Analyzing mobility patterns from social networks and social, economic and demographic open data

Caio Libânio Melo Jerônimo, Claudio E. C. Campelo, Cláudio de Souza Baptista

Systems and Computing Department, Federal University of Campina Grande (UFCG) - Campina Grande, Pb - Brazil

caiolibanio@copin.ufcg.edu. br {campelo, baptista} @ computacao.ufcg.edu.br

***Abstract**. With the increased facility in acquiring georeferenced data from social networks, the interest in studying human mobility based on these data has grown, bringing new challenges and opportunities for knowledge discovery in GIScience. Even with this favorable scenario, few studies have attempted to analyze how the information produced from these networks may correlate with other aspects of human life. This paper presents an approach to extracting mobility patterns from Twitter messages and to analyzing their correlation with social, economic and demographic open data. The proposed model was evaluated using a dataset of georeferenced Twitter messages and a set of social indicators, both related to Greater London. The results revealed that social indicators related to employment conditions present higher correlation with the mobility patterns than any other social indicators investigated, suggesting that these social variables may be more relevant for studying mobility patterns.*

1. Introduction

Urban Mobility Patterns represent models of human behavior in an urban environment (Luo et. al., 2016) and are especially relevant as the analysis of these patterns may influence public transportation systems, public safety, traffic engineering, health systems, and many other fields related to the planning of urban centers (Noulas et al, 2012; Wilson and Bell, 2004). Mobility patterns are also present in studies related to recommendation systems (Hao et al., 2010; Ye et al., 2011; Zheng et al., 2010) as well as in researches on trajectories (Bagrow and Elin, 2012; Hsieh et al., 2012).

There are many studies addressing this subject, however, most of them are concerned with data from cell phone networks (Gonzalez et al., 2008; Jiang et al., 2013; Palchykov et al., 2014), Wifi networks (Chaintreau et al., 2007; Zhang et al., 2012) or GPS signals (Rhee et al., 2011; Zhao et al., 2015). Although these studies help understand mobility patterns, they have restrictions on privacy, as well as on precision, specially when using cell phone networks data.

The massive usage of social networks by different classes of people have lead to an increase in the amount of data that are generated in the internet, enabling the rise of new studies related to knowledge discovery. This phenomenon, allied to the usage of smartphones in daily life, and the capability of these devices to generate georeferenced data, have also favored studies related to mobility patterns, specially in urban centers.

It is known that large cities might have considerable economic, social and demographic discrepancies among their regions, which can influence the way locals move within such urban centers. Analyzing how these factors influence urban mobility is a major challenge to be considered, for example, in Points of Interest (POI) recommendation, in route prediction, or in urban planning systems.

In this paper we propose a model to extract mobility patterns from georeferenced data collected from social networks and to find statistical correlations between these patterns and social, economic and demographic open data from an urban center. Georeferenced information obtained from social media is typically imprecise and fragmented. For example, many people only post messages from certain locations (e.g from home, from work), even though they visit many other places; many users remain inactive for long periods. Thus, this research aims at verifying the feasibility of extracting mobility information from social media data that is relevant enough to be correlated with other open social data.

To evaluate our model, we collected Twitter[1] messages from Greater London for one year, totalizing 19,456,798 messages. For the social data, we used the public platform London Datastore (http://data.london.gov.uk/). These data were supplied to our model, allowing the discovery of some correlations between these informations, which indicates the practicability of using a model of this kind. Through the experiments we conducted, we found that the social variables related to employment conditions tend to correlate with the mobility patterns analyzed in this study, specially the following variables: economically inactive people, employment rate, unemployment rate and persons with no qualifications.

The remainder of this paper is organized as follows. The next section presents related work. Section 3 describes the mobility pattern properties used and some relevant related concepts. Then the experiments conducted to validate our model are addressed in Section 4. Section 5 discusses obtained results. Section 6 concludes the paper and points to future directions of this research work.

2. Related work

Most studies related to mobility patterns use data derived from cell phone networks, RFID devices, GPS based data, or Wifi networks. Recently, studies have addressed the task of extracting and identifying mobility patterns from social media data. This tendency is a consequence of the way that these networks offer their data, since this information is mostly available for public access, reducing financial costs applied to research projects.

Yuan et al. (2013) proposed a probabilistic model called W4 (Who + Where + When + What) to extract from Twitter messages aspects of mobility related to the users of this social network. The authors considered the spatial and temporal dimensions, and also the activities performed by the users.

Considering social networks as new data sources for current and future research in many different fields, some researchers have analyzed the suitability of this kind of data source. Jurdak et al. (2015) analyze mobility patterns considering spatial and temporal aspects related to the major cities of Australia, in order to demonstrate that

[1] Twitter - https://twitter.com

Twitter can be quite efficient when working with mobility patterns using georeferenced messages, where similar features were found by both Twitter and mobile phone networks data. Similarly, the research presented by Hawelka et al. (2014) only considers spatial and temporal aspects of Twitter data, however, they deal with global mobility scales (between countries). The study aims at revealing global mobility patterns related to these messages, demonstrating that these data have similar properties to other kinds of data sources used in different studies.

Hasan et al. (2013) categorize mobility patterns through user's activities around three major cities: New York, Chicago and Los Angeles. The differential in their research is that they consider, in addition to spatial and temporal aspects, the semantics of displacements. To do this, they analyze georeferenced messages from Twitter, using links to the Foursquare platform, which allows them to identify and categorize the check-ins as: (1) at home; (2) work; (3) meal; (4) entertainment activity; (5) recreation and (6) shopping.

The vast majority of researches relating mobility patterns and social network data only focus on the spatial and temporal aspects of these patterns. However, other aspects might be considered when studying mobility patterns, specially economic, social and demographic factors. In this context, the works that consider these dimensions are restricted and limited to a few variables in a social context.

Shelton et al. (2015) provide a conceptual and methodological framework to analyze inequalities in different regions of Louisville, Kentucky. The authors explore the spatial imaginaries of the citizens to divide the city into areas that they think it is more or less segregated in comparison to the rest of the city, mainly because of economic and racial/ethnic factors. They use Twitter messages to analyze how citizens travel around those divided areas and conclude that the popular imaginary could not be confirmed.

Cheng et al. (2011) investigate georeferenced Twitter messages, considering, in addition to spatial and temporal aspects, variables related to income, popularity on the social network, and the content of the messages. The authors try to find out some relations between these variables and the mobility patterns encountered in the Twitter messages. They conclude that people who live in cities with a higher average income, tend to get around for longer distances. Luo et al. (2016) investigate, in addition to spatial and temporal aspects, the following variables: ethnicity, age and gender. These social variables have been inferred from the users' profiles on Twitter and from public information provided by the government. The authors analyze how these three variables influence the mobility patterns extracted from the messages of Twitter related to the city of Chicago. They conclude that ethnicity was the most determining factor with regard to mobility patterns, possibly because this variable may express some socioeconomic characteristics of these users, demonstrating some level of segregation imposed to foreign people.

The social networks have made informations about mobility publicly available by the use of API's designed to explore the data generated everyday by these networks. Additionally, the governments are being encouraged to keep their populational data open for public access, generating a scenario where the information is easily accessible for any researcher who wants to discover new and hidden knowledge. Even with this favorable context, information related to population income, value of the properties of a region, crime rates and others are not analyzed nor correlated with mobility patterns, bringing the necessity of models that are able to perform such analysis, which could help in understanding the cities dynamics, guiding further studies related to urban centers and how people behave in these centers.

3. Mobility patterns and social analysis

Figure 1 shows an overview of our approach to detecting correlations between mobility patterns and social, economic and demographic data.

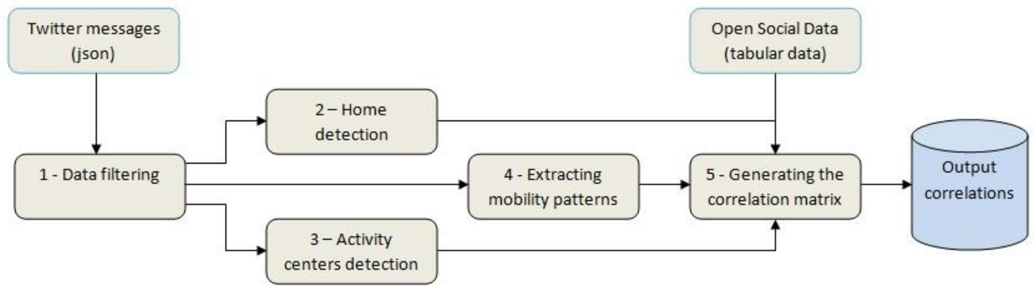

Figure 1 - Basic flow process of the proposed model

Our model accepts as input a set of Twitter messages in *json* format. The first step consists in filtering the Twitter messages. Initially, the model filters out messages without geographic coordinates information and those whose latitude/longitude coordinates do not point to a location inside the area of interest. Then, still in this filtering process, messages posted by stationary users are also filtered out. We consider as stationary those users whose messages are located within the same area, based on a predefined radius. Filtering this kind of users is important due to the fact that these users generally represent companies that report community information such as weather or traffic conditions, which are not relevant for the proposed model. Finally, similarly to Birkin et al. (2014), we remove all users with less than 20 messages, avoiding noise in the data provided by users with few posts.

In the second and third steps, the model detects the home and the activity centers (most visited regions) for each user. These concepts are used by the model to identify social characteristics of the users. The fourth step extracts statistical properties to express mobility patterns, that are: radius of gyration and user displacement distance (Luo et al., 2016; Cheng et al., 2011; Gonzalez et al., 2008; Hasan et al., 2013). These concepts will be explained in further sessions.

The fifth step consists in the correlation analysis, where the model receives the mobility patterns extracted by the previous process, and accepts tabular data containing the social variables and the polygon related to each region of the city/region to be analyzed, allowing the model to calculate the correlations between mobility and social data.

3.1. Radius of Gyration

The radius of gyration represents the standard deviation of distances between points of a trajectory and the center of mass of these points. This metric can measure how far and how frequently a user moves. A low radius of gyration indicates that a specific user tends to travel mostly locally, with few long-distance checkins, while a high value of this metric generally indicates that the user moves predominantly for long distances (Cheng et. al., 2011). This metric can be formalized in Equation 1 as:

$$r = \sqrt{\frac{1}{m}\sum_{i=1}^{m}(\mathrm{p_i} - \mathrm{p_c})^2}$$

Equation 1. Calculus of radius of gyration

Where:

- r represents the value for the radius of gyration for a user;
- m is the number of messages for a user;
- $\mathrm{p_i}$ represents a particular point where the message was posted;
- $\mathrm{p_c}$ represents the user center of mass (centroid);
- $(\mathrm{p_i} - \mathrm{p_c})$ is the distance between a particular point from the user's centroid;

3.2. User displacement distance

The user displacement distance represents the sum of the distances between all consecutive messages or checkins, reflecting the total distance traveled by the user around the analyzed geographic region.

Cheng et al. (2011) suggest that the behavior of user's displacement for Twitter messages follows a Lévy Flight distribution, which is characterized as a mixture of short and random displacements, with occasional long jumps. Shin et al. (2008) find similar results for the displacement distribution by analysing GPS data in different scenarios, such as metropolitan area and college campuses. In opposition to prevailing Lévy flight random walk models, Gonzalez et al. (2008), by analysing data from cell phone calls, highlight that human displacements have a significant level of temporal and spatial regularity, mainly because they tend to return to a few highly frequented locations.

3.3. Activity centers

An activity center can be defined as a location that a user frequently visits. The locations for an activity center could be a restaurant, home, place of work or any location where the user post his/her messages with some frequency. This concept appears to be an important parameter to express life patterns of a user, indicating the user's preferences for certain places or regions in a city.

To identify such activity centers, we used a popular clustering algorithm called DBSCAN (Density Based Spatial Clustering of Applications with Noise) (Ester et al. 1996). This is a density based clustering algorithm that clusters points that are closely located, and it can deal with points that are located in a low density regions, treating these points as noise. The algorithm has two parameters: the maximum radius of the neighborhood to be considered to form a cluster (ε); and the minimum number of points that a cluster must have (minPts). This algorithm presents advantages over other clustering algorithms, such as: it does not require the user to specify the number of clusters for the execution; it can deal with noise data efficiently; it can find a cluster surrounded by another cluster; and it relies on two variables only.

3.4. Home detection

The region of a user's home represents an important characteristic that needs to be considered in a study that analyze social, economic and demographic data. This is mainly because the user's home location may express social conditions, and these conditions might, in part, influence how the citizens move across an urban center.

Detecting home location based on user's center of mass of all checkins can lead to problems of splitting-the-difference, where a user that travels to distant regions over the city will have her home located at the middle of these regions (Cheng et al., 2011).

In this study, we consider home locations as the most intense activity centers during the night time, following other existing works (Luo et. al., 2016; Huang et al., 2014). For this, we select the messages posted between 8pm and 6am (on weekdays only); then we apply the DBSCAN algorithm to cluster the points of these messages; and finally we select the cluster with the greatest number of points as the user home.

4. Experimental Evaluation

This section presents the experimental evaluation conducted to validate our proposed model for identifying possible correlations between mobility patterns extracted from Twitter messages and social open data provided by governmental organizations.

4.1. Social data and maps

To perform the experiments, the city of London was chosen as a case study, specially due to the large volume of social, economic and demographic data publicly available for this city. In this study, data were collected from the governmental platform "London Datastore" (http://data.london.gov.uk/).

The social data collected for these experiments is related to the year of 2011, since this was the most recent data for the majority of the collected variables. Observed variables are related to the following categories: population age, family structure, ethnic groups, country of birth, house prices, economic activity, qualifications, health, car or van availability and religion. Additionally, we used the data level of LSOA (Lower Super Output Area), which is the smallest area used to divide the city of London for the available data. Each LSOA has an average population of 1,722 inhabitants. Figure 2 shows the map of the city of London used in this research. The map is divided by LSOA regions, and is originally made by the ONS (Office for National Statistics). This is under the terms of the Open Government Licence (OGL)[2] and UK Government Licensing Framework and is therefore free for this kind of use. This map was also acquired from the London Datastore platform.

[2]Contains National Statistics data © Crown copyright and database right [2012] and Contains Ordnance Survey data © Crown copyright and database right [2012].

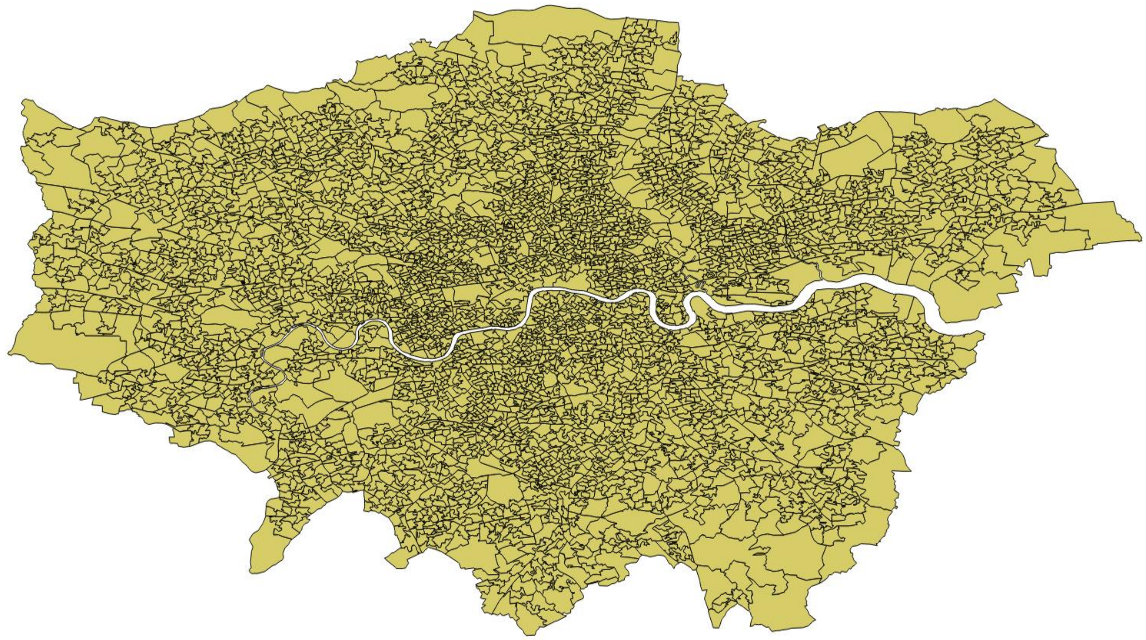

Figure 2 – City of London divided by LSOA regions

4.2. Twitter dataset

For this set of experiments, we collected georeferenced Twitter messages from the city of London via the Twitter API. The messages were collected from November 26th 2014 to November 22nd 2015, totalizing 19,456,798 messages from the region of London. From the initial set of messages, our model filtered out the messages without geographic coordinates information, resulting into 7,680,200 georeferenced messages and a total of 351,656 who have posted these messages. Then, after removing messages whose locations are outside the region of interest, messages posted from stationary users, and messages posted by light users, the dataset was further reduced to 6,203,474 georeferenced messages and 52,974 users, with an average of 117.1 messages per user.

4.3. Design of Experiments

The experiments executed in this research had the prior objective of answering the following questions:

- *Research question (Q1):* Do the characteristics of a user's home region influentiate his/her mobility patterns?
- *Research question (Q2):* Do the characteristics of the regions of a user’s activity centers influentiate his/her mobility patterns?

To answer these questions, the proposed model performs the Kendall’s correlation test to calculate possible correlations between user’s mobility patterns and open social data. We chose this test due to the fact that it neither requires a specific distribution of data nor a linear relation among the variables within the dataset.

We divided the experimental evaluation into two experiments. Experiment 1 consisted in analysing the first two metrics (radius of gyration and displacement distance). For this, we performed the correlation test between these metrics and all social data variables related to users' home locations (previously detected), selecting the correlations with τ >= 0.25 and with statistical significance (*p-value* < 0.05). This approach enables us to identify situations where the social, economic and demographic status of the region where a user's home is located may influentiate his/her mobility patterns, allowing us to answer Question *Q1*.

Experiment 2 has been conducted aiming at answering Question *Q2*. For this, we used the concept of activity centers described in Section 3.3. First, we clustered the dataset of points for each user with the DBSCAN algorithm using ε = 45 meters and *minPts* = 4 in order to find the user's activity centers. After this clustering process, we calculate the medians for all social data variables related to the regions in which these clusters were formed, for each user. Then we performed the correlation tests over these medians and the users' mobility patterns properties, allowing us to answer Question *Q2*. For both experiments, we adopted the value of 45 meters for filtering stationary users.

To perform these experiments, we divided the users into three categories: Category 1, 2 and 3. Respectively, they group the users who have posted at least 1,000 messages (679 users); 2,500 messages (168 users); and 5,500 messages (33 users). This division was made with the objective of identifying correlations that can only be found for heavy users, possibly due to the imprecision and fragmentation of messages posted in the Twitter network.

5. Results and Discussion

The histograms of the two mobility variables extracted by the model and used in this study are shown in Figure 3. In these histograms, it is possible to visualize the frequency at which the values of both variables occur in the extracted mobility dataset. The displacement distance histogram is shown in a log10 scale while the radius of gyration histogram is represented by its original values. Both variables were calculated in meters. In the first histogram, it can be seen that the majority of users have their radius of gyration between 3,000 meters and 4,000 meters, totalizing 7,554 users. The second histogram shows that most users have their displacement distance between 100,000 meters and 316,228 meters, representing 19,514 users.

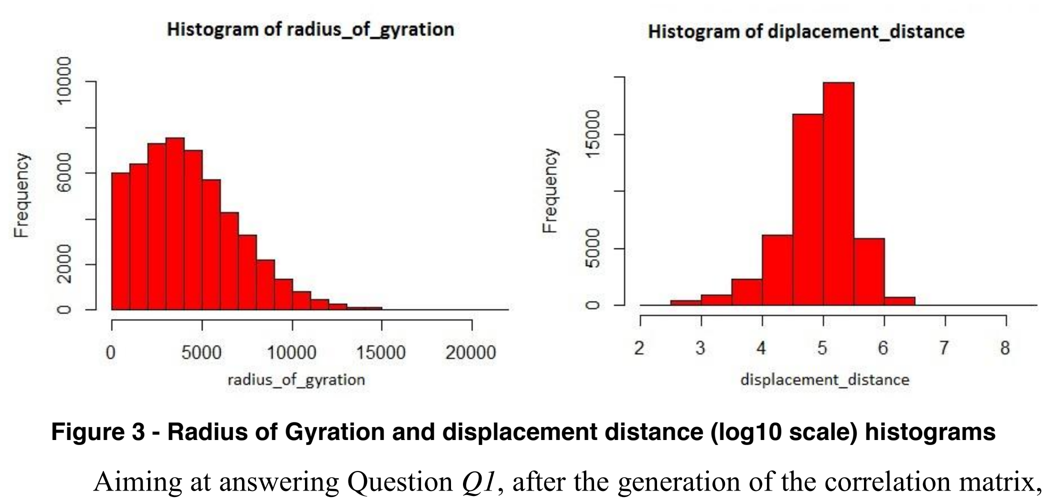

Figure 3 - Radius of Gyration and displacement distance (log10 scale) histograms

Aiming at answering Question *Q1*, after the generation of the correlation matrix, we selected the most relevant correlations found by the model. No significant correlation has been found for users from Categories 1 and 2 (users with at least 1,000 and 2,500 messages); however, some significant correlations were found for users from Category 3. The most significant correlations found for this group are shown in Table 1. In this table, we used the notation of a tuple (τ, *p-value*), where the first element represents the Kendall's *tau* (the correlation coefficient for this test), and the second represents the significance of the executed test, where a *p-value* less than 0.05 allows the rejection of the null hypothesis for the correlation test, denoting that there may be a real correlation between the variables of mobility patterns and social data.

Table 1 - Results for Experiment 1 (correlating radius of gyration and displacement distance with social variables). Values for users from Category 3.

Mobility / Social Data	Age 0-15	Couple with children	Economically inactive people	Employment rate	Unemployment rate	Persons with no qualifications
Radius of Gyration	-	-	(-0.35, 0.003)	(0.26, 0.03)	-	(-0.27, 0.02)
Displacement Distance	(-0.27, 0.02)	(-0.31, 0.01)	(-0.28, 0.02)	(0.29, 0.01)	(-0.38, 0.001)	(-0.29, 0.01)

From the results shown in Table 1 (which are related to Experiment 1), it can be observed that the highest correlation values were found for the analysis of social variables related to employment conditions. For instance, for the correlation between "Displacement Distance" and "Unemployment rate", we obtained τ = -0.38. This negative correlation expresses that as the "Displacement Distance" of a user increases, the value for the variable "Unemployment rate" (related to the user's home region) tends to decrease, indicating that users with longer traveled distances tend to live in regions with low unemployment rate. A similar finding is shown for the variables "Radius of Gyration" and "Economic Inactive People", with τ = -0.35. For the social variable "Employment Rate", when tested with the "Displacement Distance", we found a positive correlation, with τ = 0,29, denoting that the greater is the displacement distance of users, the greater is the employment rate of the region they live.

The results obtained for Experiment 1 allowed us to answer Question *Q1*, since we could find correlations with some statistical significance, denoting that there are correlations between users' mobility patterns and social aspects of their home location, specially for variables related to employment conditions.

For Experiment 2, where we analyze possible correlations between the two mobility metrics (radius of gyration and displacement distance) and the social variables related to users' activity centers, we found no significant correlations for users from Categories 1 and 2. Again, the most significant correlations were found for user from Category 3.

Table 2 - Results for Experiment 2 (correlating mobility metrics and social variables related to users' activity centers). Values for users from Category 3.

Mobility / Social Data	Employment rate	Unemployment rate	Persons with no qualifications
Radius of Gyration	(0.25, 0.03)	-	(-0.31, 0.01)
Displacement Distance	(0.36, 0.002)	(-0.32, 0.007)	(-0.28, 0.02)

The results obtained for Experiment 2 are shown in Table 2. These results show some conformance with the first experiment. Here, again, the social variables related to employment conditions presented higher correlation coefficients. For the variable "Displacement Distance" the highest correlation was found with the variable "Employment Rate", with a τ = 0.36. For the "Radius of Gyration", the highest correlation found was with the variable "Persons with no Qualifications" (also related to employment conditions), with τ = -0.31. Given these results, we can answer Question *Q2* by stating that users' mobility patterns may be correlated with social attributes of regions that they visit in an urban center. For example, we found that users that have higher "Displacement Distance" tend to visit regions with higher "Employment rates" (τ = 0.36). Moreover, for these same users, the incidence of activity centers in regions with highest "Unemployment rates" tend to decrease when the "Displacement Distance" increases (τ = -0.32).

It is important to note that even finding possible correlations between the mobility patterns and social data, these correlations were not classified as strong correlations, as the highest value was of τ = -0.38. We believe that the fragmented nature of Twitter messages can add some inaccuracies to the results. For example, poor users might post significantly more than users from rich locations, bringing imprecisions to the results. This kind of problem was partially mitigated by the segmentation of users based on the number of messages they have posted (Categories 1, 2 and 3). For users in Category 3, which provided the best results, we could visually observe that they were homogeneously distributed over the city of London. Furthermore, it is not possible to extrapolate the results obtained from the correlations to the whole population of London, as these results were based on certain Twitter profiles.

6. Conclusions

This research presented a model to allow the identification of correlations between mobility patterns and social, economic and demographic variables. This model identifies mobility patterns from georeferenced Twitter messages, detect users' home locations and activity centers, then looks for correlations with the social data supplied to the model. An experimental evaluation was conducted using data from the city of London. This city was chosen due to the high availability of Twitter messages in the time interval where these messages were collected and also for the availability of many social indicators for this city.

This study confirms that it is possible to identify some correlations between mobility patterns extracted from social media and social indicators. In the results obtained from our experiments, relevant correlations were found for variables associated with employment conditions (economically inactive people, employment rate, unemployment rate and persons with no qualifications).

Additionally, the fragmented nature of Twitter messages makes the task of finding correlations even challenging, forcing us to reduce the number of users to be considered in the experiments (only those with a large number of posts). This indicates the need of performing additional experiments involving more heavy users, to make the correlations more significant. Further work also includes applying this model to the analysis of other regions in the world and comparing the results between them. Moreover, we intend to formulate additional mobility metrics, enhancing the analysis of mobility patterns and the discovery of relevant correlations.

References

Luo, F., Cao, G., Mulligan, K., & Li, X. (2016) "Explore spatiotemporal and demographic characteristics of human mobility via Twitter: A case study of Chicago", *Applied Geography*, *70*, 11-25.

Noulas A, Scellato S, Lambiotte R, Pontil M, Mascolo C (2012) "A tale of many cities: universal patterns in human urban mobility", PLoS ONE;7(5).

Wilson T, Bell M. (2004) "Comparative empirical evaluations of internal migration modelsin subnational population projections", Journal of Population Research;21(2):127-160.

Zheng, V. W., Zheng, Y., Xie, X., & Yang, Q. (2010) "Collaborative location and activity recommendations with gps history data", In Proceedings of the 19th international conference on World wide web (pp. 1029-1038). ACM.

Jurdak, R., Zhao, K., Liu, J., AbouJaoude, M., Cameron, M., & Newth, D. (2015) "Understanding human mobility from Twitter", *PloS one*, *10*(7), e0131469.

Birkin, M., Harland, K., Malleson, N., Cross, P., & Clarke, M. (2014) "An Examination of Personal Mobility Patterns in Space and Time Using Twitter", International Journal of Agricultural and Environmental Information Systems (IJAEIS), 5(3), 55-72.

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

Yuan, Q., Cong, G., Ma, Z., Sun, A., &Thalmann, N. M. (2013) "Who, where, when and what: discover spatio-temporal topics for twitter users", In Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 605-613). ACM.

Hasan, S., Zhan, X., &Ukkusuri, S. V. (2013) "Understanding urban human activity and mobility patterns using large-scale location-based data from online social media", In Proceedings of the 2nd ACM SIGKDD international workshop on urban computing (p. 6). ACM.

Cheng, Z., Caverlee, J., Lee, K., & Sui, D. Z. (2011) "Exploring Millions of Footprints in Location Sharing Services", ICWSM, 2011, 81-88.

Hao Q, Cai R, Wang C, Xiao R, Yang J, Pang Y, Zhang L (2010) “Equip tourists with knowledge mined from travelogues”, In: Proceedings of the 19th international conference on World Wide Web, pp 401–410

Bagrow, J. P., & Lin, Y. R. (2012) “Mesoscopic structure and social aspects of human mobility”, PloS one, 7(5), e37676.

Hsieh H-P, Li C-T, Lin S-D (2012) “Exploiting large-scale check-in data to recommend time-sensitive routes”, In: Proceedings of the ACM SIGKDD international workshop on urban computing. ACM, New York, pp 55–62

Gonzalez, M. C., Hidalgo, C. A., &Barabasi, A. L. (2008) “Understanding individual human mobility patterns”, Nature, 453(7196), 779-782.

Jiang, S., Fiore, G. A., Yang, Y., Ferreira Jr, J., Frazzoli, E., & González, M. C. (2013) “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 (p. 2). ACM.

Palchykov, V., Mitrović, M., Jo, H. H., Saramäki, J., & Pan, R. K. (2014) “Inferring human mobility using communication patterns”, Scientific reports, 4.

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

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

Chaintreau, A., Hui, P., Crowcroft, J., Diot, C., Gass, R., & Scott, J. (2007) “Impact of human mobility on opportunistic forwarding algorithms”, Mobile Computing, IEEE Transactions on, 6(6), 606-620.

Zhang, Y., Wang, L., Zhang, Y. Q., Li, X. (2012) “Towards a temporal network analysis of interactive WiFi users”, EPL (Europhysics Letters), 98(6), 68002.

Ester, M., Kriegel, H.-P., Sander, J., & Xu, X. (1996) “A density-based algorithm for discovering clusters in large spatial databases with noise”, In Proceedings of 2nd international conference on knowledge discovery and data mining (vol. 96).

Huang, Q., Cao, G., & Wang, C. (2014) “From where do tweets originate?: a GIS approach for user location inference”, In Proceedings of the 7th ACM SIGSPATIAL International Workshop on Location-Based Social Networks (pp. 1-8). ACM.

Shin, R. M., Hong, S., Lee, K., & Chong, S. (2008) “On the Levy-walk nature of human mobility: Do humans walk like monkeys?”, In Proc. IEEE INFOCOM (pp. 924-932).

Shelton, T., Poorthuis, A., & Zook, M. (2015). Social media and the city: Rethinking urban socio-spatial inequality using user-generated geographic information. Landscape and Urban Planning, 142, 198-211.