The amenity mix of urban neighborhoods

ARTICLE IN PRESS

Habitat International xxx (xxxx) xxx

Contents lists available at ScienceDirect

Habitat International

journal homepage: http://www.elsevier.com/locate/habitatint

The amenity mix of urban neighborhoods

César A. Hidalgo [a,b,c,*], Elisa Castañer [d], Andres Sevtsuk [e]

[a] ANITI Chair, University of Toulouse, France
[b] Alliance Manchester Business School, University of Manchester, UK
[c] School of Engineering and Applied Sciences, Harvard University, USA
[d] Collective Learning Group, The MIT Media Lab, USA
[e] Department of Urban Studies and Planning, Massachusetts Institute of Technology, USA

ABSTRACT

Advances in computational urbanism have stimulated the rise of generative and parametric approaches to urban design. Yet, most generative and parametric approaches focus on physical characteristics, such as a neighborhoods walkability, energy efficiency, and urban form. Here, we study the colocation patterns of more than one million amenities in 47 U.S. cities to model the amenity mix of neighborhoods, and to identify the amenities that are over- or under-supplied in a neighborhood. We build this model by combining a clustering algorithm, designed to identify amenity-dense neighborhoods, and a network, connecting amenities that are likely to collocate. Our findings extend generative and parametric urban design approaches to the amenity mix of neighborhoods, by leveraging the idea of relatedness from the economic geography literature, to evaluate and optimize a neighborhood's amenity mix.

1. Introduction

In recent years urbanism has witnessed the growth of parametric and generative design approaches (Mehaffy, 2008; Nagy et al., 2018; Schumacher, 2009). These approaches, which leverage new computational tools and big data, are helping optimize the form (Alonso et al., 2019; Noyman et al., 2019), energy efficiency (Nagy et al., 2018), and walkability of neighborhoods (Rakha & Reinhart, 2012; Sonta & Jain, 2019, pp. 454–461). Yet, it is unclear whether such approaches could be applied to more qualitative neighborhood characteristics, such as the mix of amenities that populate the commercial streets and corners of amenity rich neighborhoods.

In a parallel stream of literature, scholars in regional studies, economic geography, and complex systems, have developed methods to predict changes in economic structure using data on the colocation or coproduction of economic activities (R. Boschma et al., 2015; R. A. Boschma, 2005; Frenken et al., 2007; Guevara et al., 2016; Hidalgo et al., 2018, 2007; Jara-Figueroa et al., 2018; Neffke et al., 2011; Neffke & Henning, 2013). The consensus of this literature is that the collocation patterns of economic activities follows *the principle of relatedness* (Hidalgo et al., 2018), a statistical principle that can be used to predict the activities that a location is more likely to enter or exit in the future.

The idea behind the principle of relatedness is simple and powerful: economies are more likely to enter, and less likely to exit, economic activities that are related to those already present in a location. This principle explains the emergence and coherence of economic clusters and changes in international specialization patterns (Hidalgo et al., 2007). Yet, to bring these ideas to the neighborhood scale, and use them in generative urban design approaches, requires solving a few technical challenges.

First, there is the challenge of defining the spatial unit of observation (e.g. defining each neighborhood). Unlike regional or international data, which comes with well-defined statistical or administrative boundaries, neighborhoods are not administrative units, and hence, have boundaries that need to be learned directly from spatial data. Here, we overcome this challenge by introducing a simple clustering algorithm that can be used to identify the boundaries of amenity dense neighborhoods and the amenities that belong to them. We apply this method to a dataset containing data on over one million amenities, and use it to identify amenity dense neighborhoods in 50 U.S. cities. Once we identify neighborhoods, we estimate the number of amenities of each type we expect in each neighborhood by leveraging the *principle of relatedness.* Technically, we build a multivariate model that estimates the number of amenities of each type we expect in a neighborhood based on the collocation between that amenity and other amenities. The model predicts the number of amenities we expect to find in a neighborhood (e.g. number of restaurants, hotel, or hair salons), from data on the other types of amenities that are already present in it. We use this model on data from Boston, finding that the model identifies neighborhoods where specific amenities are over- or under-supplied. These methods provide a quantitative

* Corresponding author. Artificial and Natural Intelligence Toulouse Institute (ANITI), University of Toulouse, France.
E-mail addresses: cesar_hidalgo@fas.harvard.edu, cesifoti@gmail.com (C.A. Hidalgo).

https://doi.org/10.1016/j.habitatint.2020.102205
Received 11 September 2019; Received in revised form 28 June 2020; Accepted 2 July 2020

mean to evaluate the amenity mix of neighborhoods and help extend the tools of generative and parametric design to a neighborhood's amenity mix.

1.1. Amenities and place

In recent years, agglomerations of amenities, such as restaurants, shops, and libraries, have become increasingly valued by people, as reflected in real estate prices, and city governments, as reflected in place-making initiatives.

Amenity clusters are not only desirable, but booming. In Washington D.C. and Atlanta, walkable districts account for less than one percent of the total metro land mass, but attracted fifty percent of the metro area's new office, retail, hotel, and apartment square footage between 2009 and 2013. In the U.S., multiple inner cities are witnessing a return of the middle class fueled in part by a growing attraction to amenity clusters, the desire to attract creative individuals (Florida, 2019; Li et al., 2019; You & Bie, 2017), the effect of amenities in real estate prices (Jang & Kang, 2015), and due to the convenience of pedestrian life (Ehrenhalt, 2012). Surveys and rents show that both U.S. millennials and elders prefer to live in places with easy access to retail, food and services (Hanowell, 2017; The National Trust for Historic Preservation, 2017). For instance, offices in walkable neighborhoods have been found to command a 74 percent rent-per-square-foot premium over offices in drivable suburban areas (Leinberger & Lynch, 2014).

Yet, we still have much to learn about the mix of amenities that create a viable neighborhood.

On the one hand, there is a vast theoretical literature on agglomeration of businesses (Christaller & Baskin, 1966; Eaton & Lipsey, 1982; Fujita & Krugman, 2004; Hotelling, 1929; Krugman, 1993; Losch, 1954; Marshall, 1890; Mulligan, 1984), but this literature is too abstract compared to the actual business types that make up a neighborhood's actual amenity mix (e.g. pet store, bakery, car repair).

Agglomeration theories focus on coarser categories, and can be divided into exogenous and endogenous explanation for the presence of business agglomerations.

Exogenous clustering is generally thought to emerge when multiple businesses co-locate around a commonly attractive external resource (Berry, 1967). In cities with heavy-rail transit systems (e.g. Tokyo, London), for instance, retail, service and food businesses often cluster around transit stations. Stations tend to attract a large number of daily riders and proximity to these riders can produce "spillover" visits to stores. Having a number of restaurants around a popular destination such as a metro or train station, for instance, is partly explained by a common attraction to the same exogenous resource among multiple stores. This form of clustering can emerge among either competing or complementary businesses and can occur regardless of whether there are endogenous externalities involved between the stores themselves. Other examples of exogenous destinations that attract retail, service and food amenities include dense residential or employment areas, highly connected street intersections, highly trafficked roads or highways, as well as frequently visited public facilities, open spaces (e.g. waterfronts), institutions and tourist attractions.

Whether proximity to an exogenous destination can economically sustain an amenity cluster depends in part on the level of customer access at the location, and in part, on the location of competing clusters. Competing clusters reduce the market area that any one cluster can claim. A model for such competition was laid out by Walter Christaller in the 1930s as part of his Central Place Theory, who stipulated that in the long run, amenity clusters will divide the available market into equal catchments (Christaller & Baskin, 1966). Christaller's theory also divided goods and services into higher and lower order clusters, depending on their frequency of purchases. This meant that goods that are acquired less frequently (e.g. furniture) locate in fewer clusters with large catchment areas.

There are also endogenous reasons that lead amenities to co-locate in clusters. These reasons fall into two distinct types: complimentary and competitive clustering. Complementary clustering of amenities refers to co-location of businesses that do not directly compete and which are often acquired during the same outing—theaters and ice-cream shops, for instance, are complementary since they are often consumed during the same trip (Eppli & Benjamin, 1994).

Complementary clustering is generally explained by savings in transportation costs and time. Acquiring complementary goods on a single trip saves patrons costs that would otherwise be associated with multiple trips (Hernández & Bennison, 2000; Nelson, 1958). Customers therefore have an incentive to visit clusters that offer a wider choice of complementary goods. A person needing to buy clothes, ship a package, and purchase a meal, is more likely to visit a cluster that allows her to take care of all these errands instead of undertaking three separate trips. This in turn, motivates shop owners to locate in such clusters.

While complementary clustering refers to planned store visits, customer spillovers or positive demand externalities between stores can also lead to unplanned or "impulse" visits (Hernández & Bennison, 2000). Hence, complementary clustering can also emerge when less popular stores catch spillover customers by locating next to more popular stores (Brueckner, 1993). This dynamic is most poignantly exhibited in centrally managed shopping centers that offer cheaper leases to large "anchor" stores while charging smaller stores and amenities higher rents as a way to resell the anchor store externality.

Clustering is also common among competitive stores. Many cities have main streets, where numerous restaurants co-locate next to each other. A visitor to one of these clusters is unlikely to visit more than one restaurant at a time, but he or she might return to the cluster on another occasion.

As a way to formalize many of these ideas, neo-classical retail location theory suggests that competitive clustering is explained through two key factors: lower risk about the actions of competitive stores, and lower search costs for consumers and lower prices. The idea of risk reduction was introducing by Hotelling, in its classic model of ice cream vendors on a beach (Hotelling, 1929). Hotelling showed that the socially optimal equilibrium, where the two vendors divided up the market by locating at the 1/3rd and the 2/3rd points of the beach, was unstable. Instead, a stable equilibrium exists where the vendors cluster together in the middle of the beach, requiring customers to walk significantly longer distances, but eliminating the risk of their competitors moving closer to them, "stealing" some of their catchment area.

More commonly, competitive clustering can emerge due to customers' desire to compare the quality and prices of merchandises or services before making a purchase (Eaton & Lipsey, 1975; Scitovsky, 2013). A cluster of competing restaurants, for instance, offers patrons an opportunity to check the menus and prices of adjacent options and to make a more informed and better suited dinner choice. A similar logic applies to clothing stores, shoe stores, and hobby stores. Clusters of competing businesses in these cases save patrons time and search costs. Competitive clustering is therefore most commonly observed among "comparison goods"—goods that are similar but not identical and where quality and prices tend to vary. Note that convenience goods, such as groceries or liquor stores, do not tend to cluster with each other since there is little gained from comparing standardized merchandise.

The role of comparison shopping in influencing competitive clustering has also been validated empirically. Nevin and Houston found that the variety of retail merchandise for comparison shopping is a strong predictor of shopping-center sales (Nevin, 1980). Hise et al. found that the number of competitive retailers at a shopping center is significantly correlated with the income of the center (Hise et al., 1983). Ingene has shown that the level of assortment of similar merchandise can be one of the strongest predictors for customers' choice of shopping destinations (Ingene, 1984). In a survey of 1200 individuals in six malls in the U.S., Bloch and his colleagues found that visits to shopping areas without buying plans, together with visits to look at goods that might be bought in the future, constituted 62% of all trips (Bloch et al., 1991).

Another force contributing to the formation of competitive clusters has to do with prices. Dudey has shown that agglomeration between multiple competing stores can lead to lower prices, which further attracts more customers (Dudey, 1990). From the store owners point of view, locating in competitive clusters only makes sense if the additional customer draw exceeds the loss of customers due to competition and lower prices (DiPasquale & Wheaton, 1995).

1.2. Data

We use data from the Google Places API containing the latitude, longitude and type of amenity (i.e. cafe, restaurant, library, etc.). We crawled more than 1.26 million amenities across 47 US cities. The original data set was collected in 2014 and contains 95 different types of amenities. However, we merge amenities that have similar functionality (Table 1) and exclude amenities that are scarce or ambiguous to obtain a data set composed of 74 different types of amenities. The amenities we exclude are: taxi stand, campground, store, subway station, RV park, movie rental, and shopping mall. The resulting amenities are shown in Table 2.\

The data from Google's Places API is not free of biases and limitations. The data on amenities registered in Google Places focuses on customer-facing businesses and places of interest (from hair salons and bakeries to airports and cemeteries), and hence, fails to include information on other forms of economic activity, such as manufacturing. Also, the data might have coding issues (e.g. occasionally showing a restaurant registered as a bar) and can include business that have closed down. The biggest limitation is that the data is not dynamic, and hence, cannot be used to construct panels like those used in other relatedness studies. Yet, despite these limitations, the Google Places API is accurate enough to be the backbone of the world's most popular mapping service (Google Maps) and is used daily by millions of individuals to find the location of businesses. This imperfect dataset is nevertheless an attractive source to study the spatial organization of amenities at the intra-city scale.

2. Methods

2.1. Clustering amenities

Our first goal is to identify neighborhoods and the amenities belonging to each of them. Here we introduce a clustering technique based on aggregating the number of amenities that are near each amenity, and then find the "peaks" and "valleys" of this scalar landscape. This is related, but not equivalent to methods that cluster locations based on co-visitation patterns (Cranshaw et al., 2012).

Table 1
The left column shows the amenities that were merged into a new amenity type, shown in the right column.

Original Amenities	New Amenities
Hindu Temple	Religious Center
Mosque	
Place of worship	
Synagogue	
Church	
Meal Delivery	Restaurant
Meal Takeaway	
Food	
Restaurant	
Health	Doctor
Doctor	
Finance	Finance
Bank	
Roofing Contractor	Construction Contractor
Electrician	
Plumber	
Painter	
General Contractor	

We start by defining the *effective number of amenities* of location i, A_i as the number of amenities that can be reached from amenity i using the following accessibility index (Handy & Niemeier, 1997) A_i:

$$A_i = \sum_{j=1}^{N} e^{-\gamma d_{ij}}, \tag{1}$$

where d_{ij} is the straight-line distance in kilometers between amenity i and amenity j, γ is a decay parameter that discounts amenities based on their distance to location i, and N is the total number of amenities in a city.

To interpret A it is useful to note that an amenity located where the measurement is taking place (i.e. with $d_{ii} = 0$) contributes one to the effective number of amenities in that location. If that was the only amenity available in the dataset then $A_i = 1$, meaning that there is only one amenity that can be reached from there. Yet, what the accessibility index does is add the contribution of other amenities to that location discounted by distance. For instance, an amenity at distance $d_{ij} = \ln(2)/\gamma$ will contribute only 1/2 to that location's effective number of amenities (A_i).

We find that our algorithm finds meaningful neighborhoods when we set $\gamma = 16$, which implies that the contribution of an amenity halves every 62.5 m and becomes negligible at about 500 m (a reasonable radius for a walkable neighborhood). This is consistent with research showing that pedestrian trips rarely exceed a 10-min walk (Handy & Niemeier, 1997; Sevtsuk, 2014).

Because amenities that are far from a location i have an insignificant contribution to A_i, we need to only calculate the contribution of the k closest amenities to each location. This significantly facilitates the burden of the numerical computation while leaving our results unchanged. Going forward, we set $k = 2{,}000$, meaning we discard terms of order of magnitude smaller than 10^{-13}. This guarantees that the effective number of amenities at a location always converges before summing the kth amenity.

Fig. 1 a-c illustrates the methodology and shows the neighborhoods identified in Boston. Fig. 1 b show the "peaks" and "valleys" defined by the effective number of amenities A_i. We clearly see that the effective number of amenities peaks in well-known amenity dense neighborhoods (Fig. 1 c), such as Harvard Square or Boston's North End.

We then use A to identify the amenities belonging to each neighborhood using the following steps. First, since we are interested only in amenity dense neighborhoods, we remove the 10% of amenities that have the lowest value of A. For the remaining 90% of amenities we identify local peaks on the landscape defined by A. Fig. 1b shows the peaks in this landscape. To avoid identifying nearby peaks as different clusters, we require each peak to be a maximum among n of its neighbors, with n increasing with the size of the peak. We find that the heuristic $n_i = 3A_i + 50$ works very well at helping us avoid secondary peaks.

After we identified peaks, we assign amenities to them using an iterative greedy procedure. First, (i) we initialize neighborhoods by assigning to each peak all amenities that are in close proximity to it (less than 500 m). This reduces computational time by allowing us to focus the computation on the cluster boundaries, which is the more difficult computational problem. Then, (ii) we calculate the distance between each amenity that has not been assigned to a neighborhood and all amenities that have been assigned to a neighborhood, and (iii) we assign to a neighborhood only the amenity that is closest to an amenity that has already been assigned to a neighborhood. After having added one amenity to one cluster, we (iv), recalculate the distance between assigned and unassigned amenities by repeating step (ii). We repeat steps (ii) to (iv) until all amenities have been assigned to a cluster.

Fig. 1c shows an example of the clusters found for the city of Boston. Fig. 2 shows the procedure applied to New York and San Francisco. In all cases, the identified amenity clusters reflect the on the ground experience. For instance, in the case of Boston (Fig. 1), the clusters identified

Table 2
Total number of each type of amenity in the Google Places data set in the 47 US cities in our study.

Amenity	Points	Amenity	Points	Amenity	Points
Accounting	17280	Dentist	26071	Movie Theater	1232
Airport	1535	Department store	3515	Moving Company	12744
Amusement park	1017	Doctor	153772	Museum	2161
Aquarium	492	Electronics store	11876	Night Club	5675
Art gallery	5358	Embassy	688	Park	25723
ATM	30753	Finance	32221	Parking	5527
Bakery	9255	Fire station	2050	Pet Store	2270
Bar	21506	Florist	5102	Pharmacy	15204
Beauty salon	41851	Funeral home	2761	Physiotherapist	7929
Bicycle store	1409	Furniture store	12379	Police	1613
Book store	3417	Gas station	2552	Post Office	2723
Bowling alley	366	Grocery or supermarket	15206	Real Estate Agency	39484
Bus station	110642	Gym	5934	Religious Centers	58468
Cafe	9485	Hardware store	4595	Restaurant	112430
Car dealer	11603	Home goods store	29537	School	46516
Car rental	2968	Hospital	7942	Shoe Store	8612
Car repair	40215	Hotel and lodging	11452	Spa	2843
Car wash	3202	Insurance agency	27866	Stadium	1245
Casino	172	Jewelry store	6751	Storage	5849
Cemetery	2386	Laundry	14391	Train Station	1262
City hall	140	Lawyer	37611	Travel Agency	7394
Clothing store	29806	Library	3466	University	6597
Construction contractor	86044	Liquor store	7948	Veterinary Care	5373
Convenience store	13818	Local Government Office	10081	Zoo	114
Courthouse	717	Locksmith	2182	**Total**	1262374

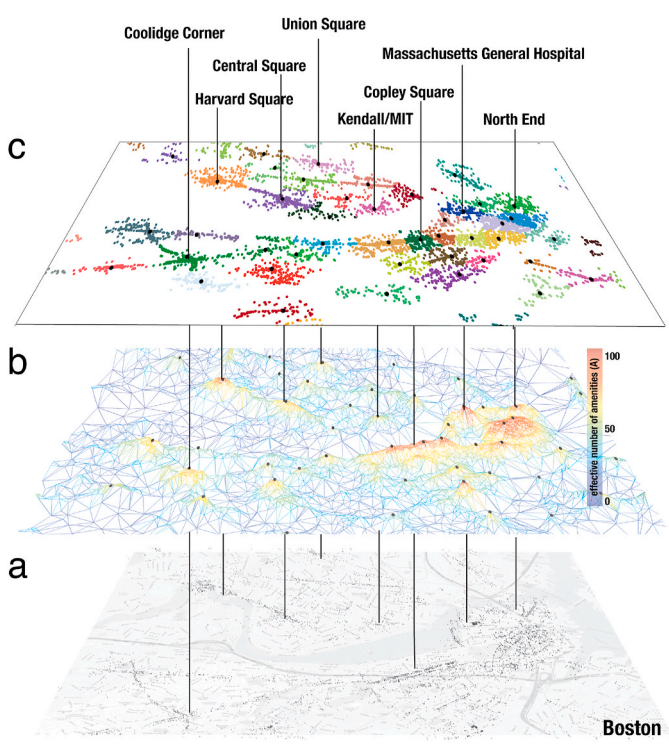

Fig. 1. Clustering algorithm. a Map of Boston with each amenity indicated as a grey dot. **b** The number of effective amenities (A) at each location where an amenity is present in Boston. Peaks represent locations with a high number of effective amenities and valleys represent locations with a low number of effective amenities. The black dots in the peak of the hills represent local maxima identified by our clustering algorithm, which we use as neighborhood centers. **C** Neighborhoods identified after the 90% of points with highest A has been assigned to a location using our clustering algorithm. Neighborhoods are shown as sets of dots of the same color. Neighborhood centers are also marked by black dots. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

correspond to well-known centers of urban activity.

3. Results: the amenity space

Armed with our clustering method, we now construct a network connecting amenities that are likely to locate in the same clusters. This network will be the backbone of our predictive model.

We construct the network of amenities, or Amenity Space, using spearman's rank correlation to measure the collocation of amenities across all clusters. That is, we create a weighted graph W_{ij}, connecting amenity types i and j using $W_{ij} = \text{spearman}(N_{ic}, N_{jc})$, N_{ic} is the number of times amenity i appears in cluster c. The correlation runs across all clusters. We use a rank based correlation (Spearman), instead of Pearson's, to avoid issues arising from mismatches in the distribution of amenities of different types.[1] This is important since some amenity types, such as restaurants, are extremely common, while others, like Zoos and Aquariums, are rare.

Fig. 3a shows a visualization of the network containing amenities that tend to locate in the same clusters. To avoid visual clutter, we visualize the network's Maximum Spanning Tree (Hidalgo et al., 2007) and add links with a pairwise Spearman correlation equal to or larger to 0.3 and which are also statistically significant (see Appendix for the full correlations matrix). This visualization technique avoids visual clutter and reveals amenities that tend to collocate with others. The size of circles illustrates the overall number of such amenities across all 47 cities. Amenity types with higher Spearman's rank correlations are shown closer to each other in the tree.

For example, the network shows that car repair shops collocate with car dealers (Spearman's $\rho = 0.45$), and religious centers collocate with schools (Spearman's $\rho = 0.46$). In fact, we find several well-defined clusters of amenities, that belong to different typologies. At the center of the network we find a food cluster (in green) that connects restaurants, bakeries, bars and cafes. That cluster is connected to a retail or

[1] Pearson's correlation is sensitive to the underlying distribution of the data and requires that of X and Y to be normally distributed for correlation between X and Y to be properly defined. Since Spearman's correlation focuses on rankings, it is robust to the underlying distribution. In our case, the distributions are not perfectly normal.

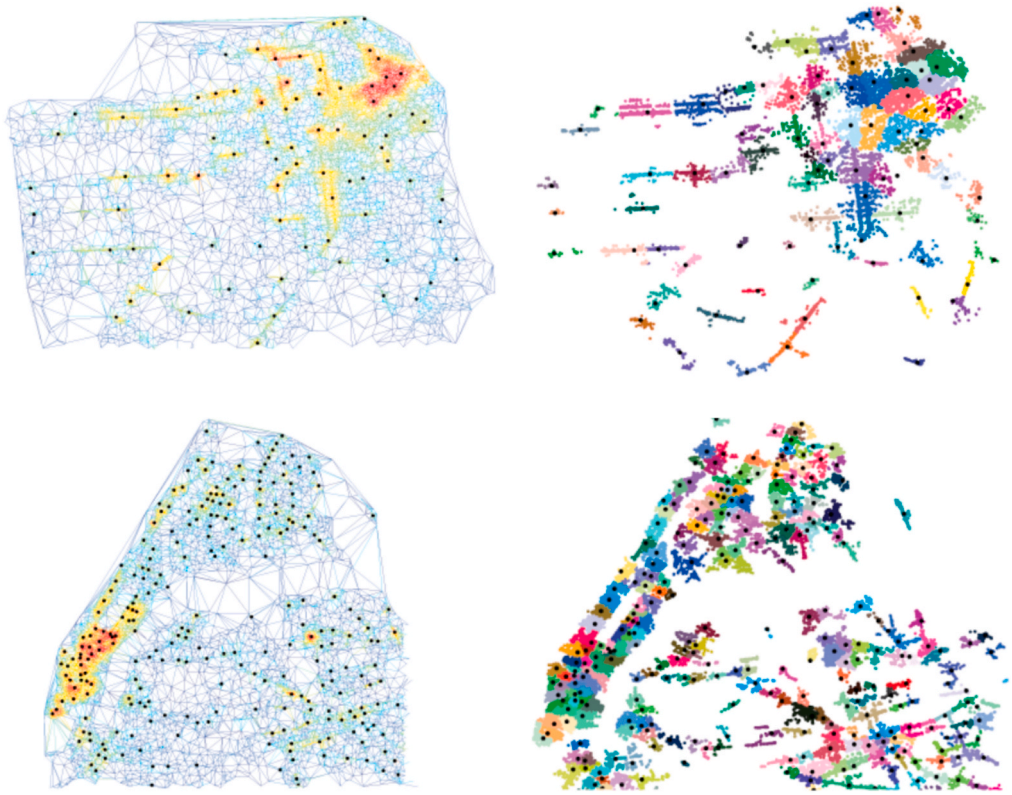

Fig. 2. Amenity clusters in San Francisco (top) and New York City (bottom). Red lines correspond to areas with a high effective number of amenities and blue lines correspond to areas with a low effective number of amenities. The black dots represent the locations we assign as neighborhood centers. The figures on the left show the corresponding assignment of amenities to neighborhoods. Each dot represents an amenity and sets of dots of the same color constitute a neighborhood. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

shopping cluster (orange) that includes clothing stores, shoe stores, jewelry stores, electronic stores, and pharmacies. The health cluster (pink), is populated by dentists, doctor offices, and hospitals. There is also a service cluster (cyan) populated by insurance agents, lawyers, beauty salons, banks (finance), and real state, and a second service cluster, more focused on cars, that includes car repair, car wash, and storage. Other clusters include the park, religious centers, and school clusters (lilac), and the government services clusters (yellow), which includes the city hall, fire station, and police station.

Interestingly, some pairs of amenities tend to repel each other, even if they are part of the same cluster. This is seen in non-transitive triads. For instance, religious centers are connected to schools and funeral homes, but funeral homes are not correlated with schools (at least not in the maximum spanning tree, or with the $\rho = 0.3$ threshold that we demand for a connection). Similarly, stadiums and libraries are correlated with universities, but they are not correlated with one another.

We can use this network to tell us which combinations of amenities should predict the presence of others. This can help us estimate a way to detect amenities that are under or over represented in a neighborhood given that neighborhood's current pattern of specialization. For instance, the network tells us that neighborhoods that specialize in beauty salons, accountants, and dentist, also tend to specialize in real-estate agents, but not car rentals. These results shed light on the degree of complementarity among businesses and institutions in urban settings. They unfortunately do not tell us whether the complementarity works in one or in both directions—we can observe doctors' offices near dentists' offices but we cannot tell from the data if this is because only dentists prefer to locate near doctors or if both types of establishments are attracted to each other.

We explore the utility of the Amenity Space by using it to build a parsimonious recommendation system (Maes, 1995; Resnick & Varian, 1997) for each type of amenity. We build this recommendation algorithm using multivariate regressions and a forward selection algorithm that iteratively includes new types of amenities to the regression until the contribution of a new type of amenity is statistically insignificant (characterized by a p-value of more than 0.001 (see Appendix)). In addition, we control for over-fitting by using both Akaike's Information Criterion (AIC) and Bayes' Information Criterion (BIC). For instance, to predict the number of amenities of a given type in a cluster (e.g. the number of beauty salons), we find the amenity that correlates more strongly with beauty salons (e.g. clothing stores). Then we search for the next amenity that contributes more to the predictive power of the model (measured using R^2). We continue adding amenity types, one by one, until the increase in R^2 obtained by adding the extra amenity is not statistically significant.

To test the accuracy of the model, we compare the predicted number of amenities of each type in a cluster with a simple benchmark where we use only the total number of amenities in a cluster as a predictor. This benchmark is inspired by the literature on urban scaling laws, which has shown that many important urban characteristics (such as GDP or crime) scale with city size (L. M. A. Bettencourt et al., 2007; L. Bettencourt & West, 2010; Youn et al., 2016).

Fig. 3b compares the R^2 of the models constructed using the amenity space with the models using only cluster size. In most cases (66/74 = 89%), the BIC test chooses the regression using the amenity space over the regression using only cluster size (the exceptions are airports, aquariums, bus stations, car rentals, casinos, convenience stores, gas stations, and zoos). This is likely due to the fact that such amenities either beget a cluster around them or within them (Airports, Casinos, and Aquariums), and are very rare, or because they simply follow aggregate demand and do not interact closely with neighboring amenities (gas stations, convenience stores). Overall, our findings suggest that the presence of an amenity in a cluster is connected more strongly to the presence of other amenities, than to the size of the cluster, except for the case of airports, aquariums, bus stations, casinos, convenience stores, gas stations, and zoos. Also, we note that the differences between the two models are not just statistically significant, but also characterized by strong size effects. On average, for the 66 amenity types in which the Amenity Space model works better, the R^2 of the Amenity Space model is twice that of the model using size only cluster size ($R^2 = 17\%$ on

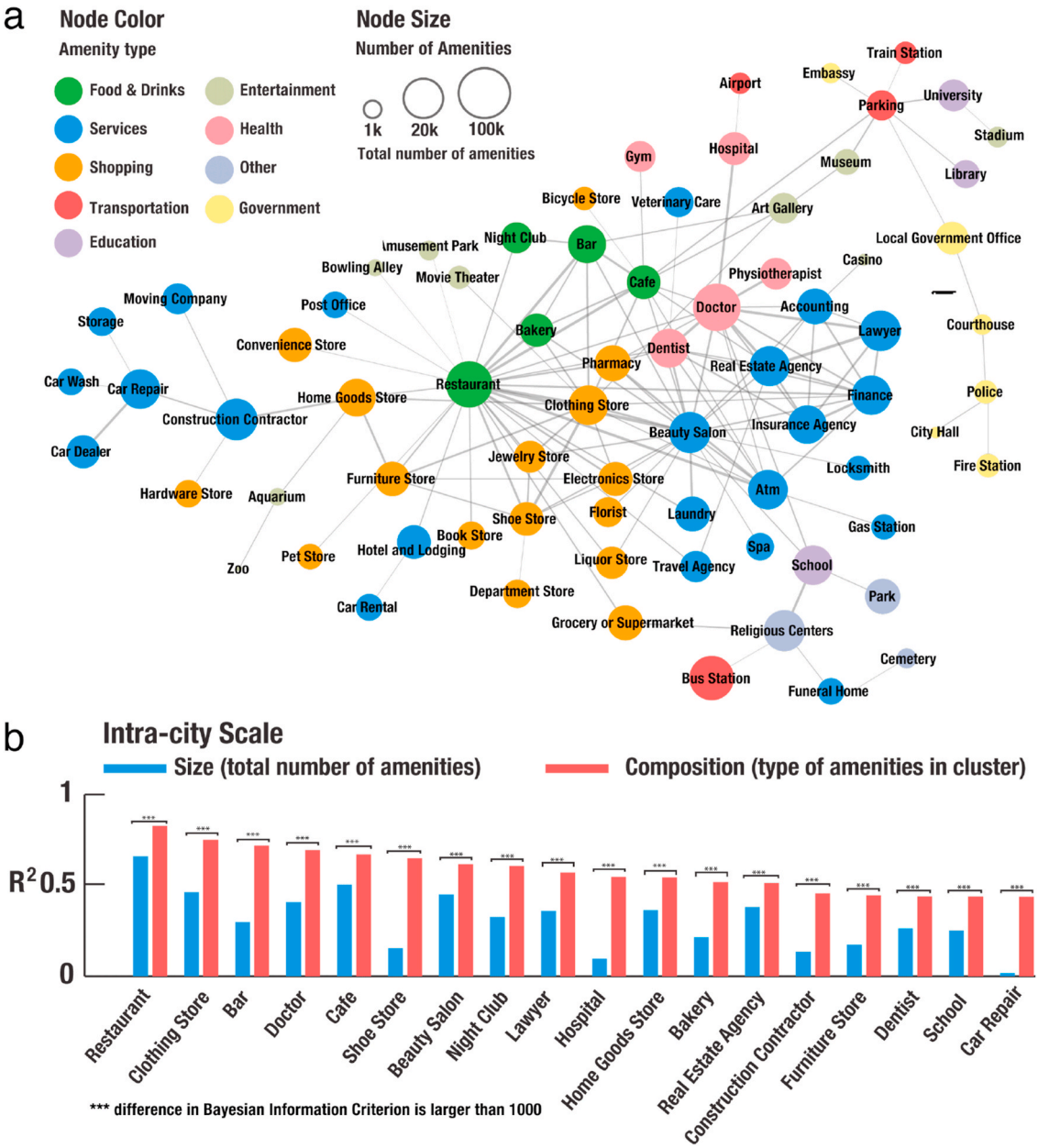

Fig. 3. **a** Network of amenity collocations. The nodes in the network represent different types of amenities and the edges connect amenities that are likely to collocate in the same neighborhood (see Appendix). The width of the edges connecting a pair of nodes is proportional to the spearman correlation obtained from the collocation of the two types of amenities across all neighborhoods. The size of a node is proportional to the number of times that an amenity is present in our data set. The color of each node represents the category that the amenity belongs to. **b** Comparison of the accuracy of two models used to predict the total number of amenities of each type on a neighborhood. The light-blue bars show the R^2 of a model predicting the number of amenities of each type in a neighborhood using only the total number of amenities in that neighborhood. The red bars show the R^2 of a model using information on the number of amenities of other types that are present in a neighborhood. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

average using size vs. $R^2 = 35\%$ on average using amenity space). This means that the increase in predictive power obtained by considering the types of amenities that locate in a neighborhood is not only statistically significant, but also substantial.

Finally, we study deviations from the regression predictions as a means to identify amenities that are either under or over represented in a cluster. For illustration purposes, we focus on amenity clusters for the city of Boston, but a similar diagnostic result can be produced for any city in the dataset. Fig. 4 compares the number of amenities observed in a neighborhood with those predicted by the presence of other amenities.

Fig. 4 a-c compare, respectively, the number of car parks, hotels, and beauty salons, observed and predicted, for each amenity cluster in Boston. Points above the line, such as Harvard Square in the case of car parks (Fig. 4a), the North End for hotels (Fig. 4b), and Central Square for Beauty Salons (Fig. 4c), indicate amenities that are underrepresented in that location, given that location's current pattern of specialization. Points below the lines such as Boston's Theatre District in car parks, Coolidge Corner in hotels, and Winthrop in beauty salons, suggest instances of excess supply.

Of course, the diagnostics of the model should be taken with care–they simply refer to the differences between the observed and "typical" amenity mix. For instance, our model identifies a shortage of parking in Harvard square, but it does not suggest that the city of Cambridge should consider adding more parking spaces around Harvard. Deviations in the

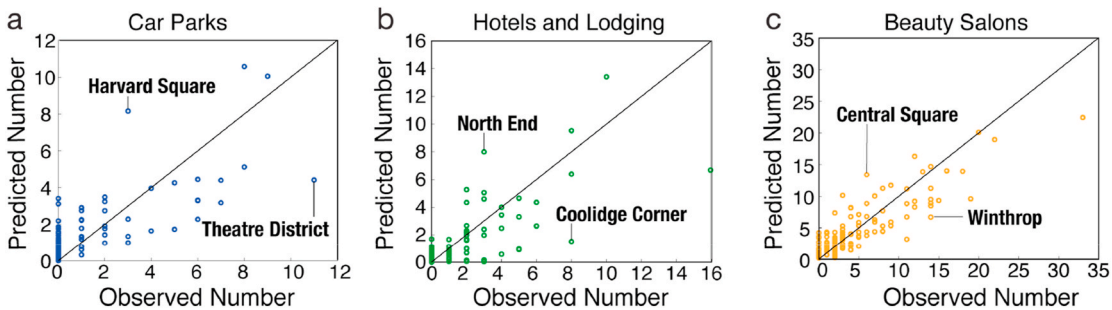

Fig. 4. Prediction of amenities in Boston's neighborhoods. a Observed vs. predicted number of car parks, **b** hotels, and **c**, beauty salons for each neighborhood in Boston. Points above the lines represent neighborhoods where the predicted number of amenities is higher than the observed, suggesting instances of under supply. Points below the lines represent neighborhood where the predicted number of amenities is lower than the observed, suggesting instances of excess supply.

amenity mix of a neighborhood can be explained by place characteristics that are not included in our model, such as architectural or historic value (Been et al., 2016; De Nadai et al., 2016; Naik et al., 2017, 2016; Salesses et al., 2013) or environmental externalities of car use. The lesson here is that the model successfully detects a known reality of Harvard Square, which is a lack of parking relative to the number of amenities it hosts. Similarly, Fig. 4b shows that the model detects a lack of hotels in the North End, a well-known tourist area in downtown Boston populated by a handful of hotels. This could mean that there is potential for hotel development or conversion in the North End, but again, such a conclusion would require additional factors that go beyond the scope of our contribution.

4. Conclusion

Generative and parametric design approaches are an important tool for modern urban design. Yet, extending these approaches to the amenity mix of neighborhoods has been challenging. Here, we leverage the principle of relatedness, from the economic geography literature, to create a method to describe, evaluate, and optimize a neighborhood's amenity mix.

Our method results in an amenity network that passes common sense muster, but also, that provides a quantitative description of neighborhood scale clusters. This adds to the literature on relatedness (Hidalgo et al., 2018), which has mapped networks of related products (Hausmann et al., 2014; Hidalgo et al., 2007), industries (Jara-Figueroa et al., 2018; Neffke et al., 2011; Neffke & Henning, 2013), occupations (Alabdulkareem et al., 2018; Muneepeerakul et al., 2013), patents (R. Boschma et al., 2015; Kogler et al., 2013), and research areas (Guevara et al., 2016). Of course, the amenity space has a structure that is different from these other networks. For instance, it is unlike the research space (Guevara et al., 2016), which is characterized by a ring like structure, circling from biology, to chemistry, to physics, computer science, economics, psychology, and back to biology. It is more similar to the product space, in which it has a clear core, in this case composed of food, retail, and personal services, surrounded by a periphery that includes cemeteries, bus stations, car washes, and airports. Yet, unlike the coagglomeration of products and industries, which is usually measured at a coarser scales reflecting the presence of common inputs (Ellison & Glaeser, 1997), the coagglomeration of amenities tend to respond more to shared demand, local real estate prices, and zoning constraints, so the similarity in structure (core-periphery), should not be interpreted as a similarity in input requirements.

Our results are also subject to numerous limitations and should be interpreted in the narrow context of the data from which they were derived. For instance, our work does not distinguish between social and retail amenities, which could cluster due to different agglomerative forces. In fact, the agglomeration of social amenities may well be due to gentrification and physical urban change (Naik et al., 2017). The data we used also comes from an online mapping service that is not officially verified or released by individual cities, and the analysis we performed was applied only to 47 U.S. cities. Moreover, because we lack the data and resources needed to estimate travel time between millions of pairs of points, we do not estimate distances using information about traffic, road conditions, or means of transport, but as "the crow flies." This source of error, however, is somehow mitigated by the fact that we are focusing on relatively short distances (<500 mts), that are less likely to be interrupted by geographic features (such as rivers) and that do not require car travel. Still, despite these numerous limitations, the data is in line with recent research leveraging online data to characterize neighborhoods (Dong et al., 2019). The question of whether the results can be generalized to other locations and whether these results hold for other datasets, remains to be studied in future work.

Beyond the data biases described above, the models presented above are limited by their simplicity, and bounded by the total amount of variance in the presence of amenities that we can explain. Our statistical model is based on linear regressions that could be potentially improved by using more complex functional forms or other machine learning approaches. Also, they could be enhanced by adding information on population density, the historic, cultural or environmental appeal of a neighborhood, foot traffic, site density restrictions (Fu & Somerville, 2001), and seasonal variations in traffic as captured by mobile phone data (Dong et al., 2017; Gonzalez et al., 2008).

Keeping the limitations and shortcomings in mind, however, the results and methods presented point to interesting new avenues of urban research and possible interventions. Our results could help inform what types of business permits or incentives ought to be further considered to help balance the amenity mix of particular neighborhoods, and also, how to alter travel behavior by improving walkability to help promote the development of clusters. There is a substantial body of planning literature on urban retail or food deserts, which compares the presence or lack of particular types of businesses across cities and urban districts without taking into account the co-location dynamics of such stores with other amenities (Schuetz et al., 2012). For instance, the correlation matrix of amenities identified here suggests that supermarkets and grocery stores collocate with restaurants and beauty salons, but also with liquor stores, clothing stores and bakeries. Our results can inform such analysis, by adding a layer of collocation dynamics, thereby yielding a deeper understanding of inequality in access to amenities.

Additionally, longitudinal data sources for both amenities and the built environment that surrounds them could be used to explore interaction between amenities that locate in an area and private real estate or public infrastructure investment that takes place. This could lead to a better understanding of the externalities and multiplier effects that the evolution of amenity clusters creates for neighborhoods or vice versa.

Together, our results, contribute to the growing toolbox of generative urban design.

CRediT authorship contribution statement

César A. Hidalgo: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Supervision, Validation, Visualization, Writing - original draft, Writing - review & editing. **Elisa Castañer:** Data curation, Formal analysis, Methodology, Visualization. **Andres Sevtsuk:** Writing - review & editing.

Appendix

Additional Description of Methods.

CLUSTERING

A. Effective number of amenities

We begin our clustering procedure by calculating the effective number of amenities at each location. The effective number of amenities, A_i, in a location i represents the number of amenities that can be reached by walking from that location. We define A_i as:

$$A_i = \sum_{j=1}^{N_c} e^{-\gamma d_{ij}} = \sum_{j=1}^{k} e^{-\gamma d_{ij}} + \sum_{j=k+1}^{N_c} e^{-\gamma d_{ij}} = \sum_{j=1}^{k} e^{-\gamma d_{ij}} + \varepsilon \tag{2}$$

where d_{ij} is the distance (in km) between amenity i and amenity j, and N_c is the total number of amenities in a city c. γ is a decay parameter that discounts amenities based on their distance to location i,. We set $\gamma = 16$, meaning that the contribution of an amenity to the effective number of amenities at a location roughly halves every 62.5 m and becomes negligible at about 500 m. Moreover, k determines the number of amenities that we use, instead of N_c, to calculate the effective number of amenities, I, at a location. Theoretically all of the amenities in a city should contribute to the effective number of amenities at a location in the city. However, because amenities that are far from a location i have an insignificant contribution to the effective number of amenities at that location i, we only calculate the contribution of the k closest amenities to each location. This yields an error ε in the effective number of amenities. We set $k = 2000$, which is a large enough so that the effective number of amenities at a location always converges before summing the kth amenity.

B. Identifying cluster centers

We continue our clustering procedure by identifying the center of each neighborhood as the local peaks on the landscape defined by A. We identify local peaks by searching for locations that have an effective number of amenities, A_i, larger than their n_i nearest neighbors. We define n_i as: $n_i = 3I_i + 50$, i.e. a function of the effective number of amenities at location i, so that the centers of very dense neighborhoods are required to have a large A_i to be considered a peak. By setting n_i proportional to I_i we avoid assigning multiple neighborhood centers to areas with high density of amenities, and we avoid not assigning any neighborhood center to areas with a low density of amenities.

C. Assigning points to clusters

Finally, we assign points to a neighborhood using the peaks we obtained. First, we remove the 10% of the points in each city with the lowest effective number of amenities, to eliminate isolated amenities that are not part of an agglomeration. After that, we assign all amenities that are within a distance of 0.5 km of a neighborhood center to that neighborhood. Then, we calculate the distance from each unassigned point to each assigned point using the following algorithm:

1. Choose the unassigned point, u, which is closest to an assigned point, a.
2. Assign point u to the neighborhood point a belongs to.
3. Calculate the distance from each unassigned point to the newly assigned point u.

The algorithm finalizes once all points have been assigned to a neighborhood. Figs. 1 and 2 in the main text show the effective number of amenities in the cities of Boston, San Francisco, and New York, and the corresponding assignments of amenities to neighborhoods.

COLLOCATION OF AMENITIES

To study the collocation patterns of amenities, we calculate the spearman correlation between all pairs of amenities across neighborhoods. We show the resulting correlations in the form of a network, where nodes represent amenity types and edges connect amenities that are highly correlated across neighborhoods. To construct this network, we first create a Maximum Spanning Tree (MST) of the network and then add edges only between amenities that have a pairwise correlation equal or larger than 0.3.

Here, we show the values of all spearman correlations between amenities across neighborhoods in the form of a matrix (Figure A1). We cluster amenities using Ward linkages.

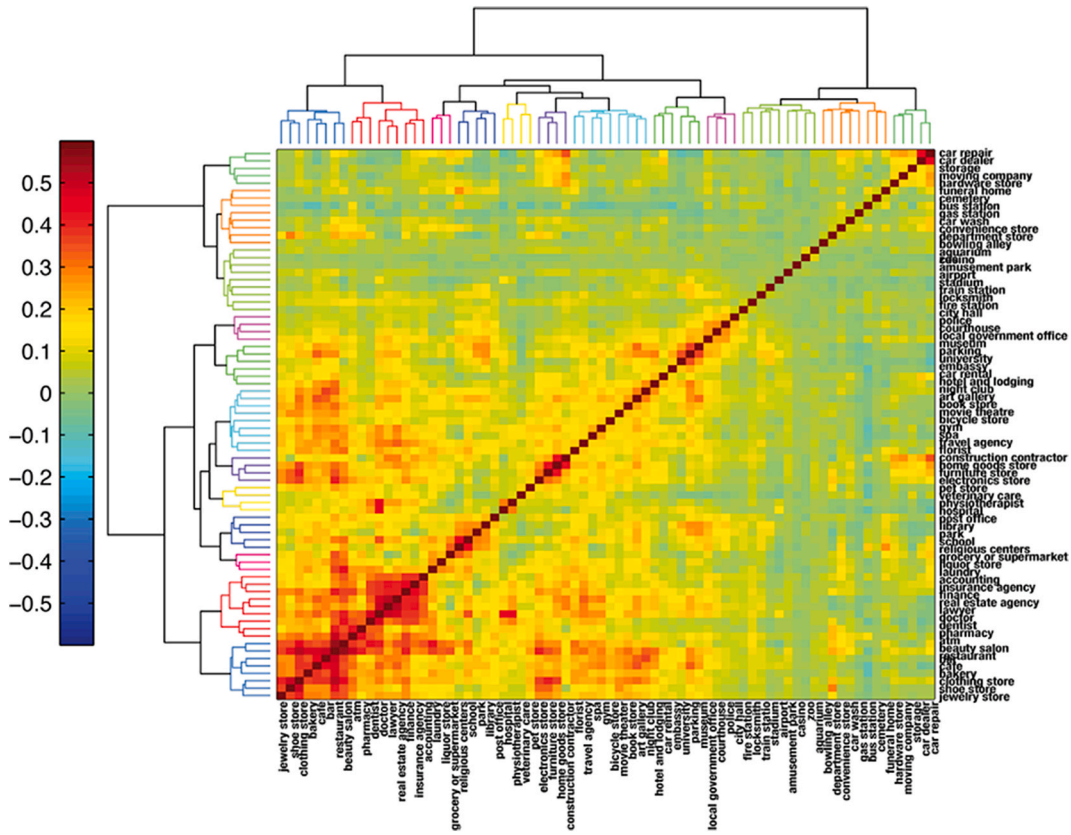

Fig. A1. Amenities correlations. Matrix showing the Spearman correlation between each pair of amenities. Amenities are clustered using Ward linkages.

PREDICTIONS

We develop four models to predict each type of amenity in the intercity and intra-city scale using two different metrics. In the intercity scale, we create a model (a linear regression) that uses the total number of amenities in a city to predict the number of each type of amenity in that city, and a model (regression using forward selection with p-enter value of 0.001), that uses the composition of amenities in a city to predict the number of each type of each amenity in that city. In the intra-city scale, we create a model (linear regression) that uses the total number of amenities in a micro-cluster to predict the number of each type of amenity in that micro-cluster, and a model (regression using forward selection with p-enter value of 0.001), that uses the composition of amenities in a micro-cluster to predict the number of each amenity in that micro-cluster. Table A1 shows the R^2 obtained for each of these models.

Table A1

R^2 of each of the models for each amenity. Given that these four models use a different number of samples and parameters, we calculate the Akaike Information Cirterion (AIC) and Bayesian Information Criterion (BIC) of each of the models. These criteria allow us to differentiate the models: the lower the AIC and BIC values, the more desirable the model (better fit and less overfitted). The AIC and BIC values obtained for each model are summarized in Table A2.

	Inter-City Model		Intra-City Model	
	Size	Composition	Size	Composition
Accounting	0.946	0.985	0.291	0.448
Airport	0.575	0.816	0.016	0.114
Amusement Park	0.382	0.724	0.002	0.005
Aquarium	0.709	0.880	0.014	0.028
Art Gallery	0.603	0.930	0.114	0.271
ATM	0.911	0.967	0.320	0.465
Bakery	0.777	0.980	0.364	0.543
Bar	0.649	0.966	0.462	0.750
Beauty Salon	0.952	0.989	0.449	0.615
Bicycle Store	0.594	0.919	0.080	0.183
Book Store	0.878	0.980	0.245	0.344
Bowling Alley	0.478	0.702	0.004	0.014
Bus Station	0.242	0.431	0.023	0.237
Cafe	0.649	0.956	0.505	0.670
Car Dealer	0.608	0.850	0.003	0.231
Car Rental	0.831	0.942	0.042	0.118
Car Repair	0.867	0.976	0.016	0.437

(*continued on next page*)

Table A1 (*continued*)

	Inter-City Model		Intra-City Model	
	Size	Composition	Size	Composition
Car Wash	0.828	0.970	0.005	0.071
Casino	0.016	0.000	0.002	0.008
Cemetery	0.126	0.585	0.001	0.015
City Hall	0.379	0.449	0.031	0.151
Clothing Store	0.884	0.993	0.298	0.718
Construction Contractor	0.824	0.978	0.135	0.456
Convenience Store	0.629	0.928	0.042	0.134
Courthouse	0.676	0.738	0.088	0.446
Dentist	0.954	0.974	0.262	0.439
Department Store	0.673	0.945	0.016	0.200
Doctor	0.957	0.986	0.408	0.694
Electronics Store	0.924	0.966	0.224	0.355
Embassy	0.102	0.419	0.046	0.114
Finance	0.953	0.983	0.424	0.610
Fire Station	0.490	0.632	0.018	0.058
Florist	0.889	0.981	0.207	0.259
Funeral Home	0.476	0.787	0.018	0.146
Furniture Store	0.912	0.980	0.173	0.444
Gas Station	0.443	0.777	0.000	0.028
Grocery or Supermarket	0.791	0.955	0.116	0.377
Gym	0.911	0.984	0.229	0.339
Hardware Store	0.896	0.953	0.020	0.194
Home Goods Store	0.908	0.986	0.213	0.517
Hospital	0.958	0.979	0.096	0.546
Hotel and Lodging	0.795	0.824	0.250	0.435
Insurance_agency	0.825	0.981	0.234	0.433
Jewelry Store	0.902	0.978	0.208	0.352
Laundry	0.933	0.984	0.180	0.354
Lawyer	0.871	0.894	0.359	0.570
Library	0.610	0.937	0.180	0.416
Liquor Store	0.753	0.815	0.175	0.301
Local Government Office	0.901	0.937	0.181	0.567
Locksmith	0.671	0.752	0.033	0.053
Movie_theater	0.780	0.952	0.125	0.190
Moving Company	0.721	0.931	0.012	0.131
Museum	0.499	0.951	0.221	0.412
Night Club	0.735	0.957	0.326	0.606
Park	0.669	0.745	0.149	0.320
Parking	0.666	0.938	0.374	0.610
Pet Store	0.812	0.943	0.077	0.192
Pharmacy	0.878	0.949	0.169	0.371
Physiotherapist	0.863	0.931	0.081	0.260
Police	0.681	0.866	0.052	0.201
Post Office	0.859	0.964	0.090	0.130
Real Estate Agency	0.835	0.952	0.381	0.513
Religious Centers	0.744	0.868	0.171	0.430
Restaurant	0.921	0.995	0.659	0.826
School	0.948	0.976	0.251	0.438
Shoe Store	0.916	0.966	0.153	0.648
Spa	0.784	0.940	0.182	0.297
Stadium	0.613	0.749	0.010	0.107
Storage	0.632	0.912	0.010	0.123
Train Station	0.099	0.414	0.047	0.087
Travel Agency	0.813	0.931	0.292	0.402
University	0.238	0.351	0.020	0.328
Veterinary Care	0.814	0.966	0.020	0.115
Zoo	0.343	0.680	0.001	0.011

Table A2
Akaike Information Criterion (AIC) values and Bayesian information Criterion (BIC) values of each model. The models with smaller AIC and BIC values are preferred.

	Inter-City Scale				Intra-City Scale			
	Size		Comp.		Size		Comp.	
	AIC	BIC	AIC	BIC	AIC	BIC	AIC	BIC
Accounting	387.1	389.0	610.2	615.7	7233.6	7240.7	5467.9	5630.0
Airport	283.6	285.4	534.2	536.0	−14564.4	−14557.4	−14565.1	−14480.5
Amusement Park	252.4	254.3	492.8	496.5	−6923.4	−6916.4	−6940.3	−6933.2
Aquarium	160.8	162.6	395.3	399.0	−24141.2	−24134.2	−23744.2	−23709.0
Art Gallery	404.3	406.2	600.4	605.9	15448.2	15455.3	13780.7	13893.4
Atm	458.2	460.1	614.1	617.8	14260.7	14267.7	12446.0	12664.5
Bakery	437.6	439.5	479.8	483.5	2507.6	2514.7	−349.5	−208.5
Bar	508.4	510.2	724.1	731.5	20416.0	20423.0	14125.5	14358.1

(*continued on next page*)

Table A2 (*continued*)

	Inter-City Scale				Intra-City Scale			
	Size		Comp.		Size		Comp.	
	AIC	BIC	AIC	BIC	AIC	BIC	AIC	BIC
Beauty Salon	471.4	473.3	625.0	628.7	19820.0	19827.0	16895.3	17113.8
Bicycle Store	268.8	270.6	283.5	285.4	−16203.0	−16196.0	−17083.1	−16970.3
Book Store	278.1	280.0	510.0	517.4	−7584.9	−7577.8	−8461.1	−8313.1
Bowling Alley	132.9	134.8	414.8	418.5	−28639.3	−28632.3	−28784.8	−28756.6
Bus Station	667.5	669.3	735.0	736.9	34335.6	34342.7	34766.9	34936.1
Cafe	443.1	445.0	461.9	465.6	4533.6	4540.6	1134.4	1338.8
Car Dealer	461.3	463.1	714.4	716.2	10190.0	10197.1	8802.5	8873.0
Car Rental	290.1	291.9	443.9	449.4	−3146.7	−3139.7	−3181.3	−3110.8
Car Repair	521.0	522.8	731.4	738.8	22908.0	22915.1	18230.6	18371.6
Car Wash	293.8	295.6	460.0	467.4	−11654.7	−11647.6	−11747.7	−11663.2
Casino	216.3	218.2	443.5	443.5	−35421.5	−35414.5	−35127.2	−35113.1
Cemetery	341.0	342.9	586.3	590.0	−21285.1	−21278.0	−21423.3	−21402.2
City Hall	76.2	78.0	293.2	295.0	−37437.7	−37430.7	−38618.9	−38562.5
Clothing Store	500.4	502.2	592.9	600.3	31184.7	31191.8	23911.6	24024.4
Construction Contractor	591.7	593.6	574.4	578.1	23556.0	23563.1	20067.1	20243.3
Convenience Store	455.2	457.1	570.9	572.8	1726.9	1733.9	2246.0	2394.0
Courthouse	160.9	162.8	404.9	406.8	−13810.3	−13803.3	−17901.4	−17788.6
Dentist	436.7	438.6	714.8	718.5	19519.7	19526.7	17163.9	17311.9
Department Store	327.4	329.2	510.3	514.0	−6448.4	−6441.4	−7219.3	−7064.3
Doctor	578.7	580.5	607.2	614.6	42180.3	42187.3	36517.8	36672.9
Electronics Store	386.1	388.0	587.0	590.7	3688.0	3695.0	2189.0	2322.9
Embassy	329.3	331.1	632.8	634.7	−2578.3	−2571.2	−3268.8	−3205.4
Finance	440.6	442.5	615.9	623.3	20231.0	20238.1	16860.1	17036.3
Fire Station	293.9	295.8	600.6	602.4	−17404.5	−17397.5	−17771.8	−17722.5
Florist	332.2	334.1	524.5	530.0	−4705.5	−4698.5	−5303.1	−5197.4
Funeral Home	336.3	338.1	568.3	570.1	−10028.7	−10021.7	−10844.2	−10703.3
Furniture Store	389.5	391.4	441.7	447.2	10673.1	10680.1	7599.0	7697.7
Gas Station	333.2	335.0	543.6	545.5	−13926.6	−13919.5	−11923.4	−11867.0
Grocery or Supermarket	467.7	469.6	592.6	596.3	9280.9	9288.0	6500.8	6677.1
Gym	321.2	323.1	463.7	469.2	−2721.9	−2714.9	−4013.8	−3837.6
Hardware Store	299.0	300.9	516.4	522.0	−8239.8	−8232.8	−9960.1	−9854.4
Home Goods Store	469.2	471.0	645.5	651.0	17169.9	17177.0	13170.9	13290.7
Hospital	310.3	312.1	428.4	433.9	11386.1	11393.2	5907.4	6027.2
Hotel and Lodging	413.6	415.5	734.4	736.3	12585.9	12592.9	10292.7	10483.0
Insurance Agencyncy	496.8	498.7	692.3	697.9	14861.7	14868.7	12397.0	12538.0
Jewelry Store	353.3	355.1	444.2	447.9	14860.2	14867.3	13143.0	13269.9
Laundry	400.6	402.4	566.7	570.4	4144.1	4151.2	2146.9	2316.0
Lawyer	479.9	481.8	728.9	730.8	38846.6	38853.6	35662.3	35831.4
Library	343.4	345.3	429.1	432.8	−5993.1	−5986.1	−8949.8	−8808.9
Liquor Store	405.4	407.2	632.6	634.4	−1736.3	−1729.2	−2355.6	−2242.8
Local Government Office	331.1	332.9	481.5	487.0	12849.6	12856.6	7505.0	7638.9
Locksmith	309.2	311.0	542.0	543.9	−14495.9	−14488.8	−14640.9	−14591.5
Movie_theater	223.5	225.4	352.4	356.1	−16822.2	−16815.1	−17422.3	−17337.7
Moving Company	443.6	445.5	628.4	632.1	300.1	307.2	−457.0	−372.4
Museum	318.3	320.1	385.0	392.4	−4793.4	−4786.3	−6985.0	−6872.2
Night Club	377.0	378.9	545.3	550.8	6321.7	6328.7	1774.4	1922.5
Park	504.8	506.7	683.4	685.3	11027.2	11034.2	10194.1	10363.3
Parking	372.0	373.9	596.1	599.8	5373.6	5380.6	1963.1	2153.4
Pet Store	285.8	287.6	452.0	455.7	−13001.7	−12994.6	−14005.4	−13885.6
Pharmacy	429.4	431.2	643.7	647.4	7366.7	7373.7	5035.9	5176.9
Physiotherapist	353.1	355.0	667.6	673.1	791.0	798.0	−544.4	−459.9
Police	252.8	254.6	349.2	352.9	−15255.4	−15248.4	−16701.6	−16602.9
Post Office	269.4	271.3	504.4	508.1	−12492.7	−12485.7	−12755.2	−12670.6
Real Estate Agency	515.7	517.6	668.7	672.4	20820.6	20827.7	19273.1	19449.4
Religious Centers	565.0	566.9	729.1	732.8	24793.2	24800.3	21728.3	21883.4
Restaurant	602.3	604.2	745.4	752.8	32182.1	32189.1	26651.6	26912.4
School	488.4	490.3	741.3	745.0	15330.2	15337.2	13283.1	13445.2
Shoe Store	363.4	365.2	607.3	611.0	18001.5	18008.6	11416.1	11528.9
Spa	307.5	309.4	456.4	460.1	−8683.6	−8676.6	−9852.6	−9732.7
Stadium	225.0	226.8	553.0	554.9	−13695.6	−13688.5	−13931.8	−13875.4
Storage	394.6	396.4	582.3	586.0	−6999.5	−6992.4	−7697.7	−7606.1
Train Station	334.7	336.6	545.1	547.0	−10105.8	−10098.8	−10424.3	−10389.0
Travel Agency	398.7	400.6	545.0	548.7	3926.0	3933.1	2523.5	2671.5
University	403.4	405.3	557.5	559.3	24047.2	24054.3	21500.1	21627.0
Veterinary Care	336.8	338.6	619.2	624.8	−3348.5	−3341.5	−3679.7	−3538.8
Zoo	70.8	72.7	144.4	148.1	−43402.7	−43395.6	−42907.2	−42879.0

References

Alabdulkareem, A., Frank, M. R., Sun, L., AlShebli, B., Hidalgo, C., & Rahwan, I. (2018). Unpacking the polarization of workplace skills. *Science Advances, 4*(7), Article eaao6030. https://doi.org/10.1126/sciadv.aao6030.

Alonso, L., Zhang, Y. R., Grignard, A., Noyman, A., Sakai, Y., Elkatsha, M., Doorley, R., & Larson, K. (2019). Data-driven, evidence-based simulation of urban dynamics. use case volpe. *Unifying Themes in Complex Systems, IX*, 4.

Been, V., Ellen, I. G., Gedal, M., Glaeser, E., & McCabe, B. J. (2016). Preserving history or restricting development? The heterogeneous effects of historic districts on local

housing markets in New York city. *Journal of Urban Economics, 92*, 16–30. https://doi.org/10.1016/j.jue.2015.12.002.

Berry, B. (1967). *Geography of market centers and retail distribution*. Prentice Hall. https://www.amazon.com/Geography-market-centers-retail-distribution/dp/B0000CNPHS.

Bettencourt, L. M. A., Lobo, J., Helbing, D., Kühnert, C., & West, G. B. (2007). Growth, innovation, scaling, and the pace of life in cities. *Proceedings of the National Academy of Sciences, 104*(17), 7301–7306. https://doi.org/10.1073/pnas.0610172104.

Bettencourt, L., & West, G. (2010). A unified theory of urban living. *Nature, 467*, 912–913. https://doi.org/10.1038/467912a.

Bloch, P. H., Ridgway, N. M., & Nelson, J. E. (1991). *Leisure and the shopping mall*. ACR North American Advances. NA-18 http://acrwebsite.org/volumes/7200/volumes/v18/NA-18.

Boschma, R. A. (2005). Does geographical proximity favour innovation? *Économie et institutions*, 6–7, 111–128 https://doi.org/10.4000/ei.926.

Boschma, R., Balland, P.-A., & Kogler, D. F. (2015). Relatedness and technological change in cities: The rise and fall of technological knowledge in US metropolitan areas from 1981 to 2010. *Industrial and Corporate Change, 24*(1), 223–250. https://doi.org/10.1093/icc/dtu012.

Brueckner, J. K. (1993). Inter-store externalities and space allocation in shopping centers. *The Journal of Real Estate Finance and Economics, 7*(1), 5–16. https://doi.org/10.1007/BF01096932.

Christaller, W., & Baskin, C. W. (1966). *Central places in southern Germany*. Prentice-Hall.

Cranshaw, J., Schwartz, R., Hong, J., & Sadeh, N. (2012). *The livehoods project: Utilizing social media to understand the dynamics of a city*. Sixth International AAAI Conference on Weblogs and Social Media.

De Nadai, M., Vieriu, R. L., Zen, G., Dragicevic, S., Naik, N., Caraviello, M., Hidalgo, C. A., Sebe, N., & Lepri, B. (2016). Are safer looking neighborhoods more lively?: A multimodal investigation into urban life. *Proceedings of the 24th ACM International Conference on Multimedia*, 1127–1135. https://doi.org/10.1145/2964284.2964312.

DiPasquale, D., & Wheaton, W. C. (1995). *Urban economics and real estate markets* (1 edition). Pearson.

Dong, L., Chen, S., Cheng, Y., Wu, Z., Li, C., & Wu, H. (2017). Measuring economic activity in China with mobile big data. *EPJ Data Science, 6*(1), 1–17. https://doi.org/10.1140/epjds/s13688-017-0125-5.

Dong, L., Ratti, C., & Zheng, S. (2019). Predicting neighborhoods' socioeconomic attributes using restaurant data. *Proceedings of the National Academy of Sciences, 116*(31), 15447–15452. https://doi.org/10.1073/pnas.1903064116.

Dudey, M. (1990). Competition by choice: The effect of consumer search on firm location decisions. *The American Economic Review, 80*(5), 1092–1104.

Eaton, B. C., & Lipsey, R. G. (1975). The principle of minimum differentiation reconsidered: Some new developments in the theory of spatial competition. *The Review of Economic Studies, 42*(1), 27–49. https://doi.org/10.2307/2296817. JSTOR.

Eaton, B. C., & Lipsey, R. G. (1982). An economic theory of central places. *The Economic Journal, 92*(365), 56–72. https://doi.org/10.2307/2232256. JSTOR.

Ehrenhalt, A. (2012). *The great inversion and the future of the American city*. Knopf Doubleday Publishing Group.

Ellison, G., & Glaeser, E. L. (1997). Geographic concentration in US manufacturing industries: A dartboard approach. *Journal of Political Economy, 105*(5), 889–927.

Eppli, M., & Benjamin, J. (1994). The evolution of shopping center research: A review and analysis. *Journal of Real Estate Research, 9*(1), 5–32. https://doi.org/10.5555/rees.9.1.l043247511363870.

Florida, R. (2019). *The rise of the creative class*. Basic books.

Frenken, K., Oort, F. V., & Verburg, T. (2007). Related variety, unrelated variety and regional economic growth. *Regional Studies, 41*(5), 685–697. https://doi.org/10.1080/00343400601120296.

Fujita, M., & Krugman, P. (2004). The new economic geography: Past, present and the future. In R. J. G. M. Florax, & D. A. Plane (Eds.), *Fifty years of regional science* (pp. 139–164). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-662-07223-3_6.

Fu, Y., & Somerville, C. T. (2001). Site density restrictions: Measurement and empirical analysis. *Journal of Urban Economics, 49*(2), 404–423. https://doi.org/10.1006/juec.2000.2200.

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

Guevara, M. R., Hartmann, D., Aristarán, M., Mendoza, M., & Hidalgo, C. A. (2016). The research space: Using career paths to predict the evolution of the research output of individuals, institutions, and nations. *Scientometrics, 109*(3), 1695–1709. https://doi.org/10.1007/s11192-016-2125-9.

Handy, S. L., & Niemeier, D. A. (1997). Measuring accessibility: An exploration of issues and alternatives. *Environment and Planning A: Economy and Space, 29*(7), 1175–1194. https://doi.org/10.1068/a291175.

Hanowell, B. (2017, August 7). Walkable neighborhoods aren't just for millennials. https://www.aplaceformom.com/blog/walkable-neighborhoods-for-seniors/.

Hausmann, R., Hidalgo, C. A., Bustos, S., Coscia, M., Simoes, A., & Yildirim, M. A. (2014). *The atlas of economic complexity: Mapping paths to prosperity*. MIT Press.

Hernández, T., & Bennison, D. (2000). The art and science of retail location decisions. *International Journal of Retail & Distribution Management*. https://doi.org/10.1108/09590550010337391.

Hidalgo, Balland, P.-A., Boschma, R., Delgado, M., Feldman, M., Frenken, K., Glaeser, E., He, C., Kogler, D. F., Morrison, A., Neffke, F., Rigby, D., Stern, S., Zheng, S., & Zhu, S. (2018). The principle of relatedness. In A. J. Morales, C. Gershenson, D. Braha, A. A. Minai, & Y. Bar-Yam (Eds.), *Unifying themes in complex systems IX* (pp. 451–457). Springer International Publishing.

Hidalgo, C. A., Klinger, B., Barabási, A.-L., & Hausmann, R. (2007). The product space conditions the development of nations. *Science, 317*(5837), 482–487. https://doi.org/10.1126/science.1144581.

Hise, R., Kelley, P., Gable, M., & Mc Donald, J. B. (1983). Factors affecting the performance of individual chain store units: An empirical analysis. *Journal of Retailing, 59*(2), 22–39.

Hotelling, H. (1929). Stability in competition. *The Economic Journal, 39*(153), 41–57. https://doi.org/10.2307/2224214. JSTOR.

Ingene, C. A. (1984). Structural determinants of market potential. *Journal of Retailing, 60*(1).

Jang, M., & Kang, C.-D. (2015). Retail accessibility and proximity effects on housing prices in seoul, korea: A retail type and housing submarket approach. *Habitat International, 49*, 516–528. https://doi.org/10.1016/j.habitatint.2015.07.004.

Jara-Figueroa, C., Jun, B., Glaeser, E. L., & Hidalgo, C. A. (2018). The role of industry-specific, occupation-specific, and location-specific knowledge in the growth and survival of new firms. *Proceedings of the National Academy of Sciences, 115*(50), 12646–12653. https://doi.org/10.1073/pnas.1800475115.

Kogler, D. F., Rigby, D. L., & Tucker, I. (2013). Mapping knowledge space and technological relatedness in US cities. *European Planning Studies, 21*(9), 1374–1391. https://doi.org/10.1080/09654313.2012.755832.

Krugman, P. (1993). On the number and location of cities. *European Economic Review, 37*(2), 293–298. https://doi.org/10.1016/0014-2921(93)90017-5.

Leinberger, C. B., & Lynch, P. (2014). Foot traffic ahead: Ranking walkable urbanism in America's largest metros. https://trid.trb.org/view/1314064.

Li, H., Wei, Y. D., & Wu, Y. (2019). Urban amenity, human capital and employment distribution in Shanghai. *Habitat International, 91*, 102025. https://doi.org/10.1016/j.habitatint.2019.102025.

Losch, A. (1954). *The economics of location: A pioneer book in the relations between economic goods and geography*. Yale University Press.

Maes, P. (1995). Agents that reduce work and information overload. In R. M. Baecker, J. Grudin, W. A. S. Buxton, & S. Greenberg (Eds.), *Readings in human–computer interaction* (pp. 811–821). https://doi.org/10.1016/B978-0-08-051574-8.50084-4. Morgan Kaufmann.

Marshall, A. (1890). *Principles of economics*. Macmillan and Company.

Mehaffy, M. W. (2008). Generative methods in urban design: A progress assessment. *Journal of Urbanism, 1*(1), 57–75.

Mulligan, G. F. (1984). Agglomeration and central place theory: A review of the literature. *International Regional Science Review, 9*(1), 1–42. https://doi.org/10.1177/016001768400900101.

Muneepeerakul, R., Lobo, J., Shutters, S. T., Goméz-Liévano, A., & Qubbaj, M. R. (2013). Urban economies and occupation space: Can they get "there" from "here"? *PloS One, 8*(9), Article e73676.

Nagy, D., Villaggi, L., & Benjamin, D. (2018). Generative urban design: Integrating financial and energy goals for automated neighborhood layout. *Proceedings of the Symposium on Simulation for Architecture and Urban Design, 25*.

Naik, N., Kominers, S. D., Raskar, R., Glaeser, E. L., & Hidalgo, C. A. (2017). Computer vision uncovers predictors of physical urban change. *Proceedings of the National Academy of Sciences, 114*(29), 7571–7576. https://doi.org/10.1073/pnas.1619003114.

Naik, N., Raskar, R., & Hidalgo, C. A. (2016). Cities are physical too: Using computer vision to measure the quality and impact of urban appearance. *The American Economic Review, 106*(5), 128–132. https://doi.org/10.1257/aer.p20161030.

Neffke, F., & Henning, M. (2013). Skill relatedness and firm diversification. *Strategic Management Journal, 34*(3), 297–316.

Neffke, F., Henning, M., & Boschma, R. (2011). How do regions diversify over time? Industry relatedness and the development of new growth paths in regions. *Economic Geography, 87*, 237–265.

Nelson, R. L. (1958). *The selection of retail locations*. F.W. Dodge Corp.

Nevin, J. R. (1980). *Image as a component of attraction to intraurban shopping areas*. Graduate School of Business, University of Wisconsin-Madison, [1980]. https://search.library.wisc.edu/catalog/999759587602121.

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. *Environment and Planning B: Urban Analytics and City Science, 46*(8), 1480–1498.

Rakha, T., & Reinhart, C. (2012). Generative urban modeling: A design work flow for walkability-optimized cities. *Proceedings of SimBuild, 5*(1), 255–262.

Resnick, P., & Varian, H. R. (1997, March 1). Recommender systems. Communications of the ACM. https://link.galegroup.com/apps/doc/A19284343/AONE?sid=lms.

Salesses, P., Schechtner, K., & Hidalgo, C. A. (2013). The collaborative image of the city: Mapping the inequality of urban perception. *PloS One, 8*(7), Article e68400. https://doi.org/10.1371/journal.pone.0068400.

Schuetz, J., Kolko, J., & Meltzer, R. (2012). Are poor neighborhoods "retail deserts"? *Regional Science and Urban Economics, 42*(1), 269–285. https://doi.org/10.1016/j.regsciurbeco.2011.09.005.

Schumacher, P. (2009). Parametricism: A new global style for architecture and urban design. *Architectural Design, 79*(4), 14–23.

Scitovsky, T. (2013). *Welfare & competition*. Routledge.

Sevtsuk, A. (2014). Location and agglomeration: The distribution of retail and food businesses in dense urban environments. *Journal of Planning Education and Research, 34*(4), 374–393. https://doi.org/10.1177/0739456X14550401.

Sonta, A. J., & Jain, R. K. (2019). Optimizing neighborhood-scale walkability. *Computing in civil engineering 2019: Data, sensing, and analytics*. VA: American Society of Civil Engineers Reston.

The National Trust for Historic Preservation. (2017). Reclaiming the past in bricks and mortar: New study reveals millennials' desire to connect with historic places |

national Trust for historic preservation. https://savingplaces.org/press-center/media-resources/new-study-reveals-millennials-desire-to-connect-with-historic-places.

You, H., & Bie, C. (2017). Creative class agglomeration across time and space in knowledge city: Determinants and their relative importance. *Habitat International, 60*, 91–100. https://doi.org/10.1016/j.habitatint.2016.12.010.

Youn, H., Bettencourt, L. M. A., Lobo, J., Strumsky, D., Samaniego, H., & West, G. B. (2016). Scaling and universality in urban economic diversification. *Journal of The Royal Society Interface, 13*(114), 20150937. https://doi.org/10.1098/rsif.2015.0937.