Rodríguez, L.; Palanca Cámara, J.; Del Val Noguera, E.; Rebollo Pedruelo, M. (01-2). Analyzing urban mobility paths based on users' activity in social networks. Future Generation Computer Systems. 102:333-346. https://doi.org/10.1016/j.future.2019.07.072

Analyzing urban mobility paths based on users' activity in social networks

L. Rodríguez[a], J. Palanca[a], E. del Val[a,b], M. Rebollo[a]

[a] *Valencian Research Institute for Artificial Intelligence*
Universitat Politècncia de València
Camino de Vera s/n 46022 Valencia (Spain)
lurod1@posgrado.upv.es
{jpalanca,mrebollo}@dsic.upv.es
[b] *Department of Computer Science and Engineering of Systems*
University of Zaragoza
Escuela Universitaria Politcnica de Teruel
c/ Atarazana 2, 44003 Teruel (Spain)
edelval@unizar.es

Abstract

This work presents an approach to model how the activity in social media of the citizens reflects the activity in the city. The proposal includes a gravitational model that deforms the surface of the city based on the intensity of the activity in different zones. The information is extracted from geolocated tweets ($n = 1.48 \times 10^6$). Furthermore, this activity affects how people move in a city. The path a user follows is calculated using the geolocation of the tweets that he or she publishes along the day. Several models are evaluated and compared using the Hausdorf's distance (d_H). The combination of gravitational potential with attraction to the destination points provides the best results, with $d_H = 1176$ against the Manhattan ($d_H = 1203$) or the geodesic ($d_H = 1417$) alternatives. Finally, the analysis is repeated with the data segmented by gender (n=2,826 paths, men=1,910, women=916). The results validate (p=0.000334) the studies that affirm that men travel longer distances ($d_M = 4.73$ km, $\alpha_M = 26.1°$) with rectilinear trajectories, whereas women have shorter and more angled paths ($d_W = 4.5$ km, $\alpha_W = 32.2°$), obtaining pvalues p=0.0014 for the significance in the differences in path lengths and p=0.006 in the angles.

Keywords: complex network, social media, mobility, gender, smart cities

1. Introduction

In last years, the analysis of people mobility is an area of research that has been reinforced by the availability of a huge volume of geo-tagged and accessible information. Due to the use of mobile phones, a large amount of data related to mobility is obtained through the use of different media: WiFi connections, communications between individuals by text messages, phone calls, or social networks. The availability of such amount of data has led to the analysis of mobility patterns of individuals at an unprecedented scale regarding the area covered by the trajectories as well as the number of individuals involved in the studies [1].

Researchers have applied the analysis of human mobility to a wide range of contexts. For instance, understanding individuals' mobility facilitates the detection of frequent activities [2], and abnormal or exceptional events [3, 4], the estimation of future movements [5, 6], migratory flows, urban planning [7, 8, 9] or traffic forecast [10]. One of the challenges in the analysis of human mobility is the proposal of models that can determine individual trajectories at a local scale. There are many research efforts on proposing mobility models based on new tracking technologies such as mobile phones [11] or GPS devices that provide low-resolution data records [12]. However, these data sources involve privacy concerns and data access restrictions [13].

Recently, as an alternative, large online social systems have been used as providers of useful data about human dynamics for mobility models. Specifically, Twitter allows users to geotag their publications with their current location providing high position resolution down to 10 meters.

There are studies that state that Twitter is a reliable source for studying human mobility patterns [14] despite some open issues, such as (i) potential sampling bias due to the communication modality; (ii) bias with respect to the amount of located tweets; (iii) those related to the profile of people who activate their location, and (iv) certain types of events in which users are more likely to share their location. The majority of the studies that analyze human mobility

focus on urban and inter-city trajectories/movements without tracing the points that an individual follows through a path. Moreover, many of the proposals solely analyze human mobility from an analytical perspective. Mobility is a complex phenomenon, and therefore not only the individual but also his or her demographic characteristics (i.e., age, gender or other contexts) must be considered [15].

There are studies that evidence a difference in the mobility patterns between men and women. An additional objective of this work is to contrast this hypothesis with empirical data provided from social networks. Furthermore, having these differences into account the mobility necessities of each collective can be considered in a fairer way.

The studies that deal with the analysis of gendered mobility are only based on qualitative data [16, 17, 18, 19]. For instance, Blumen et al. [20] analyzed the trajectories of men and women in work trips. The results show that there are significant differences between genders. Basaric et al. [21] and Maffiiny et al. [22] studied the type of transport used by genre. The results throw a difference in the media: a higher use of private cars for men and public transport and walking for women. Although this factor has not been studied explicitly in this work, the obtained results are coherent with the mentioned studies. As a consequence, the gender perspective affects the sustainability of transport in the cities, which has been considered by Hanson [23].

The interest of this research is related with the fact that mobility plays an important role in the population's quality of life and urban dispersal. Problems like accessibility to the center of the town, quality of transport service and environmental issues may be addressed with a proper analysis of mobility patterns [24].

In this paper, we present a gravitational potential model that shows how the activity in social networks is a reflection of the activity in a city. This model is represented as a contour map that shows how the population of a city is attracted to each point in the city map by means of a gravitational–like force. The initial hypothesis of this work is that these potential lines influence how people

tend to move along a city. To validate this hypothesis, we propose two strategies for movement description (i.e., the attraction strategy and the geodesic strategy), based on our gravitational model, considering users' geotagged activity in Twitter. Finally, we performed a set of experiments with more than one million geotagged tweets to empirically validate the strategies.

A second experiment is focused explicitly on detecting differences in mobility taking into account the gender of the subjects under study. This experiment tries to validate the studies that state that men and women move differently in daily trips [18, 19, 20].

The rest of the paper is organized as follows. Section 2 presents previous works that analyze human mobility using different data and models. Section 3 presents the gravitational potential approach to model the activity of a city. Section 4 describes the proposed strategies based on the previous model to estimate a path followed by a user. Section 5 evaluates the strategies for path estimation using different methods for pre-processing the real geo-located data from Twitter. This section also analyzes path differences by gender. Finally, conclusions and future work are presented.

2. Related Work

Obtaining and analyzing information on people's traces is fundamental to understand human mobility and to make strategic decisions about the design of urban infrastructure or transport systems [25]. Human mobility is a topic that has attracted the attention of many researchers for years [26, 27, 28, 6, 29]. Researchers in this area have analyzed and proposed theoretical models that facilitate the description of joint mobility flows using different sources of information such as mobile phone calls [30, 31], GPS data [32], and geotagged information from social networks [33].

A first group of the research proposals are focused on analyzing data from different sources (i.e., GPS, social media, mobile traces) to understand mobility behavior. For instance, Lima et al. [34] analyze the GPS traces generated

by 526 private cars over an 18-month period and explore how their routing behavior unfolds in four cities. Other works use the data provided by mobile phones to extract spatial patterns of mobility within the metropolitan area and their relationships to neighborhood characteristics [30]. In this line, Gonzalez et al. [28] also studied the trajectory of anonymized mobile phone traces. The authors found that human trajectories tend to show a high degree of temporal and spatial regularity [35]. Individual's movements can be determined by a time-independent characteristic travel distance and a significant probability to return to locations they visited before. Lee et al. [36] analyze the influence of street morphology on travel routes in several cities through a metric that captures the tendency of these routes to orbit towards or away from the city center. They proposed a metric called *inness* that considers the direction, orientation and length of routes, thus revealing the morphology of connectivity in street networks, including the distribution of implicit socioeconomic forces that may inform routing choices.

The use of mobile or GPS data provides accurate information about users' movements. However, this information is not always accessible. For this reason, many studies consider geo-positioned data from social networks as a suitable source to analyze mobility patterns. Flicker [37], Twitter [38], Brightkite or Foursquare [39] among others offer the possibility of tracking users' location through the geotagged information contained in the users' posts as well as understanding their environmental context. Bejar et al. [40] used Twitter data to analyze spatiotemporal patterns of users that visit or live in a city. Other works study the places frequently visited by a user and how these places are connected [41], or analyze daily movements to identify different areas of the city [42]. Other works analyze mobility at international scale [43].

A second group of research proposals try to generalize mobility patterns of users through mathematical models. One of these models is based on Newton's *gravity* force. The idea of this *gravity model* was introduced by Zipf in 1946 [26] to estimate the size of the flow between two areas. The gravity model considers that the flow of people traveling between two locations is directly proportional to

the attraction force based on the population of the source and destination, and decays as the distance between them increases. This gravity model is static and requires mobility data to fit its parameters. Several works have used this model to deal with the problem of describing human mobility patterns [44, 45, 46]. For instance, Smith et al. [45] identified deprived areas of the city describing the movement of passengers between transport stations using the gravitational model.

An alternative model was proposed by Simini et al. [47]. They present a *radiation model* that considers human movements as diffusion processes that depend on the population distribution over the space. This model is static and in its original formulation has not parameters to fit, but requires accurate knowledge of the spatial population distribution.

Stouffer et al. [27] proposed the *intervening-opportunities* model. The model states that the number of people going at a given distance is directly proportional to the number of opportunities at that distance and inversely proportional to the number of intervening opportunities. This model requires accurate knowledge about the place where it is going to be applied.

Other models study the interplay between the regular trajectories and if they are random and, therefore, unforeseeable. For instance, Jia et al. [48] state that human mobility is mainly attributable to hierarchical and individual preferences. They implement an agent-based model that considers three main features: (i) the scaling and hierarchical properties of the clusters with the targets of the individuals which serve as the underlying spatial structure; (ii) the individual preferred trajectories that capture the same Levy flight pattern as it is observed in the participants; and (iii) the jumping factor, which is the probability that one person may cancel their regular mobility schedule and explore a random place. They validated their model using GPS traces of 258 volunteers.

Some works use social relationships to establish a relationship between regular and random trajectories. Grabowicz et al. [49] propose a model that integrates physical locations and social relationships. The part of the model related to mobility takes into account recurring visits to the same location and

exploration of new places. Specifically, the model considers that one person performs a first stage action. He or she can travel to a randomly selected location of a friend with a certain probability. They can also travel to a new location considering a distance (i.e., obtained from a distribution of jump lengths) and a direction (i.e., chosen proportionally to the population density at the target distance). The model was tested considering datasets from Twitter, Gowalla, and Brightkite. The model requires four input parameters and the distribution of jump lengths.

Table 1 summarizes the previous studies presented in this section, which frequently suffer from some of the following limitations. First, they are mainly focused on human collective flows (i.e., people movements from one place to other) without distinguishing between particular users. They do not consider either the specific path that a user follows to reach its destination. Therefore, inferring more personal mobility models might be useful to come up with more customized services. Secondly, most works analyze social geolocated data focusing on the processing of spatiotemporal aspects of the data, but without taking into account other context-aware aspects like the gender of the users. As a result, these works do not fully take advantage of social media datasets.

The work presented in this paper proposes a set of strategies based on a gravitational potential model that allows us to model how people in a city is going to move along its surface. This analysis is based on users geotagged activity in Twitter that fully considers the limitations listed above:

- it takes into account individual user activity, obtained from the locations at which they have tweeted,
- we segregate the paths by gender, since there are studies that affirm that such difference exists and it is significant.

Author	Analysis	Scope	Model	Source	Context
Lima [34]	Car routes	Inter	Spatial probability distribution that bounds the route selection space within an ellipse, (origin and the destination as focal points)	GPS	–
Calabrese [30]	Trip length	Intra	n/a	Mobile phone traces	–
González [28]	Trajectories	n/a	Lèvy-Flight (LF) or Truncated Lèvy-Flight (TLF)	Mobile phone traces	–
Lee [36]	Origin-destination pairs	Intra	Innessa function of both the direction and spatial length of routes	–	–
Béjar [40]	Individual events and the connections among them	Intra	Leader clustering algorithm	Twitter, Instagram	–
Terroso-Senz [41]	Personal mobility graphs hierarchy	Intra	Density-based clustering Complex Event Processing	Mobile phone traces, Twitter	✓
Sila-Nowicka [42]	Trajectories and movement modes	Intra	Neural networks	GPS	✓

Author	Analysis	Scope	Model	Source	Context
Hawelka [43]	Global mobility patterns origin-destination pairs	Global	Mobility rate, radius of gyration, diversity of destinations, and inflowoutflow balance	Twitter	–
Beiró [44]	Human flows between pairs of geographical nodes	Inter-city	Gravity model Radiation model	Flickr	–
Smith [45]	Passenger flow patterns origin-destination pairs	Intra-inter city	Gravity model	Public transport system records	–
Simini [47]	Mobility fluxes	Intra-inter city	Radiation model (require as input only information on the population distribution)	Census data	–
Jia [48]	Mobility patterns	Intra-inter city	Levy fligth: (i) scaling and hierarchical properties of the purpose clusters (ii) the individual preferential behaviors (iii) jumping factor	GPS traces	–
Grabowicz [49]	Origin-destination pairs	Inter-city	Travel and Friend (TF) Model	Twitter, Gowalla, Brightkite	✓
Molas [50]	Mobility patterns	Intra-inter-city	Gravitational model	Twitter, Census	–

Author	Analysis	Scope	Model	Source	Context

Table 1: Overview of mobility approaches.

3. Modelling the activity of a city

In this section, we present a new way of modeling the surface of a city based on the idea of the gravitational potential and how a path is defined following this surface model and considering different strategies (i.e., attraction and geodesic).

3.1. Surface modelling of a city

In this work, we propose a surface of a city model based on the idea of the gravitational potential. Other approaches use the Newtonian idea of gravity, where two objects are attracted as a function of their mass and the distance that separates them. Unlike those models, we propose to use Einstein's relativistic idea, where gravity is defined as a deformation of space-time as a function of the mass of objects [51]. This general relativity theory is used as an inspiration to imagine how the activity in a city generates a field that attracts people moving through the city to the essential nerve points at any given moment. We can imagine that the *mass* of the geolocated tweets in the city *deforms* its surface (in this case we are going to leave the dimension of time to one side, only for the purposes of the metaphor used) creating an irregular plane where the zones with more gravitational potential are deeper, attracting people who move near them. In Figure 1, we can see how this surface would look like in three dimensions (top) and in a contour map representation (bottom). The novelty of this representation in contrast to a heat map is that we have the most relevant points of the city along with information about the pendent around each point. With this representation, we can determine the attraction of a person towards a specific place or area of the map.

In this work, we have used a dataset of tweets, but it should be noted that any kind of geolocated activity would be representable with this model. A tweet i generates a deformation caused by its weight $G_i(x, y)$, where x and y are the

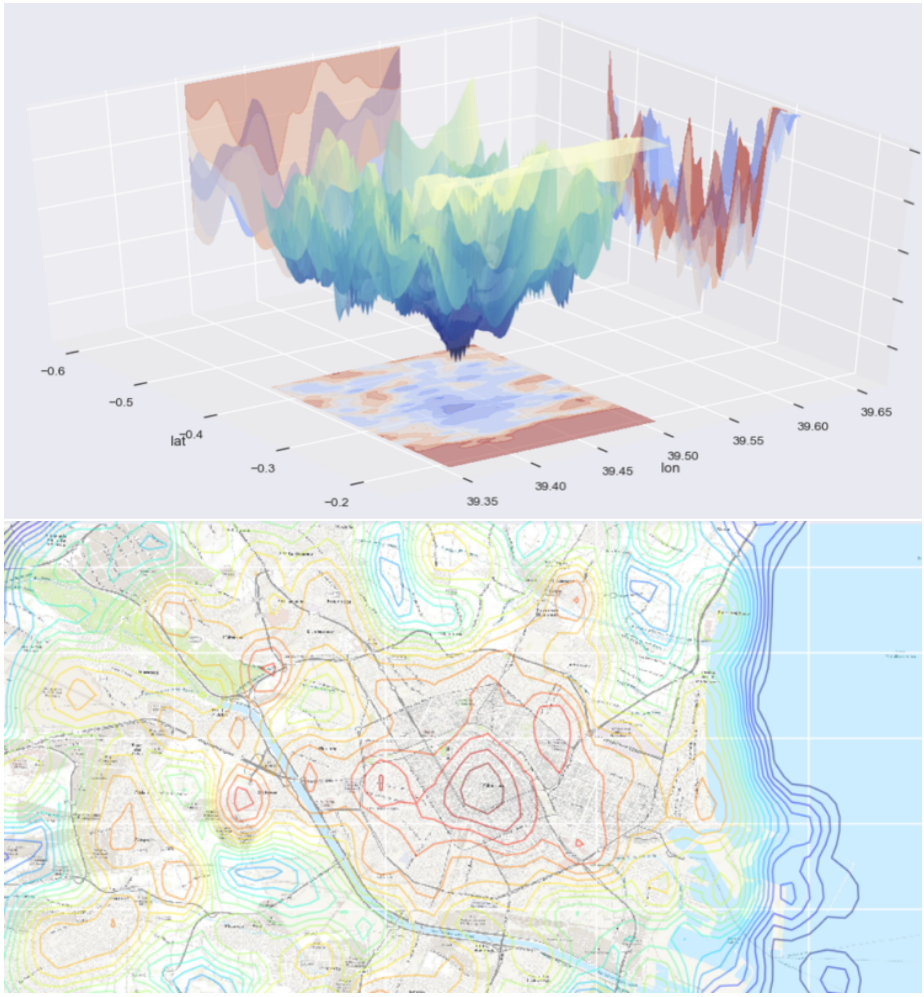

Figure 1: 3D representation (top) and contour map (bottom) of the gravitational model over the city of Valencia.

coordinates of a point in the map, that is calculated based on a renormalized, bi-dimensional Gaussian distribution centered on the coordinates of the tweet such that $G_i(lat(i), lon(i)) = -1$. A tweet i has a gravitational potential value in its surrounding area defined by Equation 1:

$$G_i(x, y) = -e^{-\frac{1}{2}\left(\frac{x-lat(i)}{\sigma}\right)^2} \times e^{-\frac{1}{2}\left(\frac{y-lon(i)}{\sigma}\right)^2} \quad (1)$$

where σ represents the area of influence of a tweet, and $lon(i)$ and $lat(i)$ represent the longitude and latitude of the tweet i.

The accumulated effect of all the tweets analyzed in the area is the sum of the gravitational potential of every single tweet. We use the logarithm for exponential smoothing the values (Equation 2).

$$G(x, y) = \log \sum_i G_i(x, y) \tag{2}$$

In the next section, we are going to present some strategies that we have defined to model paths that people follow when they move along the deformed surface of the city, generated by the density of the publications in different areas of the city.

3.2. Path modelling

The gravitational potential model presented in the last section allows us to create a surface of the city with *force attractions* that model how an object (in our case of study, people) would move along such surface. We define S_u as the ordered sequence of tweets emitted by a user u (Equation 3).

$$S_u = \langle i_0, i_1, \dots i_n \rangle \tag{3}$$

A path P_u of a user u is the ordered sequence of the coordinates of the tweets in S_u (Equation 4).

$$P_u = \langle (x_0, y_0), (x_1, y_1), \dots (x_n, y_n) \rangle \tag{4}$$

We consider the point $A = (x_0, y_0)$ as the origin point of a path and the point $B = (x_n, y_n)$ as the destination point of a path.

A first approach would be to follow a pure gradient descent over $G(x, y)$. If we imagine a small ball being dropped on a point on the surface, it moves quickly to its nearest local minimum if there is no other force pushing it. This strategy would certainly not be valid for predicting the path a person takes in a city since this path has a point of origin and (and this is the key) a point of

destination. For this reason, the mere fact of dropping the ball on the surface is indeed not enough to achieve the desired goal. Figure 2 shows what would happen in this case. The path followed by the person we are studying would be very different from that of our imaginary ball, which would quickly fall down in a pit or even leave the map without approaching the target in any case (except by accident).

For this reason, in this work we propose two strategies based on this gravitational potential to model the path that a person follows to go from point A to point B. These strategies aim to wisely use the information included in the gravitational potential to model with as little error as possible the paths that citizens take, starting from an origin to a destination. These strategies are called: *attraction* and *geodesic*.

3.2.1. Attraction Strategy

This strategy aims to solve the problem of the absence of influence of the destination point $B = (x_n, y_n)$ included in $G(x, y)$. This strategy adds some weight at the destination point. This modification forces the gradient vectors of

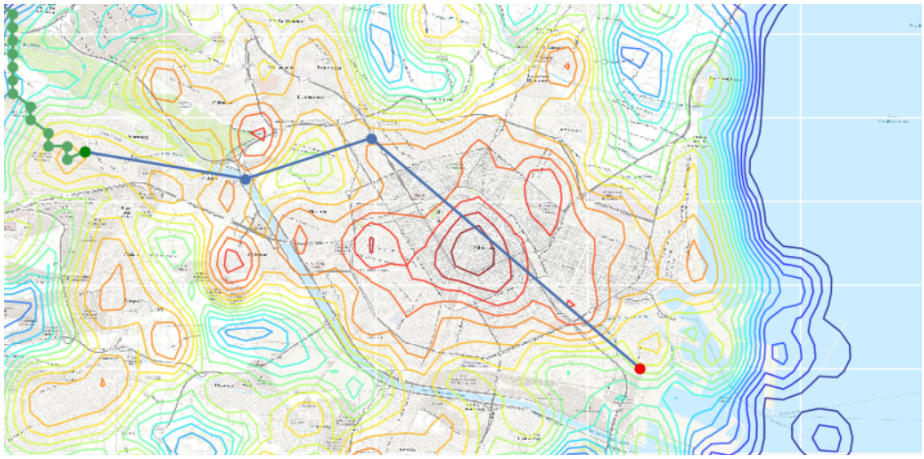

Figure 2: Real path (blue) compared to obtained path (green) when a pure gradient descent is performed. In this case, the path takes a wrong direction, guided exclusively by the gradient generated by the tweets.

the field to point to that destination, although still maintaining the roughness of the surface that the gravitational model has defined. The *attraction strategy* considers that the destination point B has an important weight and, therefore, attracts the path P_u towards itself. Equation 5 modifies Equation 1 considering the weight of the destination point B that depends on an attraction parameter called Λ. This parameter regulates the attraction that point B generates over the original gravitational map by introducing a radial gradient field centered on B. If Λ has a low weight, this means that B has a low influence in the path. However, if Λ has a high weight, B has a high influence, and the path becomes a straight line from point A to B, which is not desired either.

$$G_\Lambda(x, y) = G(x, y) - \Lambda\sqrt{(x_n - x)^2 + (y_n - y)^2} \tag{5}$$

The attraction path

$$P_\Lambda = \big\langle (x_0, y_0), (x_1, y_1), \dots (x_n, y_n) \big\rangle \tag{6}$$

has the same A and B as initial and final points, and any intermediate point (x_i, y_i) is obtained from the application of the gradient defined by $G_\Lambda(x, y)$ to the previous point (x_{i-1}, y_{i-1}).

Figure 3 shows the effect of the attraction of point B (red) in the path both in 3D (top) and 2D (bottom). The vectors represent the gradient direction of the attraction field generated by all the tweets.

3.2.2. Geodesic Strategy

The *geodesic strategy* makes a different approach to the solution by looking for the shortest path in three dimensions that connects two points in a curved space. We pretend to minimize the length of the path between two points:

$$D(P_u) = \sum_{i=0}^{n-1} d_{i,i+1} \tag{7}$$

where $d_{i,i+1}$is the Euclidean 3D-distance between two consecutive points in the path P_u.

$$d_{i,i+1} = \sqrt{(x_{i+1} - x_i)^2 + (y_{i+1} - y_i)^2 + (G(x_{i+1}, y_{i+1}) - G(x_i, y_i))^2} \tag{8}$$

Since the calculation of the shortest path in a 3D surface is an NP-Hard problem, we use a common approach that discretizes the gravitational field into a grid and applies a graph-based algorithm to find the shortest path [52]. We consider the vertices of the grid as nodes in the graph. The irregularities of the

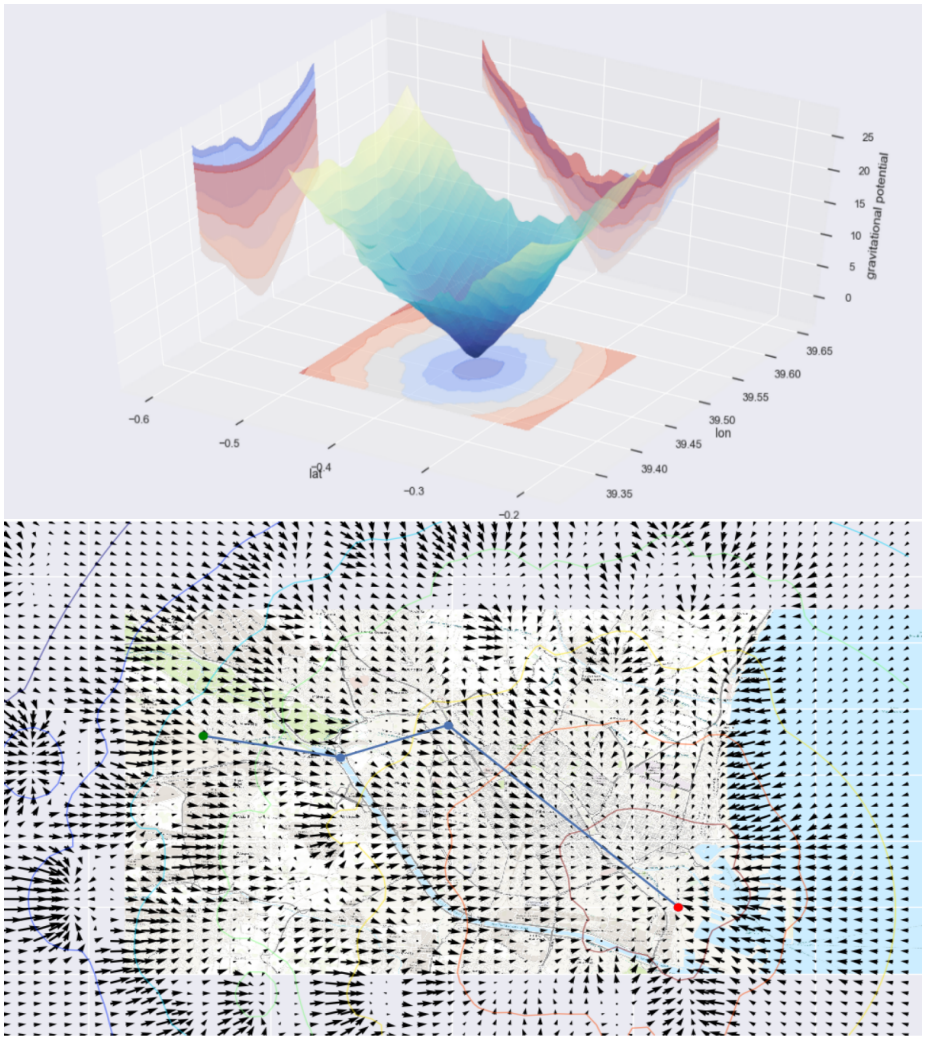

Figure 3: (top) 3D representation of the attraction strategy. The destination point B generates a radial gradient that attracts the walker although it conserves the roughness generated by the gradient of the tweets. (bottom) The vector field representation of the combined gradient, along with the path of a walker following the gradient.

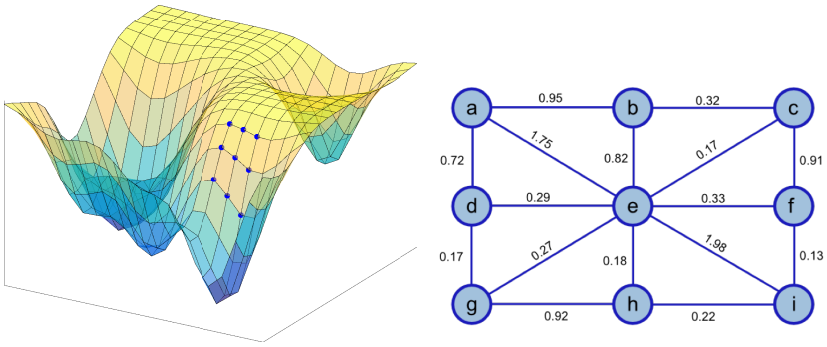

Figure 4: Graph that represents the area under analysis, where a node is a cell of the grid and the edges are the slopes with adjacent nodes.

surface are the difference of heights between two adjacent vertices. These slopes are included as weights of the corresponding edges of the graph (see Figure 4). The geodesic strategy finds the most straight and flat path between two points that avoids high slopes. To determine the path with the lowest cost, we use the Dijkstra's shortest path algorithm [53]. We denote as P_G the shortest path from A to B.

Figure 5 shows an example of a path generated with the geodesic strategy (green path) and the user's original path (blue path). The points of the path generated by the geodesic strategy are usually in adjacent level curves (i.e., points at a similar height).

4. Materials and Methods

Next, we need to evaluate the strategies we have presented in the previous section by comparing them with geolocalized data collected from social networks. Since data from social networks are sequences of coordinates we need to interpret these sequences of coordinates as real paths made by users. These real paths will be used to compare and evaluate our strategies. To do this, in this section we have iterated through three methods, which we have evaluated through experiments, in order to find the best representation of the real path from these data. Following is the Experimental Setup and the three methods

Figure 5: Contour map of the geodesic strategy. The geodesic strategy generates the green path. The original path is represented in blue color.

mentioned. Finally, an algorithm will be proposed to process the data extracted from social networks and prepare them for experiments that evaluate geodesic and attraction strategies.

The goal of this work is to validate how similar is our generated path for each strategy with the original path. For the evaluation, we considered the error between real paths and generated paths using the Haussdorf distance. Finally, in addition to the analysis of the predicted paths based on the gravitational potential model, we also study the differences between women and men path trajectories.

4.1. Setup

In this block, we describe the steps to retrieve and pre-process the dataset used in our experiments in order to reproduce them. These steps are: Information Retrieval from Twitter and Data Exclusion of redundant and incomplete data.

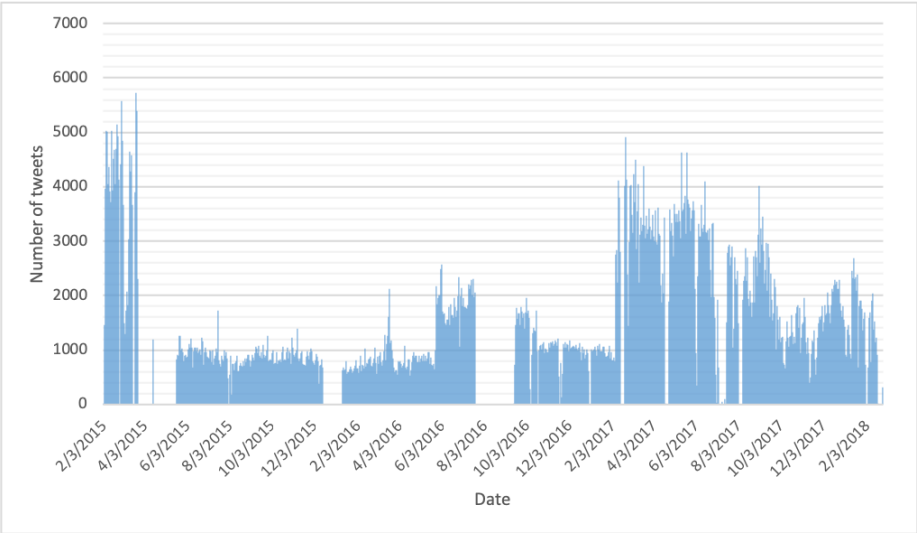

Figure 6: Distribution of geolocated tweets between 04-Feb-2015 and 26-Feb-2018].

Information Retrieval

The dataset is formed by a set of tweets collected during several years. The tweets are geolocated in the city of Valencia (Spain) between the dates 04-Feb-2015 and 26-Feb-2018 [54]. We analyzed a total of 1,483,338 tweets from 72,717 different users. Figure 6 shows the distribution of tweets over time. The intervals without tweets represent a drop from the tweet collection server due to technical problems. However, this does not affect the experiments carried out, since the biggest time interval that we use in the experiments is a day. The information that we have retrieved from each tweet is its geolocation, date, user identifier, and user name. The rest of the fields that the Twitter API provides were discarded.

The dataset of tweets is used to extract the sequence of coordinates that users publish along a day in the city. Each user u has associated a sequence S_u with the tweets that he or she has published. A path P_u is generated based on the latitude and longitude where a user published the tweet. These points allow us to know where a user has been in the city along the day. Linking the dots with the location of these tweets by straight lines we can generate the path that

has followed approximately every user. The "day" time unit has been chosen for each user to see how he has moved throughout a 24-hour period through the city. However, as will be shown below, certain subdivisions of paths have been made taking into account their shape.

Data Exclusion

For each of these users, we remove the repetitions of consecutive tweets with the same geolocation. Therefore, always there is a variation from point i to point $i+1$ in the path P_u.

For each day, all users who have published at least three tweets are considered. Finally, paths with less than two points were eliminated, since they do not provide useful information.

To evaluate the precision between a path generated by one of the proposed strategies and a real path, we have used the Hausdorff distance $d_H(P,Q)$ [55]:

$$d_{\mathrm{H}}(P,Q) = \max\{ \sup_{p\in P} \inf_{q\in Q} d(p,q), \sup_{q\in Q} \inf_{p\in P} d(p,q) \}, \tag{9}$$

where P is the subset of points of a real path and Q is the subset of points of a path generated by a strategy; $d(p,q)$ is the distance between a point $p \in P$ and a point $q \in Q$; sup represents the *supremum* and inf the $infimum$. The function $d_H(P,Q)$ calculates the greatest of all the distances from a point in a real path to the closest point in the generated path. The lower value of a Hausdorff distance, the lower error of the strategy that is predicting the path.

We compared the results of the proposed strategies (i.e., attraction and geodesic) with a set of baseline methods. We considered the following baseline strategies:

- Euclidean distance-based: this method creates a straight-line between the starting point A and the destination point B.
- Manhattan distance-based: this method proposes that the movements should be discretized in a grid along axes at right angles onto the coordinate axes.

5. Experiments and results

5.1. Experiment 1. Path segmentation.

This experiment tries to fix the problem of predicting a path that has closed angles in it. When we observe this type of paths, where there is a round trip, we can assume that it represents two different paths (see Figure 7).

We evaluate the precision of the proposed strategies (i.e., the distance between the real path and the path generated by the gravity-based strategies and the baseline strategies). To solve the problem mentioned above of round-trip paths, we first proceed with a path segmentation process.

Segmentation is the process of dividing a movement trajectory into sub-trajectories, called segments, where each segment fulfills certain criteria [56]. In this scenario, paths that contain a transition between three points that form a closed angle were divided in two, creating two segments.

Let's consider

$$P_u = \big\langle (x_0, y_0), \ldots, (x_{i-1}, y_{i-1}), (x_i, y_i), (x_{i+1}, y_{i+1}), \ldots, (x_n, y_n) \big\rangle$$

as a path which has a closed angle α ($\alpha < \frac{\pi}{2}$) at point i. Therefore, path P_u should be split into two segments $P_{u_1} = \langle (x_0, y_0), \ldots, (x_{i-1}, y_{i-1}), (x_i, y_i) \rangle$ and $P_{u_2} = \langle (x_i, y_i), (x_{i+1}, y_{i+1}), \ldots, (x_n, y_n) \rangle$ (see Figure 7). The i point that joins two segments a and b can be automatically detected when $\arccos(\frac{ab}{|a||b|}) < \frac{\pi}{2}$.

Figure 7 shows an example of a sequence of temporarily ordered geo-located tweets generated by a user during a day. Points 2, 3, and 4 form a closed angle. Therefore, we split the path $P_u = \langle 0, 1, 2, 3, 4, 5, 6 \rangle$ into two paths. The first path is composed of points $P_{u_1} = \langle 0, 1, 2, 3 \rangle$ and the second path is composed of $P_{u_2} = \langle 3, 4, 5, 6 \rangle$. Notice how both resulting paths share the point that formed the angle.

Based on the tweets dataset and considering the previous restriction, 6,972 real paths were generated. Most of these paths (nearly 97%) are three or four points long. Table 2 compares the total and average length of the paths grouped by their number of steps. Figure 8 shows the length of paths in kilometers. On

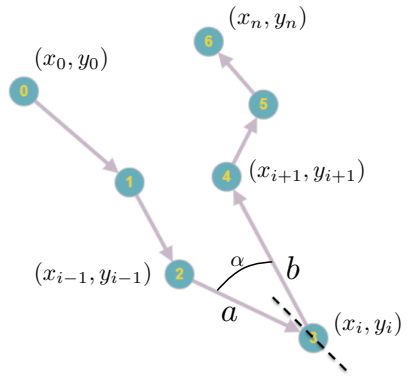

Figure 7: An example of path with a closed angle that is split into two paths.

the whole, 57% of path lengths consist of a few kilometers (i.e., between 0.5 to 4 km).

Path Length	**3**	**4**	**5**	**6**	**7**	**8**
# Paths	5,876	879	162	43	9	6
Percentage	84.24%	12.60%	2.32%	0.62%	0.13%	0.09%
Average (km)	4.73	5.66	6.18	4.38	10.66	3.83

Table 2: Analysis of real paths generated.

In Table 3, it can be observed that the Euclidean strategy obtains good results (i.e., a low value of Hausdorff metric) compared to any of the other strategies. In the attraction strategy, the best results occur when the attraction factor Λ is very high, generating almost a straight line at one step. This result could correspond to the linearity of the real paths. As paths are very short, the straight line is always prevailing.

To overcome this weakness, we also evaluate the precision of the strategies grouping paths by their minimum length. Table 4 shows the average Hausdorff distance ($\overline{d_H(X,Y)}$) obtained by grouping these paths by minimum length.

By considering the restriction of the minimum number of points for a path, the number of paths generated in each case decreases drastically. In the same

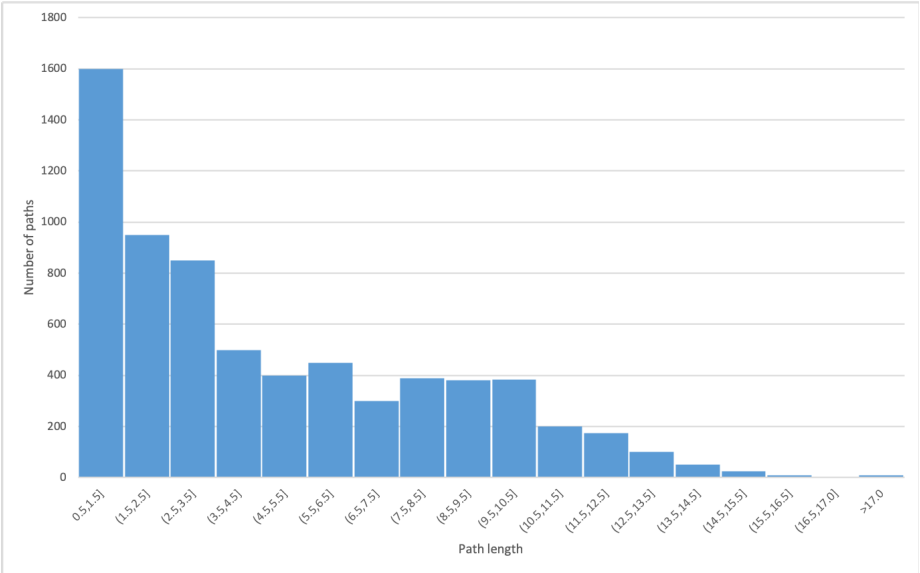

Figure 8: Distribution of number of paths over distances in kilometers.

Strategies	$\overline{\mathbf{d_H(X,Y)}}$
Euclidean	683.2
Manhattan	1,235.5
Attraction	1,285.9
Geodesic	1,380.8

Table 3: Hausdorff values obtained with each strategy using the path segmentation process.

way, Hausdorff values increase a little, although it is not determinant. This result comes from the fact that, possibly, paths of 3 or 4 points were the ones that most tended to be a straight line. By removing those paths, the ones that remained were more complex, generating a higher average of Hausdorff values.

5.2. Experiment 2. Segmentation and reassembly of paths.

In this experiment, we evaluate the precision of the proposed strategies (i.e., the distance between the real path and the path generated by the strategies). The difference with the previous experiment is that in this scenario, to generate

Minimum Path Length	**3**	**4**	**5**	**6**	**7**
# paths	6,972	1,096	218	55	13
Euclidean	683.2	970.8	1,099.5	1,008.7	1,619
Manhattan	1,235.5	1,393.4	1,378.6	1,135.4	1,852.9
Geodesic	1,380.8	1,557.9	1,603.4	1,390.7	1,835.4
Attraction (best Λ parameter)	1.285,9 (400)	1.467,3 (300)	1.456,6 (300)	1.273,2 (300)	2.011 (300)

Table 4: Hausdorff values grouping paths by minimum length using the path segmentation process.

more realistic paths, we add a second stage that reassembles the path segments after making a model for each of the segments that have been split. Specifically, first we split the sequence of points if the angle of three consecutive points is less than 90 degrees ($\frac{\pi}{2}$). Then, we model each part of the original path using the proposed strategies. Finally, for the evaluation, we reassembly the original path (i.e., without splitting). In this way, all the tweets that a user has made during a day are part of the path.

Figure 9 illustrates an example of this process for building and comparing paths generated by the proposed strategies and the real paths. The original path $P_u = \langle 0, 1, 2, 3, 4, 5, 6, 7\rangle$ is split into three segments at the time of prediction: $P_{u_1} = \langle 0, 1, 2, 3\rangle$, $P_{u_2} = \langle 3, 4, 5, 6\rangle$ and $P_{u_3} = \langle 6, 7\rangle$. Then, the predictions of these three paths are made separately. These predictions are reconstructed from the points of separation $\langle 3\rangle$ and $\langle 6\rangle$ as well as the original path, returning to $P_u = \langle 0, 1, 2, 3, 3, 4, 5, 6, 7\rangle$. The original reassembled path is then compared to the path reconstructed using the proposed strategies and the process described above.

The average Hausdorff distance obtained with the proposed strategies and considering the reassembling process described above shows that the Euclidean strategy obtained much better values than the rest of the strategies (see Table 5). The values obtained are still not as good as expected. The deviation appears

because it is necessary to maintain many two-point paths to guarantee the reassembly of paths. This approach is not a realistic one at the time of the prediction since there is no better predictor than the straight line. This effect is accentuated by the fact that 83.2% of paths are two-point length ones.

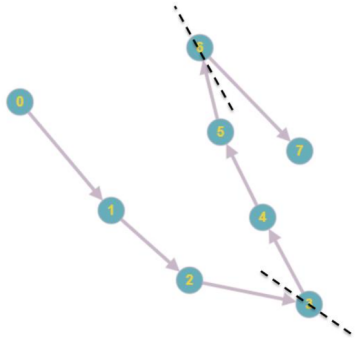

Figure 9: Example of an original path $\langle 0, 1, 2, 3, 4, 5, 6, 7 \rangle$ divided into three partial paths for the prediction: $\langle 0, 1, 2, 3 \rangle$, $\langle 3, 4, 5, 6 \rangle$ and $\langle 6, 7 \rangle$.

Strategy	$\overline{\mathbf{d_H(X, Y)}}$
Euclidean	164.2
Manhattan	1,167.4
Attraction	1,037.2
Geodesic	1,298.8

Table 5: Hausdorff values obtained with each strategy using the split and reassembly process.

5.3. Experiment 3. Realistic paths.

Previous experiments have led to the conclusion that users' tweets are not enough information for the generation of the original paths that they have done [57]. Joining the geo-located tweets with straight lines is merely wrong, it does not represent reality. To deal with this problem, we decided to replace the way in which two consecutive points are joined.

Instead of using a straight line, a more detailed path is used. The path between two points is generated by a route planning service, which is a much more realistic approach to generating paths from consecutive coordinates. This generator ignores factors such as bottlenecks, preferences, and others that the model tries to predict, so it gives a good starting point. We used a server with an instance of Project OSRM [58] to generate realistic paths that follow

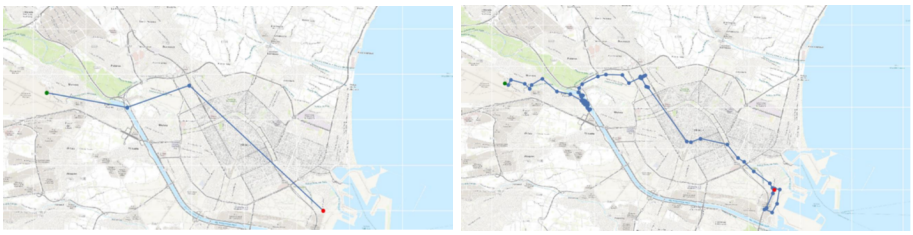

Figure 10: (Left) Example of a path generated considering only the information of tweets. (Right) Example of a path generated considering the information of tweets and information from the OSRM Service.

the directions and streets that the city has. In this way, the original paths are replaced by more detailed ones. These new realistic paths share the same intermediate points that P_u has, but include much more points between the original ones which define trajectories that are feasible paths through the streets of a city, taking into account the layout of streets and their directions. Figure 10 (Left) shows a path generated by tweets joined with straight lines, as previously done, while Figure 10 (Right) shows a path generated by OSRM data using tweets as intermediate points.

Creating these realistic paths as the set of paths to be compared to the paths that our strategies modeled gives us a dataset with better path lengths, where there are fewer short length paths and their distribution is much more regular, most focusing on length 26 and away from straight line paths (Figure 11). Results of the Hausdorff distance between these realistic paths and the four strategies we are comparing in this paper (see Table 6) show that the attraction model obtains better results than the other strategies.

5.4. Algorithm

The previously presented methods have been combined in the form of an algorithm to prepare the data to be compared with the gravitational potential proposal (both the attraction strategy and the geodesic strategy) in order to generate a model of the movements of users in a city following our gravitational approach. This algorithm is shown in Algorithm 1.

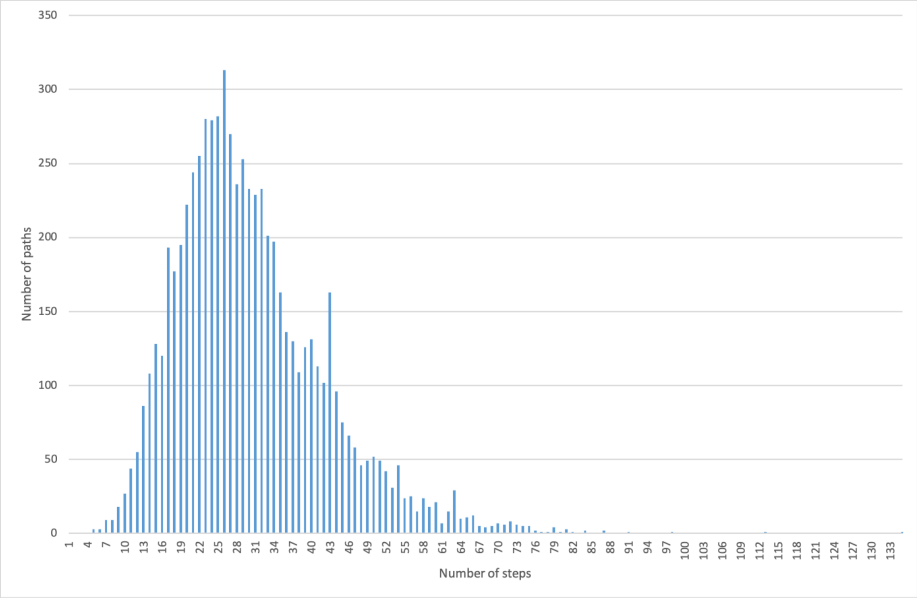

Figure 11: Distribution of realistic paths lengths after OSRM processing.

After applying this algorithm to the original data we are ready to use it to compare with the model presented in this work.

5.4.1. Parameter tuning

From the tests of previous subsections emerged the best strategy to model paths, which is the strategy of attraction with realistic paths.

Next, we performed more tests to determine the best attraction parameter (Λ). It is interesting to analyze the number of paths that do not reach their

Strategy	$\overline{\mathbf{d_H}(\mathbf{X}, \mathbf{Y})}$
Euclidean	1,185.1
Manhattan	1,203.9
Geodesic	1,417.9
Attraction	1,176.7 ($\Lambda = 1,000$)

Table 6: Hausdorff values obtained with each strategy considering realistic paths.

Result: A set of points that represent the movement of a user in a city.

1. Convert list of tweets to paths: $S_u \rightarrow P_u$
2. Segmentation: $P_u \rightarrow \langle P_{u_1}, P_{u_2}, \ldots, P_{u_n} \rangle$
3. Adjust each segment to city layout (OSRM): $P_{u_i} \rightarrow P_{R_i}$
4. Reassembly: $P_R = P_{R_1} + P_{R_2} + \ldots + P_{R_n}$

Algorithm 1: Adjust tweets to city movements

destination with $\Lambda = 0$: 6,675 out of 6,972. This failure happens because the paths fall into local minimums that they cannot leave because they obey the gradients of the model. However, this number decreases as the attraction parameter increases (see Table 7).

When calculating metrics with different attraction parameters, a decrease in the number of failed paths can be observed as well as the decreasing of the Hausdorff average distance. The best result comes at the inflection point where almost all paths reach the target, as shown in Table 7. An attraction factor of 40 generates all but two paths successfully. It is interesting to see that, if we increase too much the value of the attraction parameter, the error would grow a little, even though all paths are still generated correctly. This effect is related to the fact that, if the factor is extremely large, the paths generated would be

Λ	**-10**	**1**	**10**	**20**	**30**	**40**
$\mathbf{d_H}$	5,704.5	2,907.1	1,561.2	1,286.5	1,212.5	**1,197.5**
Failed	6,169	3,777	990	137	16	2
Λ	**50**	**100**	**500**	**1000**	**2000**	**5000**
$\mathbf{d_H}$	1,201.8	1,210.9	1,203.9	1,202.8	1,202.5	1,202.8
Failed	0	0	0	0	0	0

Table 7: Haussdorf distance and number of failed paths considering different values for the attraction parameter for paths generated with the gravitational model and the attraction strategy .

straight lines that would not conform to the real path.

Finally, Figure 12 shows examples of paths generated using the attraction strategy with this Λ value compared to the real path.

5.5. Gender analysis

The Twitter dataset used in the experiments allows us to classify and analyze paths by gender. For the classification of paths by gender, profile information was taken from users. As the structure of a tweet does not provide gender information, we considered the profile name of the users that generated a geo-tagged tweet. The name of each user account was compared with a list of names and genders taken from the GenderReader project [59]. This file provides an international list of names, with their most likely associated gender.

However, not all users publish their real name when creating a Twitter account. Therefore, gender information could not be obtained from all users. Of the total number of previously real paths generated (i.e., 6,972 paths) only 2,826 could be classified (40%). Of those paths, 1,910 paths (27%) were classified as paths traveled by men and 916 by women (13%).

First, we analyzed whether there are differences between paths considering the users' gender. Figure 13 shows the distribution of paths lengths for men and women, respectively. It can be observed that they have a similar distribution, with a slight inclination to long paths on the part of the male gender. Men have an average of 4.73 km, while women have an average of 4.15 km. The standard deviation for men is 4.1 km and 3.8 km for women.

Figure 12: Examples of an original path (blue) and a generated path (green).

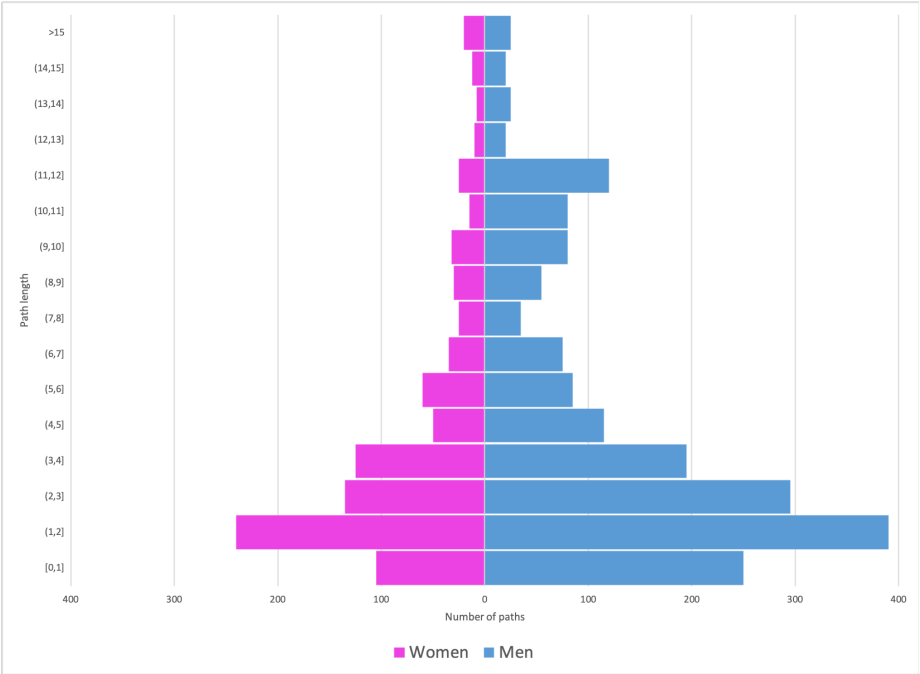

Figure 13: Path length distribution for men and women.

When performing the Student's t-test to analyze the distance distributions of the paths of these two groups, a p value of 0.000334 is obtained. Therefore, the null hypothesis (men and women paths are equal) is rejected, and we can conclude that there are statistically significant differences between the lengths of paths of men and women.

We also evaluated the average Hausdorff distance in paths for each gender using the attraction strategy, which turned out to be the one with the best results. These distances are different for men and women and the results are shown (together with the summary of the rest of the results of this experiment) in Table 8.

In addition, to better understand the previous results, we calculated the average sinuosity of paths [60]. This metric is the index that represents how much the layout of a path deviates from the straight line. The calculation of

sinuosity ϑ for path P is given by:

$$\vartheta_P = \frac{len(P)}{\sqrt{(P_{x_0} + P_{x_n})^2 + (P_{y_0} + P_{y_n})^2}} \tag{10}$$

where len(P) is the length of the path P; P_{x_0} is the latitude of the starting point; P_{x_n} is the latitude of the end point; P_{y_0} is the longitude of the starting point; P_{y_n} is the longitude of the endpoint. The average sinuosity results for men are 1.168 and for women 1.201. Therefore, women's paths are more sinuous than those of men. The more sinuous a path means that the movement is not direct from point A to point B, but that there are several stops along the way.

We also analyzed a related metric, the average of the angles of each path. That is, for each path, three consecutive points were taken, and the angle generated by the two segments was accumulated. Figure 14 shows an example of a path, whose total accumulated angle would be $\alpha + \beta$ and its average $(\frac{\alpha+\beta}{2})$. We calculated this metric for all paths, and it was obtained that the average angle of men was 26.1 degrees, while that of women was 32.2 degrees.

The result metrics of this experiment provide clues to support the hypotheses about gender differences in types of urban mobility. They show that men usually take longer paths, from narrower angles, but less sinuous. On the other hand, women take more open angles but with more sinuous paths [18], albeit from

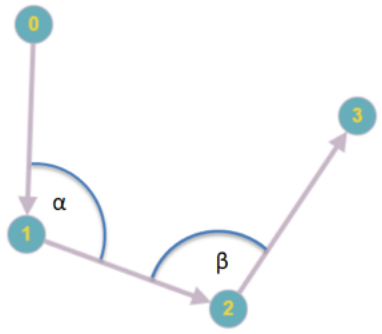

Figure 14: Path angle.

	distance	sinuosity	angle
men	1253,9	1168	26,1
women	1121,7	1201	32,2

Table 8: Summary of gender analysis results

slightly shorter roads [19].

6. Conclusions

In this paper, we explored urban human mobility using a collection of geotagged messages from Twitter ($n = 1.48 \times 10^6$). Based on this data, we proposed a gravitational potential model that is represented as a contour map. This map shows how the population of a city is attracted to each point in the city considering the slopes around certain points. To navigate this map, we propose two strategies: the attraction strategy and the geodesic strategy. In order to evaluate these strategies we first had to convert the real data extracted from the dataset to realistics paths to compare with (n=2,826 paths). We proposed a set of methods, combined in the form of an algorithm, which have also been incrementally tested one by one. This algorithm has been used to compare our strategies with baseline strategies (the euclidean and Manhattan strategies). The combination of gravitational potential with attraction to the destination points provides the best results, with $d_H = 1176$ against the Manhattan ($d_H = 1203$) or the geodesic ($d_H = 1417$) alternatives. Therefore, we can conclude that the proposed strategies are more suitable for predicting long realistic paths than short paths that could be simplified by a straight line.

Moreover, we also studied the paths by gender. We analyzed the length and the sinuosity of men and women paths. The results show that in general, women's paths are characterized by short distances ($d_M = 4.73$ km, $d_W = 4.5$ km, with p=0.0014) and high sinuosity ($\vartheta_M = 1.17$ km, $\vartheta_W = 1.2$ km, with p=0.042). These empirical results coincide with the results obtained through surveys in other studies. Therefore, we can also conclude that the results ob-

tained in experiments with Twitter data show that social networks can be used as an alternative social sensor to traditional information sources.

The benefits of this work can be relevant to different stakeholders. For example for city managers it will lead to improved public transport services as well as the urban organization of the city. For citizens, it will allow them to detect points of interest or temporary concentration of pedestrians to help them choose alternative routes. In addition, applications that suggest routes to their users may incorporate additional, more accurate information on the use of urban spaces in real time.

The findings of this paper open new lines for future investigations of human mobility. From the obtained results, we can state that urban human mobility depends on urban structures and morphology. Although this study focuses on the city of Valencia, the analysis can be extrapolated to other cities with different geographical scales. Moreover, in addition to the urban structure and gender, other sociodemographic factors from user profiles in Twitter could be considered. For instance, we can analyze the paths differences between citizens and tourists. This can be useful for assisting on urban decision-making processes. These sets of analysis will be performed in future work.

Acknowledgements

This work is partially supported by Spanish Government Project TIN2015-65515-C4-1-R and the Post-doc grant Ref. SP20170057

References

[1] H. Barbosa, M. Barthelemy, G. Ghoshal, C. R. James, M. Lenormand, T. Louail, R. Menezes, J. J. Ramasco, F. Simini, M. Tomasini, Human mobility: Models and applications, Physics Reports 734 (2018) 1–74.

[2] C. M. Schneider, V. Belik, T. Couronné, Z. Smoreda, M. C. González, Unravelling daily human mobility motifs, Journal of The Royal Society Interface 10 (84) (2013) 20130246. doi:10.1098/rsif.2013.0246.

[3] G. Sagl, M. Loidl, E. Beinat, A visual analytics approach for extracting spatio-temporal urban mobility information from mobile network traffic, ISPRS International Journal of Geo-Information 1 (3) (2012) 256–271.

[4] L. Gabrielli, S. Rinzivillo, F. Ronzano, D. Villatoro, From tweets to semantic trajectories: mining anomalous urban mobility patterns, in: Citizen in Sensor Networks, Springer, 2014, pp. 26–35.

[5] S. Hasan, C. M. Schneider, S. V. Ukkusuri, M. C. González, Spatiotemporal patterns of urban human mobility, Journal of Statistical Physics 151 (1-2) (2013) 304–318.

[6] C. Song, Z. Qu, N. Blumm, A.-L. Barabási, Limits of predictability in human mobility, Science 327 (5968) (2010) 1018–1021.

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

[8] J. Jordán, J. Palanca, E. del Val, V. Julian, V. Botti, A multi-agent system for the dynamic emplacement of electric vehicle charging stations, Applied Sciences 8 (2) (2018) 313.

[9] J. Jordán, J. Palanca, E. del Val, V. Julian, V. Botti, Using genetic algorithms to optimize the location of electric vehicle charging stations, in: The 13th International Conference on Soft Computing Models in Industrial and Environmental Applications, Springer, 2018, pp. 11–20.

[10] G. Xue, Z. Li, H. Zhu, Y. Liu, Traffic-known urban vehicular route prediction based on partial mobility patterns, in: Parallel and Distributed Systems (ICPADS), 2009 15th International Conference on, IEEE, 2009, pp. 369–375.

[11] S. Jiang, G. A. Fiore, Y. Yang, J. Ferreira Jr, E. Frazzoli, M. C. González, A review of urban computing for mobile phone traces: current methods,

challenges and opportunities, in: Proceedings of the 2nd ACM SIGKDD international workshop on Urban Computing, ACM, 2013, p. 2.

[12] K. Zhao, M. Musolesi, P. Hui, W. Rao, S. Tarkoma, Explaining the power-law distribution of human mobility through transportation modality decomposition, Scientific reports 5 (2015) 9136.

[13] H. Xu, S. Gupta, M. B. Rosson, J. M. Carroll, Measuring mobile users' concerns for information privacy, International Conference on Information Systems, ICIS 2012 3 (2012) 2278–2293.

[14] R. Jurdak, K. Zhao, J. Liu, M. AbouJaoude, M. Cameron, D. Newth, Understanding human mobility from twitter, PloS one 10 (7) (2015) e0131469.

[15] T. P. Uteng, T. Cresswell, Gendered mobilities: towards an holistic understanding, in: Gendered mobilities, Routledge, 2016, pp. 15–26.

[16] P. Gordon, A. Kumar, H. W. Richardson, Gender differences in metropolitan travel behaviour, Regional Studies 23 (6) (1989) 499–510.

[17] H. Best, M. Lanzendorf, Division of labour and gender differences in metropolitan car use: an empirical study in cologne, germany, Journal of Transport Geography 13 (2) (2005) 109–121.

[18] E. Zucchini, Género y transporte: análisis de la movilidad del cuidado como punto de partida para construir una base de conocimiento más amplia de los patrones de movilidad. el caso de madrid, Doctoral dissertation, Universidad Politcnica de Madrid (2015).

[19] A. Mena, M. Soler, La movilidad urbana de hombres y mujeres en la ciudad de valencia, Tech. Rep. 978-84-606-5641-8, Ayuntamiento de Valencia (September 2014).

[20] O. Blumen, Gender differences in the journey to work, Urban Geography 15 (3) (1994) 223–245.

[21] V. Basari, A. Vujii, J. M. Simi, V. Bogdanovi, N. Sauli, Gender and age differences in the travel behavior a novi sad case study, Transportation Research Procedia 14 (2016) 4324 – 4333.

[22] S. Maffii, P. Malgieri, C. Di Bartolo, Gender equality and mobility: mind the gap, CIVITAS WIKI Policy Analyses, online; accessed 4 June 2019 (2014).
URL https://civitas.eu/sites/default/files/civ_pol-an2_m_web.pdf

[23] S. Hanson, Gender and mobility: new approaches for informing sustainability, Gender, Place & Culture 17 (1) (2010) 5–23.

[24] J. Gonalves, M. Gomes, S. Ezequiel, Defining mobility patterns in peri-urban areas: A contribution for spatial and transport planning policy, Case Studies on Transport Policy 5 (4) (2017) 643 – 655. doi:https://doi.org/10.1016/j.cstp.2017.07.009.

[25] M. Berlingerio, F. Calabrese, G. Di Lorenzo, R. Nair, F. Pinelli, M. L. Sbodio, Allaboard: a system for exploring urban mobility and optimizing public transport using cellphone data, in: Joint European Conference on Machine Learning and Knowledge Discovery in Databases, Springer, 2013, pp. 663–666.

[26] G. K. Zipf, The p 1 p 2/d hypothesis: on the intercity movement of persons, American sociological review 11 (6) (1946) 677–686.

[27] S. A. Stouffer, Intervening opportunities: a theory relating mobility and distance, American sociological review 5 (6) (1940) 845–867.

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

[29] I. Rhee, M. Shin, S. Hong, K. Lee, S. J. Kim, S. Chong, On the levy-walk nature of human mobility, IEEE/ACM transactions on networking (TON) 19 (3) (2011) 630–643.

[30] F. Calabrese, M. Diao, G. Di Lorenzo, J. Ferreira, C. Ratti, Understanding individual mobility patterns from urban sensing data: A mobile phone trace example, Transportation research part C: emerging technologies 26 (2013) 301–313.

[31] R. Becker, R. Cáceres, K. Hanson, S. Isaacman, J. M. Loh, M. Martonosi, J. Rowland, S. Urbanek, A. Varshavsky, C. Volinsky, Human mobility characterization from cellular network data, Communications of the ACM 56 (1) (2013) 74–82.

[32] M. Lin, W.-J. Hsu, Mining gps data for mobility patterns: A survey, Pervasive and Mobile Computing 12 (2014) 1–16.

[33] L. Wu, Y. Zhi, Z. Sui, Y. Liu, Intra-urban human mobility and activity transition: Evidence from social media check-in data, PloS one 9 (5) (2014) e97010.

[34] A. Lima, R. Stanojevic, D. Papagiannaki, P. Rodriguez, M. C. González, Understanding individual routing behaviour, Journal of The Royal Society Interface 13 (116).

[35] J. Gómez-Gardeñes, D. Soriano-Paños, A. Arenas, Critical regimes driven by recurrent mobility patterns of reaction–diffusion processes in networks, Nature Physics 14 (4) (2018) 391–395.

[36] M. Lee, H. Barbosa, H. Youn, G. Ghoshal, P. Holme, Urban socioeconomic patterns revealed through morphology of travel routes, arXiv preprint arXiv:1701.02973.

[37] Y. Yuan, M. Medel, Characterizing international travel behavior from geotagged photos: A case study of flickr, PloS one 11 (5) (2016) e0154885.

[38] E. del Val, J. Palanca, M. Rebollo, U-tool: A urban-toolkit for enhancing city maps through citizens activity, in: Advances in Practical Applications of Scalable Multi-agent Systems. The PAAMS Collection, Springer, 2016, pp. 243–246.

[39] J. Frith, Turning life into a game: Foursquare, gamification, and personal mobility, Mobile Media & Communication 1 (2) (2013) 248–262.

[40] J. Béjar, S. Álvarez, D. García, I. Gómez, L. Oliva, A. Tejeda, J. Vázquez-Salceda, Discovery of spatio-temporal patterns from location-based social networks, Journal of Experimental & Theoretical Artificial Intelligence 28 (1-2) (2016) 313–329.

[41] F. Terroso-Sáenz, J. Cuenca-Jara, A. González-Vidal, A. F. Skarmeta, Human mobility prediction based on social media with complex event processing, International Journal of Distributed Sensor Networks 12 (9) (2016) 5836392.

[42] K. Siła-Nowicka, J. Vandrol, T. Oshan, J. A. Long, U. Demšar, A. S. Fotheringham, Analysis of human mobility patterns from gps trajectories and contextual information, International Journal of Geographical Information Science 30 (5) (2016) 881–906.

[43] B. Hawelka, I. Sitko, E. Beinat, S. Sobolevsky, P. Kazakopoulos, C. Ratti, Geo-located twitter as proxy for global mobility patterns, Cartography and Geographic Information Science 41 (3) (2014) 260–271.

[44] M. G. Beiró, A. Panisson, M. Tizzoni, C. Cattuto, Predicting human mobility through the assimilation of social media traces into mobility models, arXiv preprint arXiv:1601.04560.

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

[46] F. Figueiredo, B. Ribeiro, J. M. Almeida, C. Faloutsos, Tribeflow: Mining & predicting user trajectories, in: Proceedings of the 25th International Conference on World Wide Web, International World Wide Web Conferences Steering Committee, 2016, pp. 695–706.

[47] F. Simini, M. C. González, A. Maritan, A.-L. Barabási, A universal model for mobility and migration patterns, Nature 484 (7392) (2012) 96–100.

[48] T. Jia, B. Jiang, K. Carling, M. Bolin, Y. Ban, An empirical study on human mobility and its agent-based modeling, Journal of Statistical Mechanics: Theory and Experiment 2012 (11) (2012) P11024.

[49] P. A. Grabowicz, J. J. Ramasco, B. Gonçalves, V. M. Eguíluz, Entangling mobility and interactions in social media, PloS one 9 (3) (2014) e92196.

[50] Á. Molas Martín, Field theory for recurrent mobility, Master thesis, Universitat DE LES Illes Balears (2018).

[51] A. Einstein, Die grundlage der allgemeinen relativitätstheorie, Annalen der Physik 354 (7) (1916) 769–822.

[52] P. Bose, A. Maheshwari, C. Shu, S. Wuhrer, A survey of geodesic paths on 3d surfaces, Computational Geometry 44 (9) (2011) 486 – 498.

[53] E. W. Dijkstra, A note on two problems in connexion with graphs, Numerische mathematik 1 (1) (1959) 269–271.

[54] J. Palanca, E. del Val, M. Rebollo, Twitter Publications Coordinates in Valencia area, [Dataset] doi:10.6084/m9.figshare.8255855.v1 (2019).

[55] D. P. Huttenlocher, G. A. Klanderman, W. J. Rucklidge, Comparing images using the hausdorff distance, IEEE Transactions on pattern analysis and machine intelligence 15 (9) (1993) 850–863.

[56] M. Buchin, H. Kruckenberg, A. Kölzsch, Segmenting trajectories by movement states, in: Advances in spatial data handling, Springer, 2013, pp. 15–25.

[57] C. Kang, X. Ma, D. Tong, Y. Liu, Intra-urban human mobility patterns: An urban morphology perspective, Physica A: Statistical Mechanics and its Applications 391 (4) (2012) 1702–1717.

[58] D. Luxen, C. Vetter, Real-time routing with openstreetmap data, in: Proceedings of the 19th ACM SIGSPATIAL international conference on advances in geographic information systems, ACM, 2011, pp. 513–516.

[59] Genderreader, `https://github.com/cstuder/genderReader`, accessed: 2018-10-26.

[60] H.-H. Stølum, River meandering as a self-organization process, Science 271 (5256) (1996) 1710–1713.