# MobInsight: Understanding Urban Mobility with Crowd-powered Neighborhood Characterizations

Souneil Park[1], Marc Bourqui[2], Enrique Frias-Martinez[1], Nuria Oliver[1]

[1]Telefonica Research

Barcelona, Spain
{firstname.lastname}@telefonica.com

[2]School of Computer and Communication Sciences
EPFL
Lausanne, Switzerland
marc.bourqui@epfl.ch

***Abstract*—In this paper, we present *MobInsight*, an interactive visual tool for analyzing urban mobility. The tool aims to reveal the collective intelligence of the spatial choices expressed in the mobility patterns of the people that live in a city. It provides an analyst with a rich characterization of neighborhoods, enabling the analyst to compare the difference and infer possible reasons behind traveling behavior between the neighborhoods. MobInsight builds tailored neighborhood characterizations specific to the analyzed city by harnessing the geo-social annotations of the crowd. For the demonstration, MobInsight will feature Barcelona where the conference venue is located. Mobility patterns between the 70 neighborhoods of the city are extracted from real mobile network data of a large sample of residents, and the neighborhood characteristics are profiled by mining various online geo-social services and open government data.**

***Keywords—Urban informatics; mobility; neighborhood characterization; social annotations; visual analytics***

## I. Introduction

Aggregated patterns of mobility at a city scale are reflective of the collective use of the city by its dwellers. Each mobility instance is the results of a specific choice, which depends on many considerations, including the function and the character of the destination, the distance, and individual preferences. While technological advances have provided us with the means to capture human mobility, the analysis of the motivations and factors that play a role behind these choices involves multiple complexities, such as the scale of the urban areas under study, the diversity of places, and the heterogeneity of preferences among citizens. However, shedding light on these complexities can have a broad and direct impact on our understanding of urban life and can help us better manage our cities.

In this demonstration, we present *MobInsight*, a visual analytics tool for the interpretation of large-scale urban mobility. The tool allows an analyst to intuitively make sense of the frequency of mobility between neighborhoods and explore possible explanations for such mobility patterns by looking into the different characteristics between them. The core feature of MobInsight is its ability to associate urban mobility with *neighborhood characteristics*, which we analyze by comprehensively identifying existing places and harnessing rich crowd-sourced descriptions about them. Associating neighborhood characteristics with mobility provides new perspectives since they offer insights about why and for what purpose people might be moving. There has been previous work studying associations between urban mobility and a few representative statistics of the population, such as the total number of people [4] or the deprivation status [5]. While these statistics have some relation to the reasons of movements, rich characterization of neighborhoods can potentially capture more diverse and in depth aspects behind the choices of destinations.

We attempt to develop a rich characterization that reflects diverse facets of neighborhoods. Although there are general categorization schemes of areas often used in urban planning (e.g., residential, business, commercial, entertainment, etc.), we believe it is important to develop characterizations that involve the meanings given by the people in-situ in order to interpret the local mobility of the analyzed city. For the characterization, MobInsight collects crowd-sourced annotations from diverse sources specific to the target city used for the demonstration (i.e. Barcelona), including local guides, geo-social services, and also open directory data from the local government. The data set thoroughly covers the existing places in the neighborhoods and their features. We apply various language processing techniques and clustering algorithms to create a characterization scheme.

A key strength of MobInsight is the use of real telecommunication logs for mobility estimation. The broad adoption of mobile phone allows the mobility estimation to be based on a large sample of people. In addition, the data set mitigates potential sample biases given the almost universal adoption of mobile phones across different socioeconomic and demographic groups. Note that the mobile data analyzed by MobInsight is available for *all mobile phones* in the same mobile network operator, not just for smartphone users or users of a certain mobile app. The demonstration uses the call data records (CDRs) of a mobile operator with the largest market share (~40%) in the target city.

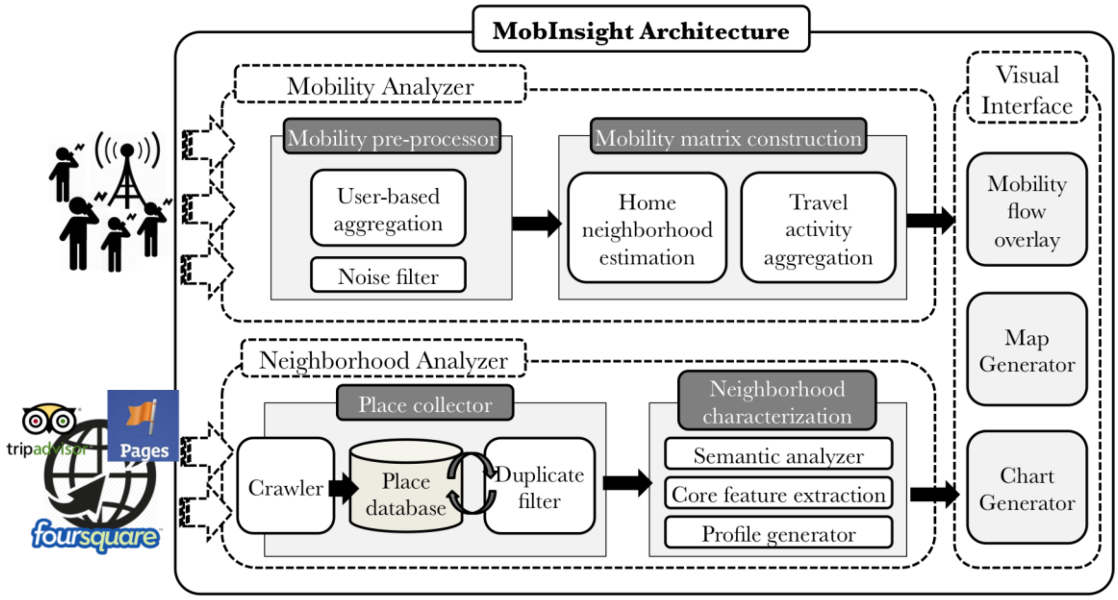


***Fig. 1. MobInsight Architecture Overview***

MobInsight expands recent work in the area of urban informatics. The area is emerging as mobile phones and location-based social services reveal mobility patterns of people, and researchers are applying data-driven techniques to study topics around how people use the urban space. While, many works have explored the topics that are related to the character of neighborhoods [1, 2, 3, 6] or mobility [4,5], few works have looked into the relationship between the two and how they can explain each other.

## II. MobInsight Design

Figure 1 provides an overview of the MobInsight architecture. MobInsight includes two data processing flows in the system: one that transforms mobile phone network logs (CDRs) into a mobility matrix, and the other that collects and aggregates social annotations about places to create neighborhood profiles. The results of the two data processing flows are merged in the visual interface, supporting analysts to associate and explore the two different types of information. Below we describe the major components of MobInsight in detail.

### *A. Mobility Analyzer*

Each entry of the CDRs includes information about the caller, receiver, the timestamp of the call, and the base transceiver station (BTS) used to connect the call which allows a rough approximation of the physical location of the caller and the receiver. From this data, the data processing flow attempts to obtain a large-set of people for whom we can estimate their home neighborhood and collect samples of their travel to other neighborhoods. By aggregating the data of all the people, a mobility matrix that estimates the frequency of travel between the neighborhoods is computed. Given a neighborhood as an instance, the mobility matrix estimates the frequency of trips made by the residents to all other neighborhoods, and conversely the frequency of travel made from other neighborhoods to this neighborhood. For the demonstration, we use 35 million records of calls made during a month in Barcelona as the input to the processing flow. In order to focus on real residents of the city and avoid tourist effects, the data set only includes the calls made by customers who have a fixed contract with the operator, which ensures that the corresponding client has a fixed address in the city. All the records are fully anonymized by hashing the phone numbers and all the results and computations are carried out in an aggregated fashion (aggregated at the neighborhood level).

- Mobility Pre-processor: The preprocessor reads the data and indexes it by the identifier (*i.e.*, hash of the phone number). Once indexed, the identifiers (IDs) with few calls are filtered out to prevent erroneous home estimations.
- Mobility Matrix Construction: First, a simple home estimation algorithm is applied for every ID. The algorithm takes into account both incoming and outgoing calls that take place at night and in the early morning (8pm-8am) during weekdays and at anytime during weekends. It counts the frequency of calls by BTSs and the neighborhood where the most frequently used BTS is located is considered to be the *home* neighborhood of the ID. If the difference in the frequency between the most frequently used BTS and the rest of BTS used by that ID is marginal, the ID is filtered out from the data set. The current demonstration uses 500,000 IDs whose home neighborhood is estimated. As a simple validation step, the Pearson correlation coefficient between the neighborhood population calculated by our method and that of the census of Barcelona [7] is 0.73.

Once the home neighborhood of an ID is identified, all the calls of the ID that are made outside the home neighborhood are considered to be samples that correspond to travelling instances of such ID. In order to construct the mobility matrix, for each neighborhood, we count the samples of residents' travel to all other neighborhoods.

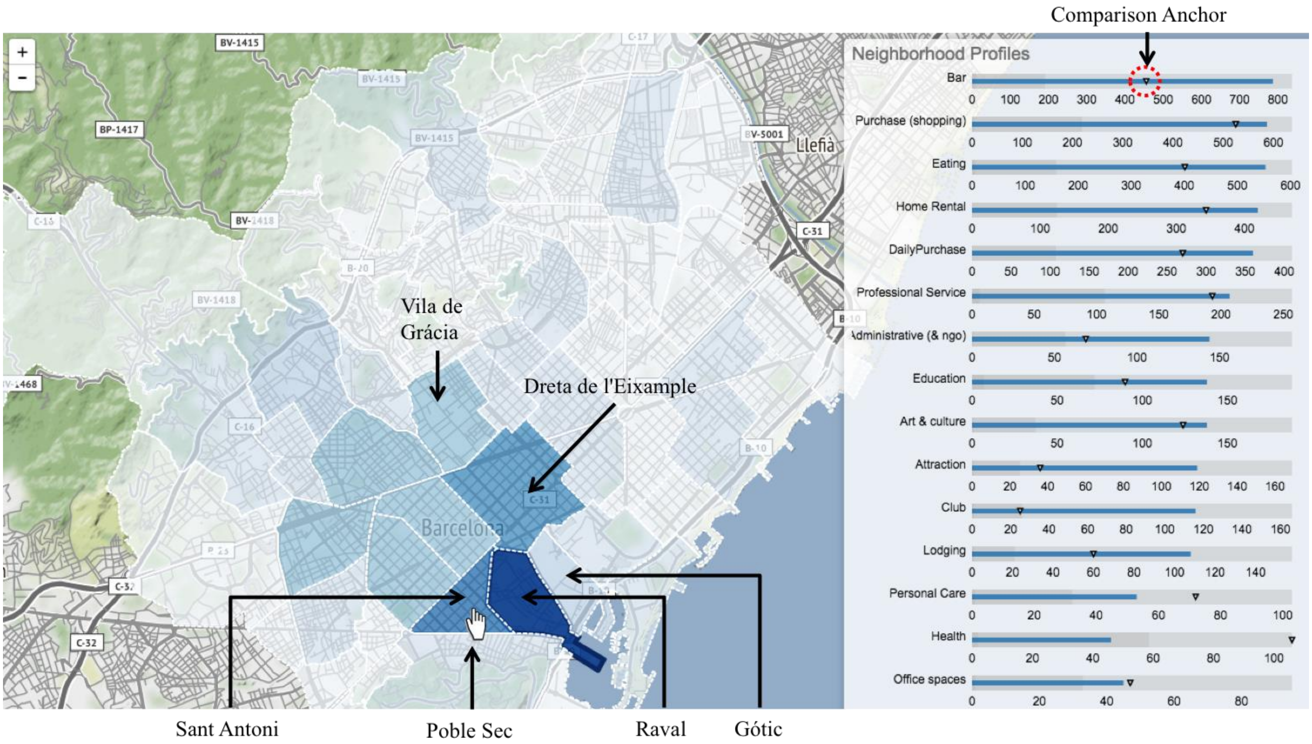


*Fig. 2. MobInsight Interface*

## B. *Neighborhood Analyzer*

The goal of the neighborhood analyzer is to create an easy-to-understand abstraction of all the places that are available in a neighborhood that an analyst can easily make sense of. The intuition behind the abstraction is that the character of a neighborhood is manifested in the types and quantity of different places that are in the neighborhood. For example, a neighborhood with many handicraft shops and art galleries is likely to have a different character than a neighborhood with department stores and chain shops. Thus, the first component of this module, the *place collector*, attempts to comprehensively identify the places in the neighborhoods and collect rich descriptions about them. The main task of the second component is to *analyze the semantics* of the descriptions and extract the *key types of places* that will be used in the characterization. Below is a more detailed explanation of the different modules in the Neighborhood Analyzer:

- *Place Collector:* We build customized crawlers to collect the place information in Barcelona from diverse sources, currently covering 14 sources including geo-social services (*e.g.*, Foursquare, Facebook pages, Google places), local guides (*e.g.*, TripAdvisor), and open directory data from the local government [7]. The collection has more than 200,000 places. The collected information includes the name and the location of the places (address and coordinates), source specific meta-data (*e.g.*, number of check-ins, stars), and reviews left on the places.

  Since the data is collected from multiple sources, duplicates exist across the sources. The duplicates are resolved by merging the places that have overlapping tokens in their names and that are in close proximity.

- *Neighborhood characterization*: We currently use the tags (crowd-sourced keywords about the places) to develop the characterizations. We are currently expanding the tool to consider other information such as ratings and reviews.

  The processing flow attempts to thoroughly identify the different types of places in the city from the tags. It first applies lemmatization and latent semantic analysis to extract the major dimensions that preserve the semantic relatedness between the tags. Then a clustering algorithm is used to obtain an overview of the types of places in the city. The algorithm clusters the places by the semantic similarity of their tags. We use the K-means algorithm and choose the $k$ based on the silhouette score metric. We additionally revise the clustering results manually (*e.g.*, merge clusters that seem redundant) by looking at the most frequent tags of the clusters, and assigning a text label that represents the cluster. As the last step, for each neighborhood, a profile is built by counting its places in the clusters. Figure 2 shows an example profile on the right side with 15 different types of places.

## III. MobInsight Interface and User Experience

MobInsight provides a map-based interface overlaying two types of information: the mobility between neighborhoods and their profiles. Figure 2 shows a screenshot of the interface for Barcelona, depicting the 70 neighborhoods in the city.

The original mobility data is thus a $70 \times 70$ matrix, which is complicated to read and make sense of from the perspective of an analyst. MobInsight supports interactive color-coding to assist the exploration of the data. Upon a click on a neighborhood of interest, the color of the neighborhoods dynamically changes according to the amount of visits made to the selected neighborhood as provided by the mobility matrix. Figure 2 shows the result of selecting Raval, the neighborhood where the conference venue is located. The analyst can easily make interesting observations. For example, the residents of Dreta de l'Eixample and Sant Antoni visit Raval more frequently than those from other neighborhoods. Some neighborhoods that are further above, such as Vila de Gràcia, have more visitors to Raval than Gótic or Poble Sec despite the distance. While not shown in the figure, clicking on the selected neighborhood again changes the color-coding to show the frequent destinations visited by the residents of the selected neighborhood.

Aside from the mobility data, the profile of the selected neighborhood is displayed on the right hand side. A list of bar charts shows the quantity of places of different types.

The key feature of the profile pane that it supports the comparison of different neighborhoods. Upon hovering over another neighborhood, each bar chart dynamically displays an anchor that represents the corresponding quantity of the neighborhood being hovered over. In Figure 2, while the blue bars represent the quantities for Raval, the anchors show the quantities for San Antoni. By reading the gap between the anchors and the blue bars, the analyst can speculate about what people might be visiting for. For example, the bigger gaps are observed for the categories “Bar”, “Club” suggesting that nightlife could be a reason for the visit to Raval from Sant Antoni.

The demonstration will require an installation of a large display or a projector to show the interface. The attendants will be invited to freely use the visual interface.

## IV. Conclusion and Future Work

In this demonstration, we present MobInsight, a tool that provides new perspective to urban mobility through the lens of neighborhood characteristics. We conclude by discussing future works and possible applications of MobInsight.

Our on-going work includes improving the neighborhood characterization and developing techniques for urban mobility prediction. There are several social meta-data fields about the places that are already collected but have not been used in the characterization. We envision a number of improvements as they offer rich hints about the places, such as popularity, quality of places and the different preferences of visitors.

We are also applying machine-learning techniques to see if neighborhood characteristics can predict the amount of mobility between the neighborhoods. The initial results are promising, showing much higher accuracy than the models that use only population and distance. One of the future directions is to see how well the models can generalize to other cities.

MobInsight can potentially assist various quantitative analyses in the context of urban studies and planning. For example, as MobInsight provides a comprehensive and detailed measurement of existing places across neighborhoods, an analyst can have a holistic view of how urban functions are distributed over a large city. The analyst can also explore unknown relationship between the functions, for example, if certain types of places are often co-located.

MobInsight can be also used to explore sociological questions related to urban space. Being aware of different neighborhood characteristics offers insights about what people might be looking for when they travel outside their home neighborhood. Thus, an analyst can explore questions about similarity and difference of demands and services offered in different neighborhoods. In addition, the analyst can further deepen the study by adding other data about neighborhoods, such as socio-economic status, and analyze the relationship with the demand of residents. The result of such an exploration can potentially provide useful implications to urban management and planning.

## References


[1] J. Cranshaw, R. Schwartz, J. I. Hong, and N. Sadeh. The livehoods project: Utilizing social media to understand the dynamics of a city. International AAAI Conference on Weblogs and Social Media, p. 58. 2012.

[2] M. De Nadai, J. Staiano, R. Larcher, N. Sebe, D. Quercia, and B. Lepri. The Death and Life of Great Italian Cities: A Mobile Phone Data Perspective. Proceedings of the 25th International Conference on World Wide Web, pp. 413-423. International World Wide Web Conferences Steering Committee, 2016.

[3] A Noulas, M. Cecilia, E. Frias-Martinez. Exploiting foursquare and cellular data to infer user activity in urban environments. IEEE 14th International Conference on Mobile Data Management, vol. 1, pp. 167-176. IEEE, 2013.

[4] A. P. Masucci, J. Serras, A. Johansson, M. Batty. Gravity versus radiation models: On the importance of scale and heterogeneity in commuting flows. Physical Review E, 88 (2), 2013:022812.

[5] C. Smith, D. Quercia, L. Capra. Finger on the pulse: identifying deprivation using transit flow analysis. Proceedings of the 2013 conference on Computer supported cooperative work (pp. 683-692). ACM. 2013.

[6] J. Yuan, Y. Zheng, and X. Xie. Discovering regions of different functions in a city using human mobility and POIs. Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining, pp. 186-194. ACM, 2012.

[7] OpenDataBCN – Open Data de l'Ajuntament de Barcelona. http://opendata.bcn.cat/opendata/