Measuring urban regional similarity through mobility signatures

Grant McKenzie, Daniel Romm

Platial Analysis Lab, Department of Geography, McGill University, Montreal, Canada
grant.mckenzie@mcgill.ca, daniel.romm@mail.mcgill.ca

Abstract

The task of identifying similar regions within and between cities is an important aspect of urban data science as well as applied domains such as real estate, tourism, and urban planning. Regional similarity is typically assessed through comparing socio-demographic variables, resource availability, or urban infrastructure. An essential dimension, often overlooked for this task, is the spatiotemporal mobility patterns of people within a city. In this work we present a novel approach to identifying regional similarity based on human mobility as proxied through micromobility trips. We use a dataset consisting of e-scooter trip origins and destinations for two major European cities that differ in population size and urban structure. Three dimensions of these data are used in modeling the spatial and temporal variability in movement between regions in cities, allowing us to compare regions through a mobility lens. The result is a parameterized similarity model and interactive web platform for comparing regions across different urban environments. The application of this model suggests that human mobility patterns are a quantifiable, unique, and appropriate characteristic through which to measure urban similarity.

Keywords: mobility, city, similarity, urban, micromobility, neighborhood

1. Introduction

An objective definition of a neighborhood or district within a city is elusive. Cities, and their sub-regions, are often described in relation to other cities, districts, or neighborhoods. For instance, a visitor to a new city is frequently heard describing their current environment by reflecting on a city with which they are more familiar. The difficulty in defining regions within a city is that they contain nuanced characteristics and qualities that resist objective definition; these features are much more easily defined in comparison to *other* places. In *The Image of the City*, Lynch (1960) describes regions of the city as thematic units that are distinctive primarily in contrast to the rest of the city. This implies that an important facet of a region's identity within an urban environment is its similarity, or lack thereof, to other parts of a city.

This task of finding similar locations is a basic function of geographic information systems. Identifying target markets in which to establish new retail locations based on the similarity of the region to other successful regions, for instance, is one of the primary tasks of industry-based GIS analysts. The basis on which *similarity* is determined ranges considerably. From the perspective of a commercial venture, similarity comparisons typically involve demographics, existing competitors, urban infrastructure, etc. In recent years, methods have expanded to consider similarity based on social themes (Adams and Janowicz, 2015), functional activity space (Gao et al., 2017), and natural language place descriptions (Kim et al., 2017), building on the rise of user-generated content and context-aware technologies. This topic garners interest because assessing similarity is a challenging process that can be approached from many different angles; cities, and the nuanced characteristics that differentiate regions within them, are multi-faceted and incredibly complex (Batty, 2008).

While demographic characteristics and urban morphology are often used to differentiate regions within and between cities, the movement of a city's urban population is also important to understanding similarities and differences between its regions. The dynamics of human movement speak to the function and form of urban regions, and modeling a city through spatiotemporal mobility patterns allows one to better understand the *pulse* of a city (Froehlich et al., 2009; Xu et al., 2019). In this work, we propose to extract of a set of *mobility signatures* from multidimensional human mobility patterns. In our analysis, these urban patterns are realized as high resolution spatial and temporal trip origins and destinations (OD) collected from a large sample of micromobility users. More specifically, we extract mobility signatures from free-floating e-scooters in two major European cities, namely Berlin, Germany and Stockholm, Sweden. Dockless e-scooters are part of the new, free-floating iteration of micromobility which eschews traditional docking stations in favor of *smart-locks* that allow users to purchase vehicle access through their mobile devices. By removing the need for a physical, fixed access-point infrastructure, users can pick up and drop off vehicles throughout a service area. The inherent flexibility of free-floating systems means that trip trajectories better model user demand – the user theoretically can access a vehicle where demand begins, and deposit a vehicle where demand ends. Trip data produced from these systems therefore closely proxies the demand of users for micromobility trips without constraints, making it valuable for examining mobility patterns in cities. While this is a rich dataset for observing dynamic mobility patterns in the real-world, it is important to note that these data are not a representative sample of a city's underlying population. The extraction of these mobility signatures is intended to be data-agnostic, meaning that this model can be applied to any spatiotemporal OD dataset. In this case, we use free-floating micromobility data to demonstrate the methodology. Once extracted, these mobility signatures form the basis of a methodology that measures the similarity between two regions. This approach allows us to compare regions within a city, and between multiple cities, based on numerous dimensions of the data.

Similarity is a complex topic but one that is not new to geographic information science and urban analytics. Early work by Tversky (1976), writing in the cognitive science literature, defined a geometric model of similarity based on metric distance between entities represented in some coordinate space. This approach has evolved to use other distance measures such as cosine similarity, Earth Mover's distance, etc. Various cognitive approaches to assessing similarity have emerged, with most recognizing that as similarity assessments are a fundamental component in reasoning and induction, the user has a governing role as they select criteria by which to assess similarity (Holt, 1999). It is with this perspective that we frame our approach to regional similarity, combining a data-driven mobility signatures methodology with a top-down, more authoritative, user-weighted approach, which accommodates human-centered methods of understanding similarity.

While the primary objective of this paper is a region-to-region similarity measure, we conclude our analysis with the development of city-wide mobility signatures that can be used for identifying *prototypical* regions. This concept draws from early work on prototype theory (Rosch, 1973) which describes the idea of conceptual categories, where certain features present a higher degree of centrality or belonging to a conceptual category than others. Prototype theory has come to play an important role in our understanding of core geographic concepts (Hahn et al., 2016; Hu, 2018). We build on this theory to identify regions of a city that best represent the city as a whole, as well as those that are least representative. With these objectives in mind, the work presented here addresses the following three research questions (*RQ*).

- RQ1 Can mobility signatures extracted from micromobility trips be used to differentiate regions within a city? Can these signatures also be used to identify similar regions *between* cities? To accomplish this, we aggregate trip origins and destinations, extracting unique dimensions of the data on which to construct a set of spatiotemporal mobility signatures.

- RQ2 Having extracted mobility signatures for each region in one or more cities, in which ways do these mobility signatures differ from one another? Does a similarity calculation based on each of these mobility signatures independently produce the same regional similarity values? We demonstrate that while the mobility signatures show some degree of overlap, there are important differences in the final results.

- RQ3 Can the identification of *prototypical* and *least representative* mobility patterns help us to identify nuanced similarities and differences between cities? For instance, what region of Berlin is most similar to the city of Stockholm as a whole? We demonstrate that cities have unique prototypical mobility patterns that can be used to differentiate them from one another.

In addition to these research questions, we designed an interactive web platform through which users can identify similar regions between cities, based on the mobility signatures and similarity model we present in this paper. The objective of this platform is to allow users to better understand the data, the weights, and the various components that contribute to our mobility-based regional similarity model. Readers are encouraged to visit the online application at `https://platial.science/citysim/` for further visual exploration.

2. Related work

Measuring similarity between regions stems from a broader body of research on spatial similarity. Yan and Li (2015) define spatial similarity relations based on the degree to which the properties of two geographic objects are identical. Crucially, spatial similarity assessments consist of both quantitative and qualitative comparisons. Qualitative similarity assessments relate to a cognitive construal of similarity, which recognizes the inconsistency between geometric models and the way humans actually understand similarity (Tversky, 1976). Bruns and Egenhofer (1996) combine the geometric and the cognitive construals of similarity by integrating three models of similarity: topological relations, distance relations which can be ordinal to accommodate qualitative reasoning, and direction relations using cardinal directions to qualitatively capture the orientation between spatial objects. Li and Fonseca (2006) build on this model by defining two types of similarity when comparing spatial scenes: relational similarity and object similarity.

Focusing specifically on the urban environment, there have been numerous data-based representations of a city used in similarity assessments. A large body of existing research has focused on the socio-demographic composition of the city. For instance, Uitenbroek et al. (1996) identified similarities and differences between three European cities based on the demographics and habitual behavior of inhabitants. A study by Thomas et al. (2012) compared the urban morphology between European neighborhoods and found that neighborhoods are often more similar across national borders than within cities. Further research compared communities in and around Vancouver, Canada with the goal of identifying regions that have similar disaster vulnerability profiles (Chang et al., 2015). This work focused on the economic, social, built environment, and natural environment capital, to facilitate policy learning between similar communities. The *livehoods* project (Cranshaw et al., 2012) defined regions within a city based on user-generated social content, identifying and differentiating spatial clusters within a city. Other work in this vein makes use of geosocial content for identifying inter-urban mobility patterns (Liu et al., 2014) and better understanding the dynamics of the city based on participatory sensing (Silva et al., 2014). Recent work by Olson et al. (2021) used restaurant reviews as the basis for understanding how a city, and regions within the city, have changed over time. Novel machine learning models have been applied to this domain with the purpose of better understanding how regions can be explained through the make-up of the place types they contain (Yan et al., 2017; Liu et al., 2019). This has continued to drive the bourgeoning field of geographic artificial intelligence research, of which spatial and regional similarity is front and center (Janowicz et al., 2020). Our work is intended to present an alternative and supplemental approach to urban similarity analysis that focuses on mobility rather than socio-demographics, urban structure, or geosocial content.

Similarity measures have a rich history in trajectory analysis and transportation studies. In recent years we have also seen a push towards better integrating movement science within the broader domains of urban studies and geographic information science (Demšar et al., 2020; Miller et al., 2019a). Human mobility research is a significant component of both domains (Siła-Nowicka et al., 2016; Kraemer et al., 2020) and the deluge of rich mobility data is serving to further integrate the two. Existing work demonstrates that human mobility patterns exhibit high degrees of temporal and spatial regularity (Gonzalez et al., 2008). Day-to-day human mobility patterns show some degree of scale variability that can partially be explained through the use of spatial *containers* that limit mobility behavior (Alessandretti et al., 2020). From a statistical perspective, *cosine similarity* is commonly used to measure similarity in data that can be represented numerically. This approach measures the angle between numerical feature vectors and reports a single similarity value. A wide range of similarity-based studies have relied on the cosine measure, including research on text classification and clustering (Li and Han, 2013; Muflikhah and Baharudin, 2009), mobility patterns Liu et al. (2020); McKenzie (2020), and tourism studies Grbovic and Cheng (2018).

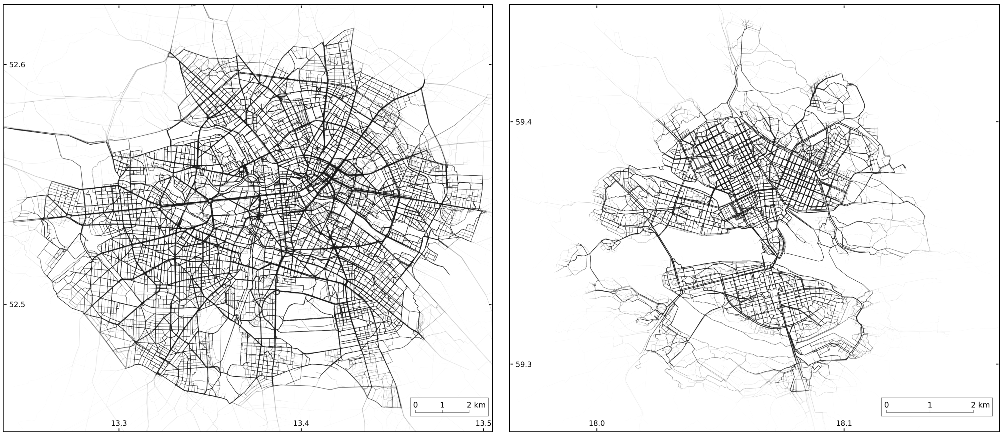
Many methods for trajectory and movement analysis, in a more general sense, employ similarity measures to make predictions based on similar, existing trajectories (Purves et al., 2014). For example, Abraham and Sojan Lal (2012) determine the similarity of trajectories through proximity to points of interest during times of interest to construct spatiotemporal similarity matrices. Kim and Chang (2009) take a different approach, instead using spatial distance and temporal distance between trajectories. Instead of focusing on individual trajectories, other approaches use similarity analysis to find common patterns embedded in trajectory data, enabling the retrieval of trajectories based on the similarity between the shape of a trajectory and candidate trajectories (Yanagisawa et al., 2003; Graser et al., 2019). These approaches are reflected in applied work, such as identifying public transit passengers who share similar travel patterns or travel behavior (Krizek and El-Geneidy, 2007; Faroqi et al., 2017). We take a mobility signatures approach to identifying similarity between regions, constructing signatures from aggregated patterns and using cosine similarity to compare signatures across regions. Other approaches to assessing similarity have treated mobility data as matrices and used different similarity methods to compare the matrices. For instance, Shoval and Isaacson (2007) used a sequence alignment method to compare human trajectories and develop clusters of patterns. A spatiotemporal alteration of the Edit Distance method was developed by Yuan and Raubal (2014) to better interpret human mobility patterns based on call detailed records. Other research in this domain has leveraged space-time prisms as a foundation on which to assess similarities in trajectories. For example, Buchin and Purves (2013) modeled trajectory speed using space-time prisms and assessed similarities between trajectories using Fréchet and equal time distance methods. Another approach assesses the similarity of space-time prisms using the previously defined concept of temporal signatures (Miller et al., 2019b). The authors calculate the semantic similarity of these patterns based on the temporal signatures, and later using network-time prisms (Jaegal and Miller, 2020). While these are all useful approaches for clustering trajectories and identifying similarities between space-time prisms, they do not directly compare aggregate mobile patterns with the purpose of quantifying the similarity of *regions*, a focus of this work.

The use of free-floating micromobility data for mobility pattern analysis is in its nascency, though a large body of literature has explored mobility patterns through traditional docking station-based bike-share platforms (Fishman et al., 2013). While many researchers have used trip data from services to make inferences about micromobility systems, few have utilized free-floating system data to examine other facets of the city. Several researchers have used free-floating system data to evaluate the relationship between street-level features, such as street morphology and greenness, and cycling frequency (Wang et al., 2020; Chen et al., 2020; Meng and Zacharias, 2020). Adopting micromobility data as a fill-in for all short-trip transportation patterns, Li et al. (2020) integrates trip data, transportation network data, road characteristics, and land use data to analyze bicycling patterns as they relate to other city features and flows. More generally, free-floating system data allows for an empirical analysis of travel dynamics over the first-and-last-mile of transportation (Yang et al., 2019). To the best of our knowledge, however, this is the first time that mobility signatures extracted from micromobility data have been used in the development of a model for identifying urban region similarity.

3. Data

Available vehicle locations were accessed for the free floating e-scooter operator, *Tier*, in two European cities, Berlin and Stockholm, through their application programming interface (API) (`https://platform.tier-services.io`). Tier, founded in 2018, is one of the larger micromobility companies operating in Europe, serving over 80 cities in 10 countries. The geographic coordinates, unique vehicle identifier, and battery life percentage of available vehicles in a given city were accessed every 60 seconds for 13 weeks starting August 28, 2020. In these data, we observed that a vehicle would appear available for a period of time and then disappear from the available vehicles dataset, indicating that it had either been rented or was being redistributed by the operating company. Some time later, the vehicle identifier would then re-emerge in the data at a different location. Trips were constructed from these data by recording the identifier, location, timestamp, and battery percentage of a vehicle directly before it disappeared from the data (origin) and again when the vehicle identifier reappeared in the dataset some time later (destination).

Criteria were set for labeling this disappearance as a *trip*. Specifically, a vehicle's location needed to move more than 100 meters and a trip could last no longer than 2 hours, the approximate duration of a vehicle's battery in full use. Furthermore, trip distance was calculated along a city's road network. The shortest path route between the origin and destination for each proposed trip were computed using the Open Source Routing Machine (`http://project-osrm.org/`) with routing set to shortest distance along the OpenStreetMap road network (including foot



(a) Berlin

(b) Stockholm

Figure 1: All trips as shortest paths between origins and destinations along respective city road networks. Trip geometries were calculated using the Open Source Routing Machine on OpenStreetMap roads and foot paths.

paths). For reference, the results of the OD routing for our two example cities are shown in Figure 1. Given the speed restriction of 20 kph placed on *Tier* e-scooters within the aforementioned cities, all routed trip distances with average speeds faster than 20 kph were removed from analysis. It was assumed that these were either redistribution trips with the vehicles traveling by truck, or data errors. For the same reason, all trips in which the battery percentage of the e-scooter increased between origin and destination were removed from further analysis. This approach to data cleaning, also used in McKenzie (2019), is conservative, in that it attempts to remove all vehicle redistribution and errors in the data at the cost of removing a few non-conforming *actual* user trips. This cleaned set of trips in both cities is the dataset used in our analysis. Table 1 provides an overview of the data from the cleaned trips.

	Berlin, Germany	**Stockholm, Sweden**
Total Trips	528,202	251,750
Unique Vehicles	8,542	2,889
Mean Vehicles/Day	2,376	1,293
Mean Trips/Day	5,804	2,766
Average Trip Length (meters)	3,004 (2,083)	2,759 (2,294)
Average Duration (seconds)	1,057 (789)	854 (687)
Population (metro area 2021)	3,567,000	1,657,000

Table 1: An overview of e-scooter trips and vehicles in Berlin and Stockholm based on cleaned data. Averages report the mean with median in parentheses.

Region delineation

For the purposes of this study, we define our *regions* based on a hexagon tessellation constructed over each city. These regions were generated over the two cities of interest at a resolution of 2 km. The spatial extent of greater Berlin is substantially larger than Stockholm, as is the spatial distribution of e-scooter vehicles. The length and duration of e-scooter trips, however, are quite similar. It was therefore determined that the size of hexagons would remain the same for both cities. Intersecting trip OD with the respective grids resulted in nearly three times as many populated grids for Berlin (354) than Stockholm (121), indicating that the spatial extent of e-scooter us is greater in Berlin. Again, the relative trip distances between grids remained quite similar.

The resolution of 2 km was selected through detailed exploration of the trip data. Each hexagon needed to be small enough to allow for differentiation between regions of the city, yet be large enough so as not to introduce sparsity issues in the spatiotemporal analysis. While not without flaws, the use of a hexagon tessellation at such a spatial resolution has a long history in spatial analysis and is supported by similar transportation and urban analyses (Ke et al., 2018; Sahr et al., 2003). This choice of region shape and size is further explored in the Discussion section.

4. Mobility Signatures

In this section we present the methodology for extracting three unique *Mobility Signatures*, which form the foundation of our regional similarity model. Each signature is based on a different dimension of the (micro)mobility data. For every trip destination region in Berlin and Stockholm, we computed the following mobility signatures.

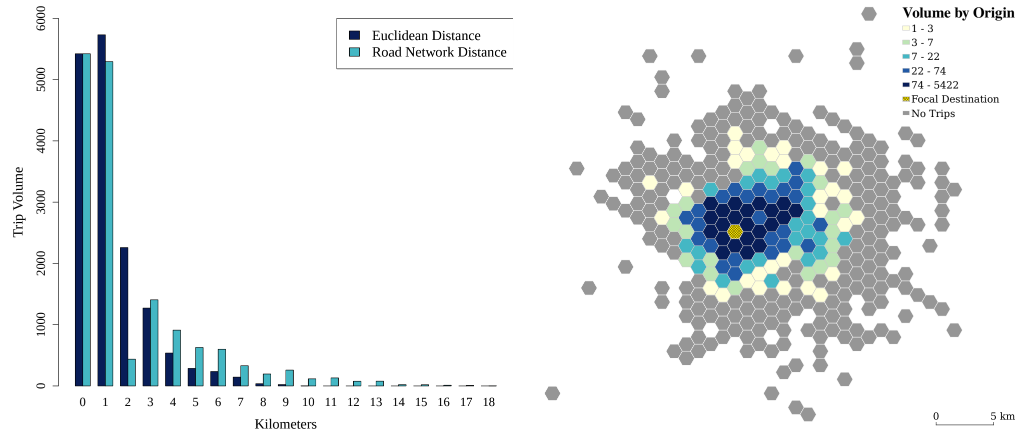
- Volume of trips by distance traveled, V;
- Number of unique trip origin regions by distance traveled, U; and
- Aggregate temporal signature by distance traveled, T.

Given the range of trip distances, they were discretized into 1 km resolution *distance bins*. For example, a 350 m long trip was assigned to the bin representing all trips 0-1 km and a 4,210 m trip was assigned to the bin representing trips between 4-5 km. In the following sections, we describe the process of extracting each of these mobility signatures.

4.1. Volume of trips by distance (V)

As a first step, we calculated the Euclidean distance between all pairs of hexagonal region centroids. Trips were then aggregated into their origin and destination hexagons and summed across 1 km distance bins. This results in a distribution of trip volumes by distance, as shown in Figure 2a. The neighboring map, shown in Figure 2b, depicts this same information cartographically to highlight the spatial distribution of trips.

We then completed the same analysis using the distance between trip origins and destinations as calculated along the city road network, including foot paths. As shown in Figure 2a, the distributions are significantly different given that trip distance along a road network is longer than Euclidean distance in virtually all cases. This highlights the fact that urban trips, micromobility or otherwise, do not exist in abstract space but rather are confined by physical geography and urban morphology. The distance calculation method chosen has a substantial effect on any mobility-based similarity model. We completed this trip volume analysis using road network-based distance for all destination regions in each of our two study cities, producing a set of V for each city.



(a) Distribution of trip volume by distance for one destination region.

(b) Map of trip volume by origin for one destination region.

Figure 2: Trip volume by distance between origin regions and a single focal destination region in Berlin, for illustrative purposes. The bar plot in Figure 2(a) shows trip volume distributed by average trip distance (1 km bins). The map in Figure 2(b) demonstrates the spatial distribution of trip origins from the sample destination region. Euclidean distance between OD regions are shown alongside average shortest path distance between OD regions on the Berlin road network.

4.2. *Unique trip origins by distance (U)*

The previous volume-based mobility signature separates trip volume by distance bins, but it does not consider the number of unique origin regions within each distance bin. This means that two destination regions could be identified as similar despite the fact that one destination consists of trips from a single origin region at each distance bin, whereas the other may consist of numerous origins within each distance bin. We construct U for each destination region by counting the number of unique trip origin regions by distance. Figure 3 shows these distributions for two sample destination regions in the city of Berlin, one near the city center and one further outside the center, demonstrating the value of this measure. Given that the total possible count of origin regions by binned distance has an upper limit dictated by the hexagonal tessellation, these values need not be normalized. This unique origin approach to constructing the mobility signature (U) complements the previous volume-based method by representing an alternative dimension of mobility, namely, the distribution of trip origins.

4.3. *Temporal signatures by distance (T)*

The extraction of mobility signatures, up to this point, has solely been concerned with the spatial distribution of trips, their origins and destinations, and the trip volume. An additional dimension of mobility that differentiates regions within a city is *time*. Existing research (McKenzie and Adams, 2017; Sparks et al., 2020) has demonstrated that regions of a city are popular at different times of day, and days of the week. Which regions are busy when depends on a range of factors such as land-use, socio-demographics, and activity affordances (Jordan et al., 1998).

We augment the previous two *spatial-only* mobility signatures with a third mobility signature that is based on the time that people travel. This was accomplished by aggregating trip start times, within each origin region, into temporal bins. Trips starting between 05:00-10:00 were classified as Morning, 10:00-15:00 as Afternoon, 15:00-20:00 as Evening, and 20:00-05:00 as Night. These times were aggregated separately for weekends and weekdays producing eight

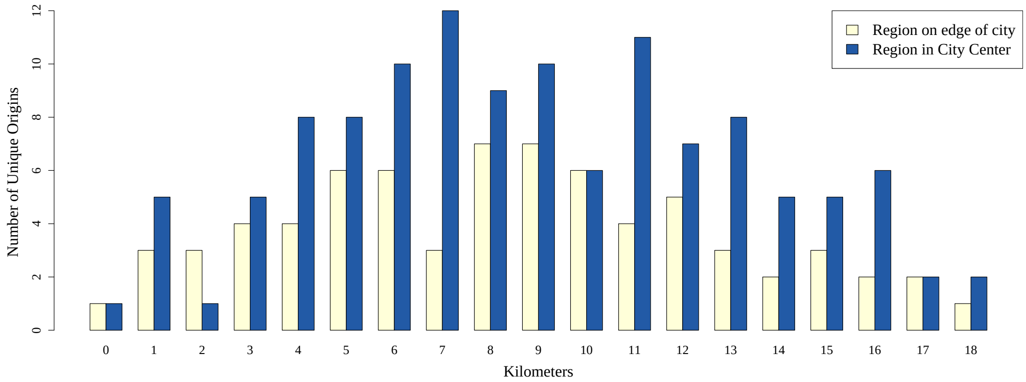


Figure 3: Number of unique origin regions by distance for two sample destination regions in Berlin.

different time periods. The resulting grouped trips for each of these time periods forms a *temporal signature*. As before, all trips were grouped into their origin and destination regions, but trips were then further split into one of the eight time windows. The temporal signatures were then normalized by the number of hours in each window so as not to artificially inflate the Night hours. Given the sparsity of the data, only those OD region pairs consisting of at least 10 trips were used in constructing the time-based mobility signature, T.

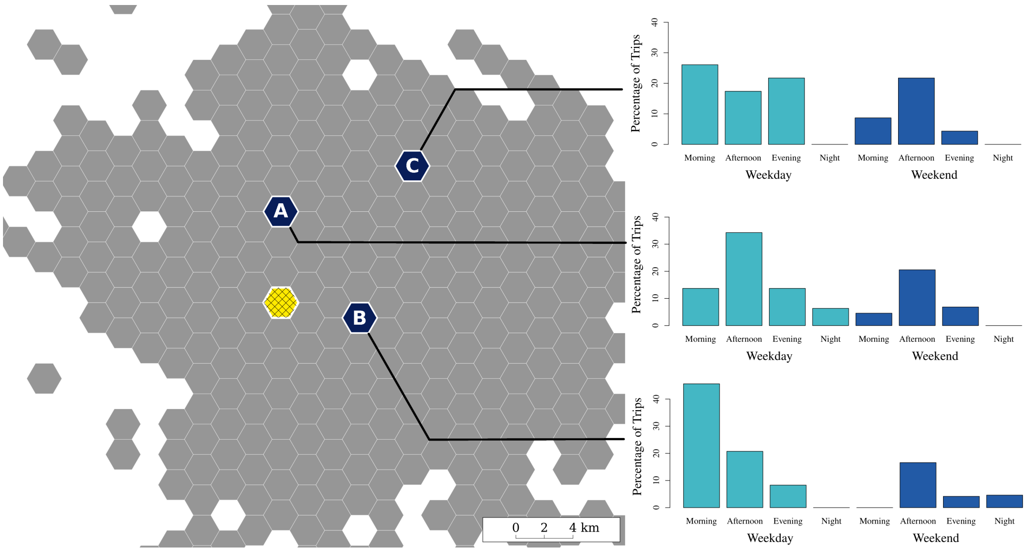


Figure 4: Three sample origin regions (A, B, C) containing trips that end within the hashed yellow destination region in Berlin. The associated temporal signatures, given in percentages of total trips, for each origin region are displayed on the right.

The inclusion of the temporal component means that for each destination region, and each 1 km distance bin, we potentially have a set of origin regions, each with its own temporal signature. We purposely elected not to average temporal signatures across all origin regions at the same distance. This is because two origin regions at the same distance from the destination region could be on opposite sides of the city, and may therefore present two very different temporal signatures. These temporal patterns are kept separate rather than averaging, which could potentially smooth any variance that exists between different regions. Figure 4 shows a simplified example temporal mobility signature, T, for a *focal* destination region shown in hashed yellow. Three sample trip origin regions (A, B, C) are featured in blue, linked to their respective temporal signatures. Origin regions A and B are within the same 5 km distance bin, but present very different temporal signatures. A shows a peak in the afternoons on weekdays and weekends whereas B presents a prominent peak on weekday mornings. Averaging these temporal signatures, based on distance, would flatten these values for a typical weekday. Instead, we produce a *set* of temporal signatures for each 1 km trip distance bin between origin and destination as shown in Equation 1. TS_i is the temporal signature for an individual origin region at D distance bins 0 - N km, where N is the longest distance between an origin and focal destination region in the dataset and M is the number of temporal signatures.

$$T = \begin{bmatrix} D(0km) = [TS_1, TS_2, TS_3, TS_M] \\ D(1km) = [TS_1] \\ \vdots \\ D(Nkm) = [TS_1, TS_2] \end{bmatrix} \tag{1}$$

A limitation of the hexagon tessellation is that a trip originating in one region may actually only be a few meters away from a trip originating in a neighboring region. This implies that the temporal signatures of hexagons in close proximity are more similar than those farther apart. To increase the robustness of our temporal signature approach, we applied a neighborhood smoothing kernel to the temporal signatures in each origin region. This involved applying a weighted average over the temporal signatures for the origin hexagon as well as the six adjacent hexagons using the weights shown in Equation 2 where TS_{H1-6} are the hexagons adjacent to the origin hexagon, TS_{OH}.

$$TS_{OH} = 0.24 \sum_{i=1}^{N=6} TS_{Hi} + 0.76 \cdot TS_{OH} \tag{2}$$

This weighted approach ensures that each neighboring region has a 4% influence on the temporal signature of the focal origin region, thus slightly reducing the impact of the arbitrarily defined hexagon boundaries produced by the tessellation. This value of 4% presented here was determined after analysis using several different weights. A sensitivity analysis determined that larger weights smoothed the focal regions too much and lost much of the nuance necessary for regional comparison. Smaller weights lead to large differences between regions in many cases. Importantly, this weighted value is a parameter that should be adjusted depending on the mobility dataset, the size of the regions, and the cities being compared.

5. Measuring similarity

The previous section presented a methodology for quantifying three different dimensions of human mobility into mobility signatures. In this section we use a combination of these signatures to calculate similarity between regions, both within a single city and between cities. Given that each of these signatures is a distribution of values (or sets of values in the case of T), we choose to approach signatures as multi-dimensional vectors, allowing us to use a *cosine similarity* approach to compare pairs of regions. Cosine similarity measures the cosine of the angle between two vectors projected in multi-dimensional space. This is equivalent to the inner product of these two vectors normalized so that both have a length of 1. The result of the cosine similarity between two vectors is a single value bounded between 0 (complete dissimilarity) and 1 (identical) (see Baeza-Yates et al. (1999) for further details). The number of dimensions in our case is the number of 1 km distance bins in our distributions. A benefit of this approach is that similarity is based on the angle and not the distance between two vectors, meaning that a difference in the magnitude of the values (i.e., volumes) does not impact the similarity model in the same way that it would with a Euclidean distance similarity method. This is particularly beneficial when comparing two cities that are substantially different in area and number of trips, as mobility signatures need not be normalized across two cities.

5.1. Within-city similarity

We start by calculating the cosine similarity between all possible region mobility signatures, V, in one city, producing a regional similarity matrix for trip volume, Sim_V^{ij}. This means that for any region i within Berlin, we can rank all other regions j based on similarity (from 1, identical, to 0, perfect dissimilarity). The same is done for U producing a regional similarity matrix, Sim_U^{ij}, based on the cosine similarity measure calculated for all pairs of unique origin region distributions. Given the low number of trips in a few of the regions on the edge of the city, only those regions that served as destinations for ten or more trips were included in further analysis. This was done to remove erroneously high similarity scores that were identified purely due to a lack of data.

The cosine similarity method becomes more complex when calculating regional similarity based on the temporal mobility signatures, T. Instead of a single attribute value (e.g., volume) at each 1 km distance, we have a potential *set* of temporal signatures. This is dealt with by first comparing all possible pairs of temporal signatures between two regions within each 1 km distance bin. Cosine similarity is again used to calculate similarity between each temporal signature pair (ts_i, ts_j) in the set of regions R, resulting in a temporal signature similarity matrix, $TempSim(d)^{ij}$, for each 1 km distance bin d in the distribution D.

The resulting similarity matrix, $TempSim(d)^{ij}$, is then ordered by the cosine similarity value, resulting in the most similar temporal signature pair listed first. The similarity matrix is then reduced to only the unique pairs of temporal signatures with the highest cosine similarity. This means that for each temporal signature in region R, only the temporal signature that results in the highest cosine similarity with another region will be recorded. There are no duplicate IDs in the resulting subset similarity matrix, $subSim$, and the length of this subset is determined by the minimum length of the two regional/distance temporal signatures.

The purpose of this approach is to identify the highest degree of temporal similarity between each region pair at each distance rather than randomly pairing temporal signatures and then calculating similarity. This method aims to assess the similarity of two regions based on the set of their most similar temporal signatures while removing less similar temporal signature pairings from further analysis. The mean of the final subset similarity matrix, $subSim$, is taken, producing a single similarity value for that region pair at that distance. Finally, we calculate the mean cosine similarity value across the entire distance vector resulting in a single similarity value for each pair of regions, i, j, and the similarity matrix, Sim_T^{ij}.

5.2. *Agreement between different mobility signature similarities*

Each of the three mobility signature-based similarity calculations mentioned above can be used independently to determine the similarity of two regions, specifically addressing the first part of RQ1. We must then ask, do these different approaches produce similar results? Does changing the underlying mobility signature (dimension of human mobility) change the resulting regional similarities (RQ2)? We address this in a variety of ways starting with an analysis of the mobility signatures themselves.

We first calculate the Pearson's correlation between each pair of the original mobility signatures. Overall, there is high positive correlation between all pairs of mobility signature similarities, with U to T showing the highest correlation, 0.841, followed by U to V at 0.671, and V to T at 0.583, all with $p < 0.01$. While one might expect that these correlations would be high, given that they are built on different dimensions of the same underlying trip data, this demonstrates that there are differences between these dimensions, and that the use of multiple mobility signatures is warranted for determining similarity between regions.

We then measure the level of agreement between the three similarity matrices (Sim_V^{ij}, Sim_U^{ij}, Sim_T^{ij}). We accomplish this by calculating the normalized *Discounted Cumulative Gain (DCG)* (Järvelin and Kekäläinen, 2002) for all combinations of the three similarity matrices. DCG is a method typically used for measuring the ranking quality of search engine query results, comparing a set of perfect (ideal) search result matches with those of a search engine's ranking algorithm. In our case we rank all regions in Sim_V^{ij} by similarity and assess how similar this ranking is to the ranking of the same regions in Sim_U^{ij}, for instance. DCG results in a single quantitative measure representing the rank-based similarity of two similarity matrices. A unique aspect of this measure is that it places greater weight on matches at higher ranks than those at lower ranks, through the inclusion of the log_2 function. Equation 3 shows how DCG is calculated where rel_i is the relevance of the region at rank position i. In our approach, the highest possible value for relevance would be $rel_i = N - i$ where N is the number of regions being compared (e.g., 153).

$$DCG = rel_1 + \sum_{i=2}^{N} \frac{rel_i}{\log_2(i+1)} \tag{3}$$

We calculate the *ideal* discounted cumulative gain (IDCG) as the ranking of one similarity matrix, e.g., Sim_V^{ij}, where the most similar region to the focal region is at position 1 with a relevance value of 152. The second most similar hexagon to the focal hexagon is at position 2 with a relevance value of 151, and so on. In this example, the IDCG is 2567.82. Next, we compare this ranking to another similarity matrix, e.g., Sim_U^{ij}. If the most similar region in Sim_U^{ij} is the second most similar region in Sim_V^{ij}, this would result in a smaller DCG value, since there is not a direct match. The *normalized* discounted cumulative gain (nDCG) is then calculated as the $DCG/IDCG$. The upper bound for nDCG is 1 (a perfect match between similarity matrices) whereas the lower bound is limited by N. Computing the nDCG for all regions in all similarity matrix pairs results in a nDCG for Sim_U^{ij}/Sim_V^{ij} of 0.932 (SD = 0.024), Sim_U^{ij}/Sim_T^{ij} of 0.963 (SD = 0.021), and Sim_V^{ij}/Sim_T^{ij} of 0.921 (SD = 0.025). These results indicate that there is a high degree of agreement between similarity assessments based on our three mobility signature measures. It also demonstrates that there are differences in the similarity of regions within a city depending on which dimension of the data one chooses. This allows us to answer RQ2 by definitively stating that each of our three mobility signatures leads to different assessments of similarity within a city.

5.3. Between-city similarity

Next, we compare regions between two different cities. This approach is virtually identical to within-city similarity with added consideration of the difference in city size and number of trips. As presented in Table 1, Berlin hosted double the number of trips and over three times the number of regions compared to Stockholm. Despite these differences we are reminded that the average distance traveled and duration of a trip were similar between both cities, suggesting there is a degree of universality to the distance and duration of scooter trips. This is supported by previous research in this area (Bai and Jiao, 2020; McKenzie, 2020).

Our assessment measure, which relies on cosine similarity, requires that two vectors have the same number of dimensions. In our case, this means that the greatest trip distance must be the same when comparing two regions. Ensuring two regions contain the same number of 1 km bins was easier for within-city comparisons, but requires further investigation for between-city assessments. In examining trips in our two cities of interest, we find that the greatest trip distance in Berlin was 32 km, with only 12 trips being longer than 16 km. By comparison, the longest trip in Stockholm was 16 km. To allow for comparison, the Berlin mobility signatures were reduced to a maximum of 16 km, in order to have two vectors of the same size for comparison.

The similarity methods discussed in the previous section were then applied to all possible pairs of Berlin–Stockholm regions producing three regional similarity matrices, one for each of the mobility signatures. Figure 5 cartographically depicts the resulting similarity measures, one map of Berlin for each of the MS, based on one selected region in Stockholm. The similarity values are notably different depending on the region and which mobility signature is visualized. This mirrors the results of the normalized discounted cumulative gain method and contributes to answering RQ2. The general trend, however, is that regions closer to the city center are determined to be more similar to the focal region in Stockholm than those further outside the downtown city center.

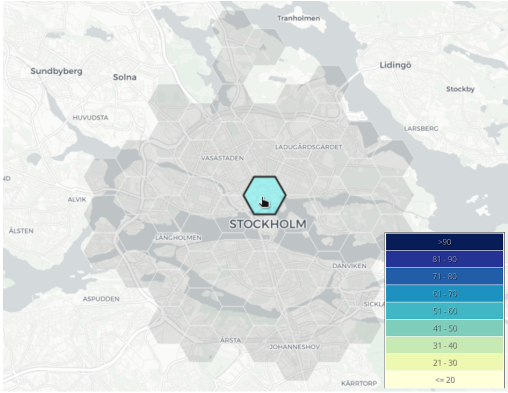
This trend of densely populated regions in one city being similar to densely populated regions in the other is seen consistently throughout a visual analysis of the data, with a few exceptions. While some of these exceptions may be explained by proximity to transit hubs, major shopping, or tourist destinations, others likely reflect latent or nuanced features of the city. Leveraging these nuanced features, identified through dimensions of human mobility behavior, is the primary objective of this research, but is also generative to further research using this methodology. Readers are encouraged to visit the online platform at https://platial.science/citysim/ for further visual exploration.

5.4. A weighted combination

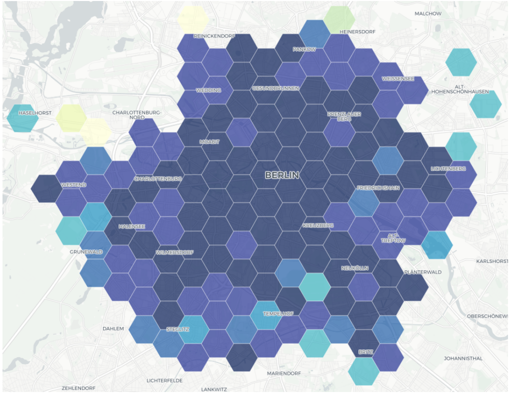
The results of comparing the different mobility signatures demonstrates that though there is often agreement as to which regions of the city each mobility signature identifies as similar, the level of agreement varies. This shows that while one dimension is useful for region similarity identification, the three different methods in combination provide a more holistic view on the similarity between urban regions. We propose a method (Equation 4) that allows a user, through our web platform, to determine how much weight should be assigned to each of the mobility signatures when calculating regional similarity, where $w_A + w_B + w_C = 1$.

$$WeightedSimilarity = Sim_V^{ij} \cdot w_A + Sim_U^{ij} \cdot w_B + Sim_T^{ij} \cdot w_C \tag{4}$$

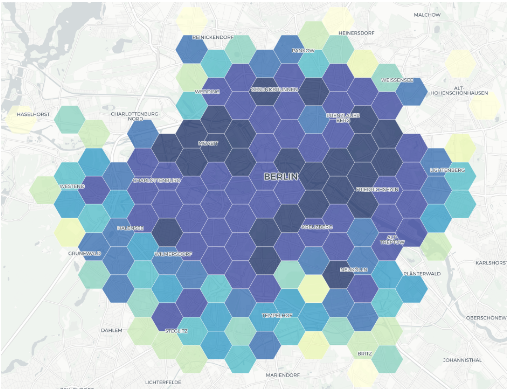
Though it is tempting to lock the weights at some combination, or *learn* the weights based on an external data source, it is important to remember that similarity, in this case as with many



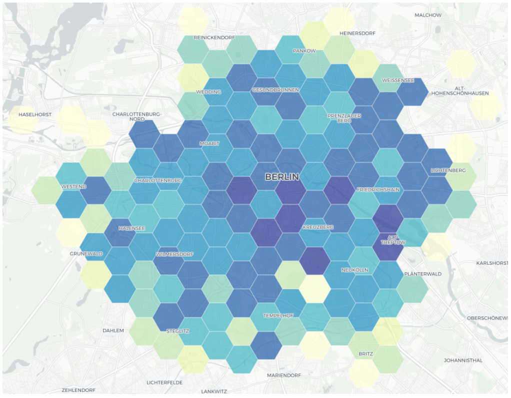
(a) Selected focal region in Stockholm



(b) Similarity to regions in Berlin based on V



(c) Similarity to regions in Berlin based on U



(d) Similarity to regions in Berlin based on T

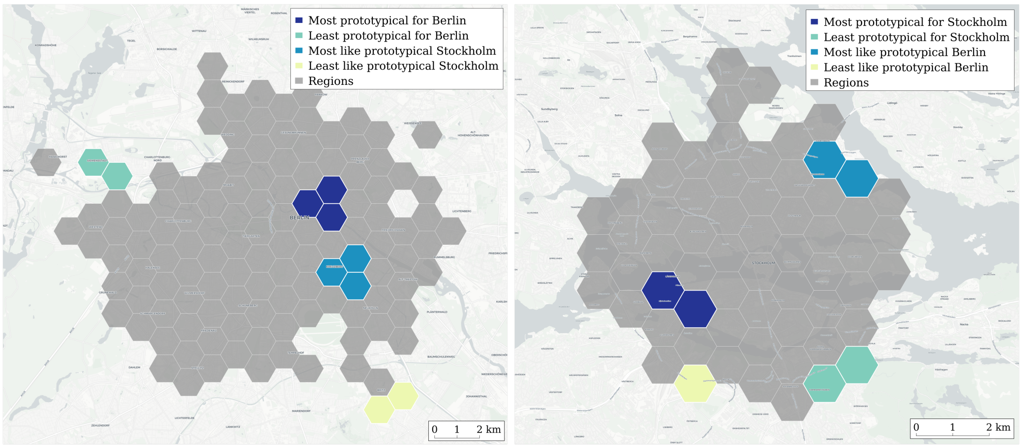
Figure 5: Regions in Berlin styled by similarity to a selected focal region in Stockholm. The three mobility signature-based similarity values are shown in Figures b-d. Screenshots taken from the interactive regional similarity platform. Base map and labels by Carto & OpenStreetMap.

others, is a subjective concept and depends on the features of a region in which one is most interested. While the work presented in this paper approaches this concept of similarity through the lens of data-driven mobility, we still want a human user to retain control over which dimensions of the data matter most to them. This accommodates the cognitive model of similarity which recognizes the importance of user governance in making similarity assessment.

5.5. *Prototypical mobility patterns*

The previous sections present an approach for selecting a single region from one city and producing a similarity-based ranking of all regions in another city. While this method is useful in many circumstances, we may also want to answer the question: What region of a city is most *prototypical* of that city? In other words, if one had to summarize an entire city by the mobility signatures of a single region, which region would that be? In contrast, what region of the city is the least representative of the city?

To answer these questions, we constructed three new mobility signatures based on the averages of the three existing mobility signatures, computed across all regions. We calculated the average trip volume, per 1 km distance traveled, for all regions and did the same for the unique number of origin regions, and all possible temporal signatures. As before, we did not average the temporal signatures in order to preserve the uniqueness of each temporal pattern within each origin region. We then compared each individual region's mobility signatures to our averaged mobility signatures using cosine similarity. The resulting three cosine similarity values were again averaged (mean) producing a single similarity value for each region in our city. For demonstration purposes, we evenly weighted each of the signatures when computing this mean. The region that showed the highest similarity to this mobility signature average was labeled the prototypical region of a city. In Figure 6, we highlight a number of different regions where the similarity values are close.



(a) Berlin

(b) Stockholm

Figure 6: Prototypical and least representative regions in Berlin and Stockholm. The most and least Berlin-like regions of Stockholm are also shown, and vice versa.

The least representative regions of a city were identified in the same way but instead of selecting the top region(s), the regions with the lowest average similarity are highlighted. Not surprisingly, in both of our example cities, the regions of the city that were least similar to the averaged mobility signature were on the edge of our analysis area, the outskirts of the city. These less representative regions consisted of fewer trips than the prototypical regions, but not fewer than many other regions, suggesting that the low similarity values were not a result of trip sparsity. For the city of Berlin, the least representative region is near the Großsiedlung Siemensstadt, a unique residential settlement in the Charlottenburg-Nord locality of Berlin. The most prototypical region is in the densely populated inner city center, near the downtown business district. The region in Berlin identified as the most prototypical is just outside the *Old Town* region of Berlin. This region is mixed land-use with residential and commercial properties combined with tourist attractions and major transit hubs. The daytime population trends slightly younger than the residential population, but mobility signatures reflect the wide range of micromobility users in this part of the city. Given that the prototypical signatures are based on an average of the mobility signatures from all the different regions, it is not surprising that this region is most similar to the average. The most prototypical region of Stockholm is the Hornstull district, in western Södermalm. While not precisely in the city center, it is within central Stockholm and contains a prominent public transit station. Again, the socio-demographic data for this region are based on a residential census, and therefore not reflective of the typical mobility user. An analysis of the land use in this regions shows predominantly high density residential with limited commercial zoning including restaurants and a shopping mall. The least representative region is to the south, across the water in the Gamla Östberga locality, an area arguably quite similar to Großsiedlung Siemensstadt in Berlin, with respect to land use and civil infrastructure. The regions both contain medium-density residential land use with a small number of amenities. Both are also well outside the downtown core of the city, yet still within the city limits.

Provided prototypical and representative regions in each of our example cities, one might instead ask: *What is the most Berlin-like region of Stockholm?* We calculated the cosine similarity between the averaged mobility signatures for Berlin and the mobility signatures for all individual regions in Stockholm. Ranking regions by similarity identifies those regions in Stockholm that are most similar to the averaged mobility signatures of Berlin. The results, along with the least representative regions, are again shown in Figure 6. Notably, the resulting regions are both outside of the downtown core of the cities and predominantly residential land use. An important finding of this analysis is that the prototypical region identified in one city is not the same as the region identified as most prototypical of a different city. This brings attention to the uniqueness of each city and the fact that while a transportation system – a single micromobility operator in our case – exists in two cities, the mobility signatures built from usage behavior are so nuanced as to clearly differentiate them.

6. Discussion

This work presents a novel methodology for identifying similarity and differences between regions based on the mobility of a city's inhabitants and visitors. The reader will note that the concept of *accuracy* as it relates to similarity is not mentioned in this work, nor are we claiming that the regional similarities presented are *correct*. Similarity is highly subjective, as what one person finds similar, another may not. Our approach offers one possible methodology for quantifying and comparing dynamic spatiotemporal mobility patterns across aggregate regions. Our intention is for this methodology to act as a substructure on which to facilitate continued discussion of city similarity, combined also with other regional similarity measures to provide a robust set of methods with which to assess urban similarity.

Within this work, we asked and answered three research questions. In response to RQ1, we demonstrate that human mobility data, as represented by micromobility trips, consists of multiple spatial and temporal dimensions. Patterns can be extracted from these dimensions, which we refer to as a set of mobility signatures. These signatures are distinct enough to differentiate regions within cities and form the basis for assessing similarity between cities. In response to RQ2, we show that regional similarity matrices built on each of our mobility signatures independently are similar, but differ in important and unique ways. We propose combining the mobility signature-based similarity measures to construct a more robust picture of regional similarity. In addressing RQ3, we present a method for identifying prototypical regions within a city, and a technique for identifying regions of a city that are most like a different city.

The spatial resolution at which we aggregate trip origins and destinations merits further discussion. As mentioned previously, we chose to aggregate all trip origins and destinations into 2 km hexagonal *regions*. We tested numerous region sizes, with the final resolution being a compromise between data sparsity and region size. We wanted regions that were small enough to fit within a typical subdistrict or neighborhood, but not so small that too few trips originated or completed within the region, limiting analysis. Other regional delineation methods were explored, such as municipally defined sub-city boundaries (e.g., neighborhoods, districts, or localities), though due to the lack of standardization between cities, a neighborhood in one city may be an order of magnitude larger than a neighborhood in another. The hexagon tessellation we ultimately selected is at a high enough resolution to allow one or more hexagons to fit within most individual neighborhoods or other localities. Agglomeration techniques to accomplish this are a topic for future work. Traffic analysis zones, postal codes, and local census boundaries were also examined but deemed unsuitable, given the range of sizes and original purposes for the construction of these boundaries (e.g., postal delivery routes, transportation planning, etc.).

Finally, the mobility signatures developed in this paper initially focus on trip *destinations* but could have instead been built from trip *origins*. The methodological approach would be virtually identical, so we elected to not repeat the analysis in a project focused on the development of the methodology. In addition, we must remember that the underlying dataset for the analysis was free-floating, short-term rental e-scooters, and a destination for one trip is most often the origin for the next. Further analysis will investigate how the results might vary with alternative datasets that do not follow this same *shared-mobility* model.

6.1. Limitations

Despite our best efforts, there are limitations to this analysis. First, the two cities used as examples in this study range considerably in size, population, road network, etc. The choice of these two cities was intentional, as one of our objectives was to demonstrate how this approach is city-agnostic. It must be mentioned, however, that differences between cities cannot be entirely quantified through the mobility of their inhabitants. There are a plethora of latent factors that can be taken into consideration, and this approach is only meant to be one possible way of measuring similarity. Furthermore, we must again acknowledge that the exemplar data used in this analysis, namely free-floating scooter trips, is not a representative sample of a city's population and should not be seen as such. Early surveys have found that e-scooter users tend to skew younger and more male than a city's population (Dill and McNeil, 2020; Christoforou et al., 2021). While useful for demonstrating the possibilities of such a mobility dataset, the actual results presented in this work are meant to reflect methodological possibilities, rather than guide urban policy decisions.

We relied on shortest-path distance along a road and foot path network in this work. It is improbable that all users took the shortest path along this network. The distances that resulted from this method are likely still more accurate than simply computing the Euclidean distance between the origin and destination. Not knowing the actual trajectory of a micromobility user is a limitation of this approach and should be considered when interpreting the final results. Furthermore, we chose to group our trips into 1 km bins in order to have enough data for comparative analysis. One alternative approach would be to use trip duration instead of distance and bin by 10 minute intervals, for example. Much of this work relied on aggregating data and calculating mean trips, times, and distances. Additional statistical approaches could be used such as calculating the variance or medians of these trips to help further refine the models.

As mentioned previously, the aggregation of point based geometries (trip origins and destinations) into spatial regions comes with a number of biases, not least of which is the modifiable areal unit problem (MAUP). The use of an equal sized hexagon tessellation is better than many other geometries, but introduces other biases. For instance, the borders between hexagons were arbitrarily determined, meaning that a cohesive neighborhood may have been split between hexagons. We attempted to mitigate this issue, as mentioned in the temporal signatures section, by applying a weighted kernel to smooth these borders. Physiographic features were ignored, meaning that land and water were both included in the hexagonal regions and civil infrastructure (e.g., rail lines or major highways) were also not considered in region delineation. Future analyses might also consider cluster-based approaches using algorithms that consider physiographic or infrastructural features, though this introduces further complications in making similarity assessments.

6.2. Future Work

This is a methods paper. Our work employs data from two European cities and demonstrates a set of methods for comparing them. Next steps will involve applying these techniques to a larger set of cities and making an effort to compare the results of the model with similarity assessments ascertained through community surveys, infrastructure-based similarity models, and other measures (see for example, Currid and Williams (2010)). Additional work will explore techniques for clustering similar mobility signatures into super-regions exhibiting similar mobility behavior, and comparing these to traditionally inhabitant-defined neighborhoods or sub-districts of a city. This would allow colloquial names to be applied to regions with similar mobility signatures, which could be used in a range of real world applications.

The approach presented here focuses purely on origins and destinations, though distance and duration are calculated through routing along a road network. An area of future investigation will involve including transitory cells to enhance the mobility signatures. In many cities, there exist transit corridors in which few trips originate or finish, but which serve an important role in differentiating parts of the city. The impact of these regions will be explored further.

Finally, the methodology presented here will be applied to a variety of mobility datasets in order to test the robustness of such an approach, and determine the variation in similarity that results from a change in the underlying data source. The eventual goal of this work is to build a more holistic model of human mobility in cities that takes into consideration multiple modes of travel, enabling a more robust similarity assessment.

7. Conclusions

Identifying similar regions within and between cities has a range of possible uses. For instance, from a policy perspective, identifying similar regions in different cities may support policy transfer between those regions. Urban planners eager to promote growth in one neighborhood may observe mobility signatures in a more vibrant neighborhood and take policy actions to encourage similar mobility usage. Transportation engineers may choose to add or remove public transit infrastructure based on perceived optimal mobility patterns of different regions. Government agencies may look to a mobility-based similarity model for actionable insight when other models are not suitable or report no discernible difference between regions. The design of this model is intended to be another *lens* through which policy makers can observe a city.

While much of the existing research on regional similarity has focused on socio-demographics and urban infrastructure, human mobility patterns present a unique, dynamic resource on which to build a model for identifying similar regions. In this work, we examined a rich set of micro-mobility trajectories in two large European cities, extracting and aggregating unique spatial and temporal features of the data as mobility signatures. The resulting set of signatures form the basic units on which urban regions are compared. A set of similarity models were constructed based on these mobility signatures producing similarity values for all regions within and between cities.

This permits us to identity which regions of a city are most and least similar as well as which regions are most representative of a given city. By identifying *prototypical regions*, we also formulate an alternative method for determining the average mobility pattern of a city, centered on regions. The approach presented in this work allows policy makers to control the weight of each mobility signature on the regional similarity model. This work contributes to the larger goal of identifying similar urban regions and makes a strong argument for the use of human mobility patterns as the basis for similarity assessment.

Acknowledgments

This research was supported in part by the Social Sciences and Humanities Research Council of Canada and the Natural Sciences and Engineering Research Council of Canada.

References

Abraham, S., Sojan Lal, P., Aug. 2012. Spatio-temporal similarity of network-constrained moving object trajectories using sequence alignment of travel locations. Transportation Research Part C: Emerging Technologies 23, 109–123.

Adams, B., Janowicz, K., 2015. Thematic signatures for cleansing and enriching place-related linked data. International Journal of Geographical Information Science 29 (4), 556–579.

Alessandretti, L., Aslak, U., Lehmann, S., 2020. The scales of human mobility. Nature 587 (7834), 402–407.

Baeza-Yates, R., Ribeiro-Neto, B., et al., 1999. Modern information retrieval. Vol. 463. ACM press New York.

Bai, S., Jiao, J., 2020. Dockless e-scooter usage patterns and urban built environments: A comparison study of austin, tx, and minneapolis, mn. Travel behaviour and society 20, 264–272.

Batty, M., 2008. The size, scale, and shape of cities. science 319 (5864), 769–771.

Bruns, H. T., Egenhofer, M. J., 1996. Similarity of Spatial Scenes. In: Proceedings of the 7th International Symposium on Spatial Data Handling: Advances in GIS Research II, 12-16 August 1996, Delft, The Netherlands: Vol. 1: SDH '96. Vol. 1. Taylor & Francis Ltd., Delft, The Netherlands, pp. 173–184.

Buchin, M., Purves, R. S., 2013. Computing similarity of coarse and irregular trajectories using space-time prisms. In: Proceedings of the 21st ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems. pp. 456–459.

Chang, S. E., Yip, J. Z. K., van Zijll de Jong, S. L., Chaster, R., Lowcock, A., Sep. 2015. Using vulnerability indicators to develop resilience networks: A similarity approach. Natural Hazards 78 (3), 1827–1841.

Chen, Y., Chen, Y., Tu, W., Zeng, X., May 2020. Is eye-level greening associated with the use of dockless shared bicycles? Urban Forestry & Urban Greening 51, 126690.

Christoforou, Z., Gioldasis, C., de Bortoli, A., Seidowsky, R., Mar. 2021. Who is using e-scooters and how? Evidence from Paris. Transportation Research Part D: Transport and Environment 92, 102708.

Cranshaw, J., Schwartz, R., Hong, J., Sadeh, N., 2012. The livehoods project: Utilizing social media to understand the dynamics of a city. In: Proceedings of the International AAAI Conference on Web and Social Media. Vol. 6.

Currid, E., Williams, S., 2010. Two cities, five industries: Similarities and differences within and between cultural industries in new york and los angeles. Journal of Planning Education and Research 29 (3), 322–335.

Demšar, U., Long, J. A., Siła-Nowicka, K., 2020. Integrated science of movement. Journal of Spatial Information Science 2020 (21), 25–31.

Dill, J., McNeil, N., Oct. 2020. Are Shared Vehicles Shared by All? A Review of Equity and Vehicle Sharing. Journal of Planning Literature, 088541222096673.

Faroqi, H., Mesbah, M., Kim, J., 2017. Spatial-Temporal Similarity Correlation between Public Transit Passengers Using Smart Card Data. Journal of Advanced Transportation 2017, 1–14.

Fishman, E., Washington, S., Haworth, N., 2013. Bike share: a synthesis of the literature. Transport reviews 33 (2), 148–165.

Froehlich, J., Neumann, J., Oliver, N., Jul. 2009. Sensing and predicting the pulse of the city through shared bicycling. In: Proceedings of the 21st International Jont Conference on Artifical Intelligence. IJCAI'09. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, pp. 1420–1426.

Gao, S., Janowicz, K., Couclelis, H., 2017. Extracting urban functional regions from points of interest and human activities on location-based social networks. Transactions in GIS 21 (3), 446–467.

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

Graser, A., Schmidt, J., Dragaschnig, M., Widhalm, P., 2019. Data-driven trajectory prediction and spatial variability of prediction performance in maritime location based services. In: Adjunct Proceedings of the 15th International Conference onLocation-Based Services. Wien.

Grbovic, M., Cheng, H., 2018. Real-time personalization using embeddings for search ranking at airbnb. In: Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. pp. 311–320.

Hahn, J., Fogliaroni, P., Frank, A. U., Navratil, G., 2016. A computational model for context and spatial concepts. In: Geospatial data in a changing world. Springer, pp. 3–19.

Holt, A., 1999. Spatial similarity and GIS: The grouping of spatial kinds. In: SIRC '99: Our Safe Living Environment : The Changing Face of Spatial Systems : Proceedings of the 11th Annual Colloquium of the Spatial Information Research Centre, 13-15 December 1999. Spatial Information Research Centre, University of Otago, Dunedin, New Zealand, p. 10.

Hu, Y., 2018. Geo-text data and data-driven geospatial semantics. Geography Compass 12 (11), e12404.

Jaegal, Y., Miller, H. J., 2020. Measuring the structural similarity of network time prisms using temporal signatures with graph indices. Transactions in GIS 24 (1), 3–26.

Janowicz, K., Gao, S., McKenzie, G., Hu, Y., Bhaduri, B., 2020. GeoAI: Spatially explicit artificial intelligence techniques for geographic knowledge discovery and beyond. International Journal of Geographical Information Science 34 (4), 625–636.

Järvelin, K., Kekäläinen, J., 2002. Cumulated gain-based evaluation of ir techniques. ACM Transactions on Information Systems (TOIS) 20 (4), 422–446.

Jordan, T., Raubal, M., Gartrell, B., Egenhofer, M., 1998. An affordance-based model of place in gis. In: 8th Int. Symposium on Spatial Data Handling, SDH. Vol. 98. Citeseer, pp. 98–109.

Ke, J., Yang, H., Zheng, H., Chen, X., Jia, Y., Gong, P., Ye, J., 2018. Hexagon-based convolutional neural network for supply-demand forecasting of ride-sourcing services. IEEE Transactions on Intelligent Transportation Systems 20 (11), 4160–4173.

Kim, J., Vasardani, M., Winter, S., 2017. Similarity matching for integrating spatial information extracted from place descriptions. International Journal of Geographical Information Science 31 (1), 56–80.

Kim, Y.-C., Chang, J.-W., 2009. A New Similar Trajectory Search Algorithm Based on Spatio-Temporal Similarity Measure for Moving Objects in Road Networks. IEICE Transactions on Information and Systems E92-D (2), 327–331.

Kraemer, M. U., Sadilek, A., Zhang, Q., Marchal, N. A., Tuli, G., Cohn, E. L., Hswen, Y., Perkins, T. A., Smith, D. L., Reiner, R. C., et al., 2020. Mapping global variation in human mobility. Nature Human Behaviour 4 (8), 800–810.

Krizek, K. J., El-Geneidy, A., 2007. Segmenting preferences and habits of transit users and non-users. Journal of public transportation 10 (3), 5.

Li, B., Fonseca, F., Mar. 2006. TDD: A Comprehensive Model for Qualitative Spatial Similarity Assessment. Spatial Cognition & Computation 6 (1), 31–62.

Li, B., Han, L., 2013. Distance weighted cosine similarity measure for text classification. In: International conference on intelligent data engineering and automated learning. Springer, pp. 611–618.

Li, W., Wang, S., Zhang, X., Jia, Q., Tian, Y., Feb. 2020. Understanding intra-urban human mobility through an exploratory spatiotemporal analysis of bike-sharing trajectories. International Journal of Geographical Information Science, 1–24.

Liu, X., Andris, C., Rahimi, S., 2019. Place niche and its regional variability: Measuring spatial context patterns for points of interest with representation learning. Computers, Environment and Urban Systems 75, 146–160.

Liu, Y., Sui, Z., Kang, C., Gao, Y., 2014. Uncovering patterns of inter-urban trip and spatial interaction from social media check-in data. PloS one 9 (1), e86026.

Liu, Z., Gong, Z., Li, J., Wu, K., 2020. Mobility-aware dynamic taxi ridesharing. In: 2020 IEEE 36th International Conference on Data Engineering (ICDE). IEEE, pp. 961–972.

Lynch, K., 1960. The image of the city. Vol. 11. MIT press.

McKenzie, G., 2019. Spatiotemporal comparative analysis of scooter-share and bike-share usage patterns in washington, dc. Journal of transport geography 78, 19–28.

McKenzie, G., 2020. Urban mobility in the sharing economy: A spatiotemporal comparison of shared mobility services. Computers, Environment and Urban Systems 79, 101418.

McKenzie, G., Adams, B., 2017. Juxtaposing thematic regions derived from spatial and platial user-generated content. In: 13th international conference on spatial information theory (COSIT 2017). Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik.

Meng, S., Zacharias, J., Sep. 2020. Street morphology and travel by dockless shared bicycles in Beijing, China. International Journal of Sustainable Transportation, 1–11.

Miller, H. J., Dodge, S., Miller, J., Bohrer, G., 2019a. Towards an integrated science of movement: converging research on animal movement ecology and human mobility science. International Journal of Geographical Information Science 33 (5), 855–876.

Miller, H. J., Jaegal, Y., Raubal, M., 2019b. Measuring the geometric and semantic similarity of space–time prisms using temporal signatures. Annals of the American Association of Geographers 109 (3), 730–753.

Muflikhah, L., Baharudin, B., 2009. Document clustering using concept space and cosine similarity measurement. In: 2009 International Conference on Computer Technology and Development. Vol. 1. IEEE, pp. 58–62.

Olson, A. W., Calderon-Figueroa, F., Bidian, O., Silver, D., Sanner, S., 2021. Reading the city through its neighbourhoods: Deep text embeddings of yelp reviews as a basis for determining similarity and change. Cities 110, 103045.

Purves, R. S., Laube, P., Buchin, M., Speckmann, B., 2014. Moving beyond the point: An agenda for research in movement analysis with real data. Computers, Environment and Urban Systems 47, 1–4.

Rosch, E. H., 1973. Natural categories. Cognitive psychology 4 (3), 328–350.

Sahr, K., White, D., Kimerling, A. J., 2003. Geodesic discrete global grid systems. Cartography and Geographic Information Science 30 (2), 121–134.

Shoval, N., Isaacson, M., 2007. Sequence alignment as a method for human activity analysis in space and time. Annals of the Association of American geographers 97 (2), 282–297.

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

Silva, T. H., De Melo, P. O. V., Almeida, J. M., Loureiro, A. A., 2014. Large-scale study of city dynamics and urban social behavior using participatory sensing. IEEE Wireless Communications 21 (1), 42–51.

Sparks, K., Thakur, G., Pasarkar, A., Urban, M., 2020. A global analysis of cities' geosocial temporal signatures for points of interest hours of operation. International Journal of Geographical Information Science 34 (4), 759–776.

Thomas, I., Frankhauser, P., Badariotti, D., Apr. 2012. Comparing the fractality of European urban neighbourhoods: Do national contexts matter? Journal of Geographical Systems 14 (2), 189–208.

Tversky, A., 1976. Features of Similarity. Psychological Review 84 (4), 327–352.

Uitenbroek, D. G., Kerekovska, A., Festchieva, N., 1996. Health lifestyle behaviour and socio-demographic characteristics. a study of varna, glasgow and edinburgh. Social science & medicine 43 (3), 367–377.

Wang, R., Lu, Y., Wu, X., Liu, Y., Yao, Y., Aug. 2020. Relationship between eye-level greenness and cycling frequency around metro stations in Shenzhen, China: A big data approach. Sustainable Cities and Society 59, 102201.

Xu, Y., Chen, D., Zhang, X., Tu, W., Chen, Y., Shen, Y., Ratti, C., May 2019. Unravel the landscape and pulses of cycling activities from a dockless bike-sharing system. Computers, Environment and Urban Systems 75, 184–203.

Yan, B., Janowicz, K., Mai, G., Gao, S., 2017. From itdl to place2vec: Reasoning about place type similarity and relatedness by learning embeddings from augmented spatial contexts. In: Proceedings of the 25th ACM SIGSPATIAL international conference on advances in geographic information systems. pp. 1–10.

Yan, H., Li, J., 2015. Spatial Similarity Relations in Multi-Scale Map Spaces. Springer International Publishing, Cham.

Yanagisawa, Y., Akahani, J.-i., Satoh, T., 2003. Shape-Based Similarity Query for Trajectory of Mobile Objects. In: Chen, M.-S., Chrysanthis, P. K., Sloman, M., Zaslavsky, A. (Eds.), MDM 2003: Mobile Data Management. Lecture Notes in Computer Science. Springer, Berlin, Heidelberg, pp. 63–77.

Yang, Y., Heppenstall, A., Turner, A., Comber, A., Sep. 2019. A spatiotemporal and graph-based analysis of dockless bike sharing patterns to understand urban flows over the last mile. Computers, Environment and Urban Systems 77, 101361.

Yuan, Y., Raubal, M., 2014. Measuring similarity of mobile phone user trajectories–a spatio-temporal edit distance method. International Journal of Geographical Information Science 28 (3), 496–520.