DOI: 10.24996/ijs.2022.63.10.32

ISSN: 0067-2904

Gravity Model for Flow Migration Within Wireless Communication Networks

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

Received: 14/9/2021 Accepted: 31/10/2021 Published: 30/10/2022

Abstract

This paper investigated an Iraqi dataset from Korek Telecom Company as Call Detail Recorded (CDRs) for six months falling between Sep. 2020-Feb. 2021. This data covers 18 governorates, and it falls within the period of COVID-19. The Gravity algorithm was applied into two levels of abstraction in deriving the results as the macroscopic and mesoscopic levels respectively. The goal of this study was to reveal the strength and weakness of people migration in-between the Iraqi cities. Thus, it has been clear that the relationship between each city with the others is based on $inflows$ and $outflows$ of mobile people. However, the COVID-19 effects on the people's migration needed to be explored. Whereas the main function of the gravity model is to clarify the migration flows through modeling spatial interaction. This was implemented using Python scripting language. It is concluded that the gravity model has a powerful ability to analyze the movement of people between cities. According to the mean of result between governorates, showing that the highest attraction was between Babil and Anbar governorates amounted to 99%, while the lowest attraction was between Wasit and Thi-Qar governorates with 85%, and the others ranged between $86\% - 98\%$.

Keywords: Gravity Model, Call Details Recorded, COVID-19, Migration, Iraq.

نمــوذج الجاذبيــة لهجـرة التدفـق ضمــن شبكــات الاتصــالات اللاسلكيــة

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

الخلاصة

حققت هذه الورقة في مجموعة بيانات عراقية من شركة كورك تليكوم للاتصالات في العراق، حيث تم تسجيل تفاصيل المكالمات لمدة ستة أشهر بين ايلول 2020 – شباط 2021. تغطي هذه البيانات 18 محافظة وهي تقع ضمن فتره كوفيد-19. تم تطبيق خوارزمية الجاذبية على مستويين من التجريد في استخلاص النتائج وهما المستويات المايكروسكوبية والمستويـات المتوسطة على التوالي. الهدف من هذه الدراسة هو الكشف عن قوة وضعف هجرة الناس بين المدن العراقية. وبالتالي، تم توضيح العلاقـة بين كل مدينة مع الآخرين بناءً على تدفقات الأشخاص الداخلين والخارجين للمدن. ومع ذلك، فقد تم استكشاف آثار

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

كوفيد–19 على الهجرة البشرية. حيث أن الوظيفة الرئيسيـة لنموذج الجاذبيـة هي توضيح تدفقات الهجرة من خلال نمذجة التفاعـل المكاني، لذلك يتم تنفيذ ذلك باستخدام لغة البرمجة النصية بايثون. ومن ثم، فقد تم استنتاج أن نموذج الجاذبيــة لديه قدرة قوية على تحليل حركــة الناس بين المدن. واستنـاداً الى المتوسط الحسـابي الناتج بين المحافظات، فقد تبين أن أعلـى جـذب كان بين محافظتي بابل والأنبـار بـلغ 99٪، بينما كان أقل جاذبيـة بين محافظتـي واسط وذي قار بنسبــة 85٪، والبعض الآخـر تراوحت بين 86٪ – 98٪.

1. Introduction

Human migration[1] includes movement between geographical locations and is represented by a group of people or moving individuals. Spatial interactions are reflected between motivations such as distance, time and a certain opportunity for movement, which affects the migration behavior and determines the choice of destination. The most popular models used for this type of study are gravity, radiation and log-linear models. So, in [2] the migration of the population is first classified into two categories, international and internal, and is often analyzed separately. In addition, the motives behind the perpetuation and initiation of both types of migration, such as economic, cultural and political migration, often differ in their importance.

Also, [3] Origin-Destination (OD) matrix can be used for flow migration or distribution of trips that has been studied for various reasons, such as prediction of traffic jams or human resources and diseases spread. Many trip models of trip distribution had been developed to rate the population flows extremely. The social media coming to light has led to a new probability in the transportation sector. Respectively, the gravity model is one of the important models in spatial interaction models. Based on the attractiveness of the thing in a place or time such as space-time. For example, presence of students in the university during the Day, the momentum is high and it is the opposite at night. Also, farmers are moving between governorates and regions according to the season to work in farms and agricultural places. This means gravity depends on two regions' production and attraction, as well the population and size of each city, meaning a big city has higher attraction. Spatial interaction modeling involves the analysis of flows from an origin to a destination, either over physical space (migration) or through abstraction space (telecommunication)[4].

However, the gravity model has some limitations. Thus, in[5], the fundamental gravity model is insufficient for simulating human mobility networks at smaller geographical and temporal dimensions. The distance variable may be the most important of the independent variables utilized in the basic gravity model. According to the findings, population density of census tracts does not have a significant influence in population migration across census tracts. Other factors such as the existence of points of attraction and land use characteristics may be more important in driving population movement across census tracts. So, in this study, the gravity model needed additional arguments not available in the data such as the destination population and the distance between the governorates and flow. Therefore some processing was required to address that issue.

Additionally, it can be said that gravity is a formula that predicts or calculates the amount of trade between two countries or two cities inside one country. In other words, it predicts the people migration between two regions. In addition, the mileage between source and destination would be from point to point[6]. Mobile data, or Call Details Recorded (CDRs), used for the study. This data is utilized in many studies, for example in socioeconomic, and it is characterized by its usefulness in terms of use. It contains calls information and its limitations is that it does not give accurate details due to privacy, but it is quick to obtain. It is available and use in a lot of research indicates its importance in the study of movement between cities, the movement of individuals and in indicating the movement of countries and cities economically and socially when festivals and celebrations and the Hajj period take place. As these occasions are very important to clarify the migration and many other huge events. So,[7] used the mobile phone data of a set of anonymous, billed records covering more than 10 million subscribers within a single European country over several years to understand if the model showed people per-hour journeys as well as mobility. In addition, interregional US tax income data may also be used from the Daily Income Statistics division of the Internal Revenue Service, which maintains all data or records of individual tax income forms each year. The outbreak of Coronavirus started in China, and quickly dispersed to countries in which migrants from South Asia concentrated, among them the Gulf, Europe and North America. Migrants came back home as the countries that migrants work in shut down their economies to protect their citizens. The COVID-19 virus spread to South Asia at the beginning of 2020, which has negative significant external costs in the migrant countries of origin. International migration was a major source of growth in South Asia, and a source of opportunities with expectation to be continued. It has been noted that 1 standard deviation is predicted in previous international out-migrants in India and Pakistan to increase by 48% in cases per capita, which is relative to a cross-sectional average in districts in each country[8]. Indeed, for two years, the COVID-19 virus became one of the main factors for changing and determining the type and time of human movement according to the indicators of the crisis cell and the Ministry of Health in Iraq. Therefore, in December 2019, the outbreak in public health, and society as a whole was impacted by the new Corona-Virus disease SARS-CoV-2, also known as COVID-19. Therefore, human mobility has received extensive study in several disciplines, including geographic location, transfer, urban design, physics, computing and human health. With the fast development of information and communication technologies (ICT) and GPS embedded devices, also the high-speed mobile telephone data is an unprecedented way to track human paths, which is advantageous for deep investigations through human mobility patterns[9].

The rest of the paper is organized as follows: section two describes the related work. Section three focuses on the spatial interaction based models and the application of the gravity model. It also focuses on the observed dataset. While section four describes and explains the produced results. Finally, section five concludes the results and presents the future work perspectives.

2. Related Work

In general, Gravity models are considered as the first modeling approaches to have been used in trade for the first time. The researchers in [10], have applied the gravity model at the first time to analyze the connection between trades, geography and technology renovation under academic qualifications. Then, tested how technology and geography affected the trades. The findings show that the advancement of information technology has reduced the impact of distance on commerce, since technical progress has made lengthy distances less significant than in the past.

Gravity model [11] for England and Wales system for both transit patterns and transport flows against the radiation model. To answer questions on human mobility and infrastructure, three different datasets were used: Composed of road, bus, coach, tube, train and plane, the complete multi-modal transport network in the UK Practices of commuting and migration at ward level for the UK. In this task the number of buses and trains working between all the cities was reduced to typical working days, 24 hours. First, the bus and train stops were assigned a ward area by spatial interceptions, and then this assignment was extended to the city areas. Finally, it depends on the scale of the problem at hand to select the best model. In

addition, one might imagine combining both models, for example using the gravity model for small distances $d < 3Km$ and the radiation model for long distances, in order to enhance accuracy.
While in [12], gravity was applied to trade using the Trade in Value-Added (TiVA) on the dataset of an Organization for Economic Cooperation and Development, which in turn distinguished between intermediate inputs and the final good in trade between the two parties. It covered 63 economies, including 35 countries as a member of the organization and 28 countries from outside the organization. This trade includes international trade and domestic trade, and thus the characteristics of both the origin and the destination are determined to control the size of the market. The gravity model structure performed well in representing bilateral commerce in both finished commodities and intermediate inputs. However, even after considering the dual nature of goods, the gravity model may perform poorly due to model misspecification and/or sample selection.

Additionally, [4] used an entropy-maximizing framework to describe the spatial interaction models as a “family” of four models, and consider gravity as the basic model. This work looked to assign the flows between the source (origin) and destination for finding the most possible arrangement of flows out of all possible configurations. Through using mutual optimization issues and inclusive information about $inflows$ and $outflows$ for each location also called “constraints”, they could find a family of four models unconstrained, production, attraction, doubly constrained.

Whereas, in [3] it was used on the Twitter dataset in NY city as the place of study indicates many trips available easily. Also, focuses on the commute trip, because it is stable temporally and represents the largest amount among populations. Meanwhile, the main accomplishment of this method is in threefold: First, discovering the chances of social media at different levels in trip distribution. Second, using machine learning technique to predict the distribution at census level. Third, compare the three models(gravity, random forest) and neural network models to detect the best model for trip distribution at census field levels. According to the results, the random forest model beats the classic gravity and neural network models. The random forest model revealed population, distance, number of Twitter users, and employment as the four most significant determinants of trip distribution.

According to [13], a gravitational distribution model was applied to Call Details Records (CDRs). The data were collected from 1.819.928 users in Portugal for 18 months for the period $from$ 2006 to 2007. Many models were obtained by computing the regression line between real migration flows within a specified time and two specific regions. Consequently, the gravitation and log-linear approaches with a straight distance (displacement) as their trip cost and districts centroids as reference points outperform the other alternatives.

A radiation model that takes district population into account outperforms the other radiation models but falls short of the gravity and log-linear models. Nonetheless, the study has significant drawbacks. To begin, only participants who had migrated once were inferred and investigated in this research. Second, due to the local non travel cost, only inter-district migration was considered in this study's trip distribution modeling. Third, there was a shortage of actual truth for the predicted migration trips, owing mostly to data scarcity. Moreover, in[14], it has been suggested that nonlinear Least Square (NLS) approaches as a major in the estimation process. Thus, since heteroscedasticity is frequently overlooked, the NLS was crucially arguable. The Ordinary Least Squares (OLS) was suggested, which is used for individual fixed effects in multiplicative estimation. Nevertheless, that approach increases the problems in zero trades, or gravity with thwart results for estimating the percent of heteroscedasticity. So, Poisson pseudo-maximum-likelihood (PPML) estimator has been applied as a new approach to solve this issue; and they have found that PPML deals with change of elasticity, false specification, and extra zeros, in contrast to previous fixed effect OLS and NLS capabilities.

In addition,[15] has examined the effect on internet latency on a broad scale of the COVID-19 pandemic. Latency is especially valuable not only because it has a deep impact on certain classes of applications, but also because it is an extremely good health status indicator for the network itself. Results, derived from an assessment of wide number of measurements, show that in terms of higher variability, the impact of increased online activity is relevant. The main changes were noted at night, the time of the day when most online actions link to entertainment. Outcomes for Italy, Spain, France, Germany, Sweden and the entire European Union show significant disparities, which may result from their network resilience and/or non-uniform limitations imposed by governments.

The research in [16] presented an analysis of a type of data (CDRs) for Kenya and the other for Namibia using gravitational spatial interaction models and some variables of the radiation model. As the study objective was the transmission of diseases, such as malaria, which resulted in the spatial accuracy of the models higher than the data itself. Similar models can be used widely for health care, mobility, and other areas. Radiation models include less parameters than gravitational models and estimate visits between overnight sites better for both nations studied. The results demonstrate that estimates for several parameters vary between nations and with geographical resolution, highlighting how poor flow data and spatial population distribution can affect model fit.

3. The Research Methodology

The goal of this research was to understand people's migration, and their transition from one place to another. The force of attraction symbolizes the migration that takes place. Analyzing the data by implementing the gravity model as a specialist and useful in this field of mobility gave several results. This procedure was implemented by using the SPatial INTeraction modeling (SpInt) module via python, which will be described in sections 3.1 and

3.1 Data Description

The flow migration needs specific types of datasets, and the mobile phone data is one of the important types in this subject. However, it is difficult to obtain because the needs to maintain privacy, and thus private information must be encrypted. Due to that, it is unable to identify the individuals and the beneficiaries from these subscriptions because it cannot distinguish them as people based on ages, names and genders. Therefore, the data retains a kind of deficiency, even though it gives useful information, such as speed and the to/from immigrant movements, where these immigrants have multiple reasons for movement. Accordingly, if there is no real data, you have to generate your own data. Some researchers worked to generate such data and specify it as they need. Also, there are several ways to generate data such as AI-GANs, Faker, MOSTLY.AI, Barnum packages, and there are several tools to generate several types of data as needed.

In this paper, a real mobile data sets used an Iraqi data set from Korek Telecom including CDRs, which was about 12 fields containing $Report\_Date$, A_Number, B_Number, $Call\_Description$, $Call\_Duration\_Seconds$, $Governorate$, $Site\_Id, Site$, $Longitude$, $Latitude$, $DateTime$, and $Destination$, in 10139733 lines for a period of $6\ months$ within from Sep 2020 to Feb 2021.Figure 1 shows the original fields of the observed data before preprocessing.

```
Index(['REPORT_DATE', ' A_NUMBER', ' B_NUMBER', ' CALL_DESCRIPTION',
       ' CAL_DURATION_SECONDS', ' GOVERNORATE', ' SITE_ID', ' SITE',
       ' LONGITUDE', '  LATITUDE', ' DATETIME', ' DESTINATION'],
      dtype='object')
```

Figure 1: Original Dataset

This dataset covers 18 Iraqi governorates. Bearing in mind that the positions pertain to the tower, they do not concern subscriber position, but it is an indication of the movement of groups of people within the areas covered by towers. In addition, it falls within the period of Covid-19, which is a well-known period that affected the movement of cities in Iraq specifically from the rest of the world.

Several preprocessing processes were implemented as integration operations and some software operations were performed on the data, because it was limited to the positions of people and coordinates. New data sets containing the coordinates of cities with names were merged with the Korek data set to obtain a complete data set. Figure 2 shows a sample of macroscopic data set after processing, while Figure 3 displays a sample of mesoscopic data set after processing.

	Src	Oi	Dst	Dj	Flows	city_id_x	lat_x	long_x	pop_x	city_id_y	lat_y	long_y	pop_y	Dij
0	Anbar	378221	Babil	92544	470765	1	32.5598	41.9196	1818318	2	32.4682	44.5502	2119403	15.283343
1	Anbar	378221	Baghdad	1489365	1867586	1	32.5598	41.9196	1818318	3	33.3152	44.3661	8340711	14.481997
2	Anbar	378221	Basra	54355	432576	1	32.5598	41.9196	1818318	4	30.5258	47.7738	2985073	19.623949
3	Anbar	378221	Diyala	29942	408163	1	32.5598	41.9196	1818318	5	33.7733	45.1495	1680328	14.731727
4	Anbar	378221	Dohuk	1429958	1808179	1	32.5598	41.9196	1818318	6	36.8632	42.9885	1326562	11.185936
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
301	Wasit	18395	Ninewa	1926577	1944972	18	32.6024	45.7521	1415034	13	36.2296	42.2362	3828197	14.456620
302	Wasit	18395	Qadissiyah	115032	133427	18	32.6024	45.7521	1415034	14	32.0437	45.1495	1325031	18.565468
303	Wasit	18395	Salahaddin	36920	55315	18	32.6024	45.7521	1415034	15	34.5338	43.4837	1637232	15.906455
304	Wasit	18395	Sulaymaniyah	1331028	1349423	18	32.6024	45.7521	1415034	16	35.5558	45.4351	2219194	16.447346
305	Wasit	18395	Thi-Qar	119796	138191	18	32.6024	45.7521	1415034	17	31.1042	46.3625	2150338	20.142749

Figure 2: MacroDataSet Sample

Where, Src and Dst refer to the labels of location i and j for individuals who are moving between them. Also, O_i and D_j are the total numbers of $outflows$ (number of humans who left location i), and $inflows$ (number of humans reached to location j). While $month$ column mentioned the date of the call. $flows$ represent the total number of inflows and outflows. In addition, the distance between the towers is represented as D_{ij}. Additionally, $city\_id\_x$ and $city\_id\_y$ columns represent ID for every governorate. Furthermore, pop_x and pop_y columns represent the population for each governorate. lastly, Lat and $Long$ represent the coordination of the nearest tower covered by the call.

	month	Src	Oi	Dst	Dj	Flows	city_id_x	lat_x	long_x	pop_x	city_id_y	lat_y	long_y	pop_y	Dij
0	DEC-20	Anbar	69777	Babil	16340	86117	1	32.5598	41.9196	1818318	2	32.4682	44.5502	2119403	15.283343
1	DEC-20	Anbar	69777	Baghdad	271058	340835	1	32.5598	41.9196	1818318	3	33.3152	44.3661	8340711	14.481997
2	DEC-20	Anbar	69777	Basra	8905	78682	1	32.5598	41.9196	1818318	4	30.5258	47.7738	2985073	19.623949
3	DEC-20	Anbar	69777	Diyala	5161	74938	1	32.5598	41.9196	1818318	5	33.7733	45.1495	1680328	14.731727
4	DEC-20	Anbar	69777	Dohuk	235947	305724	1	32.5598	41.9196	1818318	6	36.8632	42.9885	1326562	11.185936
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
1831	SEP-20	Wasit	2945	Ninewa	326428	329373	18	32.6024	45.7521	1415034	13	36.2296	42.2362	3828197	14.456620
1832	SEP-20	Wasit	2945	Qadissiyah	18088	21033	18	32.6024	45.7521	1415034	14	32.0437	45.1495	1325031	18.565468
1833	SEP-20	Wasit	2945	Salahaddin	5970	8915	18	32.6024	45.7521	1415034	15	34.5338	43.4837	1637232	15.906455
1834	SEP-20	Wasit	2945	Sulaymaniyah	225189	228134	18	32.6024	45.7521	1415034	16	35.5558	45.4351	2219194	16.447346
1835	SEP-20	Wasit	2945	Thi-Qar	15979	18924	18	32.6024	45.7521	1415034	17	31.1042	46.3625	2150338	20.142749

Figure 3: MesoDataSet Sample

3.2 Method and Material

The gravitation model was applied to the observed data sets. In this package of models, the city residents were used as a mass as an attractive formula in physics, as it is widely used in transportation and the spread of epidemics. When using it, the sample is transferred from one place to another, for example, a person or any type of virus and anything that can be move on reality[17]. In addition, the general formula displays in the general Eq. (1) [11].

$$T_{ij} = k\,\frac{S_i D_j}{d_{ij}^{\gamma}} \tag{1}$$

where, the normalization factor is k. Thus, $S_i\ and\ D_j$ in the equation are equivalent to O_i and D_j columns in the observed data. In addition, the distance between the towers is represented as D_{ij}, and equivalent to d_{ij} in the gravitation formula. The gravity model has was between the source and destination or called location $i\ and\ j$. Also, the $inflows$ and $outflows$ have been represented from the observed data. The next two formulas in Eq. (2) and Eq. (3) represent this implementation in python respectively; where, equation 2 is the library used to import the gravity algorithm, and equation 3 is the main equation for applying the gravity algorithm[4]. 'Pow' is the cost function that does the input and output for D_{ij}, or the so-called $cost$.

Also, some libraries have been used such as Geopandas for detecting the location of users, in addition to Numpy and Pandas.Figure 4 below represents the general structure of the model.

$$In\ [1]{:}\ from\ spint.gravity\ import\ Gravity \tag{2}$$

$$In\ [2]{:}\ gravity\ =\ Gravity(flows, Oi, Dj, Dij, 'pow') \tag{3}$$

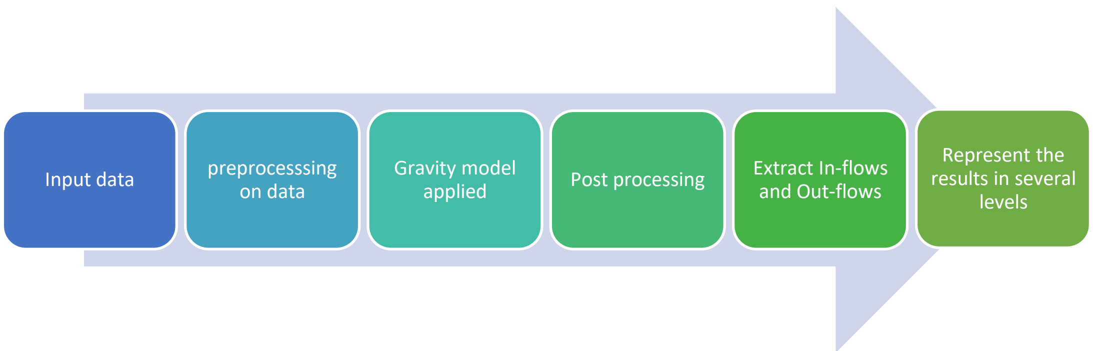


Figure 4: General structure of the Gravity model

3.3 Model Fit Statistics

There are some metrics or statistic fit to express the best model or equivalent model to the applied data set within certain fixed ratios in a fixed mathematical equation for each scale such as $R - squared$ in [18]. It is the statistical term for the fraction of variance of the dependent variable that can be predicted from the independent variable, represented by R^2 or r^2. In addition, $R - squared$ values ranged between 0 and 1. It is represented by the following Eq. (4) [4].

$$R^2_{\text{pseudo}} = 1 - \frac{\ln \hat{L}(M_{\text{full}})}{\ln \hat{L}(M_{\text{Intercept}})} \quad (4)$$

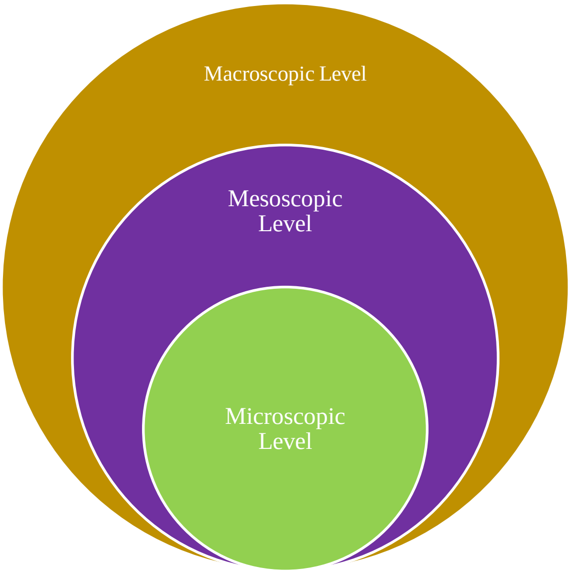


Figure 5: The models in several levels of abstraction

where, $L\hat{}$ is the likelihood of an approximate model M_{full} is an explanatory model with all variables of interest, in other words the represented data that is used for modeling. MIntercept is a model with only an intercept, or the data used in modeling.

3.4 Pedestrian Structural Simulation

For pedestrian motion simulation, there are many simulation methods and software products. Their spatial presentation might be continuous, grid, or network structure. However, their purposes might be a specific purpose or public intention. Their level of abstraction might be macroscopic, mesoscopic, or microscopic as shown in Figure 5. Therefore, these models are classifiable. However, the theory and modeling that are specifically focusing on panic conditions and emergencies are still in the early stages of development [19]. Thus, this research focused on the levels of abstraction by analyzing the observed data along the total period in macroscopic level. Meanwhile, in mesoscopic level for every single month.

4. The Results Discussion

The estimated results from applying the model to the observed data sets, are shown in Figures 6, 7 and 8 at two levels of macroscopic and mesoscopic models respectively. The macroscopic model is explored in Figure 6 that focused on the individuals in some areas and representing them in one group to clarify the general structure of the data, which can be considered as density. Thus, the obtained results show that the city of Ninewa has the lowest attraction among all Iraqi cities during the observed period. On the other hand, the most attractive cities are Diyala, Missan, Salah al-Din, and Wasit respectively. However, it is important to note that the number of in and out flows are equal because there is no information for the destination, so these results had been acquired after the preprocessing phase as mentioned above. Furthermore, the obtained results from the gravity model were projected geographically to show them as it is on the real area map.

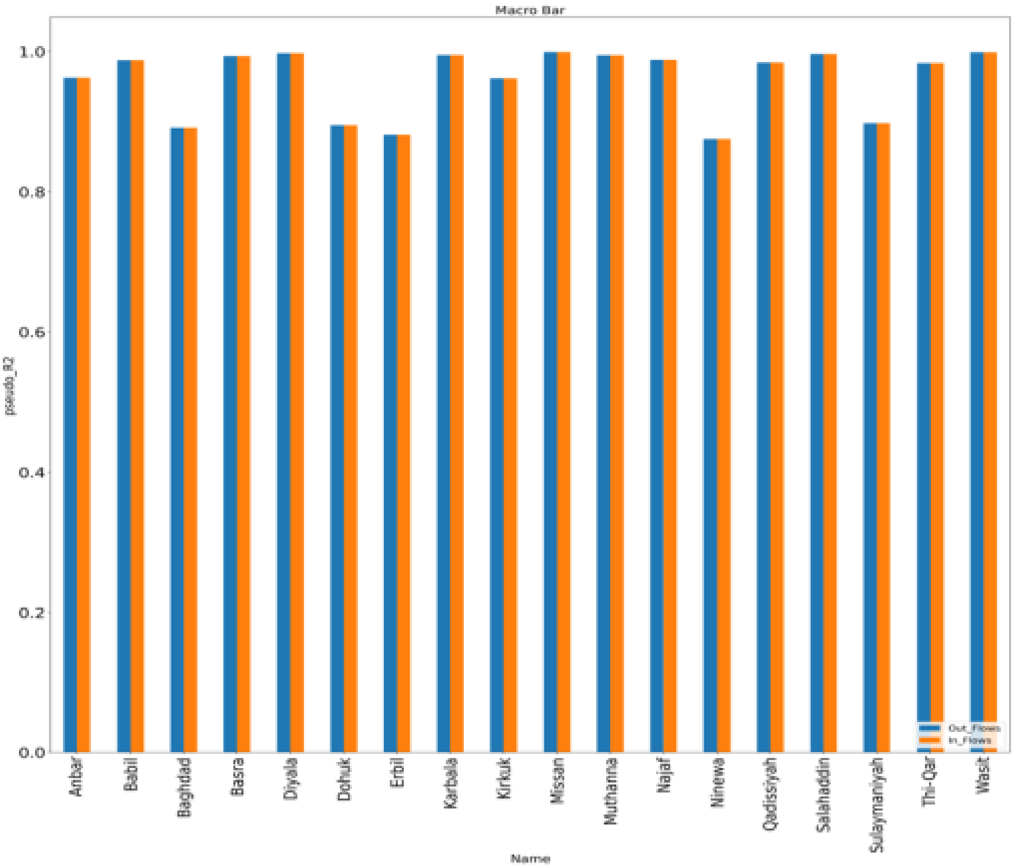
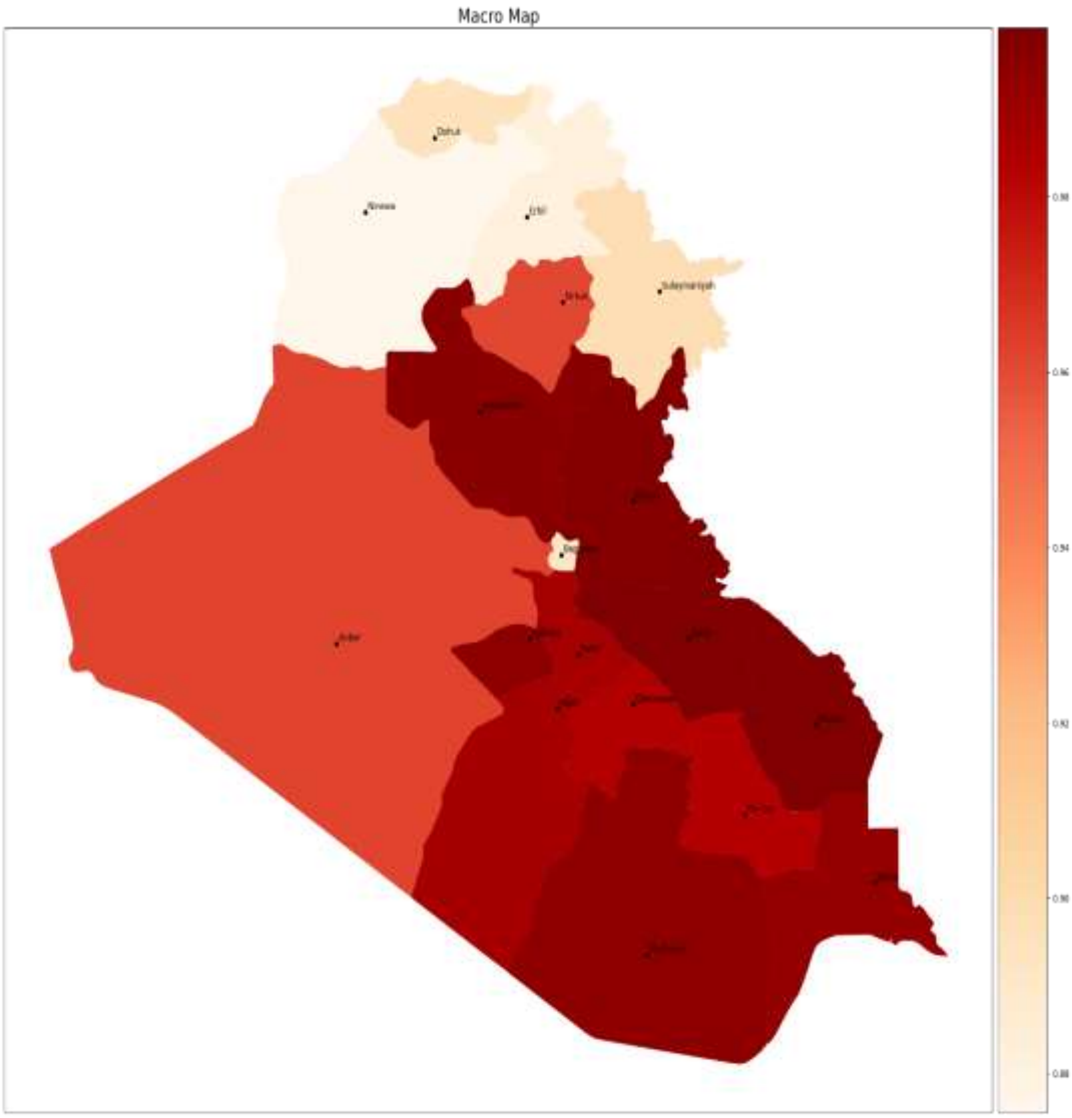


Figure 6: The Migration in-between Iraqi governorates from macroscopic level

Some software capabilities had been used to link the model and geography as shown in Figure 7. So, the Iraqi map, according to the obtained results, shows that the cities with dark color are the most attractive for migration, while the light ones are the least attractive. Thus, the mesoscopic had been used to explore the results in more detail for smaller groups, and it considered both Macroscopic and Microscopic levels. The observed data was explored in detail for each month, and is presented in one form as in Figure 8.

It has been found that there is a great diversity of migration between the governorates, so the northern governorates such as Dohuk and Sulaymaniyah are close in proportion among the rest of the provinces. Ninewa is considered to have the lowest rate of migration, and there are also high and closely related ratios, such as Missan, Salah al-Din and Wasit, as they are considered among the governorates with a medium population density and a medium area, so the movement in it is clear



.

Figure 7: The Migration map in-between Iraqi governorates from macroscopic level

However, by calculating the arithmetic mean of $inflows$ and $outflows$ between cities and based on Table 1, which represents a random sample of inter-city relations, it was concluded in Figure 9 that the line connecting $Kirkuk\ and\ Dohuk\ = 89\%$ is the least thick, which indicates the least migration between the randomly taken cities. While the thickest line is between $Missan\ and\ Diyala\ = 99\%,$ which indicates the most migration between the randomly taken cities. In addition, the color of the circle represents the governorate density, and the size of the circle represents the governorate area size. Thus, the rest of the samples table can be considered at the average level of migration. Note that these results may not reflect the true reality of typical daily life because they covered the epidemic period only. Consequently, it would be necessary to expand the dataset to cover the post-COVID-19 periods to discover the truth of migration. It is also recommended in the future to use complex networks models for studying people migration via wireless networks, it could be very useful to investigate this kind of studies.

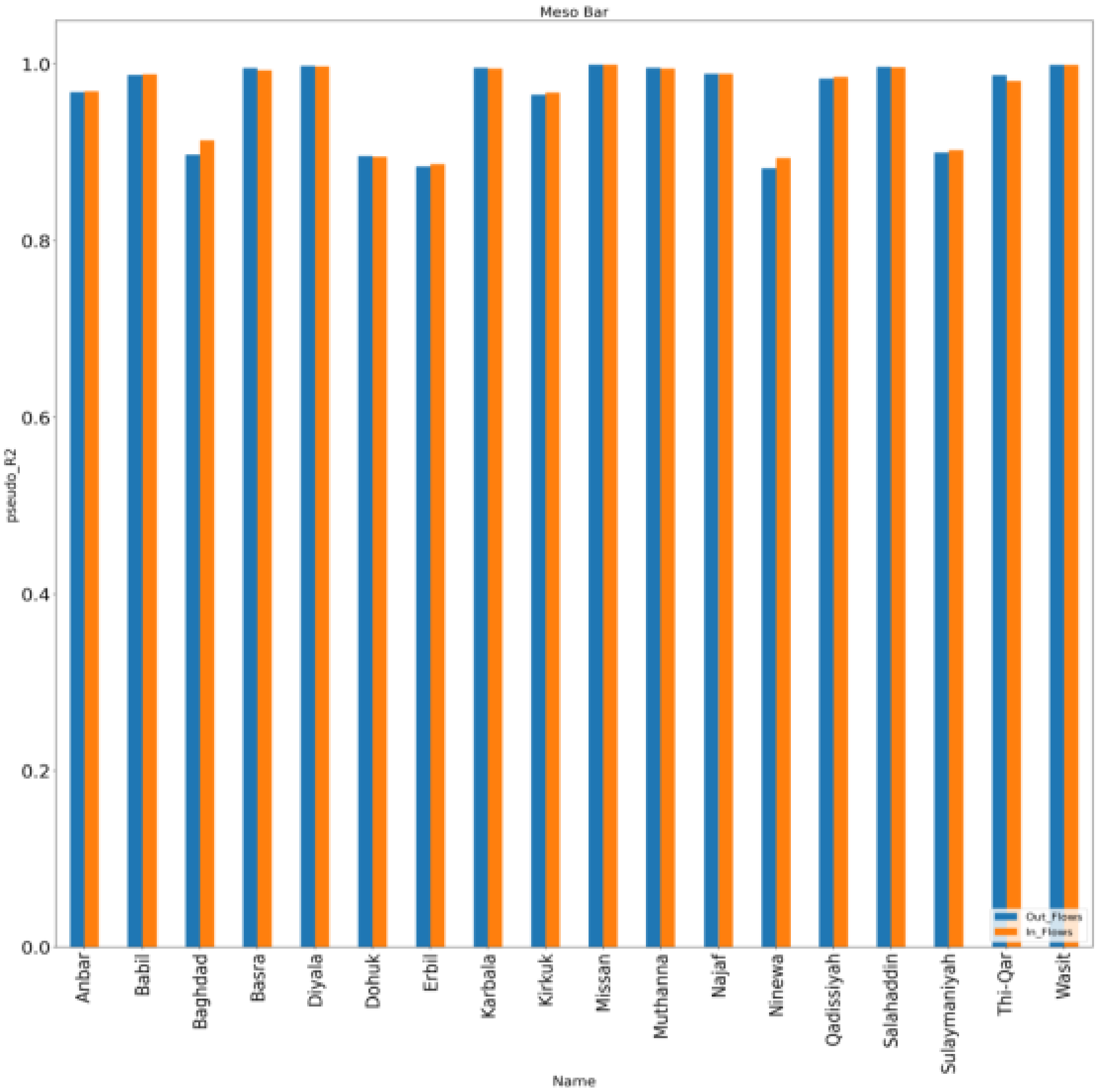


Figure 8: The Migration in-between Iraqi governorates from Mesoscopic Level

Table 1: Sample of migration between the governorates

ID	Src	Dst	Mean	ID	Src	Dst	Mean
129	Muthanna	Qadissiyah	0.982465	*4*	Anbar	Missan	0.966178
86	Dohuk	Kirkuk	0.894006	*19*	Babil	Kirkuk	0.947431
104	Karbala	Salahaddin	0.990253	*16*	Anbar	Kirkuk	0.925656
64	Diyala	Missan	0.995669	*72*	Diyala	Karbala	0.991144
112	Kirkuk	Najaf	0.948027	*121*	Missan	Sulaymaniyah	0.936327

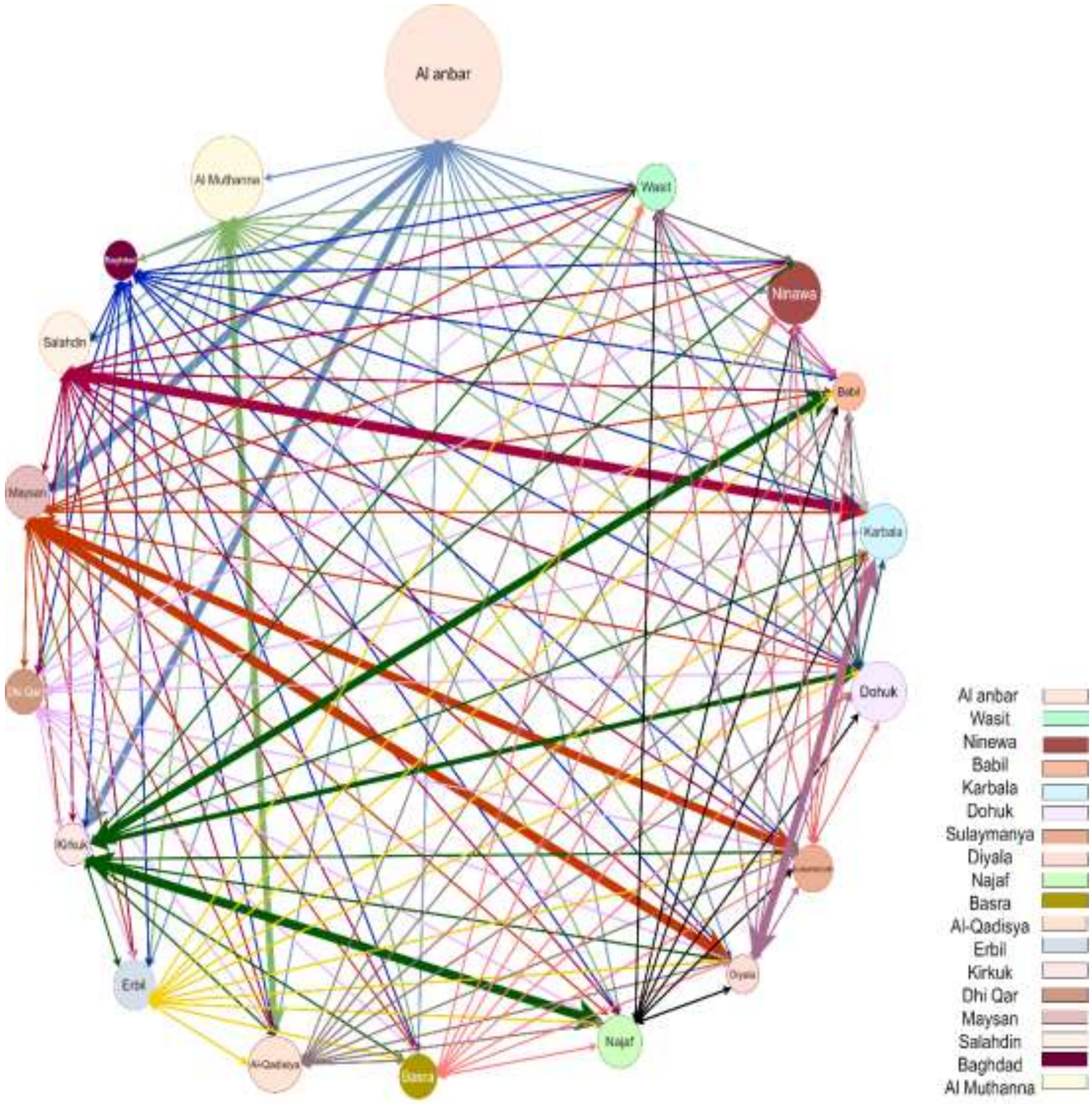


Figure 9: The Flow Migration Density

5. Conclusions and Future Work

It became clear that the phone network data are very useful in analyzing social patterns of urban life, since they aid in analyzing wide data that covers vast urban areas. In this paper, the observed data shed light on the people migration between Iraqi governorates during the COVID-19 period based on the gravity model. The extracted results showed that the highest migration cities are Diyala, Missan, Salah al-Din and Wasit, where the attraction force was highest in that period. Whereas the lowest migration was in Nineveh. Also, it had been found that this model succeeded in analyzing real CDRs data.

In addition, it is concluded that the data of the mobile network (CDRs) can aid in discovering the number of migrant people in and out of cities, also among these cities. These people movements had been reflected according to the CDRs nature of a spatio-temporal mode. Notwithstanding that there are no details data about the physical, social and economic conditions of these cities, that would have explained more reasons for the people migration. If there were more detailed groups or data, that would also help arrive to conclusions about the social and economic reasons.

Consequently, this type of data may be analyzed with complex networks models due to the diverse ratios of $inflows$ and $outflows$ among the cities within variant spatio-temporal mode. In addition, the gravity model had been tested with a calibration ratio of 77%. For the future, it is possible to deal with data for other social or economic fields, and it is even possible to deduce the momentum of the migration as to how long it will continue, and whether it will continue or not. It is recommended to expand the observed data period to cover period intervals before and after COVID-19, this would give more comprehensive perspectives.

Acknowledgement

The authors are grateful to Korek Telecom company, Iraq for the support in the research by providing the local data.

References

[1] W. R. J. Vermeulen, D. Roy, and R. Quax, "Modelling the Influence of Regional Identity on Human Migration," *Urban Sci.*, vol. 3, no. 3, p. 78, 2019, doi: 10.3390/urbansci3030078.

[2] R. Skeldon, "International Migration, Internal Migration, Mobility and Urbanization:Towards more integrated approaches," *Migr. Res. Ser. No.53 IOM*, no. August, pp. 1–15, 2018, [Online]. Available: http://www.un-ilibrary.org/migration/international-migration-internal-migration-mobility-and-urbanization_a97468ba-en.

[3] N. Pourebrahim, S. Sultana, A. Niakanlahiji, and J. C. Thill, "Trip distribution modeling with Twitter data," *Comput. Environ. Urban Syst.*, vol. 77, no. July, 2019, doi: 10.1016/j. compenvurbsys. 2019.101354.

[4] T. Oshan, "A primer for working with the spatial interaction modeling (SpInt) module in the python spatial analysis library (PySAL)," *Region*, vol. 3, no. 2, pp. R11–R23, 2016, doi: 10. 18335 /region.v3i2.175.

[5] C. W. Hsu, C. Fan, and A. Mostafavi, "Limitations of gravity models in predicting fine-scale spatial- temporal urban mobility networks," pp. 1–18.

[6] 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.

[7] 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.

[8] J. N. Lee, M. Mahmud, J. Morduch, S. Ravindran, and A. S. Shonchoy, "Migration, externalities, and the diffusion of COVID-19 in South Asia☆," *J. Public Econ.*, vol. 193, p. 104312, 2021, doi: 10.1016/j.jpubeco.2020.104312.

[9] Y. Kang, S. Gao, Y. Liang, M. Li, J. Rao, and J. Kruse, "Multiscale dynamic human mobility flow dataset in the U.S. during the COVID-19 epidemic," *Sci. Data*, vol. 7, no. 1, pp. 1–14, 2020, doi: 10.1038/s41597-020-00734-5.

[10] I. Martínez-zarzoso and L. Márquez-ramos, "International trade, technological innovation and income: a gravity model approach," *IVIE Work. Pap.*, no. June, 2005.

[11] A. P. Masucci, J. Serras, A. Johansson, and M. Batty, "Gravity versus radiation models: On the importance of scale and heterogeneity in commuting flows," *Phys. Rev. E - Stat. Nonlinear, Soft Matter Phys.*, vol. 88, no. 2, 2013, doi: 10.1103/PhysRevE.88.022812.

[12] S. Hall, M. Way, and J. Page, "University of Hawai ` i at M ā noa Department of Economics Working Paper Series Working Paper No . 12-16 The 1960 Tsunami in Hawaii : Long Term Consequences of Costal Disaster By John Lynham," vol. 2012, no. August, 2012.

[13] 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.

[14] N. T. T. Hoang, H. Q. Truong, and C. Van Dong, "Determinants of Trade Between Taiwan and ASEAN Countries: A PPML Estimator Approach," *SAGE Open*, vol. 10, no. 2, 2020, doi: 10.1177/2158244020919516.

[15] M. Candela, V. Luconi, and A. Vecchio, "Impact of the COVID-19 pandemic on the Internet latency: A large-scale study," *Comput. Networks*, vol. 182, no. July, p. 107495, 2020, doi: 10.1016/j.comnet.2020.107495.

[16] C. Ciavarella and N. M. Ferguson, "Deriving fine-scale models of human mobility from aggregated origin-destination flow data," *PLoS Comput. Biol.*, vol. 17, no. 2, pp. 1–18, 2021, doi: 10.1371/JOURNAL.PCBI.1008588.

[17] S. Isaacman, V. Frias-Martinez, and E. Frias-Martinez, "Modeling human migration paerns during drought conditions in La Guajira, Colombia," *Proc. 1st ACM SIGCAS Conf. Comput. Sustain. Soc. COMPASS 2018*, 2018, doi: 10.1145/3209811.3209861.

[18] D. Chicco, M. J. Warrens, and G. Jurman, "The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation," *PeerJ Comput. Sci.*, vol. 7, pp. 1–24, 2021, doi: 10.7717/PEERJ-CS.623.

[19] N. Shiwakoti, M. Sarvi, and G. Rose, "Modelling pedestrian behaviour under emergency conditions - State-of-the-art and future directions," *31st Australas. Transp. Res. Forum, ATRF 2008*, pp. 457–473, 2008.