Article

Urban Analytics and City Science

Environment and Planning B: Urban Analytics and City Science
0(0) 1–16

DOI: 10.1177/2399808318755146
journals.sagepub.com/home/epb
SAGE

Defining urban clusters to detect agglomeration economies

Clémentine Cottineau
University College London, UK

Olivier Finance
Université Paris 1 Panthéon-Sorbonne, France; Université Catholique de Louvain, Belgium

Erez Hatna
Johns Hopkins University, USA

Elsa Arcaute and Michael Batty
University College London, UK

Abstract
Agglomeration economies are a persistent subject of debate in regional science and city planning. Their definition turns on whether or not larger cities are more efficient than smaller ones. Here, we complement existing discussions on agglomeration economies by providing a sensitivity analysis of estimated externalities to the definitions of urban agglomeration. We regress wages versus population and jobs over thousands of different definitions of cities in France, based on an algorithmic aggregation of spatial units. We also search for evidence of larger inequalities in larger cities. This paper therefore focuses on the spatial and economic complexity of the mechanisms defining agglomeration within and between cities.

Keywords
Agglomeration economies, cities, definition, inequality, scaling

Introduction

As complex systems, cities exhibit quantitative and qualitative changes in composition as they grow in size: economies of agglomeration are one of the most debated of such transformations. Empirical evidence suggests the existence of systematic variations in productivity levels across space, but the diversity of specifications used to estimate the magnitude of agglomeration economies leads to a wide array of quantitative variations: Rosenthal and Strange (2004) for example find that doubling city size tends to increase

Corresponding author:
Clémentine Cottineau, Centre for Advanced Spatial Analysis, University College London, London W1T 4TJ, UK.
Email: c.cottineau@ucl.ac.uk

individual productivity by 3% to 8%. In a meta-analysis of the literature, Melo et al. (2009) examined the parameters influencing this estimation in 34 studies. They found some country specific effects, some industrial coverage effects (services generate more agglomeration economies than manufacturing activities) and a publication bias towards reporting positive results more systematically compared to negative results. Controlling for differences in skills also tends to lower the estimation of urban size effects on productivity.

Cities are not simply agglomerations of people: they concentrate capital, infrastructure, information and many other factors of production. The literature usually distinguishes between *localisation economies*, the positive externalities which come from the concentration of firms in a particular industry (Marshall, 1920), and *urbanisation economies* (Jacobs, 1961), the positive externalities which are provided by all the other aspects of the city. It is generally assumed that the large city provides a set of urban amenities and intra-industry linkages which results in increased firms' productivity and workers' average earnings. However, the spatial boundaries within which agglomeration economies operate are usually taken for granted. This lack of theoretical formulation on the spatial side of agglomeration economies is problematic because the urban concept is a very fuzzy one (Bretagnolle et al., 2002; Parr, 2007) and the question of which aspect of urbanity generates the productivity premium is left to speculation.

In this paper, we tackle the relation between agglomeration economies and urban definition by asking three questions. (1) Are economies of agglomeration specifically urban or is this just about people congregating within any type of geographical boundaries? (2) Are larger cities richer regardless of the city definition chosen, or does the choice of definition affect the results to the point that it is only true for certain ways of delineating cities? (3) Are richer cities also more unequal? We take the French case as an example to build a comprehensive representation of where cities extend. This allows, among other things, to compare functional labour markets with densely built environments, 'night-time cities' with 'day-time cities', using residential and workplace geographies, respectively, thus acknowledging that the location of jobs and residents do not coincide. We present the theoretical mechanisms of agglomeration economies and agglomerated inequalities in the literature as well as our own hypotheses regarding the effect of city definition in the first section. The data and methods used are described in a subsequent section. Results and conclusions follow.

Cities, scales, wealth and inequality

The concepts of cities, wealth, inequality and size have been linked through causal mechanisms at various spatial scales in the literature. Starting with the most classical of these relationships – size versus productivity, what are the theoretical underpinnings of agglomeration economies, where do they come from and at which scale do they operate? Urban economics have produced a large body of theoretical work on the spatial heterogeneity of productivity and therefore the economic existence of cities themselves. Duranton and Puga (2004) summarise this literature into three types of micro-foundations of localised increasing returns: sharing, matching and learning. 'Micro-foundations of urban agglomeration economies based on sharing mechanisms deal with sharing indivisible facilities, sharing the gains from the wider variety of input suppliers that can be sustained by a larger final-goods industry, sharing the gains from the narrower specialisation that can be sustained with larger production, and sharing risks. In discussing micro-foundations based on matching, we study mechanisms by which agglomeration improves either the expected quality of matches or the probability of matching, and alleviates hold-up problems. Finally, when we look at micro-foundations based on learning we discuss mechanisms based on the generation, the diffusion, and the accumulation of knowledge' (Duranton and Puga, 2004: 2067). In this sense, we can relate some of the risk sharing, specialised skills matching and learning processes to localisation economies, where firms benefit from the local presence of other firms in the same industry to improve their individual productivity. On the other hand, infrastructure sharing, supply-chain matching and inter-branch learning participate in urbanisation economies, where firms benefit from the amenities and diversity of the city to reduce their costs and foster innovative production.

Each of the three micro-foundations involves a different set of actors and interactions. In the case of matching, firms and workers of the whole labour market participate in the process, as the more numerous they are, the more probable efficient matches are between supply and demand. In the case of learning, firms and workers are supposed to benefit from knowledge spillovers, although mainly between close-by places of production specialised in related industries where knowledge is accumulated and diffused through face-to-face interactions. In the case of sharing, the scale and scope of the mechanism depends on what is shared. If we look at individisible facilities, they can range from very local amenities (e.g. shared office spaces, fast broadband) to neighbourhood equipments (e.g. underground station, park) and regional facilities (e.g. airport, patent registration office). On the other hand, the sharing of risk, of a wide variety of inputs and of a narrow industrial specialisation seems to indicate that urban and regional economies are involved as a whole. These different networks of agents have differing policy implications as policy will have to adapt its target and geographical scale depending on the mechanism at work. Similarly, we should not be able to measure agglomeration economies appropriately at all scales and across all city definitions if one mechanism dominated the others.

Choosing one option (e.g. intra-urban districts) or the other (e.g. metropolitan areas) is expected to affect the measured outcome because the spatial distribution of jobs and amenities is uneven within cities and between cities. Jobs and amenities are more concentrated than the resident population (Glaeser et al., 2001) and the scaling of infrastructural and socio-economic attributes with city size varies with city definition (Arcaute et al., 2015; Fragkias et al., 2013; Rybski et al., 2016). Moreover, some evidence suggests that networks of cities can spread the effects of agglomeration economies, resulting in small cities 'borrowing size' from larger neighbouring ones (Meijers and Burger, 2017). Finally, the measurement of agglomeration economies is potentially affected by the Modifiable Areal Unit Problem (MAUP) (Openshaw, 1983), as are concentration and segregation measures (Reardon, 2006; Wong et al., 1999). The spatial scale and scope of agglomeration economies have been mainly absent from theoretical debates and policy discussions, probably because of the under-defined concepts of space and time in dominant theories of urban economics (Martin, 1999). Let us see how this problem in tackled in the empirical literature.

In the seminal publications on agglomeration economies (Henderson, 1986; Moomaw, 1988; Sveikauskas, 1975), the term 'city' was used without having been explicitly defined. It was supposed to be a consensual and homogenous object, most probably because empirical investigations were limited to the USA. In this country, SMSAs[1] provided an easy and well-accepted choice of city definition, having been delineated systematically by the Census Bureau as functional commuting areas since the 1940s. This definition is convenient because it covers the spatial extent of urban labour markets. However, this property is implicit in the papers and not discussed with respect to the type of agglomeration economies studied. Therefore, the absence of positive agglomeration economies could indicate the absence of all processes of agglomeration economies, or only the ones which operate at the labour market scale, whereas the effect of learning mechanisms would still be observable at more local scales. Furthermore, when these results were transferred to other countries in which functional definitions were not as readily available as in the USA, there has not been any explicit discussion of the potential spatial biases introduced by the change of scale. For example, Ciccone (2002) or Fingleton (2006) compare their results, respectively, on European regions NUTS2 and British Local Authorities with results computed on cities, without mentioning the difference in experiment designs nor the expected variations.

What would be an ideal way of measuring agglomeration spatially and what would happen when we deviate from it? Rosenthal and Strange (2001) and Mori and Smith (2015) found evidence that estimations of localisation economies varied with geographical scale and across industries. This suggests that there is no single ideal definition but that methodology needs to adapt to theoretical questions rather than available data dictate experiment designs. Additionally, empirical evidence suggests that agglomeration economies come with agglomeration inequality as higher levels of productivity can be attained by complementing high skilled labour with low skilled labour (Eeckhout et al., 2014; Royuela et al., 2014). The dynamics of these polarised (Sassen, 1991) and segregated (Bischoff and Reardon, 2013; Cheshire et al. 2014) urban societies of large sizes would be self-reinforcing as inequality further retroacts on social equity and intergenerational mobility (Chetty et al., 2014; Piketty, 2013; Roscigno et al., 2006; van Ham et al., 2016; Watson, 2006; Wilkinson and Pickett, 2006). Along with a spatial understanding of agglomeration economies, we miss a spatial understanding of the link between city size effects, inequality and segregation. Our hypothesis is that the heterogeneous spatial structure of cities – that is, the degree of their polycentricity, the gradient of their density-decay, their patchwork of cultures – makes the boundaries chosen to delineate cities a non-trivial aspect of the potential agglomeration economies generated and measured (H1). Indeed, the mechanisms through which the urban *premium* is generated are fundamentally embedded in space and amenities are unevenly distributed. Although intra-urban characteristics are at the origin of the variations measured with different delineations, this paper is not about them specifically. For example, we do not investigate the difference between polycentric and monocentric cities with respect to agglomeration economies, despite the interesting problem it represents. Here, we expect cities delineated around day-time activity rather than residential characteristics to exhibit the strongest levels of agglomeration economies which could be explained by learning and sharing mechanisms, whereas delineations corresponding to labour markets would exhibit the strongest levels of agglomeration economies which could be related to matching (H2). Furthermore, we expect agglomeration economies to come with increased inequality and increased segregation (H3). We test hypotheses H1, H2 and H3 empirically with data on French cities.

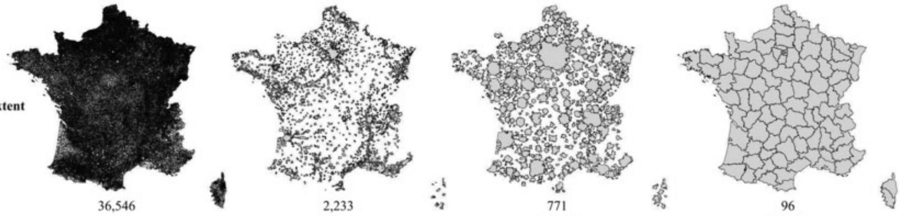
Data and methods

France was chosen because it provides an alternative example to that of the USA (dominant in the literature) and because cross-sectional economic data are available at a fine-grained level.

Population data are gathered for 36,546 local units from the 2011 French Population Census[2] and aggregated into higher levels of geography shown, for example in Figure 1.

Typically, there is empirical evidence of agglomeration economies when the added value per capita increases with city size. However, added value is seldom available at local levels of administrative geography, so wages are taken as a proxy for productivity levels

Unit Type	Local Units (COM)	City Cores (UU)	Metropolitan Areas (AU)	NUTS-3 regions (DEP)
Spatial Extent				



N	36,546	2,233	771	96
Mean Area (ha)	2,553	8,414	61,381	972,529
Min Area (ha)	1	4	18	77
Max Area (ha)	5,188	1,238,155	4,877,170	2,769,560
Mean Population	1,724	21,760	69,295	656,983
Min Population	0	605	2,169	77,156
Max Population	2,249,975	10,516,110	12,292,895	2,579,208
Orders of Magnitude	6	5	4	2
National coverage	Complete	Incomplete	Incomplete	Complete

Figure 1. Four delineations of territorial units in France. DEP (*Départements*) and COM (*Communes*) are administrative divisions. UU (*Unités Urbaines*) and AU (*Aires Urbaines*) are statistical urban delineations produced by INSEE.

(Combes et al., 2011). Wage data come from the public database CLAP[3] which provides information on firms. Data refer to total wages and the total number of employees at the firm's location in 2008, further aggregated to the local unit level of *communes* (COM). This location at the firm level rather than at the establishment level can overestimate the number of jobs in the largest cities (where the headquarters tend to concentrate), but it is more reliable in terms of average wages.

Absolute numbers of residents and jobs are used to represent the size of a geographical unit, and their density is measured per unit of urbanised surface (from CORINE Landcover 2006 raster data[4]). The aggregated wages Y of a spatial unit i is OLS-regressed against its total population or density using equations (1) and (2), which are the log-transforms of the scaling equations specified for population and density

$$log(Y_i) = a + \beta \times log(P_i) + \varepsilon_i \quad (1)$$

$$log(Y_i) = a + \gamma \times log(D_i) + \varepsilon_i \quad (2)$$

where Y_i represents the total wages of a spatial unit i, P_i is the urban population (measured in residents or firms), D_i is the urban density, a is a normalisation constant, β and γ are the scaling coefficients and ε_i the residuals.

We interpret β and γ relative to 1: an exponent equal to 1 represents the absence of economies of agglomeration (which is isometry or linear scaling in allometric terms), as the economic output grows proportionately with population or population density (linear regime). Exponents significantly greater than 1 (which is positive allometry or superlinear scaling) indicate economies of agglomeration, or rising average wages with size or density. A value significantly lower than 1 (which is negative allometry or sublinear scaling) suggests diseconomies of agglomeration.

Unlike other examples in the economics literature, we do not add instrumentation to this regression for three reasons. Firstly, we do not have access to individual data but only aggregates, which are therefore subject to ecological errors. Secondly, we want to keep the model as simple as possible since we use many combinations of economic and geographic specifications, complicating the interpretation. Finally, this is the only way to compare our results with other urban scaling studies. We are aware of the fact that Briant et al. (2010) have shown that scale and shape effects of geographical units played a role in the estimation of the French wage premium, although a less important one than the effect of specifications and controls used in the regressions. This exercise does not provide a complete account of the levels of productivity in cities, let alone a causal explanation for its spatial distribution. Instead, it focuses on city size as one strong predictor of increased productivity and inequality across city definitions, to be filtered out for further analysis.

Empirical results

Our results are organised so as to answer the three questions stated in the introduction.

Are economies of agglomeration specifically urban?

In this section, we test whether or not agglomeration effects are characteristic of urban spaces. Two types of geographical boundaries are considered: strictly urban delineations (i.e. excluding rural space from the national partition) and non-strictly-urban delineations (i.e. total coverage of the national space). In the strictly-urban category, *city cores* (UU, cf. Figure 1) are defined based on the continuity of the built-up area (<200 m between buildings), and *metropolitan areas* (AU) correspond to city cores of more than 1500 jobs to which are attached local units where more than 40% of the working residents commute to the city core (INSEE, 2010). In the second category, *local units* (COM) and *NUTS-3 regions* (DEP) provide an exhaustive coverage of the French territory, including rural areas, at respectively low and high spatial scales.

Figure 2 presents the results obtained for the estimated regressions using a combination of four geographical delineations (rows) and four economic specifications (columns). Horizontally, one evaluates the sensitivity to economic specifications, whereas vertically, the reader sees variations that depend only on the geographical delineation chosen (in terms of scale and 'urbanity').

About half of the estimations (7 out of 16) indicate a linear relation between economic output and population, even without the introduction of housing prices as instruments of the regression. Economies of agglomeration are found in eight cases, and diseconomies in one case: when wages are regressed with the density of population at the local scale of local units (*communes*).

Regarding the sensitivity of estimates to economic specifications, the referent population (residents P or jobs J) plays a decisive role in the estimation, as jobs are spatially more concentrated than residence, and less ubiquitous. There is some variation between regressions with the absolute population and regressions with the density. However, the coefficients of mass β and density γ are mathematically linked through the scaling of the urbanised area with population so their different values provide less information about possible economic mechanisms than about the way cities of different sizes sprawl.

Regarding the sensitivity of estimates to geographical specifications, the scale of analysis seems to be paramount, whereas mixed evidence is found as to the 'urbanity' of economies of agglomeration. With respect to wages and the number of jobs J (column 2), economies of agglomeration are non-urban: partitions of the territory at the local and the regional scales exhibit superlinear exponents (respectively, 1.06 and 1.09), whereas city cores and metropolitan areas produce exponents closer to 1 (respectively, 1.04 and 1.02). Counter-intuitively, larger regions (DEP) and larger units (COM) are richer, but larger cities (UU and

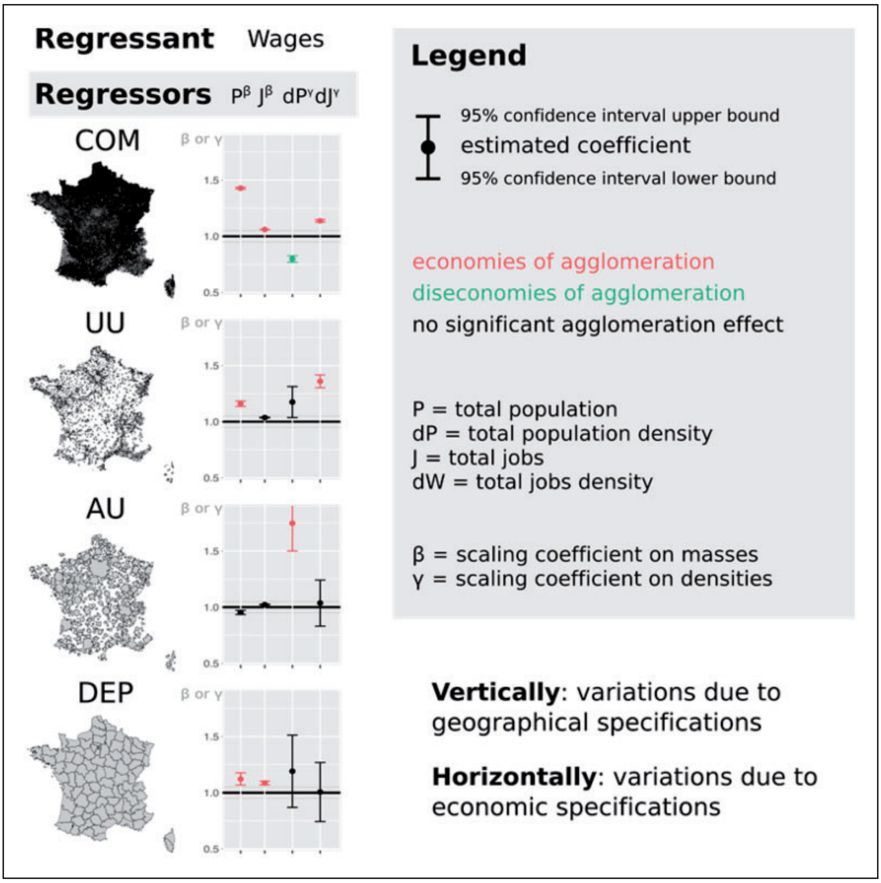


Figure 2. Variations of scaling exponents with geographic and economic specifications.
COM: Local Units. UU: Built-up areas. AU: Metropolitan areas. DEP: NUT-3 regions. N.B. We also computed exponents based on the log-likelihood method of Leitao et al. (2016) and compared them with a linear model. We find that in general, the results are very consistent. The lognormal estimation gives lower values of exponents (−2% on average), and discards the very high exponents (>1.5) obtained by regressing against density in metropolitan areas. This method being much more computationally demanding and equivalent with OLS in results, we kept the OLS simpler method to estimate further exponents.

AU) are not. This is probably due to the different composition of urban and rural in regions of different sizes than to a change in economic behaviour for spatial units of larger population. Estimates significantly over 1 are found at the local scales when considering the jobs density dJ (column 4), thus confirming the explanations involving local processes within the labour market (sharing and learning) rather than those of residential sorting, because agglomeration economies hold for high work densities in city centres and not for regional labour markets (i.e. including commuting zones into the definition of cities as in metropolitan areas AU).

To conclude, economies of agglomeration are urban insofar as 'urban' refers to the dense concentrations of jobs in central cities, in accordance with mainstream urban economics. In terms of absolute concentration of residents and workers (i.e. total population rather than density), results tend to identify regional partitions as best suited for the observation agglomeration economies. Therefore, the scale of analysis (local, urban, regional) is important in finding agglomeration effects on wages or not, because it means than the whole urban system or only parts of it are considered in the measure.

Are larger cities richer?

To disentangle these mixed conclusions, we choose to construct intermediate delineations of cities so as to explore the effect of three urban features: the central density, the integration of commuting suburbs and the population size.

Cities are delineated systematically using the method developed by Arcaute et al. (2015) and applied to France by Cottineau et al. (2017). It consists of an algorithm for clustering local units into urban cores, based on a density cutoff D: all contiguous local units of density higher than D are aggregated. A second algorithm attaches functional peripheries to these urban cores, based on the percentage of commuters of local units working in the centres F. A final criterion is used when clusters with a population of at least P inhabitants are selected. The advantage of this method is that it can produce representations of the urban system for a variety of values for each criterion.

We combined 39 density cutoffs, 21 flow cutoffs and 6 population cutoffs to build 4914 representations of the French system of cities, all made of aggregation of local units (the *communes*), from the most restrictive delineation of very dense centres with no functional periphery (close to local units, although without a complete coverage) to a very loose consideration of urban lifestyle that covers most of the French territory, thus close to a regional partition such as NUTS-3 regions. We proceed in an iterative way to produce our results (Figure 3), picking values for all parameters (the density, flow and population cutoffs which determine cities' boundaries, as well as the regressor and regressant variables) necessary to produce the corresponding clusters and their aggregate value of population, wages, etc. We then estimate the scaling parameter in the regression of interest and represent the value of the β or γ parameter on a heatmap which describes the definitional space of cities, with the value of the density cutoff D in the x-axis, the flow cutoff F in the y-axis and a new heatmap for each population cutoff P (Figure 3, step 4). The process is repeated for each of the 4914 unique combinations of definition parameters, for each unique combination of regressor and regressant variables. A selection of the resulting heatmaps is presented in Figure 4 for some variables of interest.

Each change in the definitional criteria affects the resulting urban clusters generated (see online supplementary Annex 1) and hence the measured levels of urban population, wages and jobs. However, it does not mean that the scaling relation between the economic outcome and inequality has to change with respect to size and density. If there were no dependence on city definition, we would observe a series of homogenous heatmaps throughout population cutoffs (cf. online supplementary Annex 2, pattern E and 2). In case of dependence on city definition, there is a variety of possible effects (see online supplementary Annex 2, patterns A to D) originating from the spatial definition of cities which can be interpreted in terms of the different representations of cities that particular case corresponds to. These oppositions should reflect and refine the apparent paradoxes highlighted with official delineations. We thus project these as well on our definitional space (labelled squares[5], Figure 4(a) and (b)).

Out of the 4914 regressions performed on 39 density cutoffs D, 21 flow cutoffs F and 6 population cutoffs P, we only display 3276 results by selecting the 4 population cutoffs $P = \{0, 10\,k, 20\,k, 50\,k\}$.

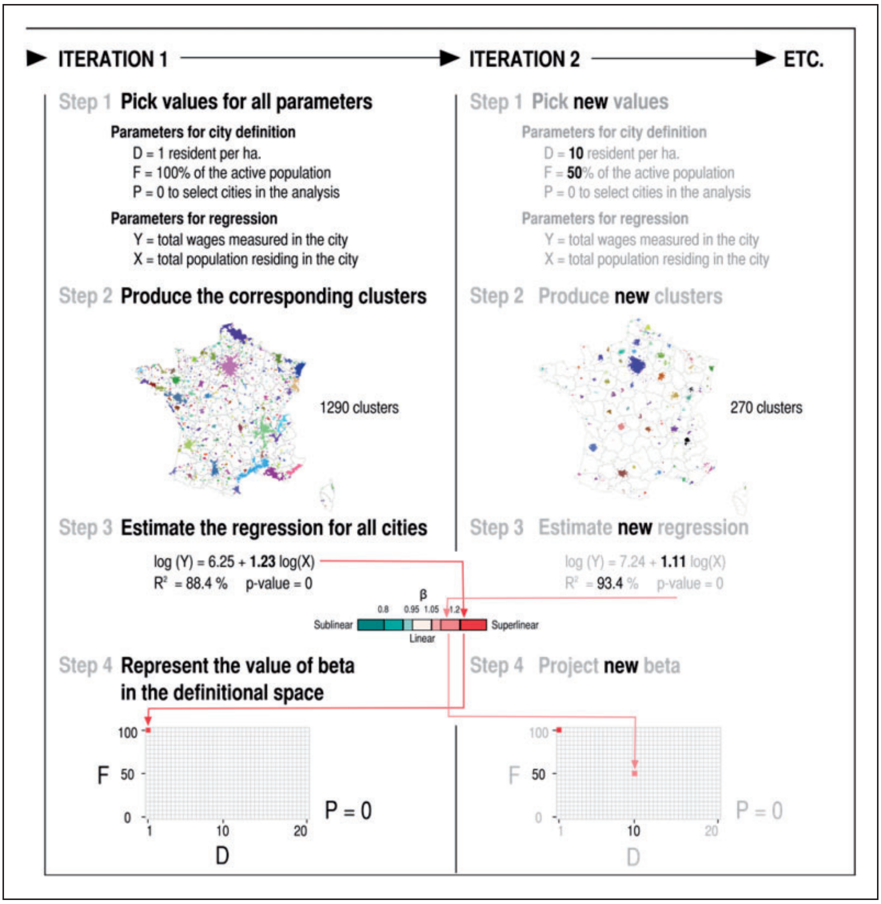


Figure 3. Algorithm used for the construction of Figure 4.

We use two different combinations of regression variables to investigate if larger cities are richer: total wages are regressed against the resident population and the number of jobs. Figure 4(a) represents the projection of selected results in the definitional space of cities, and highlights three main findings. Firstly, the size effects recorded are either positive (red) or nil (beige), meaning that *larger cities appear either richer or as rich as smaller cities, but never poorer* on average, regardless of the way cities are defined. Secondly, the spatial definition of cities impacts these measurements differently depending on the variables used to proxy city size. Considering total wages against total jobs, there is clearly no impact of city definition on scaling estimation, as all heatmaps appear homogeneously linear. In other words, larger urban economies do not provide higher wages, independently of where the boundaries for the urban aggregations are set. This goes against the matching and sharing explanations. However, with respect to the relationship between total wages and total residents, we see that larger cities, when they are defined as sprawling metropolises (bottom-right), do seem richer than their smaller counterparts. This is not true when we look at city cores only (top-left). This pattern only reflects the difference in spatial distribution of jobs and residents, with a stronger concentration of the first in the central parts of cities. We looked at another measure of wealth (total income declared by fiscal households, see online supplementary

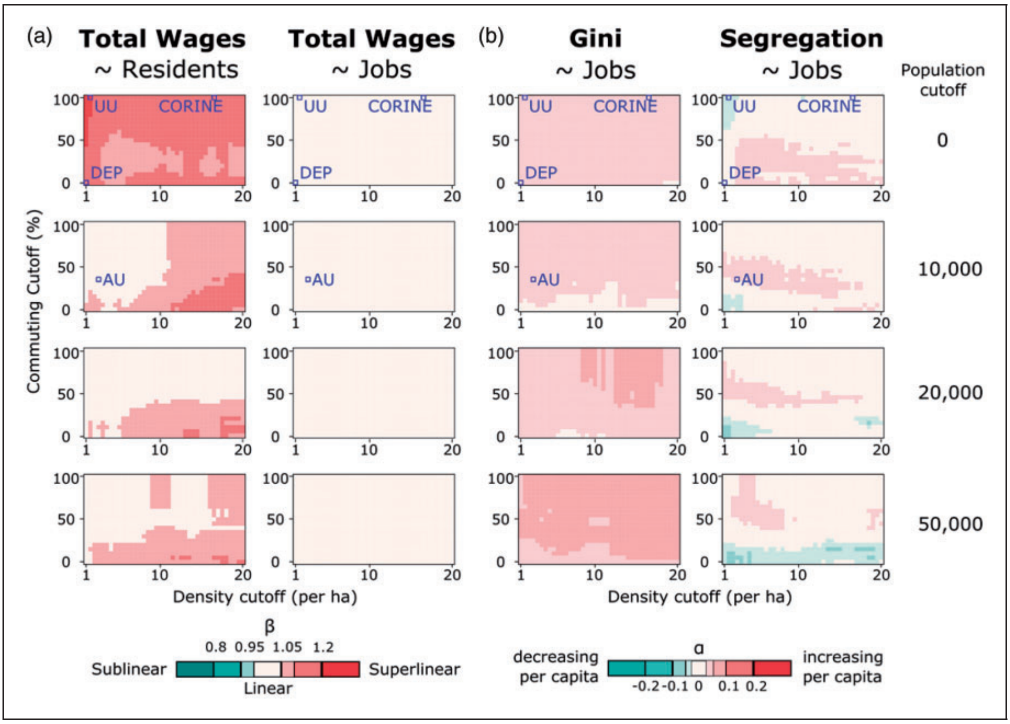


Figure 4. Variations of scaling estimations for wages and inequality with city definitions. (a) Regression on wages to test if larger cities are richer. (b) Regression on inequality indexes to test if larger cities are more unequal.

Annex 3) and found a slight evidence of a difference between the regressions estimated on cities considered as traditional cities (very dense core, no suburb) and cities in their regional extension, for high population cutoffs. The income premium in large cities is observed only in the first case.

To conclude, larger cities tend to concentrate the jobs rather than paying higher wages for the same job. In terms of income, there are positive size effects in the most urban parts of the largest cities (where density and integration are high). Therefore, larger French cities are not necessarily richer but the wages earned in central cities 'circulate' (Davezies, 2008) and end up more concentrated as income in the largest urban areas than as wages in dense areas in terms of jobs.

Are larger cities more unequal?

Our hypotheses H2 and H3 assume larger cities are more unequal and more segregated. They are tested here on a wide array of city definitions for France.

Inequality. As an aggregate measure of inequality, the Gini coefficient has the merit of being synthetic, scale-independent and comparable between distributions. Moreover, it correlates with other measures of inequality (Glaeser et al., 2009). Gini coefficients were estimated for the wages groups of Table 1, as in Fuller (1979). One value of the Gini index was computed for each city cluster of each of the ~5000 different city definitions, based on the number of

Table 1. Distribution of firms by average wages in France, 2008.

Category	Average wage (k€)	*N* firms	% Firms
D1	10–12.8	114,491	10.0
D2	12.8–14.9	114,492	10.0
D3	14.9–16.8	114,491	10.0
D4	16.8–18.7	114,492	10.0
D5	18.7–20.6	114,491	10.0
D6	20.6–22.8	114,492	10.0
D7	22.8–25.6	114,491	10.0
D8	25.6–29.5	114,492	10.0
D9	29.5–36.4	114,491	10.0
D10	>36.4	114,491	10.0
Firms from all deciles		1,144,914	100.0

Source: CLAP 2008. The number of firms indexed in CLAP (4,413,779) includes firms with no employees as well as confidential and unreliable data. It is thus much higher than the number of firms for which the mean wage is computed and used here (1,144,914).

firms of each wage category and their aggregate wages in the city. The value of the Gini index was then regressed against the log number of jobs for all cities s

$$Gini_s = \alpha * log(Job_s) + b + \varepsilon_s \tag{3}$$

In this case, the scaling parameter α is interpreted with reference to 0, which corresponds to the absence of significant size effects.

Figure 4(b) shows the values of α projected on the parameter-space of city definitions. A first conclusion is that size is not sufficient to predict the level of inequality in cities. Indeed, the statistical fitness of the regression models for all the city definitions was quite weak (typically, less than 40%). This is characteristic of per-capita measures, whereas regression models of absolute values tend to produce higher values of R^2 (Shalizi, 2011). The picture is clear though: *larger cities are either as or more unequal than smaller cities, but never more equal* on average (column 3). The coefficient belongs to the interval [0;0.015] for every definition and economic specification, meaning than the Gini index increases from 0 to $2^{0.015}$= 1.05 (in %) with every doubling of city size. The way one defines cities spatially is mainly irrelevant with respect to this relation for wages, but the way centres are defined matters for income (Wheeler, 2006; cf. online supplementary Annexes 3 and 2, pattern A): inequality is larger in larger cities when they are defined with a restrictive density value ($D > 10$). This means inequality is an observable problem when looked at in terms of labour market at a regional level, but it 'disappears' when the lens is too focused on urban spaces of high density. Online supplementary Annexes 4 and 5 confirm the increase of inequality with city size when population cutoffs apply, and reveal that it is not due so much to a polarization of urban societies but to the concentration of high-income earners in the largest cities (Sarkar et al., 2016).

Segregation. Another way of looking at inequality in cities is to look at the degree to which economic groups segregate spatially within cities. The ordinal nature of wage groups calls for ordinal measures such as the ones developed by Reardon (2009). The ordinal variation ratio index R^O corresponds to the '*proportion of variation in a population that lies between, rather* *than within, organisational units*' (Reardon, 2009: 150) in this case: the local units which compose the urban clusters. This measure does not depend on the overall inequality of each city, but only on the spatial distribution of groups within the city. One value of the segregation index was computed for each city for each city definition, based on the number of firms in each wages category in the local units (*communes*) composing the urban clusters. This value was then regressed against population (number of jobs in the case of wages) for all city-like clusters.

In contrast to claims in the American literature (Bischoff and Reardon, 2013; Logan, 2011), most of our results show *no link between city size and economic segregation*, regardless of how cities are defined[6] (Figure 4, see online supplementary Annex 3). Indeed, most scaling estimates are non-significantly different from 0 and most models have very low R^2 (comprised between 0% and 20% for wages). Despite the low level of statistical explanation of these model, an interesting result in Figure 4(b) column 4 is the slight positive relationship between wages segregation and size which appears for definitions close to that of AU, for example metropolitan areas with 30–60% of converging commuting flows from neighbouring local units. For other types of definitions (local CORINE or regional DEP), the relationship is absent. For built-up areas (UU), there is a slightly negative relationship, which means that the larger the central cities in terms of jobs, the less segregated the firms in terms of average wages. The opposition between the two definitions of cities might pertain to the higher polycentrism of metropolitan areas, which offer more opportunities and fluidity for fragmentation of the production space in larger cities, similarly to what happens in the housing market (Watson, 2006).

To conclude, larger cities do appear slightly more unequal, and expectedly, the picture would appear even clearer if housing costs were included. With respect to segregation, we found no strong evidence of scaling, as the variation between cities of similar size seems wider than the variation across sizes.

Conclusions

There is a large diversity of theories and models of agglomeration economies. A thorough evaluation of the sensitivity of empirical estimates to economic and geographic definitions helps going beyond the mixed evidence reported (or censored, cf. Melo et al., 2009) in the literature. This paper has reviewed causal mechanisms leading to agglomeration economies, questioned the specificity of cities in that respect and analysed size effects on productivity and inequality.

Evidence from French cities and administrative partitions reveal that economic specifications are crucial to answer the question about the urban specificity. Our results show that agglomeration economies measured depend on the scale of observation: we found evidence of their presence at a regional scale, although there is a productive advantage to local concentrations of workers. These results are consistent with causal mechanisms of sharing and learning in central cities, and of sorting in wider metropolitan areas. Moreover, larger cities appear either richer or as rich as smaller cities, but never poorer and larger cities appear either more or as unequal as smaller cities, but never more equal on average.

The title of this paper suggests that there might be a way to define urban clusters which would allow us to detect agglomeration economies. Indeed, why should urban definitions matter to the estimation of agglomeration economies? Firstly, because agglomeration economies are considered to be urban, but there is no agreement as to what a city is and how to delineate it. Secondly, because the estimation of other parameters changes with city definition, as infrastructural and socio-economic attributes. Thirdly, because cities are heterogeneously populated and contain activities, jobs and amenities being more concentrated than the resident population within and between cities. And finally because geolocated estimations are subject to systematic spatial biases in general. Based on the results reported in the paper, we find indeed variations due to spatial definitions. Built-up areas correspond to the delineation that produces economies of agglomeration in the most consensual understanding of the term (higher wage output with a denser working population). The absolute size effect on income also happens for dense city cores. However, this increased productivity comes with higher inequality and ignores the fact that the people working in the dense part of the city, generating economic output by interacting closely during the day, tend to commute back to other parts of the city which are more consumptive of infrastructure such as roads as they grow in size (cf. Cottineau et al., 2017). Moreover, theory has to include the impacts of borrowed size effects to explain the non-superlinear scaling of some European countries (Arcaute et al., 2015; Meijers and Burger, 2017).

Our study is itself limited in the narrow scope of the models, which leave off instrumentation and other predictors of productivity levels. This needs to be examined in further analysis. It also does not allow any conclusion regarding the cause of the observed statistical relations. However, as a first step, it demonstrates that taking the spatial structure of cities into account with respect to their inequality is crucial for economic geography, given their consequences on health, education, crime, social and spatial equity.

Acknowledgement

The authors thank Fabien Paulus for providing access to the CLAP database. We also thank anonymous reviewers and participants of the workshops and conferences where the results of this paper were presented and discussed.

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) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the EPSRC under the Urban Dynamics Lab Grant (EP/M023583/1) and by the European Research Council under the Advanced Grant GeoDiverCity (269826-ERC-2010-AdG).

Notes

1. Standard Metropolitan Statistical Areas, or their more recent equivalents: Metropolitan Statistical Areas (MSAs), Micropolitan Statistical Areas, together forming the Combined Statistical Areas (CBAs).
2. www.insee.fr/fr/themes/detail.asp?reg_id=99&ref_id=base-cc-evol-struct-pop-2011
3. www.insee.fr/fr/methodes/default.asp?page=definitions/clap.htm
4. www.statistiques.developpement-durable.gouv.fr/donnees-ligne/li/1825.html
5. The frame CORINE corresponds to the closest match of clusters with the urbanised land cover of satellite images CORINE. http://land.copernicus.eu/pan-european/corine-land-cover

6. This could be explained by cultural differences between the two countries, by the different indexes used or by the fact that homogeneously rich and homogeneously poor cities are mechanically less segregated (Dabet and Floch, 2014). The present data do not allow us to draw any conclusions about this ambiguity.

References

Arcaute E, Hatna E, Ferguson P, et al. (2015) Constructing cities, deconstructing scaling laws. *Journal of The Royal Society Interface* 12(102): 1–8.

Bischoff K and Reardon SF (2013) Residential segregation by income, 1970–2009. In: Logan JR (ed.) *The Lost Decade? Social Change in the U.S. after 2000*. New York: Russell Sage Foundation.

Bretagnolle A, Paulus F and Pumain D (2002) Time and space scales for measuring urban growth. *Cybergeo: European Journal of Geography*. DOI: 10.4000/cybergeo.3790.

Briant A, Combes PP and Lafourcade M (2010) Dots to boxes: Do the size and shape of spatial units jeopardize economic geography estimations? *Journal of Urban Economics* 67(3): 287–302.

Cheshire PC, Nathan M and Overman HG (2014) *Urban Economics and Urban Policy: Challenging Conventional Policy Wisdom*. London: Edward Elgar Publishing.

Chetty R, Hendren N, Kline P, et al. (2014) Is the United States still a land of opportunity? Recent trends in intergenerational mobility. *The American Economic Review* 104(5): 141–147.

Ciccone A (2002) Agglomeration effects in Europe. *European Economic Review* 46(2): 213–227.

Ciccone A and Hall RE (1993) *Productivity and the Density of Economic Activity*. National Bureau of Economic Research.

Combes P-P, Duranton G and Gobillon L (2011) The identification of agglomeration economies. *Journal of Economic Geography* 11: 253–266.

Cottineau C, Hatna E, Arcaute E, et al. (2017) Diverse cities or the Systematic paradox of urban scaling laws. *Computers, Environment and Urban Systems* 63: 80–94.

Dabet G and Floch JM (2014) La ségrégation spatiale dans les grandes unités urbaines de France métropolitaine. Une approche par les revenus. *Documents de travail INSEE* 2014/01.

Davezies L (2008) *La République et ses territoires: la circulation invisible des richesses*. Paris: Seuil.

Duranton G. and Puga D (2004) Micro-foundations of urban agglomeration economies. *Handbook of Regional and Urban Economics,* Amsterdam: Elsevier, 4: 2063–2117.

Eeckhout J, Pinheiro R and Schmidheiny K (2014) Spatial sorting. *Journal of Political Economy* 122(3): 554–620.

Fingleton B (2006) The new economic geography versus urban economics: An evaluation using local wage rates in Great Britain. *Oxford Economic Papers* 58(3): 501–530.

Fragkias M, Lobo J, Strumsky D, et al. (2013) Does size matter? Scaling of CO_2 emissions and US urban areas. *PLoS One* 8(6): 1–8.

Fuller M (1979) The estimation of Gini coefficients from grouped data: Upper and lower bounds. *Economics Letters* 3(2): 187–192.

Glaeser EL, Kolko J and Saiz A (2001) Consumer city. *Journal of Economic Geography* 1(1): 27–50.

Glaeser EL, Resseger M and Tobio K (2009) Inequality in cities. *Journal of Regional Science* 49(4): 617–646.

Henderson JV (1986) Efficiency of resource usage and city size. *Journal of Urban Economics* 19(1): 47–70.

INSEE (2010) Géographie administrative et d'étude. www.insee.fr/fr/methodes/nomenclatures/zonages/default.asp (accessed 5 July 2017).

Jacobs J (1961) *The Death and Life of Great American Cities*. New York: Vintage.

Leitao JC, Miotto JM, Gerlach M, et al. (2016) Is this scaling nonlinear? *Open Science* 3(7): 150649.

Logan JR (2011) Separate and unequal: The neighborhood gap for blacks, Hispanics and Asians in Metropolitan America. *Project US2010 Report*, 1–22.

Marshall A (1920) *Principles of Economics*. 8th ed. London: MacMillan.

Martin R (1999) The new 'geographical turn' in economics: some critical reflections. *Cambridge Journal of Economics* 23(1): 65–91.

Meijers EJ and Burger MJ (2017) Stretching the concept of 'borrowed size'. *Urban Studies* 54(1): 269–291.

Melo PC, Graham DJ and Noland RB (2009) A meta-analysis of estimates of urban agglomeration economies. *Regional Science and Urban Economics* 39(3): 332–342.

Moomaw RL (1988) Agglomeration economies: Localization or urbanization? *Urban Studies* 25(2): 150–161.

Mori T and Smith TE (2015) On the spatial scale of industrial agglomerations. *Journal of Urban Economics* 89: 1–20.

Openshaw S (1983) *The Modifiable Areal Unit Problem*. Norwich: Geo Books.

Parr JB (2007) Spatial definitions of the city: four perspectives. *Urban Studies* 44(2): 381–392.

Piketty T (2013) *Le capital au XXIe siècle*. Paris: Seuil.

Reardon SF (2006) A conceptual framework for measuring segregation and its association with population outcomes. *Methods in Social Epidemiology* 1(169): 169–192.

Reardon SF (2009) Measures of ordinal segregation. In Flückige Y, Reardon SF and Silber J (ed.) *Occupational and Residential Segregation (Research on Economic Inequality, Volume 17)*. Bingley: Emerald Group Publishing Limited: 129–155.

Roscigno VJ, Tomaskovic-Devey D and Crowley M (2006) Education and the inequalities of place. *Social Forces* 84(4): 2121–2145.

Rosenthal SS and Strange WC (2001) The determinants of agglomeration. *Journal of Urban Economics* 50(2): 191–229.

Rosenthal SS and Strange WC (2004) Evidence on the nature and sources of agglomeration economies. *Handbook of Regional and Urban Economics* 4: 2119–2171.

Royuela V, Veneri P and Ramos R (2014) Income inequality, urban size and economic growth in OECD regions. *OECD Working Papers*, 2014/10. DOI: 10.1787/20737009.

Rybski D, Reusser DE, Winz A-L, et al. (2016) Cities as nuclei of sustainability? *Environment and Planning B: Urban Analytics and City Science*. DOI: 10.1177/0265813516638340.

Sarkar S, Phibbs P, Simpson R, et al. (2016) The scaling of income distribution in Australia: Possible relationships between urban allometry, city size, and economic inequality. *Environment and Planning B: Urban Analytics and City Science*. DOI: 10.1177/0265813516676488.

Shalizi CR (2011) Scaling and hierarchy in urban economies. *arXiv*. 1102.4101.

Sveikauskas L (1975) The productivity of cities. *The Quarterly Journal of Economics* 89(3): 393–413.

Van Ham M, Tammaru T, De Vuijst E, et al. (2016) Spatial segregation and socio-economic mobility in European cities. *IZA Discussion Paper*, 10277.

Watson T (2006) Metropolitan growth, inequality, and neighborhood segregation by income. *Brookings-Wharton Papers on Urban Affairs* 1–52.

Wheeler CH (2006) Urban decentralization and income inequality: Is sprawl associated with rising income segregation across neighborhoods? *FRB of St. Louis Working* 4(1): 41–57.

Wilkinson RG and Pickett KE (2006) Income inequality and population health: A review and explanation of the evidence. *Social Science & Medicine* 62(7): 1768–1784.

Wong DWS, Lasus H and Falk RF (1999) Exploring the variability of segregation index D with scale and zonal systems: an analysis of thirty US cities. *Environment and Planning A* 31: 507–522.

Clémentine Cottineau is a Research Associate at the Centre for Advanced Spatial Analysis of University College London. She studied geography and economics at University Paris 1 Panthéon-Sorbonne and holds a PhD in geography from the same university. Her thesis aimed at modelling urbanisation in the post-Soviet space. Since 2014, she has been working at the Centre for Advanced Spatial Analysis on urban economics and scaling in the context of European systems of cities, using tools from complexity science to investigate inequality in cities and economic disparities between cities.

Olivier Finance holds a doctorate in geography from the University Paris I (Panthéon-Sorbonne). During his PhD, he worked in the fields of urban, economic and quantitative geography, as part of the ERC GeoDiverCity team. More specifically, he investigated the transnational firms' strategies in the French urban system through scaling laws. As a postdoctoral fellow, he now contributes to the analysis of diverse kind of human interactions in the region of Brussels and more broadly in Belgium, using community detection methods (BRUNET project, Université Catholique de Louvain, Belgium).

Erez Hatna is an Assistant Professor in the Center for Advanced Modeling. He is also honorary researcher at the Centre for Advanced Spatial Analysis (CASA) of University College London and a member of the Systems Institute. Dr Hatna obtained his PhD from the department of Geography and Human environment of Tel Aviv University in 2007 and then worked as a postdoctoral researcher at Wageningen University and research associate at University College London. Dr. Hatna is active in the fields of spatial analysis, urban dynamics and agent-based modelling.

Elsa Arcaute is a Lecturer in Spatial Modelling and Complexity at the Centre for Advanced Spatial Analysis (CASA). Elsa is a physicist with a master in Mathematics (part III of the Mathematical Tripos) and a PhD in Theoretical Physics from the University of Cambridge, UK. She has been working in the field of Complex Systems for many years. She was previously part of the Complexity and Networks group at Imperial College London, where she worked on self-regulation in social systems. The work focused on ant colonies, robots, and on the development of an intervention for the viability of an Irish eco-village. Currently, her research focuses on scaling laws and on defining cities using different methodologies; from networks and percolation theory, to GIS tools.

Michael Batty, CBE FRS FBA, is a Bartlett Professor of Planning at University College London where he is Chair of the Centre for Advanced Spatial Analysis (CASA). He has worked on computer models of cities and their visualisation since the 1970s and the work of his Centre CASA is heavily concentrated on the development of urban analytics, big data and smart cities. His blogs www.complexcity.info cover the science underpinning the technology of cities and his posts and lectures on big data and smart cities are at www.spatialcomplexity.info. His most recent book is The New Science of Cities (MIT Press, Cambridge, MA). Prior to his current position, he was Professor of City Planning and Dean at the University of Wales at Cardiff and then Director of the National Center for Geographic Information and Analysis at the State University of New York at Buffalo. In 2016, he was awarded the Gold Medal of the Royal Town Planning Institute.