The Death and Life of Great Italian Cities: A Mobile Phone Data Perspective

Marco De Nadai
Fondazione Bruno Kessler
University of Trento
Trento, Italy
denadai@fbk.eu

Jacopo Staiano
Sorbonne Universités
Paris, France
jacopo.staiano@lip6.fr

Roberto Larcher
SKIL, Telecom Italia
Trento, Italy
roberto.larcher@telecomitalia.it

Nicu Sebe
University of Trento
Trento, Italy
sebe@disi.unitn.it

Daniele Quercia
Bell Labs
Cambridge, United Kingdom
quercia@cantab.net

Bruno Lepri
Fondazione Bruno Kessler
Trento, Italy
lepri@fbk.eu

ABSTRACT

The Death and Life of Great American Cities was written in 1961 and is now one of the most influential book in city planning. In it, Jane Jacobs proposed four conditions that promote life in a city. However, these conditions have not been empirically tested until recently. This is mainly because it is hard to collect data about "city life". The city of Seoul recently collected pedestrian activity through surveys at an unprecedented scale, with an effort spanning more than a decade, allowing researchers to conduct the first study successfully testing Jacobs's conditions. In this paper, we identify a valuable alternative to the lengthy and costly collection of activity survey data: mobile phone data. We extract human activity from such data, collect land use and socio-demographic information from the Italian Census and Open Street Map, and test the four conditions in six Italian cities. Although these cities are very different from the places for which Jacobs's conditions were spelled out (i.e., great American cities) and from the places in which they were recently tested (i.e., the Asian city of Seoul), we find those conditions to be indeed associated with urban life in Italy as well. Our methodology promises to have a great impact on urban studies, not least because, if replicated, it will make it possible to test Jacobs's theories at scale.

1. INTRODUCTION

By 2030, 9% of the world population is expected to live in just 41 *mega-cities*, each of which will accommodate more than 10M inhabitants[1]. Cities are essential crucibles for innovation, tolerance, novelty, and economic prosperity [4, 13], but they also face challenges in terms of, for example, traffic, pollution, poverty, and criminality [13].

---

[1] http://www.economist.com/node/21642053

A fundamental research question that urban planners, sociologists and economists are investigating is what creates urban life. The urban sociologist Jane Jacobs in *The Death and Life of Great American Cities* [19], one of the most influential books in city planning, introduced the urban physical environment (the *urban fabric*) as an essential factor for urban vitality [19]. As the book title says, Jacobs dealt with the death and life of American cities. In its most basic form, her argument was that death was caused by elimination of pedestrian activity (e.g., by highway construction, large-scale development projects), and that life was created by the presence of pedestrians at all times of the day.

She argued that, to promote urban life in large cities, the physical environment should be characterized by *diversity* at both the district and street level. Diversity, in turn, requires four essential conditions: (i) *mixed land uses*, that is, districts should serve more than two primary functions, and that would attract people who have different purposes; (ii) *small blocks*, which promote contact opportunities among people; (iii) *buildings diverse in terms of age and form*, which make it possible to mix high-rent and low-rent tenants; and (iv) sufficient *dense concentration* of people and buildings.

Despite their importance, those conditions have not been empirically tested all together until recently, mainly because it is hard to collect data about which neighborhoods have full urban life and which have little of it. After an effort lasting more than a decade, the city of Seoul recently managed to collect pedestrian activity through surveys, and local researchers performed the first study testing Jacobs's conditions in the city [40]. The researchers found her claims to hold in Seoul too: mixed use, old buildings, and density all contributed to urban life.

To complement the lengthy and costly survey-based collection of activity data, this work put forward the use of mobile phone data. We extracted human activity measurements from mobile phone records in six Italian cities - Bologna, Florence, Milan, Palermo, Rome, and Turin. These cities are very different from the places in which Jacobs's conditions were spelled out (i.e., great American cities) and from the places in which they were recently tested (i.e., the Asian city of Seoul). Despite that, our results showed that the four conditions for promoting urban life do hold for Italian

cities as well. To produce those results, we made three main contributions:

- Building upon previous works [19, 40], we proposed a set of metrics for *urban diversity* derived from open, aggregated, and anonymous map data, and a metric for *urban vitality* derived from mobile phone activity (Section 5);

- We provided a comprehensive assessment of the predictive power of Jacobs's diversity conditions in the Italian context (Section 6);

- Finally, we identified our work's theoretical and practical implications, and its limitations (Section 7).

2. RELATED WORK

Our work is best placed in an emerging interdisciplinary field called "urban computing" [42]. This combines computer science approaches with more traditional fields like urban planning, urban economy, and urban sociology.

The idea of testing urban theories using novel sources of data (e.g., social media, online images and videos, mobile phone data) has received increasing attention [35, 36, 37]. The urban sociologist Kevin Lynch showed that people living in an urban environment create their own personal "mental map" of the city based on features such as the areas they visit and the routes they use [26]. Hence, he hypothesized that, the more recognizable a city, the more navigable the city. To test Lynch's theory, Quercia *et al.* built a web game that crowd-sourced Londoners' mental images of the city [36]. They showed that areas suffering from social problems such as poor living conditions and crime are rarely present in residents' mental images.

Researchers also investigated which urban elements people use to visually judge a street to be safe, wealthy, and attractive using web crowdsourcing games [32, 35, 37], and they also studied how to identify walkable streets using the social media data of Flickr and Foursquare (e.g., unsafe streets tended to be photographed during the day, while walkable streets were tagged with walkability-related keywords [34]).

The recent availability of large-scale data sets, such as those automatically collected by mobile phone networks, opens new possibilities for studying city dynamics at finer and unprecedented granularity [5]. Mobile phone data represents a highly valuable proxy for human mobility patterns [14, 22, 39]. Such data was recently used to map functional uses [16, 23], to identify places that play a major role in the life of citizens [18], to compare cities based on their spatial similarities and differences [24], and to predict socio-economic indicators [11, 38], including crime [6, 7, 41].

Researchers recently tested Jane Jacobs's observations about crime [6, 7, 41]. They verified her *natural surveillance* theory. She posited that people diversity and presence of visitors in a neighborhood contribute to natural surveillance and, as such, discourage crime: as people are moving around an area, they effectively become "eyes on the street" able to observe what is going on around them.

To go beyond crime, we set out to conduct a comprehensive validation of Jacobs's four conditions for urban life. To do that, we gather several sets of data.

3. SETS OF DATA

Mobile Phone Activity. Every time a mobile phone communicates, a radio base station delivers the communication through the network, and a new call data record is created. This record reports the time of the communication, and the radio base station that handled it. Different types of call records exist, recording various types of communication (e.g., incoming calls, Internet, outgoing SMS). We decided to focus on Internet activity, since it allows for the passive reconstruction of people's mobility: even in the absence of direct user activity (e.g., making/receiving a call, receiving/sending an SMS), mobile phones can be tracked since they are likely to be connected to the Internet for background traffic and push notifications.

Call records have been provided by the Semantics and Knowledge Innovation Lab of Telecom Italia Mobile, which is the largest mobile operator in Italy with 34% of the entire market share[2]. Our data is aggregated every 60 minutes, comes from both TIM customers and roaming customers in the six cities, and covers the time ranging from February to October 2014.

OpenStreetMap. OpenStreetMap[3] (OSM) is a global project that aims to collaboratively create a detailed map of the world. With more than 2.1 million contributors, it has become the most valuable and openly accessible source of information about the physical world. OSM data is composed by three data primitives: *nodes*, *ways* and *relations*. *Nodes* define points in space like Point Of Interests (which we call "places"), *ways* define roads and other linear features, and *relations* reflect how those elements are related to each other. Since we focused on the geographic unit of a *block* (i.e., the smallest area that is surrounded by street segments), we filtered out suburban elements: *ways* with tag `highways` using Overpass API[4], `footpaths`, `primary` roads, and `proposed` roads.

Census Data. The Italian National Institute for Statistics (ISTAT) provides, under open data licences, information gathered via the 2011 Italian Census[5]. We used the census' variables related to people and buildings (Table 1), which were defined at various geographic levels.

Land Use. Urban land-cover and land-use for Large Urban Zones (LUZs) is mapped by the Urban ATLAS European project[6]. This project exploits satellite images to categorize the city in 20 classes (e.g., agricultural areas, continuous urban fabric) with a precision between 0.25 and 1 hectares, and accuracy above 80%. The dataset is built and validated for 2006, while its version updated to 2012 will be released at the end of 2015. However, for Turin and Rome, the new data is already available, and we used it.

Infrastructures. In addition to the Urban ATLAS dataset, we used the ISTAT statistical atlas of infrastructures (*Atlante*

ISTAT code	Description
P1	number of residents
E2	number of used buildings
E3	number of residential buildings
E8-E16	number of buildings in 9 age bands in the range [1919, 2011]
E17-E20	number of buildings of 4 types based on their number of floors
E21-E26	number of buildings of 6 types based on their internal apartments
ADDETTI	number of employees

Table 1: ISTAT open data variables under study. The last variable comes from the census of industry and services (*censimento dell'industria e dei servizi*), while the rest come from the census of population and housing (*censimento della popolazione e delle abitazioni*).

statistico delle infrastrutture)[7], which provides details on the logistic facilities (e.g., presence of railways) in Italy.

Foursquare Data. Created in 2008, Foursquare is the world leading location-based social network. It has attracted more than 30 million users[8]. Through a gamification system, users are encouraged to *check-in* their geographical position into the social network and sharing the places they visit (*venues*). Venues are places whose geographical information is enriched with semantic labels specifying details about them. Three hierarchical levels for place categories exist in Foursquare: an abstract level grouping *venues* in macro categories (e.g., "food"); a more specific one (e.g., "restaurant"); and a third highly detailed level (e.g., "Thai restaurant")[9]. We extracted Foursquare *venues* from the public API[10] for the categories in Table 2.

Group	Foursquare category
NightLife	Brewery, Champagne Bar, Cocktail Bar, Dive Bar, Gay Bar, Hookah Bar, Lounge, Night Market, Night club, Other Nightlife, Pub, Strip Club, Whisky Bar, Wine bar, Nightlife Spot
Art-night	Laser Tag, Movie Theater, Music Venue, Performing Arts
Services	Medical Center, Library, Government Building, Military Base, Post Office
Eating-drinking	Food
Org. activity	Comedy Club, Country Dance Club, Salsa Club, Club, Community Center, Cultural Center, Library, Social Club, Spiritual
Outside	Park, Plaza, Pedestrian Plaza
Commercial	Shop & Service (excluding ATM, Construction, EV Charging, Gas Station / Garage, Newsstand)

Table 2: The Foursquare categories of the venues under study.

---

4. DEATH AND LIFE OF CITIES

In *The Death and Life of Great American Cities* [19], Jane Jacobs argued that "death" in urban areas is caused by the elimination of pedestrian activity from urban streets. By contrast, urban life is created by a continual *sidewalk ballet* consisting of a street filled with pedestrians at all times of the day. In her own words, "a well-used city street is apt to be a safe street and a deserted city street is apt to be unsafe" [19]. This means that streets must be used fairly continuously during the day, both by residents and by strangers. These "eyes on streets" generate a virtuous loop which, in turn, increases public safety, creates diverse face-to-face interactions, and contributes to the local economy.

Jacobs spelled out four diversity requirements essential for the generation and maintenance of urban life. First, a district should serve more than one primary function, preferably more than two. This attracts people with diverse purposes who end up sharing common facilities. Moreover, when primary uses are combined in a way that attracts people at different times of the day, then that positively impacts the local economy. For instance, a neighbourhood where people are present only during office hours is likely to provide only a few, if any, leisure facilities: it would be neither efficient nor attractive for residents. Conversely, should a secondary function flourish by, e.g., the presence of theaters, museums or night-life places, residents would likely benefit.

Second, Jacobs argued that street blocks must be short. She observed that large rectangular blocks are apt to thwart the effective mixture of use and people, resulting into paths that meet each other too infrequently. By contrast, smaller blocks provide more intersections, and thus slow down automobiles, which, she argued, are inclined to destroy urban vitality by discouraging the presence of pedestrian activity.

Third, in Jacobs's view, the buildings in a district should be diverse in terms of age and form. This ensures diversity of economic activity. Diverse buildings make it possible to have a mixture of jobs (i.e., high-, medium- and moderate-income jobs) and a mixture of tenants (e.g., high-rent and low-rent tenants) [12, 15, 19, 21]. Her view came from observing large-scale projects in the New York of her time: she saw that large-scale buildings did not change over time and hardly adapted to the environment that surrounded them; by contrast, old buildings helped to cultivate new primary uses in the neighborhood (e.g., new companies were likely to start and initially grow in old and low-rent buildings).

The fourth and final condition is about dense concentration of people and buildings. The idea is that density fosters "lively" district that are able to attract people for different purposes.

Jacobs emphasized that *all* four factors are necessary: density alone cannot create urban diversity, and mixed-land use would not flourish in areas with big blocks and of low density.

Further, in addition to those four conditions, Jacobs talked about 'vacuums', which are patches of land dedicated to a single use (e.g., transport facilities, parks). Those elements might be good or bad. Small railway stations, bus stops and small parks may act as hubs for the pedestrian activity; but, at the same time, if not well managed, a park could be used only at a certain time of the day, exposing it to deprivation and criminality.

City	#Districts	Size (avg)	Population (avg)
Bologna	23	3.34	15,918
Florence	21	2.89	16,633
Milan	85	1.72	14,551
Palermo	43	2.01	15,075
Rome	146	3.24	17,312
Turin	56	2.00	15,543

Table 3: Number, average size (in km^2), and population of the districts in our six cities.

5. METHODOLOGY

Our approach leverages mobile Internet activity from call data records to extract proxies for *urban vitality*, and web data from public entities (e.g., national census) and commercial ones (e.g., Foursquare) to extract *urban diversity* as per Jacobs's four conditions. Our goal is to study the relationship between urban vitality and diversity. Next, we spell out the metrics and the regression models we used to meet that goal.

In the 2011 Italian census, the smallest administrative unit of cities is represented by the census sections (*sezioni di censimento*), also called *blocks*. These sections are delimited by street segments and are grouped with other nearby sections to form a census area (*area di censimento*) called *district*. The boundaries of a district are drawn on the basis of socio-economic conditions in a way that districts can be compared with each other in terms of, for example, population: districts have between 13,000 and 18,000 inhabitants[11]. This is a loose constraint, however: city administrations may ignore it and define their own districts. Empirically, we verify that, in our data, a district covers an average area of 2.47 km^2 with an average population density of 10,000 people per km^2 (Table 3). Jacobs did not define any strict criterion concerning district size: she simply proposed that the edge of an administrative district should not exceed 2.4 km, and that each district should have a minimum population of 50,000 people. Compared to that definition, our districts are similar in terms of area and smaller in terms of population.

5.1 Jacobs's metrics

We extracted several variables to quantify the four conditions (Table 4), and we will detail them next.

5.1.1 Land use

The first of Jacobs's four conditions is to have mixed primary uses in a district [19]. Primary categories include residential buildings, offices, industrial facilities, entertainment places, education facilities, recreation facilities, museums, libraries, and galleries.

We computed Land Use Mix (LUM) [10] in district i as:

$$\text{LUM}_i = -\sum_{j=1}^{n} \frac{P_{i,j} \log(P_{i,j})}{\log(n)} \quad (1)$$

where $P_{i,j}$ is the percentage of square footage with land use j in district i, and n is the number of possible land uses (in our case, $n = 3$). If district i's land is dedicated

---

Land use	Distribution	μ
(1) Land use mix		0.73
(2) Closeness to small parks (SPs)		1×10^{-3}
(3) Residential *vs.* Non-Res.		0.67
(4) Housing types		4.98
(5) Commercial		0.30
(6) Nightlife		0.10
(7) Nightlife density		1×10^{-5}
(8) Daily		0.02
(9) 3^{rd}Places		1×10^{-4}
Small blocks		
(10) Block area		9.61
(11) Intersections density		1×10^{-4}
(12) Anisotropicity		0.38
Aged buildings		
(14) building age		50.34
(16) $\sigma_{\text{building age}}$		12.96
(17) Employees per company		6.98
Concentration		
(18) Population density		0.01
(19) Employment density		5×10^{-3}
(20) $\frac{\text{population density}}{\text{employee density}}$		3.72
(21) $\frac{\text{|internal apartments|}}{\text{|buildings|}}$		15.31
(22) Density of daily places		5×10^{-3}
(23) Density non-daily places		3×10^{-3}
Vacuums		
(24) Closeness to large parks		3×10^{-3}
(25) Closeness to railways		1×10^{-3}
(26) Closeness to highways		6×10^{-4}
(27) Closeness to water		1×10^{-3}
Mobile phone activity		
(28) Activity density		3.84

Table 4: Urban diversity metrics plus mobile Internet activity. Most variables are not normally distributed. Density measures are computed with surface area in m^2, and closeness measures are the inverse of the distance ($1/m$).

to one use only, then LUM_i is zero; instead, if the land is used equally in all n ways, then LUM_i is one. The higher LUM_i, the more mixed i's land use. We defined the three land uses following the recommendations in [28]: the first land use is "residential"; the second includes the categories "commercial", "institutional and governmental", and "resource and industrial"; and the third includes "park and recreational" and "water".

Since well-managed parks might well function as hubs for pedestrian activity, we defined the average distance from the nearest small park (area smaller than 1 km^2) for each district

---

[11]We exclude districts that are not densely populated (e.g. the national park in Rome, the volcano in Naples) These are marked by the Italian census with an identifier `ace` equal to zero.

i:

$$\text{Closeness to SM}_i = (\frac{1}{|B_i|}\sum_{j \in B_i}\text{dist}(j, \text{closest}(j, SM)))^{-1} \tag{2}$$

where B_i is the set of blocks in district i, $\text{closest}(j, Y)$ is a function that finds the geographically closest element in set Y from block j's centroid, SM is the set of small parks, and $\text{dist}(a, b)$ is the geographic distance between two elements' centroids a and b.

We also computed the Residential/Non-Residential (RNR) balance in district i as:

$$\text{RNR}_i = 1 - |\frac{Res_i - NonRes_i}{Res_i + NonRes_i}| \tag{3}$$

where Res_i is the area occupied by residential buildings in district i, and $NonRes_i$ is the area occupied by non-residential ones. The higher RNR_i, the more balanced the district in terms of residential *vs.* non-residential uses.

To go beyond horizontal land use and look at vertical development, we computed the average number of floors per building in district i and called it 'housing types' (as [40] did):

$$\text{Housing types}_i = \frac{\sum_c h_{c,i} \cdot z_c}{\sum_c h_{c,i}} \tag{4}$$

where $h_{c,i}$ is the number of buildings that are in height category c in district i, and z_c is the number of floors corresponding to height category c. The sums were repeated over all height categories (i.e., over the categories `E17-E20` in Table 1).

The previous definitions have characterized spatial use in terms of land use and building use. However, activities are important too. Jacobs argued for mixing primary uses so that people are on the street at different times of the day. To characterize spatial use in terms of activities, upon Foursquare data, we determined whether each place is used daily (e.g., convenience stores, restaurants, sport facilities) or not, and whether it is used at night or daylight [40]. Based on that, we defined:

$$\text{Commercial}_i = \frac{|\text{non daily-use places}_i|}{|\text{places}_i|} \tag{5}$$

$$\text{Nightlife}_i = \frac{|\text{nightlife places}_i|}{|\text{places}_i|} \tag{6}$$

$$\text{Nightlife density}_i = \frac{|\text{nightlife places}_i|}{\text{area}_i} \tag{7}$$

$$\text{Daily}_i = (\frac{1}{|B_i|}\sum_{j \in B_i}\text{dist}(j, \text{closest}(j, D)))^{-1} \tag{8}$$

where D is the set of places that are used daily, and B_i is, again, the set of street blocks in district i.

Also, not all activities are equal. There are activities that are more 'social' than others. To capture that, we resorted to the concept of *third places*. These are defined by Oldeburg [30] as the "great, good places" that foster community and communication among people outside home (the first place) and work (the second place): "they are places where people gather primarily to enjoy each others' company". Third Places function as unique public spaces for social interaction, providing a context for sociability, spontaneity, community building and emotional expressiveness [31]. Therefore,

we computed:

$$3^{rd}places_i = \frac{|3^{rd}\text{places}_i|}{|\text{places}_i|} \tag{9}$$

We determined whether a place is a third place or not by following the 4-category classification proposed by Jeffres *et al.* [20]. Third places fall into these four categories: *eating, drinking and talking* (e.g., coffee shops, bars, pubs, restaurants, and cafes); *organized activities* contributing to social capital [33] (e.g., places of worship, clubs, organizations, community centers, and senior centers); *outdoor* (e.g., plazas and parks); and *commercial venues* (e.g., stores, malls, shopping centers, markets, beauty salons, and barber shops).

5.1.2 Small blocks

Jacobs listed the presence of small blocks as the second necessary condition for diversity. Small blocks are believed to support stationary activities and provide opportunities for short-term and low-intensity contacts, easing interactions with other people in a relaxed and relatively undemanding way. Specifically, she stated that "lowly, unpurposeful and random as they may appear, sidewalk contacts are the small change from which a city's wealth of public life may grow". She criticized super-blocks and rectangular blocks, which constrain urban mobility with high travel distances and limited opportunities of cross-use.

The easiest way to identify small blocks is to compute the average block area among the set B_i of blocks in district i:

$$\overline{\text{Block area}}_i = \frac{1}{|B_i|}\sum_{j \in B_i}\text{area}_j \tag{10}$$

Since a district with high intersections density is likely to contribute to random contacts, we also computed:

$$\text{Intersection density}_i = \frac{|\text{intersections}_i|}{\text{area}_i} \tag{11}$$

Finally, since block size is distributed as a power law $P(A) \sim \frac{1}{A^r}$ with $r \sim 2$, we characterized a district i by its average shape anisotropicity [25] of the blocks B_i within it:

$$\text{District anisotropicity}_i = \frac{1}{|B_i|}\sum_{j \in B_i}\Phi_j \tag{12}$$

where Φ_j is the ratio between the area of the block j and the area of its circumscribed circle $\mathcal{C}_j$:

$$\Phi_j = \frac{\text{area}_j}{\text{area}_{\mathcal{C}_j}} \tag{13}$$

The quantity Φ_j is always smaller than one, and the larger its value, the less anisotropic block j, the more opportunities for contacts the block creates.

All those metrics were computed using the district *net* area $area_i$, which excludes unpopulated patches such as rivers and natural parks.

5.1.3 Aged buildings

Jacobs stressed the importance of having old buildings in a district. If a district has only new buildings, then it would have only enterprises that can support the high costs of new constructions. If it has old buildings too, instead, it would be able to incubate new small enterprises that cannot afford high rents, and that will benefit the local economy in the long run: "If the incubation is successful enough, the yield of the building can, and often does, rise" [19].

As a first measure, we computed the average building age in district i:

$$\overline{\text{Building age}_i} = \frac{\sum_b |\text{building}_{b,i}| \cdot \text{old}_b}{\sum_b |\text{building}_{b,i}|} \qquad (14)$$

This measures how old, on average, a building in district i is. Since age is expressed in age bands in the Italian census (the first band is E8 in Table 1, and the last is E16), we needed to compute the temporal length of each band. More specifically, for each band b in the range $[\text{start}_b, \text{end}_b]$, we computed:

$$\text{old}_b = \frac{(\text{last} - \text{start}_b) + (\text{last} - \text{end}_b)}{2} \qquad (15)$$

where last is the year of the latest built building in the whole census data. Then, in equation (14), we weighted age band b's temporal length (old_b) by the number of buildings in that band ($|\text{building}_{b,i}|$), and that gave us the average age of the district's buildings.

We also computed the corresponding standard deviation as:

$$\sigma_{\text{building age}_i} = \frac{\sum_b |\text{building}_{b,i}|^2}{\left(\sum_b |\text{building}_{b,i}|\right)^2} \cdot \sigma^2 \qquad (16)$$

where σ^2 is the standard deviation of the number of buildings in each band. The higher the standard deviation, the higher the district's heterogeneity in terms of building age.

Finally, we computed the average number of employees per company in district i:

$$\overline{\text{Employees per company}_i} = \frac{1}{|F_i|} \sum_{j \in F_i} \text{employees}(j) \qquad (17)$$

where F_i is the set of companies in district i.

5.1.4 Concentration

Jacobs's fourth and final condition is about having concentration of (both residential and non-residential) buildings and of people. Similarly to Sung *et al.* [40], we computed two sets of concentration measures: one for people, and the other for buildings.

First, population and employment density measures were calculated dividing the number of people (employees) by the district's net area.

$$\overline{\text{Population density}_i} = \frac{|\text{Population}_i|}{\text{area}_i} \qquad (18)$$

$$\overline{\text{Employment density}_i} = \frac{|\text{Employed people}_i|}{\text{area}_i} \qquad (19)$$

We then computed the interaction term between those two measures as the ratio between population density and employment density:

$$\frac{\text{Population density}_i}{\text{Employment density}_i} \qquad (20)$$

The higher it is, the more residents (as opposed to employees) in the district. Different values correspond to districts with different social textures.

To add to those people-based concentration measures one building-based measure, we computed:

$$\text{internal}_i = \frac{|\text{internal apartments}_i|}{|\text{buildings}_i|} \qquad (21)$$

which is the average number of apartments per building.

Finally, as we have done previously, to go beyond people and buildings and look at activities, we computed:

$$\text{Density of daily places}_i = \frac{|\text{daily-use places}_i|}{\text{area}_i} \qquad (22)$$

$$\text{Density of non-daily places}_i = \frac{|\text{non-daily-use places}_i|}{\text{area}_i} \qquad (23)$$

These two quantities are not totally anti-correlated to each other since not all places can be classified as being fully daily *vs.* non-daily.

5.1.5 Vacuums

Border vacuums are places that act as physical obstacles to pedestrian activity. For instance, parks can be a hub of pedestrian activity, if efficiently managed [19], but they could also be deplorable places in which criminality flourishes (especially at night). In a similar way, the proximity to expressways may discourage pedestrian activity or may effectively connect different parts of the city. This is what Jacobs called "the curse of border vacuums".

Therefore, we needed to identify large areas with single use. From OSM, we took parks[12], and, from the Urban ATLAS, we took railway areas, fast transit zones, rivers, lakes, and highways. That extraction allowed us to build the sets of large parks LP, railways R, highways H and water areas W.

To verify the impact of a type of vacuum area, say, that of large parks on a district, we calculated the average distance between a district's block (i.e., smallest area surrounded by street segments) and its closest large park:

$$\text{Closeness to LP}_i = \left(\frac{1}{|B_i|} \sum_{j \in B_i} \text{dist}(j, \text{closest}(j, LP))\right)^{-1} \qquad (24)$$

where $\text{dist}(j, \text{closest}(j, LP)$ is the distance between block j and its closest large park. The sum of distances is done over all blocks in district i (B_i is indeed the set of district i's blocks).

The next type of vacuum area is that of railways. Our dataset did not differentiate between rails and stations. However, in terms of pedestrian activity, the two differ: often, rails act as obstacles of activity, while stations act as hubs. To incorporate that distinction, we excluded the stations (i.e., the 600-meter buffer areas around them) from the set R of railways, and computed:

$$\text{Closeness to R}_i = \left(\frac{1}{|B_i|} \sum_{j \in B_i} \text{dist}(j, \text{closest}(j, R))\right)^{-1} \qquad (25)$$

In a similar way, we computed the average distance between a district's block and its closest highway, and its closest water area:

$$\text{Closeness to H}_i = \left(\frac{1}{|B_i|} \sum_{j \in B_i} \text{dist}(j, \text{closest}(j, H))\right)^{-1} \qquad (26)$$

$$\text{Closeness to A}_i = \left(\frac{1}{|B_i|} \sum_{j \in B_i} \text{dist}(j, \text{closest}(j, A))\right)^{-1} \qquad (27)$$

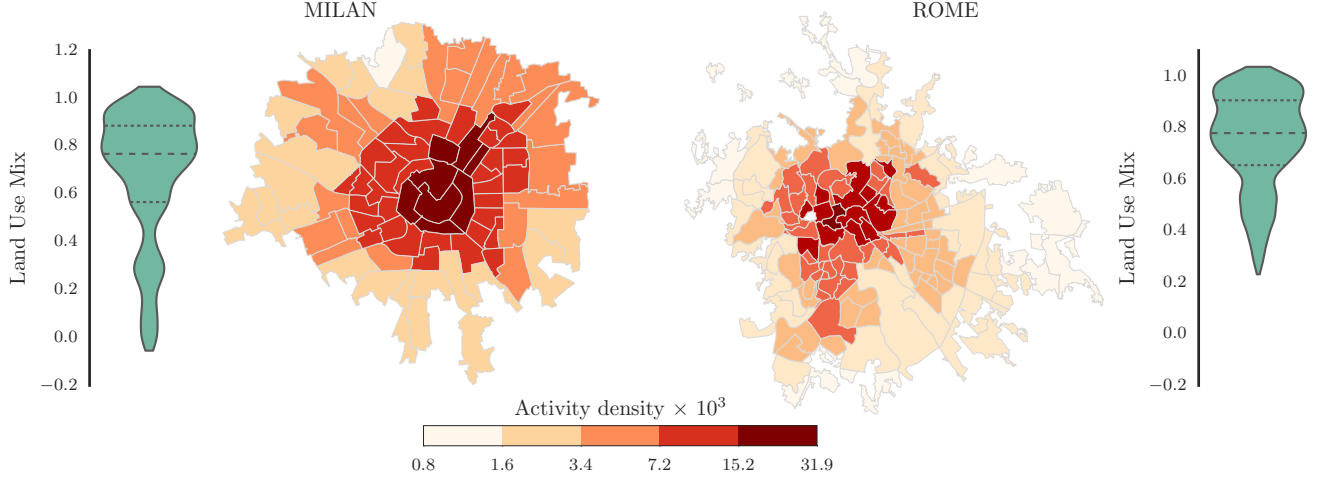[12]We took areas identified by `leisure=park`, filtering out the parks with `barrier`.

Figure 1: District activity density in Milan (left) and Rome (right), and their corresponding values of mixed land use (LUM). The colors are in logarithmic scale.

where H and A are the sets of highways and of water areas.

5.2 Activity density

As it has been extensively done in previous work, we used mobile Internet activity as a proxy for *urban vitality*.

To roughly estimate the number of Internet connections that fell into each district, we needed to represent the urban space somehow. We chose to represent it as a set of 2-dimensional, non-overlapping, and non-convex polygons. These polygons come from a Voronoi tessellation based on the radio stations' positions.

Then, to estimate the number of Internet call records $S_i(t)$ in each district i at time t, we counted the number of Internet connections over all polygons v's that fell into district i:

$$S_i(t) = \sum_v R_v(t) \frac{A_{v \cap i}}{A_v - A_{v \cap W}} \qquad (28)$$

where v is a polygon, and $R_v(t)$ is the number of Internet connections in v at time t. The count of Internet connections is weighted by $\frac{A_{v \cap i}}{A_v - A_{v \cap W}}$, which is the proportion $\frac{A_{v \cap i}}{A_v}$ of v's area that falls into district i ($A_{v \cap i}$ is v's area that falls into district i, and A_v is v's total area). From v's total area we removed sea areas denoted by W (i.e., we removed $A_{v \cap W}$).

Then, having $S_i(t)$, we computed a district's *activity density* as the average number of Internet connections throughout a typical business day, divided by the district's area. The results for Milan and Rome are shown in Figure 1. The normalization of surface area makes it possible to compare the activities of districts of different sizes.

5.3 The regression model

We used an Ordinary Least Squares (OLS) model to evaluate the relationship between each district's structural diversity (Section 5.1) and the district's activity density (Section 5.2).

Since most of the regression variables were skewed, we transformed them. More specifically, we log-transformed activity density using the natural logarithm, and transformed the structural diversity metrics using the Box-Cox method [8]. To avoid over-fitting, we did split the data into training set

(75%) and test set (25%), and repeated our measurements 1000 times using a shuffle split cross-validation.

We created six linear models. First, we created five models, each of which had *one* of the five sets of Jacobs's metrics as independent variables (i.e., we separately analyzed land use, small blocks, aged buildings, concentration, and vacuums). Those models always had activity density as dependent variable. The resulting coefficients are shown in the first five columns of Table 5. We then had a combined model by selecting the independent variables through both recursive feature elimination and stability selection [29]. The resulting selected variables and corresponding coefficients are shown in the last column of Table 5.

6. RESULTS

6.1 Land use

Mixed land use matters only in cities in which functional uses were historically separated like in the case of Milan. By looking at the first two predictors of land use in Table 5, one learns that mixed land use does not contribute neither positively nor negatively. That is because not all cities are equal: land use in Rome is mixed (Figure 1), while Milan is separated in functional areas. Consequently, in Milan, vitality is experienced only in the mixed districts. However, the highest beta coefficient concerning land use is found to be related to the presence of third places ($\beta = 0.3972$). Daily errands are important, but third places are more so (with a Spearman's correlation of 0.8337 with activity density). These places are public places in which people can hang out for good company and lively conversations, putting aside the concerns of home and work (their first and second places) [30]. Examples of such places include pubs, coffee shops, and taverns.

6.2 Small blocks

Continental European cities do not have the super-blocks typical of American cities. This is especially true in Italy, so it comes at no surprise that block size does not greatly matter ($\beta = 0.0615$ in model "small blocks" of Table 5). By

	Land use	Small blocks	Aged buildings	Concentration	Vacuums	Combined
$AdjR^2$	0.70	0.63	0.40	0.73	0.19	0.77
Land use mix (1)	0.0236					
Closeness to small parks (SPs) (2)	0.0409					
Residential *vs.* Non-Res. (RNR) (3)	-0.0180					
Housing types (4)	0.1959***					0.1854***
Commercial (5)	0.0138					
Night-life (6)	0.0110					
Night-life density (7)	-0.1107*					
Daily (8)	0.0440					
3^{rd}Places (9)	0.3972***					
Block area (10)		0.0615				
Intersections density (11)		0.7380***				0.1914***
Anisotropicity (12)		0.0462				
$\overline{\text{building age}}$ (14)			0.3689***			
$\sigma_{\text{building age}}$ (16)			0.2264***			
Employees per company (17)			-0.0778			
Population density (18)				-0.1480		
Employment density (19)				0.7401***		0.4344***
$\frac{\text{population density}}{\text{employee density}}$ (20)				0.3020*		
$\frac{|\text{internal apartments}|}{|\text{buildings}|}$ (21)				0.0220		
Density of daily places (22)				0.2046**		
Density non-daily places (23)				0.1061		
Closeness large parks (24)					0.1384**	
Closeness railways (25)					0.1745***	
Closeness highways (26)					-0.3457***	-0.1018***
Closeness water (27)					0.0991	
3^{rd}Places × Closeness to highways						0.0681**
Closeness to SPs × Closeness to highways						-0.0788**

Table 5: Linear regression models that predict district activity density. Each column is a different model. For each model's column, the table reports the model's predictive power (Adjusted R^2 in the first row) and the β coefficients. Blank cells correspond to variables that are absent from the model. As for statistical significance, we use the following notation: *p $<$ 0.05, **p $<$ 0.01, ***p $<$ 0.001.

.

contrast, the density of intersections does matter: vibrant urban areas are those with dense streets, which, in fact, slow down cars and make it easier for pedestrian to cross [9], creating what Jane Jacobs called the "sidewalk ballet". This 'ballet' goes beyond simple pedestrian activity. It is about informal contacts and public trust (e.g., children playing on the street, adults other than their parents paying attention to them, leaving home keys with shopkeepers). To increase the potential for such types of interactions, it is essential to be able to look into each others' eyes, and only small streets could foster that.

6.3 Aged buildings

In large American cities, Jacobs observed that neighborhoods with aged buildings tended to have a diversified local economy. By mingling building of varying age and conditions (i.e., mixing expensive building with "a good lot of plain, ordinary, low-value old buildings, rundown old buildings"), a neighborhood attracts not only standardized and high-earning enterprises but also ordinary and innovative ones. Aged buildings were a rarity in American cities. Back in her days, Jacobs noted: "in Miami Beach, where novelty is the sovereign remedy, hotels ten years old are considered aged and are passed up because others are newer" [19]. Decisions about which buildings get to stay and which get to go depend on culture: they are influenced by the relationship that a culture has with time and place. In a neighborhood, the patina of time may be *retained*, *imitated*, or *removed*, as Kevin Lynch puts it [27]. In Western Europe, the idea of preservation appeared about 1500 and, since then, retention is by far preferred over removal. As a result, Italian districts are defined by age. Central areas are the ancient ones and their surrounding areas followed over time - each district ended up having its own era. Consequently, in the Italian context, mixing buildings of different eras is not as important as (or, rather, as possible as) it is in the American context.

6.4 Concentration

The last of Jacobs's conditions is about dense concentration of people and buildings. In our cities, the most informative purpose is that of office work (highest beta coefficient for the column named "Concentration" of Table 5). This is reasonable as the previous three conditions (embedded in the regression) speak to the vitality contributed by *residents*, and this fourth condition adds to it the vitality contributed by *dwellers* who likely work in the district.

6.5 Border vacuums

Finally, looking at the "vacuums" model in Table 5, we observe little effect of border vacuums. Surprisingly, railways

or rivers - which might hamper pedestrian activity at times - seem, instead, to be effectively integrated in the social fabric of active districts. However, as one expects, highways are not: in general, being close to them ends up being detrimental ($\beta = -0.3457$ in the "vacuums" column and $\beta = -0.1018$ in the "combined" column).

Figure 4 shows to which extent (in terms of β coefficients) each of Jacobs's four conditions explains district activity.

6.6 Summary

Taken together, our results suggest that Jacobs's four conditions for maintaining a vital urban life hold for Italian cities as well (see Figure 4). From Figure 4 and the last column of Table 5, we see that as much as 77% of the variability of district activity is explained by simple structural and static features. As Figure 2 details, even individual features are strongly associated with activity.

Also, the extent to which the different features matter does not dramatically change across cities. To partly verify that, we took the largest and smallest cities: Rome and Firenze. As Figure 3 suggests, all the main features mattered to a very similar extent.

Based on our findings, to paraphrase the four conditions in the Italian context, we might say that:

> Active Italian districts have dense concentrations of office workers, third places at walking distance, small streets, and historical buildings.
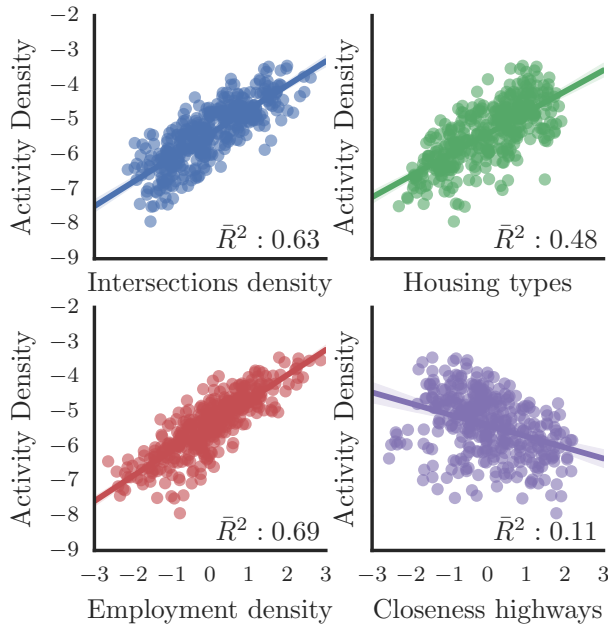

Figure 2: Plots of district activity as a function of intersection density, housing types, employment density, and closeness to highways. A district's activity is strongly associated with the district's static structural features.
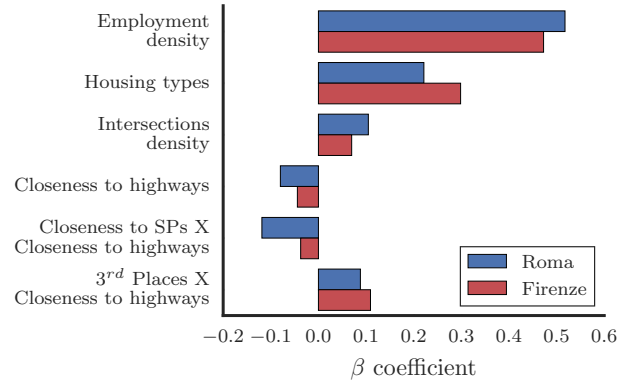

Figure 3: The β coefficients of two city-specific linear regression models for Roma and Firenze.
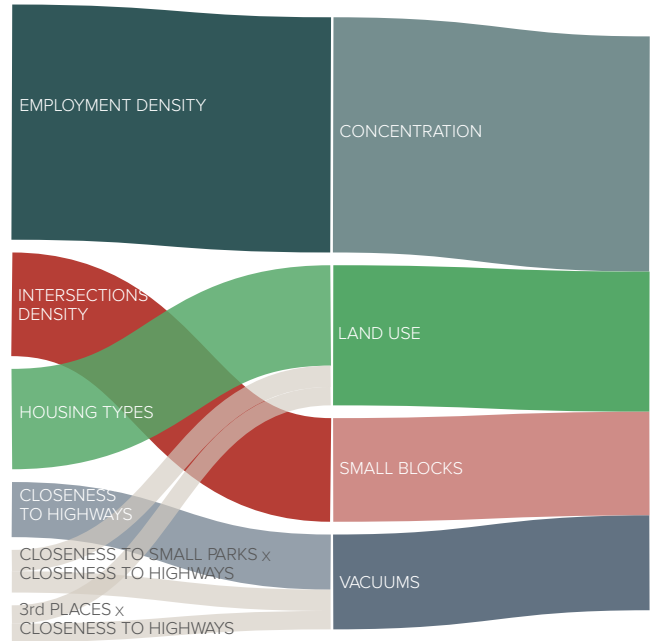

Figure 4: The predictive power of Jacobs's features. The bar size is proportional to the absolute β value in the linear regression model. The first column shows individual features, and the second shows grouped features. The three grouped features - concentration, land use, and small blocks - plus vacuum areas explain 77% of the variability of activity density.

7. DISCUSSION

After having operationalized Jane Jacobs's four conditions, we now discuss our work's practical and theoretical implications, and the questions it left open.

Practical Implications. We foresee several practical implications of our work:

City dashboards. Part of our work could inform the selection of which features should go into a city dashboard. We

have shown that a variety of structural features of the built environment are closely linked to district activity. Given the importance of those features and the ease with which they could be computed upon web and official data, city dashboards could show and track them and, in so doing, could support well-informed decisions by policy makers, urbanists, and architects.

Quantifying regulatory interventions. In his paper 'A City is Not a Tree' [1], Christopher Alexander argued that the variety and diversity essential for urban vitality was being destroyed by the implementation of zoning laws. Our work has focused on urban diversity features, the very features that zoning laws impact. To track the impact of a regulation, one now knows *which features* to contrast before and after the implementation of the regulation.

Place recommendation. How to best place retail stores? Researchers have recently tried to answer that question with semi-automatic mechanisms that identify the amenities that are missing from a neighborhood. Hidalgo *et al.* [17] analysed the economic diversity of neighborhoods to recommend new places for them. Similar analyses could profit from the addition of our work's structural features to the already present economic ones.

Theoretical Implications. Some of Jane Jacobs's theories continue to be controversial, while others have been criticized for their non-verifiability. This study showed that those theories not only are verifiable but also are still important today and valid in the European context too.

Limitations. Our measure of activity has been assumed to be a good proxy for city's vitality. However, that measure has left out one important factor that might be extracted in the future: temporal dynamics of district activity (e.g., differences between day and night). Also, if one were to have individual calls records instead of (as in our case) aggregate ones, one could well extract additional factors: duration of individual stays, radius of gyration, and identification of particular segments of the population (e.g., pedestrians *vs.* non-pedestrians, residents *vs.* dwellers). Finally, it would be difficult to *fully* replicate this study without call records. Those records were difficult to obtain in the past but have now been increasingly made publicly available by telecoms operators for research purposes [2].

8. CONCLUSION

For the first time, we verified Jane Jacobs's four conditions necessary for the promotion of urban life in the Italian context. We have done so by operationalizing her concepts in new ways: we used mobile phone records to extract a proxy for urban vitality, and web data to extract structural proxies for urban diversity. As Jacobs envisioned, vitality and diversity are intimately linked.

Until now, the research community has relied on limited data sets that reflect how a group of people experience bits of the city. Here we have found that web data and mobile records offer insights on how most urban dwellers experience entire cities. Web data (especially Open Street Map data) is available around the world, and mobile phone records

have been increasingly made publicly available by telecoms operators in search of new business models. As such, our study can be replicated at scale without resorting to lengthy and costly survey-based data collection.

By collecting bread crumbs from cell phones, social media, and participatory platforms, researchers will increasingly rely on data sets orders of magnitude richer than previous urban studies data sets [3] and, consequently, they will be able to test traditional urban theories in fine-grained detail - a totally new way to look at cities.

9. ACKNOWLEDGMENTS

10. REFERENCES

[1] C. Alexander. A city is not a tree. *Architectural Forum*, 1965.

[2] G. Barlacchi, M. De Nadai, R. Larcher, A. Casella, C. Chitic, G. Torrisi, F. Antonelli, A. Vespignani, A. Pentland, and B. Lepri. A multi-source dataset of urban life in the city of Milan and the Province of Trentino. *Scientific data*, 2015.

[3] M. Batty. Big data, smart cities and city planning. *Dialogues in Human Geography*, 3(3):274–279, 2013.

[4] L. M. Bettencourt, J. Lobo, D. Helbing, C. Kühnert, and G. B. West. Growth, innovation, scaling, and the pace of life in cities. *Proceedings of the National Academy of Sciences*, 104(17):7301–7306, 2007.

[5] V. Blondel, A. Decuyper, and G. Krings. A survey of results on mobile phone datasets analysis. *EPJ Data Science*, 4(10), 2015.

[6] A. Bogomolov, B. Lepri, J. Staiano, E. Letouzé, N. Oliver, F. Pianesi, and A. Pentland. Moves on the street: Classifying crime hotspots using aggregated anonymized data on people dynamics. *Big Data*, 3(3):148–158, 2015.

[7] A. Bogomolov, B. Lepri, J. Staiano, N. Oliver, A. Pianesi, and A. Pentland. Once upon a crime: Towards crime prediction from demographics and mobile data. In *Proceedings of the 16th International Conference on Multimodal Interaction*, pages 427–434, 2014.

[8] G. E. Box and D. R. Cox. An analysis of transformations. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 211–252, 1964.

[9] P. Calthorpe. *The Next American Metropolis: Ecology, Community, and the American Dream.* Princeton Architectural Press, 1993.

[10] R. Cervero. Land-use mixing and suburban mobility. *University of California Transportation Center*, 1989.

[11] N. Eagle, M. Macy, and R. Claxton. Network diversity and economic development. *Science*, 328(5981):1029–1031, 2010.

[12] S. S. Fainstein. New directions in planning theory. *Urban affairs review*, 35(4):451–478, 2000.

[13] E. Glaeser. *Triumph of the city: How our greatest invention makes us richer, smarter, greener, healthier, and happier.* Penguin Press, 2011.

[14] M. Gonzalez, C. Hidalgo, and L. Barabasi. Understanding individual mobility patterns. *Nature*, 453(7196):779–782, 2008.

[15] P. Gordon, S. Ikeda, et al. Does density matter? *Chapters*, 2011.

[16] S. Grauwin, S. Sobolevsky, S. Moritz, I. Gódor, and C. Ratti. Towards a comparative science of cities: Using mobile traffic records in new york, london, and hong kong. In *Computational Approaches for Urban Environments*, pages 363–387. Springer, 2015.

[17] C. A. Hidalgo and E. E. Castañer. Do we need another coffee house? the amenity space and the evolution of neighborhoods. *arXiv preprint arXiv:1509.02868*, 2015.

[18] S. Isaacman, R. Becker, R. Cáceres, S. Kobourov, M. Martonosi, J. Rowland, and A. Varshavsky. Identifying important places in people's lives from cellular network data. In *Pervasive computing*, pages 133–151. Springer, 2011.

[19] J. Jacobs. *The death and life of American cities.* Random House, 1961.

[20] L. W. Jeffres, C. C. Bracken, G. Jian, and M. F. Casey. The impact of third places on community quality of life. *Applied Research in Quality of Life*, 4(4):333–345, 2009.

[21] K. King. Jane Jacobs and "the need for aged buildings': Neighbourhood historical development pace and community social relations. *Urban Studies*, 2013.

[22] K. Kung, K. Greco, S. Sobolevsky, and C. Ratti. Exploring universal patterns in human home-work commuting from mobile phone data. *PloS one*, 9((6):e96180), 2014.

[23] M. Lenormand, M. Picornell, O. G. Cantú-Ros, T. Louail, R. Herranz, M. Barthelemy, E. Frías-Martínez, M. S. Miguel, and J. J. Ramasco. Comparing and modeling land use organization in cities. *arXiv preprint arXiv:1503.06152*, 2015.

[24] T. Louail, M. Lenormand, O. G. C. Ros, M. Picornell, R. Herranz, E. Frias-Martinez, J. J. Ramasco, and M. Barthelemy. From mobile phone data to the spatial structure of cities. *Scientific reports*, 4, 2014.

[25] R. Louf and M. Barthelemy. A typology of street patterns. *Journal of The Royal Society Interface*, 11(101), 2014.

[26] K. Lynch. *The image of the city*, volume 11. MIT press, 1960.

[27] K. Lynch. *What Time is this Place?* MIT Press, 1972.

[28] K. Manaugh and T. Kreider. What is mixed use? presenting an interaction method for measuring land use mix. *Journal of Transport and Land Use*, 6(1):63–72, 2013.

[29] N. Meinshausen and P. Bühlmann. Stability selection. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 72(4):417–473, 2010.

[30] R. Oldenburg. *The Great Good Place: Café, Coffee Shops, Community Centers, Beauty Parlors, General Stores, Bars, Hangouts, and How They Get You through the Day.* Paragon House Publishers, 1989.

[31] R. Oldenburg and D. Brissett. The third place. *Qualitative sociology*, 5(4):265–284, 1982.

[32] L. Porzi, S. Rota Bulo, B. Lepri, and E. Ricci. Predicting and understanding urban perception with convolutional neural networks. In *Proceedings of the ACM MultiMedia.* ACM, 2015.

[33] R. D. Putnam. *Bowling alone: The collapse and revival of American community.* Simon and Schuster, 2001.

[34] D. Quercia, L. M. Aiello, R. Schifanella, and A. Davies. The digital life of walkable streets. In *International Conference on World Wide Web*, pages 875–884, 2015.

[35] D. Quercia, N. K. O'Hare, and H. Cramer. Aesthetic capital: what makes london look beautiful, quiet, and happy? In *ACM conference on Computer supported cooperative work & social computing*, pages 945–955, 2014.

[36] D. Quercia, J. Pesce, V. Almeida, and J. Crowcroft. Psychological maps 2.0: A web gamification enterprise starting in london. In *Proceedings of ACM Conference on World Wide Web (WWW)*, 2013.

[37] P. Salesses, K. Schechtner, and C. A. Hidalgo. The collaborative image of the city: mapping the inequality of urban perception. *PloS one*, 8(7):e68400, 2013.

[38] C. Smith-Clarke, A. Mashhadi, and L. Capra. Poverty on the cheap: Estimating poverty maps using aggregated mobile communication networks. In *ACM conference on Human factors in computing systems*, pages 511–520, 2014.

[39] C. Song, Z. Qu, N. Blumm, and A. Barabasi. Limits of predictability in human mobility. *Science*, 327(5968):1018–1021, 2010.

[40] H. Sung, S. Lee, and S. Cheon. Operationalizing jane jacobs's urban design theory: Empirical verification from the great city of seoul, korea. *Journal of Planning Education and Research*, 35(2):117–130, 2015.

[41] M. Traunmueller, G. Quattrone, and L. Capra. Mining mobile phone data to investigate urban crime theories at scale. In *International Conference on Social Informatics*, pages 396–411, 2014.

[42] Y. Zheng, L. Capra, O. Wolfson, and Y. Hai. Urban computing: Concepts, methodologies, and applications. *ACM Transactions on Intelligent Systems and Technology*, 5, 2014.