Using mobility data as proxy for measuring urban vitality

Patrizia Sulis[1], Ed Manley[1], Chen Zhong[2], and Michael Batty[1]

[1]Centre for Advanced Spatial Analysis, University College London, United Kingdom
[2]Department of Geography, King's College London, United Kingdom

Abstract: In this paper, we propose a computational approach to Jane Jacobs' concept of diversity and vitality, analyzing new forms of spatial data to obtain quantitative measurements of urban qualities frequently employed to evaluate places. We use smart card data collected from public transport to calculate a diversity value for each research unit. Diversity is composed of three dynamic attributes: intensity, variability, and consistency, each measuring different temporal variations of mobility flows. We then apply a regression model to establish the relationship between diversity and vitality, using Twitter data as a proxy for human activity in urban space. Final results (also validated using data sourced from OpenStreetMap) unveil which are the most vibrant areas in London.

Keywords: spatial big data, mobility flows, urban vitality, temporal patterns, urban dynamics

1 Introduction

In urbanism, numerous attempts have been made at defining and calculating metrics to better describe and understand spatial dynamics in cities, and more specifically dynamics related to the presence of people in places. However, spatial information used in previous studies (including data related to human dynamics) has been commonly collected through empirical observations and surveys, resulting in very detailed yet limited data sets in terms of the spatial and temporal extents of the urban situation they were describing. A remarkable example of these attempts is the work of Jan Gehl in various cities [11, 12]. His work proposes an integrated methodology including qualitative evaluations and quantitative measurements, with data manually collected through empirical observations, of the variation of use of public space over time.

The urban data deluge that has recently become available represents an unprecedented opportunity for researchers to have access to extensive and detailed data sets of information about urban space. This opens to the possibility of unveiling urban dynamics at a finer granularity. Such informationally rich data sets, although not produced for the specific purpose of spatial analysis [20], show another representation of urban space, containing supplementary information not derived from its morphology (which is obviously the main source for urban metrics). Therefore, including these data in the analysis means introducing an additional layer of information to the complex representation of what urban space is.

The exhaustivity and fine-granularity of the data open to the remarkable possibility for researchers to support, test, and validate, from a quantitative perspective, concepts and intuitions that have become central in urban theory and planning practice. In this paper, we focus on one of these, Jane Jacobs' definition of urban vitality in the neighborhoods of US cities [19]. She understood vitality mainly in relation to street life. The variety of pedestrian flows over time is seen as an essential condition for the continuous presence of people in the street, which represents vitality in cities. According to her idea, vitality is strongly influenced by specific morphological attributes of the built environment. Recent work explored this relationship, also employing large data sets of human mobility data as a proxy for vitality. However, their results are presented at a spatiotemporal scale that appears too general to fully unveil a detailed description of the phenomenon.

In this paper, we focus on the definition of urban metrics that describe phenomena related to spatial dynamics and behavior of people in places. Different from previous works, we analyze data sets sourced from smart cards used in public transport to minutely measure the spatiotemporal variations of urban vitality in the city of London during the week. Applying a computational approach to the concept of vitality, we introduce a new set of three dynamic attributes to calculate urban diversity as the variations of human dynamics in places. Diversity is measured evaluating not only the number of people present in a place but also the continuity and the variations of this presence over specific intervals of time. Then, we establish the relationship between the diversity of flows and the urban vitality of places, applying a regression model that employs data sourced from Twitter. Results of the data analysis show the spatial distribution of vitality in London. The fine spatial and temporal granularity of data used makes it possible to observe meaningful variations among areas in the city, including interesting exceptions in relation to weekend dynamics. Furthermore, the combined metric proposed in this work highlights the component of temporal continuity, which is a significant feature for urban vitality as understood by Jacobs [19]. Finally, the regression model is evaluated and the results validated by comparing them with spatial data sourced from OpenStreetMap.

The paper is structured as follows. Section 1 presents the objectives of the work. Section 2 introduces a brief overview of current research about vitality and spatial dynamics using new large data sets. Section 3 describes the data sets used in the analysis. Section 4 illustrates the methodology applied. Section 5 presents the main results and its validation. Section 6 discusses relevance and limitation of this study. Section 7 concludes the paper.

2   State of the art

This section introduces the concept of vitality, as understood by Jacobs, and recent works that employ large spatial data sets of human mobility information to validate this concept.

It then proceeds to present a concise overview of current research exploring human dynamics and presence in urban space by employing data sets from multiple sources, highlighting their strengths and limitations.

2.1 Urban vitality: a definition by Jane Jacobs

According to Jane Jacobs [19], urban vitality is an essential requisite for a successful city in term of street life, and it is particularly associated with safety and walkability. "An intricate sidewalk ballet (p.50)" of people walking around neighborhoods, at different times for different purposes, is what makes an urban place successful and safe. Jane Jacobs strongly connects the concept of vitality to that of diversity, which she understood mainly in terms of morphological diversity related to the built form of the city and the distribution of activities. She identifies four conditions of diversity (mixed used, age of buildings, number of street intersections, and urban density) for evaluating vitality. They are essential and complementary to each other, together contributing to define the degree of diversity (and therefore vitality) of places.

In her work, Jacobs carefully describes her idea of vitality in relation to urban space; however, she never attempted to measure it. New studies propose a quantitative evaluation of vitality, exploring the relationship between human mobility data and urban morphological data, represented by a measurement of the four attributes of diversity proposed by Jacobs. In Sung et al. [34], human mobility is represented by the pedestrian presence in the streets of Seoul over several years, collected through manual surveys into extensive data sets, whereas diversity is calculated through the four attributes for specific city neighborhoods. In the work of De Nadai et al. [8], the proxy for vitality is instead represented by mobile phone data (Internet access only) from various Italian cities. Similar to [34], diversity is calculated through the four morphological attributes suggested by Jacobs, with data sourced from the Italian Census.

These works have successfully validated Jacobs' idea of the relationship between morphological diversity and urban vitality at a large scale, extending its effectiveness beyond U.S. cities. However, they also present some limitations in regards to the definition of vitality applied, particularly in relation to the spatiotemporal variations of vitality across neighborhoods.

> On successful city streets, people must appear at different times. This is time considered on a small scale, hour by hour through the day [19, p. 152] [...] The mixture of people on a street at one time of day must bear some reasonably proportionate relationship to people there at other times of day [19, p. 164].

According to Jacobs, the spatiotemporal continuity of human presence and activities in a place is a fundamental trait of vitality—its temporal variations, in terms of intensity and duration of flows, is a relevant aspect to consider in the analysis. In relation to this, the heterogeneity of flows resulting from the presence of different activities in places is also important when evaluating urban vitality in neighborhoods.

Previous works present analyses at an aggregated spatiotemporal level, and it is, therefore, difficult to extract detailed information about the continuity of urban vitality across different time intervals. However, new spatial data sets recently available make possible to analyze spatial and temporal dynamics at an unprecedented level of detail, opening to

the possibility of exploring and uncovering similarities and differences in spatial patterns of human dynamics in cities.

2.2 The availability of new spatial data

The recent deluge of location data is caused by automatically-generated data produced by sensors and other technological devices that human users interact with in urban space. This results in some advantages, such as the high-resolution of spatial and temporal information, automatically collected for wider samples and contexts with no subjective bias related to the human collection. However, it also introduces some relevant bias related to the data representativeness [20]. Some automatically-produced data (in particular social media, [2, 23]) can overestimate demographic groups that have access to specific technologies.

Numerous data sets have been employed in recent research, and three types of spatial data have been of particular interest for a number of years. The first one includes data sourced from "smart cards," mainly used for calculating commuting flows [24, 33] and identifying urban zoning [16, 31, 36]. The second type is represented by mobile phone data, aggregated and anonymized, used mainly for detecting urban dynamics and activities because of its fine granularity [3, 4, 13, 15, 22, 28]. A third variant includes social media data, which are readily available from social media companies. Spatial applications of these data are particularly relevant when data attributes include location information, as in the case of Twitter and Foursquare data [5, 6, 17, 26].

Data mentioned above have strengths and limitations that should be considered in the analysis and evaluation of results (Table 1).

Type of data	Strengths	Limitations
smart card	· good spatial precision · use of device is widely diffused	· discontinuous data · potentially multiple users per card
mobile phone	· continuous data · use of device is widely diffused	· spatial precision varies depending on data structure and content · uneven density of antennas and data · potentially multiple devices per user
social media	· easily accessible · spatially located · some content is context rich	· small percentage of geotagged data · few users, some demographic categories are under represented

Table 1: Main types of data used in current research.

Current research employing these data sets to explore human dynamics and spatial behavior in cities shows an interesting variety of applications, both in terms of data used and phenomena analyzed. Examples include works using high-resolution location data to infer specific "digital signatures" of places. The analysis of human mobility data sourced from mobile phones makes possible to detect spatial patterns of preference at the neighborhood level [4], to uncover digital footprints and preferred places of locals and tourists [13], and to detect the "digital signature" and specific temporal profiles of urban areas [29]. Using a large data set sourced from social networks (Foursquare and Twitter), Cranshaw et al. [6]

explored how the inhabitants' perceptions of urban boundaries, inferred from their daily mobility patterns, corresponded to the borders defined by the city administration. Variations of urban dynamics and routines are also analyzed in relation to specific real-world events [1,32]. Lee et al. [21] used tweets to detect small informal urban activities. Analyzing geo-tagged Twitter data, researchers monitored crowd behavior, identifying anomalies of patterns and detecting the occurrence of unusual local events.

Limitations of current research concern two main aspects: the representativeness of the data and the usage of data to predominantly explore a certain type of urban phenomena.

- **Representativeness of data.** These data sets are undoubtedly partial and represent a sample of the urban reality regardless of their extensiveness [20]. This also has implications for the type and reliability of information contained in the data (often not produced for the purpose of spatial analysis). Furthermore, current research (with meaningful exceptions [3, 22]) tends to focus on single-sourcing data sets, raising an additional concern about the reliability of results.
- **Usage of data.** Despite the wide attention received, little research([4, 10, 27, 35]) has focused on analyzing the fine-grained heterogeneity in spatial dynamics that deviates from the large-scale geographic trends that have been known and understood for decades (i.e., home-work commuting). Furthermore, few have attempted to connect this analysis to spatial applications supporting urbanism concepts and methods with valuable quantitative information.

In the light of this, we decided to focus our interest on investigating how to take advantage of the potentialities of these spatial data sets to explore new applications to well-known ideas relevant in urbanism theory and practice. We propose an computational approach to Jacobs' concepts of diversity and vitality, converting diversity to a metric that measures spatiotemporal variations in human mobility patterns across urban space.

3   Description of data

This work employs data sets containing information about human mobility and presence in places for the City of London. These data sets have multiple sources: the public transport authority Transport for London (Oyster "smart card" data), the social media platform Twitter (tweets collected through the public API), the open-source database OSM (Volunteered Geographic Information—VGI). In particular, we selected two types of data that can reliably represent human mobility (the records of individual journeys directly provided from the transport authority) and that, according to previous works, can provide a good estimation of the density of human activity in cities beyond the physical constraints of the public transport stations (Twitter data location points).

The fine granularity of the data selected make it possible to analyze spatiotemporal variations of urban flows at a highly detailed scale. Both transport and social media data sets consist of a large collection of non-aggregated records representing single events (i.e., a bus journey or a tweet), and each record contains information about a specific spatial location and a time stamp. This makes it possible to perform data analysis and measure temporal variations at several intervals, in order to obtain an in-depth knowledge about the variety of urban dynamics and spatial behavior in different areas of the city. Furthermore, we can observe in detail those variations of spatial flows that Jacobs considered an essential

feature of vitality. Particular attention is dedicated to the continuity of flows over time, rather than the simple magnitude of flows that show interest in a place during specific hours in the day (as it happens, for instance, at peak hours in central, working areas).

3.1 Oyster card data

The public transport data set is provided by Transport for London (TfL), the authority responsible for the various means of collective transport in Greater London. This data set consists of approximately one month of transaction records (collected between January and February 2014) of the Oyster card, the smart card used to get access to the public transport network. The Oyster card is used for entering and exiting through the ticket barriers on the rail network (Underground, Overground, DLR, etc.), and for boarding the buses. Each recorded transaction represents a single journey through the transport network, identified by a unique ID code and containing a large amount of information about that journey.

For the rail network, the information details recorded for each transaction and used for this analysis are:

- a unique ID code identifying the journey,
- a code identifying the transportation mode of the journey,
- a code and a name identifying the station,
- a code identifying if the user is entering or exiting the station,
- a code identifying the day of the journey,
- a time stamp for the entrance or exit of the station, identifying the start or the end of the journey.

For the buses, the Oyster card is required only when boarding. The information details recorded for each transaction and used for this analysis are:

- a unique ID code identifying the journey,
- a code identifying the transportation mode of the journey,
- a code and a name identifying the bus stop,
- a code identifying the day of the journey,
- a time stamp for the boarding of the bus, identifying the start of the journey.

The limitations of this data set are mainly related to the size of the sample and the spatial distribution of information recorded in data. The data set is provided directly by Transport for London. It is very rich in details but has constraints in time extension due to their privacy policy. Data has been cleansed and known bias has been removed from the records used in the analysis. According to related work using data from the same source [30], seasonal variations do not appear to significantly influence the data trends, therefore we consider the data reliable for this work. Regarding the spatial distribution of data, since it is recorded at specific locations, we decided to perform analysis within specific areas of the city to obtain spatially detailed results rather than aggregated ones. This limitation is common to other types of mobility data (i.e., data sourced from mobile phones). The spatial bias is intrinsically linked to the locations of sensors recording the information (ticket barriers at stations, antennas), which are generally more densely located around the central areas of the city than in the periphery.
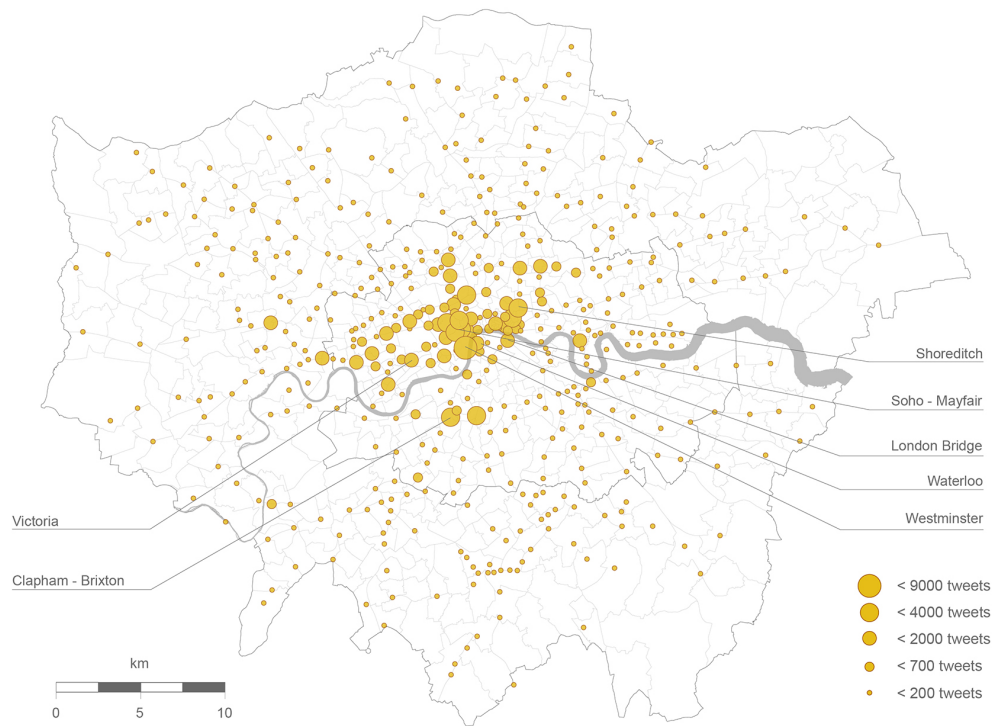
Figure 1: Number of tweets per area of research.

3.2 Twitter data

The Twitter data set consists of approximately three months of records, collected between January and March 2016, for the city of London. The data has been collected through the Twitter API [1]. The data set contains only geo-tagged tweets; that is, tweets containing spatial information (geographic coordinates expressed in latitude and longitude) about the location the message was sent from. According to literature [23], geo-tagged tweets represent approximately 1% of the total amount of tweets sent by users. Nevertheless, it represents an interesting and valuable source of information to investigate spatial dynamics across cities [5, 17]. Furthermore, Twitter data is also relatively easy to access through the public API, whereas other types of data (Foursquare, mobile phone records) are private and difficult to obtain for academic research. Each recorded tweet is identified by a unique ID code and contains additional metadata, including the user ID and name, the geographical coordinates and the time stamp the tweet was sent from, a short text (140 characters), any media attached to the text (photo, video, etc.), the language of the text, the city of the user (defined by the user in his profile). We collected around half a million geo-tagged tweets for the entire area of Greater London, setting a bounding box around the area interested in our study. Tweets were further selected to retrieve only records contained within the areas object the analysis (approximately 200,000 tweets).

[1](https://dev.twitter.com/overview/api), using the application Big Data Toolkit developed by Steven Gray at CASA, UCL (https://bigdatatoolkit.org).

The information details used for this analysis are:

- a unique ID code identifying the tweet (but not the user),
- the geographical coordinates (latitude and longitude) of the location the tweet was sent from,
- the time stamp identifying the time the tweet was sent.

In this work, we employed Twitter data as a proxy to evaluate the activity density in relation to the presence of people in various places in London (Figure 1).

The limitations of this data are mainly related to the demographic bias and the limited availability of geo-tagged tweets through the Twitter API stream (as mentioned in Section 2) [2, 23]. Another potential bias is related to the limited size of the sample. We chose to collect data from January to March because we wanted to use this set together with the Oyster card data set (provided by TfL for that specific time of the year). Therefore, we needed the two data sets to be temporally coincident. Although a preliminary analysis on a larger set of tweets suggests that data may be subject to seasonal variations in the city of London, we are confident that, for the purpose of this work, this type of bias has limited influence in affecting the validity of the methodology.

3.3 OpenStreetMap data

This data set contains information about the location of activities and Points of Interest (POI) in London, collected from OpenStreetMap through the Overpass API.

Each record contains several pieces of information, including the typology of activity (shop, pub, tourist attraction, etc.), the address, sometimes the opening hours, and more. We chose to use OpenStreetMap as a source because of the availability of data, easily retrievable through the dedicated API. However, due to the nature of the OSM database (Volunteered Geographic Information–VGI [14]), the uniformity of data attributes associated with the spatial objects may vary through the entire data set. Other sources, i.e., Ordnance Survey, may be more robust and will be considered in the future. The attributes used in our analysis are the name and address of the activity, its typology (defined by OpenStreetMap users), and the geographic coordinates (latitude-longitude).

4   Measuring urban vitality: a computational approach

In this paper, we propose to evaluate urban vitality by analyzing the diversity of mobility patterns and examining their spatiotemporal distribution and variations.

As mentioned in Section 2, the diversity of a place (which Jacobs understood in morphological terms) and its vitality are intrinsically related, with the former being an essential requisite for the latter. Consistently with Jacobs' view, we consider the continuity and variety of human mobility flows as features that essentially contribute to the vitality of a place, and more significant than the number of people (magnitude of flows) alone to obtain a reliable measure of vitality. Different from Jacobs' idea and previous related work, we consider diversity not as a measure of morphological features influencing the presence of people in places, but as a measure that actually calculates this presence and the flows affecting the areas as an object of the research.

Our measure of diversity varies in space and time. Spatially, it is calculated for different local areas, as described in Section 4.1. Temporally, the measure is calculated at daily and

hourly intervals, for weekdays and weekend days. Diversity is defined by three complementary attributes—intensity, variability, and consistency—each one representing a different granularity in temporal variation, and together contributing to a value representing how diversity temporally changes over time in the same urban place.

In this section, we present the methodology followed for calculating diversity (Dv). We first illustrate the definition of spatial areas of research in which data analysis is performed. We then describe how each of the dynamic attributes constituting diversity is calculated. Finally, we explain how to employ a regression model to extract weights for each attribute. Section 5 illustrates the final results of this methodology.

4.1 Defining the spatial units of research

We defined the areas for research considering the spatial distribution of the Oyster card data—the presence of people is recorded at specific, fixed locations in a defined moment of time. We assumed that the rail stations (Underground, Overground, DLR, etc.) and the bus stops located nearby work in synergy, with users boarding and alighting at bus stops around the stations to complete their multi-modal journeys across the network. It is noteworthy that public transport stations act as "flow aggregators," with other mobility flows converging around them, including the pedestrian flows and those generated by the bike sharing network in London. We defined our research areas considering rail stations and bus stops included within a limited radius of walking distance around each station of the network. After considering various attempts with different ranges, we found the 400 meters walking distance to be suitable for our research. It is a common standard distance used in transit planning as it is considered easily walkable by many city users [7,9]. Areas are defined using QGIS, buffer, and convex hull tools, assigning bus stops that fall within the 400 m range to the correspondent station, so as to include all the transportation points. Each area represents a spatial unit working in synergy in regard to the flows of people in that area (Figure 2). Vitality is calculated considering the presence of people recorded at the transport locations (for Oyster card data) and the geo-tagged tweets (for Twitter data) included in these areas.

A feature we carefully considered when defining these areas is the heterogeneity in the spatial distribution of stations and bus stops in the city, and the related differences in attracting the mobility flows. The access points of the transport network tend to be more concentrated in the city center, often located at the intersection of different mobility networks (acting as multi-modal nodes) and connecting different spatial scales (urban, regional, international). Due to this privileged position in the network, they are affected by higher mobility flows in comparison to the peripheral nodes of the network. Since we were more interested in the continuity of flows rather than their mere magnitude, we introduced in our calculation a metric that evaluates the position and the connections of each point to the entire urban network. Specifically, we used a metric that represents the concept of centrality, the "global integration" defined by Space Syntax [18]. Integration represents how connected ("integrated") a street is in comparison to the entire urban network, from the perspective of car-based and public transport mobility [25]. It is a measure of centrality based on the topology of the street network of the city ("axial network" [25]), highlighting the central areas in comparison to the peripheral, most "segregated" areas. We used it to normalize the total number of people recorded by the Oyster data; i.e., for a central location, we may expect big flows but also a high value of integration for the adjacent streets,
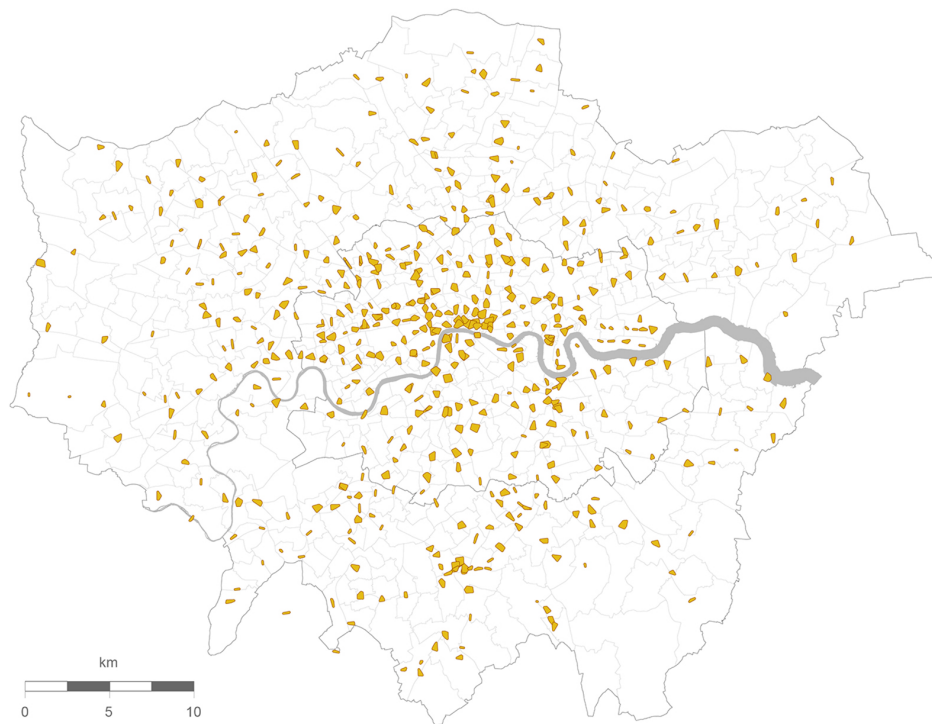
Figure 2: Spatial units of research across London

whereas for peripheral location small flows of people are common, but also lower values of integration. We assigned the value of "integration" (calculated by Space Syntax for the "axial network" of London) to each station and bus stop of the public transport network (in case of intersection with multiple axes, we assigned the maximum of the values of integration available). At the end of this process, each point of the transport network had an "integration" value assigned, which was later used as a coefficient to normalize the value of flow intensity for each day and hourly interval considered in the analysis (see Equation 2).

4.2   Calculating the three attributes of diversity

As previously mentioned, Jacobs [19] defines four morphological attributes of the built environment (age, density, mixed use, and street intersection) as essential conditions that contribute to the overall quality of urban diversity. In this work we are adopting her idea of interconnection between urban diversity and vitality, however, we decided to analyze this relationship exploring the spatiotemporal variations of human mobility, using this temporal diversity to obtain a quantitative evaluation of vitality for places in London. Therefore, we defined diversity as composed of three dynamic components representing the variation of mobility flows at different temporal scales: intensity, variability, and consistency.

These three attributes contribute to the final measure of diversity (Dv) through a linear combination, described by the Equation 1:

$$Dv = \alpha_1 * intensity(I) + \alpha_2 * variability(V) + \alpha_3 * consistency(C) \tag{1}$$

where:

a. Intensity (I) is calculated as the total number of people who enter and exit from stations and alight at bus stops included in each spatial unit at a defined time interval (Equation 2). We calculated a value of intensity for each point at an hourly lever for every weekday, and then we normalized this value using the "integration" score (as explained in paragraph 4.1). Finally, we calculated the value of intensity for each spatial unit, aggregating the data for an average working day and an average weekend day.

$$I = \frac{\sum users}{integration\ score} \tag{2}$$

b. Variability (V) is defined as the difference of user flows among different days for each station and bus stop (Equation 3). When calculating variability, we were interested in understanding the stability of temporal patterns of mobility flows: the higher the variability, the more random the flows are distributed over time, contributing to the continuity of flow in the area. Variability adds to the final value of diversity a component representing the differences of urban dynamics in relation to specific activities between the weekdays (routine patterns related to commuting) and the weekend (non-routine and irregular patterns related to leisure, shopping, etc). Variability is calculated at the daily level by comparing the correlation ($Corr()$) of flow vectors of any pair ($flow_i, flow_j$) of n days and then normalizing this value by the number of pairs. For instance, if the number of days is five, then there will be ten pairs of comparisons [37].

$$V = \frac{\sum_i^n \sum_j^n (1 - Corr(flow_i, flow_j)))}{(n * (n - 1)/2)} \tag{3}$$

c. Consistency (C) represents the hourly variation of flows during the same day for each station and bus stop (Equation 4). We used it to observe if places presented a continuous temporal pattern (indicating a constant use of the space over time) or one characterized by concentrated peaks (which will suggest a presence of people in places concentrated into specific hours, possibly driven by specific activities). This attribute contributes to highlight the variety of spatial dynamics in space across the same day (i.e., differences between morning and night flows). A higher number of outliers indicate a more irregular, non-routine temporal pattern in the use of the space, with a minor concentration of flows around specific times. This is related to the idea of continuity in street activity suggested by Jacobs. Consistency is measured calculating the hourly outliers using MAD (Median Absolute Deviation) and then normalizing this value using the daily number of outliers per station or bus stop.

$$C = \frac{outlier}{\sum daily\ outliers} \tag{4}$$

Despite the "integration" score, the more central and crowded a station or bus stop is, the more the attribute of intensity contributes to the final value of diversity (Dv). The two additional attributes (variability and consistency) represent the variations of the spatial

patterns of mobility at a different temporal resolution, and act to balance the contribution of intensity to the final value, in order to achieve a better evaluation of vitality (Vt). The three attributes (Table 2), singularly and together, uncover the existing differences in temporal patterns across the whole city of London and within the same place (Figures 3, 4, and 5). Areas showing more variations and a more dynamic temporal pattern of flows are places with a higher degree of vitality, attracting people throughout the day.

	intensity	variability	consistency
min	0.01	0.05	0
mean	0.22	0.32	0.05
max	0.58	0.61	0.15

Table 2: Description of the three attributes for the average day of the week.

Finally, the fine granularity of the data used for calculation makes it possible to measure variations and detect temporal differences in mobility patterns even at small time intervals (one-hour intervals and further). This contributes to achieving a better understanding of the spatial human dynamics.

4.3 Regression model

In order to calculate diversity Dv, the three attributes are weighted according to how much they contribute to the final value of urban vitality. To compute the weights, we used a linear regression model, and a data set sourced from Twitter as the dependent variable representing the human presence in urban space (similarly to [5, 26]).

Our intuition, when deciding to use Twitter data as the proxy for the density of human activity in the city, is that the ranking of the areas of research according to the number of geo-located tweets would not differ significantly from the ranking of the same areas according to urban vitality. It is important to stress that in this case tweets count is not considered as an exact estimation of the human presence in places (for the noted demographic bias of this type of data [2, 23]), nor as an approximation of the final values of vitality. It is instead used to rank the spatial units based on the density of human activity, a ranking that can also be descriptive of the order of the spatial units according to the presence of people in places (vitality). We judged Twitter data to be sufficiently reliable in the light of previous research, and preferable to other sources (e.g., static sample population) because it catches the temporal variations in human mobility flows. It is also a necessary choice in this study since, to our knowledge, no ground truth measuring urban vitality in various places of London is currently available (we will collect more information in the next phase of this work through a bespoke survey).

We defined the ranking assigning each tweet to the respective unit of research by spatial proximity. We then run an OLS multivariate linear regression where tweets represent the dependent variable, and the three dynamic attributes of diversity represent the independent variables of the model (standardization has been performed as required in OLS regression). Doing so, we can estimate how the final value of vitality is directly influenced by any of the three independent variables. The model follows the Equation 5 below:
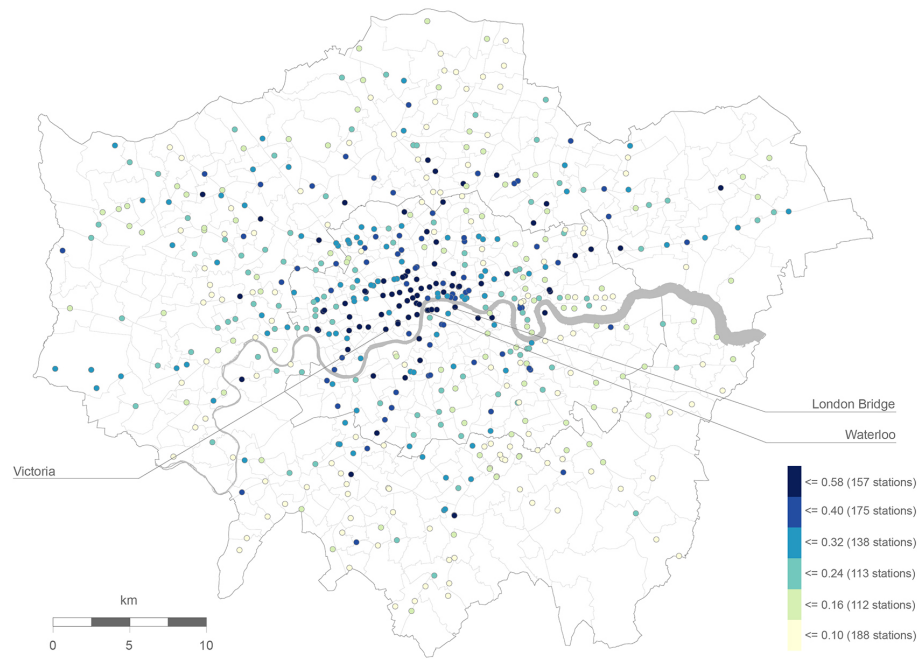
Figure 3: Spatial distribution of the intensity calculated for an average week day.

$$tweet\ count \sim \alpha_1 * intensity(I) + \alpha_2 * variability(V) + \alpha_3 * consistency(C) \qquad (5)$$

We run the regression model using different values of the variables calculated for an average weekday and an average weekend day to extract appropriate weights for different times of the week (Table 3). Once we obtained the corresponding attribute weights, we assigned them to each attribute (intensity, variability, and consistency obtained from Oyster data), and we calculated the final values of diversity (Dv) for the average weekday and the average weekend day, and for hourly intervals, as indicated in Equation 1.

Attribute	Week (Adjusted-R^2 = 0.843)	Weekend (Adjusted-R^2 = 0.831)
intensity	1.46	1.64
variability	0.51	0.39
consistency	-0.11	-0.46

Table 3: Regression results: coefficients (weights) and Adjusted-R^2.
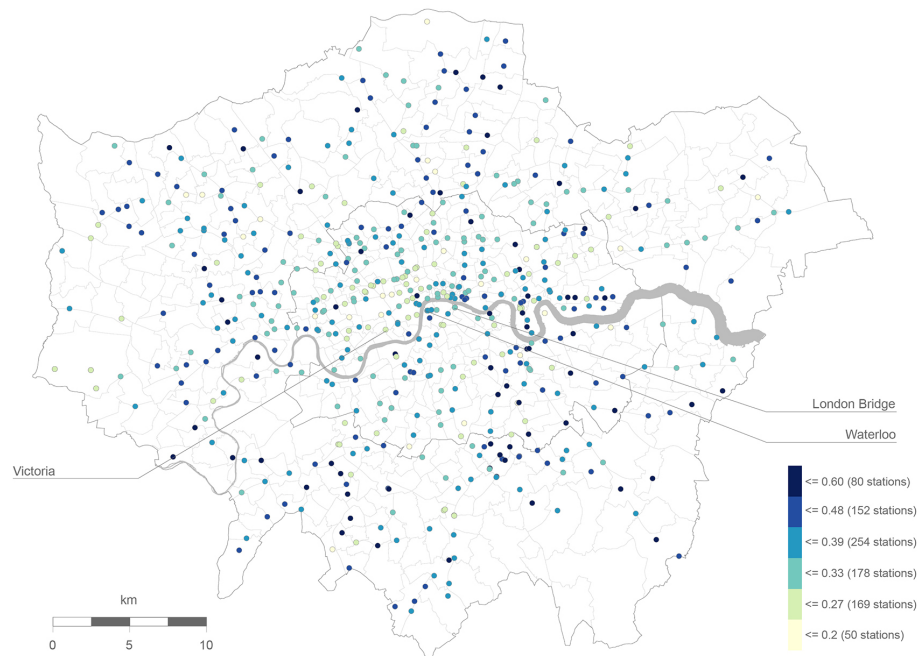
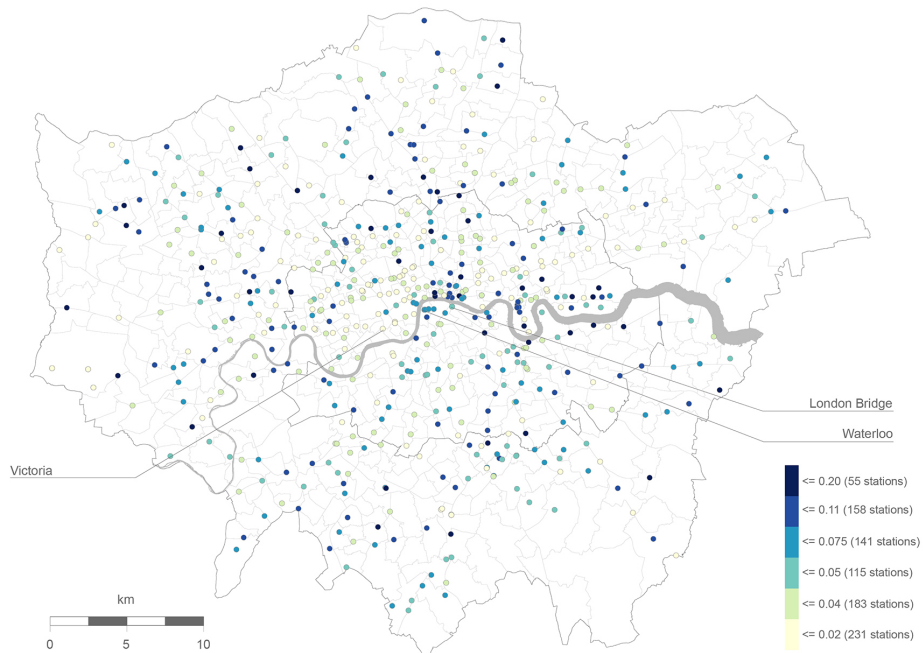Figure 4: Spatial distribution of the variability calculated for an average week day.

Figure 5: Spatial distribution of the consistency calculated for an average week day.

5 Results

5.1 Spatial distribution of vitality in London

This section presents the results of the data analysis and the metrics calculation, and shows the spatial distribution of vitality in London according to different times of the day and during the week. The fine temporal granularity of the data used in the analysis makes possible to observe meaningful variations of flows and mobility patterns, which significantly help to understand in a detailed and quantitative manner how places across London are differently used and visited over time.

Results presented here include the vitality values calculated for an average weekday (Figure 6) and for an average weekend day (Figure 7). It is worth mentioning that the data set used also includes bus records, which may explain some of the results in peripheral areas and during the weekend since rail stations (Underground, Overground, DLR, etc.) act as attractors of bus flows from the surrounding areas. Furthermore, bus data is also representing mobility flows at the local level in the overall results. This will need further analysis of the results to fully understand the implications.

For some of the areas analyzed, results confirm our empirical knowledge. This is the case for central London, and for the main multi-modal train stations (London Bridge, Waterloo, Victoria, King's Cross). We expected those places of the city (where the majority of workplaces, businesses, and tourist attractions are located) to show high values of vitality because of the presence of people in those areas. The higher values of vitality during the weekdays are strongly influenced by the magnitude of flows related to the home-work commuting journeys. Intensity presents high values for those areas, whereas variability shows lower values because of the routine behavior of flows during the week (Figure 6). Hourly variation during the day is also low, with no or few significant variations related to commuting flows (which results in low values of consistency and less vitality).

Besides that, other areas show interesting behaviors, particularly those showing a clear discontinuity in comparison to the adjacent zones, and, notably, the area including Wood Green and Turnpike Lane Underground stations (North London, Figure 6), which act as attractors of flows for the contiguous residential areas (they also include important commercial destinations). Similar patterns are evident in South London, in the areas around Clapham and Brixton stations (Figure 7), where vitality shows constant high values because the two places attract flows for the adjacent residential neighborhoods. Interestingly, variability shows low values across the whole week (weekend included). This may be related to the fact that the two stations (part of the Tube and Railway network) act as main gateways to Central and North London, and are therefore highly frequented also outside the commonly understood commuting hours.

There are also areas showing weekday patterns that appear to be influenced by types of flows other than commuting, such as tourism and leisure, for example, Richmond (South West London, Figure 6), which shows a high value of vitality during weekdays. The areas around the Greenwich Peninsula (South East London, Figure 7), with the locations of Greenwich itself and the O2 Arena, show values related to a continuous presence of people over the week. High values of intensity and variability are also maintained during the weekend, whereas in the opposite Canary Wharf the value of vitality declines, as relevant variations happen between days (and hours) of the week because of its main aspect of a business destination.
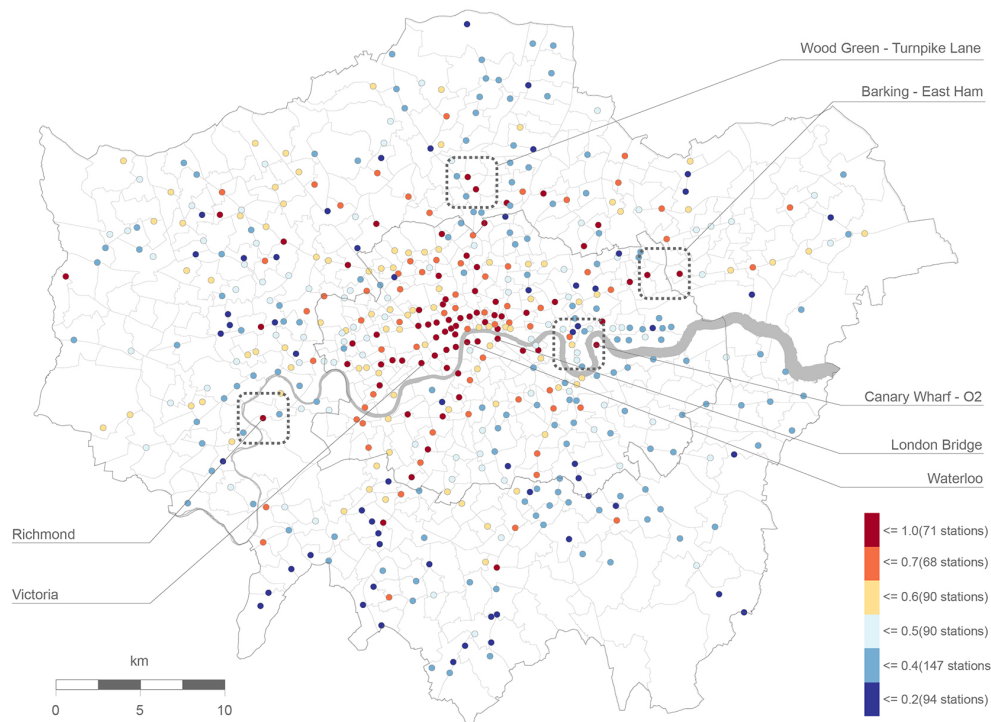
Figure 6: Vitality values for an average week day.

The higher values of vitality during the weekend reflect significant differences during week and weekend, and a multiplicity of mobility patterns during the weekend. This confirms the correct functioning of the combined index of diversity (Dv) which, although inevitably influenced by the magnitude of flows, draws the attention on the temporal variations of the spatial patterns of city users, and the continuity of them as a fundamental feature for a place to be vital.

A significant example of this is represented by the case of City of London–Shoreditch area in East London (Figure 7). It shows an insight of the variety of spatiotemporal behaviors of close areas in the city. Observing the difference between the average vitality values during the weekend and the week, some interesting patterns about the temporal use of spaces start to emerge. The area of the City of London appears less vital during the weekend, which is expected considering the presence of many offices and related facilities in the area that attract flows mainly during weekdays. At the same time, the areas around Old Street and Shoreditch High Street tube stations, which also show high values of vitality during the week, appear even more vital during the weekend. This increase is presumably related to significant variations of flows driven by the multiplicity of activities in the area, which attracts heterogeneous categories of city users in the area over a longer period of time during the day and the week. Shops, bars, and other activities related to leisure are located there (not to mention the increased popularity of the area as a tourist destination), but also different typologies of working activities as co-working places for freelancers and start-ups
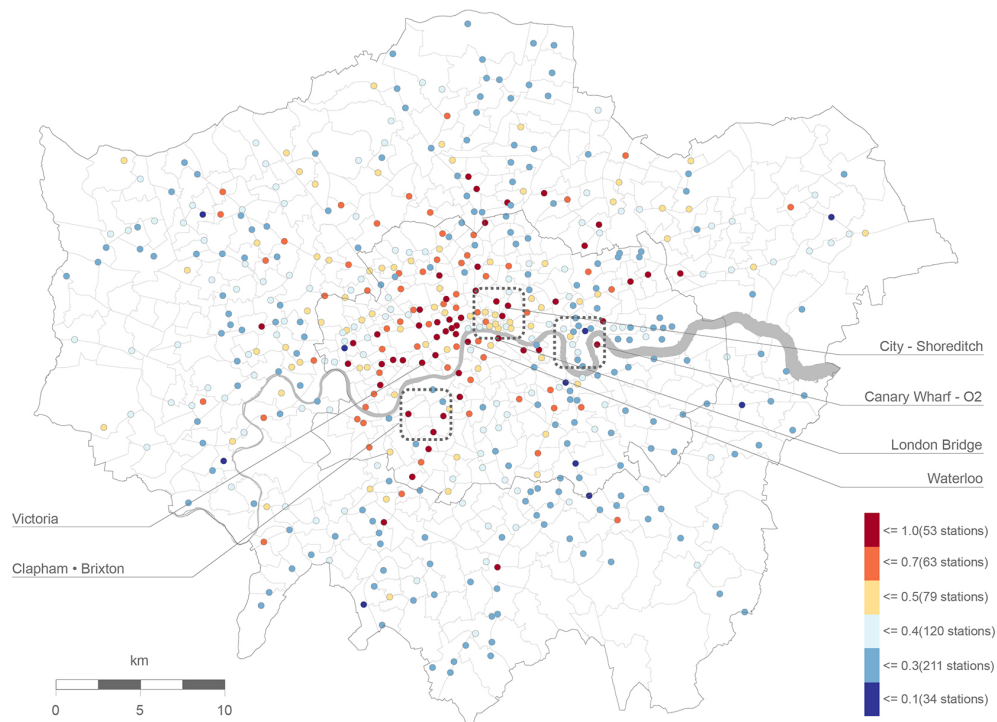
Figure 7: Vitality values for an average weekend day.

(Google has its Campus in Old Street), which generally follow a more varied patterns of working hours than the common ones (9 to 5 o'clock during weekdays).

Figure 8 shows the hourly variations of vitality of an average week and weekend day for the areas around Old Street and Shoreditch (East London), highlighting how vitality has a steady continuity in values during the whole day. Besides the different temporal patterns characterizing the week and weekend, it is interesting to recognize how in these areas the value of vitality increases during weekend also in relation to a greater variety in the temporal distribution of flows. This is not only related to commuting for work, but to the many different purposes that attract people in the areas (we can also observe the long tail of the night flows during the weekend). Figure 9 shows the hourly variations of vitality for selected areas across the city of London.

These results show how our combined metric captures urban vitality in relation to the variations of the presence of people in places, and not only to the magnitude of flows. It measures the diversity of temporal patterns in places in terms of the number of people, the continuity of their presence in space, and the temporal variations in relation to known activities (i.e., commuting to work, traveling for leisure, nightlife, etc.), all combined in one value. This represents an improvement in comparison to previous works about vitality, because it captures two elements that Jacobs considered essential to define the vitality of a place: the continuity in the presence of people in places, and the variety of these people (not only inhabitants of the neighborhood or people working there, but also "strangers").

We believe this is a significant novelty in our work. The metric represents not only the number of people present in a place at a certain moment in time, but it combines that simultaneously with two attributes that evaluate the variation of that value in the day (consistency) and the week (variability). Our metric encapsulates a temporal comparison of flow patterns and evaluates vitality according to the presence of people in places as a whole. The result is that places that are crowded only at specific times (i.e., peak-hours) are not instantly considered vital, a combination of the temporal patterns of flows in places is needed.

Another improvement of our work is the modality to calculate diversity and vitality. Rather than qualitatively estimating the presence of people in a place indirectly from its morphological features, we calculated it directly using quantitative information about human presence in space. Furthermore, we measured vitality at a higher spatiotemporal resolution than previous related works—not at the borough level but at specific locations, not as an absolute temporal value but as two average weekly values and hourly series that highlight variations in spatial patterns. Our analysis and results comprehend multiple local areas in the city for an extended interval of time. We decided to evaluate vitality for specific areas rather than averaging the results for wider areas because we consider that one of the strongest advantages of using these data sets for spatial analysis is to capture and uncover significant differences in the patterns of human spatial behavior in cities.

5.2   Validation

To evaluate the performance of our regression model, we observe the Adjusted-R^2 parameter. The value obtained for the week data set (Adjusted-R^2= 0.843) and the weekend data set (Adjusted-R^2 = 0.831) suggest that our model is working correctly.

In addition to this, we also compared the ranking of the areas of research according to the number of tweets and the final values of vitality calculated with the Oyster data, using a test data set selected when preparing the data for the regression model (random *train-test-split*. An established method to verify the relationship between two rankings (and specifically measuring their degree of similarity) is to calculate the rank correlation using the Spearman's rank correlation coefficient following the equation below (Equation 6):

$$\rho = 1 - \frac{6 \sum d_i^2}{n^3 - n} \tag{6}$$

The results (week: $\rho = 0.70$, $p < .001$, weekend: $\rho = 0.57$, $p < .001$) indicate that the ranking of the areas of research predicted by the model is consistent with the ranking obtained according to the number of tweets.

As mentioned in Section 2, we considered the continuity of flows more relevant than the intensity of flows alone when evaluating vitality. The spatial distribution of different typologies of activities can influence this continuity, attracting a variety of people with different purposes. Therefore, this aspect of the diversity of flows also needs to be considered.

In her work, Jacobs distinguishes between primary and secondary diversity [19], a mix of different activities and purposes working in synergy to attract and aggregate people in the same urban spaces. Therefore, we decided to explore a number of typologies of flows driven by multiple purposes (work, leisure, nightlife, social activities) that may significantly influence the presence of people in urban space. Previous work presents examples of validation using multi-sourced data sets. In [22] data sets sourced from mobile phones,
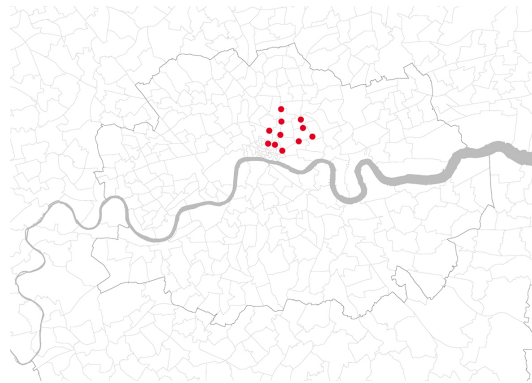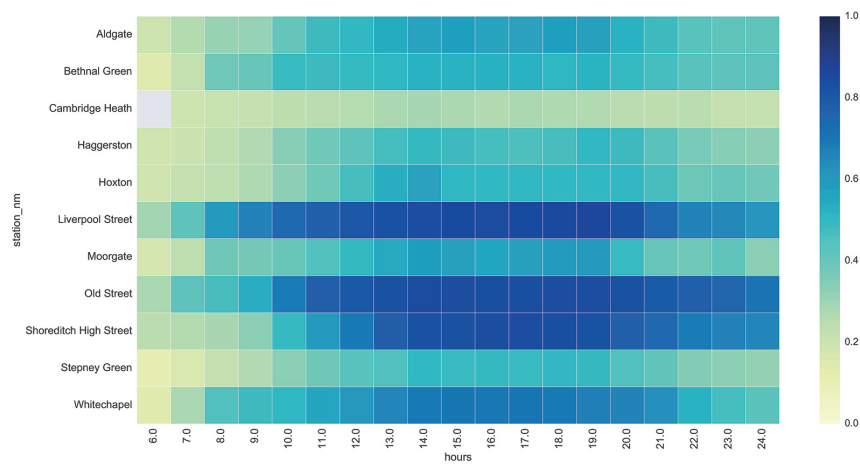
(a) selected areas
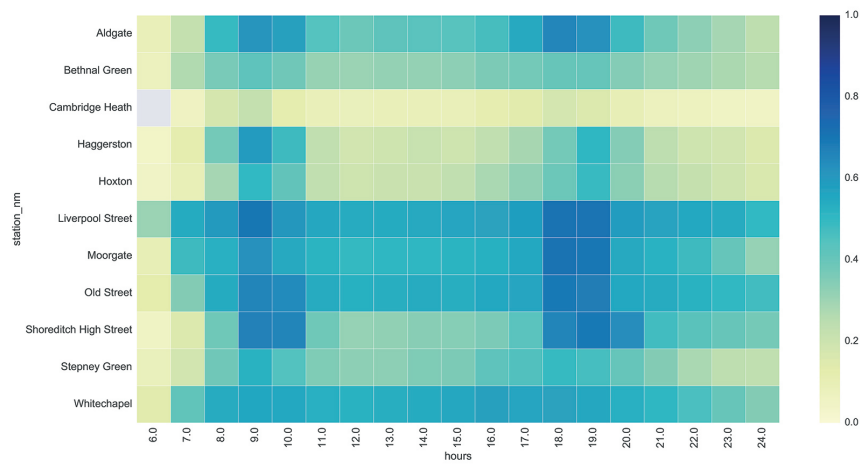
(b) weekdays

(c) weekend

Figure 8: Hourly variation of vitality in East London.

Twitter, and the Census are used to establish how precisely they are capturing human mobility flows over time. In [29] mobile phone data and location data extracted from Yellow Pages are used to establish the relationship between the spatiotemporal patterns of human mobility and the human activity places in Rome.

We conducted some preliminary analysis using points of interest (POI) extracted from OpenStreetMap to obtain an index representing the spatial distribution and diversity of activities in areas. We then calculated the correlation between the vitality metric and the total number of activities for each area of research. The results obtained so far appear incongruous ($r = 0.43$, while $p < .001$). This relation will be further investigated in future work.

6   Discussion

The main objective of this paper is to explore the possibility of building urban metrics using extensive location data sets that have become recently available. Such metrics make it possible to obtain a quantitative evaluation of particular urban dynamics and to achieve a better understanding of phenomena related to spatial dynamics and patterns of human mobility in cities.

Undoubtedly, the concept of vitality is widely understood in many different ways. Jane Jacobs defined it only through empirical observations and experience, but actually never attempted at testing or quantifying it. Our metric captures the dynamic attributes defined for diversity in relation to the variations in the pattern of flows. These elements are recognized by urbanism theory as strongly related to urban vitality, which is commonly understood more as a quality of a place. In some way, in our work we are attempting at quantifying this quality. This can be done in many different ways and we think that our method follows a coherent interpretation of Jacobs' understanding. We calculated how diverse the places are in terms of flow patterns, and this diversity and temporal variations are the constituent elements of urban vitality. Indeed, one may argue that urban vitality, as an intangible, holistic phenomenon, cannot be captured in its entirety by quantitative methods. Nevertheless, in this work we believe we have captured accurate measurements of diversity and vitality that relate to the human flows and presence of people in urban places. These metrics enable scholars and planners to analyze the phenomena with quantitative information that can be consequently used in their comprehensive evaluation of a place.

Although this can be considered as a limited measurement of a complex phenomenon such as urban vitality, nevertheless the metrics add quantitative information that can be usefully employed to make choices for planning and policy-making. Quantitative indicators can contribute to support strategies and decisions by urbanists, policy-makers, and other stakeholders involved in the process of urban planning. Understanding spatiotemporal dynamics and patterns from a quantitative perspective at this level of detail can also usefully inform transport planning and management in the circumstance of specific events that may alter the established routines of places. Regarding the spatial decision process related to zoning and masterplans, a better comprehension of the quantitative relationship between vitality and spatial distribution of activities can inform spatial developments in terms of urban policies and plans concerning land use. These metrics can be used to perform an *ante* and *post* development quantitative evaluation of the application and implementation of the chosen spatial strategies and policies. Furthermore, the validation of the

model shows an evident correlation between the order of vital areas obtained by mobility data and that obtained by social media data. This means that the same methodology can be eventually adapted and applied to other data sets in different contexts (i.e., to social media data when more reliable data about human presence in space is not available).

6.1   Limitations

The main limitation of the results obtained in this work is related to the representativeness of the data. This is a common and well-known concern in research involving these types of data, particularly in relation to Twitter data and the narrow population sample represented (Section 2). It is also important to mention that, regarding the public transport data from Oyster card, start and end of journey points are available only for the rail network (Underground, Overground, DLR, etc.), whereas for bus journeys the only spatial data available is the boarding point. Consequently, some of the measurements for specific points in time (i.e., working locations during commuting peak hours) will be only partially representing the actual dynamics of the places.

Another limitation, intrinsic of the Oyster data set, is related to the spatial distribution of the sensors in the city. This means that results cannot be extended to larger areas than those analyzed in the research without losing accuracy. The same type of spatial constraint is common to other types of data collected through sensors (i.e., mobile phone data captured by antennas). Their results are often extended to larger areas averaging finer-granularity, point-based values. For this work, we think that it is more meaningful to present detailed results instead of averaging them. We should also notice that central areas of the city may still be biased because of their location in relation to the entire urban network. The normalization of the attribute of intensity using the "integration" score does not completely cancel the influence of the magnitude of flows that interest central places in London.

A limitation in the methodology is represented by employing Twitter data as an approximation for the density of people activity in space in the regression model. Although various examples in the current literature use this type of data to measure human presence and dynamics in space, we are aware that our assumption may be prone to error in relation to the non-ideal representativeness of the Twitter data. For this specific reason, part of the future work of this research will be dedicated to build a ground truth training data set to feed the model with real-world data, using people insights and collecting their experience of vitality in urban space through surveys. Finally, it is worth pointing out that the OLS (ordinary least square) regression model selected for this research allows for negative weights, which in this particular case might seem like a counter intuitive result. However, we can simply rescale the regression to avoid negative numbers in the predictions and this serves to counter the problem of the model producing unrealistic numbers. As such, the model is indicative of vitality rather than predictive in the numerical sense. In future work, we intend to test other variants of the models such as that based on Poisson regression which would avoid such problems. We speculate that such analysis would not produce very different results, but this is for future research.

Despite the limitations discussed above, we showed that our computational application of diversity and vitality contributes meaningfully to the current research on the topic. It is not in the objective of this study to obtain an absolute representation of urban reality, particularly when dealing with a concept such as vitality, for which a definition and understanding are widely debated in urban research. The objective of this work is to demonstrate
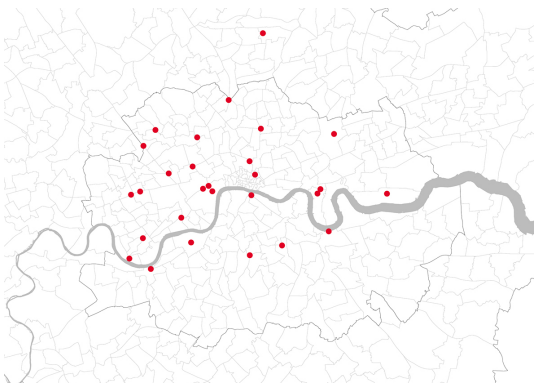
(a) selected areas
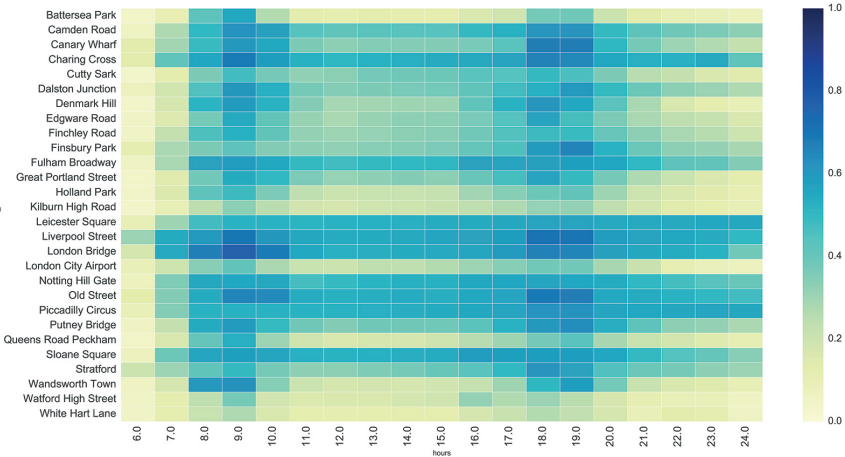
(b) weekdays
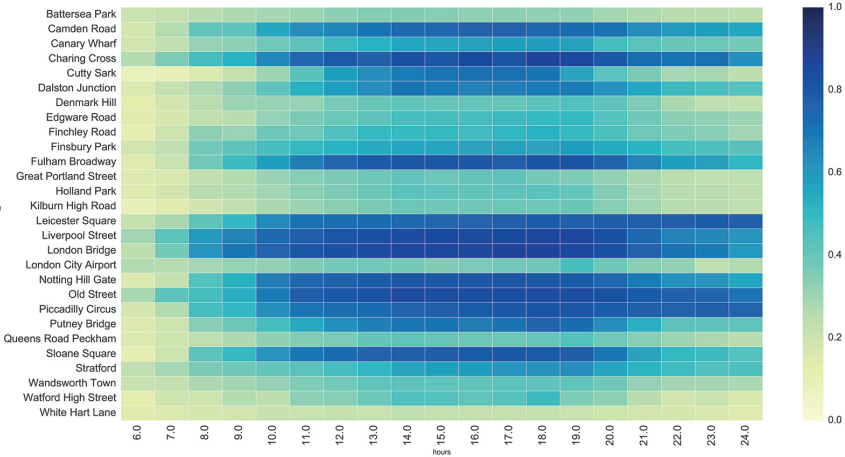
(c) weekend

Figure 9: Hourly variation of vitality for selected stations.

how it is possible to use data sets representing human mobility to build metrics about urban space that can quantitatively better support urban planning decisions and processes.

7 Conclusion

In this paper, we proposed a computational application of the concept of urban vitality defined by Jane Jacobs, using mobility data sourced from the Oyster card as a proxy for measuring indicators as diversity and vitality in relation to the spatial dynamics and the presence of people in different areas of London. Our main objective was to obtain a quantitative evaluation of these well-known urbanism concepts that, although frequently used, have not been largely explored yet from a quantitative perspective. We defined three dynamic attributes (intensity, variability, and consistency) that measure the variation in patterns of human dynamics according to the number of people, the continuity of their presence in places, and the variations over different time intervals in the week. We applied a multivariate linear regression model to obtain the weight for each attribute, and then calculated the final value of urban diversity.

The results show that vitality is higher in central places of the city, as one would expect from the daily experience. However, the high-resolution of the data sets employed in this work, and the analysis performed at a finer spatial and temporal scale, make possible to observe meaningful variations in adjacent central places that present significant differences in spatial behavior over time. Also, interesting exceptions are detectable in the peripheral areas, presumably related to flows driven by leisure and tourism. The application of the combined metric for calculating vitality, rather than only considering the magnitude of flows (as in related works), makes possible to catch variations in the mobility patterns of people, and to highlight the aspect of continuity of their presence in places, which is a significant element for vitality as understood by Jacobs. Finally, our work contributes to improving the capability of urbanism and urban analysis in the following ways.

- It provides a better understanding of spatial phenomena at a closer, more detailed scale.
- It illustrates how it is possible to develop metrics about urban space using data sets that supply additional information beyond urban morphology.
- It attempts to deviate from a data-driven approach, using well-known concepts of urbanism theory to drive the data analysis and evaluation.

Acknowledgements

References

[1] BERLINGERIO, M., CALABRESE, F., LORENZO, G. D., DONG, X., GKOUFAS, Y., AND MAVROEIDIS, D. SaferCity: A System for Detecting and Analyzing Incidents from

Social Media. In *2013 IEEE 13th International Conference on Data Mining Workshops* (2013), pp. 1077–1080. doi:10.1109/ICDMW.2013.39.

[2] BLANK, G.   The Digital Divide Among Twitter Users and Its Implications for Social Research.   *Social Science Computer Review 35*, 6 (2017), 679–697. doi:10.1177/0894439316671698.

[3] CALABRESE, F., COLONNA, M., LOVISOLO, P., PARATA, D., AND RATTI, C. Real-Time Urban Monitoring Using Cell Phones: A Case Study in Rome. *IEEE Transactions on Intelligent Transportation Systems 12*, 1 (2011), 141–151. doi:10.1109/TITS.2010.2074196.

[4] CALABRESE, F., PEREIRA, F. C., LORENZO, G. D., LIU, L., AND RATTI, C. The Geography of Taste: Analyzing Cell-Phone Mobility and Social Events. In *Pervasive Computing* (2010), Lecture Notes in Computer Science, pp. 22–37. doi:10.1007/978-3-642-12654-3_2.

[5] CHENG, Z., CAVERLEE, J., LEE, K., AND SUI, D. Z. Exploring millions of footprints in location sharing services. *Proc. the 5th International Conference on Weblogs and Social Media, ICWSM 2011 2011* (2011), 81–88.

[6] CRANSHAW, J., SCHWARTZ, R., HONG, J., AND SADEH, N. The Livehoods Project: Utilizing Social Media to Understand the Dynamics of a City. In *Proc. the 6th International Conference on Weblogs and Social Media, ICWSM 2012* (2012), pp. 58–65.

[7] DANIELS, R., AND MULLEY, C. Explaining walking distance to public transport: The dominance of public transport supply. *Journal of Transport and Land Use 6*, 2 (2013), 5–20. doi:10.5198/jtlu.v6i2.308.

[8] DE NADAI, M., STAIANO, J., LARCHER, R., SEBE, N., QUERCIA, D., AND LEPRI, B. The Death and Life of Great Italian Cities: A Mobile Phone Data Perspective. In *Proceedings of the 25th International Conference on World Wide Web* (2016), pp. 413–423. doi:10.1145/2872427.2883084.

[9] EL-GENEIDY, A., GRIMSRUD, M., WASFI, R., TETREAULT, P., AND SURPRENANT-LEGAULT, J. New evidence on walking distances to transit stops: identifying redundancies and gaps using variable service areas. *Transportation 41*, 1 (2014), 193–210. doi:10.1007/s11116-013-9508-z.

[10] FERRARI, L., MAMEI, M., AND COLONNA, M. Discovering events in the city via mobile network analysis. *Journal of Ambient Intelligence and Humanized Computing 5*, 3 (2014), 265–277. doi:10.1007/s12652-012-0169-0.

[11] GEHL, J. *Life Between Buildings: Using Public Space*. Island Press, 2011.

[12] GEHL, J., AND SVARRE, B. *How to Study Public Life*. Island Press, 2013.

[13] GIRARDIN, F., CALABRESE, F., FIORE, F. D., RATTI, C., AND BLAT, J. Digital Footprinting: Uncovering Tourists with User-Generated Content. *IEEE Pervasive Computing 7*, 4 (2008), 36–43. doi:10.1109/MPRV.2008.71.

[14] GOODCHILD, M. F. Citizens as sensors: the world of volunteered geography. *GeoJournal 69*, 4 (2007), 211–221. doi:10.1007/s10708-007-9111-y.

[15] GRAUWIN, S., SOBOLEVSKY, S., MORITZ, S., GODOR, I., AND RATTI, C. Towards a Comparative Science of Cities: Using Mobile Traffic Records in New York, London, and Hong Kong. In *Computational Approaches for Urban Environments*. Springer International Publishing, 2015, pp. 363–387. doi:10.1007/978-3-319-11469-9_15.

[16] HASAN, S., SCHNEIDER, C. M., UKKUSURI, S. V., AND GONZALEZ, M. C. Spatiotemporal Patterns of Urban Human Mobility. *Journal of Statistical Physics 151*, 1-2 (2013), 304–318. doi:10.1007/s10955-012-0645-0.

[17] HAWELKA, B., SITKO, I., BEINAT, E., SOBOLEVSKY, S., KAZAKOPOULOS, P., AND RATTI, C. Geo-located Twitter as proxy for global mobility patterns. *Cartography and Geographic Information Science 41*, 3 (2014), 260–271. doi:10.1080/15230406.2014.890072.

[18] HILLIER, B., PENN, A., HANSON, J., GRAJEWSKI, T., AND XU, J. Natural movement - or, Configuration and attraction in urban pedestrian movement. *Environment and Planning B: Planning and Design 20*, 1 (1993), 29–66.

[19] JACOBS, J. *The life and death of great American cities*. New York: Random House, 1961.

[20] KITCHIN, R. *The data revolution: Big data, open data, data infrastructures and their consequences*. Sage, 2014.

[21] LEE, R., WAKAMIYA, S., AND SUMIYA, K. Discovery of unusual regional social activities using geo-tagged microblogs. *World Wide Web 14*, 4 (2011), 321–349. doi:10.1007/s11280-011-0120-x.

[22] LENORMAND, M., PICORNELL, M., CANTU-ROS, O. G., TUGORES, A., LOUAIL, T., HERRANZ, R., BARTHELEMY, M., FRIAS-MARTINEZ, E., AND RAMASCO, J. J. Cross-Checking Different Sources of Mobility Information. *PLOS ONE 9*, 8 (2014), e105184. doi:10.1371/journal.pone.0105184.

[23] MORSTATTER, F., PFEFFER, J., LIU, H., AND CARLEY, K. M. Is the sample good enough? Comparing data from twitter's streaming API with Twitter's firehose. In *Proc. the 7th International Conference on Weblogs and Social Media, ICWSM 2013* (2013), pp. 400–408.

[24] MUNIZAGA, M. A., AND PALMA, C. Estimation of a disaggregate multimodal public transport Origin-Destination matrix from passive smartcard data from Santiago, Chile. *Transportation Research Part C: Emerging Technologies 24* (2012), 9–18. doi:10.1016/j.trc.2012.01.007.

[25] NES, A. V. *Introduction to Configurative Methods in Urban Studies*. Techne Press, Delft, 2008.

[26] NOULAS, A., SCELLATO, S., LAMBIOTTE, R., PONTIL, M., AND MASCOLO, C. A Tale of Many Cities: Universal Patterns in Human Urban Mobility. *PLOS ONE 7*, 5 (2012), e37027. doi:10.1371/journal.pone.0037027.

[27] PEREIRA, F. C., RODRIGUES, F., AND BEN-AKIVA, M. Using Data From the Web to Predict Public Transport Arrivals Under Special Events Scenarios. *Journal of Intelligent Transportation Systems 19*, 3 (2015), 273–288. doi:10.1080/15472450.2013.868284.

[28] RATTI, C., FRENCHMAN, D., PULSELLI, R. M., AND WILLIAMS, S. Mobile Landscapes: Using Location Data from Cell Phones for Urban Analysis. *Environment and Planning B: Planning and Design 33*, 5 (2006), 727–748. doi:10.1068/b32047.

[29] READES, J., CALABRESE, F., AND RATTI, C. Eigenplaces: Analysing Cities Using the Space-Time Structure of the Mobile Phone Network. *Environment and Planning B: Planning and Design 36*, 5 (2009), 824–836. doi:10.1068/b34133t.

[30] READES, J., ZHONG, C., MANLEY, E., MILTON, R., AND BATTY, M. Finding Pearls in London's Oysters. *Built Environment 42*, 3 (2016), 365–381. doi:10.2148/benv.42.3.365.

[31] ROTH, C., KANG, S. M., BATTY, M., AND BARTHELEMY, M. Structure of Urban Movements: Polycentric Activity and Entangled Hierarchical Flows. *PLOS ONE 6*, 1 (2011), e15923. doi:10.1371/journal.pone.0015923.

[32] SAKAKI, T., OKAZAKI, M., AND MATSUO, Y. Earthquake Shakes Twitter Users: Real-time Event Detection by Social Sensors. In *Proc. the 19th International Conference on World Wide Web* (2010), pp. 851–860. doi:10.1145/1772690.1772777.

[33] SEABORN, C., ATTANUCCI, J., AND WILSON, N. Analyzing Multimodal Public Transport Journeys in London with Smart Card Fare Payment Data. *Transportation Research Record: Journal of the Transportation Research Board 2121* (2009), 55–62. doi:10.3141/2121-06.

[34] SUNG, H.-G., GO, D.-H., AND CHOI, C. G. Evidence of Jacobs's street life in the great Seoul city: Identifying the association of physical environment with walking activity on streets. *Cities 35* (2013), 164–173. doi:10.1016/j.cities.2013.07.010.

[35] TRAAG, V. A., BROWET, A., CALABRESE, F., AND MORLOT, F. Social Event Detection in Massive Mobile Phone Data Using Probabilistic Location Inference. In *2011 IEEE Third International Conference on Privacy, Security, Risk and Trust and 2011 IEEE Third International Conference on Social Computing* (2011), pp. 625–628. doi:10.1109/PASSAT/SocialCom.2011.133.

[36] ZHONG, C., ARISONA, S. M., HUANG, X., BATTY, M., AND SCHMITT, G. Detecting the dynamics of urban structure through spatial network analysis. *International Journal of Geographical Information Science 28*, 11 (2014), 2178–2199. doi:10.1080/13658816.2014.914521.

[37] ZHONG, C., BATTY, M., MANLEY, E., WANG, J., WANG, Z., CHEN, F., AND SCHMITT, G. Variability in Regularity: Mining Temporal Mobility Patterns in London, Singapore and Beijing Using Smart-Card Data. *PLOS ONE 11*, 2 (2016), e0149222. doi:10.1371/journal.pone.0149222.