Analysing trajectory data to study individual mobility attitudes in New York City and Tokyo

Computational Models of Human Behaviours

Letizia Monti
University of Trento
student ID: 203960
letizia.monti@studenti.unitn.it

Abstract

Studying human mobility is important to understand individual behaviours, interests and values. The present research analyses Location-Based Social Network data of check-ins in New York City and Tokyo from Foursquare to investigate human mobility from a collective and individual point of view. Collective measures are used to discover the most frequented areas and locations, and their variation across time, while individual metrics such as the number of visits, the waiting times, the radius of gyration and the temporal-uncorrelated entropy are used to discover individual mobility attitudes. Individuals are then clustered according to their behaviours, and their similarities are assessed through the number of common unique visited locations. Then, differences and similarities are drawn from New York City and Tokyo.

1. Introduction

With the occurrence of the Information Age, the term network society has been used to refer to the changes that Internet and communication technologies have brought to society. Everything is interconnected and space assumes a new connotation: "Space does not reflect society, it expresses it, it is a fundamental dimension of society, inseparable of the overall process of social organization and social change" (Castells, 2005). In this new conceptual framework studying mobility becomes directly connected to understanding individual values, interests and behaviours, which in turn are related to issues such as the access to the world of work, social inequalities and crime.

From a more pragmatic point of view, mobility information makes possible to assess the organization of a city, the infrastructures and services provided (e.g. road traffic, public transportation, urban decay, etc). Issues like these can be addressed with the analysis of the distribution of the flows of individuals over space and time. In addition, according to urban sociologists and architects, urban planning and rebuilding should be designed to meet mobility issues. As described by Jacobs in *The Death and Life of Great American Cities* (1961), having areas on a human scale can lead to improvements in the social and economic vitality of a city.

In the last years, the analysis of human mobility has become possible thanks to the massive amount of data coming from mobile GPS sensors that can be made available for research purposes under personal consent by telephone companies or that can be extracted, after being anonymised, from social networks where people choose to share publicly their position.

The current project consists in analysing Location-Based Social Network (LBSN) data of New York City and Tokyo to explore individual mobility attitudes. The aim is to investigate human mobility in the two cities from a collective and individual point of view, and test whether the found differences can be attributed to the diversity of interests and values shared by the citizens of the two cities, or have to be considered as differences *in se* characterising the two cities.

In the following sections I will respectively make a review of the literature on related studies, provide a description of the data and methodology used, explain and report the results of the analysis and in the end, provide an interpretation of the obtained results with respect to the research question.

2. Related work

Given the wide set of theoretical frameworks in which human mobility can be studied, the literature on this topic is quite vast and diversified depending on the discipline, scope and type of analysis. In this section, I will focus on the studies that analyse how people move in ordinary life in a restricted space like a campus, district or city.

An attempt to study mobility at large is the one of Yuan et al. (2012), who provide a technique based on topic modeling for discovering regions of different functions in a city according to human mobility and points of interests (POIs), of a city. Inference is provided considering regions as documents, functions as topics, category of POIs (e.g. train station, restaurant) as metadata and human mobility patterns as words. The novelty of the paper consists in exploiting both POIs and mobility patterns information for the application of topic modeling to a non-textual framework.

Other studies provide a methodology to infer life patterns and individual location preferences from GPS data.
Ye et al. (2009) introduced a model to describe individual life patterns, the notion used to refer to ”individual's general life style and regularity”. Their approach is made of two steps: a modelling phase in which GPS data are processed and transformed into a proper format and the mining phase in which different strategies are applied in order to discover various types of patterns.
Yang et al. (2015) created a spatial temporal activity preference model that models the spatial and temporal preferences of individuals separately and combine the two dimensions in the end for the inference of preference. In this case, *personal functional regions* are derived from the location history of each individual, as people tend to be active in their most frequented areas, while preferences in time are described by temporal activity similarity among different users to overcome the problem of data sparsity.
Studies like that of Xiao et al. (2010) exploit data mining techniques to infer user similarities from location history. In this case GPS trajectories with a semantic location history (i.e. points represented by location categories) are compared and individual similarities are computed with their novel maximal travel match algorithm.

The following researches aimed at statistically modelling human behaviour from GPS data with the use of metrics to describe individual behaviours and attitudes for prediction.
Song et al. (2010), focused on demonstrating that human trajectories follow reproducible scaling laws. Song, Qu et al. (2010) described the limits of the predictability of human behaviours by considering the entropy of individual whereabouts. While Pappalardo et al. (2015) studied the behaviours of two individual profiles, the *explorers* and the *returners*, by analysing the relationship between recurrent and overall mobility.

Given the evidence provided by these last studies, concerning the possibility of some metrics to represent mobility attitudes, the present study will focus on the exploitation of some of them, as it will described in the following sections, for the characterization of the mobility attitudes of the citizens of New York City and Tokyo.

3. Methodology

The analysis is conducted on Location-Based Social Network data collected from Foursquare, a social network that allows users to check-in the visited locations, leaving reviews, discover nearby attractions, earn badges depending on the platform usage and stay connected with friends. In particular, the dataset contains check-in data in New York City and Tokyo that have been collected for 10 months, from April 2012 to February 2013, with the following information:

- User ID (anonymised)
- Latitude
- Longitude
- Date and time (UTC)
- Timezone offset
- Foursquare category (e.g. food, travel and transports, shop and service, etc) [1]
- Foursquare sub-category (e.g. fast food, subway, etc)

As described in Martì et al. (2019), the use of LBSN data for research offers several opportunities and limitations. Some of the advantages are the availability of huge amount of data from different countries and the ease of the data collection, thanks to the possibility of automatically retrieve contents. Data are characterised by accuracy and semantic richness with different extents, depending on the type of social network (e.g. Foursquare provides GPS data corresponding to real venues, enriched by venue names and categories). Representing *Volunteered Geographical Information* (Martì et al., 2019), LBSN data are an unobtrusive research method to study mobility, as individuals are not aware of the fact of being observed when sharing their information. As a consequence, the Hawthorne effect, according to which individuals tend to modify aspects of their behaviours because aware of being observed, typical of controlled experiments, is minimised.

However, LBSN data have some limitations. The most significant one is represented by self-selection bias. Not the entire population is on social networks, there are imbalances concerning socio-economic status, age and other factors that can undermine the representativeness of the data. As compared to social experiments with controlled variables, LBSN data do not contain any personal information about users, without the possibility of directly associating

[1] Retrieved from `https://developer.foursquare.com/docs/build-with-foursquare/categories`.

behaviours to other factors (e.g. gender, age, etc). In addition, individual mobility may be underestimated as movement is detected only when check-ins are made and some time may occur between consecutive visits.

In the present research the concept of mobility is explored both from the collective, via the focus on locations and visits over time, and individual point of view. This will be an attempt to give an overview of human mobility in New York City and Tokyo at the macro and micro level.

4. Analysis

In order to explore mobility attitudes and assess the existence of similarities and differences between New York City and Tokyo, data were analysed in three steps.

The first phase consisted in performing descriptive and exploratory analysis of the flows of individuals in the two cities, by the adoption of some collective measures, such as the number of visits per location, the number of visits per hour and the variation of check-ins over time.

In the second phase, individual behaviours were summarised with mobility metrics and then grouped depending on the extent of similarity of their attitudes by performing k-means clustering, an unsupervised learning technique aimed at dividing observations in groups as much different as possible. Having groups of individuals with the same mobility habits would provide a sort of segmentation of the population according to their behaviours.

In the third phase, users' trajectories, conceived as sequences of visited locations, each of which having a semantic meaning, were used to assess individual similarities inside each cluster.

Considering data of cities of two different continents, having dissimilarities at the societal, economic and political level, will also serve to empirically assess if cultural differences may be reflected in individual behaviours as declared by some theoretical studies as the one cited in the introduction.

4.1. Data preparation

The original data were cleaned and prepared for the analysis after taking some operations. Duplicated rows were deleted, the times at which check-ins occurred were adjusted according to the reported UTC time offset and general categories for each location were manually provided.

Data were preprocessed to be transformed in a proper format and to increase their accuracy. In particular, the datasets relative to New York and Tokyo were converted to trajectory dataframes, the base data structure supported by *scikit-mobility*[2] in which each row represents a trajectory's point. Then, data were filtered, deleting the points with

[2]The Python library for human mobility analysis `https://github.com/scikit-mobility/scikit-mobility`.

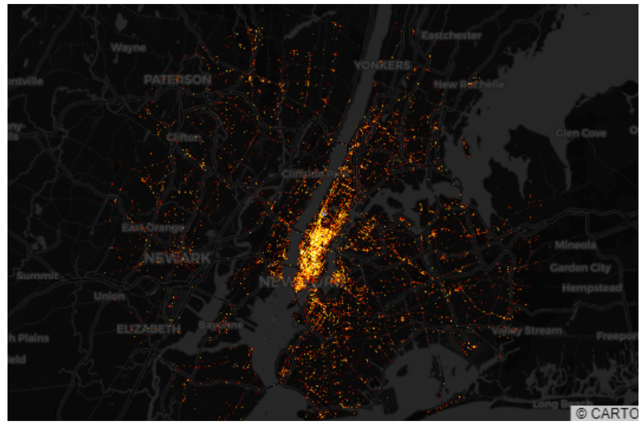

Figure 1. Foursquare check-ins in New York.

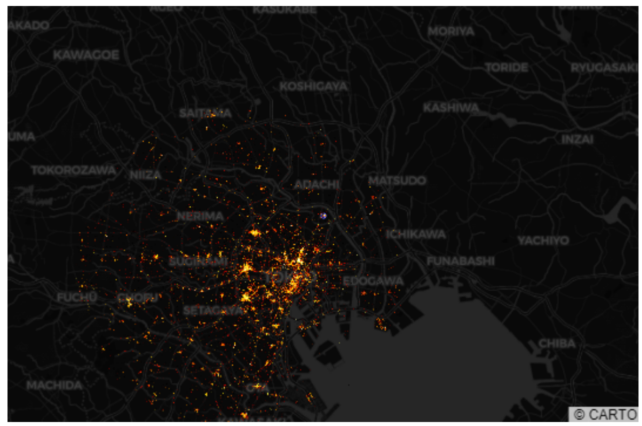

Figure 2. Foursquare check-ins in Tokyo.

speed higher than 300 km/h from the previous point and compressed, merging together all the points that were closer than 100 meters from each other. The resulting dataframes had 189694 trajectories from 1083 users in New York City and 473653 trajectories from 2293 users in Tokyo.

4.2. Human flows mobility

Analysing human mobility from a collective point of view means understanding which are the most visited areas of a city, which are the locations with the most visits (i.e. venue and category) and how check-ins vary across time.

Figure 1 shows the distribution of the check-ins in New York. It is evident that the most frequented borough, where yellow points are more concentrated, is Manhattan, while other key locations are sparse in the city. The most visited areas in Tokyo, as provided by Figure 2, correspond to the centre of the city and to other peripheral boroughs.

The relative distribution of the venue categories for each city is provided by Figure 3. The most frequented locations in New York regard food (e.g. restaurant, bakery), outdoors and recreation (e.g. park, swimming pool) and shop and service (e.g. clothes shop, bank), while travel and transports

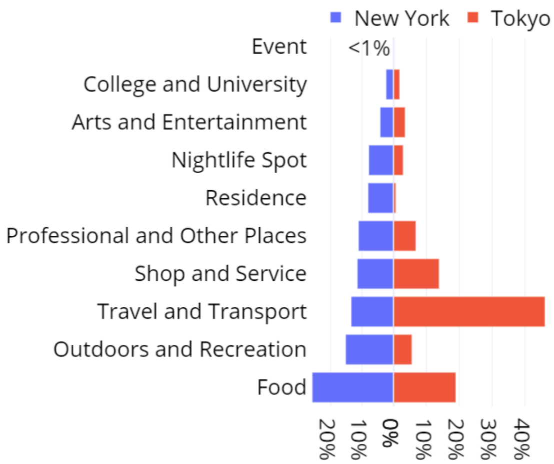

Figure 3. Relative frequency of the venue categories per city.

Figure 4. Word cloud of the venue sub-categories of New York.

Figure 5. Word cloud of the venue sub-categories of Tokyo.

(more than 40%), food and shop and service in Tokyo.

More specifically, Figure 4 and 5 provide the word clouds of the venue sub-categories of respectively New York and Tokyo. Here, the words with the highest size correspond to the most visited locations. For New York some examples are shop, restaurant, home, bar and office, that are locations whose general categories have low frequencies (e.g. nightlife spot and residence). For Tokyo instead, the most frequented venues, such as train stations, subways, restaurants and stores correspond to the most frequented location categories.

For what concerns the variation of check-ins over time, as showed in Figure 6, we observe both for New York and Tokyo the presence of many small peaks over the months.

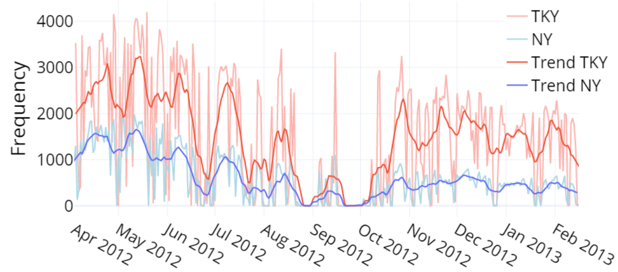

Figure 6. Number of check-ins per day and respective trends.

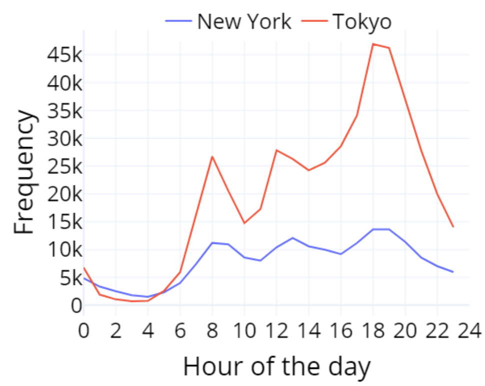

Figure 7. Number of check-ins per hour of the day.

The trends were computed on a monthly basis through an additive decomposition of the original data to delete the seasonal and random components and simplify the comparison between the two cities. It can be observed that throughout the period of investigation the number of check-ins in Tokyo is always higher than that in New York, except for the end of August and September where there is a significant drop in the number of visits with values very close to zero. Apart from this, the rises and falls in the trends of the two cities are quite similar, in some cases smoother for New York City.

Visits are characterised by some sort of variation over the day, as Figure 7 shows. Both for New York and Tokyo there are three peaks in the number of visits at around 8, 12/13 and 18. While this variation is quite smooth in New York, in Tokyo mobility is characterised by significant changes over the day, especially in the late afternoon, when the highest number of visits is registered.

During the night (from 20 to 6 in the morning) the lowest number of check-ins is registered, equal to 20% and 16% of the total number of check-ins in respectively New York City and Tokyo. Having a close look to the visits in this part of the day, we expect some differences in the type of locations visited with respect to the general evidence provided above.

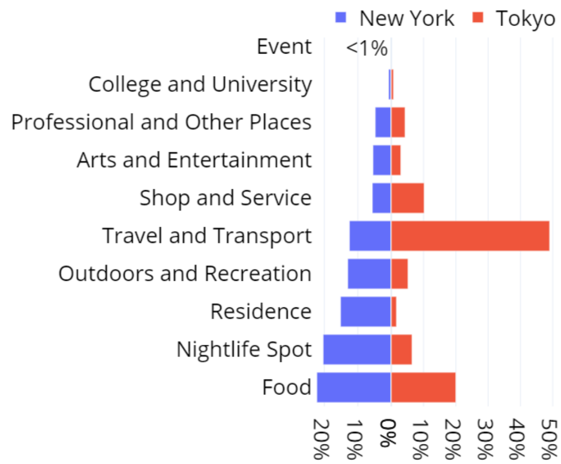

Figure 8. Relative frequency of the venue categories per city by night.

As we can see from Figure 8, the categories of the most visited locations by night are food, nightlife spot and residence for New York, while remain travel and transport, food and shop and service for Tokyo, even if the frequency of other minor categories such as residence and nightlife spot have changed.

4.3. Individual mobility

Individual mobility attitudes are detected with the use of different metrics, such as the number of visits, the waiting times, the radius of gyration, the k-radius of gyration and the temporal uncorrelated entropy, summarising different aspects of mobility.

The number of check-ins per individual is a basic metric that can be used to estimate the extent to which people move around a city. Figure 9 provides the histograms of the number of check-ins in New York City and Tokyo. In both cases this metric is flat-distributed and around 150 visits were registered by the large part of the investigated population.

An other proxy of individual mobility that covers its temporal dimension is the waiting times between the visit of two successive locations. It is a complementary metric of the number of check-ins that temporally quantifies the activity of individuals. For each individual the average of all her waiting times is computed to make possible comparisons. As showed in Figure 10, both the cities present a similar distribution and the majority of the individuals have waiting times ranging from 0 to 3 days.

The radius of gyration measures the characteristic distance travelled by an individual. It is computed with the

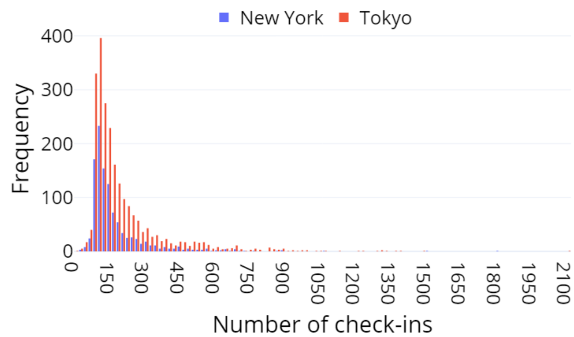

Figure 9. Number of check-ins per individual.

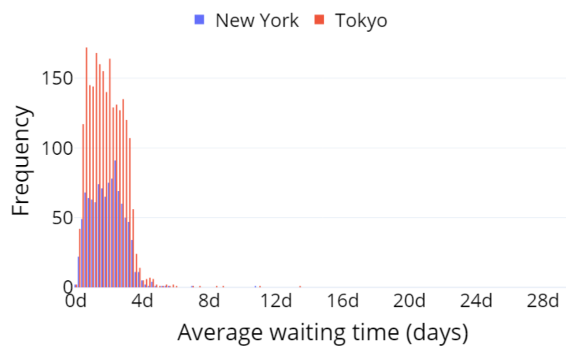

Figure 10. Individual average waiting time.

following formula:

$$r_g = \sqrt{\frac{1}{N}\sum_{i=1}^{N} n_i (r_i - r_{cm})^2}$$

where N in the number of visited locations, n_i is the number of visits to a location i, r_i is a two-dimensional vector of the geographical coordinates of location i and r_{cm} is the position vector of the centre of mass of the set of locations visited by an individual. As presented in Figure 11, the citizens of New York are characterised by an average travelled distance of 5 km, while those of Tokyo by one of 6 km. Generally, individuals in Tokyo have a higher radius of gyration than those in New York.

To understand how the k-th most frequented locations of an individual determine his characteristic travelled distance (Pappalardo et al., 2015), the k-radius of gyration defined as follows is introduced:

$$r_g^{(k)} = \sqrt{\frac{1}{N_k}\sum_{i=1}^{k} (r_i - r_{cm}^{(k)})^2}$$

where $N_k = \sum_{j=1}^{k} n^{(j)}$ is the sum of the total number of

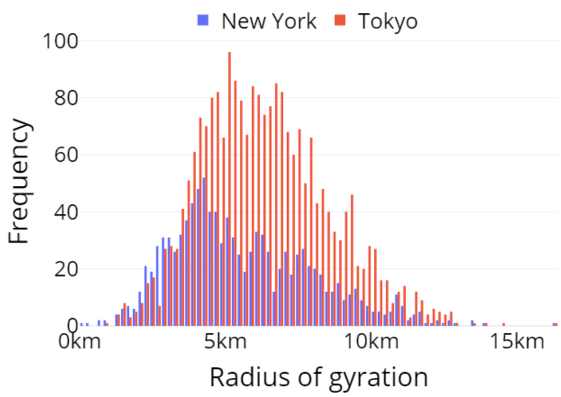

Figure 11. Individual radius of gyration.

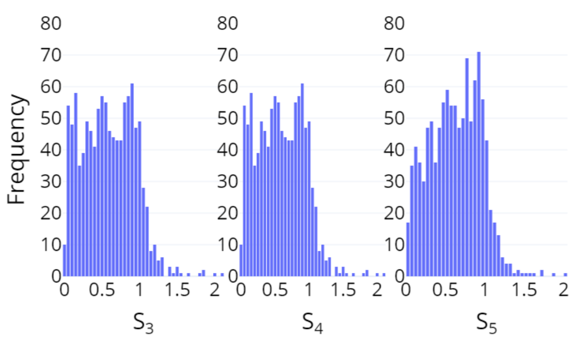

Figure 12. The S_k ratio in New York.

visits of the k-th most frequented locations and $r_{cm}^{(k)}$ is the centre of mass computed on the k-th most frequented locations.

This metric measures the radius of gyration between an individual k most frequented locations. Its relationship with the radius of gyration can be quantified by the ratio: $S_k = r_g^{(k)}/r_g$. Values of this ratio close to 0 denote the *explorers* (Pappalardo et al., 2015), the individuals whose recurrent travelled distance is significantly small with respect to the total radius of gyration, while values close to 1 represent the *returners* (Pappalardo et al., 2015), for whom the overall and recurrent mobility almost coincides.

Figure 12 and 13 respectively represent S_k for different numbers of most frequented locations (k) for New York and Tokyo. It is natural to observe that increasing k, the number of recurrent locations to consider, the proportion of explorers decreases in favour of returners. In the two cities, returners and explorers represent the extremes of a continuum along which most of the individuals positioned. A sort of balance is reached when the ratio is computed from $k = 4$ recurrent locations.

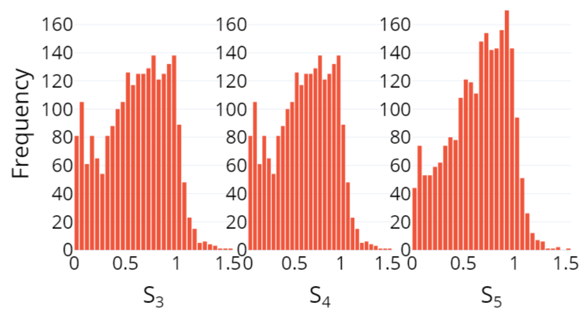

Figure 13. The S_k ratio in Tokyo.

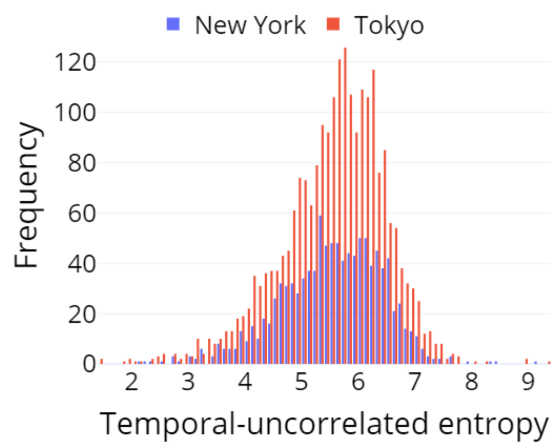

Figure 14. Individual temporal-uncorrelated entropy.

To assess the degree of uncertainty and predictability of individual mobility, the temporal-uncorrelated entropy as defined in Song et al. (2010) is used:

$$S_i^{unc} = -\sum_{j=1}^{N_i} p_i(j) log_2 p_i(j)$$

where $p_i(j)$ is the historical probability that location j was visited by individual i. Considering the visitation frequency of locations should improve the predictability of future locations. In general, the higher the value of the entropy, the higher the uncertainty regarding individual's future locations. As represented in Figure 14, the temporal-uncorrelated entropies of New York and Tokyo have overall a similar distribution with an average value of around 5.

Clustering Once individual metrics were computed, k-means clustering was used in order to group individuals depending on their mobility attitudes. This is un unsuper-

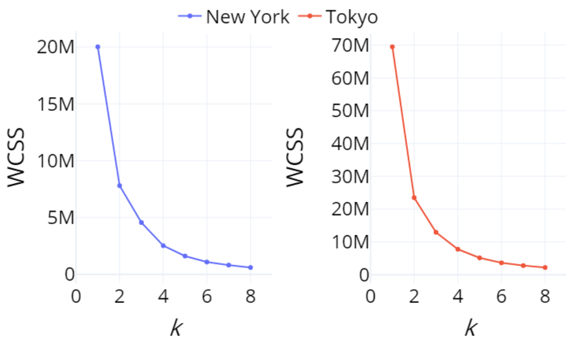

Figure 15. Elbow method for the choice of the number of clusters.

vised learning technique that aims at partitioning observations into k clusters in which similarity between observations is maximised. Each cluster is characterised by a mean vector, containing the mean values of the features investigated, relative to the observations contained in that cluster.

The features selected for clustering are the individual mobility metrics investigated above: the number of check-ins, the average waiting time, the radius of gyrations, the ratio between the radius of gyration and the 4-radius of gyration (S_4) and the temporal-uncorrelated entropy. They represent different dimensions of the concept of mobility and should be as much uncorrelated as possible to avoid the overestimation of some features during clustering. The lack of strong correlation, as can be noticed from the matrices of correlation in Table A1 and A2 of the Appendix, can also guarantee that different aspects of mobility are covered.

The number of cluster k to use is a hyperparameter that is chosen with the *elbow* method, a subjective method that consists in computing k-means clustering for different k, and for each of them estimating the within-cluster sum of squares, an error measure that quantifies the variability of the observations in each cluster. Then, the optimal k should be the one corresponding to a significant reduction of this error. Applying the elbow method to the feature vector chosen for clustering for New York City and Tokyo led to the within-cluster sum of squares plots in Figure 15. In both cases the elbow in the plots is found in case observations are grouped in three clusters.

Table 1 and 2 provide the feature mean values for each cluster for respectively New York and Tokyo. Individuals of both cities seem to be mainly grouped together according to the number of check-ins registered and the average waiting time between subsequent locations. As could be expected, the number of check-ins is inversely associated with the average waiting time. For what concerns the other features, the radius of gyration, the S_4 and the temporal-uncorrelated entropy, only a slightly variation is detected between different clusters. As a consequence, as can be seen from Figure

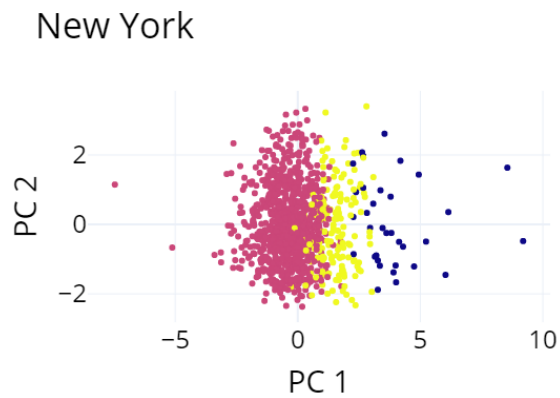

Figure 16. Clusters visualization of New York.

16 and 17[3], the clusters are not well divided and tend to overlap both for individuals in New York and in Tokyo. In addition, the sizes of the clusters are quite inhomogeneous, being the majority of the individuals part of the cluster associated with the lowest number of check-ins and the highest waiting time.

4.4. Individual mobility similarities

After clustering individuals according to their mobility attitudes and habits, an attempt to assess individual similarities based on their visits was made.

The analysis is made from users' trajectories, i.e. sequences of stay points, each of which having a semantic meaning (school, subway, bar, etc), based on their unique visited locations. Individuals belonging to the same cluster were pairwise compared in order to get the number of common visited locations. This allowed to assess if individuals with similar mobility attitudes had visited the same type of locations as well.

[3]The clusters are visualized on the first two principal components after performing Principal Component Analysis, a procedure for reducing the dimensionality of the input space by representing it with a few variables (i.e. components) that explain most of the variability of the data.

Table 1. Features mean values per cluster of New York.

	N check-ins	Waiting time	Radius of gyration	S_4	Temp-unc entropy	N individuals
0	127	2.15	5.49	0.59	5.45	883
1	319	0.89	6.04	0.55	5.85	166
2	717	0.45	6.53	0.58	6.34	34

Table 2. Features mean values per cluster of Tokyo.

	N check-ins	Waiting time	Radius of gyration	S_4	Temp-unc entropy	N individuals
0	142	2.12	6.34	0.60	5.47	1875
1	408	0.75	6.33	0.64	5.90	341
2	880	0.36	6.60	0.65	6.20	77

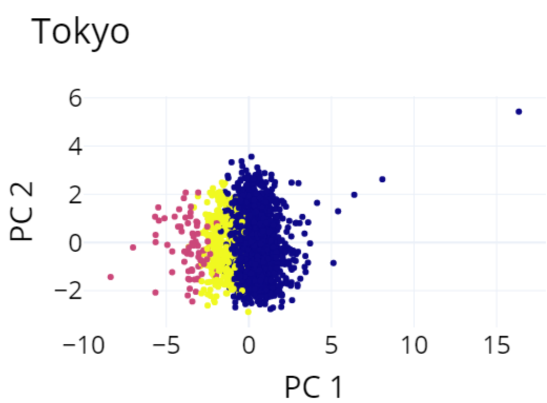

Figure 17. Clusters visualization of Tokyo.

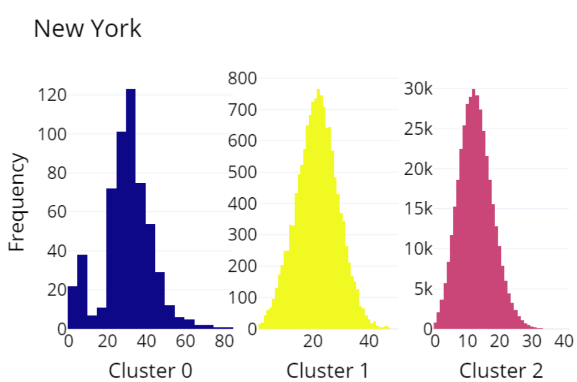

Figure 18. Number of common unique visited locations per cluster in New York.

As can be seen from Figure 18 and 19, the number of common visited locations slightly varies across different clusters. Evidences from New York City and Tokyo are very similar and show that, while individuals with the highest number of visits (cluster 0) are characterised by an average of around 30 common visited locations between each other, those with a lower number of check-ins (cluster 1 and 2) have around 20 and 10 common visited locations respectively.

The reason of this variation across clusters can be explained by the fact that having a higher number of common locations is much probable among individuals with rich trajectories, more inclined to travel around the city. On the other hand it seems reasonable that individuals with a sporadic number of check-ins are related to each other by a small number of common locations.

As a naive attempt to understand individual mobility similarities from trajectory data, more sophisticated techniques and algorithms should be adopted to improve the obtained results.

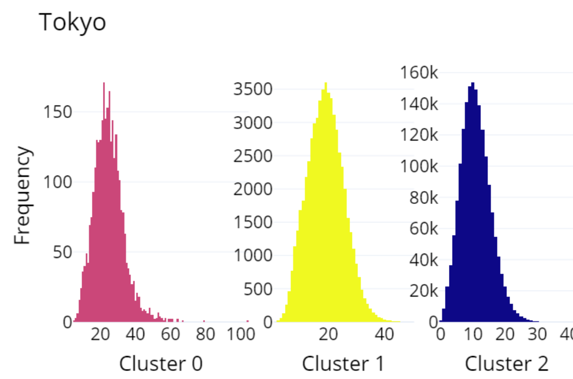

Figure 19. Number of common unique visited locations per cluster in Tokyo.

5. Discussion

The analysis of mobility in New York City and Tokyo from a collective point of view brought evidence on which were the most visited locations and areas, the most visited venue categories and their variation across time. Similarly, considering some individual mobility metrics allowed to discover the extent to which individuals move around and how often, their typical travelled distance, the relationship between recurrent and overall mobility and the degree of predictability of their whereabouts. Using these measures as a proxy of individual mobility attitudes, individuals were clustered accordingly. Then, individual similarities were assessed by detecting the number of common locations in each cluster from user's trajectories.

It is not clear whether the found differences between the two cities can be attributed to differences in the mobility attitudes and so in individual values, interests and behaviours, or rather depend on the usage of Foursquare, the source of the analysed data. Having no personal information about users makes difficult to draw conclusions from the discrepancies between the two cities. In this sense, the fact that for example more than 40% of Tokyo citizens registered at a travel and transport venue, while less than 10% of New Yorkers did the same, can possibly be just due to the preferences of the investigated population.

However, the fact that the clusters of New York city and of Tokyo, derived from the grouping of individuals from their mobility attitudes, are almost the same shows that the large part of the individuals that use Foursquare have the same characteristics.

6. Conclusion

Location-Based Social Network data of users' check-ins in New York City and Tokyo, collected for 10 months from Foursquare, were analysed in order to explore individual mobility attitudes and assess similarities and differences between the two cities both at a collective and individual level.

From an aggregated point of view, visits of New York City and Tokyo varied in space and time. Both the two cities are characterised by few locations and areas that were most visited throughout the months: airports and train/bus stations, Manhattan for New York City and few isolated suburbs for Tokyo. The difference found between the two cities in the categories of locations most visited can be motivated by the existence of actual differences in the mobility habits or, more reasonably, by a difference in the behaviours of the individuals on Foursquare and in the tendency of registering at different types of venues. For example, citizens of Tokyo were found to most frequently check-in at travel and transport, food and shop and service venues, while New Yorkers at food, outdoors and recreation and shop and service locations. This was a tendency that changed in the nightly hours, especially in New York, where nightlife spots and residence venues were the second and the third most visited location categories.

Check-ins varied over time as well, following a trend distribution with many oscillations during the year of investigation. The most anomalous finding was the drastic drop in the number of check-ins in New York City and Tokyo in August and September, probably associated to the summer holidays. In addition, for what concerns the daily variation, visits were found to peak at around 8, 12/13 and 18 of the day, more deeply in Tokyo than in New York, characterised by a smoother change in the number of visits during the day.

Individual mobility attitudes were analysed by estimating the number of visits, the average waiting times, the radius of gyration, the k-radius of gyration and the temporal-uncorrelated entropy per individual. The use of these metrics allowed to understand the extent to which and how often individuals move around the city, the overall and recurrent mobility, whether they were more *returners* or *explorers* and the predictability of their future locations. The results obtained from the distribution of these metrics were overall the same in New York and Tokyo, with the exception of the number of visits and the typical travelled distance, that were slightly higher in Tokyo.

Performing k-means clustering according to the individual mobility attitudes, represented by the computed metrics, led to the creation of three clusters. The investigated individuals were found to be segmented according to their mobility attitudes similarly in New York City and in Tokyo. The large part of them were found in the cluster with the lowest number of visits and the largest average waiting times. However, the clusters were highly inhomogeneous and quite overlapping, not ensuring a perfect division in subgroups.

In the end, the computation of the number of common unique visited locations among individuals of the same cluster provided a measure of similarity of those who share the same mobility behaviours. As could be expected, results were similar for both cities and showed that the most similar individuals were those who travelled the most and had the lowest waiting times.

Because of the limitations of the data considered, it seems problematic to generalise the results to the entire population of New York City and Tokyo. In addition, it is not certain whether the found differences between the two cities could be attributed to differences at a societal level. Surely the use of LBSN data represents a valuable framework for testing mobility models, algorithms and general techniques for the investigation of mobility. For what concerns studying citizens mobility and social dynamics, data with richer information about individuals could offer better results. To bring improvements, developments of the present research should go in this direction.

References

[1] Castells, M. (2010). *Space of Flows and Space of Places* in The Rise of the Network Society, M. Castells (Ed.).

[2] Jacobs, J. (1961). *The death and life of great American cities*. New York: Random House.

[3] Martí, P., Serrano-Estrada, L., Nolasco-Cirugeda, A. (2019). *Social Media data: Challenges, opportunities and limitations in urban studies*. Computers, Environment and Urban Systems. 74. 161-174.

[4] Pappalardo, L., Simini, F., Rinzivillo, S., Pedreschi, D., Giannotti, F., Barabasi, A. (2015). *Returners and explorers dichotomy in human mobility*. Nature Communications. 6.

[5] Song, C., Koren, T., Wang, P., Barabasi, A. (2010). *Modelling the scaling properties of human mobility*. Nature Physics. 6(10).

[6] Song, C., Qu, Z., Blumm, N., Barabasi, A. (2010). *Limits of Predictability in Human Mobility*. Science (New York, N.Y.). 327. 1018-21.

[7] Xiao, X., Zheng, Y., Luo, Q., Xie, X. (2010). *Finding similar users using category-based location history*. GIS: Proceedings of the ACM International Symposium on Advances in Geographic Information Systems. 442-445. 10.

[8] Yang, D., Zhang, D., Zheng, V., Yu, Z. (2015). *Modeling User Activity Preference by Leveraging User Spatial Temporal Characteristics in LBSNs*. Systems, Man, and Cybernetics: Systems, IEEE Transactions on. 45. 129-142.

[9] Ye, Y., Zheng, Y., Chen, Y., Feng, J., Xie, X. (2009). *Mining Individual Life Pattern Based on Location History*. Proceedings - IEEE International Conference on Mobile Data Management.

[10] Yuan, J., Zheng, Y., Xie, X. (2012). *Discovering regions of different functions in a city using human mobility and POIs*. Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.

A. Appendix

The correlation matrices relative to the features used for clustering of New York city and Tokyo are reported below.

Table A1. Correlation matrix of the features relative to New York.

	N check-ins	**Waiting time**	**Radius of gyration**	**S_4**	**Temp-unc entropy**
N check-ins	1.000	-0.549	0.091	-0.009	0.277
Waiting time	-0.549	1.000	-0.037	0.064	-0.185
Radius of gyration	0.091	-0.037	1.000	0.250	0.049
S_4	-0.009	0.064	0.250	1.000	-0.022
Temp-unc entropy	0.277	-0.185	0.049	-0.022	1.000

Table A2. Correlation matrix of the features relative to Tokyo.

	N check-ins	**Waiting time**	**Radius of gyration**	**S_4**	**Temp-unc entropy**
N check-ins	1.000	-0.535	0.012	0.035	0.254
Waiting time	-0.535	1.000	0.029	-0.009	-0.225
Radius of gyration	0.012	0.029	1.000	0.346	0.138
S_4	0.035	-0.009	0.346	1.000	-0.068
Temp-unc entropy	0.254	-0.225	0.138	-0.068	1.000