Modeling the Dynamics of City-Centers

Aparimit Kasliwal, Arun Sundaresan

9^{th} December, 2023

1 Introduction

The expansion of cities can lead to one of two distinct configurations: monocentric or polycentric, each of which represents a mutually exclusive urban state that any city may assume. Several works as [1] and [6] have classified the development of these different states as occurring because of purely attractivity-driven factors, or purely distance-driven factors.

The longstanding Central Place Theory (CPT) proposed [3] aimed to characterize cities by emphasizing the arrangement of central places and the interaction between them, facilitating the classification of cities into either mono centric (single dominant center) or poly centric (multiple centers) structures based on the distribution of services and functions across the urban landscape. According to this theory, larger, higher-order central places offer specialized and more diverse services compared to smaller, lower-order ones.

Our findings are coherent with the CPT, and we present a mathod to highlight the stark differences between a monocentric and a polycentric setting in this respect. With Boston and Los Angeles as hallmark examples of a monocentric and a polycentric city respectively, we bring out the differences in the distribution of their center sizes as a way to distinguish between these types of city formations.

In our work, we aim to characterize the distance between the different centers within a city. Our method encapsulates both, the attractivity and the distance-driven regimes directly through the data that we consider pertaining to both of these regimes. We define attractivity to be an entropy based measure of the cardinality of facilities in a geographical location, and we consider the TimeGeo [5] generated trajectories for millions of synthetic agents in Boston and Los Angeles for a measure of the distances that individuals travel in both of these cities.

We present the specific characteristics of the data that we use for this study in the next section, along with the modifications we introduced to particular methods for clustering and for calculating the radii of gyration of each user. Finally, we present the results based on the methods we used, and tie the findings back to the theoretical foundation of the previous work(s).

2 Data and Methods

2.1 Data

For our study, we build our story using two different datasets, described as follows;

2.1.1 Facility Data

Upon extraction of different types of facilities and their locations using OSMNx [2] from the widely popular Open Street Map (OSM) library, we are able to obtain the spatial distribution of these facilities for both, Boston and Los Angeles. As we describe in further detail in the next section(s), we approximate the attractivity of a center by evaluating the number of facilities located within the bounds of that center. This simple approximation aligns with the notion of attractivity and also leverages the data to its best.

2.1.2 TimeGeo Trajectory Data

The movement of multiple different individuals (or agents) is captured by the TimeGeo data that synthetically generates trajectories based on appropriate previous modelling findings. For both Boston and Los Angeles, this data provides us with latitude, longitude, type of activity, unique agent ID, and time stamps. We leverage this rich data by calculating a slightly modified version [8] of the longstanding metric of radius of gyration [4], calculation and implementation of which is described further below.

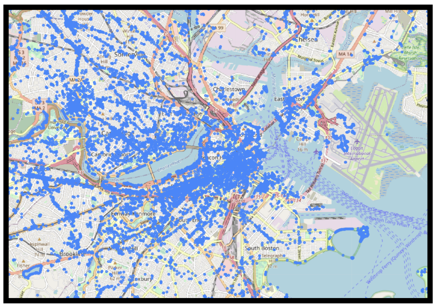

Figure 1: Facilities within Boston

	agentNumber	boston_agents_rog	lat	lon
0	1.0	20.119281	42.162677	-71.144708
1	2.0	21.462174	42.162677	-71.144708
2	3.0	20.119281	42.162677	-71.144708
3	4.0	20.119281	42.162677	-71.144708
4	5.0	17.561915	42.162677	-71.144708
...	...	...	...	...
2771378	3506078.0	13.743821	42.252704	-71.075862
2771379	3506079.0	16.983577	42.252704	-71.075862
2771380	3506080.0	47.466896	42.252704	-71.075862
2771381	3506081.0	14.703995	42.252704	-71.075862
2771382	3506082.0	18.620884	42.252704	-71.075862

2771383 rows × 4 columns

Figure 2: Calculated R_g for each agent in Boston with their Home Locations

2.2 Methods

We draw on the following methods for performing the analysis for our study;

2.2.1 Modified Radius of Gyration

In the domain of human mobility analysis, the standard formulation [4] calculates the radius of gyration in the following way;

$$r_g = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(r_i - \langle r \rangle)^2} \tag{1}$$

where:

$$\begin{aligned} r_g &: \text{Radius of gyration} \\ n &: \text{Number of visited locations} \\ r_i &: \text{Position of the individual at location i} \\ \langle r \rangle &: \text{Average position of the individual} \end{aligned}$$

Note that here, $\langle r \rangle$ is the calculated "center of mass" of the trajectory if the individual. However, in our study for the calculation of the center locations in a city, we require the calculation of radius of gyration from the **home location** of each individual. Thus, the modified equation for that we use for calculating the radius of gyration is as follows;

$$rg_{modified} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(r_i - r_h)^2} \tag{2}$$

where:

$$\begin{aligned} rg_{modified} &: \text{Modified form of Radius of gyration} \\ r_h &: \text{Home location of the individual} \end{aligned}$$

2.2.2 Clustering

First, we formulate the attractivity regime by forming clusters using solely the facility data. This is done using the KMeans algorithm. Our initial foray included experiments with Agglomerative and Divisive Clustering Methods of which the latter performed appreciably, we chose KMeans to move forward because of its simplistic approach of assigning individual points to the closest center - which comes in handy later when we combine this regime with the distance-driven radius of gyration data.

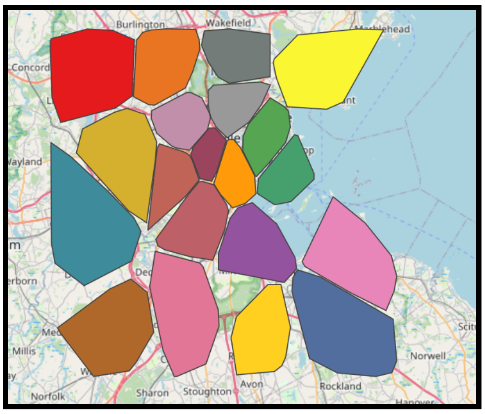

Figure 3: Tessellation of facility-clusters in Boston

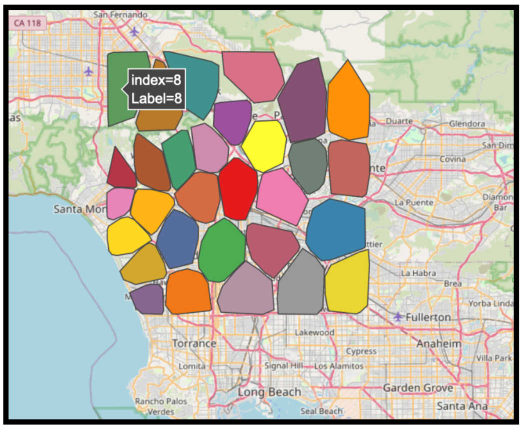

Figure 4: Tessellation of facility-clusters in LA

2.2.3 Attractivity-Regime; Evaluating K using Entropy Characterisation

In order to choose which K is the best for KMeans clustering, we implement a slightly modified version of the elbow method as follows. After running the algorithm for multiple values of K (from $1 \rightarrow 40$), we choose that value of K which maximizes the **variance of entropies** over the clusters given as an output by the KMeans algorithm. This captures the observed result that centers within cities tend to have most, if not all, kinds of facilities including shopping malls, schools, hospitals, grocery stores, etc - as is also suggested by the Central Place Theory [3].

$$H = -\sum_{i=1}^{n} p_i \log_2(p_i) \tag{3}$$

where:

$$\begin{aligned} H &: \text{Shannon entropy} \\ n &: \text{Number of facility types} \\ p_i &: \text{Probability of occurrence of the i-th facility type} \end{aligned}$$

Simply put, we leverage the elbow method but define a new metric (Shannon Entropy, as above) against which we compare the feasibility of different k values in the KMeans algorithm.

2.2.4 Distance-Regime; Radius of Gyration for Evaluating K

Below is the algorithm we use to evaluate the best K for KMeans using a distance-driven approach. Under ideal conditions, where a particular K gives the exact characterization of centers with the city, the radius of gyration for each of individual within a cluster must be equal to the distance from the cluster-center. Here, we measure the deviation from this ideal situation and use it to compare the different values of K.

Algorithm 1 Evaluating K for K-Means in the Distance-Driven Regime

Require: $N \geq 0$
  for `k in` $\{$`1,2,.....,N`$\}$ **do**
    `get clusters via K-Means`
    for `each cluster` **do**
      deviations ←(rog - distance_to_center) ▷ Deviation for each agent
      `calculate the variance of the deviations` ▷ Spread of these deviations
    end for
  end for
  `Choose K that minimizes the mean of the variances of these deviations`

3 Results

3.1 Shannon Entropy: Maximizing Diversity in Clusters

Within the attractivity regime, and in coherence with the propositions of the Central Place Theory [3], higher order centers tend to have specialized functions and thus possess higher diversity in terms of their facilities. We utilize this theoretical argument by formulating entropy of the clusters as described in the methods section.

The entropy of the clusters, however, keeps on increasing as we increase the K value for the KMeans algorithm, as we should expect. We thus implement the underlying essence of the "Elbow Method" to come up with a suitable choice for K.

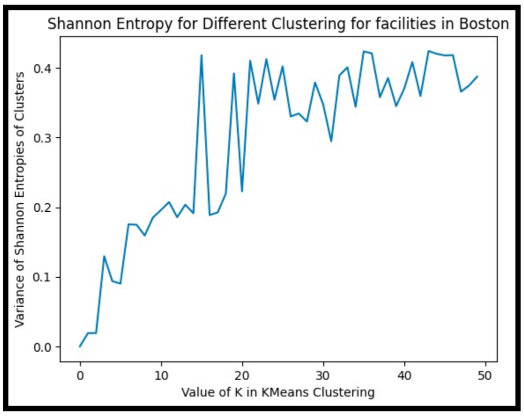

Figure 5: Shannon Entropy for Facility Clusters in Boston

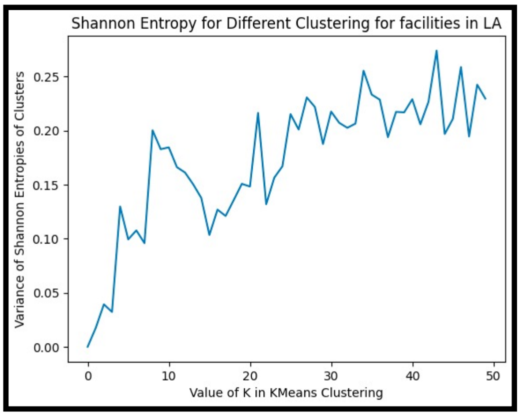

Figure 6: Shannon Entropy for Facility Clusters in Los Angeles

We make the following key observations from the figures above, which plot the Shannon Entropy for the clustering constructed for numerous facilities in the cities of Boston and Los Angeles;

- The Shannon entropy of these clusters is a monotonically increasing function of the value K in the KMeans algorithm. This means that the clusters keep on getting diversified as the total number of clusters increases, which is what we theoretically expect as well.
- However, note that this increment plateaus, as argued by the Elbow Method. Interestingly, for Boston, the entropy is maximized for around 20 clusters while the number is around 30, and thus higher, for Los Angeles.

3.2 Minimizing Deviation between ROG and Distance To Center Within Clusters

Within the distance-driven regime, each person is bound to travel to the nearest center. Thus, in an idealized distance-driven regime as characterized by our clustering, the radius of gyration of each person should be exactly equal to the distance of the center closest to them.

With the aim of minimizing these deviations to get closer to the best possible clustering, we calculate this deviation between the radius of gyration of each agent and distance to their nearest center. Leveraging the elbow method, we choose a reasonable K for both Boston and Los Angeles, as the following figures and their interpretation suggests.

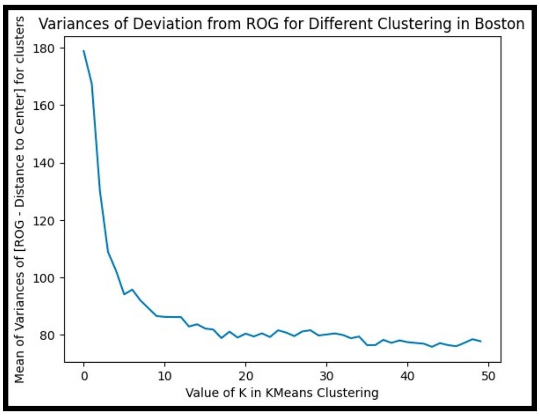

Figure 7: Shannon Entropy for Facility Clusters in Boston

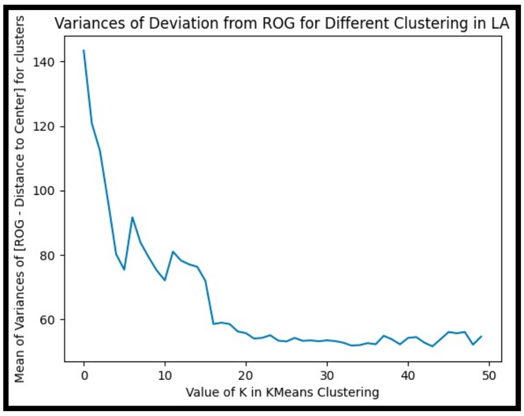

Figure 8: Shannon Entropy for Facility Clusters in Los Angeles

We make the following key observations from the figures above, which plot the deviation between the radius of gyration for each agent and the distance to the closest center, for the clusters formed for different values of K for both Boston and Los Angeles;

- This deviation is a monotonically decreasing function of the number of clusters, as we would expect from theoretical standpoint. As we increase K, the number of cluster increases and the granularity of the problem also increases - thereby rendering the radius of gyration of each agent almost the same as the distance to their closest centers.
- However, note that this decrement plateaus at some value of K, as argued by the Elbow Method. Interestingly, for Boston, this deviation is minimized at around 20 clusters, while the number is roughly around 30 for Los Angeles, which is higher than Boston.

Bringing these results together, we note that the ideal number of clusters for Boston is lesser than Los Angeles. Furthermore, the higher order centers differ more in size, function, and diversity form lower order centers for Boston, thereby arguing for its monocentric state as opposed to Los Angeles for which the size, function and diversity of centers remains more or less constant, thereby making the case for its polycentric state.

3.3 Comparative Analysis: Monocentricity Vs Polycentricity

We present three distinct formulations in terms of the parameters pertaining to our data, that distinctly draw out the differences between a monocentric and a polycentric city. The broader idea, on which our following three formulations are based, is the following.

Central Place Theory [3] suggests that there are multiple centers within any city, all of which can be distinguished based on their "order": a few centers are higher-ordered, a few more centers are lower-ordered, even more are further lower-ordered, and so on. In its essence, the higher ordered centers provide more specific, diverse set of services and functionalities.

Most importantly, cities of both types follow the same characterization - all of them have different order centers. However, the higher order centers in a monocentric setting are much more dominant compared to the other centers. On the contrary, higher order centers aren't any more dominant than the other centers in a polycentric setting, thus maintaining a uniformity in the order of centers for a polycentric setting.

We define a center to be higher-order, based on the following three formulations, all of which lead to the same conclusion.

3.3.1 Cardinality of Facilities

Relative cardinality of the facilities for each cluster is representative of whether a center is higher-ordered or otherwise. In order to compare the relative dominance of higher-order clusters in a monocentric city Vs a polycentric city, we capture the following numeric results;

$$Parameter_1 = \frac{\sum_{i=1}^{5} \text{Facilities in Top Cluster}_i}{\sum_{i=n-4}^{n} \text{Facilities in Bottom Cluster}_i} \tag{4}$$

Parameter_1 for **Boston**	8.75
Parameter_1 for **LA**	4.06

3.3.2 Area

Relative area of the facilities for each cluster is also representative of whether a center is higher-ordered, or otherwise. We compare the variance of the area for the clustering within Boston and Los Angeles - where a high variance shows that there's a much higher spread of the numerical area of clusters present, and a low variance is equivalent to clusters that are more uniform across different orders, similar to the case in a polycentric setting. We then compare these numerical values for monocentric and polycentric cities. We obtain cluster areas by computing the area of the convex hull that captures all points in that cluster.

$$Parameter_2 = \text{Var}(A) = \frac{1}{n}\sum_{i=1}^{n}(A_i - \overline{A})^2 \tag{5}$$

where:

$$A_i : \text{Area of the convex hull for each cluster}$$
$$n : \text{Number clusters}$$
$$\overline{A} : \text{Mean Area of the Convex Hulls}$$

Variance of Cluster Area for **Boston**	0.0058
Variance of Cluster Area for for **LA**	0.0018

3.3.3 Density of Facilities

Finally, a higher-order center will also have a higher Density of Facilties, which is why Density of Facilities is also representative of the ordering of centers. We compare the densities of the facilities for higher order and lower order facilities, and note the following key numerical characteristic;

$$Parameter_3 = \text{Var}(D) = \frac{1}{n}\sum_{i=1}^{n}(D_i - \overline{D})^2 \tag{6}$$

where:

$$D_i : \text{Density of the convex hull for each cluster}$$
$$n : \text{Number clusters}$$
$$\overline{D} : \text{Mean Density of the Convex Hulls}$$

Variance of Density of Facilities for **Boston**	23.65 Million
Variance of Density of Facilities for **LA**	1.32 Million

Overall, we note that the as each of the metrics above suggests, higher order centers are much more dominant in a monocentric setting as opposed to a polycentric setting. This is particularly shown to be the case for Los Angeles & Boston, which are the key examples for a polycentric and a monocentric city within USA.

Future Work and Conclusion

In this study, we bring together the facility location data and the synthetic individual mobility data generated from the TimeGeo model [5] to characterise the differences in monocentric and polycentric settings within modern cities. We note that the differences in the size of centers for monocentric cities is much larger than that for polycentric cities, where the clusters / city-centers are more or less the same size. This is also seconded by the weighted network analysis that we present towards the end of our work.

While this study focuses only on Los Angeles and Boston primarily because they are the most studied and well-known examples of a polycentric and a monocentric city in the United States, this characterisation can be extended to numerous other cities for which we have appropriate data. Furthermore, other recently proved empirical results as those in [7] can be studied further in the light of whether a city is monocentric or polycentric. Fitting mobility models and formulating the differences in individual mobility trajectories for different monocentric and polycentric cities can further inform the importance of this characterisation.

References

[1] Marc Barthélemy. Spatial networks. *Physics Reports*, 499(1):1–101, 2011.

[2] Geoff Boeing. Osmnx: New methods for acquiring, constructing, analyzing, and visualizing complex street networks. *Computers, Environment and Urban Systems*, 65:126–139, 2017.

[3] Arthur Getis and Judith Getis. Christaller's central place theory. *Journal of Geography*, 2007.

[4] Marta C. González, César A. Hidalgo, and Albert-László Barabási. Understanding individual human mobility patter. *Nature*, 2008.

[5] Shan Jiang, Yingxiang Yang, Siddharth Gupta, Daniele Veneziano, Shounak Athavale, and Marta C. González. The timegeo modeling framework for urban mobility without travel surveys. *Applied Physical Sciences*, 2016.

[6] Remi Louf and Marc Barthelemy. Modeling the polycentric transition of cities. *Phys. Rev. Lett.*, 111:198702, Nov 2013.

[7] Markus Schläpfer, Lei Dong, Kevin O'Keeffe, Paolo Santi, Michael Szell, Hadrien Salat, Samuel Anklesaria, Mohammad Vazifeh, and Carlo Ratti Geoffrey B. West. The universal visitation law of human mobility. *Nature*, 593, 2021.

[8] Yanyan Xu, Luis E. Olmos, David Mateo, Alberto Hernando, Xiaokang Yang, and Marta C. Gonzalez. Urban dynamics through the lens of human mobility, 2023.