The City Browser: Utilizing Massive Call Data to Infer City Mobility Dynamics

Fahad Alhasoun
Center for Computational Engineering
Massachusetts Institute of Technology
Cambridge MA, USA
fha@mit.edu

Abdullah Almaatouq
Center for Complex Engineering Systems at KACST and MIT
Riyadh, Saudi Arabia
amaatouq@mit.edu

Kael Greco
Senseable City Lab
Massachusetts Institute of Technology
Cambridge MA, USA
kael@mit.edu

Riccardo Campari
Senseable City Lab
Massachusetts Institute of Technology
Cambridge MA, USA
campari@mit.edu

Anas Alfaris
Center for Complex Engineering Systems at KACST and MIT
Riyadh, Saudi Arabia
anas@mit.edu

Carlo Ratti
Senseable City Lab
Massachusetts Institute of Technology
Cambridge MA, USA
ratti@mit.edu

ABSTRACT

This paper presents the City Browser, a tool developed to analyze the complexities underlying human mobility at the city scale. The tool uses data generated from mobile phones as a proxy to provide several insights with regards to the commuting patterns of the population within the bounds of a city. The three major components of the browser are the data warehouse, modules and algorithm, and the visualization interface. The modules and algorithm component utilizes Call Detail Records (CDRs) stored within the data warehouse to infer mobility patterns that are then communicated through the visualization interface. The modules and algorithm component consists of four modules: the spatial-temporal decomposition module, the home/work capturing module, the community detection module, and the flow estimation module. The visualization interface manages the output of each module to provide a comprehensive view of a city's mobility dynamics over varying time scales. A case study is presented on the city of Riyadh in Saudi Arabia, where the browser was developed to better understand city mobility patterns.

Categories and Subject Descriptors

H.2.8 [**Database Applications**]: Data Mining; H.2.8 [**Database Applications**]: Spatial Databases and GIS; H.4 [**Information Systems Applications**]: Miscellaneous

General Terms

Urban Computing, Data Analytics, City Science

Copyright is held by the author/owner(s).

UrbComp'14, August 24, 2014, New York, NY, USA.

Keywords

Call Detail Records, Mobile Applications, Urban Analysis

1. INTRODUCTION

Cities today house over 50 percent of world's population, consuming 60-80 percent of global energy and emitting almost 75 percent of greenhouse gases [18]. Some have suggested that almost 70 percent of world's population will reside in cities by 2050 [18]. With the rapid urban population growth, cities' infrastructures are being strained to the point of becoming a major hindrance to socioeconomic activity. Left unaddressed, the problem threatens to weigh down the return on investment from public projects being constructed throughout cities and adversely affect the quality of life of all residents.

Understanding the complexities underlying the emerging behaviors of human travel patterns on the city level is essential toward making informed decision-making pertaining to urban transportation infrastructures [2]. Traditional methods of assessing the social demand on transportation are expensive and take longer periods of time to conduct [10, 19, 21]. Such assessments are usually in the form of surveys with considerably small sample sizes compared to the total population of a city. Furthermore, such methods lack the accuracy and resolution in time to provide fine-grained analysis of human travel with precise time resolution.

New road counter technologies such as pressure tubes, inductive loops and other traffic counting techniques allowed for counting travelers with a finer time resolution; however, the drawback is the spatial resolution of such techniques. They are usually highly local and capture activity in a specific point in space that is miniscule with respect to the city as a whole [16]. Therefore, such techniques suffer from an inability to provide a holistic overview of the status of the system. In addition, deploying new traffic counting technologies can be extremely expensive when considering the mega-cities in the world.

An alternative approach toward capturing the social demand is by using data generated from mobile phones to model and understand the behavior of human mobility [21]. Data pertaining to mobile phone usage can be gathered at different levels within the GSM network. Telecom companies usually do not keep track of all the data traffic running across their networks; however, they store certain information for billing purposes and network development. The Call Detail Records, often referred to as CDRs, are one type of information telecom companies keep for billing purposes. Every time a user makes a phone call, sends a text message uses the Internet and even passively when the mobile communicates to the cellular network access points, the mobile network keeps a record of their usage information and location in the CDRs [11]. Therefore, such big data set can be utilized as a proxy to understand the social demand on transportation infrastructures.

The motivation behind developing the browser is derived from the demand of a tool that provides fine-grained analysis of the complexity of human travel within cites. The approach takes advantage of the existing built infrastructures to sense the mobility of people eliminating the financial and temporal burdens of traditional methods. The outcomes of the tool will assist both planners and the public in understanding the complexities of human mobility within their cities.

In this paper we will present the City Mobility Browser, a tool that facilitates a simplified understanding of human mobility across a city. The paper is divided into four sections: Section 2 describes existing methods and approaches, Section 3 presents the methodology of the browser, Section 4 describes the general architecture of the system, Section 5 describes each component of the tool in detail, and Section 7 presents results of the case study of city of Riyadh in Saudi Arabia. The contributions of this paper can be summarized into the following two points:

- We propose an architecture that combines several known techniques for data collection, storage and analysis in one framework in a meaningful context to develop the "City Browser", that can aid in simplifying the complexity of human mobility across a city.
- We examine the usefulness of the system through a case study of Riyadh, Saudi Arabia. The case study contained 100 million real mobile phone activity and demonstrates the process of analyzing massive amount of data and through visualization, distilling the bits into actionable insights.

2. BACKGROUND

Several research activities have been investigating approaches towards modeling and understanding mobility demand within cities. Traditional methods of demand modeling inferred the collective behavior of demand on transportation infrastructures through household or road surveys to gather information about user's behavior. Another approach has been to use theoretical models to estimate the number of trips and their directionality based on land use models. These approaches can be unreliable and can have financial and temporal costs. Today and with the emergence of pervasive technologies around the world, research started investigating human behavior through data gathered from mobile phones [8, 9, 12, 13]. Varying approaches have used the data as a proxy to better understand human mobility. The focus on human mobility ranges from decomposing the data onto the different dimensions to gain insights into behavioral patterns by applying algorithms and processes on models built on the data. Research investigating the dimensionality of the data includes work on utilizing the spatial decomposition of aggregate activity to understand the dynamics of cities and universal patterns of human mobility [3, 9]. On the other hand, researchers have developed techniques to gain more insights from the data by creating algorithms capturing more of the hidden patterns [7, 12, 22]. For example, researchers have been modeling the social network based on the data captured from users' interactions to better understand whether the composition of social communities is correlating with the geographical constraints [17]. Another approach was to capture users' trips from the data set and aggregate trips to get insight on the flows of people around the city towards understanding the dynamics of flows of people [4, 21]. Such understanding can help identify flawed urban planning in cities [23].

3. METHODOLOGY

The objective of the browser is to provide an understanding of the complexity underlying human mobility within a city. The browser will capture the dynamics of the distribution of the population to investigate aspects pertaining to flows of people as well as the structure of the community. Investigating population localization dynamics provides information pertaining to emerging zones with higher population densities; certain dense zones emerge on daily basis like commercial areas on weekdays while others emerge as consequence of events that are not of periodic nature. The browser will investigate whether the formation of periodic dense zones has an influence on the segregating of the population of the city into communities. On the other hand, it will provide information about how the city interacts with events in terms of population commuting flows.

The approach towards simplifying the complexity of human mobility is staged into four steps. Starting with step 1, the browser decomposes population distribution across the spatial dimension on a time resolution of a day capturing the emergence of dense zones (see Subsection 3.1). Step 2 then analyzes each individual in the CDRs to capture their home/work locations (see Subsection 3.2). Step 3 as explained in subsection 3.3 investigates the formation of communities within cities as a result of their home/work choices. Step 4 estimates people flows within the city within a day time scale (see Subsection 3.4).

3.1 Spatial-Temporal Decomposition

The first phase of the methodology decomposes the population over the spatial dimension of the city on the day scale; it will capture time series information of densities of people at every zone with time granularity in minutes. The technique quantifies the magnitudes of mobile user activities within the defined time window, generating time series data for user activity densities for each zone covered by a cell tower. Observing densities with such fine time granularity provides fine grained detail on the emergence of such populated zones by identifying when, where and how fast different dense zones emerge.

3.2 Home/Work Places Capturing

The second phase takes a larger time granularity spanning weeks to capture residential and business areas. The approach towards that is by identifying locations where users spend most of their time during day and night (i.e. home/work locations) across a sufficient time interval. Aggregating the number of users spending most of their times over a particular location captures zones that are emerging as a result of daily routine activities like regular business areas and schools.

3.3 Community Detection

To better understand the influence of where people live and work, this phase investigates the formation of segregated communities based on their home and work locations. The formation of a mobility community within the population indicates that there is a subset of the population traveling within confined bounds of the city and tend not to cross those bounds (i.e. a neighborhood or group of neighborhoods). Such analysis can provide insights on the level of heterogeneity of trips' sources and destinations.

3.4 Flows Estimation

To better understand daily commuting within a city, this phase captures flows within the city through the origin destination estimation algorithm. The algorithm captures trips generated by users around the day and then aggregates the flows of people on a specified time window. The results of the origin destination estimation algorithm will provide information about how dense zones emerge in terms of the source of the population visiting those zones.

4. GENERAL ARCHITECTURE

The general architecture of the browser is composed of three major components; data warehouse, modules and algorithms, and the visualization interface. The data warehouse contains the needed data for the modules and algorithms to produce insights and information visualized through the visualization interface. The general architecture is shown in the figure 1. The data warehouse contains data pertaining to human mobile phone usage as well as GIS information of the city and traffic counts. There are four major modules residing within the modules and algorithms component that are spatial-temporal decomposition module, home/work capturing module, community detection module and flows estimation module. Finally the visualization interface takes the results produced by the modules and algorithms together with GIS information of the city to provide a comprehensive dynamic view of human mobility within a city.

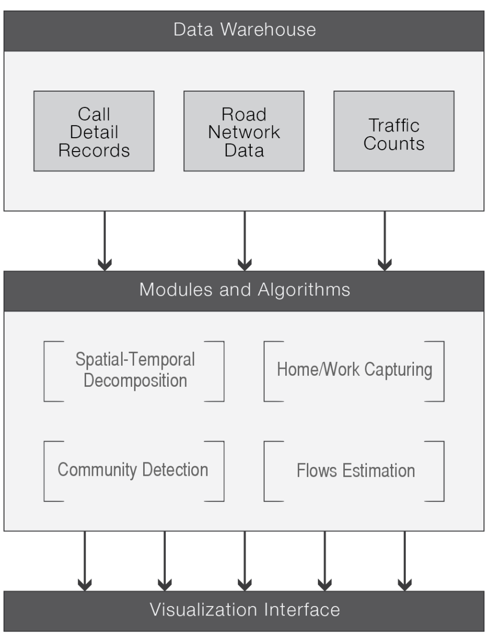


Figure 1: City Browser general architecture

"details of the implementation of the architecture"

5. COMPONENTS

The City Browser is decomposed into components following the general architecture described in section 4. This section will provide the details of each component. The breakdown of the browser into components is to allow for a more scalable, modular and simpler architecture for development. Each of the components is describes below.

5.1 Data Warehouse

The data warehouse houses several datasets containing information of the structure of the city as well as the dynamics of it. It contains a geospatial database of the city including the lookup table of the locations of the cell towers for the purpose of mapping mobile phone activity to locations. In addition, it contains information of the time series mobile phone usage data as well as traffic counts.

The major part of the data warehouse is mobile phone billing information, also known as Call Detail Records (CDRs), which are records that telecom companies usually keep for the purpose of generating bills for customers. The CDRs are generated by mobile switching centers (MSCs) within GSM networks and go through several processing methods to be usable by telecom providers. The CDRs are finally structured in a table-like format, withholding information about phone activity details. Each entry in the CDRs table is a record representing an activity generated by a user. Every time a user makes a phone call, sends a text message or accesses the Internet, the CDRs keeps a record of the cell tower that was used to facilitate activity. In addition, the data warehouse contains a lookup table for cell tower geospatial information where each cell tower is mapped to its coordinates (i.e. latitude and longitude). Each record within the databases is referred to as an activity and is described by time t, user u and cell tower c and represented as $a(t, c, u)$. For each user, the dataset contains a series of activities captured and are represented in this paper as:

$$A_u = \{a_0, a_1, a_2, ...a_n | u = u_{a_0} = u_{a_1} = u_{a_2} = ... = u_{a_n}\}$$

where a_0 is an activity record and u_{a_0} is the user generating activity a_0. The data warehouse also contains traffic volume counts at specific points on the road network. Traffic counts are usually taken for a defined period of time where pressure tubes are placed on certain links to count the number of times vehicles pass across them. Furthermore, information about the geometry of the road network is housed within the data warehouse as a spatial database. The road network spatial database contains information about the geometry of roads such as number of lanes, category, length and speed limit.

5.2 Modules and Algorithms

The Modules and Algorithms component is composed of four components: spatial-temporal decomposition module, home/work capturing module, flow estimation module, and community detection module. Each of the components is described below.

5.2.1 Spatial-Temporal Decomposition Module

The first step toward understanding the dynamics of a city on the day scale is to look at the dynamics of population densities across the city through aggregate user activities for each cell tower. This module breaks down the total activities of users on both the spatial and temporal dimensions. A similar approach was developed in [4]. For each cell tower within the city, the module generates a time series data for activity levels for a specified time granularity Δt. To capture the collective behavior of the population across the city, the module captures the aggregate activity level of users at every cell tower c_i within the city. The aggregate phone activity level denoted $AL(c_i, \Delta t)$ at cell tower c_i for a time window Δt is computed as follows from the dataset.

$$AL(c_i, \Delta t) = \sum_{c \in c_i, t \in \Delta t} a(c, t, u)$$

Where $a(c, t, u)$ is an activity generated through cell tower c at time t. Each time series data for every location c_i gives insights on the nature of the zone where the cell tower resides in terms of its use. For example, work areas within cities are expected to have a higher density of activity during work hours compared to residential areas. The module also provides insight into collective population behavioral characteristics showing when the city becomes alive in the morning. It also captures information on how users are interacting with events in terms of localization or behavior of service usage. The objective of developing this module is to provide a holistic overview of the change in population densities across space and time.

5.2.2 Capturing Home/Work Places Module

Expanding the time interval of the analysis, this module captures work zones as well as residentail zones. This is essentially capturing places where the majority of daytime calls are as a proxy to work locations. First, we segregate activity records on two time windows to capture most visited zones at daytime versus nighttime for a particular user u. Activities that would hold potential work locations are separated in a set as:

$$\begin{aligned} day_u = &\{a_0, a_1, a_2, ...a_n | u = u_{a_0} = u_{a_1} = u_{a_2} = ... = u_{a_n} \\ &\wedge t_{a_i} \in daytime\} \end{aligned}$$

Where a_0 is an activity record, u_{a_0} is the user generating activity a_0 and t_{a_i} is time tag of activity a_i. Similarly, $night_u$ is obtained with the same logic for nighttime activity. Then, $work_u$ location for user u is chosen to be the most occuring location in day_u and the same applies to $home_u$ as it is chosen to be the most occuring location in $night_u$.

After determining the $work_u$ and $home_u$ for each user. The aggregation of the resulting zones where users spend most of their times during the day and night identifies dense zones that pertain to business/residential areas since the module considers larger time granularity for the analysis. Thus, this module quantifies the extent to which a zone is considered as residentail/business zone.

5.2.3 Community Detection Module

Following on the output of section 5.2.2, this module will investigate whether there are groups within the population forming communities that have similar home and work locations. The module begins with the city-wide network of connected zones $G(N, E)$ where N is the set of cell towers within the city representing the zones and E is composed of weighted directed edges defined as the number of users who have a particular home/work pair, respectively, in the zones corresponding to the starting and terminating nodes. The adjacency matrix A of the discussed network is as follows:

$$A = \begin{pmatrix} w_{0,0} & w_{0,1} & \cdots \\ w_{1,0} & w_{1,1} & \cdots \\ \vdots & \vdots & \cdots \\ w_{m,1} & w_{m,2} & \ddots \end{pmatrix}$$

Where $w_{0,1}$ is the number of users having their $home_u$ as c_0 and $work_u$ as c_1. The algorithm then uses a modularity optimization scheme, such that sets of nodes are clustered in a way that minimizes internal arc disruption [5, 14]. Each resulting community represents an area where a large fraction of users are mostly located during the day and night.

Modularity is a standard objective function in the field of community detection; it measures how well a partition of network nodes into communities reflects the characteristics of the underlying network (in our case the commuting flow among zones). The rationale behind modularity is that a group of nodes with connections mostly directed towards its own members represent a community with higher modularity while a set of nodes with intra-community connections is what we would expect by randomly rewiring all the links.

Communities resulting from modularity optimization of telecommunication data have been empirically shown to be representative of the actual social and administrative boundaries at the level of whole countries [7].

In the case of a city, we went further and studied communities at the level of the neighborhood. The interesting results we obtained are discussed in Section 7.

5.2.4 Flows Estimation Module

To capture the directionality and mobility of the population across the city, the browser houses an algorithm that provides information about the collective behavior of human mobility through mining mobile phone activity. The module of estimating the aggregate flows of people across the city from the CDRs is a three step algorithm that has the CDRs as inputs and the aggregation of flows of people between locations at every time window Δt as its result (i.e. Origin Destination matrix). A similar approach was developed in [4]. The module starts by arranging data on a user level and considering each of their displacements as a potential trip. After that, the resulting potential trips go through a filtration process that filters out noise in the data from the potential trips generated. Finally the last step aggregates the resulting trips on both the spatial and temporal dimensions to generate an origin-destination matrix based on the provided time slice of interest.

The first step in the algorithm looks at phone activities on a user level and gathers all activities generated for each user sorted in time as follows.

$$\begin{aligned} A_u = &\{a_0, a_1, a_2, ...a_n | u = u_{a_0} = u_{a_1} = u_{a_2} = ... = u_{a_n} \\ &\wedge t_{a_0} < t_{a_1} < t_{a_2} .... t_{a_n}\} \end{aligned}$$

Where A_u is the set of all activities generated by user u, u_{a_i} is the user generating the activity a_i and t_{a_i} is the time tag of activity a_i. Every consecutive records belonging to the same user are merged into pairs of location records with their associated times representing a potential displacement of the user. The set of displacements of a user are represented as given by:

$$D_u = \{(c_{a_i}, c_{a_{i+1}}, t_{a_i}, t_{a_{i+2}}) \mid a_0, a_1, \in A_u\}$$

Where D_u is the set of all potential displacements of user u, c_{a_i} is the cell tower facilitating the activity a_i, t_{a_i} is the time tag of activity a_i and u_{a_i} is the user generating activity a_i. The set of potential displacement considers each successive user activity a potential trip though this includes noisy data such as users who did not change their locations between the successive activities but where nevertheless served by different nearby cell towers, a phenomena referred to as localization error. In order to capture user trips in which a displacement actually occurs, we apply further filtering on the set of potential displacements D_u. The goal of the filtering process is to eliminate all captured pairs of location records that are considered as noise in terms of trip-capturing. The filtration process eliminates all records that are considered as localization error, have very long time intervals or no movement detected. Entries in the data that corresponds to localization error are filtered out by eliminating all trips that are less than a specified distance of the maximum distance between any neighboring cell towers within an urban setting. Given any two neighboring cell towers that c_{a_i} and c_{a_j} , each element within D_u must satisfy the below predicate.

$$distance(c_{a_i}, c_{a_{i+1}}) > max[distance(c_{a_i}, c_{a_j})]$$

Where $disance(c_{a_i}, c_{a_{i+1}})$ is the distance between the towers c_{a_i} and $c_{a_{i+1}}$. The filter eliminates potential displacements having a distance larger than that of the maximum distance between any two neighboring towers in the city. In addition, each pair of records satisfy $t_{a_{i+1}} - t_{a_i} > \alpha$, where t_{a_i} is the time tag of activity a_i. That is a time difference between consecutive activity records being more than a threshold is filtered out of the set of displacements D_u for the purpose of reducing the uncertainty in capturing the actual departure and arrival times for trips.

The result of the filtering process is the set of displacements $\bar{D}_u$ containing all pairs of locations where movement was detected and reasonable time duration for the trip was captured. After that, the final step towards the generation of OD matrices is to aggregate the trips according to the specified time slice into the origin destination matrix given by:

$$OD(\Delta t) = \begin{pmatrix} 0 & T_{0,1} & T_{0,2} & \cdots \\ T_{1,0} & 0 & T_{1,2} & \cdots \\ T_{2,0} & T_{2,1} & 0 & \cdots \\ \vdots & \vdots & \ddots & \vdots \end{pmatrix}$$

Where each element $T_{i,j}$ gives the number of trips captured between c_i to c_j during the time slice Δt. The value of $T_{i,j}$ is computed by:

$$T_{i,j}(\Delta t) = \sum \bar{D}_u(c_{a_n}, c_{a_{n+1}}, t_{a_n}, t_{a_{n+1}})$$

Where $c_{a_n} \in i,\ c_{a_{n+1}} \in j$ and $t_{a_{n+1}} - t_{a_n} \in \Delta t$. Thus, $T_{i,j}$ quantifies the flows from zone i to zone j during the time window Δt.

6. VISUALIZATION INTERFACE

The visualization component shows the results of the modules and algorithms on two time scales depending on the nature of their outputs. It will visualize population density distribution and major flows of people across the city dynamically over the span of a day while on longer time scales it will show a static map of the communities forming around the analysis of dense zones.

The visualization will start by showing the spatial-temporal decomposition of the population over the scale of a day. A dynamic visualization with time granularity of 15-minutes will capture population density variations across the day and night. The browser shows mobile activity over a dynamic period of time broken up into 15-minute intervals as shown in figure 3. This visualization presents a rotatable, scalable map onto which a shifting, three-dimensional grid is superimposed to show locational agglomerations of cellphone activity. Grid sectors will rise and fall, and brighten and fade as people move across the city using their mobile devices.

On the same scale of a day, the visualization components shows the directionality of human mobility through the output of flow estimation module as well as the car counts stored in the data set. Major flows within the city showing the aggregate behavior of commuting around the day are visualized with a time window of 15-minutes. The component visualizes the generation of trips on each time slice by as an arc that rose from originating to terminating cell tower. As shown in figure 6, each arc embodies a variable number of trips, and to illustrate this we altered its thickness and height in correspondence to the intensity of activity along that route (on a logarithmic scale). The arcs are drawn over the same city base geography, on top of the social interaction mesh from above, in an effort to reveal unseen connections between the two results. In addition, car counts were built into the visualization as half-spheres placed at their respective intersections. Each sphere changes shape and color at an hourly rhythm in line with the measured volume.

On the longer time scale and towards visualizing the output of the community detection module, the visualization interface overlays the community network over the spatial dimension of the city to show if there are correlations between the formation of communities and the urban fabric of the city. Nodes represent zones of the city and arcs represent groups of people spending most of their times across the day/night between connected nodes. The community detection module provides the set of nodes that belong to the same community. To visualize the output of the community detection algorithm, nodes belonging to the same community are colored with the same color as shown in figure 5. Thus, areas where sub communities spend most of their time during the day and night are bounded within zones of the same color.

7. CASE STUDY

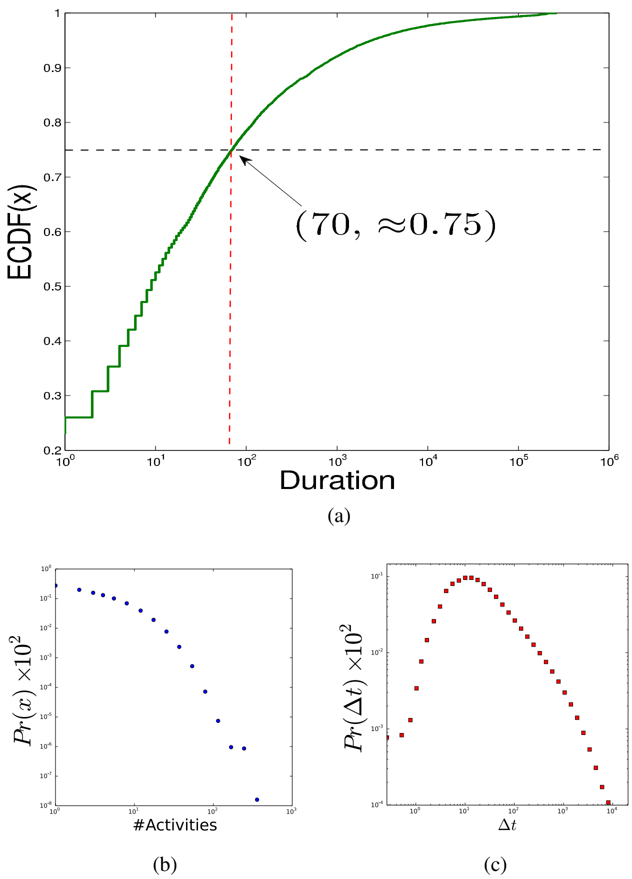
Over the past decade, Saudi Arabia has taken strong steps towards developing a diversified economy. Specifically on enhancing its Information and Communication Technology (ICT) infrastructure [1]. Today, Saudi Arabia has one of the highest Internet penetration percentages in the gulf area with current penetration at 14.7 million. It is ranked among the highest countries worldwide in mobile penetration rates with 188% of the population possess mobile phones [6]. The high penetration rate of mobile in Saudi Arabia make it an ideal candidate for utilizing the Call Detail Records (CDRs) as in situ sensors for human mobility.

The City Browser was implemented for the Urban Transportation System (UTS), a system developed to provide city planners with insights with regards to the mobility of the population. The project started with gathering information related to the structure of the city as well as the dynamics of the population. The data gathered includes Records CDRs spanning a period of the month of December, a spatial database of the road network of Riyadh city and traffic counts data on different points within the city. Currently, the data is housed within the data warehouse where several modules and algorithms are using it to generate insights on the dynamics of the city.

7.1 Data Description

Our dataset consists of one full month of records for the entire country of Saudi Arabia, with 3 billion mobile activities to over 10 thousands unique cell towers, provided by a single carrier. Each record contains an anonymized user ID, the type of activity (i.e., SMS, MMS, call, data etc), the cell tower facilitating the service, duration if its a phone call, and time stamp of the activity. Each cell tower id is spatially mapped to its latitude and longitude. For privacy concerns, user id information were completely anonymized at the telecom operator side.

Previous studies [9, 15] have shown that human communication patterns are highly heterogeneous; where some users use their mobile phone much more frequently than others. The characteristics



(a) (b) (c)

Figure 2: Communication patterns in the CDRs Dataset. Fig 2a shows the Empirical Cumulative Distribution Function (ECDF) of the activities duration. We find that almost 75% of the users conduct activities that last for 70 seconds or less. Fig 2b shows the statistical distribution of the number of communication records generated by the users for a single day. Fig 2c shows the inter-event time distribution $Pr(\Delta t)$ of calling activity, where Δt is the time elapsed between consecutive communication records (outgoing phone calls and SMS) for the same user.

of the dynamics of individual communication activity obtained in Fig 2 supports such hypothesis.

7.2 City Spatial-temporal Decomposition

The first step towards understanding the data in the city of Riyadh is to decompose cellular activity on the spatial and temporal dimensions. The visualization in figure 3 shows cellular activity through color, transparency, and height (in logarithmic scale) gridded across the metropolitan expanse of Riyadh. As opposed to seeing the cell towers as discrete points in the city, we show network traffic interpolated over a 100 by 100 grid. In this sense, each grid cell is assigned an intensity based on its distance to surrounding antennas and their activity levels using a Gaussian smoothing function. The temporal activity is interpolated in a similar manner, showing smooth transitions between each time-slice in the dataset.The city's downtown core quickly becomes clouded in smog of network activity early in the morning that hangs over region for the entire day. Clear sub centers emerge that follow construction density, and these sub-centers appear to be partitioned by the roadway network itself.

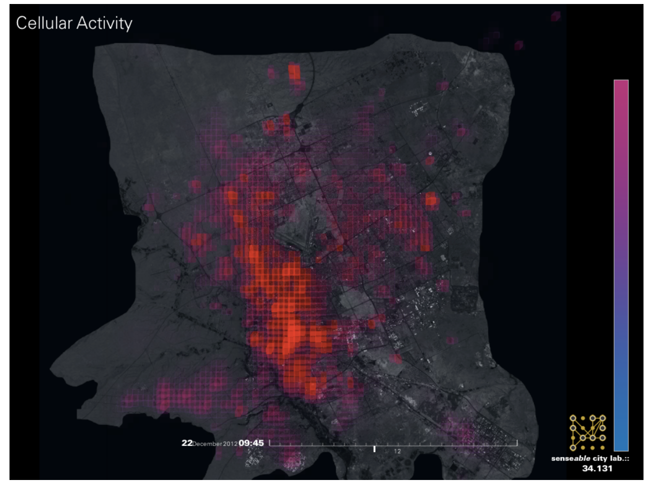


Figure 3: Spatial-Temporal Decomposition out for a single time slice. The figure demonstrates the time-cumulative spatial mobile activity conducted between 9:45am to 10:00am.

The city's shifting activity profile also highlights a rich temporal signature of communication that is all Riyadh's own. Watching the oscillations of the activity landscape, we see that Riyadh comes alive at around 6:15am. We also see strong regional delineation: the residential neighborhoods to the southwest and northeast of the downtown core come alive well before the rest of the city, and experience the strongest inter-hour fluctuations throughout the course of the day. Finally, we see some peculiar discontinuities in aggregate talk throughout the day almost as if all phone traffic was suddenly halved at strange intervals.

7.3 Capturing Home/Work Places

A fundamental quality of mobility behavior is to analyze the emergence of zones with higher densities along a wider time granularity to understand the distribution of residential and business zones. Expanding our time intervals to capture broader day and night variation we can begin to differentiate dense business areas and schools versus dense residential neighborhoods.

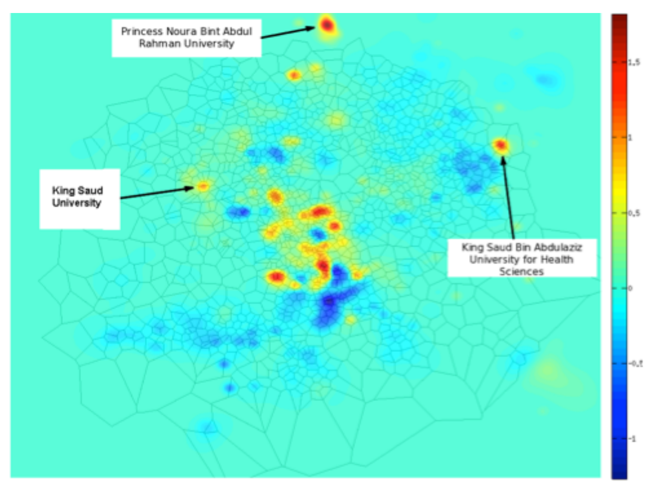


Figure 4: Dense work zones during the day versus home locations during the night. We observe high day-densities at the periphery where major universities are located.

The map in Figure 4 highlights the discrepancy between the purely day zones shifting towards the red color and the purely night dense zones shifting towards blue color, showing some mono-centrically clustered day hotspots that follow the overall spatial logic of the city. At the periphery we also see a number of universities show up strongly as day locations. Lastly, we see high agglomerations of residences to both the south and east of the city, with smaller pockets scattered throughout.

7.4 Detecting Mobility Communities

The work/home dense zones visualizations shown in section 7.3 point to an organizational logic of the city. Conceptualizing the totality of day/night commutes as a city-wide mobility network, we can conceivably break this network into sub-communities by applying a regional delineation algorithm.

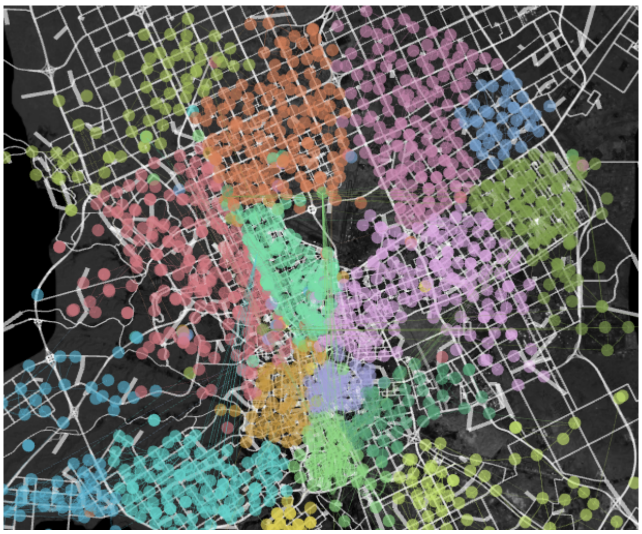


Figure 5: Community Detection Module results plotted by Latitude and Longitude on the map of Riyadh. We find support to the commonly held belief that heavily trafficked streets, on many levels, are instruments of segregation and control.

By overlaying the results of the community detection module on geography of the city (see Figure 5), a number of interesting relationships are revealed between the detected communities and the built form of the city. Most strikingly, the resulting clusters closely correlate to the main arterials of city's roadway infrastructure. Mobility communities seem to be partitioned by the street network itself, underscoring the city's dependence on highway infrastructure, while also supporting the commonly held belief that heavily trafficked streets, on many levels, are instruments of segregation and control, or, perhaps more optimistically: good streets make good neighbors.

7.5 Flow Estimation

The approach toward understanding flows that contribute dense-zone emergence on smaller time granularity unveils rich information pertaining to the sources of dense zones as well as the distribution of flow over time. By collecting and filtering each user's mobile activity as sequence of cell tower locations and then aggregating collective users' trips, we are able to estimate flows in terms of origins and destinations of trips. We've observed that these estimated flows contributed to the emergence of high density zones in the city of Riyadh; however this approach includes the added benefit of capturing travel demand at highly dynamic time slices ranging from seasonal variations to hourly fluctuations. Such a high temporal resolution has the potential to transform our understanding of urban mobility [20].

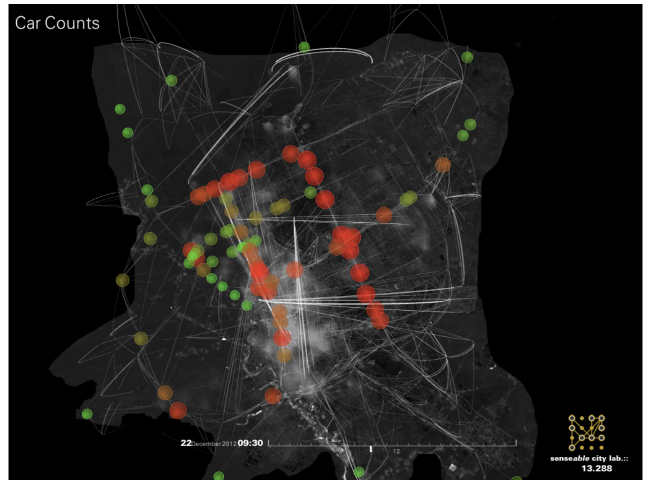


Figure 6: The extracted Origin Destination (OD) matrix across Riyadh at the time slice of 9:30-9:45am. The height of the line corresponds to the number of trips between a specific OD.

The resulting dynamic maps held a striking similarity to the local intuition of vehicular flows across the city (see Figure 6). Overall flows correspond quite closely to the underlying street network. Most notably, Figure 6 shows intense activity along the city's main arterials; King Fahd Road and the Northern and Eastern Ring roads. This agrees with the local community's subjective understanding of commute patterns across the city. But to further validate our results, we compared them against the best ground-truth measurements of roadway activity: car count volumes captured by pressure-tube sensors placed at multiple intersection across the city.

8. SUMMARY AND FUTURE WORK

In this paper, we have presented a new tool addressing the complexity of city human mobility and showed its application to the city of Riyadh the capital of Saudi Arabia through the UTS project. The project developed the Riyadh Mobility Browser by implementing several modules that mined data generated from mobile phones to provide a coherent understanding of the dynamics of the interaction between its social structure and transportation infrastructures. At the current stage, the browser is built to work with historical data and thus would provide an after-the-fact analysis and does not allow for the parsing and analysis of the data in real time. A potential future work would be investigating the possibility of enabling the browser to parse such big data in real time through establishing a live connection of data feed with GSM network operators.

The city mobility browser synthesizes and extends existing algorithms to provide a holistic decomposition of the complexity of mobility across multiple dimensions. Although the browser captures the dynamics of the demand on transportation, it does not map the demand over the road network of the city.

We also acknowledge that some of the explanations and conclusions proposed in this work might lack rigorous validations and this is due to the nature of the CDRs where it lacks sufficient granularity in space and time. Spatially, the data is mapped to the locations of cell towers and not the exact locations of users and therefore the coordinates of cell towers are used as a proxy to the exact locations of users. Temporally, users have a bursty phone usage behavior where activities are clustered around different times of the day rather than spread out around the day to enable a more comprehensive understanding of mobility in this case. However, we believe that our analysis of human mobility can describe well the current trends and phenomenon of human mobility and can be leveraged in planning the city and transportation operations.

The visualizations provided by the tool give a dynamic qualitative understanding of the spatial attributes of the city as well as its population directionality across different times of the day. The city mobility browser is envisioned to be a tool that can provide planners, engineers and the public with an easy to understand analysis while capturing fine grained details about the city. Future work could also enable the visualization interface to provide quantitative analysis and a better understanding of emerging patterns.

9. REFERENCES

[1] A. Almaatouq, F. Alhasoun, R. Campari, and A. Alfaris. The influence of social norms on synchronous versus asynchronous communication technologies. In *Proceedings of the 1st ACM International Workshop on Personal Data Meets Distributed Multimedia*, PDM '13, pages 39–42, New York, NY, USA, 2013. ACM.

[2] M. Batty, K. Axhausen, F. Giannotti, A. Pozdnoukhov, A. Bazzani, M. Wachowicz, G. Ouzounis, and Y. Portugali. Smart cities of the future. *The European Physical Journal Special Topics*, 214(1):481–518, 2012.

[3] D. Brockmann, L. Hufnagel, and T. Geisel. The scaling laws of human travel. *Nature*, 439(7075):462–465, Jan. 2006.

[4] F. Calabrese, M. Colonna, P. Lovisolo, D. Parata, and C. Ratti. Real-time urban monitoring using cell phones: A case study in rome. *IEEE Transactions on Intelligent Transportation Systems*, 12(1):141–151, 2011.

[5] A. Clauset, M. E. J. Newman, and C. Moore. Finding community structure in very large networks. *Physical Review E*, 70:066111, 2004.

[6] Communications and I. Commission. Ict indicators, q2-2012, 2012.

[7] Y.-A. de Montjoye, C. A. Hidalgo, M. Verleysen, and V. D. Blondel. Unique in the Crowd: The privacy bounds of human mobility. *Scientific Reports*, 3, Mar. 2013.

[8] N. Eagle and A. (Sandy) Pentland. Reality mining: Sensing complex social systems. *Personal Ubiquitous Comput.*, 10(4):255–268, Mar. 2006.

[9] M. C. Gonzalez, C. A. Hidalgo, and A.-L. Barabasi. Understanding individual human mobility patterns. *Nature*, 453(7196):779–782, June 2008.

[10] J. Herrera, D. Work, R. Herring, X. Ban, Q. Jacobson, and A. Bayen. Evaluation of traffic data obtained via GPS-enabled mobile phones: The Mobile Century field experiment. *Transportation Research Part C*, 18(4):568–583, August 2010.

[11] S. Jiang, G. A. Fiore, Y. Yang, J. Ferreira, Jr., E. Frazzoli, and M. C. González. A review of urban computing for mobile phone traces: Current methods, challenges and opportunities. In *Proceedings of the 2Nd ACM SIGKDD International Workshop on Urban Computing*, UrbComp '13, pages 2:1–2:9, New York, NY, USA, 2013. ACM.

[12] S. Jiang, J. F. Jr, and M. González. Discovering urban spatial-temporal structure from human activity patterns. In *Proceedings of the ACM SIGKDD International Workshop on Urban Computing*, pages 95–102, 2012.

[13] J. R. Kwapisz, G. M. Weiss, and S. A. Moore. Activity recognition using cell phone accelerometers. *SIGKDD Explor. Newsl.*, 12(2):74–82, Mar. 2011.

[14] M. E. Newman. Modularity and community structure in networks. *Proc Natl Acad Sci U S A*, 103(23):8577–8582, June 2006.

[15] J. Onnela, J. Saramaki, J. Hyvonen, G. Szabo, D. Lazer, K. Kaski, J. Kertesz, and A. L. Barabasi. Structure and tie strengths in mobile communication networks. *Proc. Natl. Acad. Sci. USA*, 104(18):7332–7336, 2007.

[16] S. Peeta and P. a. Zhang. Counting device selection and reliability: Synthesis study. Technical report, Purdue University.

[17] C. Ratti, S. Sobolevsky, F. Calabrese, C. Andris, J. Reades, M. Martino, R. Claxton, and S. H. Strogatz. Redrawing the map of great britain from a network of human interactions. *PLoS ONE*, 5(12):e14248, 12 2010.

[18] P. Rode and R. Burdett. Cities: investing in energy and resource efficiency. In *United Nations Envrionment Programme, (corp. ed.) Towards a Green Economy: Pathways to Sustainable Development and Poverty Eradication*, pages 453–492. United Nations Environment Programme, 2011.

[19] W. Shen and L. Wynter. Real-time traffic prediction using GPS data with low sampling rates: A hybrid approach. In *91st Transportation Research Board Annual Meeting*, number 12-1692, Washington, D.C., January 2012.

[20] J. L. Toole, S. Colak, F. Alhasoun, A. Evsukoff, and M. C. Gonzalez. The path most travelled: Mining road usage patterns from massive call data. *arXiv preprint arXiv:1403.0636*, 2014.

[21] P. Wang, T. Hunter, A. M. Bayen, K. Schechtner, and M. C. González. Understanding Road Usage Patterns in Urban Areas. *Scientific Reports*, 2, Dec. 2012.

[22] J. Yuan, Y. Zheng, and X. Xie. Discovering regions of different functions in a city using human mobility and pois. In *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 186–194. ACM, 2012.

[23] Y. Zheng, Y. Liu, J. Yuan, and X. Xie. Urban computing with taxicabs. In *Proceedings of the 13th international conference on Ubiquitous computing*, pages 89–98. ACM, 2011.