

## **Amenity complexity and urban locations of socio-economic mixing**

Sándor Juhász, Gergo Pintér, Ádám Kovács, Endre Borza, Gergely Mónus, László Lorincz, and Balázs Lengyel

Papers in Evolutionary Economic Geography

# 22.32



Utrecht University

Human Geography and Planning

# Amenity complexity and urban locations of socio-economic mixing

Sándor Juhász <sup>1,2,3,\*</sup>, Gergő Pintér <sup>2</sup>, Ádám Kovács <sup>3</sup>, Endre Borza <sup>3</sup>,  
Gergely Mónus <sup>2</sup>, László Lőrincz <sup>2,3</sup>, and Balázs Lengyel <sup>2,3</sup>

<sup>1</sup>Complexity Science Hub Vienna, Vienna, Austria

<sup>2</sup>NETI Lab, Corvinus University of Budapest, Budapest, Hungary

<sup>3</sup>ANET Lab, ELKH Center for Economic and Regional Sciences, Budapest, Hungary

\*Corresponding author: [juhasz@csh.ac.at](mailto:juhasz@csh.ac.at)

## Abstract

Cities host diverse people and their mixing is the engine of prosperity. In turn, segregation and inequalities are common features of most cities and locations that enable the meeting of people with different socio-economic status are key for urban inclusion. In this study, we adopt the concept of economic complexity to quantify the ability of locations – on the level of neighborhoods and amenities – to attract diverse visitors from various socio-economic backgrounds across the city. Utilizing the spatial distribution of point of interests inside the city of Budapest, Hungary, we construct the measures of amenity complexity based on the local portfolio of diverse and non-ubiquitous amenities. We investigate mixing patterns at visited third places by tracing the daily mobility of individuals and characterizing their socio-economic status by the real-estate price of their home locations. Results suggest that measures of ubiquity and diversity of amenities do not, but amenity complexity correlates with the diversity of visitors to neighborhoods and to actual amenities alike. We demonstrate that, in this monocentric city, amenity complexity is correlated with the relative geographic centrality of locations, which in itself is a strong predictor of socio-economic mixing. Our work combines urban mobility data with economic complexity thinking to show that the diversity of non-ubiquitous amenities, central locations, and the potentials for socio-economic mixing are interrelated.

**Keywords:** urban mobility, economic complexity, amenities, social mixing

**JEL codes:** R10, R30, Z13

## 1 Introduction

Diversity is the key ingredient of successful and resilient cities (Jacobs [1961](#)). The spatially concentrated interaction of people from various social and economic background create environments that foster creativity (Florida [2004](#)), support inclusion (Benton-Short and Short [2013](#)) and in general, make cities vivid and prosperous (Glaeser [2012](#)). At the same time, cities show high levels of

segregation such that individuals from different socio-economic background are separated from each other in the urban space (Musterd [2020](#)). This phenomenon limits social mobility for many (Mayer and Jencks [1989](#)) and induced inequalities can expose segregated groups to health or climate crises (Torrats-Espinosa [2021](#); Loughran and Elliott [2022](#)) and can imply radicalization and populism (Abadie [2006](#); Engler and Weisstanner [2021](#)).

Recent studies leverage GPS mobility data to study socio-economic segregation and mixing patterns in visited urban locations (Cagney et al. [2020](#)). This growing literature frequently reports that people in cities visit and interact with locations that are similar to their residential neighborhood in terms of income, education, ethnicity or other socio-economic features (Wang, Nolan Edward Phillips, and Sampson [2018](#); Dong et al. [2020](#); Bokányi et al. [2021](#); Hilman, Iñiguez, and Karsai [2022](#)). However, the places, services or amenities that individuals visit in the city exhibit different levels of experienced segregation, as some locations mix different socio-economic groups while others do not (Athey et al. [2021](#); Moro et al. [2021](#)).

In this study we search for urban locations that foster socio-economic mixing. In other words, we aim to identify locations that present less experienced segregation and are visited by people from diverse strata. To do so, we emphasize two aspects of urban locations that can influence observed socio-economic mixing. First, category of amenities available at a location determine its purpose and function and therefore can influence social mixing. Noyman et al. ([2019](#)) illustrates through individual GPS trajectories that urban locations are visited by a more diverse set of people in case they offer entertainment amenities, services or natural water features. Athey et al. ([2021](#)) describes that individuals can experience relative low segregation at outdoor places like parks, sports fields and playgrounds, or at commercial establishments such as restaurants, bars and retail stores. They find that places of entertainment, like theaters and accommodations, like hotels are the least segregated urban locations. Moro et al. ([2021](#)) shows that the category of places is a strong predictor for experienced income segregation and unique places in cities, such as arts venues, museums or airports tend to be highly integrated, while places that primarily serve local communities, such as grocery stores or places of worship are generally more segregated by income. Despite previous empirical efforts, systematic examination on how the mixture of amenities at specific urban locations contribute to social mixing is still missing from the literature.

Second, specialized locations that serve the specific needs of the wider public and therefore can attract people from diverse neighborhoods tend to situate in the center of cities. The central place theory originally developed for the inter-urban scale by Christaller ([1933](#)) and Lösch ([1954](#)) explains the hierarchy of cities and towns through their size and the range of functions that they provide. Higher-order centers share most of the functions (goods and services) of lower order centers and some specialised functions that attract population from a larger area. Lower-order center will locate relatively close to one another for efficiency reasons as people do not want to travel far for their everyday needs such as grocery shopping. However, people would travel further for infrequent purchases or specialized goods and services, which would be located in higher-order centers that are farther apart. Building on the central place theory, Zhong et al. ([2017](#)) combines

density, the number of people attracted to locations and diversity, the range of activities that they engage with at these locations in a single centrality measure to identify urban centers in Singapore and illustrate their evolution over time. Noyman et al. (2019) shows that urban locations with higher centrality in urban road networks attract more diverse visitors. On the contrary, Moro et al. (2021) presents that urban locations with higher average travel distance to them tend to be less segregated than locations that are highly accessible. While most of the studies highlight that accessible, central locations attract more diverse visitors, yet, the nature of the available amenity mix might be related to the urban location, which has not been focused on so far.

Here, we aim to extend the above literature by investigating how the amenity mix and central position of urban locations are related to experienced segregation or, put it differently, to the mixing of people from diverse socio-economic strata. A new contribution is the application of the economic complexity framework to urban amenities (Hidalgo 2021) to quantify the ability of locations – on the level of neighborhoods and amenities as well – in attracting visitors of diverse socio-economic status from across the city.

The concept of economic complexity is originally developed by Hidalgo and Hausmann (2009) who defined complexity of economies by the diversity of their non-ubiquitous products and services. Economic complexity is indicative of countries economic growth, income level, emissions and inequalities (Hidalgo 2021). By now, the concept is applied to different data sources such as patents, occupations or scientific publications and to diverse spatial scales from countries to cities (Balland et al. 2022; Magalhães et al. 2023). Here we adopt the same logic to uncover the amenity complexity of urban locations. We argue that the complex amenity mix of an urban location should offer diverse amenities and the available amenities should be barely present at other locations.

Consequently, amenity complexity of an urban location is connected to social mixing of people from diverse socio-economic background for two reasons. First, diverse amenity mixes can attract people with diverse demands and second, locations with non-ubiquitous amenities can attract people from diverse neighborhoods, as the particular service is only available at the location in question. Therefore, our hypothesis is that the diverse mix of non-ubiquitous amenities can create an inclusive, multi-purpose neighborhood that is most likely to be attractive for a wide-variety of people. While the contribution of amenity mix to the socio-economic diversity of visitors at urban locations has been rarely unveiled, diverse amenities are argued to concentrate in and attract people to central places of cities (Zhong et al. 2017). To better understand the connection between amenity complexity, urban centrality and socio-economic mixing, we test the influence of distance from center and the components of amenity complexity on the diversity of visitors in parallel.

We build on the work of Hidalgo, Castañer, and Sevtsuk (2020) that created the amenity space by utilizing the co-location of amenities inside cities and construct the indicator(s) of amenity complexity. Utilizing the geographic distribution of point of interests (POIs) in urban neighborhoods, we measure the amenity complexity for both neighborhoods and amenity categories, analogously to economic complexity of regions and product complexity of economic outputs. We use these measures to illustrate that urban locations with complex structure of amenities attract diverse vis-

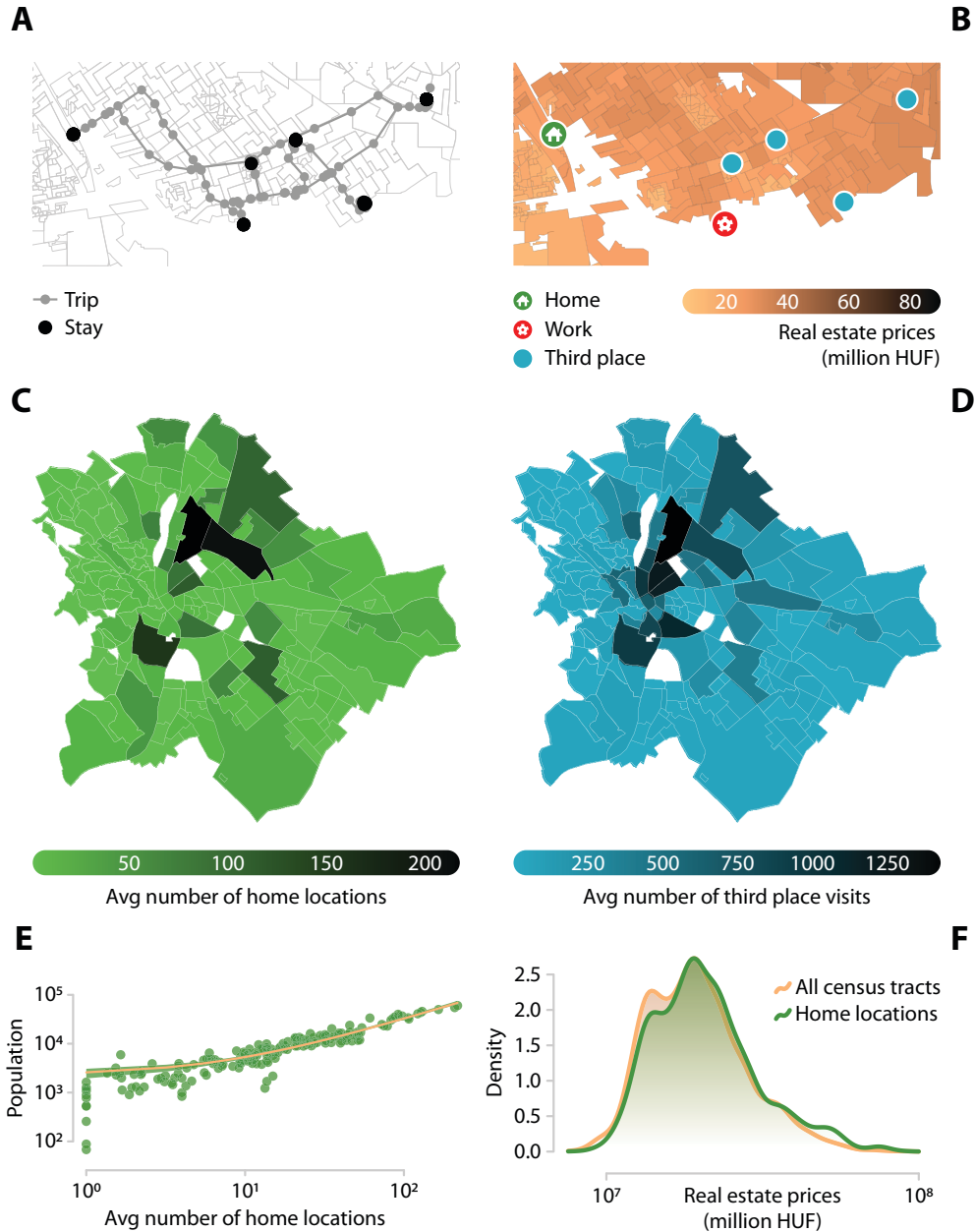
itors from across the city. We test this argument by combining POI data and mobility data from individual GPS trajectories inside Budapest, the capital of Hungary. More precisely, we collect information on all POIs in Budapest from the Google Places API to have a detailed understanding on the portfolio of amenities present in urban neighborhoods through the world’s most popular location service platform. We use data from a GPS aggregator company to trace individual mobility patterns. We identify home, work and third place visits (Oldenburg [1999]) in daily trajectories inside Budapest for 24 months by clustering the pings in geographical space and over time. We combine the information of predicted home locations with real estate prices at the census tract level. This allows us to infer the socio-economic diversity of visitors in each urban neighborhoods and in each actual amenity by investigating third place visits.

Our results illustrate that urban locations with a more complex amenity mix are visited by a larger diversity of socio-economic groups. We demonstrate that, in the monocentric city of Budapest, amenity complexity is correlated with the relative geographical centrality of locations and distance from the center is a strong predictor of socio-economic mixing. The contribution of this paper is that it combines urban mobility data with the concept of economic complexity to show that the diversity of non-ubiquitous services, central locations, and the potentials for socio-economic mixing are interrelated.

## 2 Tracing mobility inside cities

We capture socio-economic mixing at urban locations by identifying the diversity of visitors. In order to do so, we study urban mobility patterns of individuals. We rely on a raw GPS data from a data aggregator company. This data set allows us to trace the daily mobility of 5.2 million devices in Hungary over 24 months (between 2019 June and 2021 May). We initially filter this data to focus on devices that appear inside Budapest and have at least 20 GPS pings in total after discarding pings which indicate unreasonably high speeds of device mobility. Detailed description on the mobility data preparation process can be found in section S1 of the Supplementary information.

We process raw trajectories of individuals by applying the Infostop algorithm (Aslak and Alessandretti [2020]). It enables us to effectively detect the stationary points of individual movements and cluster GPS pings around stop locations. Figure 1A-B illustrates the raw data and the outcome of stop detection through an example device. The algorithm gives each stop a label indicating a place that can reoccur along the trajectory of the device. We focus on devices with at least 2 distinct places and 10 stops in a month inside Budapest. Using the monthly recurrence of stops and places by each device, we label places as home, work or third place visits. We do this in two steps. First, we categorize each visited place as potential home or work based on the part of the day it is visited, the duration of visits and their reappearance in the daily trajectory. A place is identified as potential home location in case the device spends the most amount of time there between 20:00 pm and 8:00 am on a given week and the combined time spent at the place exceeds 8 hours. Places where devices spend most of their time between 8:00 am and 17:00 pm in a week



**Figure 1:** Identifying home locations and third places visits from daily mobility trajectories. **(A)** Example trajectory to illustrate the stop detection process. **(B)** Identified places from the stop detection results and their function as home, work or third places. **(C)** Average number of home location over 24 months by urban neighborhoods of Budapest. **(D)** Average number of third place visits over 24 months by urban neighborhoods of Budapest. **(E)** The relationship between average number of home locations over 24 months and population of urban neighborhoods in Budapest. **(F)** Real estate prices at census tracts of identified home locations and across all census tracts of Budapest.

(exceeding at least 3 hours) is identified as the potential workplace of the device.

Second, we time-aggregate device trajectories to monthly visitation patterns. Thus, we identify home and work of a device in a month by the mean coordinate pairs of weekly potential home and work places, but only in case a device stops at the place at least 10 times over a month and the standard deviation of both latitude and longitude coordinates are smaller than 0.001 (about 100 meter in Budapest) over the respective month. We categorize every other visited place as a third place, in case it is labeled by the stop detection algorithm as a unique place, but it is not the home or the work place of the device in the respective month. Figure 1C presents the average number of devices with identified home location (and at least one visited third place) and Figure 1D illustrates the average number of third place visits over the 24 month period aggregated to the level of urban neighborhoods.

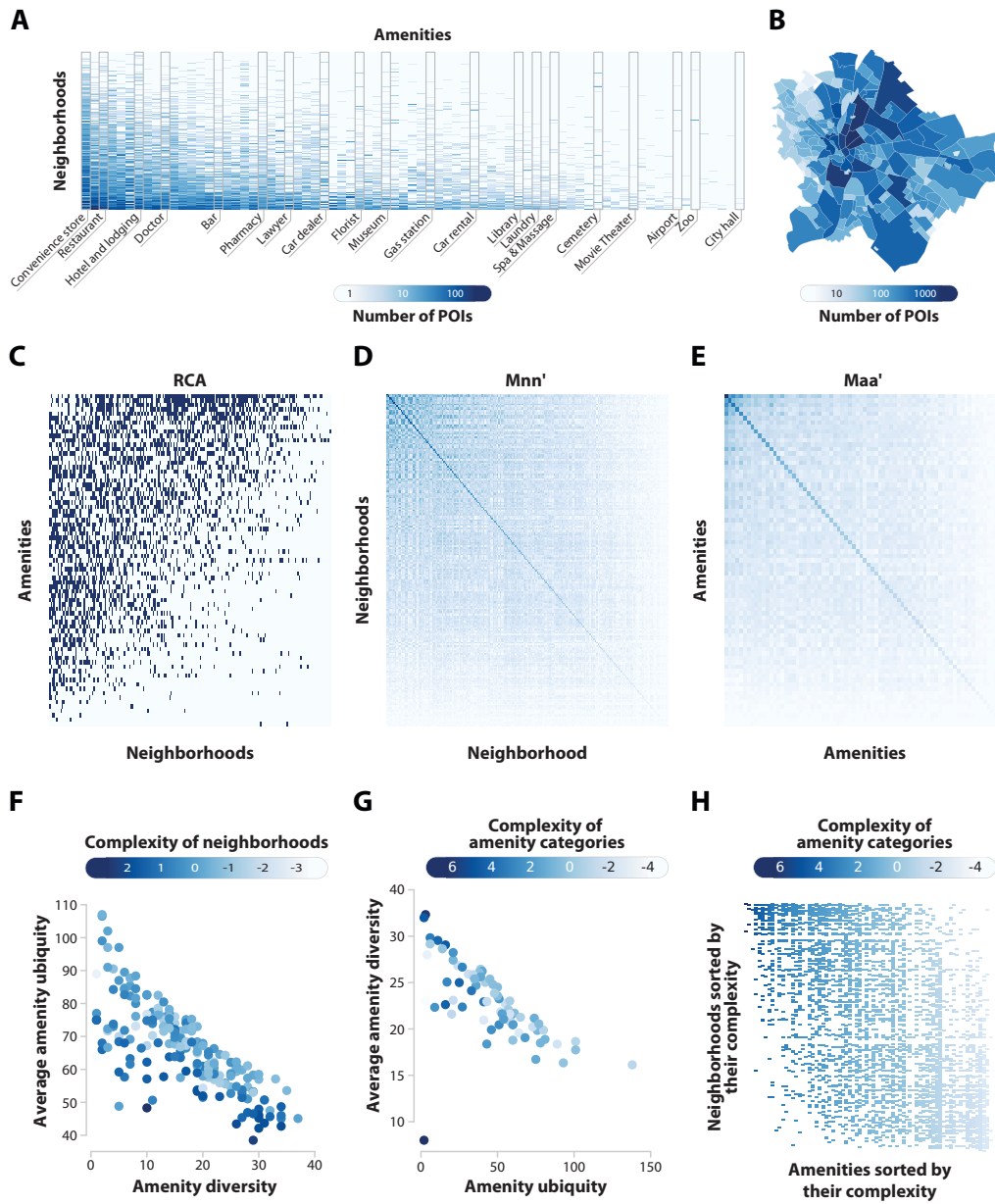
To join home locations and third places to other data sources with spatial reference, we rely on Uber’s Hexagonal Hierarchical Spatial Index (H3) (Uber Technologies, Inc. 2022). The applied indexes of size 10 H3 hexagons refer to an average 15.000 m<sup>2</sup> area, which is close to the buffer area of a point with a 70 meter radius. We connect all the identified home locations and third places to hexagons and split each neighborhood or census tract level polygons to the same hexagon size for efficient combination.

To infer on the socio-economic status of the followed devices, we join home locations to census tract level real estate prices. In Hungary, information on income is not part of the census data collection. We rely on residential real estate sales contracts from 2013-2019 collected by the Hungarian Central Statistical Office and predict real estate prices to each census tract of Budapest. Section S2 in the Supplementary information introduces the prediction process in detail. Figure 1F presents that real estate prices at the identified home locations and across all census tracts are closely align.

### 3 Measuring amenity complexity

To describe the attraction of urban locations for visitors, we construct the measures of amenity complexity. These indicators are based on the spatial distribution of amenities, which is studied through point of interest (POI) data from the Google Places API. Besides its limitations in terms of time scale and POI categorization, it is one of the world’s most popular mapping service supporting applications worldwide and helping millions of individuals on a daily basis to find the location of businesses. This makes Google data attractive to study the spatial organization of amenities inside cities (Hidalgo, Castañer, and Sevtsuk 2020; Kaufmann et al. 2022; Heroy et al. 2022).

We collected information as the latitude, longitude and amenity category for all the POIs around the city of Budapest in early 2022. The resulted data set contains 63.601 POIs in 78 different amenity categories. We removed the frequently appearing and ambiguous categories of ATM (1.054 POIs) and Parking (729 POIs) and filter out the category Casino with less than 2 POIs in Budapest. We use this data to illustrate the amenity profile of the 207 urban neighborhoods of Budapest (Hungarian Central Statistical Office 2022). Neighborhoods of Budapest are in be-



**Figure 2:** Constructing the amenity complexity measures. **(A)** Distribution of point of interests (POIs) across neighborhoods and amenity categories. **(B)** Map of urban neighborhoods colored by the number of observed POIs. **(C)** Revealed comparative advantage (RCA) values transformed to a binary specialization matrix (M). **(D)** Similarity matrix of neighborhoods based on their specialization in amenity categories. This matrix is used to measure the amenity complexity of urban neighborhoods. **(E)** Similarity matrix of amenities based on their specialization in neighborhoods. This matrix is used to measure the complexity of amenity categories. **(F)** Relationship between amenity diversity and average amenity ubiquity in neighborhoods. Dots (neighborhoods) are colored by their amenity complexity value. **(G)** Relationship between ubiquity and average diversity of amenities. Dots (amenity categories) are colored by their amenity complexity value. **(H)** Neighborhoods with higher amenity complexity are specialized in amenity categories that have a higher complexity value. Each cell in the matrix represents a neighborhood specialized in an amenity category and cells are colored by the complexity of amenity categories.



tween districts and census tracts in the spatial hierarchy, which makes them a suitable spatial scale for our analysis (Natera Orozco et al. 2020). They have an average population of 10.000 people (standard deviation around 10.000), have an average area of 2.5 km<sup>2</sup> (standard deviation around 3.9) and on average they consist of 41 lower level census tracts (standard deviation around 50). Further description about the urban neighborhoods of Budapest can be found in section S3 of the Supplementary information.

We consider every neighborhood in Budapest with at least 2 amenity categories with minimum 2 POIs. Figure 2A presents the resulted 75 amenity categories and the number of POIs across the focal 200 neighborhoods of Budapest. The most frequent categories are convenient store (5.989 observations), beauty salon (4.461 observations) and restaurant (3.727 observations), while we observe less than 10 amusement park, bowling alley and city hall. Figure 2B illustrates the unequal spatial distribution of POIs on the map of neighborhoods in Budapest.

To describe the relative importance of amenity categories and illustrate the differences between the amenity structure of urban neighborhoods, we adopt the economic complexity index (ECI) (Hidalgo and Hausmann 2009). The ECI is successfully used to describe the economic development of countries and regions (Hidalgo 2021) and its approach is adoptable to amenities and urban neighborhoods. We measure *amenity complexity* the following way. We normalize the matrix of Figure 2A to make comparisons appropriate between neighborhoods and amenity categories and compute the revealed comparative advantage (RCA) of neighborhoods in amenity categories by the following standard equation (also known as the Balassa index):

$$RCA_{n,a} = (P_{n,a}/P_a)/(P_n/P) \quad (1)$$

where  $P_{n,a}$  is the number of POIs in neighborhood  $n$  in amenity category  $a$  and missing indices indicate summed variables such as  $P_a = \sum_a P_{n,a}$ .  $RCA \geq 1$  suggests that neighborhood  $n$  is specialized in amenity category  $a$ . In other words, an amenity category is overrepresented in a neighborhood in case its  $RCA$  value is above or equal to 1. We use the  $RCA$  values to create a binary specialization matrix  $M_{n,a}$  the following way:

$$M_{n,a} = \begin{cases} 1 & \text{if } RCA_{n,a} \geq 1 \\ 0 & \text{if } RCA_{n,a} < 1 \end{cases} \quad (2)$$

Figure 2C illustrates the resulted binary  $RCA$  matrix of neighborhoods and amenity categories in Budapest. Sum of rows in this matrix present the number of amenity categories a neighborhood has comparative advantage in (amenity diversity) and the column sums give the number of neighborhoods where an amenity category is overrepresented (amenity ubiquity).

$$\text{Amenity diversity} = M_n = \sum_a M_{n,a} \quad (3)$$

$$\text{Amenity ubiquity} = M_a = \sum_n M_{n,a} \quad (4)$$

In geographic matrices like  $M$  the average ubiquity of the activities present in a location tends to correlate negatively with the diversity of activities in a location. This is the result of the matrix property known as nestedness and this feature is utilized to explain that more complex activities are only available at a handful of locations with a diverse portfolio of activities (Hidalgo and Hausmann 2009; Balland et al. 2020).

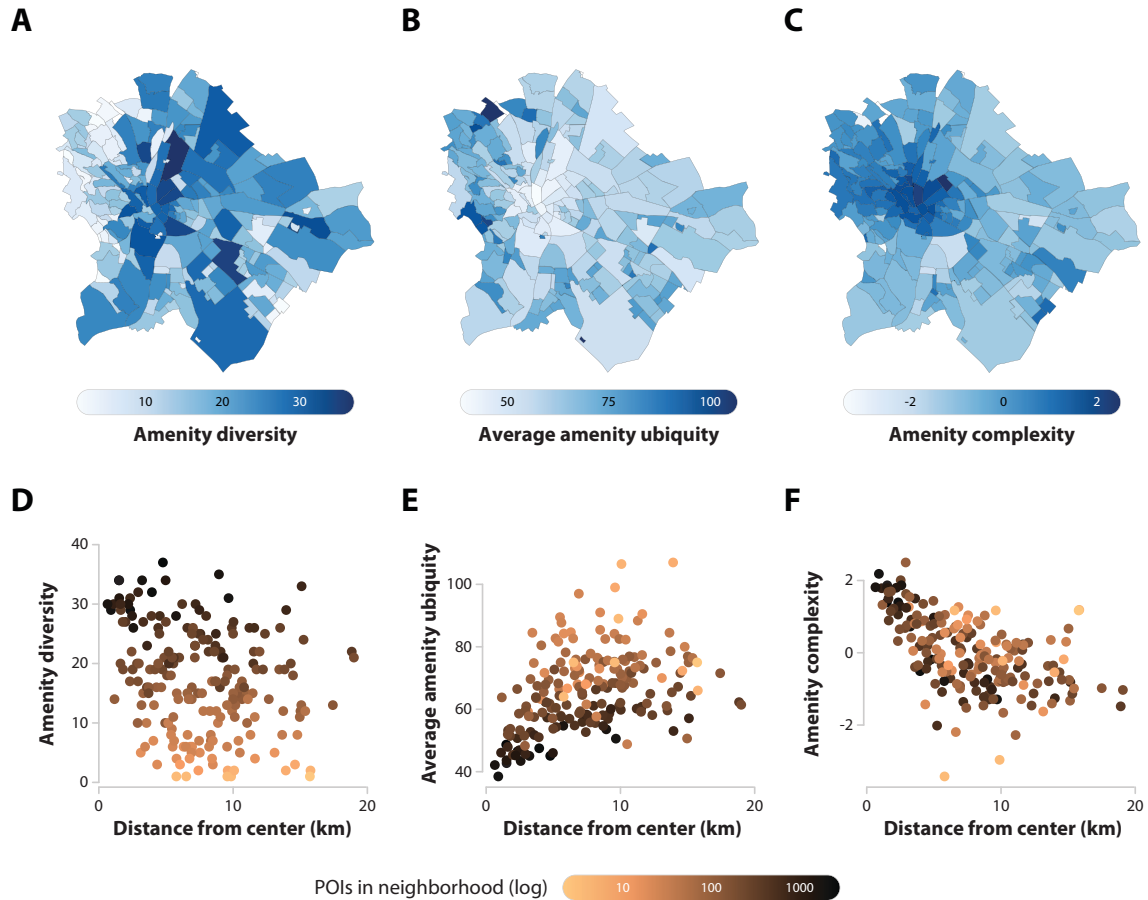
$$\text{Amenity complexity}_{\text{neighborhoods}} = K_n = \frac{1}{M_n} \sum_a M_{n,a} K_a \quad (5)$$

$$\text{Amenity complexity}_{\text{amenities}} = K_a = \frac{1}{M_a} \sum_n M_{n,a} K_n \quad (6)$$

The economic complexity index (ECI) that describes the production structure of economies were originally defined through the iterative, self-referential algorithm of the 'method of reflection' (Hidalgo and Hausmann 2009). The algorithm calculates the above explained diversity and ubiquity vectors and then recursively uses the information in one to correct the other in equation (5) and (6). Later it was presented that the method of reflection is equivalent to finding the eigenvectors of the similarity matrix  $M_{nn'}$  (Mealy, Farmer, and Teytelboym 2019; Hidalgo 2021), which in our case is defined from the original binary neighborhood-amenity matrix  $M$  as  $M_{nn'} = M^T * M$ . The neighborhood-neighborhood similarity matrix used to construct the amenity complexity of locations is visualized by Figure 2D. To measure the complexity of amenity categories based on their geographic distribution across neighborhoods, we also create an amenity-amenity similarity matrix as  $M_{aa'} = M * M^T$ , visualized by Figure 2E. Applying the most common approach to measure complexity from geographical matrices, we take the second eigenvector of  $M_{nn'}$ , which is the leading correction to the equilibrium distribution and is the vector that is the best at dividing neighborhoods into groups based on the amenities that are present in them. Similarly, we take the second eigenvector of  $M_{aa'}$  to get the complexity values of amenity categories. This process to measure complexity is similar to dimension reduction techniques (singular value decomposition) that provide ways to explain the structure of a matrices (for an overview, see Hidalgo (2021)).

Figure 2F illustrates the relationship between amenity diversity and average amenity ubiquity of neighborhoods. Each point represents a neighborhood and is colored by the neighborhoods amenity complexity. Besides the expected negative correlation between amenity diversity and average amenity ubiquity (Hidalgo 2021), the amenity complexity of locations and the diversity of amenities at these locations shows remarkable variance. Figure 2G presents the relationship between the ubiquity of amenity categories and their average diversity. Each point stands for an amenity category and is colored by the complexity of the category. Overall, we observe that more complex amenity categories are non-ubiquitous and on average appear in more diverse areas. However, the figure indicates a clear outlier (bottom left corner), zoo, which is very non-ubiquitous and at the same time appears in less diverse neighborhoods. Figure 2H visualizes the connection between the amenity complexity of neighborhoods and the complexity of amenities present in these neighborhoods. The figure makes it clear that complex neighborhoods mostly

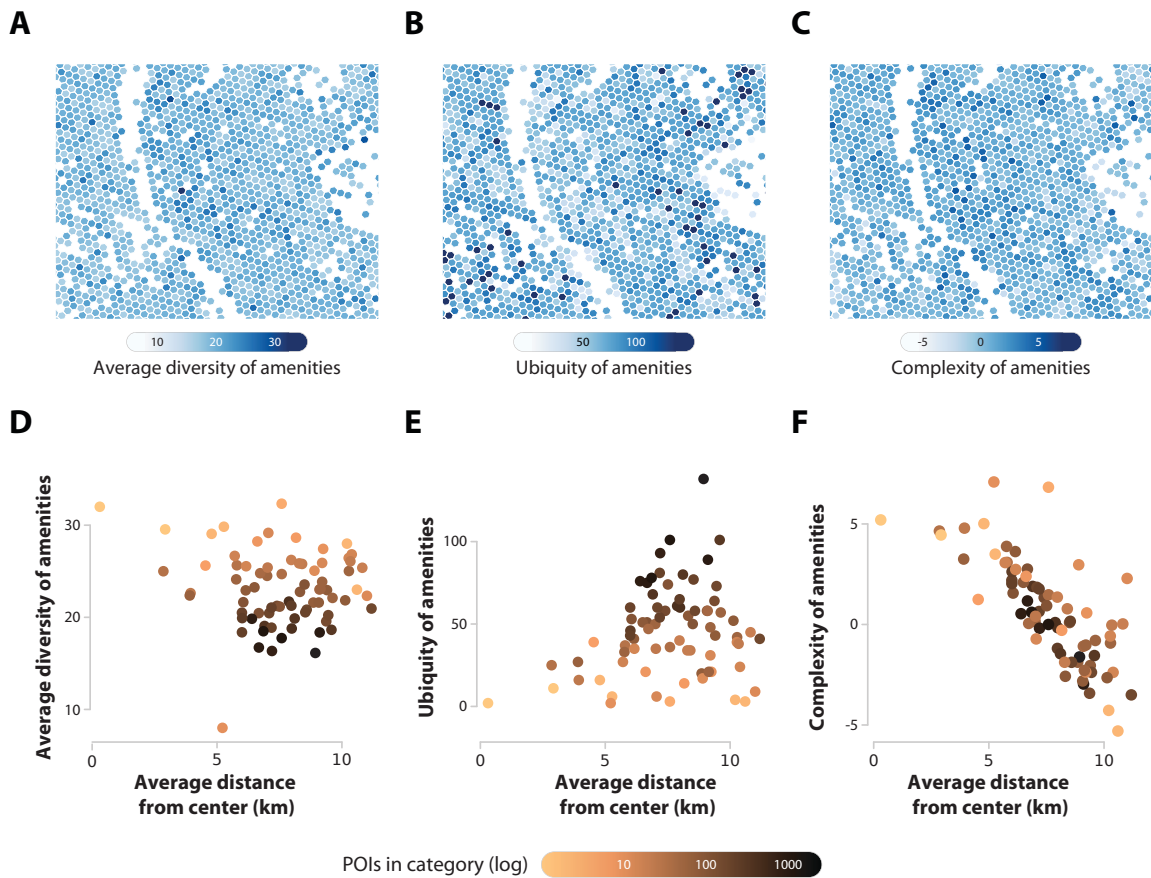
have complex amenities. These findings are in line with the patterns revealed by Mealy, Farmer, and Teytelboym (2019) for countries and exported products. Section S4 in the Supplementary information present the ranking of amenity complexity measures for amenity categories and for the urban neighborhoods of Budapest.



**Figure 3:** Components of amenity complexity in neighborhoods and their connections to the distance from the center of Budapest. (A) Map of Budapest colored by the amenity diversity in neighborhoods. (B) Map of Budapest colored by the average amenity ubiquity in neighborhoods. (C) Map of Budapest colored by the amenity complexity of neighborhoods. (D) Relationship between the distance from center and amenity diversity in neighborhoods. (E) Relationship between the distance from center and average amenity ubiquity in neighborhoods. (F) Relationship between the distance from center and amenity complexity of neighborhoods.

Figure 3A, B and C presents amenity diversity, average amenity ubiquity and amenity complexity of neighborhoods on the map of Budapest, while Figure 3D, E and F illustrates the connection of these factors to the distance from the city center defined as Deák Ferenc square (section S5 of the Supplementary information illustrates the choice of the city center). The figures suggest that average amenity ubiquity and amenity complexity in neighborhoods correlate with distance from the center, but the correlation is stronger for the complexity measure. Figure 4A, B and C illustrates

actual amenities on a zoomed in map of inner Budapest through size 10 H3 hexagons colored by the average diversity, ubiquity and complexity of the amenity category present at the location. At dense inner locations of the city, some hexagons contain amenities in multiple amenity categories. The identification of the dominant amenity category is detailed in section S6 of the Supplementary information. Figure 4D, E and F shows that more complex amenity categories have lower average distance from the center, while average diversity and ubiquity of amenities have no clear connection to distance from the center. The two figures suggests that there is a strong connection between geographic centrality and amenity complexity of urban locations in our case and we further discuss this feature in the following.

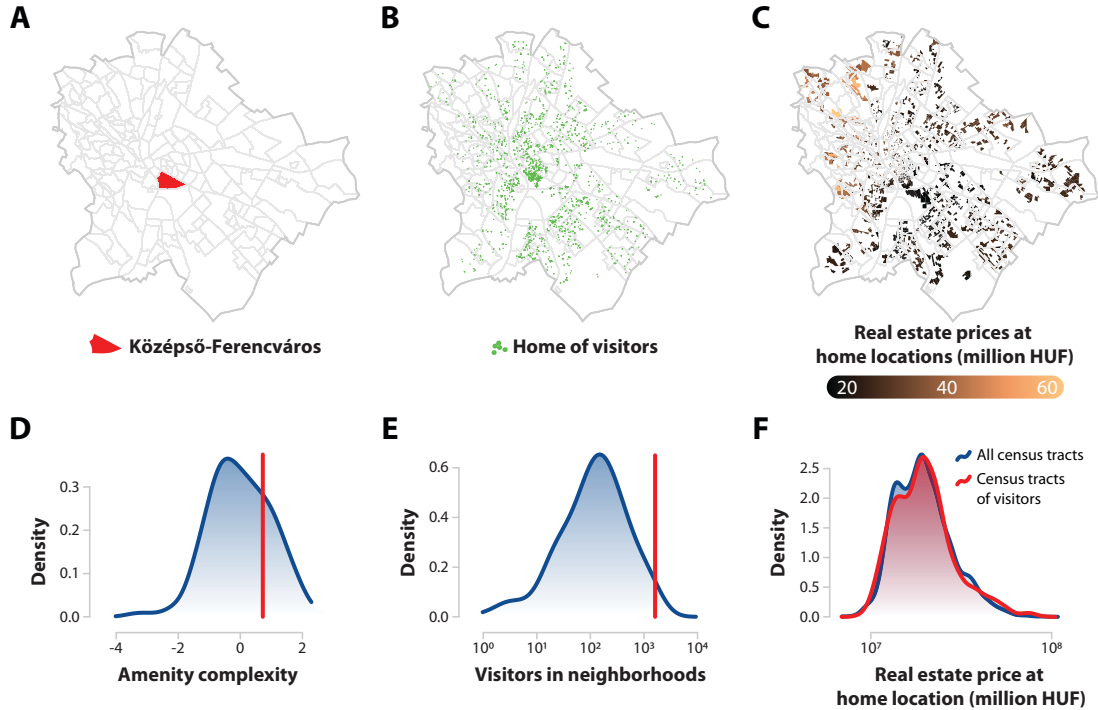


**Figure 4:** Components behind the complexity of amenity categories and their connection to the distance from the center of Budapest. (A) Amenities around Budapest colored by the average diversity of amenity category. The map is zoomed to the city center for illustration. (B) Amenities around Budapest colored by the ubiquity of amenity category. The map is zoomed to the city center for illustration. (C) Amenities around Budapest colored by their complexity. (D) Relationship between average distance of amenity categories from the center of Budapest and average diversity of amenity categories. (E) Relationship between average distance of amenity categories from center of Budapest and ubiquity of amenity categories. (F) Relationship between average distance of amenity categories from the center of Budapest and complexity of amenity categories.

## 4 Results

### 4.1 Diversity of visitors to complex urban neighborhoods

To illustrate the properties of urban locations that attract people of diverse socio-economic status, we combine amenity complexity measures at the neighborhood level with more granular visitation patterns from mobility data. Figure 5 presents our process to join different data sources through the example neighborhood of Középső-Ferencváros in Budapest.



**Figure 5:** Amenity complexity of a selected neighborhood and its visitors in January 2020. (A) Selected urban neighborhood of Középső-Ferencváros. (B) Home location of devices visiting Középső-Ferencváros. (C) Real estate prices at the home location of visitors. (D) Distribution of amenity complexity at the level of neighborhoods. The red vertical line indicates the complexity of the selected neighborhood of Középső-Ferencváros. (E) Distribution of observed visitors in neighborhoods. The red vertical line indicates the number of visitors in the selected neighborhood. (F) Distribution of real estate prices across all census tracts and at the home census tracts of visitors to the selected neighborhood.

Figure 5A presents the location of the selected neighborhood, while Figure 5B visualizes the home location of devices that visited any third places inside Középső-Ferencváros during the month of January 2020. We connect the home location of visitors to census tracts as Figure 5C indicates. This allows us to infer on the socio-economic status of visitors reflected by the real estate prices at the census tract of their home location. Figure 5D shows that the amenity mix at the selected neighborhood is relatively complex, while Figure 5E and F shows that Középső-Ferencváros is visited by more devices than most neighborhoods in January 2020 and its visitors come from diverse census tracts from all around Budapest.

	Coefficient of variation			
	(1)	(2)	(3)	(4)
Amenity complexity	0.126*** (0.036)			
Amenity diversity		-0.059 (0.084)		
Avg amenity ubiquity			0.019 (0.052)	
Distance from center (log)				-0.139*** (0.023)
Population (log)	-0.089*** (0.024)	-0.104*** (0.024)	-0.109*** (0.026)	-0.044* (0.024)
Nr visitors (log)	0.085*** (0.029)	0.111*** (0.029)	0.110*** (0.030)	0.053* (0.028)
Nr POIs (log)	0.004 (0.024)	0.020 (0.036)	0.006 (0.028)	-0.016 (0.023)
Constant	0.468*** (0.066)	0.534*** (0.072)	0.548*** (0.070)	0.590*** (0.057)
Observations	185	185	185	185
R <sup>2</sup>	0.197	0.145	0.143	0.288
Adjusted R <sup>2</sup>	0.179	0.126	0.124	0.272

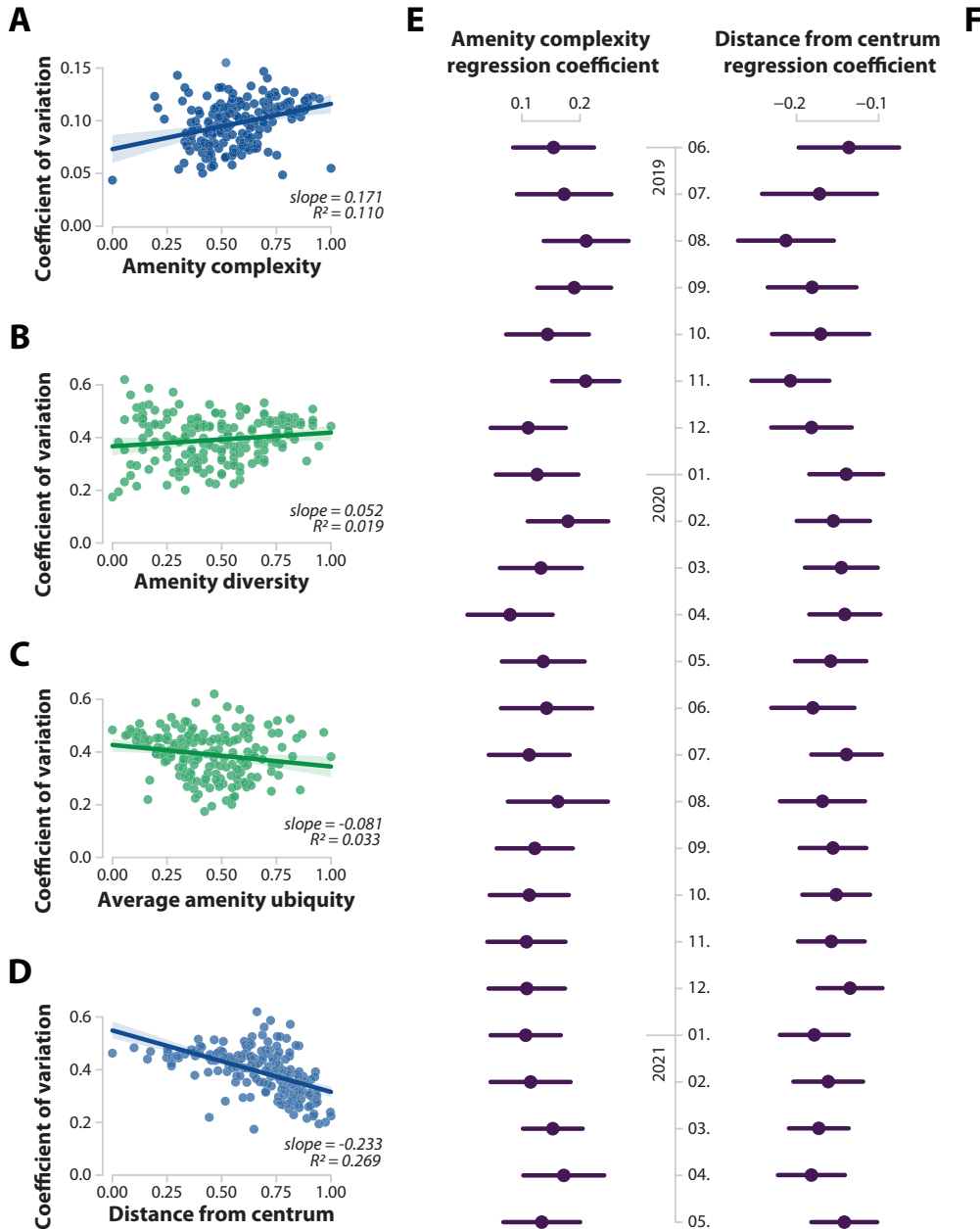
Note: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

**Table 1:** Controlled correlations between the socio-economic diversity of visitors and the amenity complexity of neighborhoods

To capture socio-economic mixing at urban locations, we measure the diversity of visitors in each neighborhood for every month by calculating the coefficient of variation (ratio of standard deviation to the mean) of the real estate prices at the home census tracts of visitors. We focus only on neighborhoods with at least 10 observed visitors in the focal month to get meaningful measures. Table 1 presents controlled correlations where we test the relationship between diversity of visitors and the amenity structure of neighborhoods by simple OLS regressions. Model (1) suggests that even after controlling for population, number of visitors and number of POIs in neighborhoods, amenity complexity is still positively correlated to the diversity of visitors. Interestingly, amenity diversity and average amenity ubiquity does not correlate with visitation patterns in our case (see Model (2) and (3) in Table 1). This suggests that amenity complexity captures the ability of mixing divers socio-economic groups than its elements: diversity and ubiquity of amenities.

Assessing the influence of central location to the diversity of visitors, we test the relationship between distance from center, measured as the overhead distance between the geometric center of neighborhoods from the center of Deák Ferenc square (unlike the center of gravity, this square can be considered as Budapest's central point in urban mobility). Model (4) suggests that distance from the center has a strong negative connection to the diversity of visitors. The  $R^2$  values of the different models suggest that distance from the center of Budapest has a higher explanatory power for the diversity of visitors than amenity complexity. Using the Gini coefficient or the Theil index to capture the diversity of visitors, we get the same results. Related model outputs can be found in section S7 of the Supplementary information.

Figure 6A-D presents the direct relationship between the four key explanatory variables and coefficient of variation. The univariate models are in line with the results of Table 1. Figure 6E presents the coefficient of amenity complexity estimated for the available 24 months by the same model presented in Table 1. The figure suggests that amenity complexity of neighborhoods have a positive and significant relationship with the diversity of visitors in neighborhoods in all 24 months. Figure 6F illustrates that stable negative and significant relationship of distance from center on the diversity of neighborhoods visitors. The results of Figure 6E and F are especially robust in case we consider the influence of COVID-19 related mobility restrictions in 2020. We run the same models presented in Table 1 on the visitation patterns of non-local users only. In this setting we only consider users living outside of the focal neighborhood, we observe similar patterns. Related models and figures can be found in section S8 of the Supplementary information.

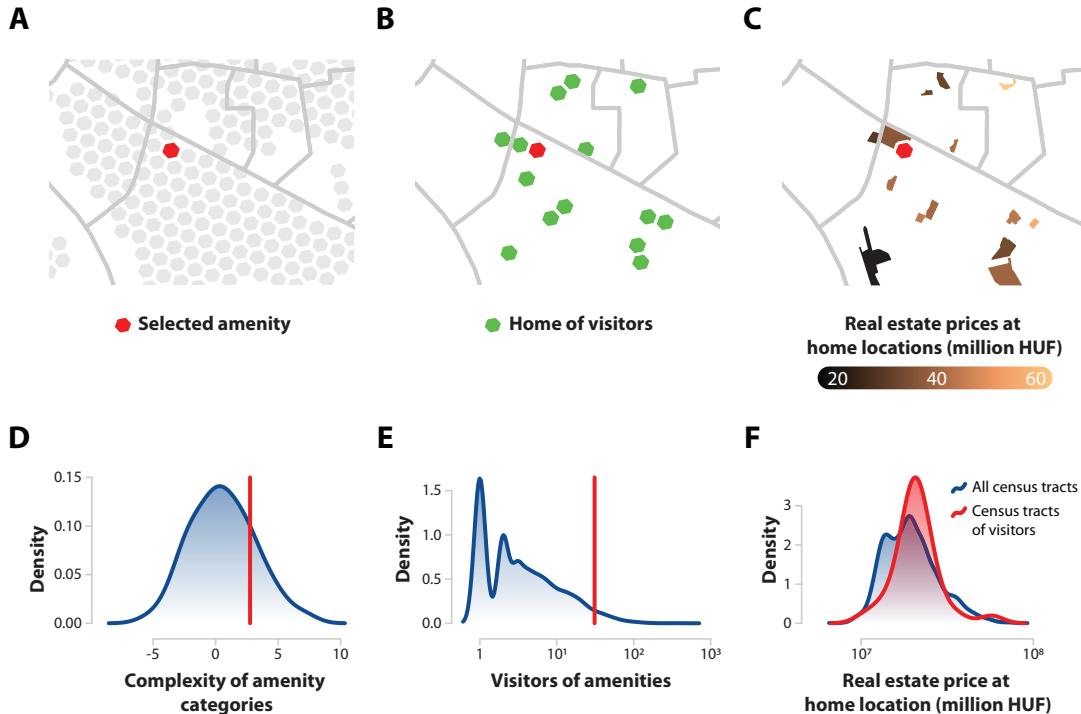


**Figure 6:** Univariate relationship and long time connection between our key variables and socio-economic diversity of visitors at the level of neighborhoods. **(A)** Relationship between amenity complexity in neighborhoods and socio-economic diversity of visitors measured by the coefficient of variation of the real estate prices at the home locations. **(B)** Relationship between amenity diversity and socio-economic diversity of visitors on a scatter plot with regression line. **(C)** Relationship between average amenity ubiquity and socio-economic diversity of visitors on a scatter plot with regression line. **(D)** Relationship between distance from city center and socio-economic diversity of visitors on a scatter plot with regression line. All scatter plots reflect only the observations from January 2020. **(E)** Regression coefficient of amenity complexity estimated for 24 months by the model presented in Table 1. **(F)** Regression coefficient of distance from center estimated for 24 months by the model presented in Table 1.



## 4.2 Diversity of visitors to complex amenities

To go beyond the level of urban neighborhoods, we combine amenity complexity measured at the amenity category level with visitations to actual amenities derived from our fine-grained mobility data. Figure 7 presents our process to join different data sources at the level of amenities through an example. The selected example is a bar in the neighborhood of Középső-Ferencváros, Budapest. Figure 7A shows the selected amenity and all the surrounding amenities on a zoomed-in map in size 10 H3 hexagons. Figure 7B illustrates the home location of visitor devices in 2020 January from the surrounding area in size 10 H3 hexagons. We connect the home location of visitors to census tracts to proxy the socio-economic status of visitors by real estate prices. Figure 7D shows that the selected example bar is from a relative complex amenity category and Figure 7E suggests that it is relate frequently visited in comparison to other observed amenities in 2020 January. Moreover, visitors from census tracts with middle and higher real estate prices are over-represented in 2020 January, as suggested by Figure 7F.



**Figure 7:** Connecting complexity of amenities to visitor diversity. (A) Selected bar on a map. Light red color hexagons indicate other nearby amenities. (B) Home location of the visitors of the selected amenity on a map. (C) Real estate prices at the census tracts of visitor home locations. (D) Distribution of amenity complexity values at the level of categories. The red vertical line indicates the complexity of the selected amenity in the category of bars. (E) Distribution of visitors to observed amenities in 2020 January, Budapest. The red vertical line indicates the number of visitors at the selected bar. (F) Distribution of real estate prices across all census tracts and at the home census tracts of visitors.

We measure the socio-economic diversity of visitors to each amenities for every month by cal-

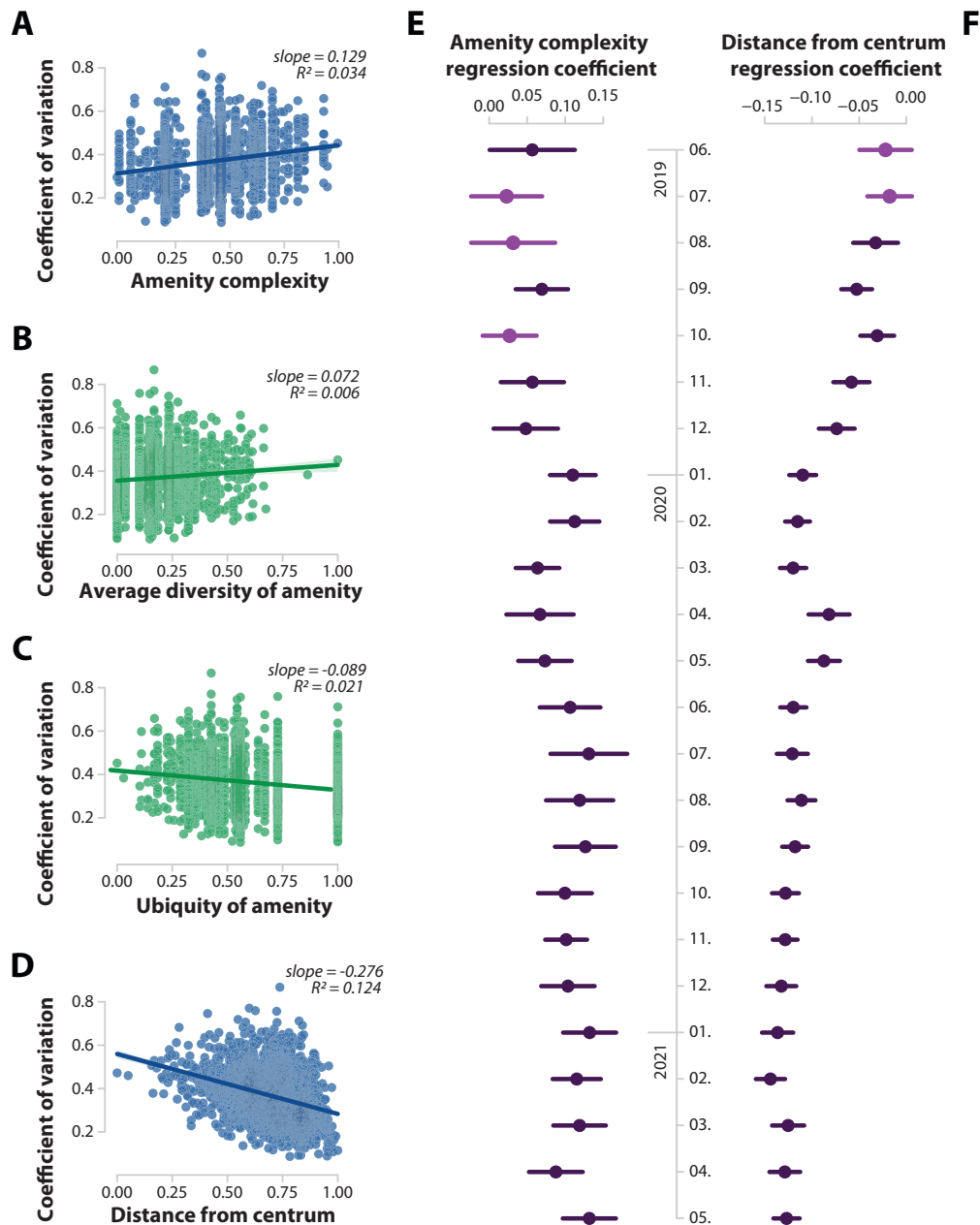
culating the coefficient of variation (ratio of standard deviation to the mean) of the real estate prices at the home census tracts of visitors. To do so, we focus only on amenities with at least 10 observed visitors in the focal month that helps us avoid meaningless values of the indicator. Table 2 presents simple OLS models to illustrate the relationship between the socio-economic diversity of visitors and components of amenity complexity at the level of amenities in 2020 January. Model (1) suggests that even after controlling for the total number of POIs in the respective amenity category around Budapest and the number of observed visitors in the focal month to the amenity, the complexity of the amenity category still has a positive and significant relationship with the socio-economic diversity of visitors. The negative and significant coefficient of the ubiquity of amenity category in Model (2) suggests that rare amenities are visited by more diverse groups of people. This result is consistent with the findings of Moro et al. (2021). The positive and slightly significant coefficient on the average diversity of amenities in Model (3) indicates that amenity categories that mostly appear in diverse neighborhoods attract visitors with different socio-economic status.

	Coefficient of variation			
	(1)	(2)	(3)	(4)
Complexity of amenity	0.110*** (0.015)			
Ubiquity of amenity		-0.112*** (0.018)		
Avg diversity of amenity			0.061** (0.026)	
Distance from center (log)				-0.110*** (0.007)
Nr POIs in category (log)	0.001 (0.008)	0.029*** (0.010)	0.001 (0.010)	0.00000 (0.007)
Nr visitors (log)	0.073*** (0.010)	0.074*** (0.010)	0.082*** (0.010)	0.039*** (0.010)
Constant	0.223*** (0.030)	0.241*** (0.030)	0.249*** (0.039)	0.382*** (0.027)
Observations	2,226	2,226	2,226	2,226
R <sup>2</sup>	0.056	0.051	0.037	0.124
Adjusted R <sup>2</sup>	0.055	0.049	0.035	0.123

Note: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

**Table 2:** Controlled correlations between the socio-economic diversity of visitors and the complexity of amenities

To measure the influence of central location on the socio-economic diversity of visitors to amenities, we test distance from center, measured as the overhead distance between the amenity



**Figure 8:** Univariate relationship and long time connection between our key variables and the socio-economic diversity of visitors at the level of amenities. **(A)** Relationship between complexity of amenities and the socio-economic diversity of visitors measured by the coefficient of variation of the real estate prices at the home locations. **(B)** Relationship between average amenity diversity and the socio-economic diversity of visitors on a scatter plot with regression line. **(C)** Relationship between amenity ubiquity and the socio-economic diversity of visitors on a scatter plot with regression line. **(D)** Relationship between distance from city center and the socio-economic diversity of visitors on a scatter plot with regression line. All scatter plots reflect only the observations from January 2020. **(E)** Regression coefficient of amenity complexity estimated for 24 months by the model presented in Table 1. Light colors indicate insignificant coefficients. **(F)** Regression coefficient of distance from center estimated for 24 months by the model presented in Table 1. Light colors indicate insignificant coefficients.

and the center Budapest (see section S5 of the Supplementary information for more details) in the same model setting. The negative and significant coefficient in Model (4) suggests that the further away amenities are from the center, the less diverse their visitors are in terms of socio-economic status. The  $R^2$  values of the different models indicate that distance from the center of Budapest has the highest explanatory power on the socio-economic diversity of visitors. As a robustness check, we use the Gini coefficient and the Theil index to capture the socio-economic diversity of visitors and we get the same results. Related regressions can be found in section S7 of the Supplementary information.

Figure 8A-D illustrates the direct relationship between the four key explanatory variables and the coefficient of variation at the level of amenities on scatter plots with univariate regression lines. The figures suggest similar results to Table 2. Figure 8E presents the coefficient of amenity complexity estimated for the available 24 months by the same model presented in Table 2. Lighter colors on Figure 8E indicate insignificant estimates. The figure suggests that the complexity of amenities has a positive and significant relationship with the socio-economic diversity of visitors with the exception of 3 months in 2019. Figure 8F illustrates a stable negative and significant relationship between distance from center and the socio-economic diversity of visitors to amenities with the exception of 2 months in 2019. Despite the COVID-19 related restriction in 2020, our estimates on Figure 8E and F proved to be stable.

## 5 Discussion

In this work we bring the ideas behind economic complexity to the urban problems of experienced segregation and social mixing. We measure amenity complexity by utilizing the spatial distribution of point of interests (POIs) inside a city. Then, we combine the information on the complexity of amenities with fine-grained mobility data to illustrate the relationship between amenity complex and visitor attraction. Focusing on the urban neighborhoods of Budapest, Hungary, we find that neighborhoods populated with a more complex amenity mix attract a bigger diversity of socio-economic groups. Applying the same logic to actual amenities inside Budapest, we also show that POIs of more complex amenity categories are visited by larger diversities of strata. However, components of amenity complexity, such as amenity diversity and amenity ubiquity, are only connected to the socio-economic diversity of visitors at the level of amenities and do not correlate with socio-economic mixing at the level of neighborhoods.

Considering centrality inside the monocentric city of Budapest, we find that distance of locations from the center of the city is strongly connected to their ability to attract socio-economically diverse visitors. While amenity complexity seems to correlate with visitation patterns, central location turned out to be a more influential factor for socio-economic mixing. Our empirical work illustrates that diversity and ubiquity of amenities do not have a strong connection to urban centrality, however, amenity complexity and distance from center are strongly correlated in case of the unequally distributed city of Budapest.

The general contribution of our paper is that we combine economic complexity concepts with urban mobility research. Constructing the measures of amenity complexity allows us to systematically test the contribution of certain amenity categories to socio-economic mixing in cities. Moreover, we contribute to the line of research on segregation patterns inside cities by illustrating in a direct fashion based on fine-grained mobility data that centrality of urban locations largely influence socio-economic mixing.

Our empirical work has several limitations, but offers promising future research directions. The study only focuses on the city of Budapest. Budapest is the only large city in Hungary and it clearly has a monocentric structure. Therefore, our findings are limited to this specific context and similar empirical works in cities with different size, geography and urban structure are necessary to assess the generality of our conclusions.

We construct our amenity complexity measures by mapping the distribution of POIs across amenity categories and urban neighborhoods in Budapest. We believe that the level of neighborhoods is the appropriate spatial scale to construct amenity complexity metrics for two reasons. First, the size of the applied spatial units can influence the nestedness of the location-amenity matrix used to construct complexity indexes. Co-occurrence of POIs in different amenity categories are less likely in case we consider smaller geographical areas. Neighborhoods are proved to be large enough to produce intuitive results. Second, neighborhoods are very important spatial units of urban life. They are argued to be the environment that can influence social capital accumulation and social mobility (Chetty et al. 2022; Chetty, Hendren, and Katz 2016). Moreover, they have clear administrative borders and people can identify with them, which makes the interpretation of amenity complexity results more appealing. However, alternative spatial scales are necessary to be tested in the future.

In our empirical exercise, we adopted the most commonly used economic complexity indicator to amenities and neighborhoods. However, several modifications have been suggested to improve economic complexity measurement (Tachella et al. 2012; Mealy, Farmer, and Teytelboym 2019) and the adoption of these methods to intra-urban scale is an apparent future research direction.

## Acknowledgements

The authors acknowledge the help of Orsolya Vásárhelyi and Luis Guillermo Natera Orozco with the original POI data collection. We wish to thank Tom Broekel and Frank Neffke for their comments and suggestions. We acknowledge the significant assistance of our illustrator Szabolcs Tóth-Zs. in finalizing our primary figures.

## References

- Abadie, A. 2006. "Poverty, political freedom, and the roots of terrorism." *American Economic Review* 96 (2): 50–56.
- Aslak, U., and L. Alessandretti. 2020. "Infostop: Scalable stop-location detection in multi-user mobility data."
- Athey, S., B. Ferguson, M. Gentzkow, and T. Schmidt. 2021. "Estimating experienced racial segregation in US cities using large-scale GPS data." *Proceedings of the National Academy of Sciences* 118 (46): e2026160118.
- Balland, P.-A., T. Broekel, D. Diodato, E. Giuliani, R. Hausmann, N. O'Clery, and D. Rigby. 2022. "The new paradigm of economic complexity." *Research Policy* 51 (3): 104450.
- Balland, P.-A., C. Jara-Figueroa, S. G. Petralia, M. P. A. Steijn, D. L. Rigby, and C. A. Hidalgo. 2020. "Complex economic activities concentrate in large cities." *Nature Human Behaviour* 4:248–254.
- Benton-Short, L., and J. R. Short. 2013. *Cities and Nature*. London: Routledge.
- Bokányi, E., S. Juhász, M. Karsai, and B. Lengyel. 2021. "Universal patterns of long-distance commuting and social assortativity in cities." *Scientific Reports* 11:20829.
- Cagney, K. A., E. York Cornwell, A. W. Goldman, and L. Cai. 2020. "Urban Mobility and Activity Space." *Annual Review of Sociology* 46 (1): 623–648.
- Chetty, R., N. Hendren, and L. F. Katz. 2016. "The Effects of Exposure to Better Neighborhoods on Children: New Evidence from the Moving to Opportunity Experiment." *American Economic Review* 106, no. 4 (April): 855–902.
- Chetty, R., M. O. Jackson, T. Kuchler, and et al. 2022. "Social capital I: measurement and associations with economic mobility." *Nature* 608:108–121.
- Chi, B., A. Dennett, T. Oléron-Evans, and R. Morphet. 2021. "Shedding new light on residential property price variation in England: A multi-scale exploration." *Environment and Planning B* 48 (7): 1895–1911.
- Christaller, W. 1933. *Die zentralen Orte in Süddeutschland*. Jena: Gustav Fischer.
- Dong, X., A. J. Morales, E. Jahani, E. Moro, B. Lepri, B. Bozkaya, C. Sarraute, Y. Bar-Yam, and A. Pentland. 2020. "Segregated interactions in urban and online space." *EPJ Data Science* 9:1–20.
- Engler, S., and D. Weisstanner. 2021. "The threat of social decline: income inequality and radical right support." *Journal of European Public Policy* 28 (2): 153–173.
- Florida, R. 2004. *Cities and the Creative Class*. New York: Routledge.

- Glaeser, E. 2012. *Triumph of the city: How our greatest invention makes us richer, smarter, greener, healthier, and happier*. New York: Penguin.
- Heroy, S., I. Loaiza, A. Pentland, and N. O'Clery. 2022. "Are neighbourhood amenities associated with more walking and less driving? Yes, but predominantly for the wealthy." *Environment and Planning B: Urban Analytics and City Science*, 1–25.
- Hidalgo, C. A. 2021. "Economic complexity theory and applications." *Nature Reviews Physics* 3:92–113.
- Hidalgo, C. A., E. Castañer, and A. Sevtsuk. 2020. "The amenity mix of urban neighborhoods." *Habitat International* 106:102205.
- Hidalgo, C. A., and R. Hausmann. 2009. "The building blocks of economic complexity." *Proceedings of the National Academy of Sciences* 106 (26): 10570–10575.
- Hilman, R. M., G. Iñiguez, and M. Karsai. 2022. "Socioeconomic biases in urban mixing patterns of US metropolitan areas." *EPJ data science* 11 (1): 32.
- Hungarian Central Statistical Office. 2022. *Regional Atlas*. (Online; accessed on 10 December 2022). [https://www.ksh.hu/regionalatlas%5C\\_administrative%5C\\_units](https://www.ksh.hu/regionalatlas%5C_administrative%5C_units).
- Jacobs, J. 1961. *The Death and Life of Great American Cities*. New York: Random House.
- Kaufmann, T., L. Radaelli, L. M. Bettencourt, and E. Shmueli. 2022. "Scaling of urban amenities: generative statistics and implications for urban planning." *EPJ Data Science* 11:50.
- Lösch, A. 1954. *The Economics of Location*. New Haven, CT: Yale University Press.
- Loughran, K., and J. R. Elliott. 2022. "Unequal retreats: how racial segregation shapes climate adaptation." *Housing Policy Debate* 32 (1): 171–189.
- Magalhães, L., M. Kuffer, N. Schwarz, and M. Haddad. 2023. "Bringing economic complexity to the intra-urban scale: The role of services in the urban economy of Belo Horizonte, Brazil." *Applied Geography* 150:102837. ISSN: 0143-6228.
- Mayer, S. E., and C. Jencks. 1989. "Growing up in Poor Neighborhoods: How Much Does it Matter?" *Science* 243 (4897): 1441–1445.
- Mealy, P., J. D. Farmer, and A. Teytelboym. 2019. "Interpreting economic complexity." *Science Advances* 5 (1): eaau1705.
- Moro, E., D. Calacci, X. Dong, and A. Pentland. 2021. "Mobility patterns are associated with experienced income segregation in large US cities." *Nature Communications* 12 (1): 1–10.
- Musterd, S. 2020. *Handbook of urban segregation*. Edward Elgar Publishing.

- Natera Orozco, L. G., D. Deritei, A. Vancso, and O. Vasarhelyi. 2020. "Quantifying Life Quality as Walkability on Urban Networks: The Case of Budapest." In *Complex Networks and Their Applications VIII*, edited by H. Cherifi, S. Gaito, J. F. Mendes, E. Moro, and L. M. Rocha, 905–918. Cham: Springer International Publishing.
- Noyman, A., R. Doorley, Z. Xiong, L. Alonso, A. Grignard, and K. Larson. 2019. "Reversed urbanism: Inferring urban performance through behavioral patterns in temporal telecom data." *Environment and Planning B* 46 (8): 1480–1498.
- Oldenburg, R. 1999. *The Great Good Place: Cafes, Coffee Shops, Bookstores, Bars, Hair Salons, and Other Hangouts at the Heart of a Community*. Boston, MA: Da Capo Press.
- Snijders, T. A., and R. J. Bosker. 2011. *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling. 2nd Edition*. London, UK: Sage Publishers.
- Tacchella, A., M. Cristelli, G. Caldarelli, A. Gabrielli, and L. Pietronero. 2012. "A New Metrics for Countries' Fitness and Products' Complexity." *Scientific Reports* 2 (723): 1–7.
- Torrats-Espinosa, G. 2021. "Using machine learning to estimate the effect of racial segregation on COVID-19 mortality in the United States." *Proceedings of the National Academy of Sciences* 118 (7): e2015577118.
- Uber Technologies, Inc. 2022. *H3*. (Online; accessed on 28 November 2022). <https://h3geo.org/>.
- Wang, Q., M. L. S. Nolan Edward Phillips, and R. J. Sampson. 2018. "Urban mobility and neighborhood isolation in America's 50 largest cities." *Proceedings of the National Academy of Sciences* 115 (30): 7735–7740.
- Zhong, C., M. Schlapfer, S. Müller Arisona, M. Batty, C. Ratti, and G. Schmitt. 2017. "Revealing centrality in the spatial structure of cities from human activity patterns." *Urban Studies* 54 (2): 437–455.



## Supplementary information

### S1 Mobility data preparation

Our GPS based mobility data is provided by a data aggregator company that collects and combines anonymous location data from users' smartphone applications. The sample we use for our analysis is an unbalanced panel of GPS pings from 5.2 million devices in Hungary between 2019 June and 2021 May. Our raw data consists of a device identifier, a time stamp, a latitude, and a longitude coordinate, where GPS pings occur. Pings are logged in case an application on a device requests location information. Sometimes this is the result of an active behavior such as using a navigation application or requesting local weather information. In other cases, pings could be the result of an application requesting information while running in the background. As a consequence, pings occur at irregular intervals. To identify visitation patterns such as home, work or third place visits to urban locations, we transform and filter our raw GPS trajectory data in several steps. We detail these steps in the following.

As the data contains some pings attributed to the same device that indicate unreasonable behavior, we start by iteratively removing all ping pairs that signals a movement over 300 kilometers per hour speed. Additionally, we discard all devices that have fewer than 20 pings remaining after this initial speed-based filter. These two steps reduce the total ping count from 3.18 billion to 3.13 billion and the total unique device count from 5.2 million to 1.88 million.

To focus on locations where devices stopped for some time, we run the Infostop stop detection algorithm (Aslak and Alessandretti [2020](#)). In short, it classifies GPS pings to trips or stays and clusters stationary points into stops in an effective way. We apply the algorithm with the following parameter set.  $r1$ , the maximum roaming distance allowed for two pings within the same stay is set to 370 meters.  $r2$ , the typical distance between two stays in the same destination is set to 140 meters.  $t_{min}$ , the minimum duration of a stay is set to 270 seconds, while  $t_{max}$ , the maximum time difference between two consecutive pings to be considered within the same stay is set to 7200 seconds. The minimum number of GPS pings required for stationary points is set to 2.

The parameters are calibrated using the Google location history data of 7 consenting individuals from Budapest, Hungary. We ran our stop detection algorithm on the sample trajectories with a wide range of parameters. We compared the results of the stop detection process to the personal, anecdotal experiences and to the semantic stop detection extracted from Google accounts. This experiment confirmed that the stop detection algorithm and the parameters produce a reasonable set of trips, stays, and destinations. Moreover, the subsequent home and work detection process built on it produced accurate results for our small sample.

Using the output of the stop detection process, we further filter the data to devices that have at least 2 unique destinations with over 4 different stays in each. This reduces the total stop count from 100 million to about 80 million, and the unique device count from 1.75 million to about 240.000. However, even in this final form, over 2.1 billion pings are used from the original 3.18 billion. Our empirical exercise in the end only uses information for each month from devices with

identified home, work and at least a single visited third place inside Budapest.

## S2 Socio-economic status from census tract level real estate prices

We infer on the socio-economic status of individuals living in Budapest by connecting their identified home location to residential real estate prices at the census tract level. Approximating socio-economic status through real estate prices has several benefits in comparison to the prevalent solution of using household income statistics of urban locations. First, real estate prices are by definition connected to places, while it is harder to connect income to locations. Second, real estate statistics in the census are comprehensive, while income information only reflects on the status of active employees.

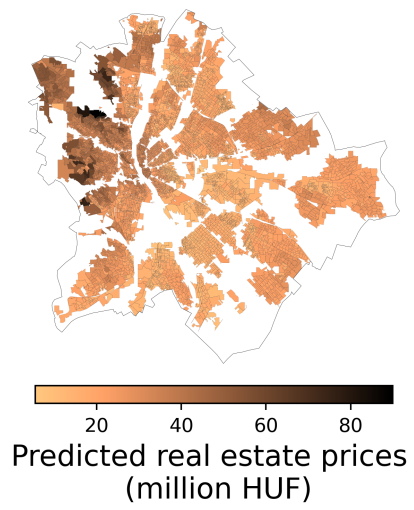
The Hungarian Central Statistical Office collects data on all residential real estate sales contracts and derives information on transaction prices for the entire country. As not every real estate is on the market and observed contracts sometimes suffer from missing information on the parameters of given properties, direct measurement on lower geographical level is difficult. By utilizing the fact that real estate prices tend to follow a strong multi-level hierarchy as location (and especially neighborhoods in Budapest) explains much of the price differences, we train a multi-level random slope regression model on the observed transaction prices (Chi et al. 2021; Snijders and Bosker 2011). To do so, we use real estate transaction contracts between 2013 and 2019. We create a pooled setting by correcting prices through the city level house price index published yearly by the National Bank of Hungary. Our model can be written as:

$$\begin{aligned}
 h_{i,j} &= \beta_{0,j} + \beta_{1,j}s_{i,j} + \varepsilon_{i,j} \\
 \beta_{0,j} &= \beta_0 + n_{0,j} \\
 \beta_{1,j} &= \beta_1 + n_{1,j}
 \end{aligned}
 \tag{7}$$

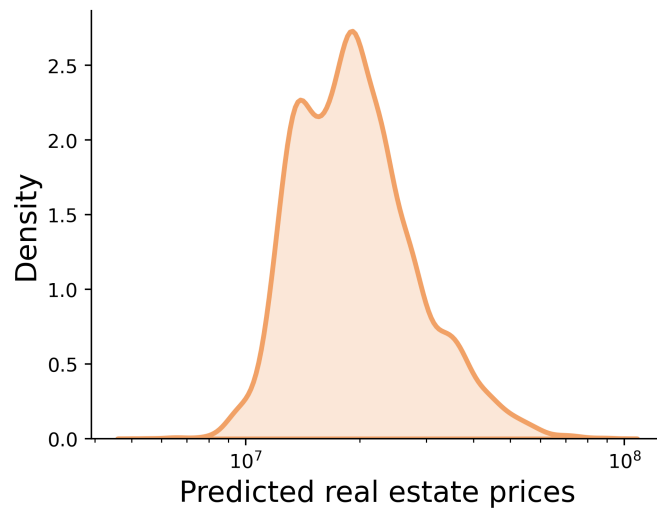
Here  $h_{i,j}$  is the logarithm of the individual price of real estate  $i$  in neighborhood  $j$ .  $\beta_{0,j}$  represents how much the estimated mean house price differs by neighborhoods. Estimated mean neighborhood prices are decomposed to  $\beta_0$ , the city level mean, and  $n_{0,j}$ , the neighborhood deviation from this value. To be able to capture differences within neighborhoods, we apply the individual level parameter  $s_{i,j}$  that refers to the size (floor area) of real estate  $i$  in neighborhood  $j$ . Since the effect of floor area can vary between neighborhoods, we train a random slope model.  $\beta_{1,j}$  represents the effect of floor area in neighborhood  $j$ .  $\beta_{1,j}$  is decomposed to a city level slope  $\beta_1$ , and the deviation of neighborhood slopes around this value  $n_{1,j}$ .

Utilizing this model, we predict prices for every real estate captured by the last Hungarian census in 2010. By taking the mean of the predicted real estate prices at the census tract level, we get a highly granular socio-economic status map for the entire city of Budapest. Figure 9A illustrates the predicted prices aggregated to the census tract level on the map of Budapest, while Figure 9B shows the distribution of predicted real estate prices.

**A**



**B**



**Figure 9:** Predicted real estate prices at the census tract level around Budapest (A) and the distribution of real estate prices aggregated to the census tract level in Budapest (B)

### S3 Urban neighborhoods of Budapest, Hungary

We use urban neighborhoods as geographic units to construct our amenity complexity measures. Budapest consists of 207 urban neighborhoods and they are in between districts and census tracts in terms of area and population. The unequal size distribution of urban neighborhoods are illustrated in Figure 10 and Figure 11. The correlation between population of urban neighborhoods and the number of census tracts per urban neighborhoods is strong as Pearson's R is 0.979. Further details about neighborhoods can be found at the website of Hungarian Central Statistical Office 2022.

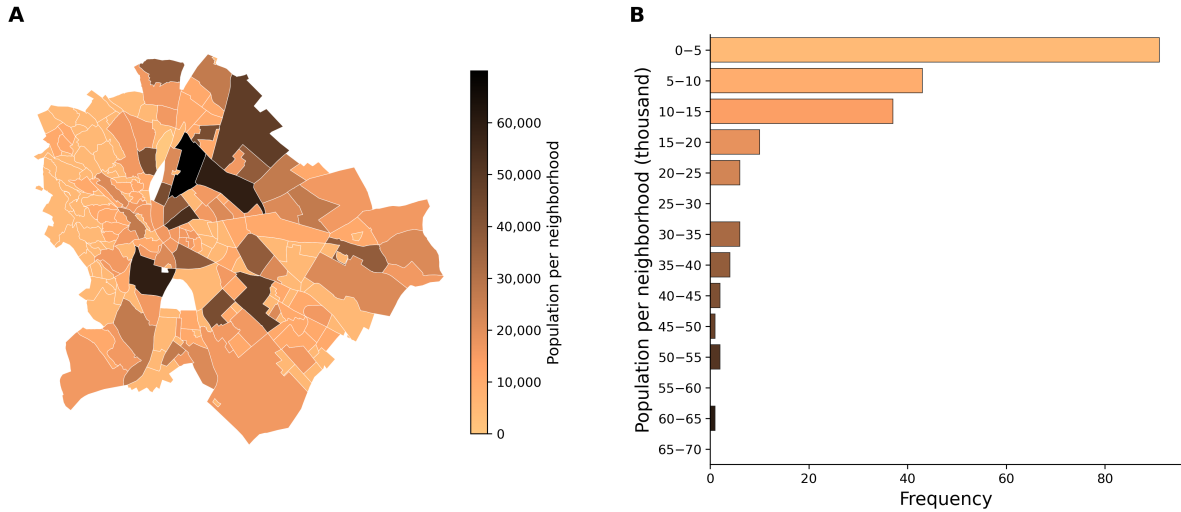


Figure 10: Population of urban neighborhoods based on the census of 2010 on the map of Budapest (A) and as a distribution plot (B).

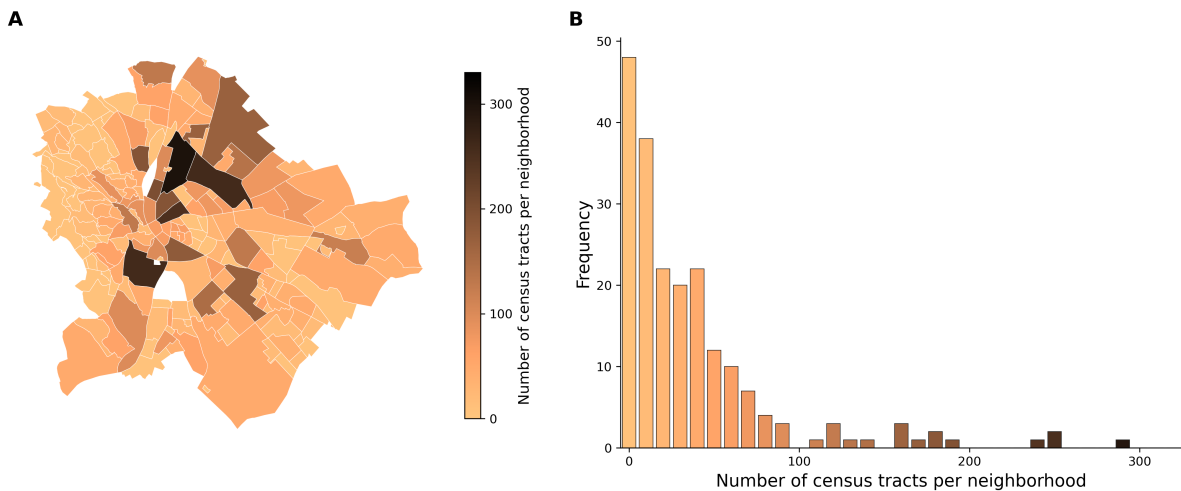


Figure 11: Number of census tracts per urban neighborhood based on the census of 2010 on the map of Budapest (A) and as a distribution plot (B).

## S4 Amenity complexity rankings

Ranking of all amenity categories (Figure 12) and all neighborhoods by their amenity complexity values (Figure 13).

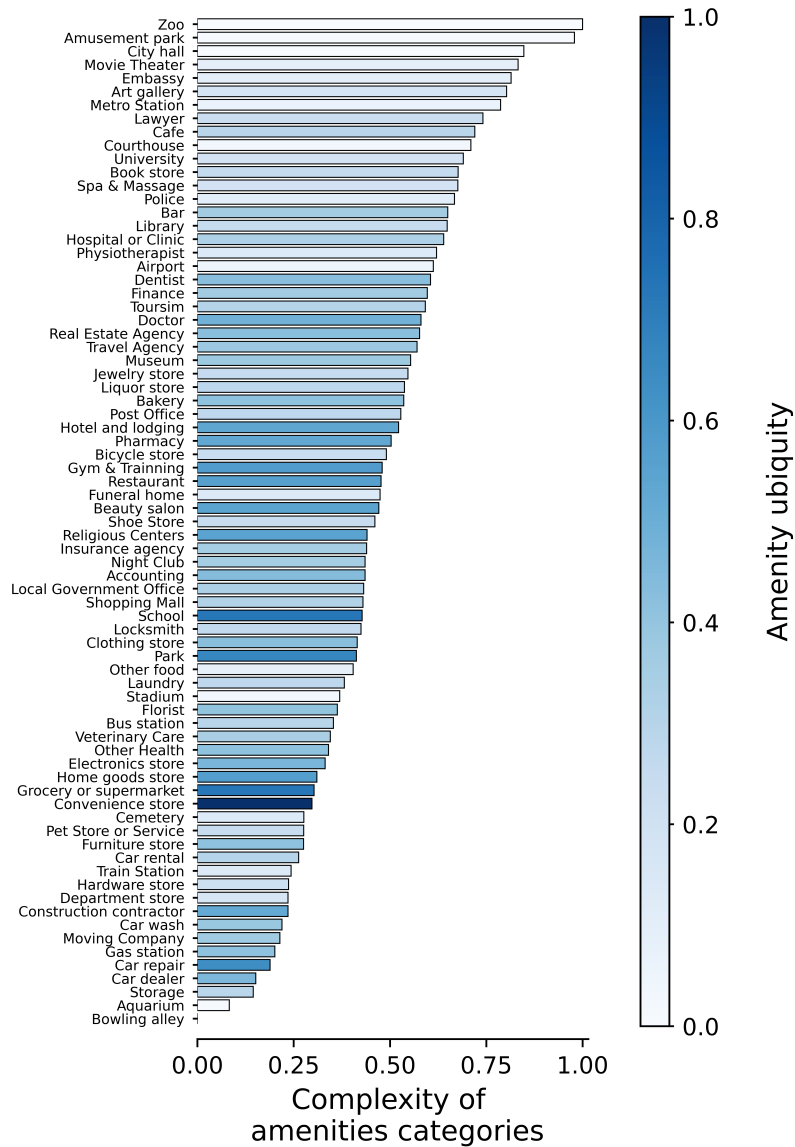
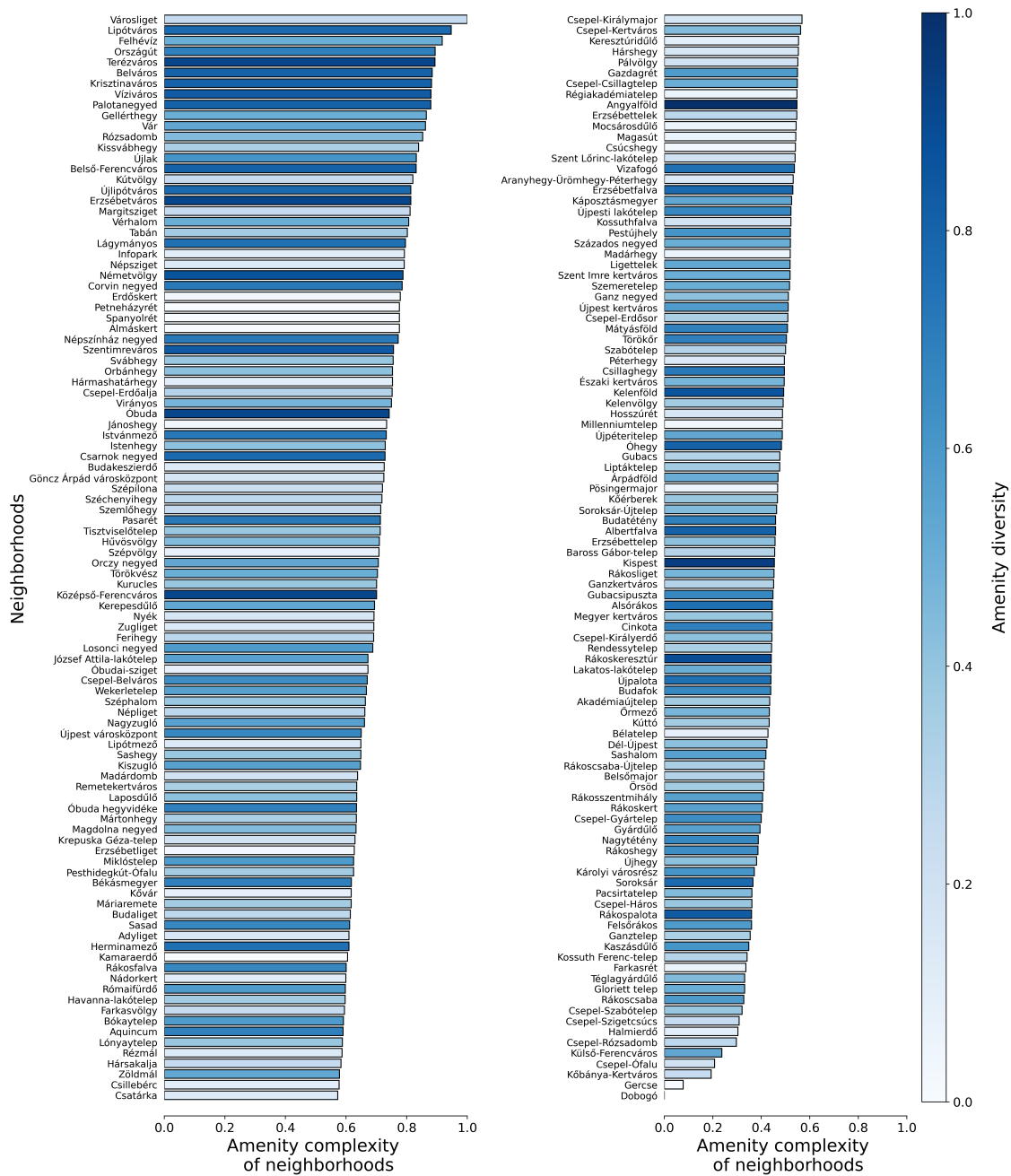


Figure 12: Amenity categories ranked by their amenity complexity value. Categories are colored by their ubiquity across neighborhoods. Complexity and ubiquity values are normalized to 0-1 scale for visualization purposes.



**Figure 13:** Neighborhoods of Budapest ranked by their amenity complexity value. Neighborhoods are colored by the diversity of their amenities. Complexity and diversity values are normalized to 0-1 scale for visualization purposes.

## S5 Center of Budapest, Hungary

We define the center of Budapest on a functional basis as Deák Ferenc tér. This square is the hotspot of public transportation in the heart of the inner city. Figure 14 shows the location of the selected central square. We tested several alternatives for city center such as Landmark zero sculpture that serves as the transportation center of the city or the Hungarian parliament building and our results remain the same. Figure 14 illustrates the overhead distance distribution for neighborhoods and amenities in our sample to Deák Ferenc tér.

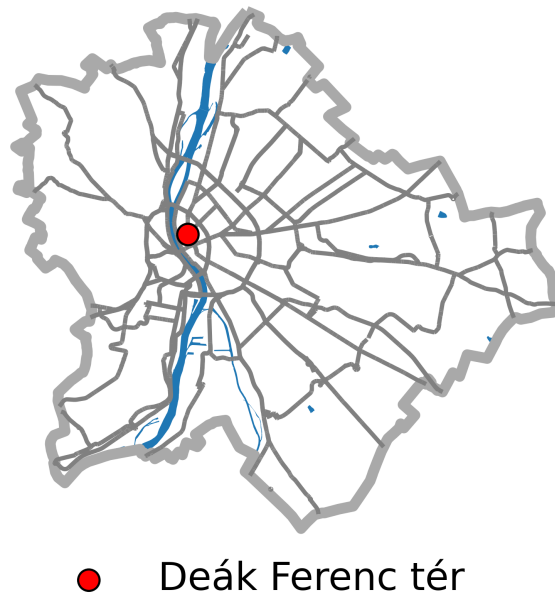


Figure 14: The location of Deák Ferenc tér, the center point.

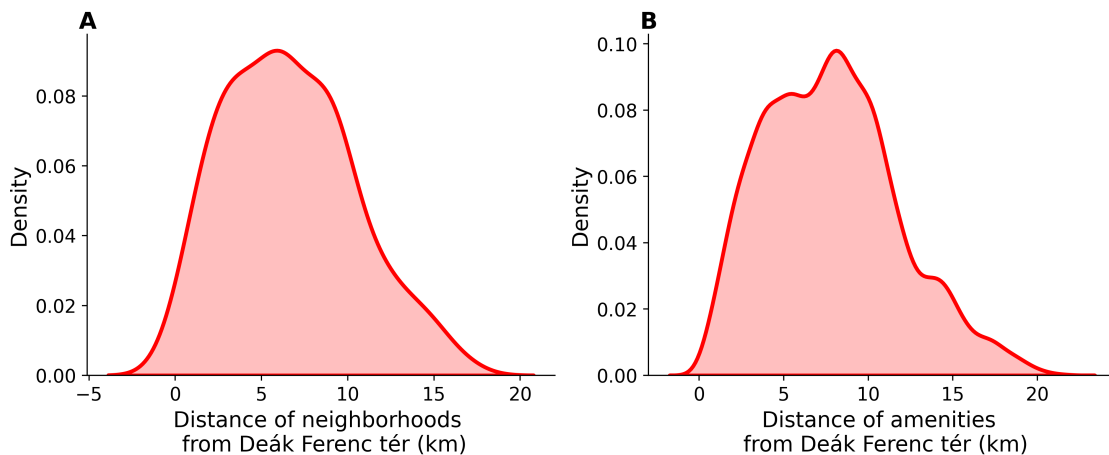
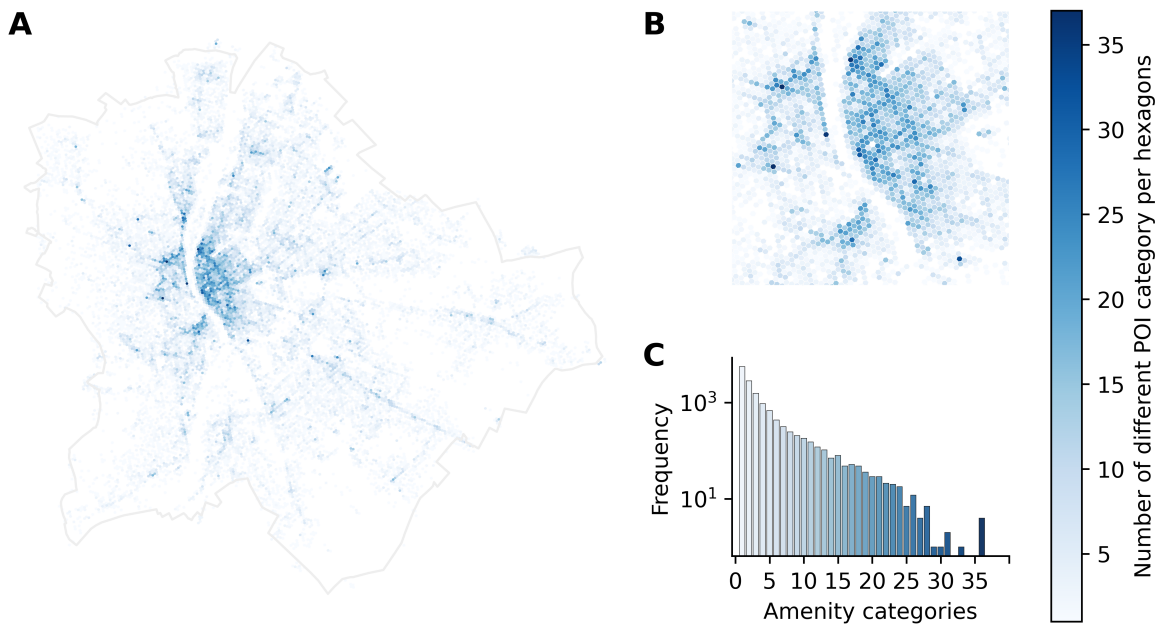


Figure 15: Distance distribution for neighborhoods (A) and amenities (B) to the center of Deák Ferenc tér.

## S6 Dominant amenity categories of H3 hexagons

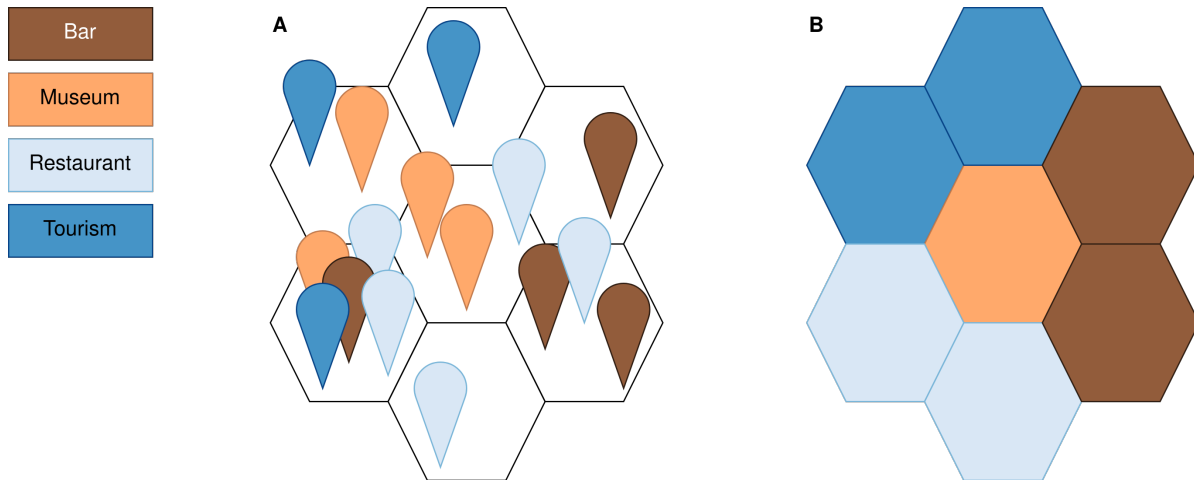
The second part of our empirical exercise (section 4.2 Diversity of visitors to complex amenities) connects visitors to actual amenities. This is done by mapping each point of interest (POI) from the Google Places API to a size 10 H3 hexagons (Uber Technologies, Inc. 2022). These hexagons are on average  $15.000\text{ m}^2$  area, which is close to the buffer area of a point with a 70 meter radius. As we use the amenity complexity values calculated for amenity categories (see Equation (6) in the main text) to explain the diversity of visitors to size 10 H3 hexagons, we need to assign a single amenity category to each hexagon with POIs. Figure 16 illustrates that most hexagons only have amenities in a single amenity category, however, hexagons at dense, central locations often contain multiple POIs from different categories.



**Figure 16:** Number of different amenity categories per size 10 H3 hexagons on the map of Budapest (A) and on the map of the inner city (B). Distribution of different amenity categories per hexagons (C). The colorbar applies for all subplots.

We assign a dominant amenity category for each hexagon based on the local frequency of POIs in amenity categories. Figure 17 visualizes our dominant category selection process. 35.15% of the hexagons have ambiguous amenity category dominance and in these cases we simply choose the first category listed. Table 3 illustrates in comparison to Table 2 of the main text that our key findings are the same in case we focus only on amenities in H3 hexagons with a single amenity category or in case we only consider amenities in H3 hexagons with an unambiguously dominant amenity category (Table 4).





**Figure 17:** Illustration of the dominant amenity category selection. Different POIs over the hexagons (A) and the determined hexagon category (b).

	Coefficient of variation			
	(1)	(2)	(3)	(4)
Complexity of amenity	0.163*** (0.061)			
Ubiquity of amenity		-0.239** (0.094)		
Avg diversity of amenity			0.115 (0.159)	
Distance from center (log)				-0.170*** (0.050)
Nr POIs in category (log)	-0.037 (0.031)	0.061 (0.047)	0.007 (0.058)	-0.000 (0.030)
Nr visitors (log)	0.076 (0.070)	0.081 (0.070)	0.059 (0.073)	0.100 (0.069)
Constant	0.296** (0.133)	0.168 (0.149)	0.208 (0.225)	0.403*** (0.131)
Observations	107	107	107	107
R <sup>2</sup>	0.079	0.074	0.021	0.117
Adjusted R <sup>2</sup>	0.052	0.047	-0.007	0.091
Note:	*p<0.1; **p<0.05; ***p<0.01			

**Table 3:** Controlled correlations between the diversity of visitors and the complexity of amenities in case we only consider H3 hexagons with a single amenity category

	Coefficient of variation			
	(1)	(2)	(3)	(4)
Complexity of amenity	0.115*** (0.018)			
Ubiquity of amenity		-0.126*** (0.022)		
Avg diversity of amenity			0.083*** (0.032)	
Distance from center (log)				-0.110*** (0.008)
Nr POIs in category (log)	-0.002 (0.009)	0.027** (0.012)	0.001 (0.013)	-0.006 (0.008)
Nr visitors (log)	0.073*** (0.012)	0.072*** (0.012)	0.083*** (0.012)	0.038*** (0.012)
Constant	0.233*** (0.036)	0.258*** (0.035)	0.243*** (0.048)	0.404*** (0.031)
Observations	1,477	1,477	1,477	1,477
R <sup>2</sup>	0.068	0.063	0.045	0.140
Adjusted R <sup>2</sup>	0.066	0.061	0.044	0.138
<i>Note:</i>	*p<0.1; **p<0.05; ***p<0.01			

**Table 4:** Controlled correlations between the diversity of visitors and the complexity of amenities in case we only consider H3 hexagons with an unambiguously dominant amenity category

## S7 Alternative measures on the diversity of visitors to neighborhoods and amenities

Table 5 supports the findings presented in Table 1 of the main text. It uses the same model setting to illustrate the relationship between the diversity of visitors to neighborhoods, amenity complexity and distance from the center of the city, but applies different measurements of the dependent variable. Besides the coefficient of variation, the Gini index and the Theil index are used to measure the diversity of visitors and all model version suggest similar connections. Table 6 supplements the findings of Table 2 in the main text in a similar fashion at the level of amenities.

	Coeff var (1)	Gini (2)	Theil (3)	Coeff var (4)	Gini (5)	Theil (6)
Amenity complexity	0.126*** (0.036)	0.106*** (0.018)	0.059*** (0.012)			
Distance from center (log)				-0.139*** (0.023)	-0.084*** (0.012)	-0.051*** (0.008)
Population (log)	-0.089*** (0.024)	-0.034*** (0.012)	-0.023*** (0.008)	-0.044* (0.024)	-0.010 (0.012)	-0.009 (0.008)
Nr visitors (log)	0.085*** (0.029)	0.029** (0.014)	0.021** (0.010)	0.053* (0.028)	0.015 (0.014)	0.011 (0.009)
Nr POIs (log)	0.004 (0.024)	-0.002 (0.012)	-0.003 (0.008)	-0.016 (0.023)	-0.014 (0.011)	-0.011 (0.008)
Constant	0.468*** (0.066)	0.212*** (0.033)	0.090*** (0.022)	0.590*** (0.057)	0.307*** (0.029)	0.144*** (0.019)
Observations	185	185	185	185	185	185
R <sup>2</sup>	0.197	0.241	0.200	0.288	0.300	0.269
Adjusted R <sup>2</sup>	0.179	0.224	0.182	0.272	0.284	0.253

Note:

\*p<0.1; \*\*p<0.05; \*\*\*p<0.01

**Table 5:** Controlled correlations between different measures on the diversity of visitors and the amenity complexity of neighborhoods

	Coeff var	Gini	Theil	Coeff var	Gini	Theil
	(1)	(2)	(3)	(4)	(5)	(6)
Complexity of amenity	0.110*** (0.015)	0.055*** (0.007)	0.035*** (0.005)			
Distance from center (log)				-0.110*** (0.007)	-0.056*** (0.003)	-0.033*** (0.002)
Nr POIs in category (log)	0.001 (0.008)	-0.003 (0.003)	0.00001 (0.002)	0.00000 (0.007)	-0.003 (0.003)	-0.001 (0.002)
Nr visitors (log)	0.073*** (0.010)	0.032*** (0.004)	0.016*** (0.003)	0.039*** (0.010)	0.014*** (0.004)	0.006* (0.003)
Constant	0.223*** (0.030)	0.132*** (0.014)	0.031*** (0.010)	0.382*** (0.027)	0.212*** (0.012)	0.081*** (0.009)
Observations	2,226	2,226	2,226	2,226	2,226	2,226
R <sup>2</sup>	0.056	0.061	0.042	0.124	0.147	0.100
Adjusted R <sup>2</sup>	0.055	0.060	0.040	0.123	0.146	0.099

Note:

\*p<0.1; \*\*p<0.05; \*\*\*p<0.01

**Table 6:** Controlled correlations between different measures on the diversity of visitors and the complexity of amenities

## S8 Diversity of non-local visitors to neighborhoods

Table 7 supports the findings presented in Table 1 of the main text, but the models are based only on non-local visits. This means that we excluded all third places visits inside the home neighborhoods of devices. Results are very similar to our main models.

	Coefficient of variation (non-local visitors only)			
	(1)	(2)	(3)	(4)
Amenity complexity	0.126*** (0.036)			
Amenity diversity		-0.079 (0.084)		
Average amenity ubiquity			0.057 (0.052)	
Distance from center (log)				-0.130*** (0.023)
Population (log)	-0.079*** (0.024)	-0.093*** (0.024)	-0.105*** (0.025)	-0.037 (0.025)
Nr visitors (log)	0.068** (0.029)	0.095*** (0.029)	0.098*** (0.030)	0.040 (0.028)
Nr POIs (log)	0.015 (0.024)	0.037 (0.036)	0.027 (0.028)	-0.003 (0.023)
Constant	0.448*** (0.066)	0.505*** (0.072)	0.504*** (0.070)	0.567*** (0.057)
Observations	185	185	185	185
R <sup>2</sup>	0.184	0.134	0.135	0.258
Adjusted R <sup>2</sup>	0.166	0.114	0.116	0.242
Note:	*p<0.1; **p<0.05; ***p<0.01			

**Table 7:** Controlled correlations between the diversity of non-local visitors and the amenity complexity of neighborhoods