










# Amenity complexity and urban locations of socio-economic mixing

Sándor Juhász<sup>1,2,3\*</sup> , Gergő Pintér<sup>2</sup> , Ádám J. Kovács<sup>3</sup> , Endre Borza<sup>3,4</sup> , Gergely Mónus<sup>2</sup> ,  
László Lőrincz<sup>2,3,5</sup>  and Balázs Lengyel<sup>2,3,5</sup> 

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

<sup>1</sup>Complexity Science Hub Vienna,  
Josefstaedter Strasse 39, 1080,  
Vienna, Austria

<sup>2</sup>Laboratory for Networks,  
Technology and Innovation,  
Corvinus University of Budapest,  
Fővám tér 8, 1093, Budapest,  
Hungary

Full list of author information is  
available at the end of the article

## 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 sophistication of amenity supply at urban locations. We propose that neighborhood complexity and amenity complexity are connected to the ability of locations to attract diverse visitors from various socio-economic backgrounds across the city. We construct the measures of amenity complexity based on the local portfolio of diverse and non-ubiquitous amenities in Budapest, Hungary. Socio-economic mixing at visited third places is investigated by tracing the daily mobility of individuals and by characterizing their status by the real-estate price of their home locations. Results suggest that measures of ubiquity and diversity of amenities do not, but neighborhood complexity and amenity complexity are correlated with the urban centrality of locations. Urban centrality is a strong predictor of socio-economic mixing, but both neighborhood complexity and amenity complexity add further explanatory power to our models. 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; Segregation

## 1 Introduction

Diversity is the key ingredient of successful and resilient cities [1]. The spatially concentrated interaction of people from various social and economic background create environments that foster creativity [2], support inclusion [3] and in general, make cities vivid and prosperous [4]. 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 [5]. This phenomenon limits social mobility for many [6] and induced inequalities can expose segregated groups to health or climate crises [7, 8] and can imply radicalization and populism [9, 10].

Recent studies leverage GPS mobility data to study socio-economic segregation and mixing patterns in visited urban locations [11]. This growing literature frequently reports

© The Author(s) 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

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 [12–15]. 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 [16, 17].

In this study we characterize urban locations that foster socio-economic mixing and lower experienced segregation by attracting people from diverse strata. To do so, we emphasize two aspects of urban locations that can influence observed socio-economic mixing: their amenity portfolio and geographical centrality in the city.

The type of amenities available at a location determine its purpose and function and therefore is related to experienced segregation. Noyman et al. [18] illustrates through individual GPS trajectories that urban locations offering entertainment amenities, services or natural water features are visited by a more diverse set of people. Athey et al. [16] describes that individuals can experience relatively low experienced 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. [17] 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 integrative, while places that primarily serve local communities, such as grocery stores or places of worship are generally more segregated by income. Yet, urban locations can be hardly described by single amenity types; instead, they typically host more types of amenities. 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.

Specialized amenities 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 [19] and Lösch [20] explains the hierarchy of cities and towns through their size and the range of functions that they provide and has been used to study the functions of locations within cities too (see for example [21]). Higher-order centers attract population from a larger area, because they not only share most of the functions of lower order centers, but also host some more specialized functions too. Building on the central place theory, Zhong et al. [22] 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. [18] shows that urban locations with higher centrality in urban road networks attract more diverse visitors. On the contrary, Moro et al. [17] 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 position of locations, which has not been focused on so far.

Here, we aim to extend the above literature by investigating how the available amenities 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 [23] to

quantify the sophistication of local amenity supply. We argue that more complex neighborhoods and amenities attract visitors of diverse socio-economic status from across the city.

The concept of economic complexity is originally developed by Hidalgo and Hausmann [24] 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 [23]. 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 [25, 26]. Here we adopt the measurement technique to uncover the complexity of neighborhoods and amenities. We propose that a neighborhood has a complex amenity mix in case it offers diverse set of amenities of those types that other locations are not specialized in. On the contrary, complex amenities are those that only few neighborhoods are specialized into and are co-located with diverse sets of similarly non-ubiquitous amenities. Unlike in the original framework of economic complexity that captures the knowledge and capabilities required to achieve economic outputs [24], our approach does not address the productive and operative knowledge that a given location has accumulated [23]. Instead, we measure the sophistication of local amenity supply that can serve a wide range of unique needs.

The rationale to apply neighborhood and amenity complexity to understand mixing of people is based on two reasons. First, diverse amenity mixes in neighborhoods can attract people with diverse demands. Second, locations with non-ubiquitous amenities can attract people from diverse neighborhoods, as the particular service is hard to find elsewhere. Consequently, complex amenities combining diversity and non-ubiquity are also expected to attract diverse visitors. Therefore, our hypothesis is that the diverse mix of non-ubiquitous amenities can create an inclusive, multipurpose 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 rarely been unveiled, diverse amenities are argued to concentrate in and attract people to central places of cities [22]. To better understand the connection between neighborhood and amenity complexity, urban centrality and socio-economic mixing, we measure their correlation with the socio-economic diversity of visitors.

We test this argument in Budapest, the capital of Hungary by combining point of interests (POI) data collected from the Google Places API and individual mobility trajectories collected by a GPS aggregator company. Building on the work of Hidalgo et al. [27], we construct the indicators of neighborhood and amenity complexity by utilizing the geographic distribution of POIs in neighborhoods. We identify home, work and third place visits in daily mobility trajectories for 24 months by clustering the geolocated pings of devices in geographical space and over time [28]. We combine the information of predicted home locations with real estate prices at the census tract level. This allows us to investigate third place visits and to infer the socio-economic diversity of visitors in each urban neighborhood and in each actual amenity.

Our results illustrate that, in the monocentric city of Budapest, specialization and diversity of amenities do not, but neighborhood complexity and amenity complexity are correlated with urban centrality. We find that socio-economic mixing increases as neighborhood complexity grows, and that amenity complexity is also associated with lower levels of experienced segregation. These suggest that the combination of mobility data with eco-

conomic complexity thinking can provide new insights to the research of urban segregation and mixing.

## 2 Tracing mobility inside cities

Urban mobility of individuals are studied by using raw GPS data from a data aggregator company. We can 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 Sect. 1 of Additional file 1.

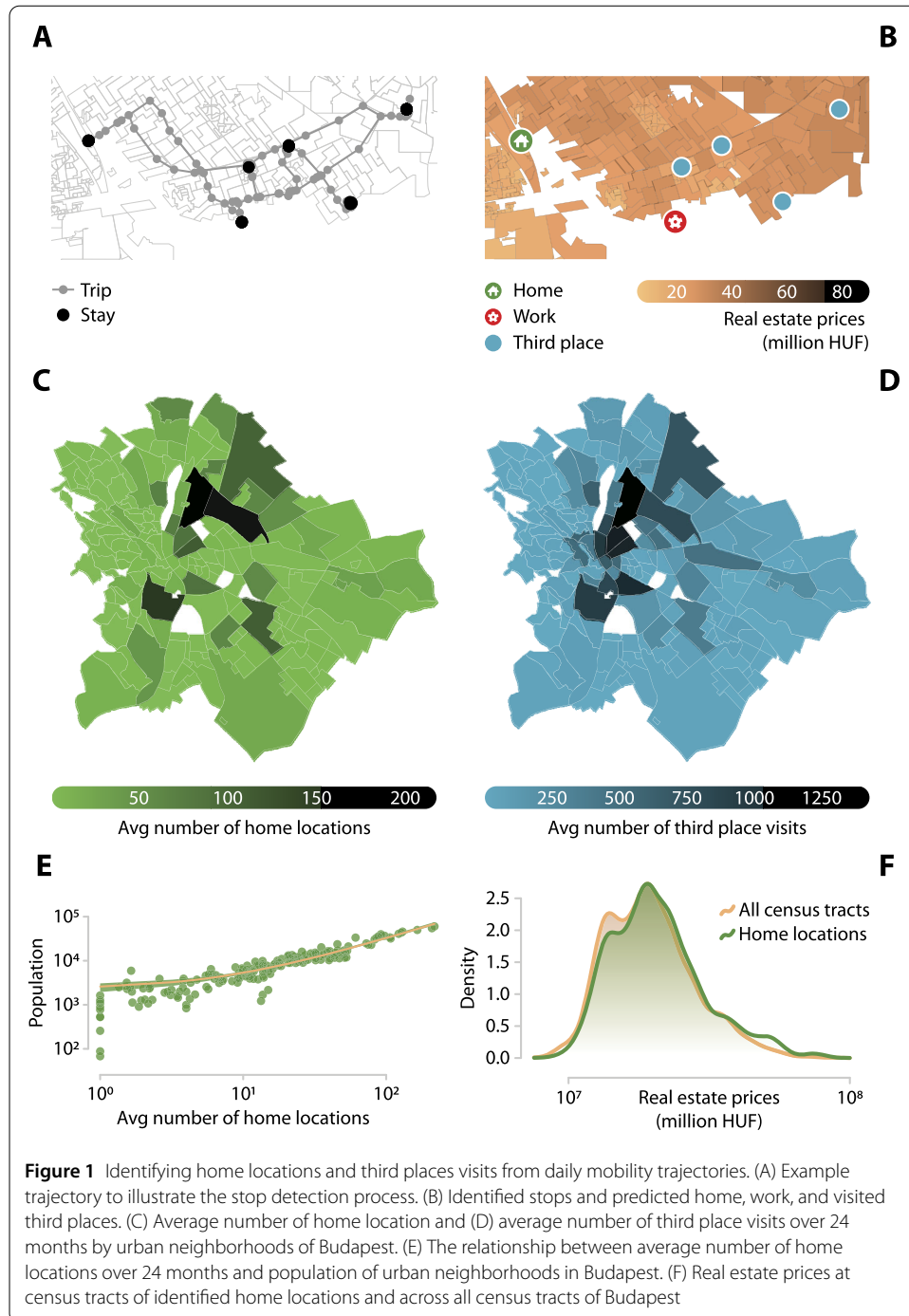
We process raw trajectories of individuals by applying the Infostop algorithm [29]. It enables us to detect the stationary points of individual movements and cluster GPS pings around stop locations. Figure 1A-B illustrates the raw data and the outcome of stop detection through an example device. The 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 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. The potential home is where the device spends the most time between 8:00 pm and 8:00 am on weekdays or at any time during the weekend, and the cumulative time spent at the place exceeds 8 hours per week. Places where devices spend the most time between 9:00 am and 5:00 pm on weekdays (at least 3 hours a week) are considered as potential workplaces.

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 meters 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 Fig. 1D illustrates the average number of third place visits over the 24-month period aggregated to the level of urban neighborhoods.

Home locations and third places are joined to other data sources with Uber's Hexagonal Hierarchical Spatial Index (H3) [30]. 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 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. Therefore, 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 2 in Additional file 1 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 attractiveness of urban locations, 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 limi-

tations in terms of timescale 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 [27, 31, 32].

We collected GPS coordinates 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 [33]. Neighborhoods are in between districts and census tracts in the spatial hierarchy, which makes them a suitable scale for our analysis [34]. 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 neighborhoods of Budapest can be found in Sect. 3 of Additional file 1.

Every neighborhood in Budapest with at least 2 amenity categories that have minimum 2 POIs are considered in the analysis. Alternative specifications and their influence on amenity complexity measurement can be found in Sect. 4 of Additional file 1. Figure 2A presents the resulted 75 amenity categories and the number of POIs across the focal 200 neighborhoods. 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 parks, Bowling alleys and City halls. 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 neighborhoods, we adopt the economic complexity index (ECI) and the product complexity index (PCI) [24]. The ECI is successfully used to describe the economic development of countries and regions [23] and its approach is adoptable to amenities and urban neighborhoods. We measure neighborhood complexity and amenity complexity the following way. We normalize the matrix of Fig. 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):

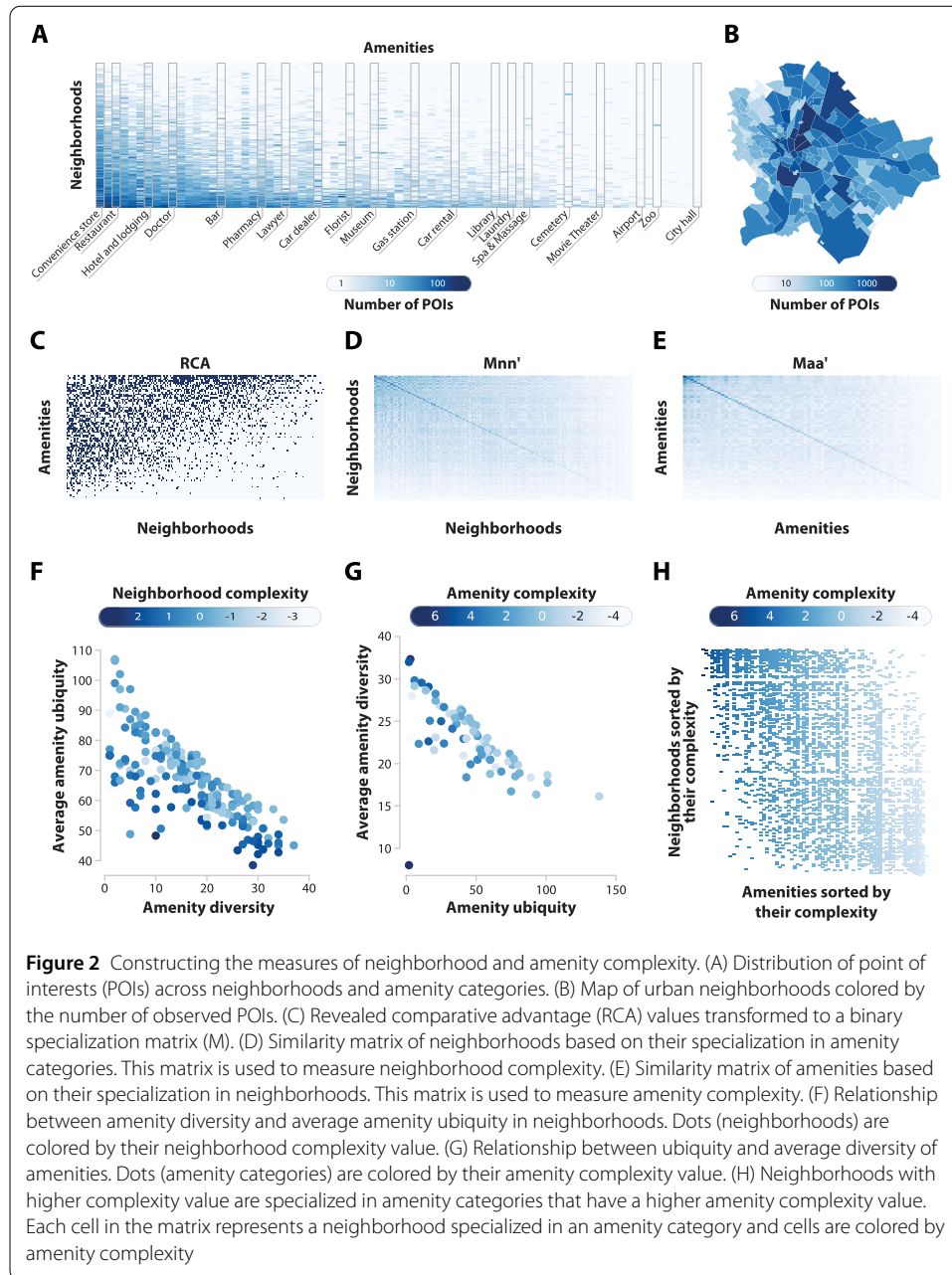
$$RCA_{n,a} = (P_{n,a}/P_a)/(P_n/P), \tag{1}$$

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:

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

Figure 2C illustrates the resulted binary  $RCA$  matrix of neighborhoods and amenity categories in Budapest. Sum of rows in this matrix presents 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).

$$\text{Amenity diversity} = M_n = \sum_a M_{n,a}, \tag{3}$$

$$\text{Amenity ubiquity} = M_a = \sum_n M_{n,a}. \tag{4}$$

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 [24, 35].

$$\text{Neighborhood complexity} = K_n = \frac{1}{M_n} \sum_a M_{n,a} K_a, \tag{5}$$

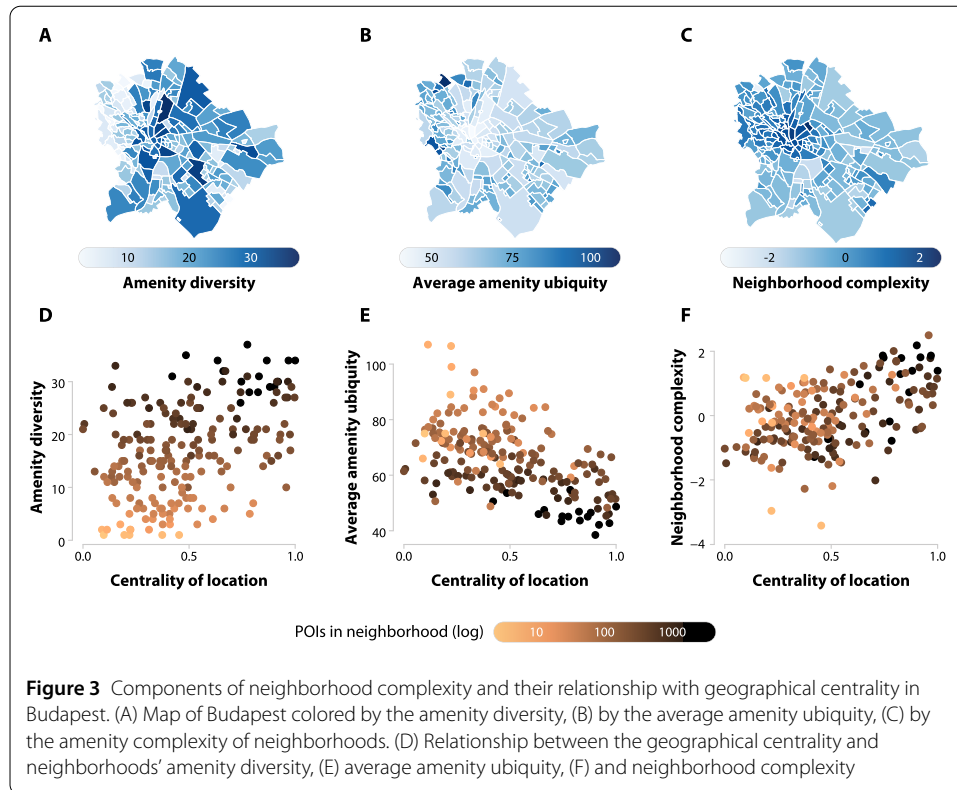
$$\text{Amenity complexity} = K_a = \frac{1}{M_a} \sum_n M_{n,a} K_n. \tag{6}$$

The economic complexity index (ECI) that describes the production structure of economies and the product complexity index (PCI) that describe the complexity of products were originally defined through the iterative, self-referential algorithm of the ‘method of reflection’ [24]. The algorithm calculates the above explained diversity and ubiquity vectors and then recursively uses the information in one equation to correct the other (see (5) and (6)). Later it was presented that the method of reflection is equivalent to finding the eigenvectors of the similarity matrix  $M_{nn'}$  and  $M_{aa'}$  [23, 36]. In our case  $M_{nn'}$  is defined from the original binary neighborhood-amenity matrix  $M$  as  $M_{nn'} = M^T * M$ . The neighborhood-neighborhood similarity matrix used to construct our *neighborhood complexity* measure is visualized by Fig. 2D. Neighborhood complexity is analogous to economic complexity in terms of measurement and it captures the amenity complexity of neighborhoods. To measure the complexity of amenity categories based on their geographic distribution across neighborhoods, we create an amenity-amenity similarity matrix as  $M_{aa'} = M * M^T$ , visualized by Fig. 2E. The network representation and the clustered version of this matrix can be found in Sect. 5 in Additional file 1. Our *amenity complexity* measure is constructed in a similar way to the product complexity index. As discussed in the Introduction, neighborhood complexity and amenity complexity are interpreted as measures of the sophistication of the local amenity supply that can serve a wide range of unique needs.

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 amenity 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 matrices (for an overview, see [23]).

Figure 2F illustrates the relationship between amenity diversity and average amenity ubiquity of neighborhoods. Each point represents a neighborhood and is colored by the derived neighborhood complexity values. Besides the expected negative correlation between amenity diversity and average amenity ubiquity [23], neighborhood complexity 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 derived amenity complexity values. 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 mechanical relationship between

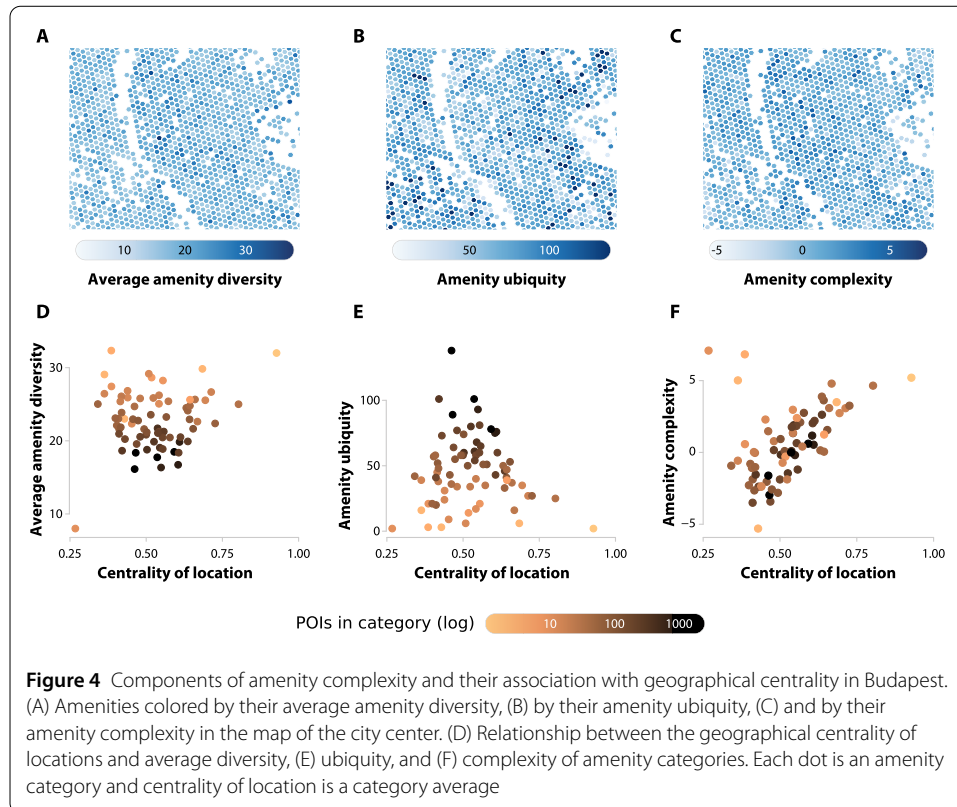




amenity complexity and neighborhood complexity. The figure makes it clear that complex neighborhoods have complex amenities. These patterns are in line with the ones revealed by Mealy et al. [36] for countries and exported products. Section 6 in Additional file 1 presents the ranking of neighborhoods and amenities in Budapest by their neighborhood complexity and amenity complexity values.

Figure 3A, B and C presents amenity diversity, average amenity ubiquity and neighborhood complexity on the map of Budapest, while Fig. 3D, E and F illustrates their correlation with the geographical centrality of neighborhoods in the city. The geographical centrality of locations (both in case of neighborhoods and actual amenities) is determined by the inverse of the average distance to reach the centroid of the location (on a logarithmic scale) from the center of every census tract in Budapest. Since census tracts are relatively homogeneous in terms of population, but heterogeneous in their area, this measure gives us higher values for more densely populated, central locations around the historical city center. To facilitate interpretation, the measure is normalized to a scale of 0-1. As this metric is not based on heuristics or local knowledge, it can be applied to other cities and is motivated by the approach taken by Moro et al. [17], who showed in detail that the average travel distance to locations is related to the diversity of visitors. In Sect. 7 of Additional file 1, we provide further details on our measure of geographic centrality, compare our results with the use of several other centrality metrics, and discuss their differences in terms of expectations about their social mixing abilities. Figure 3 suggests that all three variables correlate with central location, but the correlation is stronger for the average amenity ubiquity (−0.538) and neighborhood complexity (−0.464).

Figures 4A, B and C illustrate actual amenities on a zoomed in map of inner Budapest through size 10 H3 hexagons colored by the average diversity, ubiquity and amenity com-



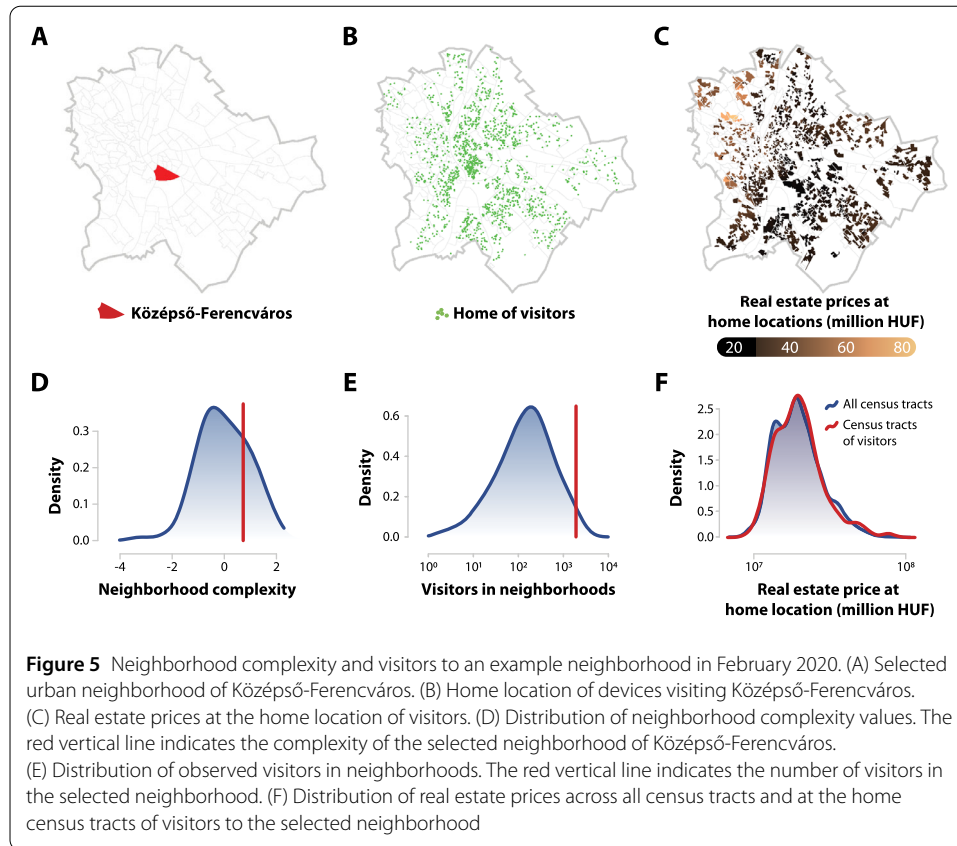
plexity of amenity categories at the location. At dense inner locations of the city, some hexagons contain amenities in multiple amenity categories. The identification of the dominant amenity category is detailed in Sect. 8 of Additional file 1. Figures 4D, E and F show how average amenity diversity, amenity ubiquity and amenity complexity are associated with central location. While average diversity and ubiquity of amenities have no clear connection to urban centrality (correlations are 0.180 and  $-0.056$ ), Fig. 4F suggests that complex amenities tend to be located around the city center (correlation is 0.451, with three remarkable outliers).

## 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 our neighborhood complexity index with more granular visitation patterns from mobility data. Figure 5 presents our process to join data sources through the example neighborhood of Középső-Ferencváros.

Figure 5A presents the location of the selected neighborhood, while Fig. 5B visualizes the home location of devices that visited any third places in Középső-Ferencváros during the month of February 2020. We connect the home location of visitors to census tracts as Fig. 5C illustrates. This allows us to infer 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 Fig. 5E and F show that Középső-Ferencváros is visited by more devices than most neighborhoods in February 2020 and its visitors come from diverse census tracts from all around Budapest.



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 testing the relationship between the diversity of visitors and the amenity structure of neighborhoods using simple OLS regressions on February 2020 data. Model (1) is our baseline model that illustrates the relationship between the diversity of visitors and the centrality of neighborhoods, while controlling for population, number of visitors and number of POIs in neighborhoods. The positive and significant coefficient for urban centrality suggests that central neighborhoods that are on average less distant from the census tracts of Budapest are visited by more diverse people. Model (2) builds on the same model structure, but includes neighborhood complexity as an explanatory variable. It suggests that neighborhood complexity has positive and significant correlation with the diversity of visitors, while taking into account, among other things, the urban centrality of neighborhoods. Interestingly, the diversity of amenities shows a mere negative correlation, while the average ubiquity of amenities is not correlated with socio-economic diversity of visitors in our case (see models (3) and (4) in Table 1). Model (5) includes all the key explanatory variables and highlight the stable, significant connection between neighborhood complexity and the diversity of visitors.

The average variance inflation factor (VIF) is below 10 in all of the above models, and the VIFs of our main explanatory variables are below 2 in all cases, indicating no serious

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

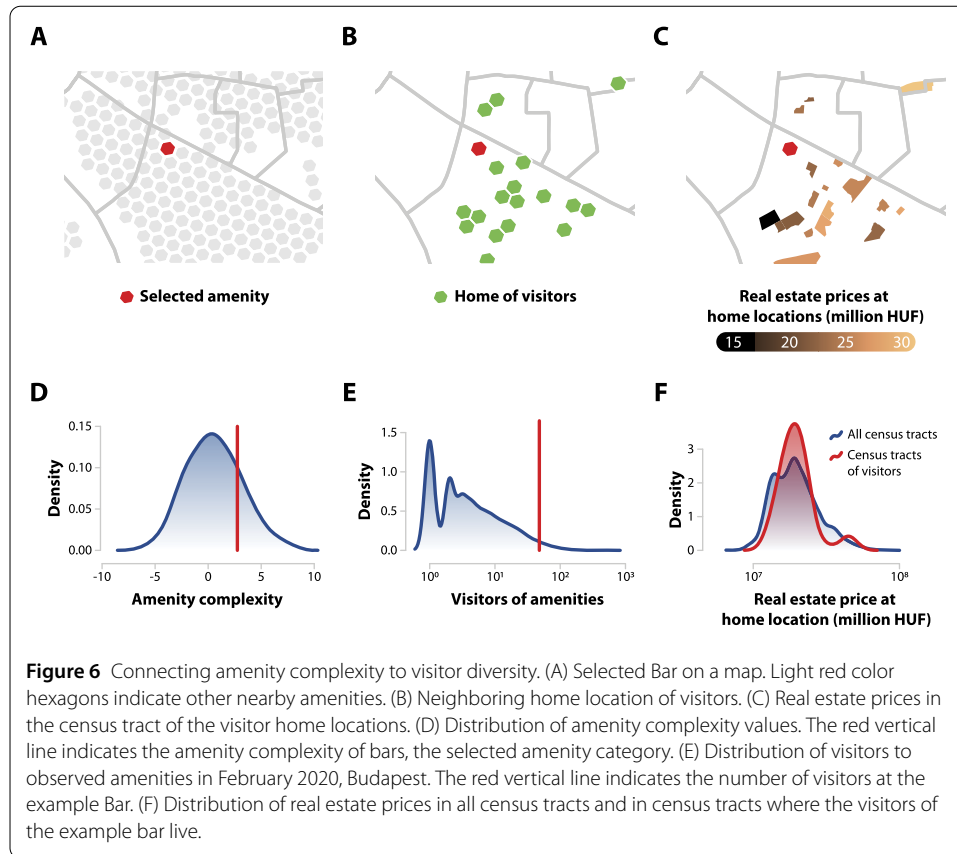
	Coefficient of variation				
	(1)	(2)	(3)	(4)	(5)
Neighborhood complexity		0.125*** (0.036)			0.125*** (0.038)
Amenity diversity			-0.148* (0.076)		-0.162** (0.074)
Avg amenity ubiquity				-0.049 (0.049)	-0.017 (0.050)
Centrality of location	0.130*** (0.027)	0.089*** (0.028)	0.129*** (0.026)	0.123*** (0.027)	0.087*** (0.028)
Population (log)	-0.053** (0.022)	-0.050** (0.022)	-0.049** (0.022)	-0.047** (0.023)	-0.043* (0.022)
Nr visitors (log)	0.047 (0.029)	0.039 (0.029)	0.054* (0.029)	0.045 (0.029)	0.046 (0.029)
Nr POIs (log)	-0.001 (0.023)	0.006 (0.023)	0.047 (0.033)	-0.013 (0.026)	0.054 (0.034)
Constant	0.424*** (0.056)	0.363*** (0.057)	0.360*** (0.064)	0.460*** (0.066)	0.306*** (0.075)
Observations	186	186	186	186	186
R <sup>2</sup>	0.257	0.303	0.273	0.261	0.321
Adjusted R <sup>2</sup>	0.241	0.284	0.253	0.241	0.295

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

problems of multicollinearity. In Sect. 7 of Additional file 1, we illustrate the robustness of our results by using a number of alternative measures of location centrality. Models using centrality metrics based on local knowledge show slightly different results, but for our main models in Table 1 we use the most general and adoptable centrality measure. 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 Sect. 9 of Additional file 1. We run the same models presented in Table 1 on the visitation patterns of non-local users only and observe similar results. In this setting we only consider users living outside the focal neighborhood. Related models are presented in Sect. 10 of Additional file 1.

Furthermore, the relationship between neighborhood complexity and socio-economic diversity of visitors is estimated for each of the available 24 months using the setting of model (2) in Table 1. The related figure in Sect. 11 of Additional file 1 suggests that neighborhood complexity has a positive and significant relationship with the diversity of visitors to neighborhoods in 18 out of the available 24 months. The possible reasons behind the uneven coefficients are discussed in Sect. 5.

Both Fig. 3 and Fig. 4 suggest that our neighborhood and amenity complexity measures are correlated to urban centrality. Indeed, neighborhood complexity and amenity complexity are derived from the spatial distribution of amenities and then used to explain visits to spatial units, which may raise spatial autocorrelation and endogeneity problems [37, 38]. We address potential endogeneity issues by applying two different instrumental variable (IV) approaches. Results of IV regressions presented in Sect. 12 of Additional file 1 further strengthen our argument that neighborhood complexity is connected to the socio-economic diversity of visitors.



#### 4.2 Diversity of visitors to complex amenities

To go beyond the level of neighborhoods, we combine amenity complexity measured at the amenity category level with visitations to actual amenities derived from our fine-grained mobility data. Figure 6 presents our process to join data sources at the level of amenities through an example bar in the neighborhood of Középső-Ferencváros, Budapest. The selected amenity is surrounded by other amenities (Fig. 6A) and by detecting the home location of visitor devices (in Fig. 6B we use February 2020 data and the surrounding area in size 10 H3 hexagons), we can observe the socio-economic status of visitors proxied by real estate prices (Fig. 6C). Our Bar example is a relatively complex amenity category (Fig. 6D) and is very frequently visited in comparison to other observed amenities in February 2020 (Fig. 6E). In addition, visitors from census tracts with medium or higher real estate prices are over-represented in February 2020, as shown in Fig. 6F.

We measure the socio-economic diversity of visitors to each amenity 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. 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 February 2020. All our models at the level of amenities present clustered standard errors at the level of amenity categories.

We find in Model (1) that geographical centrality is a significant predictor of socio-economic mixing at amenities, even after controlling for the total number of POIs in the

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

	Coefficient of variation				
	(1)	(2)	(3)	(4)	(5)
Amenity complexity		0.069*** (0.016)			0.040*** (0.015)
Amenity ubiquity			-0.071*** (0.015)		-0.051** (0.018)
Avg amenity diversity				0.047** (0.024)	-0.007 (0.018)
Centrality of location	0.126*** (0.011)	0.116*** (0.011)	0.118*** (0.010)	0.125*** (0.011)	0.115*** (0.010)
Nr POIs in category (log)	-0.013 (0.009)	-0.006 (0.007)	0.015* (0.009)	-0.001 (0.009)	0.009 (0.008)
Nr visitors (log)	0.073*** (0.007)	0.069*** (0.008)	0.068*** (0.008)	0.071*** (0.007)	0.067*** (0.008)
Constant	0.220*** (0.033)	0.172*** (0.030)	0.180*** (0.027)	0.173*** (0.036)	0.170*** (0.028)
Observations	2,742	2,742	2,742	2,742	2,742
R <sup>2</sup>	0.102	0.108	0.108	0.104	0.110
Adjusted R <sup>2</sup>	0.101	0.106	0.107	0.102	0.108

Note: \* $p < 0.1$ ; \*\* $p < 0.05$ ; \*\*\* $p < 0.01$ .  
Standard errors are clustered at the amenity category level.

respective amenity category across Budapest and the observed number of visitors to the amenity in the focal month. Model (2) presents that amenity complexity has an additional positive and significant relationship with the socio-economic mixing. The negative and significant coefficient of amenity ubiquity in Model (3) suggests that amenities that many urban locations are specialized in are visited by less diverse, while amenities that only few neighborhoods specialized in are visited by more diverse groups of people. This result is consistent with the findings of Moro et al. [17]. The positive and slightly significant coefficient on the average diversity of amenities in Model (4) indicates that amenity categories that mostly appear in diverse neighborhoods attract visitors with different socio-economic status. Model (5) includes all the key explanatory variables and highlight the stable, significant connection between amenity complexity and the diversity of visitors. While the effect of amenity diversity remains stable, the significance of average amenity diversity disappears in the final model.

The average VIF value is below 3 in all of the above models, and the VIFs of our main explanatory variables are close 1 in all cases, indicating no serious problems of multicollinearity. Results with alternative measures for location centrality can be found in Sect. 7 of Additional file 1. Amenity complexity is significantly related to socio-economic mixing using almost any centrality indicators. Changing the dependent variable to the Gini coefficient or the Theil index to capture the diversity of visitors to amenities, we obtain the same results. Related model outputs can be found in Sect. 9 of Additional file 1. The relationship between amenity complexity and socio-economic diversity of visitors is estimated for each of the available 24 months using the setting of model (2) in Table 2. The related figure in Sect. 11 in Additional file 1 presents that amenity complexity has a positive and significant relationship with the diversity of visitors to amenities in 16 out of the available 24 months.



## 5 Discussion

In this work we bring the ideas behind economic complexity metrics 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. We combine the information on the complexity of neighborhoods and of amenity categories with fine-grained mobility data to illustrate the relationship between the complexity of amenities available in a location and the socio-economic diversity of its visitors. Focusing on the urban neighborhoods of Budapest, Hungary, we find that neighborhoods that concentrate 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, the diversity and ubiquity of amenities, the two components of amenity complexity, show a less clear relationship with the socio-economic diversity of visitors. Diversity of amenities shows a surprising negative correlation with visitor diversity, but only at the neighborhood level, while ubiquity of amenities is related to visitor diversity only at the amenity category level.

The geographical centrality of urban locations is a strong predictor of socio-economic mixing. Our results illustrate that both neighborhood complexity and amenity complexity correlate with the geographical centrality of locations. Contrary to previous works [22], we find that diversity of amenities is less correlated to urban centrality, while amenity ubiquity is only associated with centrality of locations at the neighborhood level. The relationship between amenity complexity and centrality of locations has inspired a number of robustness checks, including the use of instrumental variable regressions (details can be found in Sect. 7 and Sect. 12 of Additional file 1), which further confirm our key finding that amenity complexity is associated with socio-economic diversity of visitors.

The general contribution of our paper is that we combine economic complexity concepts with urban mobility research. Constructing the measures of neighborhood and amenity complexity allows us to systematically test the contribution of certain amenity categories and also the amenity portfolio at certain locations 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. This 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.

To construct the amenity complexity measures, we rely on the specific neighborhood structure of Budapest, however, alternative spatial scales in different urban settings are necessary to be tested in the future. 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 metrics. 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 [39, 40]. Moreover, they have clear

administrative borders and people can identify with them, which makes the interpretation of amenity complexity results more appealing.

Inside Budapest, we observe that central location is correlated to both neighborhood complexity and amenity complexity and these factors are all connected to the socio-economic diversity of visitors. To understand this relationship clearly, we test several different measures of central location, with mixed results. For our main empirical exercise we choose the centrality metric based on the average distance to reach neighborhoods and amenities from any census tracts inspired by Moro et al. [17], as this measure does not require local knowledge and is easy to adopt for other cities. However, further research is needed to better understand these relationships, as a recent work based on similar measures adopted to Paris, France, shows that amenity complexity is not exclusively linked to a single city center [41].

Mobility data for our empirical analysis are produced on a monthly basis. Our findings are valid for 16 of the 24 months available and COVID-19 did not clearly affect the relationship between amenity complexity and the diversity of visitors to urban locations. However, our data does not contain enough observations to provide significant results for all the 24 months under study. Confirming the results with better mobility data is an important future research direction.

The counting of POIs within neighborhoods does not allow differentiation between the capacity or quality of amenities. We were not successful in using the number of visitors to amenity categories within neighborhoods as an input rather than the sheer number of POIs. In the future, similar but more sophisticated data would be needed to measure amenity complexity more accurately.

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 [36, 42] and the adoption of these methods to the neighborhood scale in urban environments is an apparent future research direction.

## Supplementary information

**Supplementary information** accompanies this paper at <https://doi.org/10.1140/epjds/s13688-023-00413-6>.

**Additional file 1.** Supplementary information accompanies this paper (PDF 8.6 MB)

## Acknowledgements

Sándor Juhász worked on the paper as a Marie Skłodowska-Curie Postdoctoral Fellow at the Complexity Science Hub Vienna (grant number 101062606). The work of Balázs Lengyel was financially supported by Hungarian National Scientific Fund (OTKA K 138970). 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, Frank Neffke, Gergő Tóth and László Czaller for their comments and suggestions. We acknowledge the significant assistance of our illustrator Szabolcs Tóth-Zs. in finalizing our primary figures. The authors thank the Social Science Computing Unit Budapest, and the Data Bank of the Centre for Economic- and Regional Studies for the contribution in data management. We acknowledge the help of the Hungarian Central Statistical Office for providing access to the real estate data.

## Funding

Not applicable.

## Abbreviations

GPS, Global positioning system; POI, Point of interest; ECI, Economic complexity index; PCI, Product complexity index; RCA, Revealed comparative advantage; OLS, Ordinary least squares; VIF, Variance inflation factor; IV, Instrumental variable.

### Availability of data and materials

Data sufficient to reproduce all results in this paper will be made available upon request. Code that supports the findings of this study is available at <https://github.com/sscu-budapest/mobility> and <https://github.com/sandorjuhasz/amenity-complexity>.

### Declarations

#### Competing interests

The authors declare that they have no competing interests.

#### Author contributions

All authors discussed and designed the experiments as well as contributing to the write up of the paper. SJ, GP, AK, EB and GM carried out the computational and analytical tasks; SJ and BL wrote the manuscript. All authors approved the final manuscript.

#### Author details

<sup>1</sup>Complexity Science Hub Vienna, Josefstädter Strasse 39, 1080, Vienna, Austria. <sup>2</sup>Laboratory for Networks, Technology and Innovation, Corvinus University of Budapest, Fővám tér 8, 1093, Budapest, Hungary. <sup>3</sup>Agglomeration and Social Networks Research Group, ELKH Centre for Economic and Regional Studies, Tóth Kálmán utca 4, 1097 Budapest, Hungary. <sup>4</sup>Center for Collective Learning, Corvinus University of Budapest, Fővám tér 8, 1093, Budapest, Hungary. <sup>5</sup>Institute for Data Analytics and Information Systems, Corvinus University of Budapest, Fővám tér 8, 1093, Budapest, Hungary.

Received: 14 December 2022 Accepted: 22 August 2023 Published online: 06 September 2023

### References

1. Jacobs J (1961) The death and life of great American cities. Random House, New York
2. Florida R (2004) Cities and the creative class. Routledge, New York
3. Benton-Short L, Short JR (2013) Cities and nature. Routledge, London
4. Glaeser E (2012) Triumph of the city: how our greatest invention makes us richer, smarter, greener, healthier, and happier. Penguin, New York
5. Musterd S (2020) Handbook of urban segregation. Edward Elgar, Cheltenham Glos
6. Mayer SE, Jencks C (1989) Growing up in poor neighborhoods: how much does it matter? *Science* 243:1441–1445
7. Torrats-Espinoso G (2021) Using machine learning to estimate the effect of racial segregation on Covid-19 mortality in the United States. *Proc Natl Acad Sci* 118(7):2015577118
8. Loughran K, Elliott JR (2022) Unequal retreats: how racial segregation shapes climate adaptation. *Hous Policy Debate* 32(1):171–189
9. Abadie A (2006) Poverty, political freedom, and the roots of terrorism. *Am Econ Rev* 96(2):50–56
10. Engler S, Weisstanner D (2021) The threat of social decline: income inequality and radical right support. *J Eur Public Policy* 28(2):153–173
11. Cagney KA, York Cornwell E, Goldman AW, Cai L (2020) Urban mobility and activity space. *Annu Rev Sociol* 46(1):623–648
12. Wang Q, Phillips NE, Small ML, Sampson RJ (2018) Urban mobility and neighborhood isolation in America's 50 largest cities. *Proc Natl Acad Sci* 115:7735–7740
13. Dong X, Morales AJ, Jahani E, Moro E, Lepri B, Bozkaya B, Sarraute C, Bar-Yam Y, Pentland A (2020) Segregated interactions in urban and online space. *EPJ Data Sci* 9:1
14. Bokányi E, Juhász S, Karsai M, Lengyel B (2021) Universal patterns of long-distance commuting and social assortativity in cities. *Sci Rep* 11:20829
15. Hilman RM, Iñiguez G, Karsai M (2022) Socioeconomic biases in urban mixing patterns of us metropolitan areas. *EPJ Data Sci* 11(1):32
16. Athey S, Ferguson B, Gentzkow M, Schmidt T (2021) Estimating experienced racial segregation in us cities using large-scale gps data. *Proc Natl Acad Sci* 118(46):2026160118
17. Moro E, Calacci D, Dong X, Pentland A (2021) Mobility patterns are associated with experienced income segregation in large us cities. *Nat Commun* 12:1–10
18. Noyman A, Doorley R, Xiong Z, Alonso L, Grignard A, Larson K (2019) Reversed urbanism: inferring urban performance through behavioral patterns in temporal telecom data. *Environ Plan B* 46:1480–1498
19. Christaller W (1933) Die Zentralen Orte in Süddeutschland. Fischer, Jena
20. Lösch A (1954) The economics of location. Yale University Press, New Haven
21. Wang M (2021) Polycentric urban development and urban amenities: evidence from Chinese cities. *Environ Plan B: Urban Anal City Sci* 48(3):400–416
22. Zhong C, Schlapfer M, Müller Arisona S, Batty M, Ratti C, Schmitt G (2017) Revealing centrality in the spatial structure of cities from human activity patterns. *Urban Stud* 54(2):437–455
23. Hidalgo CA (2021) Economic complexity theory and applications. *Nat Rev Phys* 3:92–113
24. Hidalgo CA, Hausmann R (2009) The building blocks of economic complexity. *Proc Natl Acad Sci* 106:10570–10575
25. Balland P-A, Broekel T, Diodato D, Giuliani E, Hausmann R, O'Clery N, Rigby D (2022) The new paradigm of economic complexity. *Res Policy* 51(3):104450
26. Magalhães L, Kuffer M, Schwarz N, Haddad M (2023) Bringing economic complexity to the intra-urban scale: the role of services in the urban economy of belo horizonte, Brazil. *Appl Geogr* 150:102837
27. Hidalgo CA, Castañer E, Sevtsuk A (2020) The amenity mix of urban neighborhoods. *Habitat Int* 106:102205
28. Oldenburg R (1999) The great good place: cafes, coffee shops, bookstores, bars, hair salons, and other hangouts at the heart of a community. Da Capo Press, Boston
29. Aslak U, Alessandretti L (2020) Infostop: scalable stop-location detection in multi-user mobility data
30. Uber Technologies, Inc. (2022) H3. <https://h3geo.org/>. Online. Accessed 28 Nov 2022

31. Kaufmann T, Radaelli L, Bettencourt LMA, Shmueli E (2022) Scaling of urban amenities: generative statistics and implications for urban planning. *EPJ Data Sci* 11:50
32. Heroy S, Loaiza I, Pentland A, O'Clery N (2022) Are neighbourhood amenities associated with more walking and less driving? Yes, but predominantly for the wealthy. *Environ Plan B: Urban Anal City Sci*, 1–25
33. Hungarian Central Statistical Office (2022) Regional Atlas. [https://www.ksh.hu/regionalatlas\\_administrative\\_units](https://www.ksh.hu/regionalatlas_administrative_units). Online. Accessed 10 Dec 2022
34. Natera Orozco LG, Deritei D, Vancso A, Vasarhelyi O (2020) Quantifying life quality as walkability on urban networks: the case of Budapest. In: Cherifi H, Gaito S, Mendes JF, Moro E, Rocha LM (eds) *Complex networks and their applications VIII*. Springer, Cham, pp 905–918
35. Balland P-A, Jara-Figueroa C, Petralia SG, Steijn MPA, Rigby DL, Hidalgo CA (2020) Complex economic activities concentrate in large cities. *Nat Hum Behav* 4:248–254
36. Mealy P, Farmer JD, Teytelboym A (2019) Interpreting economic complexity. *Sci Adv* 5(1):1705
37. Broekel T (2019) Using structural diversity to measure the complexity of technologies. *PLoS ONE* 14(5):1–23
38. Salinas G (2021) Proximity and horizontal policies: the backbone of export diversification and complexity. IMF Working Paper, 1–31
39. Chetty R, Jackson MO, Kuchler T et al (2022) Social capital I: measurement and associations with economic mobility. *Nature* 608:108–121
40. Chetty R, Hendren N, Katz LF (2016) The effects of exposure to better neighborhoods on children: new evidence from the moving to opportunity experiment. *Am Econ Rev* 106(4):855–902
41. Robertson C, Suire R, Dejean S (2023) Unpacking and measuring urban complexity evidence from amenities in Paris. *Pap Evol Econ Geogr* 23(25)
42. Tacchella A, Cristelli M, Caldarelli G, Gabrielli A, Pietronero L (2012) A new metrics for countries' fitness and products' complexity. *Sci Rep* 2(723):1–7

### Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Submit your manuscript to a SpringerOpen<sup>®</sup> journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

---

Submit your next manuscript at ► [springeropen.com](https://www.springeropen.com)

---