Special Issue: Spatial Inequalities and Cities

B Urban Analytics and City Science

EPB: Urban Analytics and City Science
2023, Vol. 0(0) 1–17

DOI: 10.1177/23998083231209073
journals.sagepub.com/home/epb
Sage

Spatiotemporal gender differences in urban vibrancy

Thomas Collins
University of Exeter, UK

Riccardo Di Clemente
Northeastern University London, UK

Mario Gutiérrez-Roig
University of Essex, UK

Federico Botta
University of Exeter, UK; The Alan Turing Institute, UK

Abstract

Urban vibrancy is the dynamic activity of humans in urban locations. It can vary with urban features and the opportunities for human interactions, but it might also differ according to the underlying social conditions of city inhabitants across and within social surroundings. Such heterogeneity in how different demographic groups may experience cities has the potential to cause gender segregation because of differences in the preferences of inhabitants, their accessibility and opportunities, and large-scale mobility behaviours. However, traditional studies have failed to capture fully a high-frequency understanding of how urban vibrancy is linked to urban features, how this might differ for different genders, and how this might affect segregation in cities. Our results show that (1) there are differences between males and females in terms of urban vibrancy, (2) the differences relate to 'Points of Interest' as well as transportation networks, and (3) there are both positive and negative 'spatial spillovers' existing across each city. To do this, we use a quantitative approach using Call Detail Record data – taking advantage of the near-ubiquitous use of mobile phones – to gain high-frequency observations of spatial behaviours across Italy's seven most prominent cities. We use a spatial model comparison approach of the direct and 'spillover' effects from urban features on male-female differences. Our results increase our understanding of inequality in cities and how we can make future cities fairer.

Keywords

urban vibrancy, urban gender segregation, mobile phone data, spatial data science

Corresponding author:
Thomas Collins, Department of Computer Science, University of Exeter, Exeter EX4 4PY, UK.
Email: trc207@exeter.ac.uk

Data Availability Statement included at the end of the article

Introduction

As the world continues to urbanize at an unprecedented rate, the lives of city inhabitants are transforming, with both unprecedented opportunities but also growing challenges and complexities that cannot be ignored. The United Nations reported that, by 2050, 68% of the world's population will be living in cities (UN DESA 2019) and that, while urban populations are increasing, rural populations are in decline. This trend toward urban life is thought to be related to economic development alongside changes in social organization (Harris 1990; Bettencourt et al. 2007), how humans use land (Liu et al. 2020), and the drastic changes in the patterns of collective human behaviour (Bettencourt et al. 2007; Thomas 2008). Rapid urban growth is thought to make cities more innovative and generate wealth but can cause large-scale social issues for people and communities. These include reduced housing affordability (Nelson et al. 2002), environmental degradation (El Araby 2002), high crime rates with negative effects on economics, education, and health (Glaeser and Sacerdote 1999), greater disease incidence (Connolly et al. 2021; Santana et al. 2023), and traffic congestion (Zheng et al. 2014; Bassolas et al. 2019; Kalila et al. 2018; Xu et al. 2021).

Tóth et al. (2021) reported that rapid city growth can increase segregation and inequality in urban areas (Tóth et al. 2021). Indeed, in cities in the United States, life expectancy has generally increased in the middle classes whereas, in poorer classes, it has remained the same (Bor et al. 2017). Within the spatial structure of cities, some neighbourhoods have become differentially desirable. More expensive locations force lower-income inhabitants away, and in some cases, to the fringes of cities or areas with increased levels of criminality or poverty (Weller and Van Hulten 2012). This can generate a powerful reinforcement loop thought to block those wishing to move to the area and hinder social mobility further.

One way to understand and quantify socio-spatial segregation in cities has been to use traditional data, like a census. However, by being based only on where people live, such data only ever 'scratch at the surface' regarding the quantification of the fascinating details of urban environments and the relationship to our social lives and our 'quality-of-life' (Entwisle 2007; King 2013). Thus, city planners increasingly look to new technologies to study collective human behaviour and, especially, to characterize mobility patterns (Steenbruggen et al. 2015). Data on broad movement behaviours are now accessible due to widespread interaction with technological systems (González et al. 2008) and computer technology can help to reveal patterns in human behaviour. The world's near-ubiquitous uptake of mobile phone technology and social media generates huge amounts of data on our behaviour and mobility (Lazer et al. 2009; Vespignani 2009; Salesses et al. 2013; Botta et al. 2015; Seresinhe et al. 2016; Preis et al. 2020). From shopping habits (Di Clemente et al. 2018; Bannister and Botta 2021; Xu et al. 2019) to transportation (Su et al. 2022), there is an unlimited array of uses afforded to us due to this new ability to track and record the movements of citizens. This new direction has provided an extraordinary new understanding of urban environments and cities (Batty 2013; Pan et al. 2013; Botta et al. 2015, 2020; Barthelemy 2016).

Mobile phone data can support the study of *urban vibrancy* or *urban vitality*, which measures the activity of urban environments (Sulis et al. 2018; Botta and Gutiérrez-Roig 2021; Wang et al. 2021). Urban vibrancy is a concept that has been extensively theorized. Jane Jacobs was hugely influential in highlighting how urban design could encourage urban vibrancy and her arguments often focused on the maintenance and provision of social interactions in cities (Jacobs 1961). Her greatest addition to theory is an understanding that density and diversity in the physical structure of an urban place might affect its functional use (Moroni 2016) and that locations that are more diverse – or more concentrated, in terms of their street networks, buildings, or 'Points of Interest' – may be the most vibrant locations. Thus, city planners should consider diversity and social accessibility because diversity provides social cohesion and supplies opportunities for spontaneous interactions, subsequently allowing high levels of creativity and activity that are accessible to the inhabitants and also maintain the community with a diverse socioeconomic background (Perrone 2019).

We follow on from previous work by Botta and Gutiérrez-Roig (2021) that found that *third places* – that is, places that humans use, that are neither home (*first places*) or work (*second places*) places and are specific in that they are used for social interactions – are important predictors of urban vibrancy levels across age groups. Here, we study gender segregation and urban inequality through the lens of urban vibrancy. We explore the link between urban features and urban vibrancy and whether this differs for different genders, resulting in spatial segregation. We use a large data set that contains the presence of males and females in urban spaces, as measured via mobile phone activity data, as well as *OpenStreetMap* geographical data, and residential census data. We use mobile phone activity data as a proxy measurement for urban vibrancy and analyse which urban features contribute to urban vibrancy for different social groups, particularly males and females. We find that there are differences between males and females in terms of urban vibrancy. Indeed, the differences relate to 'Points of Interest' and transportation networks; however, there are both positive and negative spatial 'spillovers' that exist across each city. We discuss how these differences could be accounted for in urban planning and design, and how human interaction with large technological systems provides a wealth of data that can complement that derived from more traditional methods of monitoring populations. This could allow social problems, such as spatial segregation, to be measured more accurately, and at faster rates, so that social problems might be solved more easily by policymakers and urban planners for the cities of the future. We aim to further the understanding of gender differences (Vaitla et al. 2017) and segregation in urban spaces.

Data

We use three main data sources: (1) Italian census data that contains detailed information on where people live, (2) Call Detail Records data (CDR), derived from mobile phones, containing information on where people are at a high temporal granularity, and (3) *OpenStreetMap* (OSM 2017). We use OSM data because it provides measurable features of urban environments. For each data type, we gather data for only the 'metropolitan areas' of each city because these areas are the most densely populated areas of cities (see Supplementary Materials for summary statistics Table SI 1). Metropolitan cities are areas that are linked to the city in terms of its culture and economy as well as its geographic proximity. Data for the metropolitan area boundary were gathered from the Italian Office for National Statistics ('ISTAT') (ISTAT 2023). The mobile phone data can be made available on request. We outline the data sources below.

Italian census data

The *ISTAT* census (ISTAT 2023) is conducted every decade in Italy. Census sections are small, typically with 250 households, and provide total population and gender counts per section. We use 2011 census data downloaded from the ISTAT to confirm resident locations during key times of the day.

Call detail record data

We use mobile phone Call Detail Record (CDR) data from *Gruppo TIM* (formerly *Telecom Italia*) that was made available as part of their '*Big Data Challenge 2015*' (described in GruppoTIM (2015)). The CDR data is available for seven different cities: *Milan*, *Rome*, *Turin*, *Naples*, *Venice*, *Palermo*, and *Bari*, covering approximately 2 months (from 23:00 GMT on 2015-02-28 to 21:45 GMT on 2015-04-30). Each CDR data set has a corresponding grid designed by *Telecom Italia*

(GruppoTIM 2015). Each grid takes into account the topology of each city and the potential communication load. Each spatial grid was also designed to maintain the privacy of the inhabitants of each city. Subsequently, the grid polygons change shape in relation to the underlying mobile cells: cells typically get smaller the closer they are to the centre of each city. The activity of the data has a granularity of fifteen-minute intervals; however, there are time points in the CDR data that contain no records; this may occur because the number of users drops below three, and no data is recorded to preserve privacy, but there may also be further issues such as cutouts or problems in the collection of the data. The data contain the gender of the user who generated the CDR within the network (see CDR summary statistics Table SI 2). The data contain a value for each gender and the gender is derived from the registration of the SIM card of male and female mobile phone users. Therefore, the data shows male and female use across time.

A related data set has already been used to understand how different age groups interact with cities (Botta and Gutiérrez-Roig 2021). This is possible because this type of data allows us to analyse and investigate the existence of differences across social groups instead of aggregating across a population, where information concerning the differences between social groups is neglected or reduced.

We utilize the mobile phone CDR data set as a vector-based proxy for measuring urban vibrancy in the cities under investigation because prior research findings demonstrate consistent effectiveness in studying urban vibrancy (Sulis et al. 2018; Wang et al. 2021; Botta and Gutiérrez-Roig 2021). Though graph-based metrics are also used to measure urban vibrancy (Wang et al. 2018), we approximate urban vibrancy in a vector-based way because the method allows fine temporal and spatial granularity and because we can separate the data according to gender. We intend to use these data to understand differences between social groups to understand how urban features contribute to a vibrant environment with respect to gender. We have aggregated the cell-user data across an array of time periods such that we get metrics pertaining to urban vibrancy in each grid cell.

OpenStreetMap *data*

To understand cities, we must first create representations of their characteristics. Creating such representations has only recently been made easier thanks to collaborative projects such as *OpenStreetMap* (OSM 2017). OSM is an open-source data repository generated and collected by volunteer collaborators. The data that is formed consists of large-scale geographic data that is made freely available to users. It is possible to download data on an array of city attributes including the networks, systems, and features of urban landscapes. Here, we retrieve data for each study area; however, it is important to note that the data that was downloaded was the most up-to-date version of the urban features. What we derive from these data is explained systematically in the next section.

Methods

A proxy for urban vibrancy: Call detail records

As a proxy measure for urban vibrancy, we use CDR data as it indicates the presence of inhabitants throughout the day. We calculate gender differences by subtracting the male value from the female value:

$$\Delta_i = M_i - F_i \tag{1}$$

where Δ_i represents the vector of differences, M_i is the vector of male users, and F_i is the female users.

Gender identity was only available for users who disclosed it when acquiring their SIM cards. Users who did not disclose their gender were excluded from the analysis. As described above in the Call Detail Records data Section, gaps in the data occur when the number of users falls below three to protect their anonymity. To fill in these gaps, we assumed a value of zero for these time points. The grid consists of cells of varying sizes, so we normalized the raw data by the area of each grid cell to obtain a population density that accounts for the usage and topology of each cell in relation to the city.

Cities as networks: Independent variables

We downloaded for each static grid cell of each city a range of features shown to be related to urban vibrancy by previous research (Sung et al. 2013; Botta and Gutiérrez-Roig 2021; Yu et al. 2022; Chen et al. 2022). We processed these features in two ways as outlined below.

Density *in urban features.* By Jacobs (Jacobs 1961), increased feature density was a promoting factor in urban vibrancy because of the increased activity. The density of buildings, highways, networks, intersections, or 'Points of Interest' in a place, all have the potential to provide more opportunities for activities because of the increased number of users of those locations, or the increased vehicular or pedestrian access; however, importantly, these might differ across genders. Here, we define density as the concentration of a feature type within a given area. To arrive at the density value, we first download features from the free geographical database: *OpenStreetMap* (OSM; see Data Section). We construct feature collections of the buildings, transport networks, and 'Points of Interest' found in each cell of each city. We took the total count per cell and divided it by the total area of that cell to give a value of the feature density. For the networks, we calculated the average length of transport networks or the average number of intersections that are accessible for (1) pedestrians, (2) cyclists, and (3) drivers. We used the following calculation to determine feature density:

$$\rho = \frac{N}{A} \tag{2}$$

where ρ is the density, N is the total number of features in the geometry, and A is the area of the geometry. The values were added to the grid cells for each city (see Figure SI 2).

Diversity *in urban features.* According to Jacobs, it is the diversity in features that increases and encourages a location's vibrancy (Jacobs 1961). Similarly to above, in a place, the diversity of buildings, highways, or 'Points of Interest' all have the potential to provide more opportunities for activities because of the increased usage of those locations; however, these might differ across genders. To gather data on urban feature diversity we use the same downloaded OSM feature collections. We use the Shannon-Wiener diversity index (Shannon 1948) to calculate the diversity of features. The diversity index is calculated as follows:

$$H = -\sum_{i=1}^{M} P_i log_2 P_i \tag{3}$$

where H is the diversity index, M is the total number of categories in the geometry feature, and P_i is the frequency of the i^{th} category. The values were added to the grid cells for each city (see Figure SI 2). We detail the 'Points of Interest' variables below.

Points of interest. 'Points of Interest' are important features that directly relate to urban vibrancy. We collected all 'Points of Interest' from OSM found under the amenity, building, leisure, shop, and sport tags in the OSM database to construct a 'Points of Interest' collection for each city. We manually labelled the points using the same label collection as Moro et al. (2021) and Fan et al. (2022), based on the Foursquare classification system, which includes 14 categories. This taxonomy of labels is as follows: (1) *Arts/Museum*, (2) *City/Outdoors*, (3) *Coffee/Tea*, (4) *College*, (5) *Entertainment*, (6) *Food*, (7) *Grocery*, (8) *Health*, (9) *Residential*, (10) *Service*, (11) *Shopping*, (12) *Sports*, (13) *Transportation*, and (14) *Work*. We considered these labels because they represent the most frequently visited locations and are likely to be important for segregation (Moro et al. 2021).

Third places. We constructed a collection of '*Third Places*' (Oldenburg and Brissett 1982). Third Places are locations that are neither home nor workplaces and are considered to be vitally important in terms of urban vibrancy because they allow for impromptu everyday gatherings in urban locations that result in positive effects for communities (Jeffres et al. 2009; Botta and Gutiérrez-Roig 2021) and because people spend a significant fraction of their free time in third places. For this analysis, we considered that there may be differences between genders in how they use amenities like shops, pubs, cafés, or community centres. And that this was likely to be used differently depending on general discrepancies in urban mobility diversities and, amongst others, related to socioeconomic characteristics.

Here, using the same 'Points of Interest' tags used in OSM (i.e. 'amenity', 'building', 'leisure', 'shop', and 'sport'), we calculated density and diversity for third places (as defined in the two subsections above) across all grid cells and cities. We manually labelled third places based on Jeffres et al. (2009)'s categorization of (1) *eating and drinking*, (2) *organized activities*, (3) *outdoor*, and (4) *commercial venues*, and added a fifth label of *commercial services*. Commercial services are based on the locations where people might receive a service or go with the intent of buying something, but where you may also have opportunities for social interactions that might be brief compared to the other groupings. We included this label to capture the potential for brief social interactions in locations like banks and pharmacies that would have otherwise been removed due to Jeffres et al. (2009) definition. Only those locations that fit within these categories were considered third places.

Statistical approach

In this analysis, the Call Detail Record (CDR) data and census data have different spatial grids. We use the geometry of CDR data as our main reference and extract OSM data for each cell. We interpolate census data to the same spatial grid using areal interpolation (Comber and Zeng 2019; Bergroth et al. 2022) to enable correlation analysis and spatial linear regression at the grid-cell level (for an overview of the methodology, see SI 1).

We perform a correlation analysis to test the data's representativeness by comparing the CDR data (we used the nighttime values only) with census counts that have been converted to density estimates, matching the CDR data in terms of spatial scale and intensive property. We compare both male and female nighttime CDR values with their respective interpolated census data and use Kendall's correlation coefficient as it is distribution-free and more suitable for spatial data (Hamed 2011). We carry out this analysis in all cities.

We aim to model male-female differences for each city while also building an aggregated model to identify common trends. We refer to this aggregated model in many of the analyses below for clarity. To ensure comparability, we standardize all variables. We then construct an ordinary least squares (OLS) model as a baseline method for estimating the regression β coefficients and evaluating the importance of spatial extensions to OLS. We use the following linear function to explain male-female differences as a function of a set of separate features denoted by X (See Sections above):

$$Y = \beta_i X_i + \epsilon \tag{4}$$

where Y represents the male-female differences as a response variable consisting of a value proportional to the users per gender used here as a proxy for activity and a measure of vibrancy (see Sections above), β_i are the regression coefficients, X_i is the independent variables, and ϵ is the error.

Our data may have spatial autocorrelation, which violates some assumptions of a basic regression model. We address this issue in the following sections.

We diagnosed spatial dependence by analysing OLS model residuals. The error terms may not be independent due to a spatial relationship, so we checked for spatial structure and the need for spatial models using Moran's I analyses from Ward and Gleditsch (2019). Spatial clustering implies spatial dependence, and so requires spatial models.

We use maximum likelihood estimation to create spatial lag and spatial error models (Anselin et al. 2006; Ward and Gleditsch 2019). The Lagrange multiplier and AIC difference from OLS were calculated. We utilized *Queen's contiguity* based on grid-cell geometry for spatial weights' matrix, connecting centroids of observations to those with shared vertices (Rey and Anselin 2010). We used row transformation to normalize the weights' values to achieve an average of variable values in each observation's neighbourhood. We chose Queen's contiguity due to the irregular grid size. To assess the sensitivity of the choice of spatial weights, we replicate the same analysis using both Rook's contiguity and k-nearest neighbours' distance as weighting schemes, and we find qualitatively similar results (see SI 3).

We consider the spatial error model (SEM) as our first model, which incorporates spatial dependence using a spatially lagged error term:

$$Y = X_i \beta_i + u, u = \rho W u + \epsilon \tag{5}$$

where Y is the male-female differences as a response variable, X_i represents the explanatory variables, β_i are the regression coefficients, u is the first error term, ρ is a scalar of the spatial lag parameter, W is the weights' matrix (Queen's contiguity), and ϵ is the spatially independent error term.

We also utilized spatial lag models (SAR) where dependent variables were spatially lagged, providing coefficients for both *direct* and *indirect* effects of independent variables on the response and mean activity of neighbouring grid cells (Lesage and Fischer 2008). The SAR model terms are as follows:

$$Y = X_i \beta_i + \rho W Y u + \epsilon \tag{6}$$

where Y is the vector of the response variable, X represents the explanatory variables, β is the regression coefficients, ρ is a scalar of the spatial lag parameter, WY is the weights' matrix (Queen's contiguity), and ϵ is the spatially independent error term.

For each model, we calculated the direct effect, indirect effects, and total effects due to the challenges associated with interpreting predictors unit change under 'spillover' effects (Anselin and Rey 2014). Spatial spillover effects are indirect effects and refer to secondary impacts that result from the direct effects. To compute these values, we first derived estimated coefficients for exogenous variables in the model, which yielded the direct effect that shows the influence of one spatial unit on another (Lesage and Fischer 2008). Direct effects are calculated as

$$DE = \beta \tag{7}$$

where DE are the direct effects, β are the coefficients of the SAR model without the spatial lag term.

Contained within the direct effects are both the indirect effects and the total effects (GuoMeng et al. 2021). To split these apart, we extract the spatial lag term from the coefficients and divide the coefficients by 1 minus the spatial lag term multiplied by the largest eigenvalue of the spatial weights' matrix which effectively normalizes the coefficients (Lesage and Fischer 2008; Bivand and Piras 2015). Total effects take into account the full range of impacts that a particular change may have on the spatial system as a whole (Lesage and Fischer 2008). Total effects are calculated as

$$TE = \beta/(1 - \rho\lambda) \tag{8}$$

where TE are the total effects, β are the coefficients of the exogenous variable in the spatial lag model, ρ is the spatial lag term, and λ is the maximum eigenvalue from the spatial weights matrix.

Finally, we calculate the indirect effects. These refer to the secondary impacts that result from the direct effects. Indirect effects are a measure of any spillover effects that occur beyond the immediate spatial units (Lesage and Fischer 2008). Indirect effects are calculated as

$$IE = DE - TE \tag{9}$$

where IE are the indirect effects, DE are the direct effect and TE are the total effects.

Using the above methods and data, we created a hierarchy of models that aggregated CDR data at different time periods. These included (i) All daytime data (08:00–20:00), (ii) Weekdays (Monday–Thursday) versus Weekends (Friday–Sunday) within the all-day time period, and (iii) Twenty-four-hour day data divided into Morning (06:00–12:00), Afternoon (12:00–18:00), Evening (18:00–00:00), and Night (00:00–06:00) categories. Models were run for individual cities and for all cities combined. This approach enabled us to study how vibrancy relates to gender and urban features at different times and investigate variations over time.

Results

First, we use Kendall's rank correlation coefficient (Kendall 1938) to compare the census and the Call Detail Record (CDR) data to check representativeness. Across all cities, we find results were positive and significant at the 5% level (see Figure SI 5). The lowest τ for females is 0.55 (Bari), whereas the highest τ for females is 0.72 (Torino). For males, the lowest τ is 0.56 (Bari), whereas the highest τ for males is 0.7 (Torino). These strong correlations between the CDR – when selecting only the nighttime values – and census data show that the CDR data are broadly representative of the census data. This suggests that the values of the daytime should also be representative of the movements of the general population.

Next, we use Kendall's rank correlation coefficient to compare each urban feature with the male-female differences. We adjust our results using false discovery-rated detection to correct for the rate of type I errors in the null hypothesis. We found that, except for a few instances, all of our results were positive and were highly significant at the 5% level ($p < 0.05$). Also, between density and diversity metrics, diversity often had a smaller association contrasting with previous work on age groups (Botta and Gutiérrez-Roig 2021), we found that smaller cities had larger associations with male-female differences. See Table 1 for the full correlation analysis results.

We use Moran's spatial autocorrelation analysis to determine the global Moran's I of the data and to further understand the data in terms of its spatial dependence. We extract residuals of an ordinary least squares (OLS) model at each level of the model hierarchy. We calculate Moran's I statistics for the male-female differences in each city. For clarity, we report only the minimum and maximum in the lowest level of the hierarchy (this does not include the aggregated model which is excluded from this analysis due to its obvious spatial clustering in different cities), that is, the daytime data (08:00–20:00; see Statistical Approach Section above). The lowest value for Moran's I was 0.013, the

Table 1. Correlation between male-female differences and all urban features. For each city and the aggregation of all cities, we calculate the correlation coefficient and associated p-value. We have separated the variables into categories of density and diversity. We use Kendall's tau (τ) rank correlation coefficient because of its particular suitability for spatial data (Hamed 2011). We can see both the associational strength and directionality of each correlation. We also calculated the average for each variable based on each city but not the aggregation of all cities. p-values are shown using their significance stars, where [0–0.001] is '***', [0.001–0.01] is '**', [0.01–0.05] is '*', [0.05–0.1] is '.', and [0.1–1.0] has no symbol. Cities are ordered by the number of grid cells.

	Aggregated		Milano		Roma		Torino		Napoli		Venezia		Palermo		Bari		
Variable	τ	p	τ	p	τ	p	τ	p	τ	p	τ	p	τ	p	τ	p	Average
Density																	
Building	0.41	***	0.25	***	0.49	***	0.42	***	0.40	***	0.60	***	0.64	***	0.62	***	0.53
Road network length	0.37	***	0.19	***	0.46	***	0.48	***	0.41	***	0.51	***	0.58	***	0.51	***	0.49
Cycling network length	0.36	***	0.21	***	0.52	***	0.47	***	0.43	***	0.53	***	0.53	***	0.46	***	0.49
Walking network length	0.36	***	0.27	***	0.55	***	0.48	***	0.45	***	0.53	***	0.55	***	0.44	***	0.5
Highway	0.41	***	0.34	***	0.65	***	0.57	***	0.65	***	0.74	***	0.71	***	0.62	***	0.66
Road network node	0.40	***	0.20	***	0.51	***	0.52	***	0.51	***	0.54	***	0.62	***	0.55	***	0.54
Cycling network node	0.39	***	0.25	***	0.55	***	0.52	***	0.53	***	0.56	***	0.62	***	0.54	***	0.56
Walking network node	0.38	***	0.29	***	0.58	***	0.52	***	0.51	***	0.57	***	0.63	***	0.53	***	0.56
Point of interest	0.42	***	0.29	***	0.58	***	0.50	***	0.13	*	0.65	***	0.61	***	0.66	***	0.52
Third place	0.36	***	0.28	***	0.55	***	0.51	***	0.45	***	0.65	***	0.61	***	0.49	***	0.54
Diversity																	
Building	0.08	***	0.14	***	0.29	***	0.32	***	−0.04		−0.23	**	−0.16		0.42	***	0.1
Highway	0.00		−0.10	**	0.19	***	0.02		0.01		0.04		−0.02		0.35	***	0.1
Point of interest	0.10	***	0.11	**	0.32	***	0.03		0.44	***	0.44	***	0.13		0.10		0.24
Third place	0.06	**	−0.04		0.27	***	−0.04		0.06		0.22	**	0.12		0.20	*	0.14
Commercial venue (third place)	0.30	***	0.27	***	0.42	***	0.46	***	0.27	***	0.54	***	0.48	***	0.34	***	0.42
Organized activity (third place)	0.20	***	0.10	**	0.40	***	0.15	***	0.20	**	0.46	***	0.22	*	0.46	***	0.31
Outdoor (third place)	0.26	***	0.17	***	0.46	***	0.38	***	0.29	***	0.57	***	0.32	***	0.31	***	0.39
Eating drinking (third place)	0.31	***	0.26	***	0.50	***	0.41	***	0.22	***	0.53	***	0.36	***	0.33	***	0.39
Commercial service (third place)	0.27	***	0.24	***	0.45	***	0.40	***	0.24	***	0.46	***	0.42	***	0.37	***	0.39

maximum was 0.8, and the mean was 0.2. Of the seven cities in the analysis, two were non-significant at the 5% level, these were Bari and Napoli; all the rest were significant ($p < 0.05$). These values confirm the presence of spatial clustering and spatial dependence; however, these values alone cannot provide a full account of the presence or absence of spatial clustering or fully understand the spatial structure of the data; this, however, provides evidence that spatial models may be more appropriate. See Supplementary Materials Figure SI 4 and Table SI 3 for full Moran's *I* analysis results.

To further test the presence of spatial clustering, we calculate the Lagrange multiplier test statistic (both non-robust and robust) for the OLS model of each city and at each level of the analysis. We did this to identify the type of model most suitable for the analysis. This provides a measure to inform us whether to use a simple linear model or use either the spatial lag models (SAR) or the spatial error models (SER) (LeSage and Pace 2009). At the same time, we calculate the Akaike information criterion (AIC) of each of the models. We find that the SAR models most often contain the highest values (Minimum LMerr = 0.062, p is NS, Maximum LMerr = 410.104; Minimum LMlag = 9.56, $p < 0.001$, Maximum LMlag = 539.642). The same relationship is also found for the robust methods. The differences between the AIC of the OLS and the AIC of the spatial models were consistently greater for the SAR models. Because these tests consistently pointed toward using the SAR models, and we observe spatial clustering, we continue the rest of our analysis with SAR models.

We find fairly consistent results across this hierarchy. Firstly, we find that smaller cities had larger amounts of error in the estimates than larger cities and with more variation in the coefficients, most likely due to the size and number of cells in the smaller cities. Secondly, there are no strong gender differences between the night and day (see Figure 1, SI 6 and SI 7). In the larger cities, we find a significant positive indirect effect between male-female differences and third place density (see Figure 1); this relationship is consistent across the three largest cities and the aggregated model. We also found a significant negative indirect effect between male-female differences and the density of 'Points of Interest' (see Figure 1). The pattern was again similar across, this time, the four largest cities and the aggregated model. We also find a relationship between highway density and all three of the intersection variables; however, they did not share the direction. Both cycling and walking intersection density were significant with positive indirect effects; however, for road intersections, though there was a negative indirect effect, this was not significant. The highway density was only significant in the aggregate model (see Figure 1). We found that diversity metrics generally were not significant; however, we found that the 'organised activity' and 'outdoor' third place categories were significant; this was not the case for third places generally (see Figures SI 6 and SI 7).

We find that the strongest effect, which was also the most consistent, was the positive indirect effect of the third place density. We consider this an interesting finding when coupled with the significant negative indirect effect found in the 'Points of Interest' variable because the directionality of the 'Points of Interest' variable – without accounting for the social aspect of the third places – are opposing one another. This could suggest that locations that fall under our five categories for third places are not equally used by each gender, whereas, the 'Points of Interest' as a whole are used more equally.

For each level of the nested hierarchy, we report the results of the pseudo-r-squared values from the models. The pseudo-r-squared is the squared correlation between the dependent variable and the predictions of the dependent variable (Anselin 1988). These values are a measure of goodness-of-fit and are used to understand the relationship between the model variables. Our models exhibited relatively high values consistently across cities and across the hierarchy of models highlighting a good relationship between independent and dependent variables (see Figure 1, SI 6, and SI 7).

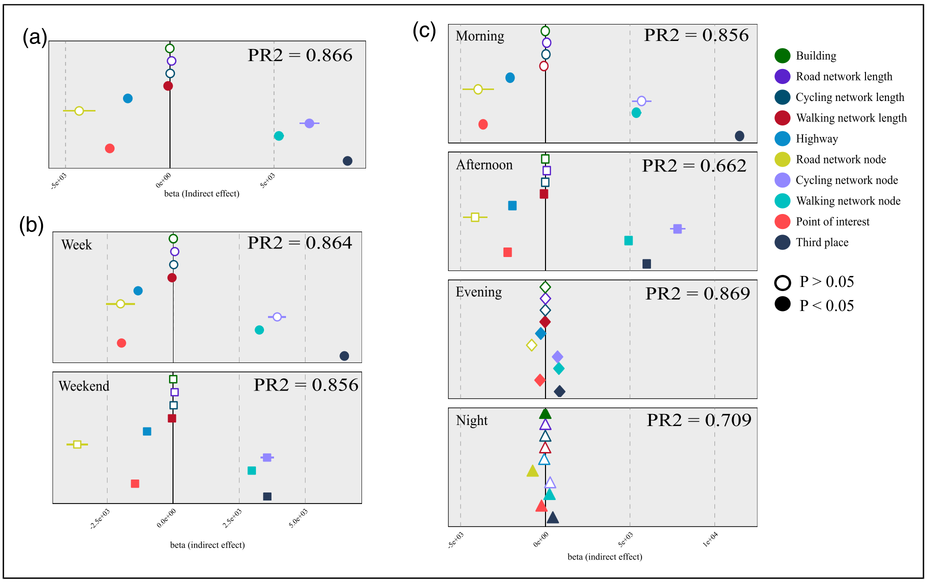

Figure 1. The relationship between density in features and male-female differences for the spatial model aggregating all cities. The plot shows the significance of the direct effect; however, the β coefficients represent the indirect effect (Methods Section for definitions). The plot displays all density variables, with the y-axis showing the variables for each model and the colour representing each variable. Panel (A) shows all daytime data between 08:00 and 20:00; Panel (B) displays weekday (Monday–Thursday) versus weekend (Friday–Sunday) data within the same time period; Panel (C) shows data averaged into four categories: Morning (06:00–12:00), afternoon (12:00–18:00), Evening (18:00–00:00), and Night (00:00–06:00). To aid comparison, shapes denote categories in Panels (b) and (c). Significance is indicated by closed and open shapes ($p < 0.05$ and $p > 0.05$, respectively), and each shape shows the error bars as a horizontal line. The pseudo-r-squared is reported for each panel.

Discussion

In this study, we have focused on modelling *urban vibrancy* – a measure of the dynamic activity of human beings in urban environments. For this, we have considered seven of the largest cities in Italy. We asked how urban features might contribute to a vibrant environment and how they might vary across social groups, especially concerning gender. We hypothesized that there would be differences for different genders because, firstly, heterogeneity exists generally in how people interact with urban environments (De Palma and Papageorgiou 1988) but, secondly, that similarities might correlate most closely with groups such as gender due to similarity in socioeconomic characteristics or general behaviours. We used a computational approach to reveal any potential socio-spatial segregation across our study areas, and we used a range of relevant urban features taken from urban vibrancy theory (Sung et al. 2013; Botta and Gutiérrez-Roig 2021; Yu et al. 2022; Chen et al. 2022). To model urban vibrancy, we used data showing the presence of mobile phone users as a proxy – another established methodology (Jia et al. 2019; Botta and Gutiérrez-Roig 2021).

We have uncovered a variety of findings. First, we have been able to study urban vibrancy – and potential segregation in urban vibrancy – by using high-frequency Call Detail Record (CDR) data and open-source geographical data. We show that it is possible to do this with high predictive power and goodness-of-fit, and we do this across a model hierarchy that accounts for movement behaviours in order to reflect the reality of life in cities. Second, we have furthered discussions from previous works that focused on the importance of third places in cities: we found significant evidence that an increase in the density of third places in a given area is associated with larger male-female differences in urban vibrancy, whereas an increase in the density of 'Points of Interest' overall (i.e. without specifically considering third places) is associated with smaller male-female differences. The evidence that third places are associated with larger differences could suggest that locations that we have defined as third places, that is, locations that fall under our five-category system (see Methods Section), are not equally used by each gender. This evidence does not necessarily mean that increases in the density of third places increase differences; however, it may be that certain types of third places are unequally used across genders. Reasons for this could be based on cultural differences or socioeconomic factors. Another part to consider is the clustered nature of third places in cities: a positive indirect effect indicates increases in male-female differences but also that the variable is positively correlated with the neighbouring values of male-female differences. It is important to consider that third places are places that are likely to cluster geographically with other factors such as economic activity or environmental conditions. More evidence would be needed to uncover further details but a similar methodology could be used with extensions and additional analyses. One such methodological adaptation could be the use of a Geographically Weighted Regression. This would help to understand the predictive power across cities whilst exploring potential spatial biases in the data.

Within our analysis, we can identify a number of limitations, and it is important to acknowledge these and discuss them here. Firstly, our CDR data is used as a proxy for urban vibrancy measurement; these data are from Telecom Italia, just one provider. Though this is the largest provider in Italy, these data do not capture the entire population and so may contain unknown biases. A full account of the general population could be gained by using multiple providers and may improve our overall analysis. It is also the case that the data were only derived from phone calls; this clearly misses a breadth of other communication methods and could potentially hide biases in the data due to the myriad of different ways people communicate today. Furthermore, gender information is derived from SIM purchases, but this is likely to not be an exact representation of the gender of users. However, we also note that our validation with the census data shows a good correlation with the mobile phone data, suggesting that these issues may be relatively limited (see Figure SI 5). A second potential problem is that the data are from differing time periods: we have taken census data from 2011, CDR data from 2015, and *OpenStreetMap* data from 2022. This undoubtedly introduces some biases in the analysis; however, we expect them to be small and not affect the overall results.

In this study, we have considered the modelling of *urban vibrancy* with respect to gender differences. We found that the density of different collections of 'Points of Interest' are simultaneously associated with both decreases and increases in male-female differences. This was the case when we gave a broad-scale use-category ('Points of Interest') and a fine-scale use-category that considers the social context of a place (third place). This adds further evidence for the importance of characterizing third places when studying urban environments and urban vibrancy. This evidence also suggests that comparing different collections of 'Points of Interest' could hold interesting avenues for further research relating to urban vibrancy. We have shown that this could also provide details on the potential segregation we find existing in cities today.

To conclude, our analysis provides further evidence and support for the use of CDR and crowdsourced data to understand large-scale movement behaviours and how we can use these data to understand the social fabric of urban life. In turn, this could provide evidence for the design of our future urban environments.

Acknowledgements

The authors wish to express deep appreciation toward Mario Gutiérrez-Roig and extend our heartfelt condolences to his family. Mario played a pivotal role in shaping the direction of this work. While he is no longer with us, his influence continues to guide our academic pursuits and serves as a lasting reminder of his commitment to our growth as researchers. This work is a tribute to Mario's legacy, and he will always be remembered.

R.D.C. acknowledges Sony CSL Laboratories in Paris for hosting him during part of the research.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iDs

Thomas Collins https://orcid.org/0000-0003-1791-7002
Riccardo Di Clemente https://orcid.org/0000-0001-8005-6351
Federico Botta https://orcid.org/0000-0002-5681-4535

Data availability statement

The mobile phone data used in this manuscript was made openly available to researchers as part of Telecom Italia Big Data Challenge 2015. The full data set is no longer available via Telecom Italia, but aggregate data to reproduce our results can be made available from the authors upon reasonable request. OpenStreetMap data is openly available at www.openstreetmap.org. The Italian census data used here is available at www.istat.it/it/archivio/222527.

Supplemental Material

Supplemental material for this article is available online.

References

Anselin L (1988) *Spatial Econometrics: Methods and Models, Studies in Operational Regional Science*. Dordrecht: Springer Netherlands. DOI: 10.1007/978-94-015-7799-1.

Anselin L and Rey SJ (2014) *Modern Spatial Econometrics in Practice: A Guide to GeoDa, GeoDaSpace and PySAL*. Chicago, IL: GeoDa Press LLC.

Anselin L, Syabri I and Kho Y (2006) ***GeoDa***: an introduction to spatial data analysis. *Geographical Analysis* 38(1): 5–22. DOI: 10.1111/j.0016-7363.2005.00671.x.

Bannister A and Botta F (2021) Rapid indicators of deprivation using grocery shopping data. *Royal Society Open Science* 8(12): 211069. DOI: 10.1098/rsos.211069.

Barthelemy M (2016) *The Structure and Dynamics of Cities: Urban Data Analysis and Theoretical Modeling*. Cambridge, MA: Cambridge University Press. DOI: 10.1017/9781316271377.

Bassolas A, Barbosa-Filho H, Dickinson B, et al. (2019) Hierarchical organization of urban mobility and its connection with city livability. *Nature Communications* 10(1): 4817. DOI: 10.1038/s41467-019-12809-y.

Batty M (2013) *The New Science of Cities*. Cambridge, MA: The MIT Press

Bergroth C, Järv O, Tenkanen H, et al. (2022) A 24-hour population distribution dataset based on mobile phone data from Helsinki Metropolitan Area, Finland. *Scientific Data* 9(1): 39. DOI: 10.1038/s41597-021-01113-4.

Bettencourt LMA, Lobo J, Helbing D, et al. (2007) Growth, innovation, scaling, and the pace of life in cities. *Proceedings of the National Academy of Sciences of the United States of America* 104(17): 7301–7306. DOI: 10.1073/pnas.0610172104.

Bivand R and Piras G (2015) Comparing implementations of estimation methods for spatial econometrics. *Journal of Statistical Software* 63: 1–36. DOI: 10.18637/jss.v063.i18.

Bor J, Cohen GH and Galea S (2017) Population health in an era of rising income inequality: USA, 1980–2015. *Lancet (London, England)* 389(10077): 1475–1490. DOI: 10.1016/S0140-6736(17)30571-8.

Botta F and Gutiérrez-Roig M (2021) Modelling urban vibrancy with mobile phone and OpenStreetMap data. *PLoS One* 16(6): e0252015. DOI: 10.1371/journal.pone.0252015.

Botta F, Moat HS and Preis T (2015) Quantifying crowd size with mobile phone and *Twitter* data. *Royal Society Open Science* 2(5): 150162. DOI: 10.1098/rsos.150162.

Botta F, Preis T and Moat HS (2020) In search of art: rapid estimates of gallery and museum visits using Google Trends. *EPJ Data Science* 9(1): 12–14. DOI: 10.1140/epjds/s13688-020-00232-z.

Chen L, Zhao L, Xiao Y, et al. (2022) Investigating the spatiotemporal pattern between the built environment and urban vibrancy using big data in Shenzhen, China. *Computers, Environment and Urban Systems* 95: 101827. DOI: 10.1016/j.compenvurbsys.2022.101827.

Comber A and Zeng W (2019) Spatial interpolation using areal features: a review of methods and opportunities using new forms of data with coded illustrations. *Geography Compass* 13(10): e12465. DOI: 10.1111/gec3.12465.

Connolly C, Keil R and Ali SH (2021) Extended urbanisation and the spatialities of infectious disease: demographic change, infrastructure and governance. *Urban Studies* 58(2): 245–263. DOI: 10.1177/0042098020910873.

De Palma A and Papageorgiou YY (1988) Heterogeneity in states and urban structure. *Regional Science and Urban Economics* 18(1): 37–56. DOI: 10.1016/0166-0462(88)90004-X.

Di Clemente R, Luengo-Oroz M, Travizano M, et al. (2018) Sequences of purchases in credit card data reveal lifestyles in urban populations. *Nature Communications* 9(1): 3330. DOI: 10.1038/s41467-018-05690-8.

El Araby M (2002) Urban growth and environmental degradation. *Cities* 19(6): 389–400. DOI: 10.1016/S0264-2751(02)00069-0.

Entwisle B (2007) Putting people into place. *Demography* 44(4): 687–703. DOI: 10.1353/dem.2007.0045.

Fan Z, Su T, Sun M, et al. (2022) *Diversity beyond Density: Experienced Social Mixing of Urban Streets*. Oxford: Oxford Academic

Glaeser EL and Sacerdote B (1999) Why is there more crime in cities? *Journal of Political Economy* 107(S6): S225–S258. DOI: 10.1086/250109.

González MC, Hidalgo CA and Barabási AL (2008) Understanding individual human mobility patterns. *Nature* 453(7196): 779–782. DOI: 10.1038/nature06958.

GruppoTIM (2015) *Dataset Source: TIM Big Data Challenge*. Rome, Italy: GruppoTIM. https://www.gruppotim.it/en.html

GuoMeng, cheng DZ and Shukai C (2021) An analysis of the "direct effect" and "indirect effect" of urban housing prices on the upgrading of industrial structure—based on data of 285 cities. *Journal of Physics: Conference Series* 1792(1): 012017. DOI: 10.1088/1742-6596/1792/1/012017.

Hamed KH (2011) The distribution of Kendall's tau for testing the significance of cross-correlation in persistent data. *Hydrological Sciences Journal* 56(5): 841–853. DOI: 10.1080/02626667.2011.586948.

Harris N (1990) Urbanisation, economic development and policy in developing countries. *Habitat International* 14(4): 3–42. DOI: 10.1016/0197-3975(90)90002-I.

ISTAT (2023) *Borders of Administrative Units for Statistical Purposes as of 1 January 2023*. Rome, Italy: ISTAT. Available at: https://www.istat.it/it/archivio/222527.

Jacobs J (1961) *The Death and Life of Great American Cities*. New York, NY: Random House.

Jeffres LW, Bracken CC, Jian G, et al. (2009) The impact of third places on community quality of life. *Applied Research in Quality of Life* 4(4): 333–345. DOI: 10.1007/s11482-009-9084-8.

Jia C, Du Y, Wang S, et al. (2019) Measuring the vibrancy of urban neighborhoods using mobile phone data with an improved PageRank algorithm. *Transactions in GIS* 23(2): 241–258. DOI: 10.1111/tgis.12515.

Kalila A, Awwad Z, Di Clemente R, et al. (2018) Big data fusion to estimate urban fuel consumption: a case study of riyadh. *Transportation Research Record: Journal of the Transportation Research Board* 2672(24): 49–59. DOI: 10.1177/0361198118798461.

Kendall MG (1938) A new measure of rank correlation. *Biometrika Trust*. Oxford, UK: Oxford University Press

King K (2013) Jane Jacobs and 'the need for aged buildings': neighbourhood historical development pace and community social relations. *Urban Studies* 50(12): 2407–2424. DOI: 10.1177/0042098013477698.

Lazer D, Pentland AS, Adamic L, et al. (2009) Life in the network: the coming age of computational social science. *Science (New York, N.Y.)* 323(5915): 721–723. DOI: 10.1126/science.1167742.

Lesage JP and Fischer MM (2008) Spatial growth regressions: model specification, estimation and interpretation. *Spatial Economic Analysis* 3(3): 275–304. DOI: 10.1080/17421770802353758.

LeSage J and Pace RK (2009) *Introduction to Spatial Econometrics*. New York: Chapman and Hall/CRC

Liu W, Wu W, Thakuriah P, et al. (2020) The geography of human activity and land use: a big data approach. *Cities* 97: 102523. DOI: 10.1016/j.cities.2019.102523.

Moro E, Calacci D, Dong X, et al. (2021) Mobility patterns are associated with experienced income segregation in large US cities. *Nature Communications* 12(1): 4633. DOI: 10.1038/s41467-021-24899-8.

Moroni S (2016) Urban density after Jane Jacobs: the crucial role of diversity and emergence. *City, Territory and Architecture* 3(1): 13. DOI: 10.1186/s40410-016-0041-1.

Nelson A, Pendall R, Dawkins CJ, et al. (2002) *The Link between Growth Management and Housing Affordability: The Academic Evidence*. Washington, DC: The Brookings.

Oldenburg R and Brissett D (1982) The third place. *Qualitative Sociology* 5(4): 265–284. DOI: 10.1007/BF00986754.

OSM (2017) *OpenStreetMap Contributors*. https://www.openstreetmap.org/.

Pan W, Ghoshal G, Krumme C, et al. (2013) Urban characteristics attributable to density-driven tie formation. *Nature Communications* 4(1): 1961. DOI: 10.1038/ncomms2961.

Perrone C (2019) 'Downtown Is for People': the street-level approach in Jane Jacobs' legacy and its resonance in the planning debate within the complexity theory of cities. *Cities* 91: 10–16. DOI: 10.1016/j.cities.2018.12.023.

Preis T, Botta F and Moat HS (2020) Sensing global tourism numbers with millions of publicly shared online photographs. *Environment and Planning A: Economy and Space* 52(3): 471–477. DOI: 10.1177/0308518X19872772.

Rey SJ and Anselin L (2010) PySAL: a Python library of spatial analytical methods. In: Fischer MM and Getis A (eds), *Handbook of Applied Spatial Analysis: Software Tools, Methods and Applications*. Berlin, Heidelberg: Springer, pp. 175–193

Salesses P, Schechtner K and Hidalgo CA (2013) The collaborative image of the city: mapping the inequality of urban perception. *PLoS One* 8(7): e68400. DOI: 10.1371/journal.pone.0068400.

Santana C, Botta F, Barbosa H, et al. (2023) COVID-19 is linked to changes in the time–space dimension of human mobility. *Nature Human Behaviour*. DOI: 10.1038/s41562-023-01660-3.

Seresinhe CI, Preis T and Moat HS (2016) Quantifying the link between art and property prices in urban neighbourhoods. *Royal Society Open Science* 3(4): 160146. DOI: 10.1098/rsos.160146.

Shannon CE (1948) A mathematical theory of communication. *Bell System Technical Journal* 27(3): 379–423. DOI: 10.1002/j.1538-7305.1948.tb01338.x.

Steenbruggen J, Tranos E and Nijkamp P (2015) Data from mobile phone operators: a tool for smarter cities? *Telecommunications Policy* 39(3): 335–346. DOI: 10.1016/j.telpol.2014.04.001.

Su T, Sun M, Fan Z, et al. (2022) Rhythm of the streets: a street classification framework based on street activity patterns. *EPJ Data Science* 11(1): 43. DOI: 10.1140/epjds/s13688-022-00355-5.

Sulis P, Manley E, Zhong C, et al. (2018) Using mobility data as proxy for measuring urban vitality. *Journal of Spatial Information Science*(16): 137–162.

Sung HG, Go DH and Choi CG (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. DOI: 10.1016/j.cities.2013.07.010.

Thomas S (2008) *Urbanisation as a Driver of Change Sustainable City 2008*. Greece: Skiathos, pp. 95–104.

Tóth G, Wachs J, Di Clemente R, et al. (2021) Inequality is rising where social network segregation interacts with urban topology. *Nature Communications* 12(1): 1143. DOI: 10.1038/s41467-021-21465-0.

UN DESA (2019) *World Urbanization Prospects: The 2018 Revision*. New York: UN DESA

Vaitla B, Bosco C, Alegana V, et al. (2017) *Big Data and the Well-Being of Women and Girls Applications on the Social Scientific Frontier*. Medford, MA: Tufts University.

Vespignani A (2009) Predicting the behavior of techno-social systems. *Science* 325(5939): 425–428. DOI: 10.1126/science.1171990.

Wang P, Fu Y, Zhang J, et al. (2018) Learning urban community structures: a collective embedding perspective with periodic spatial-temporal mobility graphs. *ACM Transactions on Intelligent Systems and Technology* 9(6): 1–63. DOI:10.1145/3209686

Wang P, Liu K, Wang D, et al. (2021) Measuring urban vibrancy of residential communities using big crowdsourced geotagged data. *Frontiers in Big Data* 4: 690970.

Ward MD and Gleditsch KS (2019) Spatial regression models. *Quantitative Applications in the Social Sciences*. 2nd edition. Thousand Oaks, CA: Sage Publications

Weller S and van Hulten A (2012) Gentrification and displacement: the effects of a housing crisis on Melbourne's low-income residents. *Urban Policy and Research* 30(1): 25–42. DOI: 10.1080/08111146.2011.635410.

Xu S, Di Clemente R and González MC (2019) Mining urban lifestyles: urban computing, human behavior and recommender systems. *Big Data Recommender Systems: Application Paradigms*. London, UK: Institution of Engineering and Technology, pp. 71–81

Xu Y, Xu Y, Xie XF, et al. (2021) Anti-platelet aggregation of Panax notoginseng triol saponins by regulating GP1BA for ischemic stroke therapy. *Chinese Medicine* 16(1): 12. DOI: 10.1140/epjds/s13688-021-00267-w.

Yu B, Cui X, Liu R, et al. (2022) Intergenerational differences in the urban vibrancy of TOD: impacts of the built environment on the activities of different age groups. *Frontiers in Public Health* 10: 994835.

Zheng Y, Capra L, Wolfson O, et al. (2014) Urban computing: concepts, methodologies, and applications. *ACM Transactions on Intelligent Systems and Technology* 5(3): 1–55. DOI: 10.1145/2629592.

Thomas Collins is a doctoral researcher at the University of Exeter in the Department of Computer Science. The aim of his research is to model urban environments using novel born-digital data sources to visualise human behaviour. His current research uses an interdisciplinary approach that incorporates computational methodologies to increase understanding of the relationship between urban vibrancy–the energetic activity in an urban environment–and gender differences.

Riccardo Di Clemente is an Associate Professor at the Network Science Institute at Northeastern University London. He has been Newton International Fellow of the Royal Society and served as an Alan Turing Fellow of The Alan Turing Institute. Riccardo's research primarily focuses on the analysis of the digital traces we leave behind to understand human urban behavior and social dynamics and interactions. He employs network science, complexity theory, and data science to develop data-driven approaches for spatial planning, poverty reduction, and disaster resilience in developing countries in collaboration with the World Bank, the Bill & Melinda Gates Foundation, and Data2x.

Mario Gutierrez-Roig is a Lecturer in Data Science at University of Essex. His current research, within the field of Computational Social Science, focuses on using Data Science and Machine

Learning methodologies to analyse and model urban systems and online human behaviour. Over the years he has developed a highly multidisciplinary profile with contributions on topics such as Complex Systems, Social Dilemmas, Financial Markets, Pedestrian Movement, Computational Archaeology and Urban Analytics.

Federico Botta is a Senior Lecturer in data science in the department of computer science at the University of Exeter, and also a fellow at the Alan Turing Insitute, the UK national institute for data science and AI. His research aims to provide a deeper understanding of human behaviour, both at the collective and individual level, and society, by using novel data streams. He uses tools from data science, network theory, behavioural and computational social sciences to analyse these data sets and investigate different aspects of human behaviour.