DOI: 10.24996/ijs.2022.63.5.36

ISSN: 0067-2904

A Review of Flow Migration Through Mobile Networks

Manhal K. Alqaysi*, Suhad Faisal Behadili
Department of Computer Science, College of Science, University of Baghdad, Baghdad, Iraq

Received: 14/4/2021 Accepted: 12/7/2021 Published: 30/5/2022

Abstract
The interesting new sources of data for official statistics are cell phone data. Electronic media has defined the way of research human behavior rapidly over the last decade. As data storage and sensing technology progressed, electronic records now cover a diverse variety of human activities from localized data (phone) to open source contributions on Wikipedia and the Open Area Map. Electronic records now encompass the numerous fields of activity. The ad hoc vehicle network is a research community-based wireless technology for the implementation of intelligent transport applications. It is necessary to estimate migration flows and predict future trends to understand the causes and effects of migration and to enforce policies to deliver certain services. Several studies have exposed in this review with their datasets such as Credit card records (CCRs) provide deep insight into buying behavior; Call Details Records (CDRs) present new possibilities for under-implemented human mobility. Therefore, various forms of transportation and other travel behavior, various travel-related events such as "Home-Tour-Work -Tour-Tour" and the corresponding travel-related motifs have also been distinguished by the inclusion of land use details in the GIS data. This review also investigated migration trips residence between cities and inside a single city. The review concluded clear results for the adoption of the mentioned data, for example, mobile phone data (CDRs), because it is very useful as it provides real big data or real time big data without additional cost and is available in telecommunications companies, from which it is possible to analyze the movement of communities and deduce the activity of a particular city. In the future, there is a tendency to use this type of data from Korek Telecom Company in Iraq to the flow migration of Iraq governorates by using the gravity model. As well as, an attempt to study the conditions of cities and the movement of individuals in urban places to clarify the needs of the city in its need for new improvements.

Keywords: Human mobility, Migration, CDRs, Flow, GPS, GSM.

مراجعه لترحيل تدفق الناس بأستخدام بيانات الهاتف المحمول

منهل خالد عدنان*, سهاد فيصل شيحان
قسم علوم الحاسوب, كلية العلوم, جامعة بغداد, بغداد, العراق

الخلاصة
مصادر البيانات الجديدة المثيرة للاهتمام للإحصاءات الرسمية هي بيانات الهاتف الخليوي. حددت وسائل الإعلام الإلكترونية طريقة البحث في السلوك البشري بسرعة خلال العقد الماضي. مع تقدم تقنية تخزين البيانات والاستشعار ، تغطي السجلات الإلكترونية الآن مجموعة متنوعة من الأنشطة البشرية من البيانات

*Email: Manhal.adnan1201@sc.uobaghdad.edu.iq

المحلية (الهاتف) إلى مساهمات المصادر المفتوحة على ويكيبيديا وخريطة المنطقة المفتوحة. تشمل السجلات الإلكترونية الآن العديد من مجالات النشاط. شبكة المركبات المخصصة هي عبارة عن تقنية لاسلكية قائمة على المجتمع البحثي لتنفيذ تطبيقات النقل الذكية. من الضروري تقدير تدفقات الهجرة والتنبؤ بالاتجاهات المستقبلية لفهم أسباب وآثار الهجرة وإنفاذ سياسات لتقديم خدمات معينة. كشفت العديد من الدراسات في هذه المراجعة مع مجموعات البيانات الخاصة بها مثل سجلات بطاقات الائتمان (CCRs) التي توفر رؤية عميقة لسلوك الشراء؛ تقدم سجلات تفاصيل المكالمات (CDRs) إمكانيات جديدة للتنقل البشري قيد التنفيذ. مما يقود، لشكل مختلف من وسائل النقل وسلوك السفر الأخرى. تم أيضًا تمييز العديد من الأحداث المتعلقة بالسفر مثل "جولة منزلية – عمل – جولة – جولة" وما يقابلها من عناصر متعلقة بالسفر من خلال تضمين تفاصيل استخدام الأراضي في بيانات نظم المعلومات الجغرافية. وقد حققت هذه المراجعة أيضًا في رحلات الهجرة والإقامة بين المدن ويتقاطع مع المدينة الواحدة. أفضت المراجعة إلى نتائج واضحة لاعتماد البيانات المذكورة، على سبيل المثال، بيانات الهاتف المحمول (CDRs)، لأنها مفيدة جدًا حيث انها توفر بيانات كبيرة حقيقية أو بيانات كبيرة في الوقت الفعلي دون تكلفة إضافية ومتاحة في شركات الاتصالات، ومنها يمكن تحليل حركة المجتمعات واستنتاج نشاط مدينة معينة. في المستقبل، هناك اتجاه لاستخدام هذا النوع من البيانات من شركة كورك تليكوم، العراق لدراسة تدفق الهجرة بين محافظات العراق بأستخدام نموذج الجاذبي (gravity model) . وايضاً محاولة دراسة ظروف المدن وحركة الافراد في الاماكن الحضرية لتوضيح احتياج المدينة لتوزيع جديد.

1. Introduction

Digital traces are created by communication networks (mobile telephone networks and social media platforms). By analyzing those traces over cities, good data on everyday activities of residents can be given, so urban planners have relevant guidance for city design and developing decision making.

The Call Details Records (CDRs) often offer precious spatial and temporal information at city or national level. In the background of the observation and during activities, the CDRs may be investigated for extraction of patterns of life and mobility of individuals in the observed urban zone[1]. Comprehension of traffic patterns is important to urban and transport planning. The field of travel demand modeling for which the sequence four-step model has historically been used in transportation forecasts, such as predicting the number of vehicles on a planned route, the riding on a railway line, and the number of bus passengers at the airport, is studied in terms of human mobility from the transport perspective[2]. The movement of both citizens and organizations in short term and long distances. The way people live has always been inextricably associated with their movement. In former movements, factors such as climate change and unfriendly environments were primarily affected, while the current movements were largely driven by socio-economic factors, such as employment, work conditions and food [3].

The rest of this paper organized as follows: section 2 explain related work, section 3 describes the Geographical and locating side. While, section 4 represents some mobility techniques and models with a little bit of description. Then, section five displays some of the datasets with their manipulation methods with tables to understand what sides, which entered. Finally, conclusion section represents the summary of the article content.

2. Related Work

Many studies examined mobile data will be presented in this section. Thus, in [4], a pilot study presented the key move for using mobile data in official statistics: where someone lives, to detect his place of residence. Identification of important locations which are places where people spend substantial time. Most famous example is home or work. Such places need actively focused on the sensed position traces users involved [5]. The complementary of national data is one conceptual application of cell phone indicators. The combination with census data, which usually provide contextual information that, opens up fascinating research possibilities by capturing activity in large populations. Concentrate explicitly on three different spatial factors that affect confusion and errors during development or using mobile phone indicators [6]. In [7], modeling individual lifestyles together is an issue of greater heterogeneity in relation to aggregate activity in city areas, which is more complex. GPS traces provide precise spatiotemporal information for people, but generally, GPS-assisted surveys remain limited in their sample size and monitoring period [8]. Human mobility is being studied using numbers of proxies such as CDRs, GPS, Wi-Fi, travel studios. For predictive models such as Markov chains, Naïve Bayes, Artificial Neural Systems, and time series analysis, a variety of techniques were proposed [9]. For optimization of road, most studies using GPS data for estimating fuel consumption based on individual user fuel consumption [10]. In [11] would like to discuss the community structure and relationship of social networks in cities. Draft study to investigate the role in social networks of social and geographical distances. In addition, social groups described as network communities that are dense sub-networks locally. So, in comparison, advanced technology such as GPS, Location-Based Services (LBS), Location-Based Social Networks (LBSN), Online Social Networks (OSN), and mobility data can all sensor person movement passively for large numbers of citizens and at high resolution spatiotemporal [12]. The precise capture of population movement between various regions is one of the greatest challenges modeling companies face when it comes to consider and replicate the spatial distribution of an infectious disease epidemic [13]. In emerging economies, transport networks are stressing high population growth in urban areas. Road congestion is a common reality in many urban cites, especially on the main roads which feed the city workforce day-to-day [14]. Non-traditional sources of data include information produced by internet users or devices actively or passively [15]. However, understanding the patterns of human mobility was an important subject study on transportation [16]. To consider the demand for travel, the time-space distribution of these operations must be understood places. The movements of people from place to place can be grouped according to their time and space features [17]. Urban life is dependent on weekdays and its time variability [18]. The key distinguishing characteristic is working days from weekends [18]. More complex movement characteristics are also present for each day, including Friday and Monday effects on the weekends; Saturday effect is affected by working days or the mindset of shopping days [18]. The mobility of people influences our current society and the world different scale trends. The long distance and long journeys consist of rare and irregular occurrences such as international or traveling flights between towns. In comparison, short journeys consist mostly of intercity journeys such as job or shopping for food [19]. Hence, Commute (move between the workspace and the place of residence) is the most common journey of a common person who makes together profound patterns of mobility that often characterize the area central mobility characteristic – morning flows generally (residence to workplace) and night movement (workplace to residential) [20]. Intentionally and periodically, group travel behavior (GTB) is classified as two or more switches from one place to another, moving together [21]. The fundamental concept is: if two or more commuters tap at the same bus/metro stop in a very short time window and at least one such syncing takes place they are known as community travelers a certain number of times[21]. The study of migration therefore covers different fields of science, including anthropology, sociology, economics, mathematics, physics and more recently informational. It has the so-called "social big data" of people created by mobile, web, Online Social Networks (OSN), web-based devices via things [22]. This investigation focus on three separate migration phases: the journey -analysis of stocks and flows of migration; Stay - study of multiculturalism and changes in participating communities; return - study of returning migrants to their country of origin [21]. The car is a vital part of Vehicle Ad hoc Network (VANET), which is an evolving sub-class of MANET. On the road, where vehicles are mobile nodes, VANET deployed. Hence, a routing node provides information on vehicle path, road pattern, and speed and vehicle density. Inter- and intra-vehicle contact in VANET. The contact between vehicles means communication between vehicles (inter-) and the communication inside the vehicle implies intra-vehicle communications [23].

3. The Telecommunication and Geographies

The Geography is the analysis of places and their relationships with people. Thus, the Geographers explore physical characteristics of the Earth surface and the human's societies that are spread over it. They also analyze the interaction of human culture with natural surroundings, and the way that places can influence people. However, the Geography tries to recognize where the stuff found, why it is, and how it develops and changes over time. Actually, regarding to focus on the position details every 15 minutes, for one sample Smartphone frequency. Since, every location sample has a timing, latitude, longitude, and precision meaning [9]. Such allocation normally occurs with home detection algorithms incur the cell tower that is more plausible to protect the home of a customer on reasonably straightforward heuristics such as maximum nighttime work [24]. Commuting networks from census surveys in three census surveys, one for each studied country is taken from the census commuting networks: Portuguese, Spanish and French [13]. Every survey will track the number of people traveling daily between two places in the country for work or studying reasons [13]. Networks are created by establishing a directional weighted connection between two nodes, showing where the origin and destination are, and weight shows how many passengers travel on that connection on a daily basis [13]. In order to make outgoing and incoming calls via a cell phone calls must report regularly the nearby cell towers their presence, and thereafter record their location within the geographical cell protected by one of the towers [25]. Generally, it is more difficult to use GPS data for an effective communication study because users need constant consent to obtain this data and quickly drain the battery, particularly because of the long time it takes to gain the signal [26]. The following user cell phone records are used to identify whether the user stays or passes by those tasks at a single location, en-route to your destination [17]. Listed the most important places each user visited such as home for users, work and other sites. Then the journey to these sites is related to the activities occupational or home or other [17]. Analysis of data was achieved by allocating a single SIM in each date at the divisional secretarial division (DSD) level home and workplace [27]. Social networking profiles are normally connected to different kinds of Data, including the present and previous locations of users and demographics, such as age and gender [15]. Mobile locations are usually close to the collected movement data using various means: travel plans, GPS, counters, video and questionnaires tracking [18]. The data includes two or three coordinates of position (x, y, z) and time Coordinate (t)[18]. The home and work location has been checked manually via questionnaire and place data compared [18]. Location-based social networks (LBSNs) are social platforms online let their users share with friends and the public their locations [28]. Each check-in provides information that shows who spend time (which user), where (what position), whenever (what daytime, what weekday), and do what (depending on the type of location: grocery shopping, restaurant dining, etc.)[28]

4. The Mobility Models

Regarding the mobility models, either analytical or simulation models can be used to describe behavior or activity of user movement. The contribution to models of analytical mobility simplifies assumptions in respect to user movements. These models classify mobile user movements over a period regarding their location, speed, and direction. Therefore, this section would explain and discuss several types of models and techniques that used in mobility study.

4.1 Mobility techniques

The emerged technology now forms an important part of everyday life and work. Nowadays, the network technology takes big part of our lives like Internet, Social Networks, News, Media and 3G/4G networks. It has been developed with IoT, it is possible to attach multiple sensor devices in our living areas to cyber and physical objects [29]. The data generated by smart phones, tablets and wearable have begun to be used by geo-crowdsourcing with the new way of sensing called Mobile Crowd Sensing (MCS). In reality, MCS is viewed as an IoT and volunteer paradigm. In this concept is proposed a new model of human-human-related objects, things and things [30]. Whereas, VANET refers to an ad-hoc network where various passing cars and other connected systems contact each other via a wireless media and exchange valuable information. A small network with vehicles and other equipment as nodes in the network is created at the same time. So, [23] VANET plays a significant role in mobility of automobiles. The VANETMobiSim & Bonnmotion method is used to produce traces of practical mobility including intelligent driver model-intersection management (IDM-IM), IDM–lane changing (LC) and ManhattanGrid. Hence, the models of mobility imitate actual traffic movement. There are multiple factors affecting mobility such as traffic type, speed limits, path layout, random vehicle delivery, intersection assistance and contact disruption as a result of obstruction [23]. Continuous track places are found by the Geographical Information System (GIS) analysis techniques to common algorithms of K-means, Bayesian nonparametric approaches, radio system, and fingerprinting [5]. In [31] the study used a simple model based on the previous data to simulate a user location most frequently, and calculate the exact likelihood of prediction (Precision). To find these sites, break the week into hours, tantalizing $7 * 24\, slots = 168$ slots per week for each hour as in Figure 1.

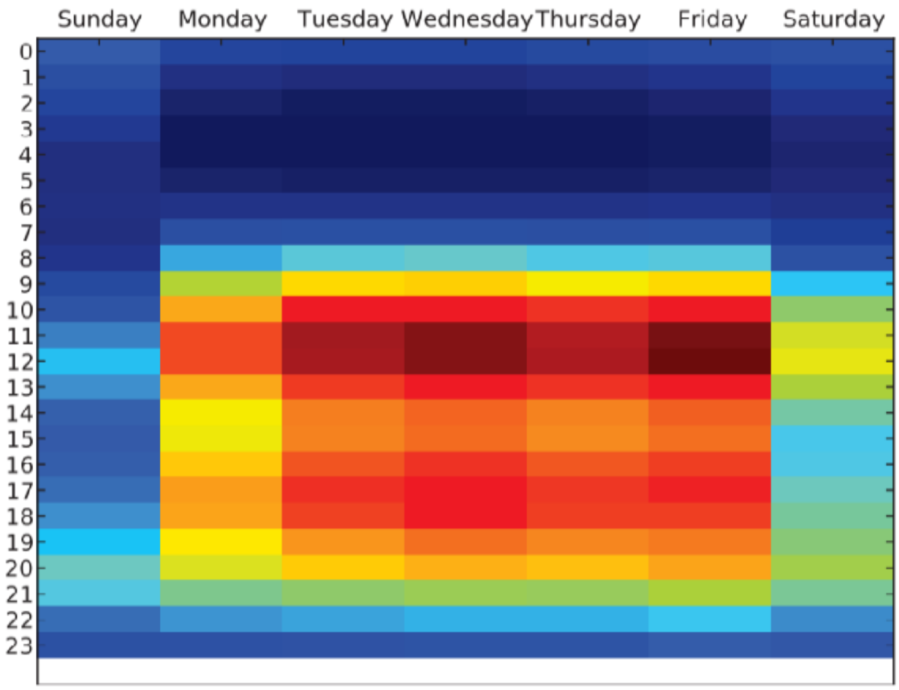

Figure 1- Averaged over a span of five months, call allocation according to the day and hour of the week [31]

As well as, a Multivariate Hawkes cyclical process (MHPs) model was proposed, which in turn captured the rates of different activities among the movements of people [32]. As well as, in [9] has been considered for two models of approaches, the density based clustering, where every user processed individually and Markov chain model, where the prediction of the future behavior linked with the last one. Moreover, the road environment has multi-lane highways, unusual intersections and traffic lights. Thus, the speed and slow-down models are considered because vehicles sharply split and travel [23]. The Random Way Point (RWP) for the first proposed the time and was used for wireless network applications by several researchers [33]. The mobile node randomly designs the movement, the definition random speed and coordinates. So, moving node by algorithm moves when the description coordinates are reached to arbitrarily redefine random coordinates at a random speed [30]. The Model of Versatility was used the most frequently [33][[34]. Furthermore, Random Walk (RW) model which has been explained mathematically by Einstein on 1926 as the first time. The mobile node is used in this model randomly between $0\ and\ 2\pi$ a path at a random rapid rate between $Vmin\ and\ Vmax$, respectively. Thus, randomly redefined route, speed and time and the visit continues then. The mobile node goes on to visit till established time initially. So, multiple mobile nodes can exist. This model was simplified; hence each mobile node was identical Speed [30]. The RW is an extremely simple model of RWP and shown in Figure 2. Meanwhile, the random direction model is equal to the RW model. But, the difference is that RD is moving in a random way at a given random direction ($0\ and\ 2\pi$) at range between $Vmin\ and\ Vmax$ speed until a defined border area is reached. When the given area boundary is reached and the travel then continued. The RD is a kind of RW where the random direction and speed are redefined [30]. In order to spread these points more evenly in the field of simulation, the RD mobility model was designed to solve the issue of the RWP model [30] The RD is presented in Figure 3. As well as, Random point (RP) is the developed model of RP which inspired by a RWP model. During a random speed of mobile nodes between and a mobile node is active to a random coordination (x_0, y_0) in this model, it differs from RWP it followed the sense of the path. The mobile node has finished the initial phase of the model by equalizing the random x or y with a view to reach the randomly established coordinate (x_0, y_0) [30]. The RP showed in Figure 4.

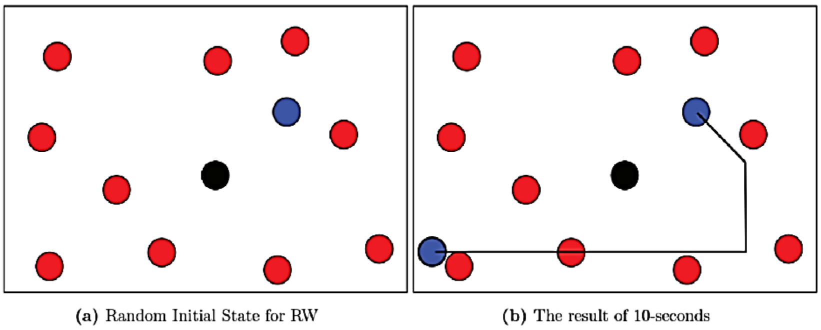

Figure 2-Random walk (RW) model [30]

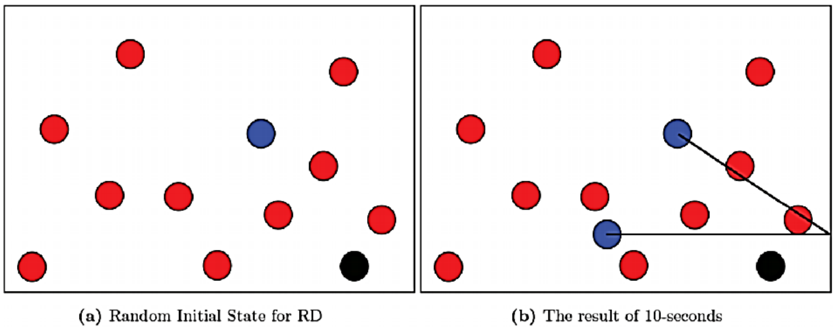

Figure 3-Random direction (RD) model [30]

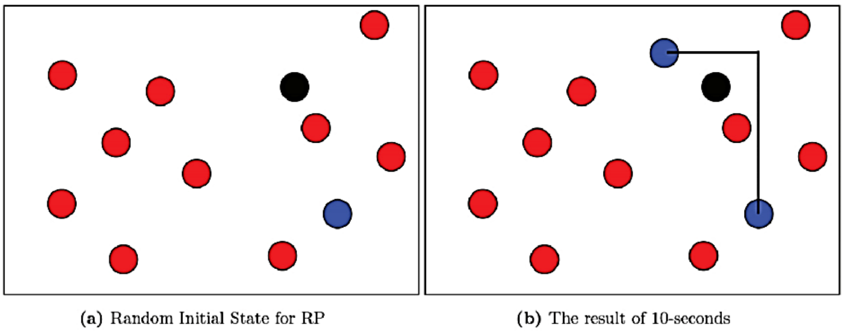

Figure 4-Random point (RP) model [30]

Random Journey (RJ) is a model form based on the RW model. In this model, the mobile node travels in a random direction $(0-2\pi)$ to a randomly determined pixel measurement. The period of the navigation simulation continues until a number of pixels for a random, separate path are completed randomly when hitting the appropriate point [30]. Also, Inference Route Choice [2], the role of this technique could be to determine the line or direction most likely to go to, with a set of possible routes as well. The method of interpolation depends on the way points and the choice of the road may lead to a difference in the number of points for the road. Curved paths contain an additional path and the path divisions are equal in number of segments with the same waypoint divisions. This technique was carried out using mobile data CDRs for Portugal between 2006 and 2007, covering millions customers. As shown in the Figure 5.

All in all[30], the performance in the experimental environment was compared by three different measurements that are the Nodes visit rates, access rates, and the number of messages they transmitted. The biggest increase in different areas of the model of RP at various numbers of nodes was 2.6% compared to RWP, 7% compared to RW and 46.34% compared to RD, while for the number of nodes visited these values were 3% compared with RWP, 1.5% compared with RW and 17.67% versus RD. The model collected more messages, 1465.4, 2933.46 and 7260.12, in the same conditions as those collected by the number of messages, compared to those in RWP, RW and RD respectively. In terms of reaching the basis, the biggest increase in different sizes and numbers of nodes of the RJ model is 1% versus RWP and 3.5% versus RW and 25% versus RD, whereas the visited number of nodes was 0.75% compared with RWP, 2% compared to RW and 21.4% versus RD. The model collected $1109.56, 1534.26\ and\ 4488.5$ messages more than those in RWP, RW and RD respectively under the same terms in the metric of the number of messages.

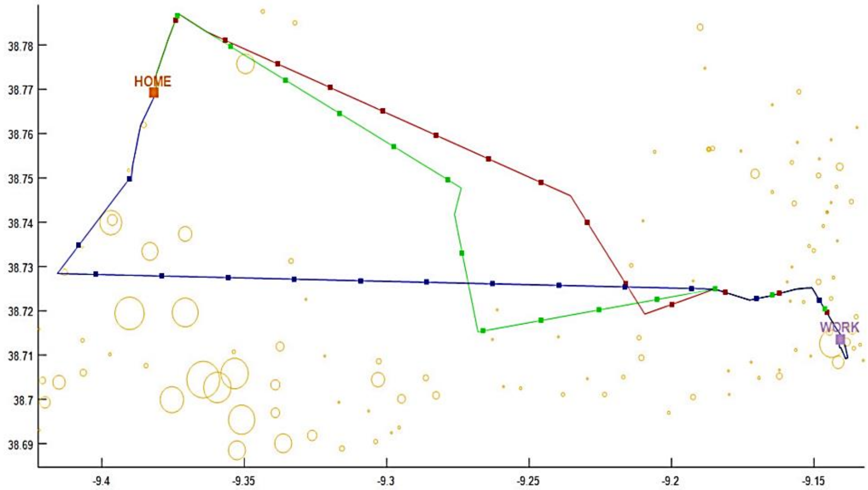

Figure 5-Inferring route choice of equal number of waypoints [2]

On the other side, the gravity model summarizes the volumes of trips between two areas, inspired by analogies to the gravity law in Newton, which can be considered as proportional to its populations but inversely proportional to their costs of travel. The relationship is mathematically defined by Eq. (1) [3]

$$T_{ij} = A_i O_i B_j D_j f\left(C_{ij}\right) \quad \ldots\ldots\ldots (1)$$

Where, the number of tours between district i and district j is T_{ij} , $O_i$and $D_j$is the total end of the tours of districts $i \: and \: j \: and \: f(C_{ij})$ is the general cost of travel between districts $i \: and \: j$.
For the analysis of value within contingency tables, the log-linear model has applied. The analysis of the $O - D$ matrix flow among 18 districts could consider as a two-way eventuality table. The multiplier portion of log-linear model can be written as shown in Eq. (2), describing the complete system [3].

$$T_{ij} = t t_i^o t_j^D t_{ij}^{oD} \ldots \quad \ldots\ldots (2)$$

Where, t is the principal effect representing migration level, t_i^o and t_j^D is the principal effects of origin and destination as represented by categorical data. Given some limitations in the gravity model[35], for example, the model requires previous traffic data or traffic to predict mobility in poor areas of regular traffic, and these areas are of great importance in the movement and modeling of some infectious diseases. The inability of the gravity model to predict or explain fluctuations between the numbers of travelers to the two sites. It also predicts that the number of passenger increases with the increase in the population of the destination and the increase will be without limits, however, it is not possible for the number of passengers to exceed the number of residents at the source, and this highlights the analytical contradiction in the model. We turn to the radiation model, which solves many of the constraints or problems involved in the gravitational model. The radiation model comes from the dynamics of dissemination and inspired by Stouffers interference framework. The radiation model explains the quantity of trips made by two areas as the selection process consisting of the selection of jobs and the selecting of jobs. The jobs selection takes number of opportunities in each area into account, while the selecting of jobs whose criteria is to select the job that is next best available in the resident region, and with a benefit higher than in the resident region[3]. Mathematically, represented by Eq. (3):

$$T_{ij} = T_i \frac{m_i n_j}{(m_i + s_{ij})(m_i + n_j + s_{ij})} \quad \ldots\ldots\ldots (3)$$

Where, T_{ij} represents the numbers of trips between origin I and destination j, T_i is the total number of journeys leaving place I, mi is the population of region I, ni the population of region j and s_{ij} the enclosed community of the distance between zones in the circle with radius excluding the zones. Additionally, the model of radiation builds a synthetic network that is based on stochastic decision making method that assigns work locations to each potential commuter, so that regular flows across the country are calculated [13]. Finally, Table 1 summarized the common mobility techniques [3][9][23][30][31][33].

Table1- Summary of mobility techniques

Model name	Highlights
MHPs	Captured the rates of different activities among the people movements
Density based clustering	Every user processed individually
Markov chain model	The prediction of the future behavior linked with the last one
Speed and slow-down models	Specify the speed of vehicles in the street
RWP	Used for wireless network applications. Moving node by algorithm moves when the description coordinates are reached to arbitrarily redefine random coordinates at a random speed.
RW	The RW is an extremely simple model of RWP. The mobile node goes on to visit randomly until established time initially.
RD	The difference is that RD is moving in random way at given random direction (0 and 2π) at range between Vmin and Vmax speed until a defined border area reached.
RP	It differs from RWP it followed the sense of the path.
RJ	It is a model form based on the RW model. The period of the navigation simulation continues until a number of pixels for a random, separate path is completed randomly when hitting the appropriate point
Gravity model	Summaries the total of trips between two areas and can call them productions and generations.
Log-linear model	Based on the statistical estimation of flows, that includes the characteristics of origin and destination zones.
Radiation model	The radiation model explains the quantity of trips made by two areas as the selection process consisting of the selection of jobs and the selection of jobs.
Inference Route Choice	Identify the route that the subject most likely took from his/her pattern of mobile usage.

5. Simulation and Modeling of Mobility

Data from migration flows capture the number of migrants visiting a country ($inflow$ and $outflow$) for a particular period, a year or month. Migration data are vital for understanding the patterns of international migration, how different factors and policies can relate to flows in the countries of origin and destination. Therefore, in this section will explain some of methods, algorithms, and datasets to analyze and capture the number of migrants in different directions of migration (groups and individual) and movement**.**

5.1 Data description

Actually, in [3][2], CDRs data from $1{,}8{,}000{,}000$ cell phones were obtained customers in Portugal, from April 2006 to the following year: March 2007. The data obtained from one of Portugal biggest telecom operators. The caller ID, Callee ID and caller link Cell tower ID and Callee linked cell tower ID are included in any record during the call size, and time[3]. The datasets analyzed include the CDRs in a European country of October 2006 for cell telephone contact for customers of a mobile phone operator (call it O). O has a share market in the country of around 30[36]. In Mexico City, a six-month model used credit card data. Then, action forms are purchasing categories in these datasets such as tolls, quick food restaurants, pharmacies [32]. So, the CDRs are the most common cell phone data gathering information on each user call or text. They show the population presence on a national basis is estimated scale, and form an interesting complement for any given time Statistics on population and migration [37]. Resident and workplace Inferences especially interested to travel, where home and work place or school is the origin and destination. Therefore, the first details we needed before we could examine the route were the place of each subject home and the workplace. Choices of routes until every subject are at his or her place of business, a collection of possible route choices inferred from the use of the Google Maps API Addresses. It is asked for potential routes via the API for each topic, choices by vehicle and public transit [2]. While, in [7] the main data sets are two sets anonymized for Mexican citizens over five months in 2015: CDRs and CCRS records. For more than 800 users at the Technical University of Denmark (DTU) the project obtained mobile sensing data from Smartphone. The data sources contain GPS, Bluetooth, SMS, telephone contacts, Wi-Fi, and communication via Face book [9]. The study evaluate the smart card transactions for such a study over a period of three months of 626 public transit users anonymous from London, United Kingdom [38]. The Face book Network can be used as a non-traditional data source to estimate the "migrants," because it is the only wide-spread social network classifying users on the basis of their previous domicile to our best knowledge[15].

In addition, focuses on a collection of real-world tweets in the city of London metropolitan area give a good result in extracting spatio-temporal features [39]. Additionally, in [21], the analysis has used fully the city-scale data obtained from the Automated Fare Collection (AFC) system. To determine motifs, three separate datasets are used: a Paris survey and data billing for cell phones and a Chicago survey [19]. Also, Official stats obtained by organizations are census data and surveys, the following are collected in census data relating to immigration; ethnicity, country of birth, last residence and stay [22]. Surveys do gather information on immigrant flows, stocks, and more frequently than census data are collected. In comparison to census data, information about families, labor markets or neighborhoods typically collected, depending on its primary objective. Recent studies have shown a growing use of social big data for the study of immigrant travel [22]. This classification can include different types of data. It can also be social media data, online services, and smart phones, transaction records for the store and on and on. Furthermore, Twitter data helps researchers to target immigrant flows and stocks and to track recent patterns prior to release of the official statistics. In the context of international migration trends, data from Skype Ego networks can also be used[22]. In addition, OSN data may contribute to study different subjects, for example social integration, language usage, local language changes, immigrant sentiment, etc. [22]. The data collection from human movement named mobility datasets images is mostly taken as space events from two different viewpoints (independent space and time) and as trajectories of individuals (e.g. photography owners)[28]. As well as, in [40] using data of two types: (1) Nanjing Metro intelligent card data and Nanjing public cycle sharing services automatic fare collection (AFC) system; (2) the Nanjing City GIS data that includes a land-use layer map and network of highways. Finally, one of the most important aspects of this review is the mobile phone data CDRs, as the following Figure 6 shows the details and uses of the data to be added to the mentioned aspects and algorithms.

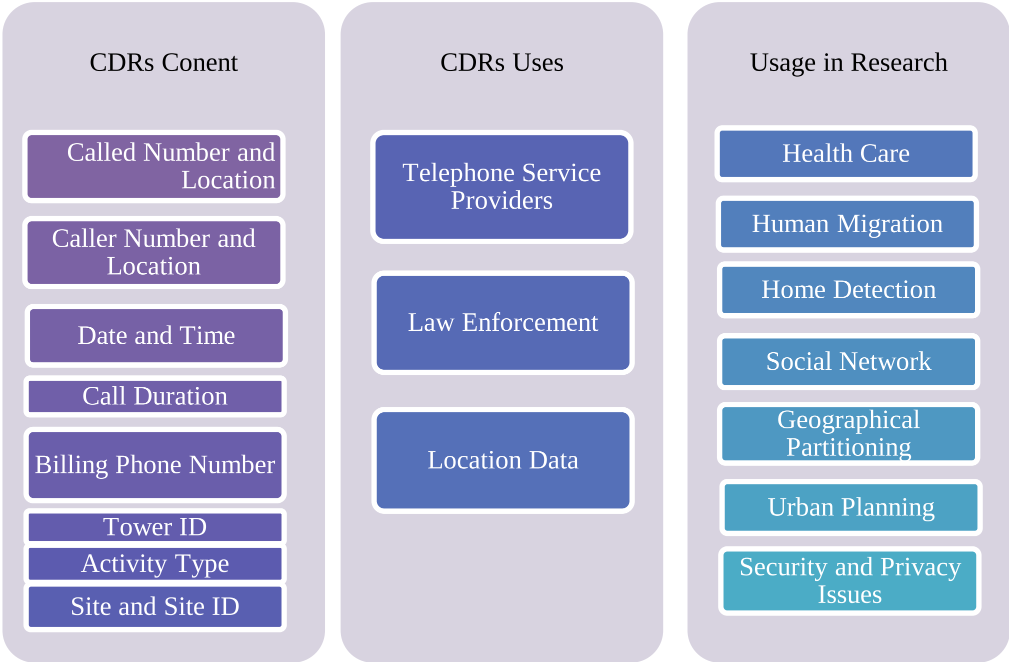

Figure 6-Common Principles of CDRs

5.2 Methods

Generally, the migration involves moving people for a certain amount of time between two locations. However, an accurate description for migration is difficult to determine. The problem is how long a person has to move and how long he or she has to keep away from the destination. As well as, migration trips focus about changing the residence, so firstly need to determine the locations of live or residence for each time and identify if there was change. That is applying for at least 14 months. Whereas, [41] used the same method, locate the most frequently towers used at night in range $10\ PM\ and\ 7\ AM$ as the approximation location of live. Then, noticed the call just at night every month. Each of these subjects has variances results of people residence locations, different from one to 14 months or more. At last, only one place has detected since 14 months which there was no change of residence [3]. Also, there were some people who moved inside the same city, that's addition to the migration flow between regions [3]. Also, the Density-based spatial clustering of applications with noise in certain areas, which combines points, which narrowly connected (with several nearest neighbors) as outliers, which lie alone in low-density regions (whose nearest neighbors are too far away). As well as, [28] the identification of user locations is one of the major tasks in mobility data analysis and a required preparatory stage for many applications. Correct detection of these private positions is an important issue in a huge range of methods, which usually includes grid partitioning or generic clustering algorithms like DBSCAN or OPTICS. So, the study of the individual mobility of users in terms of their individual locations, that is to say places or regions where they stop for all manner of activities is a fundamental task in mobility data analyses. Therefore [28] proposed TOSCA (two-step clustering algorithm) the concept behind TOSCA lies in the necessity of efficiently detecting user locations without compromising the clustering efficiency and, above all, without any feature tuning phases. Consequently, provision on mobile user's future positions is a useful challenge, which enables us to operate an array of various applications that require this sort of information efficiently. So, [28] MyWay is a system that predicts a user exact future positions while he/she flows. According to the framework of personal data analysis, MyWay indicators leverage individual structural behaviors, the individual systematic behaviors and a hybrid mix of all users of the method. MyWay takes advantage of the ability to make use of two levels of ability (individual and the group). Whereas, each user day-to-day production of mobility data such as GPS and GSM via mobile devices helps to improve daily living by custom mobility services. The predictability of human behavior is the foundation of these mobility services. In addition,[28] Routinary Actions Mobility Agenda (RAMA) is a method that pulls the customer own mobility model and makes use of the approach to replicate the user own mobility agenda which represents the expected locations where the user executes his/her work all day long. Table 2 represents the summaries for these tools.

Table 2- Tools of algorithms

Tools	Highlights
TOSCA	Detecting user location without clustering efficiency
MyWay	Predict the next location of user (individual) when moving
RAMA	Used to extract the customers own mobility to reuse it in another location.

If there are many points, or they can be represented by people based on the methods used in analyzing the movement of people, it is very important to distribute or separate these points into several groups to benefit from each group on a specific topic that is determined by the need for it. As well as, [36] proposed K-means Clustering where the k-means algorithm divides objects into groups of N clusters within the distances between them, if the character of each object is distinguish by a features with N elements generally use Euclidean distance between the vectors as an interval among objects. Indeed, modeling of travel demand aims at establishing explicitly the spatial distribution of travel through a suitable zone system. As much as, [3] used a travel demand model, a model of tour production then used to predict each county total productions and attractions for the journey, then a model for the distribution of trips from the model for the development of trips can used within destinations. In addition, [32] offer methodologies for estimating parameters, simulation and hypothesis testing of the Multivariate Periodic Hawkes Process (MPHP). Also, introduce a Maximum a Posteriori (MAP), Expectation-maximization (EM) to estimate parameters for MPHP that are appropriate for large datasets and allow the use of introductions to provide the necessary regulation. Recently, the artificial intelligent and networks with mobile network data play an important role in the world by analyzing movement people and migration. Although,[10] using Street Smart to estimate personal vehicle fuel consumption. The energy required by vehicles for various movement conditions has been comprehensively established [10]. On the other hand, [8] explained the proposed methods with two separate data sets from urban areas Vienna, Austria and Boston, USA, and compare the results. For Vienna cell phone traffic used for the analysis, including the network connections of the devices idle, For Boston, the analysis relied on anonymous call detail records CDRs. Although, there is a two-fold contribution. First, suggest a method to detect stays robustly and turn raw mobile phones into a series of trips and activities, including estimates of arrival times and stay times. Secondly, an unsupervised learning approach is proposed in order to expose mobile activity patterns[8]. In order to present a simple method of urban movement, two models have a separate option for making mobility decisions by an individual: what places to visit, how much time to spend in each of these places [38].

A variety of techniques used to deliver the message by using smart phone records for six months for three countries Spain, Portugal and France. It had been defined each criterion of the experiments, where RAN used depth first search (DFS) in a routing algorithm, random routing is used as a base-line comparison, so that infinite loops can effectively be avoided. GEO, this technique gives the message to the friend nearest to the target geographically.

DEG, in this scenario, the message will send to the friend of most of the members. COM, Groups within the network is detected in order to imitate social attributes (school, work) [11]. It is important to detect the start point and end point of people in analyzing some data. Thus, [26] studied the detection of traffic flow by delegated traffic generation, in order to detect traffic on the road network; an OD matrix containing the principal stay points for users in time-series was first developed. One of the key steps of the suggested solution was to extract main areas of activity from the Smartphone use of subscribers through activity hubs extraction through DBSCAN clusters. This accomplished using the clustering. OD Matrix Trip Generation, which has agreed based on our OD matrix generation, namely home and workplace, two key areas of operation per user. The routes from Open Street Maps (OSM) between their origin and their destination have chosen by travel generation, route choice, and traffic allocation. The next step in a pipeline was to predict traffic from a fixed time for the specified place, using Multi-Layer Perception Classifier (MLPC).
Furthermore, [27] built a methodology and SIM to and from their homes and workplaces, in this case at the DSD level. The first step in identifying Divisional Secretarial Division (DSD) places at home and at work was to evaluate the time period between the places in which an average person would work in that locations. A three-step procedure was used to extract mobility chart data from the datasets by means of a typical working day: measured the average location, every hour of the day (24 points total) is a (latitude and longitude) for each human SIM. The distance traveled during each individual SIM is achieved by measurement of the distance from two consecutive average hourly positions in the Euclidean (the straight line). For all the SIMs that displayed movement over a species hour and traced the results in a Figure 7, the average hours-wise distance measure was achieved.

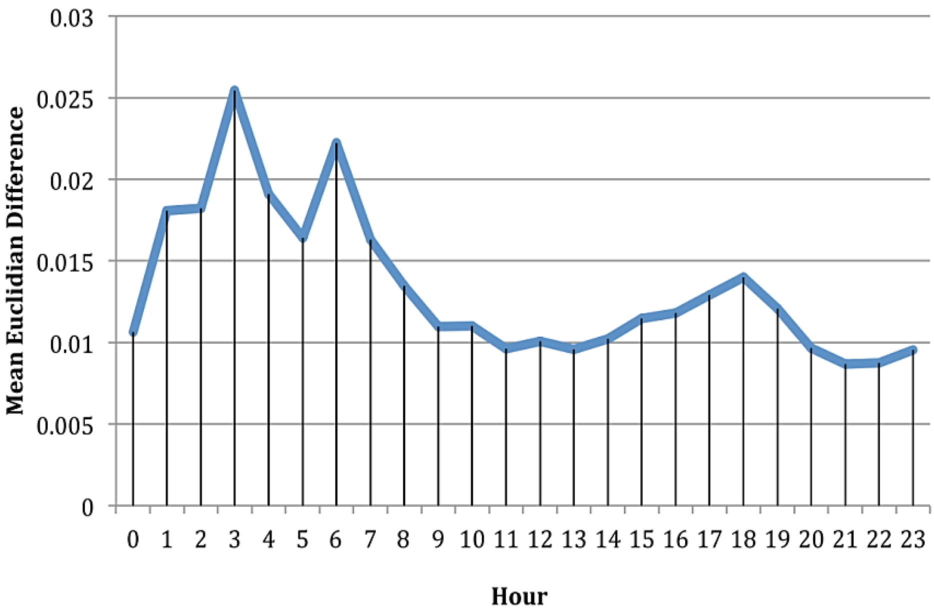

Figure 7-Average SIM mobility in one day for each hour [27]

Consequently, detection of physical activity is the digital recording of person movements. Human movement can collect by any camera, and input for processing in a computer vision engine. In [39], they explained some main steps of methodology presented to detect the human paths from geo-referenced posts of online communication network. Firstly, the geo data from twitter have to be collected. Then, discover the semantic site. Next, extract the tracks and then mining the recurring travel paths using the chain-mining pattern. Finally, extract spatial/temporal features from all frequent trends in order to capture the factors driving user motion. Moreover, in [2] there is another method to measure the shortest distance between the route segment and the visits cell tower locations instead of the distance to way-point. Voronoi Cell-based approach for Voronoi is a popular diagram for dividing space into thread based on a number of predetermined points in space. The generated cells Voronoi that define shown in Figure 8 Countrywide service coverage areas for all 6511 cell tower locations. Filter of noise although traveling to/from other places aside from home and workplace constitutes a majority of individual travels. These trips are present in connectivity logs as well as from CDRs data. These trips are not available for travelers. However, since major focus is on travel, noise can consider in relation to the mobile network connectivity of those non-travel trips.

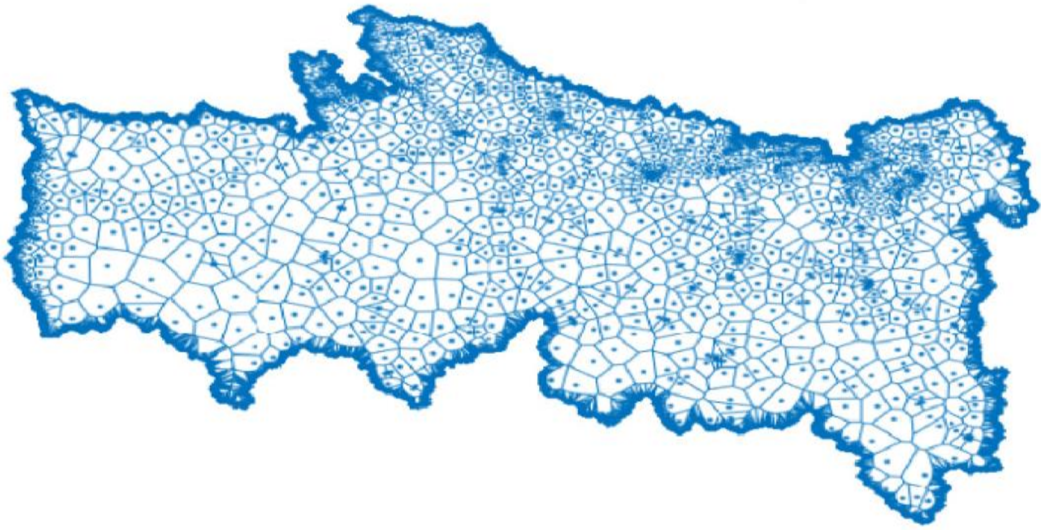

Figure 8- Example of Voronoi Cell [2]

Accordingly, by using various migration phases such as *journey*, research migration flows, and stocks, big data helped with migration, giving examples in which big data could affect. The second phase describes *stay*, i.e. integration of migrants in the country of destination. The final phase concerns migration effects on the countries of origin and the *return* of migrants. Furthermore, commuting is regular repeated journeys between the place where the traveler leaves his home community and the place of residence or study at. As well as, [20] concluded the commuting route, presents two methods: method A is to choose the optimal path, at minimum distance, for the positions of cells visited or used by the user. This approach helps users to communicate from many points during their mobile path during the year via their mobile phones. Method B: maximum overlap A new method was introduced that interpolates and extrapolates waypoints to normalize the space between them in order to solve the problem of inconsistent waypoint distribution along the road. That has done by actually making use of grids and generating a new data point that represents a waypoint at the center of the grid that leads along the route. Additionally, a framework for simulating spatio-temporal patterns of human mobility is DITRAS (Diary based TRAjectory Simulator) in Figure 9. The time features of human mobility are separated by DITRAS through two operating stages. DITRAS operates In the first phase, DITRAS uses a diary generator to produce a diary, that is, an algorithm that captures human mobility templates by specifying the time of arrival and time spent in each of the places that each person visits. Through a mechanism to explore mobility locations, DITRAS will transform the mobility diary into a track in the second step[42]. Moreover, [15] examines how feasible non-traditional sources of data can use to fill existing migration statistics gaps. For this, data from Facebook advertising platform is anonymous and publicly available. If you previously had lived at Country and now live in a different country, your advertising platform classifies its User as "Live in Country". Based on statistics on

Facebook Network users, who have lived abroad (Facebook, Instagram, Messenger and the Audience Network), and apply sample bias correction methods, calculated the number of "migrants" on Facebook in 119 countries and in 2 periods by $sex, age\ and\ gender$ in 119 countries and previous residential country. The following in the preparations of data for Facebook Network (FN): first, clean the data for FN; second, the lower or higher boundaries of the rounded response to the Facebook API; third, the estimation of the rate of penetration in FN. Finally, Table 3 summarizes the explained methods and approaches in this review.

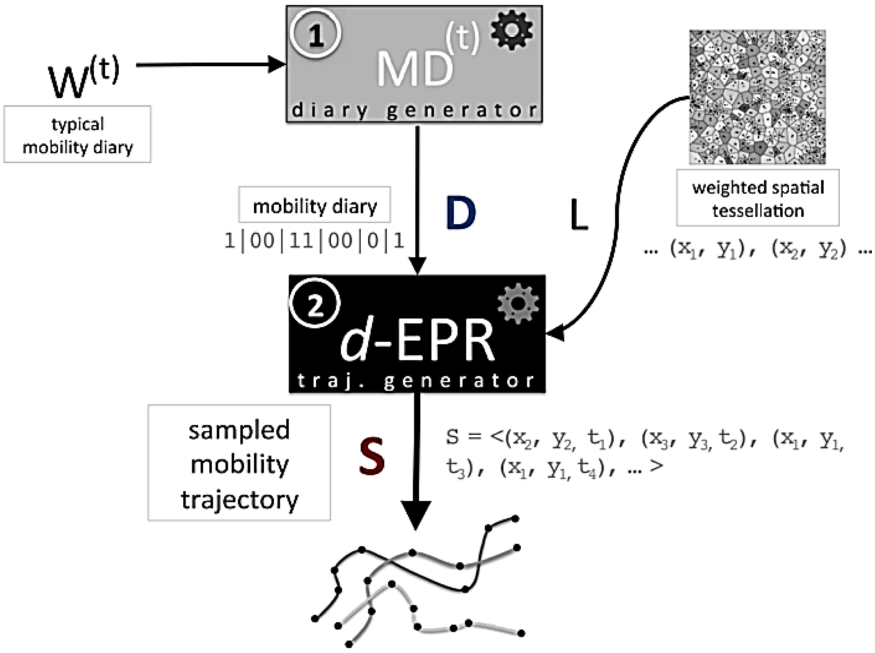

Figure 9-DITRAS Framework overview [42]

Table 3- Summery of the methods and approaches explained in the review

Highlights	**Datasets**	**Results**
Detecting home and work location by using coordination of user [3][41].	CDRs	It has concluded that 80% of the travel reaches only 20 km from the nearest social links at home. In a geo-social range of 45 km, this rises to 90%. The workplaces of ties also have a similar result
K-means algorithm divides objects into groups of N clusters within the distances between them by using Euclidean distance [36].	CDRs	Clustering used to divide the people into groups and comparing between them.
This recorded, at the individual level, the spatial clustering of human activity, the structure of interdependence and the co-excitement between various activities, and the periodic effects of every week rhythms [32].	Credit Card Data	In overall, results show the efficiency and frequency of tasks priorities in clarifying credit card transactions activities
Street Smart to estimate personal vehicle fuel consumption [10].	CDRs , GPS	Approximate fuel consumption for mainly three traffic conditions times. The hours are from morning to 9 am, from midday to midday 12 to 13 pm and from evening to weekdays 7-18 pm. For each edge of the road network, fuel consumption per car can calculate based on every matched speed profile.
Suggested a method to detect stays robustly and turn raw mobile phones into a series of trips and activities. And	CDRs, cell phone traffic	The method produces robust results over several days, while working days and weekends show different patterns that reflect the well-known

an unsupervised learning approach proposed in order to expose mobile activity patterns [8].		differences in journey behaviors.
Introduce the spatiotemporal patterns for the mobility of individuals in cities using smart metro card transaction data [38].	Smart card transaction	Presented a simplistic movement pattern in order to predict where people visit using the popularity of places in the city as an exchange parameter among people.
Several techniques of data mining and artificial intelligent have focused on traffic flow measurement and traffic flow prediction for a number of locations with mobile data use over time [26].	Mobile phone dataset	The results showed that the built method is more efficient than the actual traffic flow rate for measuring and predicting traffic flow demand for particular locations.
Utilizing the huge mobile data for transport planning to solve the economic problems faced by countries in transport regulations [27].	Mobile network big data	The analysis suggests that population movements can be complemented by uncommon surveys with a larger rate and a higher spatial resolution than was previously possible.
Investigate these movements for people and behavior in the community [39].	Smartphone GPS Interface	The analysis the most popular routes followed by people voyages and the spatial-temporal distribution of those journeys were provided with useful insights regarding common movements followed in the city.
Used Voronoi and waypoints methods to measure the shortest distance between the route segment and the visits cell tower location [2].	CDR	Based on cell tower locations, these methods could indicate service coverage zones and other techniques.
Provides a way to extract information from mobile phone communication logs from commuting flow and route selection [20].	CDR	By using visualization tool that is useful to understudied the movement and pattern of people in the whole day.
Used big data to discuss the immigration phases such as stay, return and journey [22].	Facebook, Twitter, Census data and Surveys, Skype, CDR.	Three phases have discussed to understand the migration phenomena in several types of data sets.
As the only social network to classify their users based on their previous residences, Facebook have chosen to use as non-traditionally data source for "migratory" estimation [15].	FN	Corrected FN migrants aged $15-64$ are strongly related to the migrant supplier of official sources ($R2 = 0.735, p < 0.001$) for each gender, country of former residence and country of current residency.
DITRAS a framework extract the trajectory of people by detecting the time arrival and time time spent in every space[42].	Mobile phone data, GPS	The proposed algorithm reproduces in the most precise way the statistical characteristics of actual trajectories, making a step towards an understanding of the origin of human spatio-temporal mobility patterns.

6. Conclusions

This paper focused on several types of important and useful researches in the field of human mobility and migration, and revealed multiple results using multiple models including models of mobility for actual traffic movement; which it depends on many factures as mentioned in previous sections. Moreover, the model of radiation for detecting the position of every passenger. In addition, the time and network randomly affected the movement and directions such as RD, RW models. In addition, CDRs can be useful data as explained in this review because they have several information about users to analyze the movement between cites or inside a single city like socio analysis, socio economic, home detection and so on. Furthermore, these data are secure; give approximately the users locations, although the cell phone signal would be disappeared when the cell phone is shouting down. There is another data type such as credit card data for online uses, smart phone data, Facebook data to estimate the migration as mentioned before, GPS, OSN, and so on. However, this review described several movement methods or algorithms for analysis datasets such as TOSCA, Voronoi CellBased, K-means Clustering, MPHP, MyWay, and RAMA. Nevertheless, the most important one is focus on migration trip of residence. Hence, for future work, it intended to investigate a dataset composed of CDRs for flow migration analysis among several countries with several networks data. The intended dataset has several indicators such as id number caller and called, activity type, geometry, coordinates. In addition, it will focus on some modeling of spatiotemporal migration flows such as Gravity model, Log- linear model, and Radiation model.

References

[1] S. F. Behadili, "Adaptative modeling of urban dynamics with mobile Suhad Faisal Behadili To cite this version : HAL Id : tel-01668513 Adaptive Modeling of Urban Dynamics with Mobile Phone Database Présentée et soutenue par," no. November 2016, 2017.

[2] P. Sakamanee, S. Phithakkitnukoon, Z. Smoreda, and C. Ratti, "Methods for inferring route choice of commuting trip from mobile phone network data," *ISPRS Int. J. Geo-Information*, vol. 9, no. 5, pp. 1–23, 2020, doi: 10.3390/ijgi9050306.

[3] S. Hankaew, S. Phithakkitnukoon, M. G. Demissie, L. Kattan, Z. Smoreda, and C. Ratti, "Inferring and Modeling Migration Flows Using Mobile Phone Network Data," *IEEE Access*, vol. 7, pp. 164746–164758, 2019, doi: 10.1109/ACCESS.2019.2952911.

[4] M. Vanhoof, F. Reis, T. Ploetz, and Z. Smoreda, "Assessing the quality of home detection from mobile phone data for official statistics," *J. Off. Stat.*, vol. 34, no. 4, pp. 935–960, 2018, doi: 10.2478/jos-2018-0046.

[5] M. Vanhoof, F. Reis, Z. Smoreda, and T. Ploetz, "Detecting home locations from CDR data: introducing spatial uncertainty to the state-of-the-art," 2018, [Online]. Available: http://arxiv.org/abs/1808.06398.

[6] M. Vanhoof, T. Ploetz, and Z. Smoreda, "Geographical veracity of indicators derived from mobile phone data," no. Figure 1.

[7] S. Xu, R. Di Clemente, and M. C. Gonzalez, "Big Data Recommender Systems - Volume 2: Application Paradigms," *Big Data Recomm. Syst. - Vol. 2 Appl. Paradig.*, no. November, 2019, doi: 10.1049/pbpc035g.

[8] P. Widhalm, Y. Yang, M. Ulm, S. Athavale, and M. C. González, "Discovering urban activity patterns in cell phone data," *Transportation (Amst).*, vol. 42, no. 4, pp. 597–623, 2015, doi: 10.1007/s11116-015-9598-x.

[9] A. Cuttone, S. Lehmann, and M. C. González, "Understanding predictability and exploration in human mobility," *EPJ Data Sci.*, vol. 7, no. 1, 2018, doi: 10.1140/epjds/s13688-017-0129-1.

[10] A. Kalila, Z. Awwad, R. Di Clemente, and M. C. González, "Big data fusion to estimate urban fuel consumption: A case study of Riyadh," *Transp. Res. Rec.*, vol. 2672, no. 24, pp. 49–59, 2018, doi: 10.1177/0361198118798461.

[11] [11] C. Herrera-Yagüe *et al.*, "The anatomy of urban social networks and its implications in the searchability problem," *Sci. Rep.*, vol. 5, no. June, 2015, doi: 10.1038/srep10265.

[12] [12] M. Vanhoof, W. Schoors, A. Van Rompaey, T. Ploetz, and Z. Smoreda, "Comparing Regional Patterns of Individual Movement Using Corrected Mobility Entropy," *J. Urban Technol.*, vol. 25, no. 2, pp. 27–61, 2018, doi: 10.1080/10630732.2018.1450593.

[13] M. Tizzoni *et al.*, "On the Use of Human Mobility Proxies for Modeling Epidemics," *PLoS Comput. Biol.*, vol. 10, no. 7, 2014, doi: 10.1371/journal.pcbi.1003716.

[14] S. Lokanathan, G. Kreindler, N. D. de Silva, Y. Miyauchi, and D. Dhananjaya, "Using Mobile Network Big Data for Informing Transportation and Urban Planning in Colombo," *SSRN Electron. J.*, no. January, 2014, doi: 10.2139/ssrn.2526642.

[15] S. Spyratos, M. Vespe, F. Natale, I. Weber, E. Zagheni, and M. Rango, "Quantifying international human mobility patterns using Facebook Network data," *PLoS One*, vol. 14, no. 10, 2019, doi: 10.1371/journal.pone.0224134.

[16] Y. Xu, S. L. Shaw, Z. Zhao, L. Yin, Z. Fang, and Q. Li, “Understanding aggregate human mobility patterns using passive mobile phone location data: a home-based approach,” *Transportation (Amst).*, vol. 42, no. 4, pp. 625–646, 2015, doi: 10.1007/s11116-015-9597-y.
[17] M. G. Demissie, S. Phithakkitnukoon, L. Kattan, and A. Farhan, “Understanding human mobility patterns in a developing country using mobile phone data,” *Data Sci. J.*, vol. 18, no. 1, 2019, doi: 10.5334/dsj-2019-001.
[18] R. Ahas, A. Aasa, S. Silm, and M. Tiru, “Daily rhythms of suburban commuters’ movements in the Tallinn metropolitan area: Case study with mobile positioning data,” *Transp. Res. Part C Emerg. Technol.*, vol. 18, no. 1, pp. 45–54, 2010, doi: 10.1016/j.trc.2009.04.011.
[19] C. M. Schneider, V. Belik, T. Couronné, Z. Smoreda, and M. C. González, “Unravelling daily human mobility motifs,” *J. R. Soc. Interface*, vol. 10, no. 84, 2013, doi: 10.1098/rsif.2013.0246.
[20] T. Jundee, C. Kunyadoi, A. Apavatjrut, S. Phithakkitnukoon, and Z. Smoreda, “Inferring commuting flows using CDR data: A case study of Lisbon, Portugal,” *UbiComp/ISWC 2018 - Adjun. Proc. 2018 ACM Int. Jt. Conf. Pervasive Ubiquitous Comput. Proc. 2018 ACM Int. Symp. Wearable Comput.*, pp. 1041–1050, 2018, doi: 10.1145/3267305.3274159.
[21] X. Tian and B. Zheng, “Study Group Travel Behaviour Patterns from Large-Scale Smart Card Data,” *Proc. - 2019 IEEE Int. Conf. Big Data, Big Data 2019*, pp. 1232–1237, 2019, doi: 10.1109/BigData47090.2019.9005575.
[22] A. Sîrbu *et al.*, “Human migration: the big data perspective,” *Int. J. Data Sci. Anal.*, no. April, 2020, doi: 10.1007/s41060-020-00213-5.
[23] B. Yelure and S. Sonavane, “Performance of routing protocols using mobility models in VANET,” *Adv. Intell. Syst. Comput.*, vol. 1162, no. January, pp. 272–280, 2021, doi: 10.1007/978-981-15-4851-2_29.
[24] T. Yasseri, R. Sumi, and J. Kertész, “Circadian patterns of Wikipedia editorial activity: A demographic analysis,” *PLoS One*, vol. 7, no. 1, 2012, doi: 10.1371/journal.pone.0030091.
[25] J. Candia, M. C. González, P. Wang, T. Schoenharl, G. Madey, and A. L. Barabási, “Uncovering individual and collective human dynamics from mobile phone records,” *J. Phys. A Math. Theor.*, vol. 41, no. 22, 2008, doi: 10.1088/1751-8113/41/22/224015.
[26] M. Saliba, C. Abela, and C. Layfield, “Vehicular traffic flow intensity detection and prediction through mobile data usage,” *CEUR Workshop Proc.*, vol. 2259, no. December, pp. 66–77, 2018.
[27] S. Lokanathan, G. Kreindler, N. De Silva, Y. Miyauchi, D. Dhananjaya, and R. Samarajiva, “The Potential of Mobile Network Big Data as a Tool in Colombo ’ s Transportation and Urban Planning,” pp. 1–14, 2013.
[28] G. Andrienko *et al.*, *(So) Big Data and the transformation of the city*, vol. 340330706, no. April. Springer International Publishing, 2020.
[29] D. Zordan, B. Martinez, I. Vilajosana, and M. Rossi, “On the performance of lossy compression schemes for energy constrained sensor networking,” *ACM Trans. Sens. Networks*, vol. 11, no. 1, 2014, doi: 10.1145/2629660.
[30] M. Bilgin, “Novel random models of entity mobility models and performance analysis of random entity mobility models,” *Turkish J. Electr. Eng. Comput. Sci.*, vol. 28, no. 2, pp. 708–726, 2020, doi: 10.3906/elk-1904-102.
[31] N. B. Ponieman, C. Sarraute, M. Minnoni, M. Travizano, P. Rodriguez Zivic, and A. Salles, “Mobility and sociocultural events in mobile phone data records,” *AI Commun.*, vol. 29, no. 1, pp. 77–86, 2016, doi: 10.3233/AIC-150687.
[32] S. Xu, S. Morse, and M. C. González, “Modeling Human Dynamics and Lifestyle Using Digital Traces,” pp. 1–15, 2020, [Online]. Available: http://arxiv.org/abs/2005.06542.
[33] J. Broch, D. A. Maltz, and D. B. Johnson, “A Performance Comparison of Multi-Hop Wireless Ad Hoc Network Routing A Performance Comparison of Multi-Hop Wireless Ad Hoc Network Routing Protocols,” no. March 2001, 2012, doi: 10.1145/288235.288256.
[34] D. A. Maltz, “Dynamic Source Routing in Ad Hoc Wireless Networks in Ad Hoc Wireless Networks,” no. May 1999, 2012.
[35] F. Simini, M. C. González, A. Maritan, and A. L. Barabási, “A universal model for mobility and migration patterns,” *Nature*, vol. 484, no. 7392, pp. 96–100, 2012, doi: 10.1038/nature10856.
[36] A. Stoica, Z. Smoreda, and C. Prieur, “A Local Structure-Based Method for Nodes Clustering: Application to a Large Mobile Phone Social Network,” no. August 2015, pp. 157–184, 2013, doi:

10.1007/978-3-7091-1346-2_7.

[37] M. Vanhoof, C. Lee, and Z. Smoreda, “Performance and Sensitivities of Home Detection on Mobile Phone Data,” *Big Data Meets Surv. Sci.*, no. April 2019, pp. 245–271, 2020, doi: 10.1002/9781118976357.ch8.

[38] S. Hasan, C. M. Schneider, S. V. Ukkusuri, and M. C. González, “Spatiotemporal Patterns of Urban Human Mobility,” *J. Stat. Phys.*, vol. 151, no. 1–2, pp. 304–318, 2013, doi: 10.1007/s10955-012-0645-0.

[39] C. Comito, D. Falcone, and D. Talia, “Carmela Comitoa,∗, Deborah Falconeb, Domenico Talia,” *Pervasive Mob. Comput.*, vol. 33, no. June, pp. 91–107, 2016, doi: 10.1016/j.pmcj.2016.06.005.

[40] D. Lei, X. Chen, L. Cheng, L. Zhang, S. V. Ukkusuri, and F. Witlox, “Inferring temporal motifs for travel pattern analysis using large scale smart card data,” *Transp. Res. Part C Emerg. Technol.*, vol. 120, no. September, p. 102810, 2020, doi: 10.1016/j.trc.2020.102810.

[41] S. Phithakkitnukoon, Z. Smoreda, and P. Olivier, “Socio-Geography of Human Mobility : A Study Using Longitudinal Mobile Phone Data,” vol. 7, no. 6, pp. 1–9, 2012, doi: 10.1371/journal.pone.0039253.

[42] L. Pappalardo and F. Simini, *Data-driven generation of spatio-temporal routines in human mobility*, vol. 32, no. 3. Springer US, 2018.