A Multi-Scale Multi-Cultural Study of Commuting and its Applications for Characterizing Road Usage Patterns

Marta C. González *

*Massachusetts Institute of Technology, Cambridge, USA

Abbreviations: POI, point of interest;OD, Origin-Destination; MDS, Major Driver Source

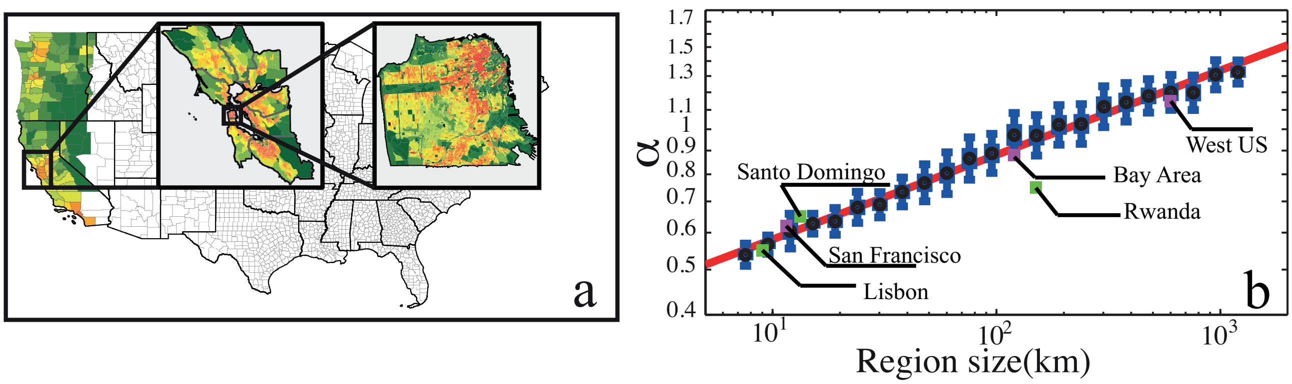

Fig. 1. (a) Three sample scales of the commuting study. (b) Functional relationship between α and the region size. 1000 regions with random centers and sizes are seleted and in each case the corresponding α value is measured. The relationship between α and the size of the region is red line and the error bar is in blue. Three special cases in the US: San Francisco, the Bay Area, and The West US are marked in magenta. The α values for the cell phone users in Lisbon, Santo Domingo, and Rwanda are in green. All of them, except Rwanda, conform to the functional relationship. In Rwanda the actual α value is smaller than the predicted one because the actual study region size, which cannot be measured should be much smaller than the whole country's size

Human mobility modeling is an essential component of various areas of study ranging from epidemiology to urban and transportation planning. Commuting flows take up a large proportion of the total flow of a population. The mainstream models for commuting flow prediction include the gravity model and the intervening opportunity model. These models require previous Origin-Destination matrices as input for parameter calibration. The recently proposed radiation model, which is parameter-free and has a closed analytical form, has the potential to become a universal model for mobility patterns. We show that at large scales the radiation model's performance is comparable to models relying on adjustable parameters. But if we zoom in from the intra-city scale to the inner-city scale some extension must be applied to the model. We extend the radiation model by adding one parameter α which reflects the influence of the size of the study region. The extended radiation model gives close commuting flow predictions to the census data at all the scales. For regions without detailed census data but with available cell phone records, we propose a cell phone user OD matrix expansion model so that we could gain insights into these regions' commuting flow characteristics from cell phone records. This method is validated in the Bay Area and then applied to three different countries: Rwanda, Dominican Republic, and Portugal, through which some special commuting flow characteristics are observed.

In this study, we first compared the performance of the doubly constrained gravity model with the radiation model in its original form at different scales from San Francisco to the entire west coast of the US. Each of them is applicable to certain scales. Then based on these two models, we proposed an extension of the radiation model with only one parameter α but obtaining good performances at all scales. Moreover, we found a functional relationship between the parameter and the scale of the study region. We did this by randomly selecting 1000 different regions in the US with different region centers and region sizes (ranging from a few kilometers to 1000 kilometers); then used the census commuting OD statistics to calibrate the best α value for each region. The result shows a clear functional relationship between the parameter α and the size of a region. This makes the model's parameter predictable and thus applicable to regions without empirical OD matrices for parameter calibration.

Usually the gravity model, the intervening opportunities model and the radiation model all use population density as a proxy for both trip generation and trip attraction rates. We show that while this approximation is reasonable at large scales, at inner city scale it does not hold. When a city is divided into block groups population density can only represent the trip generation but not the attraction rate. However the availability of various kinds of urban digital traces provide us with more choices of data sources. We find that digital geo-located information such as the point of interests (POIs) is a good representation of trip attraction rates. In addition, since a detailed census is time consuming and costly, many countries lack accurate population and commuting profiles. However, cell phone service has covered almost all populated regions in the world. We show that cell phone records alone are sufficient to provide a seed commuting OD matrix, which can be expanded to recover the full commuting OD matrix for the whole population under study. Thus cell phone records, together with population and POI densities which are available worldwide, could allow us to model the commuting flow patterns of regions that lack traditional survey data.

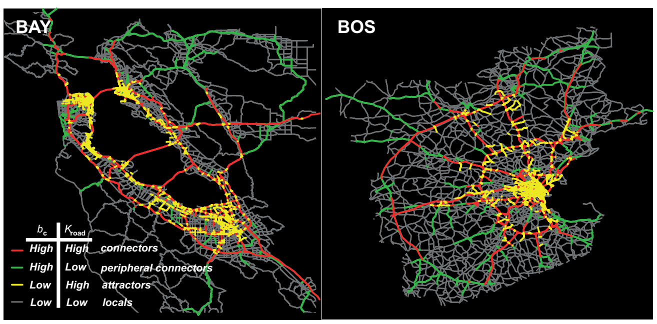

Fig. 2. Types of roads defined by b_c and K_{road}. The road segments are grouped by their betweenness centrality b_c and degree K_{road}. The red lines (connectors) represent the road segments with the top 25% of b_c and K_{road}; they are topologically important and diversely used by drivers. The green lines (peripheral connectors) represent the road segments in the top 25% of b_c, but with low values of K_{road}; they are topologically important, but less diversely used. The road segments in yellow are those with low values of b_c, but within the top 25% K_{road}; they behave as attractors to drivers from many sources (attractors). The road segments in grey have the low values of b_c and K_{road}, they are not topologically important and are locally used (locals).

For each mobile phone user that generated the OD, we additionally locate the zone where he or she lives, which further on we will call the driver source. Connecting ODs with driver sources allows us for the first time to take advantage of mobile phone data sets in order to understand urban road usage. A driver source is calculated from the mobile phone data based on the regularity of visits of mobile phone users at each time of the day. This regularity is time dependent, and peaks at night when most people tend to be reliably at a home base with an average probability of 90%. Thus, we make a reasonable assumption that a driver source is the zone where the user is mostly found from 9pm to 6am in the entire observational period. We present the analysis of the road usage characterization in the morning period as a case study.

The traditional difficulty in gathering ODs at large scales has until now limited the comparison of roads in regard to their attractiveness for different driver sources, which inherently controls the formation of traffic flows in a road network. To quantify this aspect, for each road segment with $V > 0$, we calculate the fraction of traffic flow generated by each driver source, and rank these sources by their contribution to the traffic flow. Consequently, we define a road segments major driver sources (MDS) as the top ranked sources that produce 80% of its traffic flow. We next define a bipartite network, called here the network of road usage, formed by the edges connecting each road segment to their MDS. Hence, the degree of a driver source K_{source} is the number of road segments for which the driver source is a MDS, and the degree of a road segment K_{road} is the number of MDS that produce the vehicle flow in this road. To better understand a roads functionality, we classify roads in four groups according to their b_c and K_{road} in the transportation network (see Fig. 2). We define the connectors, as the road segments with the largest 25% of b_c and the attractors as the road segments with the largest 25% of K_{road}. The other two groups define the highways in the periphery, or peripheral connectors, and the majority of the roads are called local, which have both small b_c and K_{road} (Fig. 2). By combining b_c and K_{road}, a new quality in the understanding of urban road usage patterns can be achieved. Future models of distributed flows in road networks will benefit by incorporating usage patterns via the defined four roles. This novel framework of defining the roads by their connections to their MDS can trigger numerous applications. As a proof of concept, we also present how these findings can be applied to mitigate congestion.