Exploring the relationship between urban vitality and the distribution of amenity typologies

Patrizia Sulis[1], Ed Manley[2]

[1,2]*Centre for Advanced Spatial Analysis - UCL*
patrizia.sulis.14, ed.manley@ucl.ac.uk

Abstract: The availability of large datasets containing spatiotemporal information about human mobility in cities can reveal useful insights about how people use urban space across different times, and be employed to improve urban qualities. Urban vitality represents a critical quality, strictly related to the continuity of the presence of people in places. Vitality may be significantly driven by the types of amenities distributed in the city. This work explores the quantitative relationship between urban vitality and the variety of amenities in London, employing datasets sourced from smart cards, social media and activity location. Vitality is calculated as the temporal variation of patterns in the presence of people in a place. It is compared with the amenity typologies located in such place, combined through a clustering analysis of the spatial distribution of amenities across the city. Results suggest that urban vitality increases in areas with a variety of typologies acting in synergy to attract people, rather than in areas characterised by a predominance of a specific type of activities. This paper enhances the theory-informed quantitative understanding of urban space, integrating data analysis methods with urban planning: this approach has significant potential for informing robust and holistic policy and decision-making.

Keywords: urban vitality; urban amenities; cluster analysis; spatial distribution.

Introduction

Urban amenities represent points of attraction in the mobility patterns of cities and may act as drivers for the liveliness of places. For this reason, the spatial information about urban amenities is commonly used in urban analysis as an established and reliable source of spatial information for understanding dynamics in cities.

This work attempts at exploring the quantitative relationship between the urban vitality and the location and spatial distribution of amenities in the city, also following the suggestion made by Jacobs (1961) regarding the spatial features fundamental for vitality. In order to evaluate the results obtained using the new sources of human mobility data recently available, this paper selected some urban data sets that are widely used in urban analysis and are well-established as reliable sources for the description and the understanding of how urban space works.

This paper investigates how the different typologies of the amenities located in a city might influence the variation of its vitality over time. The values of hourly temporal trends of urban vitality, calculated as in (Sulis et al., 2018) employing human mobility data sourced from the smart card and social media data sets, are compared to the spatial distribution of amenity typologies using POIs sourced from Ordnance Survey. In particular, vitality values and temporal trends are compared to specific combinations of amenity typologies, which are calculated through a cluster analysis that associate areas showing the similar spatial distribution of typologies. Evaluating the quantitative relationship between these two spatial features is relevant to understand if the vitality of a place may be driven by a specific typology of activities, or if the synergy amongst different types contributes to the continuity of the liveliness and presence of people across the day.

This work represents an example of how it is possible to combine new and standard data sourced and methods and integrating them into the procedures employed for analysing and planning urban space. Quantifying the relationship amongst urban metrics that also describe different spatial domains (i.e., mobility and morphology) possibly result in unveiling how spatial elements and phenomena characterising a place interact and influence each other. This aspect represents a significant factor to consider when planning and designing any development in cities: tasks such as urban strategy, modelling and policy-making would all benefit from the results of this analysis.

This paper is structured as follows: after a presentation of previous work, the data description and methodology employed in the work are illustrated. After that, the paper describes the results related to the correspondence between the temporal patterns of vitality and those of amenity distribution in a place, highlighting the main results. The discussion of the results, the relevance of the approach and the limitations of the methodology concludes the paper.

Previous work

The recent deluge of spatial information (Batty, 2013) coming from different technological sources and devices recording the human-space interaction has made possible for researchers to explore urban dynamics in cities from an unprecedented level of detail, how people move around in cities, also from a temporal perspective. Numerous studies explored the spatiotemporal patterns of human mobility in cities employing different data sources (Isaacman et al., 2010; Noulas et al., 2012; Hasan et al., 2013; Hawelka et al., 2014; Lenormand et al., 2014; Louail et al., 2014; Grauwin et al., 2015; Zhong et al., 2016).

Besides more general approaches, some investigated very specific ideas related to human mobility, i.e., the digital signature of mobility patterns of specific city users as tourists (Girardin et al., 2009), the patterns of preferences of people within a neighbourhood (Calabrese et al., 2010), or the patterns of routine of users in the district (Cranshaw et al., 2012), unveiling patterns at a very detailed scale. All these studies contribute to unveil urban space in terms of how it is used, somehow adding a quantitative perspective to the quality of spaces that were estimated with more empirical tools (Gehl, 2010).

Amongst urban qualities, liveliness or urban vitality (Jacobs, 1961) represents a critical one to understand how successful a city is. Recent work attempted at quantifying this quality using extensive data sets about human mobility recently available (Sung et al., 2013, De Nadai et al., 2016), also considering the fundamental features of temporal continuity highly regarded by Jacobs (Sulis et al., 2018).

To explore more human mobility and what drives it, some work also employed multiple data sources to explore the relationship between different urban features such as human mobility and the locations of urban activities (leisure, commercial etc) to establish if human activity is driven by that and to which measure. Previous work shows interesting results in terms of comparing the average values of human activity (from mobile phone data) and location of urban amenities (Reades et al., 2009), or the correspondence of absolute values of vitality (as a proxy of human activity) and specific categories of amenities related to leisure (De Nadai et al., 2016).

Following these work, and the idea suggested from Jacobs of the importance of variety in amenities in relation to urban vitality, this work employs the hourly values of urban vitality calculated as the temporal variation of the presence of people in urban places in (Sulis et al., 2018): the variation is calculated at three different temporal scales (hourly, daily and weekly), employing human mobility data sourced from the Oyster smart card and the

Twitter platform. The vitality values are compared to the spatial distribution of urban amenities related to the leisure domain, in the attempt of evaluating the quantitative relationship of these features and the possibility of amenity location driving the level of the liveliness of places in the city of London.

Data description and methodology

Description of data

This work investigates the quantitative relationship between urban vitality and the distribution of POIs in cities. The analysis explores how the different combinations of amenity typologies located in a place (the mixed-use mentioned by Jacobs, 1961) influence the temporal patterns of vitality in cities. This approach may reveal more accurate insights about a specific combination of amenities possibly driving the vitality of a place. The metric of vitality was calculated using different sources of human mobility data as presented in (Sulis et al., 2018). These values are here compared to spatial information about city amenities collected from Ordnance Survey.

The data set sourced from Ordnance Survey contains information about the location of Points of Interest (POI) in London. The POI categories selected for this analysis are related to the leisure domain, since the focus of this analysis is on urban phenomena influencing the presence of people in urban public places: only the categories related to non-shopping leisure (e.g., eating out, tourist attraction, entertainment) are selected, following the idea of streets as public space. Figure 1 shows the spatial distribution of the data included in the Ordnance Survey data set, and it illustrates the density of amenities within the London wards.

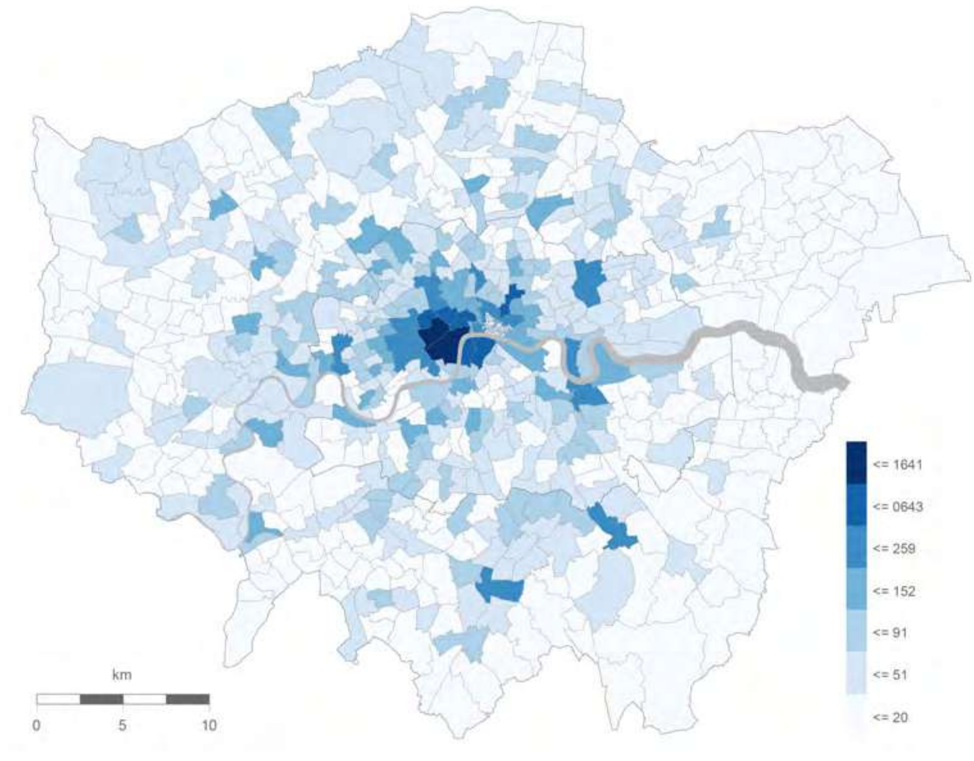

Figure 1: spatial distribution of urban amenities

The main advantage of employing this type of data is that it represents a source of spatial information about cities that is frequently used in urban planning and considered a well-established and accurate source of data for investigating the use of places and the mobility of people in cities. Furthermore, the information supplied by this type of data, especially the geographical location and the category of the activity, can be combined with other metrics about urban places, e.g., human activity density obtained from the other data sets, making possible to investigate the relationship amongst the different types of data describing the same urban phenomena.

The possibility of enriching the data with additional information (e.g, opening hours) and the detailed categorisation of amenities also makes possible to explore how specific typologies drive or influence more than others the human mobility patterns and the presence of people in places during specific times, and the impact of this possible influence.

Methodology

At first, a profile of the amenity distribution characterising each area of the study is extracted from the data set: this profile contains the total number of amenities for each typology present in that area. These distribution profiles are then used to run a cluster analysis to inferr the similarity of the areas according to their patterns of amenity distribution, assigning each area to a specific group. This approach represents a more accurate method for representing the specific characteristic of each area concerning the spatial distribution of urban amenities than the simple sum of activities located around each place. The labels assigned to the areas are subsequently compared to the temporal variation of vitality, to observe if they are recurrently associated, and eventually verify the correspondence between a specific combination of amenity typologies and the liveliness of a place.

The distance metric considered to be the most appropriated to be applied in this specific case is the Jensen-Shannon Divergence (JSD), applied to calculate the distance metric amongst the amenity profiles of the research areas. This method is specifically designed to measure the similarity between two (or more) probability distributions P, Q: in this case, it is measuring how similar the spatial distribution of amenity typologies is between two different areas in the city. The JSD metric is calculated according to Equation 1 below:

$$JSD(P\|Q) = \frac{1}{2} \cdot D(P\|M) + \frac{1}{2} \cdot D(Q\|M)$$

The square root of the divergence is the equivalent distance metric required for the cluster analysis. In this case, the JSD is applied to a matrix of vectors, with each vector representing the POI distribution calculated for each research area. The matrix employed in the analysis has been normalised in advance by the sum of the activities located in each area.

After calculating the distance metrics, the analysis requires a suitable algorithm to cluster the distributions of POI typologies. In this case, the one chosen for the cluster analysis is HDBSCAN. Results of the cluster analysis are presented in the following section.

The labels obtained in the cluster analysis of the profiles of amenity distribution are compared to the trends of the temporal variation of vitality. This comparison detects the occurrence of the same association between amenity labels and vitality trends in the areas, which may be used to unveil if the specific distribution of amenities in an area influences the liveliness and the continuous presence of people in a place. To compare these two features, the analysis employs a confusion matrix. The algorithm counts each time that the same pair of vitality trend and amenity label are associated. This value may be employed to evaluate the relationship between the vitality of an area and the diversity of amenities in the same area.

The code used to calculate the confusion matrix is the following:

```
L = len(np.unique(df['vitality_labels']))
B = len(np.unique(df['activities_labels']))

labels_M = np.zeros((L, B))

for i in df.index:
    x = df.iloc[i]['vitality_labels']
    y = df.iloc[i]['activities_labels']

    labels_M[x, y] = labels_M[x, y] + 1
```

The confusion matrix is calculated for different combinations of labels:

- temporal variation of vitality (week) + distribution of POI typologies;
- temporal variation of vitality (weekend) + distribution of POI typologies.

The results (illustrated in the following section) show which amenity labels are more frequently associated to a specific vitality trend and pattern of human activity in a place, corresponding to a specific distribution and amount of amenity typologies. This correspondence might suggest that a particular combination of typologies influences and possibly drives a certain level of vitality in an area. These results, also showing different label associations occurring during week and weekend days, can then be used to further analysis and empirical investigation.

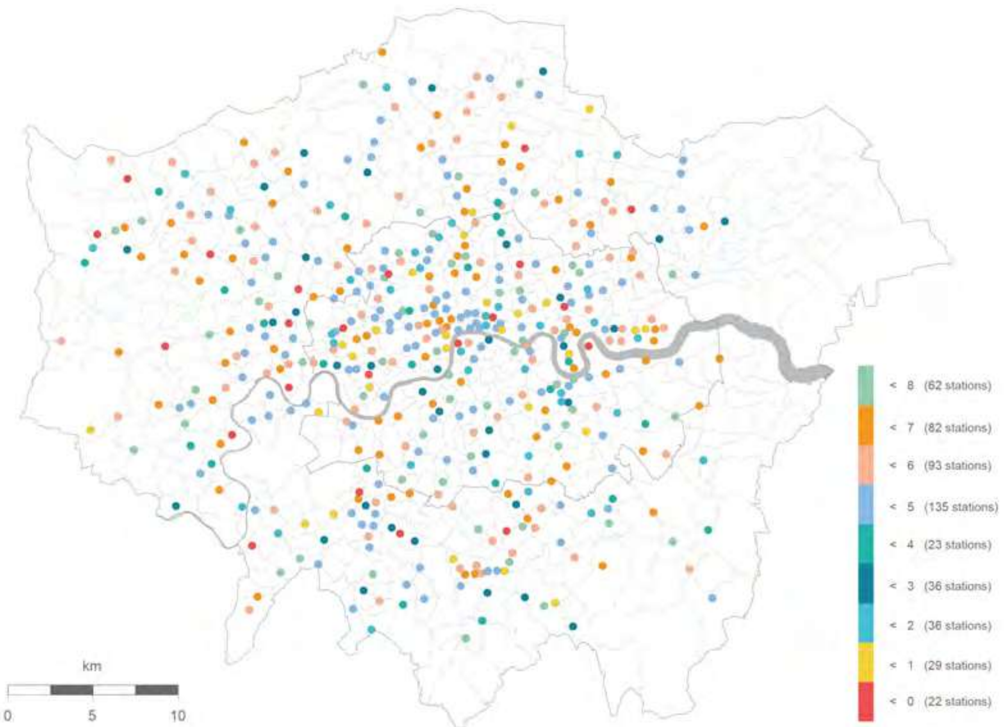

Figure 2: spatial distribution of labels in London

Results

This section presents the results of the cluster analysis applied using the Jensen-Shannon distance and HDBSCAN combination. Observing the spatial distribution for each label (Figure 2), it is possible to obtain revealing insights about the patterns of similarity of amenity distribution in the city of London.

Labels 5 and 9 appears to be the more varied and balanced ones, with the presence of many typologies related to city attractions and entertainment. Label 5 appears to be the most common label, with an even distribution around the city. This aspect may be related to the fact that the analysis considered areas very closed to the public transport network, which generally acts as an attractor for activity locations. Other interesting labels to observe in relation to the combination of amenity typologies are the labels 1 and 6. Label 1 shows a spike in the presence of restaurants in the profile, and it appears to be mainly located in the inner area of the city. Label 6 instead appears to be quite evenly distributed on the map, whereas the profile shows a significant presence of libraries amongst the other typologies.

These labels are compared to the temporal trends of vitality using a confusion matrix. The results show a neat distinction in the label associations during the weekdays. The majority of the associations occurs with label 5 and label 7, which correspond to two rather different amenity distribution profiles. The temporal trends of vitality associated to these two labels are also different: trends representing a temporal continuity in the vitality are associated to label 5, which presents a nicely balanced distribution of a variety of amenity typologies (eating, entertainment, attraction). On the other hand, the vitality patterns characterised by a two-peak profile are predominantly associated with label 7, showing a presence of amenities pertaining exclusively to the eating category. In the weekend, the majority of label associations are still occurring with label 5 and 7; however, the distinction in the vitality trends appears less defined, therefore more analysis is required.

The most representative associations of vitality trends and activity distribution labels are illustrated in Figures 3 - 4 below.

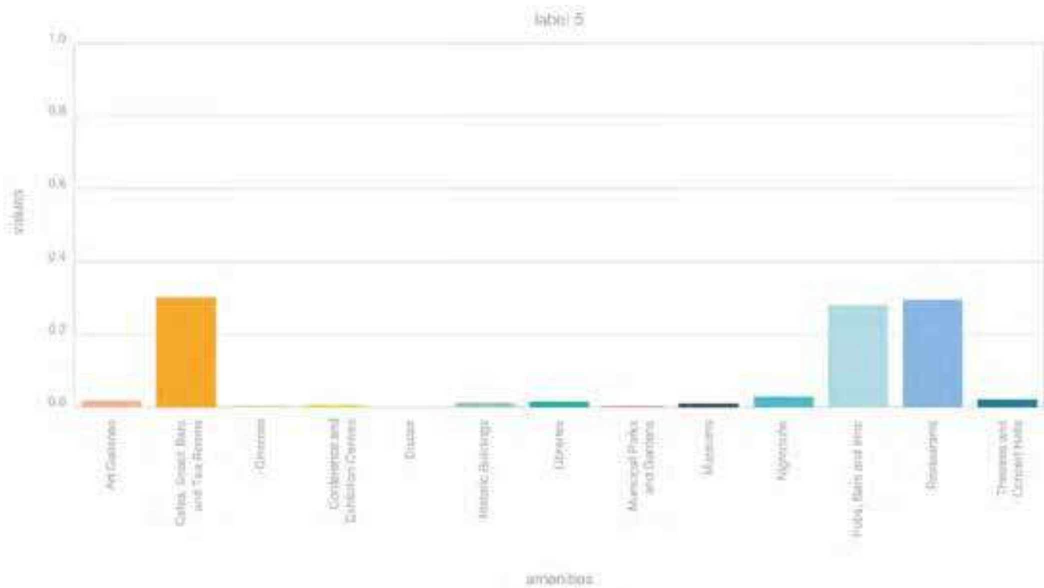

Figure 3: typology distribution for label 5

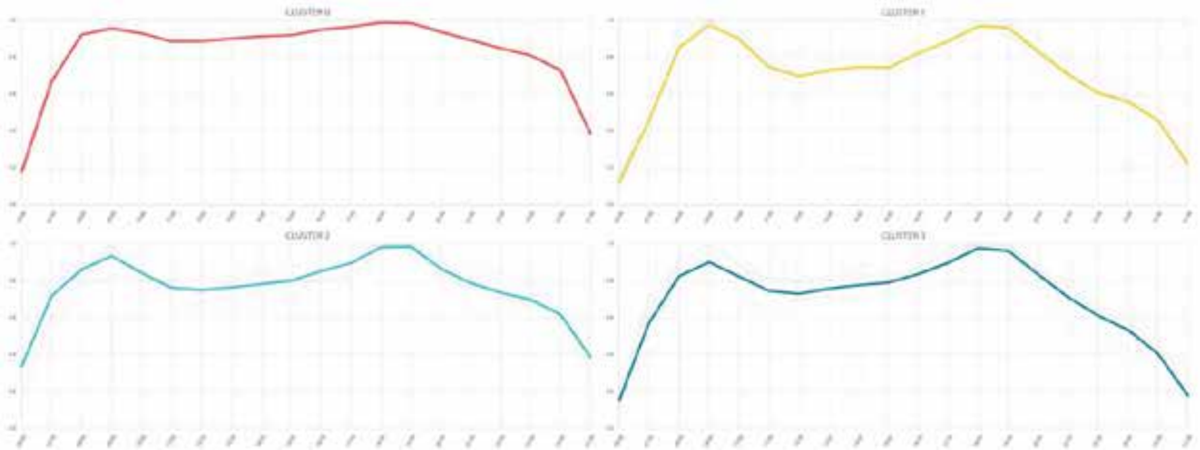

Figure 4: temporal trends of vitality associated with label 5

Discussion

This section discusses the results obtained in the analysis, also in light of the relevance and the potential applications in the urban planning methods. Firstly, it is important to underline how this work showed that it is possible to explore and measure relationships amongst different spatial features of the same place employing a quantitative method. Estimating this relationship represents an additional enhancement in the process of unveiling details about the spatial phenomena occurring in cities.

About the relationship between vitality trends and specific combinations of amenity typologies, results do not appear to present a specific association of precise vitality trends with activity labels. What can be noticed, however, is a clear distinction in the association between continuous and two-peak vitality trends and specific distributions of amenities. In particular, results seem to suggest that the continuity in the patterns of vitality (and therefore human activity in places) might be influenced by a balanced variety of activities located in the area (in this case, different leisure activities). This aspect may represent a hint towards what Jacobs described, that the mixed use of amenities contributes to attracting different people for different purposes, therefore ensuring their presence during the day and resulting in the vitality of a place. On the other hand, the two-peak patterns appear to be mostly associated with less varied amenity distributions in the areas. This may suggest that the presence of people in such places is mainly concentrated around specific times of the day (e.g., opening hours of certain amenities).

Unveiling such information about spatial features represents an advantageous factor to be employed in urban planning. Understanding which combination of amenity typologies (which types and in which quantities) most influence the liveliness of a place, and its success in behaving as a public space, is a very relevant element for both analysis and planning phases. It represents a condition that can also be employed in policy-making, for example to target areas that appear to be more secluded in terms of human mobility patterns and the use of the space at different times in the day. Identifying the most successful combinations of amenities that are contributing to the liveliness of a place can also be employed to intervene in other areas with specific policies targeting either the typologies of the amenities to locate in the area or the opening hours of activities already existing, in combination to other urban elements present in the area.

All the aforementioned examples about the applications of the metrics into the planning procedure must inevitably be evaluated considering the existing context and urban situation of the city, district, neighbourhood where the urban development is happening. Most importantly, a critical point to highlight here is that the metrics employed are context-sensitive and prone to a certain level of uncertainty. The urban context on which the spatial analysis is performed results into the metrics being specific to that context: they cannot be applied lightly to other urban situations, or generalised into overall standards. The approach and the methodology can be adapted and applied to other case studies and data sets, contributing to comparative analyses of the same spatial phenomena occurring in other cities.

One of the limitations of this study is related to the data sets employed in the analysis, in particular concerning the accuracy and the availability of the amenity location data obtained from the different sources. Whereas it is important to make the data available and open to everyone since it is built thanks to the users' contribution (as it is the case of OpenStreetMap), the completeness and the homogeneity of the data collected may lack in quality. On the other hand, the accuracy and details of the Ordnance Survey data set employed in this analysis come at the expense of the availability and usability of data for everyone. Another limitation is related to the technique used in the cluster analysis: the HDBSCAN algorithm labels some of the objects as noise, making necessary to reassign some of these objects to the closest cluster. Despite this limitation, this type of cluster algorithm appears to be the most suitable for the data sets employed, based on the evaluation of the preliminary results obtained with other techniques.

Conclusion

This paper illustrated an attempt of establishing the quantitative relationship between the temporal patterns of the vitality of a place and the specific combination of amenity typologies present in the same place. This relationship is explored using a comparison between the temporal variation of vitality and the spatial distribution of amenity typologies in each area.

The results obtained from the analysis suggest that a quantitative relationship might be established between the liveliness and presence of people in a place and the number of activities of that place. Results also suggest that certain combinations in the distribution of amenities (a balanced variety of different amenity typologies) appear to be associated with specific temporal trends in vitality. This might represent a hint that the former might drive or at least encourage the latter. This hint can be further explored in future work both with data analysis and empirical research in the urban space.

The method and results presented in this work also represent an example of the potentialities intrinsic in combining new and standard spatial information and techniques into the methods conventionally employed in urban analysis. Unveiling how different spatial elements characterising the same place interact and influence each other represents a relevant aspect that may inform and support urban planning in many different phases.

This approach also contributes to two other objectives for urbanism:

- integrating different types of data sets, using big and small data together, and attempting at establishing the relationship between different types of urban data pertinent to the same urban context and influencing the same urban phenomena;

- showing how new data sources can be successfully integrated with conventional data sources in urban planning and analysis.

References

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

Calabrese, F., Pereira, F.C., Di Lorenzo, G., Liu, L., Ratti, C., 2010, The geography of taste: analyzing cell-phone mobility and social events, in: *Pervasive Computing*, Springer, pp. 22–37.

Cranshaw, J., Schwartz, R., Hong, J., Sadeh, N., 2012, The Livehoods Project: Utilizing Social Media to Understand the Dynamics of a City, in: *Proceedings of the 6th International Conference on Weblogs and Social Media, ICWSM 2012,* pp. 58–65.

De Nadai, M., Staiano, J., Larcher, R., Sebe, N., Quercia, D., Lepri, B., 2016, The Death and Life of Great Italian Cities: A Mobile Phone Data Perspective, arXiv:1603.04012 [physics].

Gehl, J., Svarre, B., 2013, *How to Study Public Life*, Island Press.

Girardin, F., Vaccari, A., Gerber, A., Biderman, A., Ratti, C., 2009, Towards estimating the presence of visitors from the aggragate mobile phone network activity they generate, in: *Intl. Conference on Computers in Urban Planning and Urban Management*.

Grauwin, S., Sobolevsky, S., Moritz, S., Godor, I., Ratti, C., 2015, 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, pp. 363–387.

Hasan, S., Schneider, C.M., Ukkusuri, S.V., Gonzalez, M.C., 2013, Spatiotemporal Patterns of Urban Human Mobility, *J Stat Phys 151*, 304–318.

Hawelka, B., Sitko, I., Beinat, E., Sobolevsky, S., Kazakopoulos, P., Ratti, C., 2014, Geo-located Twitter as proxy for global mobility patterns, *Cartography and Geographic Information Science 41*, 260–271.

Isaacman, S., Becker, R., Cáceres, R., Kobourov, S., Rowland, J., Varshavsky, A., 2010, A Tale of Two Cities, in: *Proceedings of the Eleventh Workshop on Mobile Computing Systems; Applications, HotMobile '10,* ACM, New York, NY, USA, pp. 19–24.

Jacobs, J., 1961, *The life and death of great American cities*, New York: Random House.

Lenormand, M., Picornell, M., Cantu-Ros, O.G., Tugores, A., Louail, T., Herranz, R., Barthelemy, M., Frias-Martinez, E., Ramasco, J.J., 2014, Cross-Checking Different Sources of Mobility Information, *PLOS ONE 9*, e105184.

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

Noulas, A., Scellato, S., Lambiotte, R., Pontil, M., Mascolo, C., Anel, J.A., 2012, A Tale of Many Cities: Universal Patterns in Human Urban Mobility, *PLoS ONE 7*, e37027.

Phithakkitnukoon, S., Horanont, T., Di Lorenzo, G., Shibasaki, R., Ratti, C., 2010, Activity-aware map: Identifying human daily activity pattern using mobile phone data, in: *Human Behavior Understanding*. Springer, pp. 14–25.

Reades, J., Calabrese, F., Ratti, C., 2009, Eigenplaces: Analysing Cities Using the Space–Time Structure of the Mobile Phone Network, *Environ Plann B Plann Des 36*, 824–836.

Sulis, P., Manley, E., Zhong, C., Batty, M., 2018, Using mobility data as proxy for measuring urban vitality. *Journal of Spatial Information Science* 2018, 137–162.

Sung, H.-G., Go, D.-H., Choi, C.G., 2013, Evidence of Jacobs's street life in the great Seoul city: Identifying the association of physical environment with walking activity on streets, *Cities 35*, 164–173.

Zhong, C., Arisona, S.M., Huang, X., Batty, M., Schmitt, G., 2014, Detecting the dynamics of urban structure through spatial network analysis, *International Journal of Geographical Information Science 28*, 2178–2199.