Urban Scaling Revisited: Size, Scale, and Shape

Michael Batty[1] Yao Shen [2]

[1] Centre for Advanced Spatial Analysis (CASA), University College London, 90 Tottenham Court Road, London W1T 4TJ, UK, m.batty@ucl.ac.uk, t@jmichaelbatty

[2] College of Architecture and Urban Planning (CAUP), Tongji University, 1239 Siping Road, 200092, PR China, eshenyao@tongji.edu.cn

ABSTRACT

For urban scaling, we define four different sets of relationships that tie together theories and methods that describe and explain how cities and their spatial locations change as they scale. By scaling we mean changes in the size of urban phenomena such as population that take place as cities grow and more generically, how cities change over time. These relationships cover city size distributions and the rank size rule, urban density functions that relate to how dense populations are with respect to their location around the central area of cities, how gravitational interactions between locations scale with respect to distance, and finally how attributes relating to size in cities such as income scale allometrically as cities change in size. These four relationships can be associated with those who first popularised their form, which in the order we introduce them, are what we call Zipf's Law, Clark's Law, Tobler's Law, and Marshall's Law. Having described their form, we illustrate their application to the distribution of employment and populations for small areas and for whole cities in the UK, reflecting a form of spatial intelligence for planning informed by the principles of urban scaling.

Keywords: Scaling; Allometry; Size; Population; Employment

Acknowledgements

This project was part-funded by the Alan Turing Institute under the *RUBICON, QUANT* Projects 2020-2023, ELAT2/100002 and EP/W006022/1, the Chinese NSFC grant (52278074) and the Fundamental Research Funds for the Central Universities (2212024370).

1 The Context

One hundred years ago in thinking about how biological systems evolved, the gestalt psychologists who believed that our understanding of the world was composed of a unique ability to synthesise diverse elements in our environment, adopted the mantra that "… the whole is greater than the sum of its parts". It was Kurt Koffka (1935) who is accredited with this phrase although the idea goes back to classical times, to the words of Aristotle who implied that synergy is a key property in explanation. This notion that 'more can come from less' encapsulates the idea that as a system evolves, it can change qualitatively, which is often revealed in its changing shape or morphology. The most obvious and immediate example is the human form where a new born baby's head is proportionately much bigger than its body mass whereas a growing child quickly stabilises in terms of the shape of its body parts as adolescence is passed. As we grow, indeed as any object grows, one critical property is its scale. Indeed as an object gets bigger or smaller, it is said to scale and in this sense, if an object scales and changes qualitatively with respect to the ways it elements form and interact, then we say that the phenomenon is '*scaling*'. This is essentially the meaning of scaling as developed in the many contributions that comprise this book.

The simplest, and in some respects the special case of scaling, is when an object increases or decreases its size linearly, that is proportionately to its mass or volume or some other measure of its geometry. If this increase is more than proportionate, the relationship is *super linear* while if the increase is less than proportionate, the relationship is *sub linear*. In biology, super linear is often called positive allometry while sub linear is called negative allometry and there is a direct correspondence between this scaling and economies and diseconomies of agglomeration in economics. In turn these scale economies are sometimes mapped onto the positive benefits that might accrue from growth or negative benefits, and these have different interpretations and formulations (West, 2017). Although the focus is usually on growth, scaling in spatial systems is usually defined not with respect to the processes of growth but to the emergent patterns that at any particular time, display positive or negative allometry. Although systems can be seen as being organised from the top down with successive subsystems scaling in some fashion as a hierarchy of parts, this can also be seen as a series of ever larger subsystems that emerge from the bottom up. In fact scaling in this fashion is entirely coincident with the complexity sciences whose elements often display a particular motif that is replicated at ever larger scales, as fractal patterns that scale non-linearly. Although we do not have the luxury of developing these ideas here, the notion that we can define geometric scaling in mathematical terms serves to focus our initial review on the key relationships that recur again and again in the various chapters that follow. Readers can find more detailed explanations of the relationships to different morphologies, particularly cities, in Batty and Longley (1994) and Batty (2005).

2 The Basic Scaling Relations

The core relation relates the size of some set of objects, phenomena, or system to their order, and by order we mean the relationships between the objects. The simplest order is based on arranging the objects with respect to some property such as their size (or scale) and examining the way this size distribution behaves. We first define the set of n objects where each object is defined as $P_r, r = 1, 2, 3, \ldots, n$ where the obvious order is simply to associate the size of the object to its rank measured with respect to its reverse order. This relates P_r to $n-r-1$ where $P_r \propto n-r-1$ and it is clear that this is a simple linear relation where size is proportional to the reverse rank. These is no scaling in this relationship for the size of the object is the same as its reverse rank. This is almost nihilistic in its form but if we arrange the objects by size in their basic rank order, we derive the nonlinear relation

$$P_r = Kr^{-1} \quad . \tag{1}$$

K is a constant that determines the dimension or metric of the relation. This is the classic rank size relation associated with the frequency distribution of the set of $\{P_r\}$ objects and now known as the 'rank size' rule or *Zipf's Law* (Zipf, 1949). To demonstrate that the relationship is scaling, if we scale the rank r in equation (1) by the parameter λ as $(\lambda r)^{-1}$, this can be manipulated to show that

$$\lambda^{-1}P_r = K(\lambda r)^{-1} = K\lambda^{-1}r^{-1} \propto P_r \quad . \tag{2}$$

This says if we double the rank, by say $\lambda = 2$, we halve the size of the object in question and this indicates the more than proportionate decrease which is associated with the scaling.

This relation has been very widely used to examine the relative frequencies and attributes of objects that form systems where the objects compete against one another to grow; for example, the size of cities, income distribution, the frequencies of words, and the distribution of many natural and physical phenomena from animal populations, flora and fauna to geophysical distributions such as earthquakes and volcanoes. In terms of cities, Zipf (1949) was one of the first to demonstrate that city size distributions followed the pure rank size scaling as in equation (1) although many others who followed him, have noted that the inverse scaling is more likely to use a power law where the exponent α differs from 1. This is written as $P_r = Kr^{-\alpha}$ although it is likely that the 'true' relation is more like a lognormal with the inverse power only associated with the upper or heavy tail of the distributions where the largest objects exist.

The second form of scaling that we introduce is in some senses the opposite or even complement to the rank size rule. Imagine that the ranks no longer relate to the order of objects from large to small but to the distance from some origin or source. In other words, the objects are no longer ranked in terms of their size, but the distance of each object from some source now reflects the order. Whereas the rank size rule represents order in which the objects are ranked from large to small, the new rule is based on an order associated with the objects that arranges them according to their position. This is exactly the way we would examine an attribute of an object at a series of successive distances from some source. It might be the density of say population at increasing distances from the central core of the city, as reflected in what is called *Clark's Law*. Clark (1951) like some before and many after argued that the density (and/or rent) of population falls off according to some inverse exponential of distance from some point usually the central business district (CBD) of large metropolitan areas. If we now define the density of the object as P_j and the distance of the object from the CBD as $d_j, j = 1, 2, 3, \dots, n$, then Clark hypothesised that the relation should be a negative exponential, that is

$$P_j = Kexp(-\beta d_j) \quad , \tag{3}$$

where K is the dimensional constant as before and β is the scaling parameter. In fact in many applications of these kinds of model, an inverse power is used instead of a negative exponential and thus the complement to the rank size rule that we might call the rank distance rule is

$$P_j = K d_j^{-\beta} \quad . \tag{4}$$

Note that the key difference between equations (1) and (4) is the size of the populations in (1) are ordered by size from $P_1 \dots P_k \dots P_n$ while the sizes in (4) are arranged in the order given by distance from the CBD $d_1 \dots d_j \dots d_n$.

Our third scaling function also relates to position and distance but in terms of interaction between any two locations and the flow of activity between them. In the same way that Clark's Law defines how density decays from a given locational focus, the CBD, interaction defined as T_{ij} between any two locations i and j is a function of the distance between them d_{ij}. We can define this kind of relation from one of many gravitational equations typical of which is

$$T_{ij} = K P_i P_j d_{ij}^{-\gamma} \quad , \tag{5}$$

where K and γ are appropriate parameters and the basic scaling is applied to distance d_{ij} in the same manner as we demonstrated above for both Zipf and Clark's Laws. This equation reflects the fact that interaction T_{ij} falls off with distance from i or j and this has been called *Tobler's Law* after Tobler (1970) who articulated his Law as follows: He said: "… everything is related to everything else, but near things are more related than distant things …". The gravitational model in equation (5) can be generalised in many ways but variants have been derived in which all the independent variables of size and distance can be parameterised which makes them subject to differential scaling. If we now define the locational variables P_i and P_j with respect to their scale as P_i^{α} and P_i^{β} and then use these in an augmented gravity model, equation (5) becomes

$$\tilde{T}_{ij} = K P_i^{\alpha} P_i^{\beta} d_{ij}^{-\gamma} \quad . \tag{6}$$

By raising the locational variables to a power, this enables the locations to reflect economies or diseconomies of scale. If we scale the locational variables by a and b and distance by λ, then it is easy to show that the interactions $\tilde{T}_{ij}$ in the new model in equation (6) are composed of a more complex set of functions that scale the whole equation by $a^{\alpha} b^{\beta} \lambda^{-\gamma}$, again a constant scalar.

This brings us to our fourth type of scaling and in this context, physical growth of the system becomes more explicit. There are many phenomena where changes to the size of the objects in question lead to variants of linear growth as we defined them earlier. Super linear growth manifests itself if an object gets larger and some attribute of the object gets more than proportionately bigger due to its internal dynamics. Sub linear growth operates in an opposite manner for as the object grows, some attribute in question grows less than proportionately. This kind of growth was articulated by the economist Alfred Marshall (1890) in the late 19th century where he defined it implicitly as agglomeration: economics of scale if growth is super linear or diseconomies of scale if sub linear. In various economic models, such scaling also features but in the context of spatial systems such as cities, it is of a particularly simple form with the best examples being those demonstrated by West (2017) and his colleagues for different city sizes with respect to changes in overall income. We will call this *Marshall's Law* (1890) after his early discussion of economies of scale in his book *Principles of Economics.*

If we define income for a city or location i as Y_i, and assume that the size of the driver of income is population P_i, then the scaling relation that has been widely fitted to income and related population data for cities over the last 10 years or more and is reflected in several chapters in this book, is of the simplest form (Bettencourt et al. 2009). We can state this as

$$Y_i = KP_i^{\varphi} \quad , \qquad (7)$$

where the parameters K and φ act in the same way as in the other scaling relationships discussed above. This is the form of the so-called allometric equation that has been adopted in recent work on scaling in cities although there is some uncertainty about whether or not a better estimate would be of the income per capita against population, that is

$$Y_i/P_i = KP_i^{\varphi-1} \quad . \qquad (8)$$

The parameter does not vary if the form in equation (8) rather than (7) is used but the goodness of fit is perceived differently. In this last relationship, we introduce much stronger substantive issues than in the previous cases where the focus before has been very much on geometry, position and location. By introducing allometry , we open this discussion of scaling to more important issues pertaining to how city and social systems function. In this research, we will not pursue this further for others writing in this book will take up and elaborate these themes. To round out our discussion here, we will now demonstrate empirically how we might begin to estimate the various relationships we have presented.

3 Demonstrating Scaling for City Systems and Systems of Cities

The four relationships that we have just introduced are applicable generically to many types of system and although our focus has been on cities and regions, the ideas do not need to be developed for spatial systems per se. For example, rank size relationships are widely applicable to firm sizes, income distributions, and large language models while urban density functions pertain to many systems where diffusion of a rapidly decaying emission around some source needs to be modelled. Gravitational models clearly apply to how people interact within cities but they are applicable to information, demographic migration, and energy flows as well as to trips and traffic. Last but not least, allometric relations pertain directly to animal morphologies as well as to a variety of economic and infrastructural elements that make up the wider environment.

In the rest of this chapter, we will apply each of these relations to the UK urban system which is defined at the level of census tracts known as Middle-layer Super Output Areas (MSOAs) of which there are 8436 in England, Scotland and Wales which have an average population of 7791 and average employment of 2378. These are much smaller than the 63 cities in the UK greater than 110,000 population in 2023 but they represent an elemental unit against which we can test for the existence of distributions which scale with their size. The fact that Britain is an island economy and that it is almost entirely urbanised, introduces a degree of uniformity into its spatial morphology that mirrors classic industrial cities and city regions. In the examples that follow, we define 'population' as total workplace employment from the 2011 Population Census which we consider mirrors classical urban structure much better than the distribution of residential population defined from a total head count.

The first relationship we present is based on the 8436 MSOAs which we rank in terms of employment from the largest zone (in the City of London, the square mile) which has 325,874 employees to the smallest which has 14. In fact, in these examples, we will use employment density which assumes that employment exists at a point location and in this way we can assume that the unit of analysis – the employment zone (or the 'city') – is dimensionless, that is, it is normalised to account for area. Most rank size analyses of locational distributions such as population are usually done on entire countries – systems of cities – where cities are defined according to geometric considerations and density thresholds, so the analysis we indicate here is unusual in that our zones are not cities. Nevertheless we consider them to be competitors of one another for resources, notwithstanding the well-known problems of defining their physical and socio-economic boundaries (Arcaute et al., 2016).

Our first application is to fit the rank size equation $P_r = Kr^{-\alpha}$ to the ranked MSOA data ($r = 1, 2, 3, \dots, 8436$) which are ordered from the largest to the smallest employment density zone. In fact what we plot in Figures 1 (a) and (b) are the inverse power equation $P_r = Kr^{-\alpha}$ and its logarithmically transformed equivalent, $\log P_r = \log K - \alpha \log r$. Noting that the fit of the equation to the observed employment density is fairly good but modest with an $R^2 = 0.844$, it is worth making the point that not much analysis of rank size at a disaggregate scale has ever been accomplished. In fact Figure 1(b) illustrates that the long tail (to the right side of the graph) distorts the rank size linearity considerably and the heavy tail (to the left side of the graph) reduces the slope even further. Were we to chop off most of the long tail upwards from rank $r = 1000$, the slope looks much more like Zipf's Law with an $R^2 = 0.935$ but these results suggests that there are many, many different distributions which are defined at different spatial scales that resemble rank size. Moreover as we begin to trim the tails to increase linearity, then the nature of the system that such scaling is applicable to begins to change. To an extent, this probably implies that estimates of this kind are not necessarily good measures of the dynamics of rank size but simply evidence of the fact that systems with limited resources inevitably always form an equilibrium which can be described by Zipf's Law.

In Figure 1(b) we also show the pure rank size rule normalised to ensure that the intercept of the function is the maximum employment density. The straight line thus models the equation $\log P_r = \log P_1 - \log r$ and this shows quite clearly the problem with the tails. Although our first scaling law is based on the classic inverse power, there are several ways of defining this scaling associated with the spatial distribution of activities such as employment and their densities. Instead of ranking the size of activities from largest to smallest where the rank is equally spaced as $r = 1, r = 2, r = 3, \dots$, we order these sizes according to their location. In this application, we define location as the distance from central London where we define the first distance $d_1 =$

0.5 as the average distance within the MSOA which is Charing Cross. With all 8436 locations measured with respect to this zone, and with these distances ordered from the smallest to largest $d_1 \leq d_2 \leq d_3$ ..., we can examine the distribution of activities $E_1 \leq E_2 \leq E_3$ ..., and infer whether or not this distribution scales. Remember that in the rank size rule, activities are ranked from largest to smallest but in this second application, it is the distances that are ranked. In short this second application changes the order from size to distance where we plot rank and distance on the same horizontal axes. The most important issue here is that this transfer focuses the analysis on the morphology of the system in that its geometry – ranking the distance associated with the relevant activity at that location – provides an alternate way of looking at the same data, the distribution of employment densities in this case.

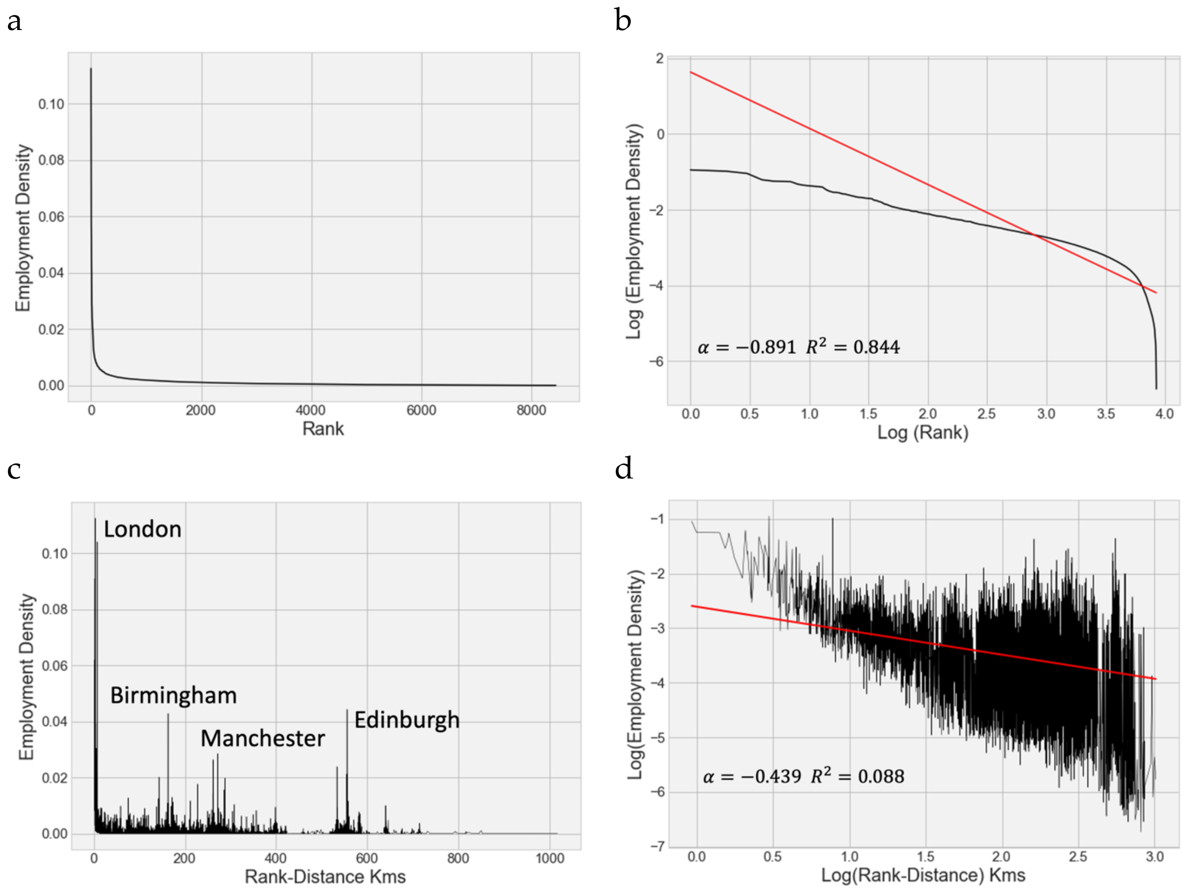
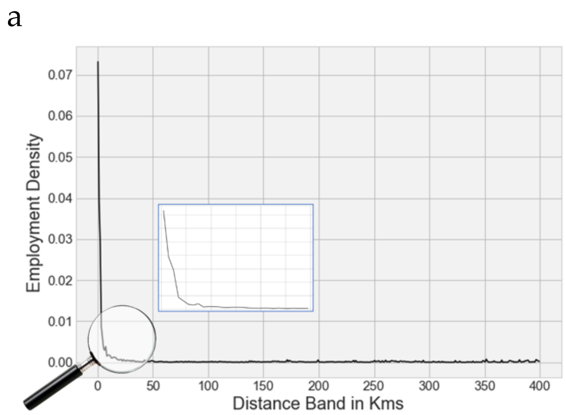


Figure 1: Classic Scaling Relationships a) and b) Rank Size, and c) and d) Rank Distance

In Figure 1(c), we plot the employment density E_j against what we call the rank distance d_j. Whereas in the rank size of employment densities in Figures 1(a) and (b), there is nothing in this distribution to link it to the morphology of employment densities but when we use rank distances, it is extremely clear that the distribution of cities in Britain measured as their distance from London, mirrors the hierarchy of the city system that is well known. In Figure 1(c), London stands out as the origin of the UK urban system as a massive density. 240 kilometres distant from London, Birmingham is located, while at 300 kms Manchester can be clearly identified. Within these distances, West Yorkshire and then Newcastle can be located before the last big urban agglomeration Glasgow-Edinburgh which is some 550 kms from London. In Figure 1(d), we plot the logarithmic transformation, $\log P_j = \log K - \alpha \log d_j$ but this transformations masks the morphology. The further away from London, the wider the variation in densities and at greater distances from the capital, the employment density falls dramatically. In fact there is a hidden 'rank size' here if we regress log P_j against $\log d_j$ with the slope estimated as $\alpha = -0.439$ but with the R^2 falling to 0.088.

The second relationship that we have defined relates density and distance to activity distributions such as employment and population in the manner formulated by Clark (1951) that we introduced previously. We have already defined these activities with respect to their distance from the UK's capital London but to demonstrate urban density scaling, we have defined rings of equal distance from the core Charing Cross in bands of 1 km distance. In each annulus, we add the total employment in the relevant MSOAs that fall predominantly in each band and then divide by the area of the annulus to extract the density. As we begin at the core point and then move successively out from London, we eventually reach the outer annulus some 1018 kms from London which includes the last zone number 8436. This method smooths the variations in density and produces a much cleaner picture of density variation in the UK. To an extent. it is an alternative picture of the densities that we plot in Figure 1(c) and (d). However, as we move past 400 kms from London, we encounter annuli which have no employment associated with them and it is therefore not possible to plot these effectively in automated way. We therefore restrict our visual analysis to this limit which does not quite include urban Scotland although this is far enough away to make little difference to the overall scaling implied in these plots.



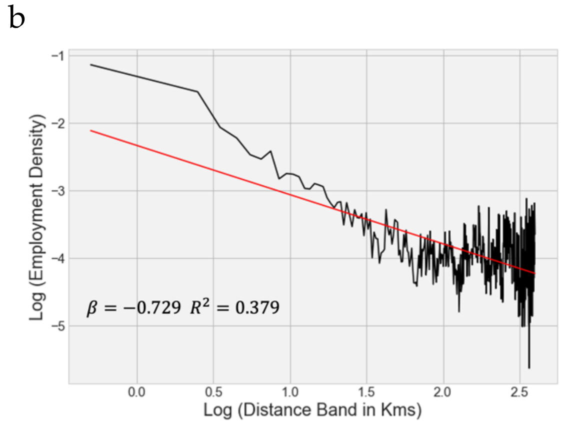


Figure 2: Urban Density Profiles for London a) Inverse Power b) Logarithmic Transformation

We show the employment density profile in Figure 2(a) which reveals a very clear inverse power law scaling based on $P_j = K d_j^{-\alpha}$. The density falls dramatically within the first 50 rings up to 50 kms as we show in the inset. To examine the stability of the relationship up to 400 kms from London, we plot the logarithmic transformation log $P_j = \log K - \alpha \log d_j$ in Figure 2(b) which reveals that the density accords to Clark's Law up to about 30 kms from the centre (Charing Cross) with an $R^2 = 0.958$. From then on, the variations get greater although over the entire system, the fit is good. From Figure 2(b), the slope of the function is $\beta = -0.729$ while the variance explained

falls to about 0.379. Figure 2(b) can be interpreted as a continuous equivalent of the rank distance relationship that links Clark's to Zipf's Law.
Our third scaling involves the way patterns of interaction between zones that we refer to as origins and destinations scale with respect to the measure of impedance or deterrence between locations. We are not able to demonstrate this here in any quick and simple manner but it is intrinsic to the kind of scaling that is reflected in the model shown in equation (5). This model is the workhorse that we have developed for the urban system based on the island of Britain where the trips between origins and destinations are simulated using a model where the inverse power law has been replaced with the negative exponential that is scaling in a somewhat different manner. Then the generic model is

$$T_{ij} = E_i P_j \exp(-\gamma d_{ij}) / \sum_j P_j \exp(-\gamma d_{ij}) \quad , \qquad (6)$$

where γ is the scaling parameter, E_i is the employment at the origin zone i and P_j is the employment at the destination zone j. The model we have built is in fact further disaggregated by different travel modes and an appropriate way of thinking about this to assume that the model is equation (6) is fitted separately to each of three modes of travel where the parameter values for car, bus and rail are estimated as 0.131, 0.072, and 0.064. When we aggregate all these trips over all modes, we can work out an average flow from any zone to all others and plot these vectors for each zone, giving the picture below in Figure 3 which clearly shows that the cities in Britain scale according to size (Batty and Milton, 2021). The picture is still impressionistic but it provides a glimpse of the fractal structure of flows that make up the urban system.

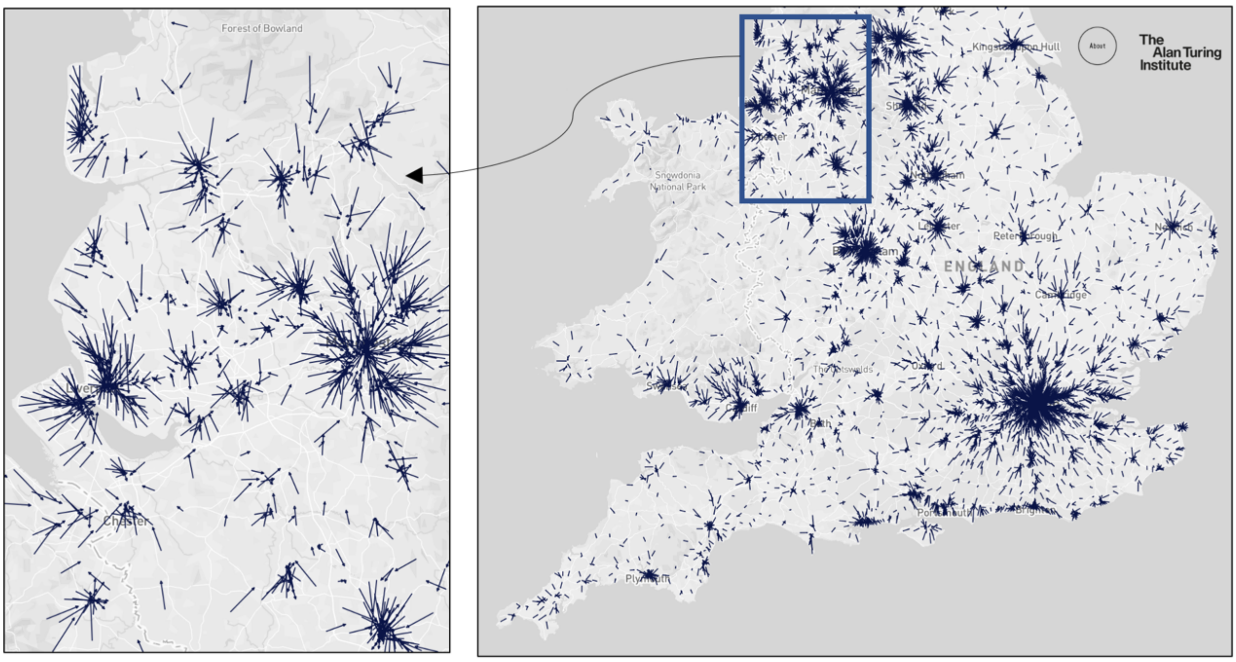


Figure 3: Spatial Interactions Patterns That Scale Over City Size

4 Urban Allometry and Scale

Our fourth and last scaling relation that we introduce here turns the relationship from a negative (inverse) one to a positive one in that it concerns the attributes of a variable as it changes in size. As cities get bigger for example, they change qualitatively as we noted earlier when we introduced the idea of allometry. The notion that one variable changes at a different rate from another but that they are intrinsically linked lies at the core of this relation. In the examples we have identified so far, we have not discussed the need for the units that we are explaining in terms of employment and population locations to be internally consistent. In fact the rank size relation is largely applied to cities that are defined to be quite separate from one another and when it comes to positive scaling, it is essential that the units are integral objects which exist in their own right. In the examples so far, we have used partitions of cities into smaller output area units but when it comes to comparing attributes such as different features of their size, we need to ensure that the units themselves are relatively independent from one another. In this sense, here we need to ensure that we are dealing with well-defined cities. Rather than 8436 MSOAs, we will now use the 63 cities defined for the UK by the Centre for Cities (CfC, 2023) which are cities greater than 110,000 in population in 2023 which can be easily compared with one another.

The idea of allometry is captured in equation (7) which implies that the attribute Y_i of the object i in question – a city, say – scales in a positive fashion with city size P_i^{φ}. The parameter φ is assumed to be positive; if it is between 0 and 1, the attribute increases sub linearly with city size but at a decreasing rate whereas if φ is greater than 1, the increase is super linear with city size, the attribute growing more than proportionately. As we noted above, this implies that growth realises economies or diseconomies of scale. This is the relation first articulated by Marshall (1890) and explored extensively by Bettencourt et al. (2007) which has motivated many of the chapters in this book. In Figure 4(a), we show the scaling relationship between the total weekly wages and population in the CfC data base and what is striking is that there is hardly any evidence of scaling; in short, Figure 4(a) implies that wages per capita barely change as city size increases. It is also clear that London is a clear outlier with a substantial premium in terms of being associated with increasing wages compared to all the other cities in the data set.

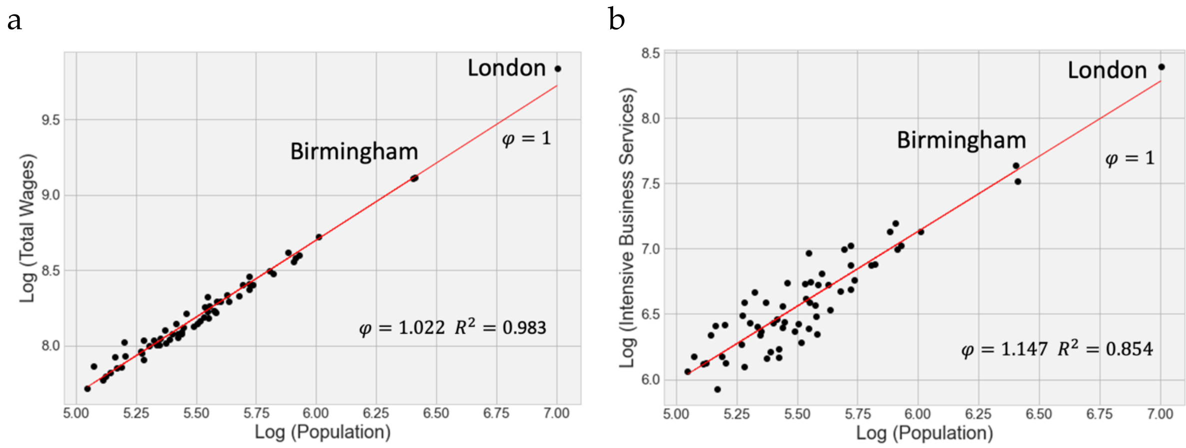


Figure 4: Linear and Super Linear Scaling for the Biggest 63 Cities in the UK

In terms of city size, an old argument that has gradually gained pace during the last 100 years is the idea that as a city grows in size, its economies of scale outweigh its diseconomies. This is best seen in terms of the fact that as more and more people come to interact in cities, then economies of scale are generated in proportion to the positive power of population, that is in proportion to population whose scaling is positive, $P_i^\alpha, \alpha > 1$. Diseconomies such as congestion and access to specialist facilities do not appear to cancel out economies. Measures of this positive power are revealed in terms of the relative growth of the service sector, measures of innovation in technology such as the acquisition of patents, and other measures of the knowledge base and the concentration of internetworking facilities that increase the accessibility of every larger cities in the global economy. Figure 4(b) demonstrates that this super liner scaling is applicable to the CfC data for intensive business services which are highly clustered in the largest cities. The CfC database reveals that some of these speculations are born out for a rather limited number of the largest UK cities in contrast to evidence from other parts of the world where positive super linear scaling seems to be much more evident (Bettencourt, 2021).

5 Conclusions: A Cornucopia of Scaling

There are however a number of open questions that are key challenges to be resolved in terms of any theory of scaling and these relate to how we define the units of analysis that we are explaining with respect to various of their attributes (Arcaute et al., 2016). Cities are notoriously hard to define with hard and fast boundaries particularly in an increasingly global world where many of the world's largest cities are dominated by waves of international migration. Many industries from primary to quaternary to the most esoteric information services which form the fast growing quinary sector are physically diffuse and spread across many cities and countries. This makes much economic activity hard to ground at unambiguous locations. The other feature that we have not explored is how city size relates to temporal scaling. Most of our focus has been on cities in equilibrium although in terms of economies of agglomeration, there are implicit relations that describe how cities change their scaling as they grow. Very little work has been developed in this regard but how we can explain for example how agglomeration economies defined by allometric parameters shift from lower to higher value as cities grow in size represents a serious challenge that any robust theory of scaling must surely embrace. These are all problems for future research.

What we need is a theory or theories that tie all these different scaling relations together. To do this, such a theory which would be based on the geometry and morphology of city systems and systems of cities, ideas first championed many years ago by Berry (1964) in an inspiring article but barely followed up in the intervening years. This as Berry argued must also be intrinsically linked to spatial economic theories where geometry is implicit, to ideas about how location theories and urban economies function. And all this must be set in a landscape of globalisation, where the scaling relations that we have defined here are under continual transformation. Glimpses of this theory are contained in the pages of this book and collectively they point to a greater understanding of how scaling can be used to define ways in which we can develop more sustainable and robust approaches to the design of better cities.

References

Arcaute, E., Molinero, C., Hatna, E., Murcio, R., Vargas-Ruiz, C., Masucci, A. P., and Batty, M. Cities and Regions in Britain through Hierarchical Percolation. *Royal Society Open Science*, 3 150691; DOI: 10.1098/rsos.150691, 2016.

Batty, M. *Cities and Complexity: Understanding Cities with Cellular Automata, Agent-Based Models, and Fractals*. Cambridge MA :The MIT Press, 2005.

Batty, M. and Milton, R. A New Framework for Very Large-Scale Urban Modelling. *Urban Studies*, *58* (15), 3071–3094, 2021.

Batty, M. and Longley, P. A. *Fractal Cities: A Geometry of Form and Function*. London and San Diego CA: Academic Press, 1994.

Berry, B. J. L. Cities as Systems Within Systems of Cities, *Papers of the Regional Science Association*, 13, 146–163, 1964.

Bettencourt, L. M. A. *Introduction to Urban Science*. Cambridge MA :The MIT Press, 2005.

Bettencourt, L. M. A., Lobo, J., Helbing, D., , and West, G. Growth, Innovation, Scaling, and the Pace of Life in Cities. *Proceedings of the National Academy of Sciences*, 104 (17) 7301-7306, 2007.

CfC Cities Data Tool, available at https://www.centreforcities.org/data-tool/#graph=map&city=show-all London: Centre for Cities, 2023.

Clark, C. Urban Population Densities. *Journal of the Royal Statistical Society, Series A* (General), *114*, 490–496, 1951.

Koffka, K. *The Principles of Gestalt Psychology*. London: Routledge and Kegan Paul, 1935.

Marshall, A. *Principles of Economics*. London: MacMillan Company, 1890.

Tobler, W. A Computer Movie Simulating Urban Growth in the Detroit Region. *Economic Geography*, 46(Supplement), 234–240, 1970.

West, G. *Scale: The Universal Laws of Life and Death in Organisms, Cities and Companies*. London: Weidenfeld and Nicolson, 2017.

Zipf, G. K. *Human Behavior and the Principle of Least Effort*. Cambridge MA: Addison-Wesley, 1949.