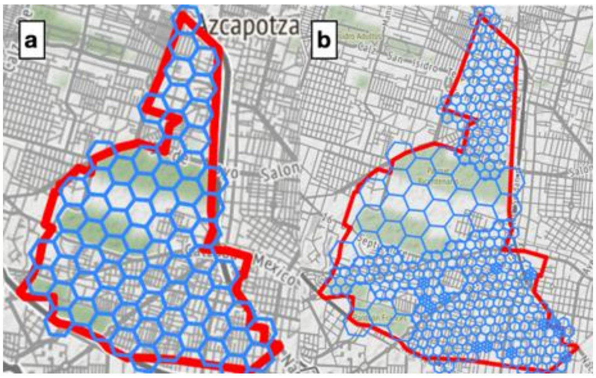

Fig. 1 Hierarchical sampling method to extract POIs. a Initial state and resolution of parsing algorithm, **b** Final state after recursively increasing resolution in hexagons that reach the API request limit.

smaller hexagons. With 16 supported resolutions, the system is flexible to most use cases. Figure 1a shows a sample resolution applied on a district in Mexico City.

Hexagons serve as good approximations of circles, while minimizing overlap between cells. This is useful as the Google Places API requires a radius parameter within which the search for POIs will be made.

2.1 Parsing Algorithm

An initial resolution for the size of the hexagons is determined. The lower the initial resolution, the more efficient the script is likely to run, as excessive requests

are avoided in sparsely developed areas. On the other hand, low resolutions also increase the marginal areas near the borders of irregular shapes unaccounted by the algorithm. Before issuing any API request, the initial resolution is tuned and visualized to balance these tradeoffs.

For each hexagon, an API request is made at the centroid. If the request reaches the limit of POIs it can return, the algorithm subdivides that hexagon into smaller hexagons. This process is repeated until each request is met without reaching the limit. In Figure 1b, some areas, such as parks and nature reserves, do not need numerous API requests. Downtown city blocks and dense neighborhoods, on the other hand, are recursively splintered.

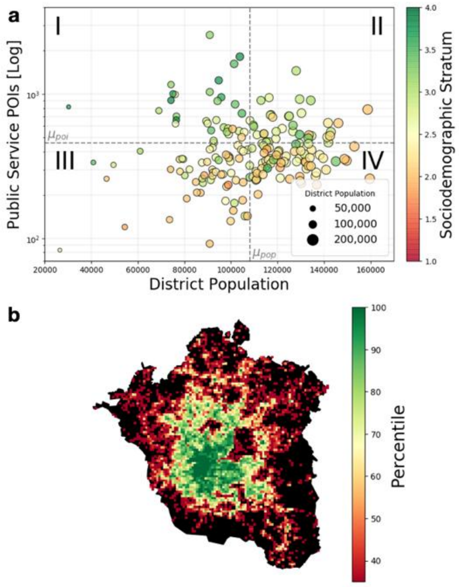

Fig. 2 Spatial distribution of population and services. **a** Relationship of the sociodemographic stratum of a district with the ratio of the number of public service establishments to the population, **b** Percentiles of the number of public service POIs for each 1 square km. block

3 Spatial Distribution of POIs

In the use case for this chapter, the parsing algorithm returned a total of over 733,000 POIs from Google Places across the Metropolitan Zone of the Valley of Mexico. These points of interest provide new dimensions to analyzing data from the travel survey that could generate insights on characteristics of the megacity.

For instance, the API requests return tags for each POI, indicating the nature of the establishment. This may include broad categories, such as 'store', or more specific labels, such as 'electronic store'. Clustering relevant tags together, POIs may be classified as either commercial or public service establishments. Combining this data with the travel survey, Figure

2a maps the relationship of the sociodemographic status of a district with the ratio of the number of public service establishments to the population.

In this case, sociodemographic strata are indices defined by the travel survey to characterize a respondent's social and economic conditions, with numbers from 1 to 4, denoting an increasing economic well-being. In Quadrant I, the number of public service establishments are above average and the population is below average, districts tend to enjoy the highest sociodemographic stratum. Quadrant II has districts of intermediate sociodemographic status, still benefiting from an above average number of POIs. Quadrant III has both less than average population and number of facilities and lower socio economic stratum. Interestingly, Quadrant IV has districts on opposite ends of the sociodemographic spectrum, possibly due to the diversity of inner cities and the efficiencies of density that allow fewer establishments to serve more people in a small amount of space. These enrich the spatial information of the surveys deserves further research.

Another advantage gained through the POIs is the spatial granularity of the collected data. Travel survey respondents are often organized by district of residence, whereas establishments on Google Places are pinpointed to streets address coordinates. Since cities and districts are not homogeneous, this level of detail provides a more realistic perspective of city dynamics, highlighting functional interaction over arbitrary political boundaries.

In Figure 2b, the coordinates of public service establishments are truncated to two decimal places, binning them to grids that are approximately a kilometer per side. Due to the orders of magnitude in difference between the urban core and more rural areas, the number of public service establishments are abstracted to intervals of 5 percentile points. As it is, mapping these establishments may have a strong dependency on population density. Nevertheless, a hidden structure to the city is revealed, with a strong urban core, some urban corridors expanding outwards from the city center, and regional centers further away from the center. Significantly, there are large regions on the outskirts of the study area where public services are sparse. Further insights may be gained when supplemented by population distribution data at a similar level of granularity.

4 Extended Radiation Model for Human Mobility

Counting the number of POIs per district is necessary for direct comparison with the 2017 travel survey data which has the smallest granularity only at the level of districts. Mapping these per district in Figures 3a and 3b, a direct comparison can be made with trip attraction reported in the 2017 travel survey.

While the correspondence is not perfect, the distribution of points of interest makes a good approximation of the distribution of Trip Attraction obtained from the travel survey. Most notably, the difference between the city center and the rest of the region is similarly stark. Plotting the relationship between trip attraction and points of

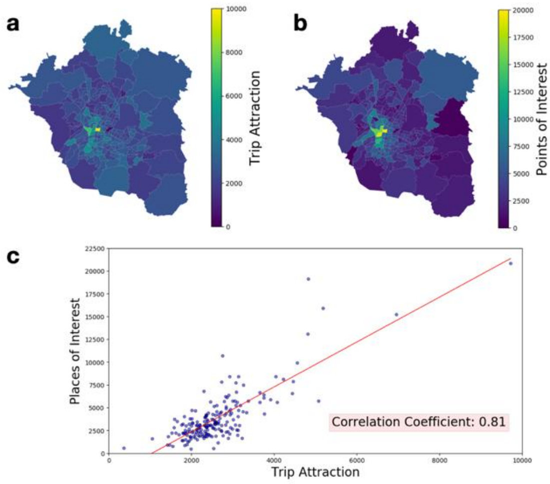

Fig. 3 Trips attraction vs. POIs. **a** Values of trip attraction, **b** Number of points of interest, **c** Correlation plot of trip attraction and points of interests.

interest in Figure 3c yields a quantitative plot, with the correlation coefficient of the two variables determined to be quite high at 0.81. This comparison will be of great relevance later, where the POIs are used to model mobility patterns in the city, in place of travel survey data.

Many models have been developed in order to predict population movement at different scales. In the context of the Greater Mexico City we want to investigate how accurate such models are and how well do they perform to reconstruct mobility patterns. The models of trip distribution can be divided in gravity model types for [9, 10, 11,12], or intervening opportunity types [12]. In this chapter we present an application of the later, named the extended radiation model [13], to estimate trip distributions in Mexico City.

The radiation model [14, 15] is based on a stochastic process that is parameter-free nature and enables, without previous mobility measurements, estimates of trip distributions in good agreement with mobility and transport patterns [14]. The original radiation model only relies on population densities and to estimate commuting patterns between U.S. counties [14].

Here we use the natural partition of the city in districts, the model states that a trip occurs based on the number of opportunities that can be found in each district if the two following steps are met: (1) an individual seeks opportunities from all districts, including his/her home district (the number of opportunities in each county is proportional to the resident population); (2) the individual goes to the closest district that offers more opportunities than his/her home district. To analytically predict the commuting fluxes with the radiation model, we consider locations i and j with population m_i and n_j respectively, at distance r_{ij} from each other. We denote with s_{ij} the total population in the circle of radius r_{ij} centered at i (excluding the source and destination population). The average flux T_{ij} from i to j, is

$$\langle T_{ij}\rangle = T_i \frac{m_i n_j}{(m_i + s_{ij})(m_i + n_j + s_{ij})} \#(4.2.1)$$

where $T_i \equiv \sum_{i \neq j} T_{ij}$ is the total number of commuters that start their journey from location i, or "trip production" of location i.

The extended radiation model aims at predicting flows without previous calibrating data. Thus, it introduces a scaling parameter $\propto$ by combining the derivation of the original radiation model with survival analysis and gives

$$\langle T_{ij}\rangle = \gamma T_i \frac{\left[(a_{ij} + m_j)^{\propto} - a_{ij}^{\propto}\right](n_i^{\propto} + 1)}{(a_{ij}^{\propto} + 1)\left[(a_{ij} + m_j)^{\propto} + 1\right]} \#(4.2.2)$$

where $a_{ij} = n_i + s_{ij}, \gamma$ is the percentage trip attractions between all places found origin and destination, and empirically sets $\propto = \left(\frac{l}{36\ [km]}\right)^{1.33}$, where l is the characteristic length of the study area, $\propto$ account for the fact that the trip distributions depend on the area of study.

The extended radiation model, was meant to be used when we lack from trip data for calibration. When there is actual trip data as in this case, one can evaluate it with the Common Part of Commuters based on the Sørensen index

$$CPC(T, \tilde{T}) = \frac{2\ \sum_{i=1}^{n}\sum_{j=1}^{n}\min\left(T_{ij}, \widetilde{T_{ij}}\right)}{\sum_{i=1}^{n}\sum_{j=1}^{n} T_{ij} + \sum_{i=1}^{n}\sum_{j=1}^{n}\widetilde{T_{ij}}} \#(4.2.3)$$

It gives a quantitative measure of the goodness of the flow estimation, 0 meaning no agreement found and 1 perfect estimation. CPC compares the model estimates T_{ij} vs. the empirical observations $\widetilde{T_{ij}}$, between all origin destination pairs.

4.1 Results

From the survey data, we extracted the different variables to run the extended radiation model. First, we extracted the 194 districts that compose Greater Mexico City with their respective population, trip attraction (number of daily trips coming to the district), trip production (number of daily trips leaving from the district), point of interest and characteristic length, given as the square root of the area of the district.

Then, we set l as the mean of the characteristic length of each district. We also constructed the distance matrix that gives for every row i and columns j the distance between the centroids of the districts i and j. Finally, γ is set to the proportion of the total number of trips over the total population.

Four different set ups are then used to compare the performance of the model based on different approximations of the trip generation from the origin districts and the trip attraction of the destination districts: (1) we use trip attraction and trip production as a baseline, (2) we use POI as a proxy for trip attraction, (3) we use population as a proxy for trip production, and (4) we combine (2) and (3). The resulting CPC values are shown in Table 1.

Origin	Trip Production	Trip Production	Population	Population
Destination	Trip Attraction	POI	Trip Attraction	POI
CPC	0.69	0.67	0.64	0.63

Table 1. Comparison of the goodness of fit depending on different input data in the model

Table 1 shows that the CPC of the estimates of the extended radiation model is close to other recently proposed models [12]. Moreover, we investigate the impact of different proxies for flow generation and attraction volumes as input in our model and find that the use of more easily acquired data sources such as population and POI density achieves nearly the same level of accuracy. POIs seem particularly interesting because they enable good estimates compared without travel surveys, but with data of much cheaper access. On the other hand, the use of population in place of trip production, aims at predicting future mobility patterns given the knowledge of γ the proportion of the total population of the system commuting, and assuming changes in this ratio. Here, we extracted γ on the 2017 survey and use it for the models. Consequently, we cannot validate the predictive power of the model but nonetheless, when distorting the population data of each district by multiplying it by γ but still observe encouraging results.

5 Analyzing Human Mobility by Mode of Transportation

This section is devoted to the analysis of individual travelers within Mexico City. One advantage of a broad user surveys is to identify types of dominant behavior in the population, respect to the modes of transportation used, their geographic distribution, and socio-demographic characteristics.

We analyze the large database collected by the Mexico City survey, containing information on individual residents, it details information on more than half a million trips. For each trip identified, we have: the mode of transportation, the districts of departure and arrival, the time of departure and arrival, the purpose of the trip, the gender of the traveler, her age and her socio-demographic stratum. As many as twenty different modes of transportation can be identified among the 196 districts of the survey.

We want to reduce the complexity of this information by groping the trips based on transportation mode, without associating the other metrics. The latter will then be involved in the analysis of clusters formed. In doing so, we seek to distinguish the main mobility behaviors, which will in turn combine various proportions of the possible transport modes and trip purposes.

By simple inspection, it is clear that all the means of transport mentioned in the database are not significantly present in the main groups of behaviors. We expect to see certain modes of transport - such as cars or walking - as the majority in certain behaviors and others, such as the category "Other means of transport", very poorly represented or even absent. It is therefore not needed such a large number of variables - initially twenty - to describe the individual trip database. We apply the Principal Component Analysis (PCA) method to determine the main variables. This will allow us to reduce computation time and complexity when using a clustering algorithm. Projecting into a lower dimensional base informs our understanding of it [16, 17].

The PCA method aims to capture as much of the total variance of the data as possible with a reduced number of variables - called Principal Components (PC). Since the criterion retains here to set the size of the new projected database is that the total variance captured by the N first PCs has to be more than a threshold of 85%, we therefore choose to keep only the first five PCs for the rest of the study [18].

To group trips around main behaviors, we use the K-means clustering algorithm. Each journey of the database is initially represented as a vector composed of zeros and ones, depending on the mode of transportation used. We only consider its projection in the PCs database when applying the K-means algorithm. K-means works iteratively to ultimately minimize the sum of the distances between each projected journey and the centroids of the clusters determined by the algorithm, and thus allow patterns to be identified within the dataset. As a result, we obtain a list that reflects the membership of each trip in a particular cluster. We also calculate the proportions of the modes of transport for each cluster to determine their average behavior. While the ideal number of clusters can be estimated via various metrics, such as in the Elbow method, the best number of clusters depends on the interpretability of the data available. In this case we decide to keep six clusters as the best results.

5.1 Results

Figure 4a in the top shows the six clusters that characterize daily mobility in Mexico City, and their percentages. They represent the main ways of moving around the city. Since the database reports journeys, several of which may have been made by the same person, and residents can have several trips. The analysis groups journeys and not individuals. Note that these journeys also have the purposes of these trips such as: going to home, going to work, errands, shopping, etc. Their average percentage is shown in the bottom of Figure 4a. In the top of Figure 4a, only the three most reported modes of transportation in each cluster are shown. Each of these components is associated in the y axis with its fraction within the cluster. The % in the x axis show the fraction of trips in each cluster.

Thus, based on cluster 2, we see that it contains 35% of all the trips in Mexico City. The fraction of the walking on the ordinate is equal to one, while that of the second most present mode of transportation in this cluster, Mexibus & Metrobus, has a fraction of 0.027. Thus, only about 2.7%

of the trips attached to this cluster combined their walking with Mexibus or Metrobus. It can therefore be said that these trips are made almost exclusively by walking.

Figure 4b shows, for each of the six clusters, the proportion, per cluster, of each of the ten purposes of the trips considered in the survey: going to home, going to work, going to school, shopping, leisure, errands, picking someone up, religion, health purposes, or all other purposes.

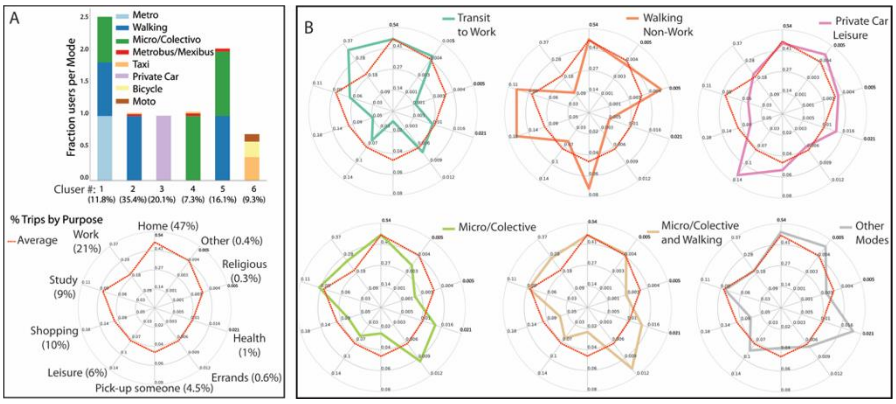

Fig. 4 Mobility groups in Mexico City. We indicate the three most important main means of transport for each group. The percentage of trips contained by each group is indicated at the bottom. Comparison of the mean of purpose trip variables within each group with respect to the median of all groups. Each letter is linked, in order, with the identified clusters of the Figure 4 (from 1 to 6 from left to right)

We compare the average percentage of trip purposes with the average within each cluster. Cluster 1 represents 11.8% of all the trips and have 33% of them with work as purpose, much larger than the 21% of the average among all trips. We see that when people walk (Cluster 2) the shopping purpose is twice as average. While about 16% of the trips associated with the second cluster are for shopping purposes, the average number for all trips is around 10% for this category. On the contrary, it seems that walking is not commonly used for commuting or going to the doctor.

In addition, since the average travel time of this cluster is about 20 minutes where the average cluster is about twice as long, this cluster can therefore be associated with local trips. This suggests that workplaces or care centers are generally located further from family homes than shops, schools or religious places.

Cluster 3 groups 20% of the trips made in Mexico City, it is exclusively composed by private car as a mode of transportation. This case has leisure in higher proportion compared to other clusters. This can be a consequence of the lack of transit to cover distance journeys or being inconvenient for such purpose.

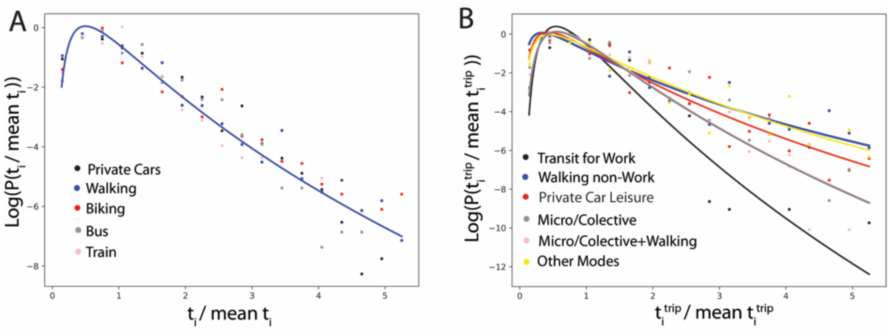

Fig 5. Comparison of travel times by mode and by cluster group. (a) Log normal fit for the scaled time-averaged travel-time distributions for different modes of transport on a logarithmic scale as reported in Ref [12] based on UK surveys. (b) Log normal fit for the scaled time-averaged travel-time distributions for the clusters found in the Mexico City travel survey.

Cluster 5 contains 16% of the trips and includes the routes that exclusively combine walking and micro collective, while Cluster 4 with 7% of the trips does not include walking. These two clusters are similar in purpose to the average and their average travel time is the longest, about one hour per trip.

The use of walking, metro and micro collective during the same journey is also observed in the first cluster. Indeed, the metro obtains a proportion equal to 1, walking 0.83 and the micro collective 0.71. Not all the journeys in this cluster therefore systematically combine these three means of transport, but this average behaviour indicates that in the vast majority of cases these three means of transport are combined. This behaviour is over-represented in the heart of the capital's historic district, where more than 55% of the trips undertaken are associated with this cluster. On the other hand, it becomes absent as soon as one moves away from this geographical area. This is due to the high concentration of metros and micro collectives in this part of the city, making travel much faster and more convenient by linking these modes of transport, particularly to get to work.

Cluster 6 is harder to interpret, because it does not represent any mode in large. However, it should be noted that it is mainly concentrated in the agricultural regions that make up some districts.

Koelbl and Helbing analyzed data from *UK National Travel Surveys* during the years 1972–98, observing that the average journey times for different modes of transport are inversely proportional to the energy consumption rates measured for the respective human physical activities. In Fig. 5a we show the distribution of the travel times per mode divided by their mean as reported in Ref. [19]. They presented five transportation modes, and they all collapse well in one lognormal distribution with parameters reported in Table 2. As further understanding of our clusters, we do the same analysis of the travel time of the individual trips divided by the mean travel time. We observe a lognormal with different parameters for each cluster, only cluster 5 have closer parameters to the ones reported in Ref. [19]. Given the challenges of mobility in Mexico City we observe larger variance among the members of each cluster, except for the trips of cluster 1, which groups higher fraction of the journeys to work. The differences between the results reported in UK and Mexico City could be related to a more strained transit service and longer commuting journeys in a vast metropolis. The universal scaling shown in different modes in Ref. 12, could still serve as a guide to target improvements in the transit system. If private car and transit presented similar travel times and variance, these could be more attractive for those that can afford traveling by private car.

Cluster	M	σ^2	Mean t_i^{trip} [min]
Transit to Work	- 0.03	0.21	89
Walking non-Work	- 0.28	0.63	20
Private Car + Leisure	- 0.19	0.69	40
Micro/Collective	- 0.07	0.41	49
Micro/Collective + Walking	- 0.11	0.41	58
Other Modes	- 0.30	0.47	30
Ref[19]	- 0.14	0.51	N/A

Table 2. Comparison of the fitted parameters for the clusters

6 Conclusion

Data informed analysis of complex human socio-technical systems has become the interest of interdisciplinary groups around the world. These techniques can inform urban planning with an analytical angle in the complex task of amending current cities and their infrastructures. This increases its relevance to better accommodate the continued expansion of major cities and metropolises around the world. The purpose of this study was to show a sample of data analysis techniques drawn from various disciplines to better characterize urban mobility. The common aim of the methods presented is to reduce the complexity of the dataset at hand, while simultaneously extracting useful information. To this end, recent growth of passively collected data lends import opportunities to the understanding and implementation of these and other methods. In particular, we analyzed and model human mobility in the Great Mexico City, one of the largest cities in the world with over 21 million people. We explore the big data set of a recent major travel survey conducted in 2017, using machine learning methods and compare with the extended radiation model for human mobility applied to Mexico City.

The overall goal is to convert data into useful information that generates understanding within urban sciences, and facilitate the decision making process. It is also critical to maintain the measures by sociodemographic stratum, as the novel methods and data should serve as tools to plan for social equity and accessibility in the world's major cities.

Acknowledgments

We are grateful to Emmanuel Landa and Irving Morales of Data Lab MX for collaborating with us in collecting data and in gaining better insights to the Metropolitan Zone of the Valley of Mexico. This content for this chapter were initiated as a class project based on the content covered in CYPLAN 257: Data Science for Human Mobility and Sociotechnical Systems. The codes and data used in this chapter are available in: https://github.com/VincentFig/urban_computing_mexico

[1] Cottrill, C. D. A., Pereira, F. C. A., Zhao, F. A., Dias, I. F. B., Lim, H. B. C., Ben-Akiva, M. E.D., et al. (2013) Future mobility survey. *Transportation Research Record, 2354*, 59–67.
[2] Ben-Akiva, M. E., & Lerman, S. R. (1985). *Discrete choice analysis: Theory and application*

to travel demand. Cambridge: MIT Press.

[3] Hall, R.W. (Ed.) (1999). *Handbook of transportation science*. International series in operations
research & management science (Vol. 23). Boston: Springer.

[4] de Dios Ortúzar, J., & Willumsen, L. G. (2011). *Modelling transport*. Chichester: Wiley.

[5] Encuesta Origen-Destino en Hogares de la Zona Metropolitana del Valle de Mexico (2017) Instituto Nacional de Estadistica y Geografia, Mexico. http://en.www.inegi.org.mx/programas/eod/2017/. Accessed 11 October 2018

[6] Jiang S, Fiore GA, Yang Y, Ferreira J. Jr, Frazzoli E, Gonzalez MC (2013) A Review of Urban Computing for Mobile Phone Traces: Current Methods, Challenges, and Opportunities. Paper presented at the 2nd ACM SIGKDD International Workshop on Urban Computing, Chicago, Illinois, 11 August 2013

[7] https://developers.google.com/places/web-service/search

[8] https://eng.uber.com/h3/

[9] Barthélemy, M. Spatial networks. Phys. Rep. 499, 1–101 (2010).

[10] Erlander, S. & Stewart, N. F. The Gravity Model in Transportation Analysis: Theory and Extensions (VSP, 1990).

[11] Jung, W. S., Wang, F. & Stanley, H. E. Gravity model in the Korean highway. EPL 81, 48005 (2008)

[12] Lenormand M, Bassolas A & Ramasco JJ (2016) Systematic comparison of trip distribution laws and models. *Journal of Transport Geography* 51, 158-169.

[13] Yang Y, Herrera C, Eagle N, & González M. C. (2014) Limits of predictability in commuting flows in the absence of data for calibration. Scientific Reports. doi :10.1038/srep05662

[14] Simini, F., Maritan, A.&Neda, Z. Human mobility in a continuum approach. PloS one 8, e60069 (2013).

[15] Simini F, González M. C., Maritan A., & Barabási A. L. (2012). A universal model for mobility and migration patterns. Nature. doi: 10.1038/nature10856

[16] Eagle N. and Pentland A. S. (2009). Eigenbehaviors: identifying structure in routine. Behav Ecol Sociobiol 63:1057-1066

[17] Dorothy C. Ibes (2015) A multi-dimensional classification and equity analysis of an urban park system: A novel methodology and case study application. Landscape and Urban Planning 137 122-137

[18] Shlens J. A tutorial on Principal Component Analysis (December 10, 2005; Version 2)

[19] Kölbl, Robert, and Dirk Helbing. Energy laws in human travel behaviour. *New Journal of Physics* 5.1 (2003): 48.