Urban Attractors: Discovering Patterns in Regions of Attraction in Cities

May Alhazzani[1], Fahad Alhasoun[2], Zeyad Alawwad[1], Marta C. González[2,3],

1 Center for Complex Engineering Systems, King Abdulaziz City for Science and Technology, Riyadh, Saudi Arabia
2 Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139.
3 Center for Advanced Urbanism, Massachusetts Institute of Technology, Cambridge, MA 02139.

* martag@mit.edu

Abstract

Understanding the dynamics by which urban areas attract visitors is significant for urban development in cities. In addition, identifying services that relate to highly attractive districts is useful to make policies regarding the placement of such places. Thus, we present a framework for classifying districts in cities by their attractiveness to visitors, and relating Points of Interests (POIs) types to districts' attraction patterns. We used Origin-Destination matrices (ODs) mined from cell phone data that capture the flow of trips between each pair of places in Riyadh, Saudi Arabia. We define the attraction profile for a place based on three main statistical features: The amount of visitors a place received, the distribution of distance traveled by visitors on the road network, and the spatial spread of where visitors come from. We use a hierarchical clustering algorithm to classify all places in the city by their features of attraction. We detect three types of Urban Attractors in Riyadh during the morning period: *G*lobal which are significant places in the city, *D*owntown which the central business district and Residential attractors. In addition, we uncover what makes these places different in terms of attraction patterns. We used a statistical significance testing approach to rigorously quantify the relationship between Points of Interests (POIs) types (services) and the 3 patterns of Urban Attractors we detected. The proposed framework can be used for detecting the attraction patterns given by type of services related to each pattern. This is a critical piece of information to inform trip distribution models.

Introduction

Understanding how different places in the city influence human mobility is significant for urban and transportation planning. A pressing need, in complex and congested cities is maintaining a robust transportation infrastructure. Understanding the patterns by which places in the city attract visitors is essential for planning and modifying the transportation system. Specifically, new data sources can help us to better predict the relation between population and trips attraction. This issue is of a particular significance to the city of Riyadh, Saudi Arabia where the largest metro project is being developed and promised to start running in 2019 [1,2]. Moreover, understanding how different types of places affects the flow of trips in the city differently helps to inform

decisions and policies related to placing and modifying concentration of services. For example, how an industrial area influences the flow of trips during different times of the day; or where to place new business stores for higher profitability [3].

Today, with the ubiquity and pervasiveness of technology, data generated from mobile phones enable data analysts to better understand the behavior of individuals across many dimensions including their mobility patterns [4, 5]. An interesting area is how patterns of human mobility are affected by different places in the city. The standard approach on categorizing urban areas classifies regions by their functionality and land use (i.e. commercial, educational, ...etc.). Recent works consider the human mobility aspects to classify regions. For example, Yuan *et al.* [6] proposed a topic modeling approach to classify districts into functional zones according to people's socioeconomic activities mined from taxi and public transport traces and points of interests (POIs) data. Pan *et al.* [7] proposed a land use classification approach based on the social functions of districts also analyzed from GPS taxi traces where districts witness change of land use class dynamically. Toole *et al.* [8] analyzed cell phone data to test cell phone activity patterns to classify for land use types. Less is known about how phone data can help to classify urban regions based on how attractive they are to different origins.

In this work, we present a novel computational framework for classifying urban places by their attraction patterns. We define attraction profiles in terms of statistical features of incoming trips on a given time window. Different places in the city attract visitors differently. Some places like universities and hospitals attract a large amount of visitors who come from all over the city and travel long distances to go there. On the other hand, some places in the city that provide local services such as restaurants, schools, and small clinics only attract few people from nearby areas. We aim to automatically identify patterns of attraction of places based on three main dimensions: how many visitors a place receives, where visitors are coming from, and how long visitors are traveling to reach that place. Secondly, we further show what makes a region attraction behavior. To accomplish that, we used statistical significance testing to automatically relate the decomposition of POI types (services) and the discovered attraction patterns. This information is useful to relate type of businesses and attraction profiles.

The main contribution to this work is as follows:

- We present a computational framework for detecting attraction patterns and further relating POI types to each pattern of attraction.

- We classify attraction patterns via the spatial dispersion of trip origins, the distribution of distances traveled by visitors through the road network and the total number of trips.

- We quantify the significance of POI types in a region using a statistical significance testing approach, which performs well in the context of phone and POI data.

Related work

Multiple studies used human mobility behavior to classify urban areas. A recent study investigated the relationship between land use and mobility [9]. The authors showed that purposes of people's trips are strongly correlated with the land use of the trip's origin and destination. Recently, the availability of dynamic sources of data allowed for dynamic segmentation of the city according to human mobility behavior. Some studies combined human mobility with land use or POIs data to segment districts in urban areas according to their functions or use. The type of data used to capture human

mobility behavior varies between individual GPS traces [10,11], taxi pick up/drop off locations as in [7,12] , Call Detail Records (CDRs) as in [2,8], social media check ins as in [13–15], and bus smart card data as in [16]. This work differs from previous studies, by being the first to classify the urban regions through attraction profiles.

Survey travel data has been used to detect the centers (significant places) of a city [17,18]. A recent study proposed a method for measuring the centrality of locations that incorporates the number of people attracted to the location and the diversity of activities in which visitors engage [17]. The proposed method was tested on survey travel data in Singapore to identify the functional centers and track their significance over time. A similar approach focused on analyzing the aggregate behavior of the population to predicted highly attractive events such as the times square during new years count down in New York [19]. Our method is based on validated Origin Destination (ODs) matrices mined from massive cell phone data that captures human mobility. More significantly, our approach incorporate not just the amount of people a place attracts, but also on where do they come from and the road distance they traveled.

Network analysis methods were used to detect hotspots based on flow patterns between locations [2,20]. A recent paper [2] used ODs matrices extracted from cell phone data to identify the signature of mobility behavior as 4 main types of movements within the city: between hotspots, to hotspots, originating at hotspots and random flows. They showed how different cities have different mobility signatures. Additionally, a recent study used Taxi drop off/pick up traces in Shanghai to create a network of flow between places. They applied community detection to extract sub regions and analyze the interaction between and within sub regions. They found that urban sub-regions have larger internal interactions, while suburban centers are more significant on local traffic.This work made the breakdown of flow patterns instead of the impact of the place in attracting visitors, which is our aim in this paper. Researchers adapted modeling approaches from Natural Language Processing (NLP) in identifying functional zones in urban areas [6,21]. One study applied a Latent Dirichlet Allocation (LDA) model on Foursquare check-ins to detect local geographic topics that indicate the potential and intrinsic relations among the locations in accordance with users' trajectories. Additionally, a recent study used LDA and POIs to detect functional zones [6]. Our work is different where we aim to analyze the attraction behavior of a place using measures that has not been used in any of the previous work.

Urban Attractors Framework

Fig.**??** shows the general structure of the process of analyzing attraction patterns in cities with the input datasets and the outputs. The first step in the process is to extract trips information from Call Detail Records (CDRs) of cell phones using the validated origin destination extraction algorithm implemented in [22]. We use the ODs as a data source for estimating human mobility, where it provides the amount of trip from each pair of origin and destination. From the ODs, we mine three statistical features that quantify how attractive a place is: the number of trips a place receives, the spatial dispersion of the origins of all incoming trips, and the distance distribution visitors traveled to visit the place on the road network. We use these attraction features to classify all regions in the city according to their attraction behavior. Finally, using a statistical significance testing approach we relate each type of POIs that are significantly concentrated in each types of attractors identified. In the following sections we explain the process and its output in detail.
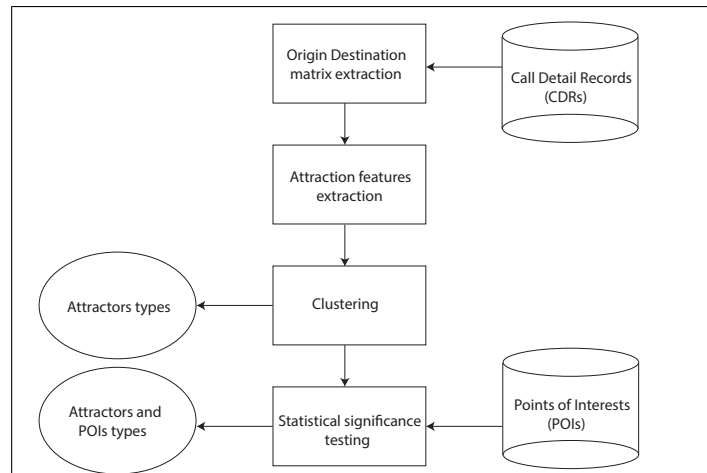
Fig 1. Urban Attractors framework

Origin Destination Matrix Extraction

The aim of this process is to extract the Origin-Destination matrices (ODs), that provides the number of trips between each pair of locations in the city for a specified time window of a typical weekday. Methods of estimating ODs range from traditional methods to more modern ones. Traditional methods include running surveys within cities and estimating the flows between locations of the city from the feedback of those surveys. Such methods consume longer periods of time and are inaccurate at times. They usually span smaller population sample sizes and thus are more prone to biases. Recent research in the domain of ubiquitous computing provided alternative methodologies for estimating more accurate ODs from user generated datasets like cell phone data. The methods proposed in [22, 23] uses mobile phone location traces (i.e. CDRs) to estimate the flows of people between areas in the city. The large scale of the cell phone data provide sufficient sample sizes and more accurate information compared to traditional methods. In this paper, we use state of the art methods of extracting OD matrices for the city of Riyadh between each pair of traffic analysis zones (TAZes) as shown in Fig.2.

Our primary source of data is one month (December 2012) of Call Detail Records (CDRs) of anonymous mobile phone users in the city of Riyadh, Saudi Arabia. Within the CDRs, each record contains an anonymized user ID of the caller and receiver, the type of communication (i.e., SMS, MMS, call, data etc), the cell tower ID facilitating the service, the duration, and a time stamp of the phone activity. Each cell tower ID is spatially mapped to its latitude and longitude where each Voronoi cell in Fig.2 correspond to a tower. The CDRs data contains more than 3 million unique users, which is a representative sample of Riyadh's population. Thus, the CDRs provide a proxy for tracking human mobility behavior in the city. However, computational steps are needed to extract clean trajectories from the CDRs.

The computational steps we take to transform raw mobile phone data (CDRs) to ODs are summarized as follows. First, we estimate transition probabilities between labeled locations such as home, work or other. Next, we filter users by number of records per day such that these locations are labeled with enough confidence (details of the method in [22]). A trip is marked by probable departures and arrivals to such locations within a specified time window. Then we scale from users to total population using census data. The output of this process is the OD matrix that indicates the

number of individuals traveling between every possible pair of locations between 7 am and 10 am on a typical weekday.

The spatial scale we used for locations is based on Traffic Analysis Zones, which is the official segmentation used in transportation planning. Conventionally segmenting the city into TAZes are based on census block information such as population per hour, where zones tend to be smaller in denser areas and larger in areas of low density. The TAZ based segmentation is more flexible and useful in analyzing places attraction patterns than using other segmentations such as neighborhood based or spatially uniform segmentation. Thus, we define our OD matrix T by aggregating cell phone towers on 1492 TAZes. The elements of the matrix are the number of trips between each pair of TAZes (i, j) in the city.
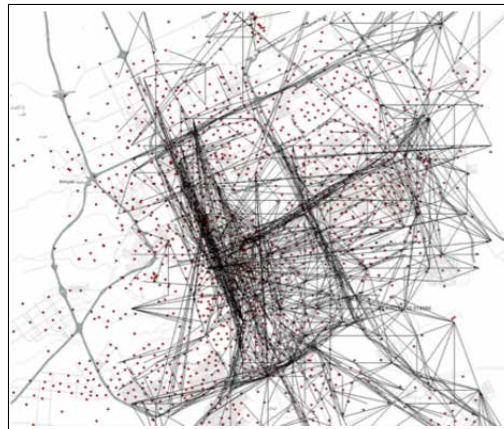

Fig 2. The ODs in Riaydh during the morning period. Each line represents a trip from a source to a destination.

Attraction Features

We aim to identify different patterns of attraction through statistical features of the inflow to each place in the city. The first feature is the total amount inflow a place receives. As the more visitors a place receives, the more attractive that place is. Additionally, a place is more attractive if it attracts people from various places in the city. Some places only attract people nearby which makes them local in terms of from where they attract people. On the other hand, some place attracts people from all over the city such as universities and hospitals. Thus, the second feature we measure is how spatially dispersed the origins are. Finally, we quantify the distances traveled to visit the place on the road network. In the following sections, we detail the calculations of each feature.

Inflow

The amount of visitors a place receives is the strongest indicator of how attractive the place is. This feature measures the attraction force of a location , where locations that have high inflow (number of visitors) are major attractors in the city. Fig.3 shows the distribution of the number of TAZes according to their inflow amount. The majority of TAZes have small to moderate inflow. However, there are few TAZes that have a very large inflow (colored red), which makes them highly significant. The inflow magnitude
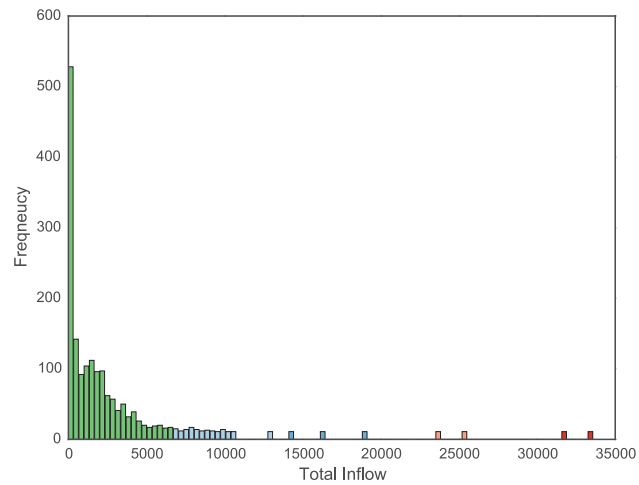
Fig 3. The distribution of total inflow received by TAZ's in Riaydh. the majority of places receive small to medium number of visitors. Few places receive very high inflow (colored red),which makes them highly attractive.
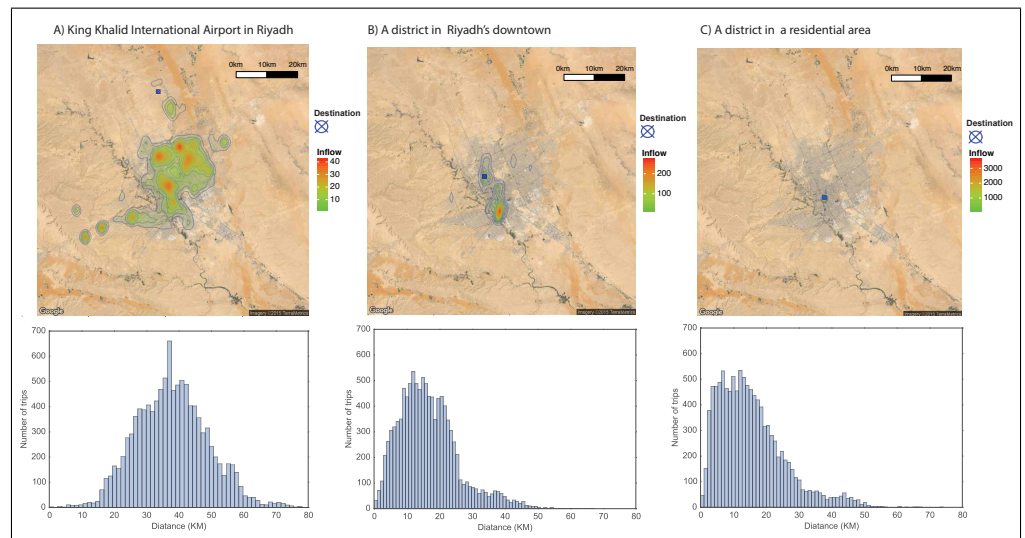

Fig 4. Spatial dispersion and distance distributions of three examples of types of attractors (marked by black dots). (A) The international airport in Riaydh. (B) A place in the downtown area. (C) A place in a residential area. The top row shows heatmaps of the origins of the inflow, where the color differs corresponds by the amount of trips from that location. The bottom row is the distribution of road distance traveled by visitors of the selected place.

of a TAZ i is simply calculated from the OD matrix as follows:

$$Inflow_i = \sum_{j=1}^{n} T_{ji} \tag{1}$$

,Where n is the total number TAZes, and T_{ji} is the number of trips from TAZ j to TAZ i.
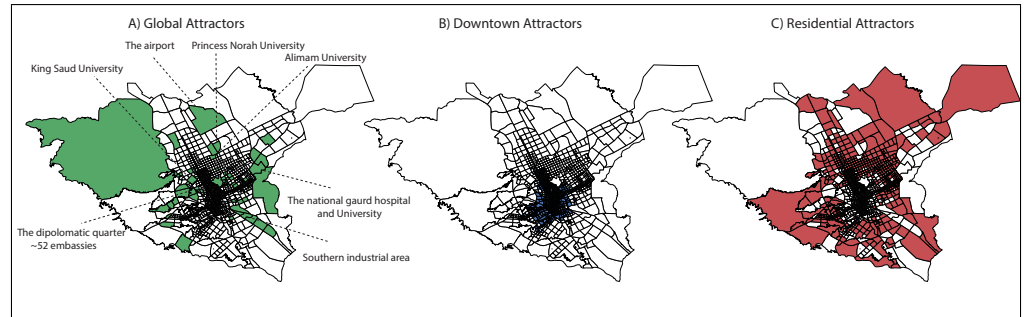
Fig 5. Types of attractors. (A) Global ~~at tractors~~, which are places with large number of visitors from all over the city traveling long distances to visit them. These include unique places in the city like the airport, large universities, hospitals, industrial areas, and the diplomatic quarter as annotated on the figure. (B) Attractors with large number of visitors due to their central location (downtown) they are more accessible and thus the distance visitors travel to reach them are shorter. (C) Less attractive places, that are located on the outer (residential) areas of the city.

Spatial dispersion

Another way to measure the popularity of a place is to measure the spatial dispersion of the origins it attracts. The spatial dispersion quantifies how spatially dispersed the locations of the origins of trips are in relation to the center of mass of all origins. A place is more attractive if it attracts visors from various and spread-out places in the city. Major attractors tend to attract people from all over the city (large spatial dispersion), while less significant attractors only attract people nearby (presenting small spatial dispersion).

We measure the spatial dispersion of visitors by calculating a weighted standard distance deviation, which is a standard method used to measure the statistical dispersion of spatial data [24]. Mathematically, the weighted spatial dispersion (SD) for a TAZ i is defined as follows:

$$SD_i = \frac{\sqrt{\sum_{i=1}^{n} w_i(X_i - X_c)^2 + \sum_{i=1}^{n} w_i(Y_i - Y_c)^2}}{\sum_{i=1}^{n} w_i} \tag{2}$$

Where n is the total number TAZes. X_i and Y_i are the spatial coordinates of the origin of a trip i. w_i is the amount of inflow from source TAZ i. X_c, Y_c are the coordinates of the spatial center of mass of all origins of all the incoming flow calculated as follows:

$$X_c = \frac{\sum_{i=1}^{n} w_i.X_i}{\sum_{i=1}^{n} w_i}, Y_c = \frac{\sum_{i=1}^{n} w_i.Y_i}{\sum_{i=1}^{n} w_i} \tag{3}$$

Fig.4 shows three examples of places with different attraction behavior. The top row shows the heat maps of the inflow sources and their concentration. The destination TAZ is labeled with a target sign on the maps. Example A shows the heat map of the international airport in Riyadh city, where the heat is spread all over the city, which indicates strong attraction. Example B is a TAZ in the downtown area, where the inflow sources are moderately spread. Example C is in a TAZ in a residential area, where it only attract visitors nearby with small spatial dispersion.

Distance distribution

Another characteristic that defines the attraction patterns is the distribution of distances traveled by all the trips the destination receives. The trip distance from each

source to the centroid of the attractor were calculated on the road network of Riyadh by using the Dijkstra shortest path algorithm [25] to find the optimal routes between all of the origin-destination pairs. This provides a more accurate estimation than the Euclidean or the Manhattan distances, as it accounts for the variation in the geometry of the road network.

The bottom row in Fig.4 shows the distance distribution of all trips a TAZ received. In Fig.4 *A* the distance distribution to the airport is unique, with long mean distance (around 40 km.) and a shift in the distribution due to the distant location of the airport in the very far north of the city. On the other hand, Fig.4 *B* has intermediate length of the mean distance with a distribution tail that corresponds to the long distance traveled by some visitors to reach downtown. Finally, in Fig.4 *C*, most of the trips correspond to short distances. Clearly, for different types of attractors the distance distributions differ. Thus, we select the mean and the standard deviation of the distribution as the features to distinguish attraction behaviors.

Clustering

To discover common patterns of inflow within cities, regions are clustered using the attraction features discussed in the previous section. We used a Hierarchical Agglomerative Clustering (HAC) approach to categorize all 1492 TAZes in Riyadh based on their attraction features. HAC classifies objects, where each object is represented as a vector of features that describe that object, based on specified similarity metrics. Here, a vector x_i represent the attraction features that describe TAZ i as follows:

$$x_i = [inflow_i, SD_i, \mu_i, \sigma_i] \tag{4}$$

Where $inflow_i$ is the inflow magnitude of TAZ i, SD_i is the spatial dispersion of the inflow sources for TAZ i, μ_i is the mean of the distances traveled to TAZ i, and σ_i is the standard divination of the traveled distance distribution.

HAC starts by assigning each single object to a separate cluster, and sequentially merge the most similar clusters until it results in one cluster. Thus, HAC requires defining how to merge clusters and how to measure the distance between them. For merging clusters, We used complete-linkage algorithm, which merges two clusters based on their most dissimilar objects a follows:

$$D(X, Y) = \max_{x \in X, y \in Y} d(x, y) \tag{5}$$

Where $d(x, y)$ is the distance between two objects $x \in X$ and $y \in Y$, and X and Y are the 2 sets of clusters. Complete-linkage algorithm is conservative when merging clusters, thus it tends to find very compact clusters, which fits our objective in in finding closely related attraction patterns. For measuring the distance between clusters' objects $d(x, y)$, we use correlation based distance metric defined as follows:

$$d(x, y) = 1 - \frac{(x - \overline{x}).(y - \overline{y})}{\|(x - \overline{x})\|_2 \|(y - \overline{y})\|_2} \tag{6}$$

Where $\overline{x}$ and $\overline{y}$ are the mean of the elements of vector x and y correspondingly, and $(x - \overline{x}).(y - \overline{y})$ is the dot product of the vectors $(x - \overline{x})$ and $(y - \overline{y})$. Correlation distance works well for finding unbalanced clusters sizes as we expect to have small number of places behaving very uniquely as strong attractors and larger number of places that are not as attractive. Additionally, the correlation score can correct for any scaling within a feature, while the final score is still being tabulated. Thus, different features that use different scales can still be used.
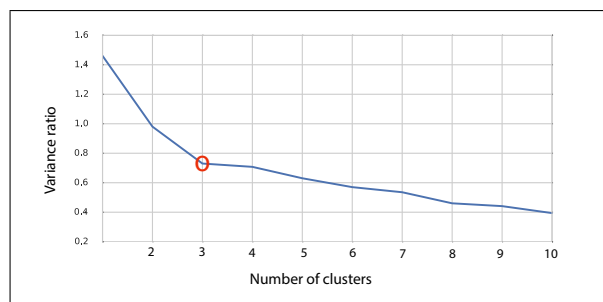
Fig 6. The ratio of the within-cluster variance to the total variance for each possible choice of K (number of clusters). The variance decreases as the data is split into more clusters until it stops decreasing significantly at $k = 3$ (marked by a red circle).
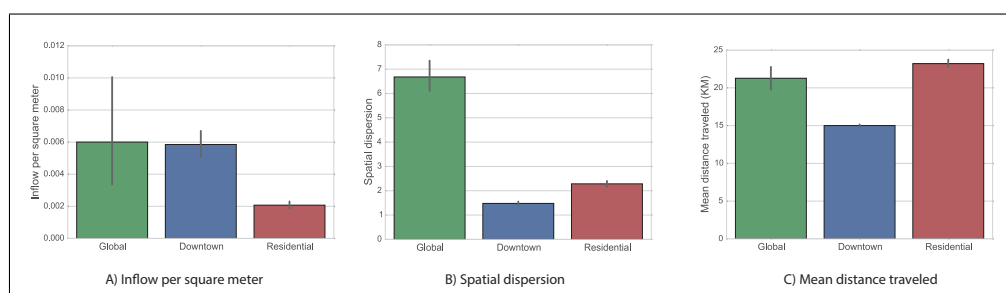

A) Inflow per square meter   B) Spatial dispersion   C) Mean distance traveled

Fig 7. The attraction features of the three detected types of attractors clusters. (A) shows the total inflow, which is high for both global and downtown attractors but extremely low for residential attractors. (B) shows the spatial dispersion of origins of visitors, where it's significantly high in global attractors. (C) shows the mean of the distance traveled by visitors, where downtown attractors exhibit smaller distance mean due to its central location and thus higher accessibility to most visitors

HAC provides a hierarchy structure of the classified regions in the city. To determine the number of clusters k that best divide the data, we calculate the ratio of the between-cluster variance to the total variance for each possible k from 1 to 10. The variance drops as k increases, until it stops decreasing significantly. We select the k that correspond to the point where the variance stops decreasing significantly, which is $k = 3$ in our case as shown in Fig.6. The classification process over all TAZes in the city of Riaydh finds three types of attractors that have distinct features. The following section extends these findings to further interpret the results.

Attractor Types

Locations in Riyadh are classified into three main types of attractors based on distinguishable attraction of trips. Fig.5 shows the 3 types of attractors classes detected, where the polygons are the TAZes. The global attractors are the ones that have significant influence on the whole city, hence their name. Unlike the remaining clusters, the locations of these places seem to be random around the city. The second detected type is the downtown attractors, which play a significant influence, after global, to attract trips. These are mostly clustered in the downtown area of the city. Finally, the residential attractors, are the least influential attractors in the morning period of typical weekdays. They are mostly located on the outer places of the city.
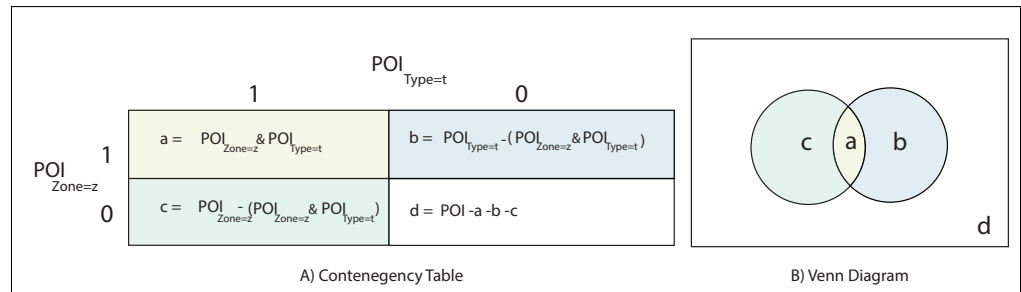
Fig 8. The contingency table and the corresponding Venn diagram for statistical significance testing method.

Global Attractors

The most distinguishable feature of global attractors is the large spatial dispersion of the incoming flows, as visitors come from all over the city to visit these places, as shown in Fig.7 A. Additionally, the amount of visitors they attract is the largest as shown in Fig.9 B, where we use inflow per square meters due to the unbalanced sizes of the TAZes. Moreover, the mean distances traveled by visitors to these locations is extremely high, which makes these places highly attractive and unique. We call these places global attractors because they strongly influence human mobility over the whole city. Global attractors always offer some unique *services* that makes them distinguished from other regions where visitors only find such services in those regions. Significant places in the city like the airport, major universities, and hospitals, that occupy TAZes on their own and are easy to identify from the map. Fig.5 A shows annotation of these major places in the city.

Downtown Attractors

The second type of attractors is the downtown attractors, shown Fig.7 B. They contain places that are mostly clustered in the central business district. These are TAZes that have relatively high inflow. However, because of their central location in the city, visitors from all over the city have short routes to access these places. They have smaller average distances compared to the other two types as shown in Fig.5 C. For the same reason, the dispersion of the origins from the center of mass of inflows is also small as shown in Fig.5 C. The significant feature of these places is that they attract great number of visitors and are accessible.

Residential Attractors

Residential attractors shown in Fig.5 C attract the smallest number of visitors. As they are located in the outer sides of the city, the visitors to these places travel long distance on average to reach them as shown in Fig.7 C. For the same reason, the dispersion of the small number of visitors is larger than the downtown attractors as shown in Fig.7 B.

Attractors and POIs

We aim to further uncover what may cause the trips to the different types of attractors discovered. Thus, we relate the classified TAZes to the composition of services offered in these TAZes. We used official Points of Interests (POIs) data that contains all places in Riyadh city, around $12,000$, offered by Arriyadh Development Authority (ADA), which

is the official entity managing all urban planning tasks in the city. POIs are classified into 23 subtypes of services offered in the city such as restaurants, schools, hospitals ... etc. In the following section, We explain the methodology by which we quantify the relationship between different types of POIs and attractors types .

Statistical significance testing

We aim to relate each type of POIs to the different attractors types to explore what makes different places have different attraction profiles. Using only counts of POIs per attractor type is not enough, because the distribution of types of POIs is unbalance. For example, restaurants are very frequent and spatially distributed, whereas universities are much fewer and are present in few districts. Thus, to identify which type of POIs are significantly concentrated in each type of attractors, we use a statistical significance testing approach. The statistical significance testing method measures the probability of observing the amount of a POI type in a spatial zone. It factors in the amount of all POIs in the spatial zone in addition to the amount of POIs with that type tag in the whole city.

We used Fisher's Exact Test (FET) to relate each type of POIs to the different attractor types. We selected FET because it works for small observations and calculates the exact probabilities rather than approximations such as in Chi square test. FET aims to test the dependency between two categorical variables given the observed data. It takes a contingency table as an input, which represent the relationship between two categorical variables in terms of their frequency distribution and their overlap. In our context, the first variable $POI_{Type=t}$ is the number of POIs that belong to a specific type t , and the second variable $POI_{Zone=z}$ is the number of POIs in a spatial zone z . We aim to test the significance of the overlap $POI_{Zone=z} \& POI_{Type=t}$,which is the amount of POI t in zone z, between these two variables. Fig.8 shows the contingency table and the corresponding Venn diagram representation. In Fig. 8, a represents the overlap, which is the amount of POIs of type t that are located in zone z, b represents the amount of POIs of type t that are not located in zone z, c is the amount of other POI types located in zone z, and d represents the number of all the rest of POIs in the city.

The objective is to test weather the amount of POI type t is significantly concentrated in a spatial zone z. FET quantifies the probability (p-value) of observing that amount of POI type t or larger in zone z by chance. The smallest the p-value is, the strongest the concentration of type t in zone z is. FET calculates the p-value p as follows:

$$p = \frac{(a+b)!(c+d)!(a+c)!(b+d)!}{a!b!c!d!n!} \tag{7}$$

Where n is the total amount of all POIs in the city. FET incorporates two crucial factors when measuring the significance. First, the significance depends on how many other POI types in zone z as a measure of purity. If a zone has a large amount of POIs of type t ,but it also has a lot of other POIs types, that makes POI of type t less significance due to this impurity in the decomposition of all POIs in that zone. Second, the significance depends on the amount of POI of type t that are not in the tested zone z as a measure of rarity. If there are large number of POIs of type t elsewhere, that makes type t insignificance in the tested zone. These two features makes FET superior to trivial methods like calculating the percentages of POIs types in spatial zones.

We find that the proposed method can capture how each type of attractor has a different composition of POI types as shown in Fig. 9. We discuss the attractors types and the significant POIs related to them in the following sections.
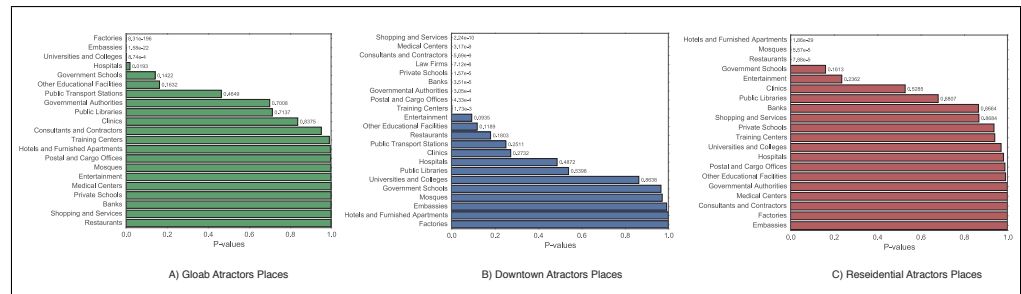
Fig 9. The types of POIs ordered by their statistical significance in the corresponding attractor's type. (A) The p-values of finding each of the POI types in TAZes classified as global attractors, where POIs types are ranked by their significance showing that factories, embassies, universities, and hospitals are type of services significantly located in global attractors. (B) POI types significantly located in the downtown attractors, which mostly includes business types of services. (C) POI types significantly present in the residential attractors that includes services and amenities such as furnished apartments, mosques (places of worship for Muslims), and public schools.

Global Attractors

Fig. 9 A shows each type of POIs ordered by their p-values, which represent how much they are concentrated in global attractors. Factories, embassies, universities, and hospitals are the top POI types that attract massive amounts of people coming from all over the city. The city has three major universities attracting a student body of $40k$ each contributing significantly to the observed global attraction. The city of Riyadh also has a major industrial area in the south to where a major number of factory workers commute. In addition, Riyadh city has the diplomatic quarter district, which hosts over 50 embassies attracting workers and visitors from all over the city to process documents. Hospitals are also expected to cause global attraction as visitors come from all over the city.

Downtown Attractors

Fig.9 B shows the types of POIs mostly located in the downtown attractors in a descendant order by significance. The first observation is that we witness a large number of POI types with strong significance (p-values¡0.01) compared to the other two types of at tractors. That is due to the richness in quantity and variation of types of POIs in this area. Thus, we expect larger number of POIs types to be significant in that area. The common theme of significant POI types in this attractor is businesses, typical in downtown regions.

Residential Attractors

Fig.9 C shows the types of POIs concentrated in the residential attractors in a descendant order. Most significant types of POIs are services needed in residential areas like furnished apartments, mosques (worship places for Muslims), restaurants, including small restaurants and fast food places, and public schools. These types of places are not unique, so each residential neighborhood has its own share of these places to serve the population living nearby.

Conclusions and future directions

We present a novel computational framework to discover different attraction patterns in cities. We proposed 3 dimensions to define attraction of urban zones: total number of incoming trips, the spatial dispersion of the origins of trips, and the distribution of distances traveled by visitors to reach that district. Further, we present a method for understanding the relation between the decomposition of the types of POIs in a spatial zone and its attraction behavior. We applied the method and discuss the results in the city of Riyadh, the capital of Saudi Arabia.

The results of implementing the discussed modules mine data from mobile phones to provide a coherent understanding of the dynamics of the interaction between the flows of people to a district and types of services (POIs) that are located in that district. We detect three attraction patterns in the city of Riyadh according to the morning mobility dynamics. Global attractors, receive large share of the visitors traveling longer distances and coming from all over the city. These attractors have places of interest that are the destination of large student bodies, factory workers, hospital associates, and embassies. The second type of attractor is that of the downtown area, which receives high inflow of people from smaller distances and spatial dispersion due to its central location in the city that makes them accessible. The most significant POIs types located in the downtown attractors are business based places like firms, shopping and service places. The least significant trip attraction is to the residential areas in the morning hours, where the amount of inflow is the lowest. Residential attractors contain common POIs that serve inhabitants such as apartments, mosques, and schools.

Several interesting directions can follow this work. One is developing a predictive model of inflows and trip distributions, taking into account set of POIs, distance between origin and destinations and their population densities. Another possible direction is to compare these patterns among various cities to learn about best urban plans that reduce traffic.

Acknowledgments

References

1. Authority AD. King Abdulaziz Project for Riyadh Public Transport; 2016. Accessed: 2016-03-30. `http://www.ada.gov.sa/ADA_e/DocumentShow_e/?url=/res/ADA/En/Projects/RiyadhMetro/index.html`.

2. Louail T, Lenormand M, Picornell M, Cantú OG, Herranz R, Frias-Martinez E, et al. Uncovering the spatial structure of mobility networks. Nature Communications. 2015;6.

3. Karamshuk D, Noulas A, Scellato S, Nicosia V, Mascolo C. Geo-spotting: mining online location-based services for optimal retail store placement. In: Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining. ACM; 2013. p. 793–801.

4. Zheng Y, Capra L, Wolfson O, Yang H. Urban computing: concepts, methodologies, and applications. ACM Transactions on Intelligent Systems and Technology (TIST). 2014;5(3):38.

5. Blondel VD, Decuyper A, Krings G. A survey of results on mobile phone datasets analysis. EPJ Data Science. 2015;4(1):1–55.

6. Yuan NJ, Zheng Y, Xie X, Wang Y, Zheng K, Xiong H. Discovering urban functional zones using latent activity trajectories. Knowledge and Data Engineering, IEEE Transactions on. 2015;27(3):712–725.

7. Pan G, Qi G, Wu Z, Zhang D, Li S. Land-use classification using taxi GPS traces. Intelligent Transportation Systems, IEEE Transactions on. 2013;14(1):113–123.

8. Toole JL, Ulm M, González MC, Bauer D. Inferring land use from mobile phone activity. In: Proceedings of the ACM SIGKDD international workshop on urban computing. ACM; 2012. p. 1–8.

9. Lee M, Holme P. Relating land use and human intra-city mobility. PloS one. 2015;10(10):e0140152.

10. Zheng Y, Zhang L, Xie X, Ma WY. Mining interesting locations and travel sequences from GPS trajectories. In: Proceedings of the 18th international conference on World wide web. ACM; 2009. p. 791–800.

11. Fan Z, Song X, Shibasaki R. Cityspectrum: A non-negative tensor factorization approach. In: Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing. ACM; 2014. p. 213–223.

12. Liu X, Gong L, Gong Y, Liu Y. Revealing travel patterns and city structure with taxi trip data. Journal of Transport Geography. 2015;43:78–90.

13. Zhan X, Ukkusuri SV, Zhu F. Inferring urban land use using large-scale social media check-in data. Networks and Spatial Economics. 2014;14(3-4):647–667.

14. Long X, Jin L, Joshi J. Exploring trajectory-driven local geographic topics in foursquare. In: Proceedings of the 2012 ACM Conference on Ubiquitous Computing. ACM; 2012. p. 927–934.

15. Bassolas A, Lenormand M, Tugores A, Gonçalves B, Ramasco JJ. Touristic site attractiveness seen through Twitter. EPJ Data Science. 2016;5(1):1–9.

16. Han H, Yu X, Long Y. Discovering functional zones using bus smart card data and points of interest in Beijing. arXiv preprint arXiv:150303131. 2015;.

17. Zhong C, Schläpfer M, Arisona SM, Batty M, Ratti C, Schmitt G. Revealing centrality in the spatial structure of cities from human activity patterns. Urban Studies. 2015;p. 0042098015601599.

18. De Nadai M, Staiano J, Larcher R, Sebe N, Quercia D, Lepri B. The Death and Life of Great Italian Cities: A Mobile Phone Data Perspective. arXiv preprint arXiv:160304012. 2016;.

19. Fan Z, Song X, Shibasaki R, Adachi R. CityMomentum: an online approach for crowd behavior prediction at a citywide level. In: Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing. ACM; 2015. p. 559–569.

20. Wu L, Leung H, Jiang H, Zheng H, Ma L. Incorporating Human Movement Behavior into the Analysis of Spatially Distributed Infrastructure. PloS one. 2016;11(1).

21. Yuan J, Zheng Y, Xie X. 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. ACM; 2012. p. 186–194.

22. Toole JL, Colak S, Sturt B, Alexander LP, Evsukoff A, González MC. The path most traveled: Travel demand estimation using big data resources. Transportation Research Part C: Emerging Technologies. 2015;58:162–177.

23. Toole JL, Colak S, Alhasoun F, Evsukoff A, Gonzalez MC. The path most travelled: mining road usage patterns from massive call data. arXiv preprint arXiv:14030636. 2014;.

24. Mitchell A. The ESRI guide to GIS analysis, Volume 2: Spatial Measurements and Statistics. Redlands. CA: Esri Press; 2005.

25. Dijkstra EW. A note on two problems in connexion with graphs. Numerische mathematik. 1959;1(1):269–271.