



Alma Mater Studiorum - Università di Bologna
DEPARTMENT OF ECONOMICS

**DID ZIPF ANTICIPATE
SOCIO-ECONOMIC SPATIAL
NETWORKS?**

Aura Reggiani
Peter Nijkamp

Quaderni - Working Paper DSE N° 816



DID ZIPF ANTICIPATE SOCIO-ECONOMIC SPATIAL NETWORKS?

Aura Reggiani* and Peter Nijkamp**

*Department of Economics, Faculty of Statistics, University of Bologna, Piazza Scaravilli 2, 40126 Bologna, Italy; email: aura.reggiani@unibo.it

**Department of Spatial Economics, VU University, De Boelelaan 1105, 1081 HV Amsterdam, The Netherlands; email: p.nijkamp@vu.nl

Abstract

An avalanche of empirical studies has addressed the validity of the rank-size rule (or Zipf's law) in a multi-city context in many countries. City size in most countries seems to obey Zipf's law, but the question under which conditions (e.g. sample size, spatial scale) this 'law' holds remained largely underinvestigated. Another complementary question is whether socio-economic networks in space also show a similar hierarchical pattern.

Against this background, the present paper investigates – from a methodological viewpoint – the relationship between network connectivity and the rank-size rule (or Zipf's law) in an urban-economic network constellation. After a review of the literature, we address in particular the following methodological issues: (i) the (aggregate) behavioural foundation underlying the rank-size rule/Zipf's law in the light of spatial-economic network theories (e.g. entropy maximization, spatial interaction theory, etc.); (ii) the nature of the analytical relationship between social-spatial network analysis and the rank-size rule/Zipf's law. We argue that the rank size rule is compatible with conventional economic foundations of spatial network models. Consequently, a spatial-economic interpretation – as well as a network connectivity interpretation – of the rank-size rule coefficient is provided.

Our methodological contribution forms the foundation for the subsequent empirical analysis applied to spatial networks in a socio-economic context. The aim here is to test the sensitivity of empirical findings for changes in scale, functional forms, time periods, and network structures. Our application is concerned with an extensive spatio-temporal panel database related to the evolution of urban population in Germany. We test the relevance of the rank-size rule/Zipf's law, and its evolution over the years, and – in parallel – the related 'socio-economic' connectivity in these urban networks. In particular, we will show that Zipf's law (i.e., with the rank-size coefficient equal to 1) is only valid under particular conditions of the sample size. The paper concludes with some retrospective and prospective remarks.

Key-words: rank-size rule, Zipf's law, entropy, spatial interaction, connectivity, German municipalities

1 Hierarchical Spatial Structures: Introduction

“How much validity and universality should be attributed to this rank-size rule is, at this stage, a matter of individual opinion and judgment” (Isard, 1956, p. 57).

The spatial organization of a system of cities has received much attention in the urban-geographic and regional science literature over the past decades. In particular, studies on Christaller-Lösch systems – and the related rank-size rules – have culminated in a wealth of literature on Zipf’s law. This law addresses essentially the issue of spatial organization of systems of places. One may distinguish two main strands of the literature that form the foundation for a hierarchical ordering of spatial units. The first paradigm emerges from gravity theory, which forms the basis for spatial interaction modelling. The gravity model is based on two basic assumptions, viz. the attraction (or centripetal) forces exerted by a certain mass or centre (population, production, amenities, etc.) – this is, in fact, an agglomeration externality – and the distance friction exerted by the geographical separation of two units – this is, in fact, cost-minimizing behaviour of actors in space. Thus, agglomeration forces and spatial cost-minimization may be seen as the foundation of spatial interaction models. Combined with heterogeneity of goods or services traded or transported between different places, this also forms a cornerstone of the New Economic Geography. It can be demonstrated that the essence of such a model is equivalent to an entropy-maximizing paradigm of a spatial flow model (see Nijkamp and Reggiani, 1992; Gordon, 2010). Zipf’s law is then a particular case of entropy, as we also demonstrate in Section 2.

The complementary foundation of Zipf’s law stems from trade theory in market network analysis. In accordance with Lösch’s spatial equilibrium analysis, geographical space is governed by spatial units (centres) with agglomeration advantages, the presence of transport costs for traded goods, and the existence of price-elastic demand curves for heterogeneous goods. This leads to spatial product specialization, in which the highest-order goods (often luxury goods, with lower price elasticity) are produced in the biggest agglomeration (i.e. the city with the highest rank), while the lowest-order goods (often daily goods, with a high demand elasticity) are found in all or most places. This hierarchical equilibrium order implies essentially a combination of individual profit maximization at the firm level and total transport cost minimization for all actors in the system (see also Isard, 1956).

A merger of spatial interaction models and the hierarchical organization of trade networks between regions can be found, inter alia, in Bos (1964) and Tinbergen (1964), where the authors demonstrate that trade specialization may generate trade revenues that increase total welfare through cost minimization of spatial transport or trade flows. Their hierarchy of open spatial systems is based on an underlying hierarchy of the economic force fields in trading regions, and highlights the importance of sectoral heterogeneity in transport and trade.

Over the years, an avalanche of publications has been produced on the foundation and measurement of spatial hierarchies of places or regions. Most of the literature has addressed the issue of a hierarchy of cities, an issue that found its main origin in the work of Zipf (1949). The practical idea was that spatial hierarchies in networks (e.g. of the Christaller-Lösch type or of the Tinbergen-Bos type) were difficult to test, unless simple measurement schemes were introduced. Zipf’s law – or in a more general sense, the rank-size rule – was an indirect test to confirm the validity of functional hierarchies in spatial systems. More recently, considerable attention has also been given to the identification of hierarchical organization of networks in a broader context (see, e.g., Barabási, 2002). Our paper aims to

advance the current debate by critically reviewing the Zipf's law assumption, by next proposing alternative (or complementary) measurement schemes, and by focussing attention more directly on structural regularities in spatial networks, both conceptually and empirically.

It is noteworthy that the regional science literature over the past decades has seen an intensive ongoing debate on the empirical rank-size distribution¹ of cities, inspired by the basic contribution by Zipf (1949). Starting from Simon (1955) and Isard (1956), who explained the existence of power laws in city-size distributions, there has been a permanent flow of interesting interpretations of the significance of rank-size distributions, and hence of Zipf's law. We refer here to, amongst others, Krugman (1996a), who questioned the economic interpretation of Simon, by arguing "*Because while a relationship like equation 1 is difficult to explain with an equilibrium story about determination of city size, it is quite easy to justify with a nihilistic story of the kind analyzed by Herbert Simon*" (p. 246)².

Many discussions have centred around the universality and the potential of power functions in city-size distributions. Clearly, other functions have been found to be suitable as well. For example, the lognormal function or the double Pareto lognormal function have also received much attention in the literature (see, e.g., Giesen et al., 2010). It is noteworthy that most studies have restricted themselves to truncated samples, by focussing mainly on large cities in a national context (see also Section 4). Given the abundant literature on these issues, our purpose is to briefly review the main results from – and to offer new thoughts on – the continuing debate on the empirics of rank-size distributions of cities and the underlying theoretical frameworks. In particular, we will highlight that:

- the rank-size distribution of cities (or rank-size rule) may also be valid for functions different from the power form, for example, for the exponential form, the Tanner function, etc.;
- the choice of the unit in a geographical space, as well as the choice of the sample size of cities, is critical;
- despite the empirical nature of the rank-size rule, a methodological interpretation of Zipf's law – and its related rank-size rule – should not only address hierarchical patterns in city size, but also socio-economic networks in space.

Starting from the above methodological propositions, a second aim of this paper is then to explore the structural dynamics of urban systems, with reference to rank-size/Zipf's rules, as well as to network analysis, over a period of different years, and for different spatial scales and population sizes. In our empirical analysis, we will consider the population of urban centres in Germany in terms of various spatial scale levels (region, district, city, municipality, city-network), for different years. The emerging different patterns will also be investigated by means of a number of functional forms in the rank-size distribution (more specifically, by adopting power, exponential, lognormal, and Tanner functions). These functions play a crucial role in spatial modelling, in particular in conventional Spatial Interaction Models (SIMs), and hence due attention is needed for the related spatial-economic underpinning of these models.

Furthermore, urban systems are usually connected – or networked – by means of physical and virtual infrastructures, and hence our analysis will also explore how rank-size rules may be linked to

¹"Rank-size distribution or the rank-size rule (or law) describes the remarkable regularity in many phenomena including the distribution of city-sizes around the world, sizes of businesses, particle sizes (such as sand), lengths of rivers, frequencies of word usage, wealth among individuals, etc. All are real-world observations that follow power laws such as those called Zipf's law, the Yule distribution, or the Pareto distribution. If one ranks the population size of cities in a given country or in the entire world and calculates the natural logarithm of the rank and of the city population, the resulting graph will show a remarkable log-linear pattern. This is the rank-size distribution" (Wikipedia, 2011).

² Equation 1 mentioned by Krugman is the rank-size rule (see expression (1) in Section 2.1).

social network analysis in space. From this latter perspective, it seems plausible that different types of infrastructure and socio-economic networks may contribute – or give rise – to different city-size distributions.

Our conclusion – both methodologically and empirically – is that the rank-size concept can be conceived as a ‘universal’ expression of many models traditionally used in urban and regional economics, and is thus able to: (i) grasp the homogeneity/heterogeneity of the spatial system concerned (at an aggregate level, by means of its rank-size coefficient); and (ii) understand the associated connectivity infrastructure or socio-economic network morphology (or socio-economic constellation) of complex spaces.

The present paper is organized as follows. Section 2 critically reviews the broader context of Zipf’s law, followed by an exposition on the relationship between Zipf’s law and socio-spatial networks in Section 3. This section also discusses the question whether the highest ranked places are also the best connected places in a network. Next, Section 4 is empirical in nature and offers a wealth of experiments on German population centres and socio-spatial networks. These empirical analyses are followed by concluding remarks in Section 5.

2 The Spatial Economics of Zipf’s Law

2.1 The Rank-Size Rule for City-Size Distribution

Despite the frequent reference in the geography literature to Zipf (1949), Auerbach (1913) was essentially the first to use the power-law concept to describe city-size distributions. The basic form of the rank-size distribution is:

$$P = AR^{-q}, \quad (1)$$

where P is a given size of a city population; R is the related rank of the city; q is the elasticity parameter, and A is a positive constant (usually the population of the biggest city). In logarithmic form, the rank-size distribution (1) shows a linear relationship as follows:

$$\log(P) = \log(A) - q \log(R). \quad (2)$$

In increasing order, the city with a rank equal to 1 is the biggest city of a country; the city with a rank equal to 2 is the second biggest city, and so on. The parameter q represents the slope of the logarithmic expression between the rank of the city and the size. The slope q may be larger, equal, or smaller than 1. Zipf (1949) stipulates that the city-size distribution follows the rank-size distribution with the q -coefficient equal or close to 1. This particular expression is known as ‘Zipf’s law’. It implies that the city with rank 1 is two times bigger than the city with rank 2, three times bigger than the city with rank 3, and so on. When q is larger than 1, it means that there is more disparity and heterogeneity between the cities, whereas when q is smaller than 1, it implies that the cities are more homogeneous (see, among others, Brakman et al., 1999). Clearly, Zipf’s law is a special case of the general rank-size distribution. It expresses a sort of ‘perfect hierarchy’ between the spatial units in geographical space. It is essentially based on spatial efficiency principles in a functional interconnected equilibrium framework.

Several researchers also refer to Zipf’s law as the Pareto distribution. This is formally correct, but we prefer to differentiate between these two expressions. In the present paper, we use formulation (1), where we have the rank on the x -axis and the city size on the y -axis, so as to comply with standard representations in regional science and geography. Thus, our empirical research is based on a rank-size expression of type (1) and its logarithmic form (2). Many authors have used this rank-size procedure; for

example, Berry and Garrison (1958) show the cases of spatial hierarchy in Korea and Washington. In more recent years, Pumain and Moriconi-Ebrard (1997) show the relevance of Zipf's law for a set of countries with data from 'Geopolis'³; Brakman et al. (1999) and Mulder and de Groot (2008) study the case of the Netherlands; Benguigui and Blumenfeld-Lieberthal (2010) present findings from Israel; and Matlaba et al. (2011) study the size distribution of urban areas in Brazil.

Clearly, other authors have used the axes in a reverse way, and then find, of course, the Pareto distribution. For example, Ioannides and Overman (2003) study the evolution of the distribution of populations of metropolitan areas for the USA; Duranton (2002) analyses the city-size distribution for the US Consolidated Metropolitan Statistical Areas (CMSAs); Nitsch (2005) provides a meta-analysis based on Zipf's law; Giesen and Sudekum (2011) explain Zipf's law in Germany at the regional and country level; and Rosen and Resnick (1980) examine the Pareto exponent at the country level. Clearly the two approaches – i.e. Zipf and Pareto expressions – are formally similar.

2.2 A Spatial-Economic Interpretation of the Rank-Size Rule

2.2.1 A Brief Review

As mentioned above, there is a large number of studies which show the validity of Zipf's law, particularly in the upper (truncated) tail (Clemente et al., 2011; Krugman, 1996a; Levy, 2009; Mansury, 2010; Rosen and Resnick, 1980). It should be noted that Soo (2005) has shown that the degree of variability of Zipf's law also depends on the estimation method used: he found contrasting results using OLS or Hill⁴ estimators. In particular, by using OLS, Zipf's law was rejected for the majority of countries analysed, while this was not the case using a Hill estimator. The present paper, however, does not aim to discuss the best estimator for analysing the rank-size rule/Zipf's law. Rather, we prefer to pay attention to the theoretical underpinning of the rank-size rule of type (1), and hence to an appropriate spatial-economic interpretation of its parameter q .

In spatial economics, Zipf's law – and its general rank-size rule – has enjoyed much popularity, mainly owing to its excellent fit in many empirical cases (for recent reviews, see, among others, Batty, 2005; Nota and Song, 2007; Terra 2009). These results were accompanied by more basic interpretations. Spatial-economic interpretations of the rank-size rule of type (1) can be positioned in the context of: Central Place theory (Berry, 1961; Krugman, 1996b); random growth (Simon, 1955; Gabaix, 1999); revealed comparative advantages (Hinloopen and Marrewijk, 2006); Markov chain models of migrating agents (Mansury, 2010); or general equilibrium theory (Brakman et al., 1999). However, in all these models the 'real' spatial arguments are missing, even though spatial interaction is taken into account, as, for example, in Hinloopen and Marrewijk (2006), and in Mansury and Gulyas (2007). In our study, we address this shortcoming in particular.

The first striking observation concerns the choice for and robustness of the power function, which has been largely accepted and empirically proven in hundreds of studies, often showing an excellent fit, with very high R^2 values. But some authors argue that this high R^2 holds, because the size of a city is intrinsically correlated to its rank (Gan et al., 2006). This reinforces the argument that Zipf's law is a statistical regularity and not an economic phenomenon. There is thus clearly a reason to pursue a more

³ "Geopolis is a data base that includes directly comparable population figures for all towns and cities in the world having 10,000 inhabitants or more" (Pumain and Moriconi-Ebrard, 1997, p. 308).

⁴ "Gabaix and Ioannides (2004) propose the Hill (1975) estimator as an alternative procedure for calculating the value of the Pareto exponent. Under the null hypothesis of the power law, it is the maximum likelihood estimator" (Soo, 2005, p. 244).

critical analysis of the economic foundations of Zipf's law. It should be added that, when considering the entire distribution of cities, including the smallest ones, the rank-size rule might not hold convincingly. Instead, a better distribution seems to be the lognormal distribution. This follows from Gibrat's law of proportionate growth (Glaeser et al., 1995; Gibrat, 1931). As a consequence, there is still an ongoing debate in the scientific literature on the difference between lognormal and power functions in the rank-size rule (see, e.g., Giesen et al., 2010; Eeckhout, 2004).

Starting from the above considerations, we will hypothesize here that the rank-size rule of type (1) is closely associated with entropy maximization, where the related power function depends on the analytical form of the constraint. Different functional forms for the rank-size rule can then emerge on the basis of different forms of the constraints (see also Section 2.2.2 below). It is noteworthy that Corominas-Murtra and Solé (2010) demonstrated that Zipf's law is the outcome of a class of stochastic systems, described in terms of statistical entropy; however, in their contribution, the value of entropy is fixed and it is neither a minimum nor a maximum. More recently, Corominas-Murtra et al. (2011) show how Zipf's law can emerge from a variational approach to the problem based on Kullbacks' minimum discrimination of information principle, in the framework of the evolution of communicating systems/human languages. In our entropy interpretation (see Subsection 2.2.2), we will show that the rank-size rule can also be derived from the entropy maximization approach as developed by Wilson (1967, 1970) and others.

2.2.2 Entropy and Rank-size Distribution

This subsection will present some central ideas on spatial entropy derived from Wilson's (1967) work, but revisited in a geographic population framework. Spatial flows are not present here, but the stock of population P_j in centre j can be considered as the outcome of a stock-flow process: namely, as the sum of the net incoming flows plus residents, so that it is intrinsically connected with spatial interaction. It is noteworthy that Parr (1985) has already interpreted the rank-size rule – and related concentration patterns – in the framework of cities' interaction: “*the process of concentration is facilitated by (and ultimately dependent on) improved interurban and interregional transportation*” (Parr, 1985, p. 208). Let us now consider E as the most likely combination of population stocks P_j , from among a very large number of realizations of independent micro-level outcomes, as follows:

$$E = \frac{P!}{\prod_j P_j!}. \quad (3)$$

The entropy E essentially refers here to the maximum probability of decentralization among random population centres. In the context of spatial entropy – as the basis for Spatial Interaction Models – a set of relevant constraints is normally imposed, e.g. on marginal totals or on the system's total costs. In our analytical framework based on Zipf's law on urban hierarchies, we may trace the aggregate implications of rank sizes of different centres as one of the cornerstones of our analysis. This then leads to the following Wilson type of entropy model for networks of centres:

$$\text{Max} \ln E = \text{Max} \left[- \sum_j P_j (\ln P_j - 1) \right], \quad (4)$$

s.t.

$$\sum_j P_j = P, \quad (5)$$

$$\sum_j P_j R_j = H, \quad (6)$$

where R_j is the rank of population centre j (with population size P_j). The expression of the constraint $\sum_j P_j R_j = H$ may be seen as some sort of aggregate production possibility frontier, a potential function which depicts the position of efficient urban centres as a product of their population size and their corresponding rank. This expression maps out the efficient distribution of population centres in a Christaller-Lösch framework. Any position away from this efficiency frontier means higher spatial costs for a movement towards this frontier. H may then be seen as the weighted agglomeration potential of the total spatial system characterized by a Löschian hierarchical efficiency principle. Thus, the rank R_j may be interpreted as a spatial (shadow) cost/benefit proxy that will be incurred to reach higher agglomeration benefits.

Problem (4), (5), and (6) can be solved by means of the Lagrangian equation:

$$L = -\sum_j P_j (\ln P_j - 1) + \lambda (P - \sum_j P_j) + \beta (H - \sum_j P_j R_j), \quad (7)$$

which leads to the following optimal equilibrium solution for P_j :

$$P_j = A \exp(-\beta R_j), \quad (8)$$

where

$$A = \exp(-\lambda). \quad (9)$$

In addition, if we consider the H constraint (6) in logarithmic terms, we obtain the following power form (with $\beta = q$):

$$P_j = A R_j^{-q}. \quad (10)$$

Formulation (10) is similar to the rank-size equation (1). In addition, by using the maximum entropy concept, we can also justify – in a rank-size rule context – analytical forms that are different from the power function (10), such as the exponential form (8) (see also Subsection 2.2.3). From a methodological viewpoint, the above analysis is clearly in line with the generalized cost minimization principle (see Nijkamp, 1975), and thus also with the ‘principle of least effort’ advocated by Zipf (1949) (although this was used by him in a language context). It should be noted here that the rank-size rule expressions (8) and (9), by being compatible with spatial interaction models of the Wilson type, are also compatible with discrete choice models, and hence find their origin in micro-economic behavioural theory. In a Lösch spatial equilibrium framework, the highest ranking centre has the highest internal efficiency (i.e. agglomeration advantages) and the lowest system’s overall interaction costs. And therefore, the exponent q may be seen as a shadow price for the transaction costs in moving from a lower rank towards a higher rank. In this framework, rank-size refers to the “*properties of the configuration of opportunities for spatial interaction*” (as the vector accessibility A_i in Weibull, 1980, p. 54).

It should be noted here that the value of $q = 1$ in expression (10) represents the specific case of Zipf’s law, in which – following Zipf’s terminology – the unification forces (influencing the masses P_j) and the opposing diversification forces (influencing the shadow prices R_j) are balanced according to a perfect equilateral hyperbola:

$$P_j R_j = A \quad (11)$$

where the constant A can be then interpreted as a sort of energy⁵ incorporating – at the equilibrium – the conflicting forces $P_j R_j$. It seems then obvious to argue that this ‘perfect equilibrium’ (11), emerging from an entropy maximization approach, may occur under specific space-time conditions, i.e., with particular samples and conditions of the variable P_j (or of any other economic variable under analysis in expression (10)), as we will show in our empirical application (Section 4).

2.2.3 *Alternative Functions for the City-Size Distribution*

The previous section has shown that the exponential form (12) in the rank-size rule can be derived from a theoretical framework related to entropy maximization (or generalized cost minimization). It should also be noted that ‘spatial’ entropy (5) can be interpreted as an indicator of the variety/diversity of a system (Nijkamp and Reggiani, 1992; Wilson, 2009; Gordon, 2010), while it can also be linked to Central Place theory (Gordon, 2010). The potential of the exponential and the power functions (expressed in formulations (12) and (14)) in mapping the homogeneity vs. heterogeneity of spatial-economic networks has also been highlighted in the various regional and urban economic studies, e.g. by Richardson (1969), Wilson (1967), Willigers et al. (2007), and recently by Reggiani et al. (2011a), in the context of spatial interaction modelling. In fact, an exponential functional form is consistent with the assumption of a constant distance decay parameter for all trip makers – who are then homogeneous with respect to this dimension. On the other hand, the power function is consistent with a gamma distribution for the distance decay parameter; that is, the population of trip makers is heterogeneous with respect to this parameter (Fotheringham and O’Kelly, 1989).

The previous methodological analysis has also highlighted the empirical relevance of rank-size rules. The power function (1), as well as the exponential function of type (8), will be adopted in our subsequent empirical tests that serve to explore the rank-size distribution of population at different spatial scales (Section 4). In addition, given the ongoing debate on the potential of the lognormal function in the presence of small units, the lognormal function will also be utilized. Finally, we also consider the Tanner function, which incorporates both the power and the exponential form. Therefore, we test the following set of alternative rank-size functions:

$$(a) \text{ the conventional power form: } P = AR^{-q}; \quad (12)$$

$$(b) \text{ the exponential form: } P = Be^{-\beta R}; \quad (13)$$

$$(c) \text{ the lognormal form: } P = Ce^{-\lambda(\log R)^2}; \quad (14)$$

$$(d) \text{ the Tanner form: } P = D(e^{-\beta R} R^{-q}). \quad (15)$$

The empirical results emerging from the estimation of the above functions (12)-(15) in the context of German regions will be presented in Section 4.

⁵ Surprisingly, the rank-size rule as in (10) resembles Einstein’s law (1905): $E = M C^2$, where E is energy, M is mass, C is the speed of light, which – like the rank R – may clearly have an economic value; here the coefficient of C assumes the value 2 (Newton’s value), instead of 1 (Zipf’s value). The population P_j in Eq. (10) may then be interpreted as mass M , and the constant A as energy E (Reggiani, 2012). It should be noted that Isard (1971) interpreted Einstein’s law in agglomeration theory; however, in his view the variable C represents flux or movement rather than a relative cost (or benefit) factor, as in the present interpretation.

The next important methodological step concerns the analysis of the validity of the rank-size rule in spatial networks. In other words, if we consider an urban system as a connected network, is the rank-size rule/Zipf's law still valid, and how is it related to network connectivity analysis? The subsequent section will offer some new thoughts on this issue.

3 Connectivity Networks under Zipf's Law

In recent years, there has been great interest in the interdisciplinary study of complex networks with special reference to the relevance of their topological structures. The idea that underlies this approach is the proposition that the topology (or architecture) of interactions between the various 'units' of the evolving networks is an essential constituent of many dynamic processes, which cannot be ignored without losing a crucial ingredient of the phenomena under consideration (Reggiani and Nijkamp, 2009; Vega-Redondo, 2007).

In this framework, several studies have been devoted to the relationship between many social and economic phenomena and their related network connectivity structures. Although the application fields are substantially different, it has been shown that these connected networks display common topological patterns. These patterns can be roughly associated with two main network models, Random Networks (RN) and Scale-Free (SF) networks, on the basis of the fundamental contributions by Barabási and his co-authors (Albert and Barabási 2000 and 2002; Barabási 2002; Barabási and Albert 1999, 2000). Several applications of these models can be found in different fields, such social science, biology, economics, technology, transport and telecommunication (see, e.g., Buchanan 2002; Reggiani and Nijkamp 2006; Reggiani et al., 2011a; Yook et al., 2002).

These network models are characterized by their (connectivity) degree distribution, i.e. the probability $P(k)$ that a chosen node in a certain network has exactly k links. In an RN context the degree distribution follows a Poisson distribution, while in SF models it follows a power-law distribution, i.e. $P(k) \sim k^{-\gamma}$, where γ represents the degree exponent. The Poisson distribution indicates that most nodes have approximately the same number of links, while in the power-law degree distribution "the probability that a node is highly connected is statistically more significant than in a random graph, the network properties often being determined by a relatively small number of highly connected nodes that are known as hubs" (Barabási and Oltvai, 2004, p. 105). It is interesting to recall here that the power-law function has also been extensively studied in biology and economics (Bak 1996; Krugman 1996b).

Interestingly, according to Barabási, the value of γ is relevant for determining specific properties of the network. "*The smaller the value of γ , the more important the role of the hubs is in the network. Whereas for $\gamma > 3$ the hubs are not relevant, for $2 < \gamma < 3$ there is a hierarchy of hubs, with the most connected hub being in contact with a small fraction of all nodes, and for $\gamma = 2$ a hub-and-spoke network emerges, with the largest hub being in contact with a large fraction of all nodes*" (Barabási and Oltvai, 2004, p. 102). In addition, Barabási and Oltvai also outline that for $\gamma > 3$ "most unusual features are absent, and in many respects the scale-free network behaves like a random one" (2004, p. 102).

We will consider the urban structure as a network, where all centres can be linked⁶, e.g. by internet, transport and social links, etc. The rank-size rule, or Zipf's law, may now be reconsidered from this new perspective, by comparing the q -coefficient in expression (1) with the connectivity degree distribution γ of the population according to Barabási's analysis. Our aim is then to infer a methodological correspondence between the rank-size rule/Zipf's law and network connectivity

⁶ According to Barabási's analysis, one flow-connection is sufficient for determining one network-link.

analysis. In this context, we will start from the following correspondence between the q -coefficient and the γ -coefficient:

$$\gamma = 1 + \left(\frac{1}{q} \right), \quad (16)$$

This expression is in agreement with Adamic (2000), who shows that the power-law γ -coefficient (emerging from the nodes' probability (frequency) distribution) is related to the rank-size q -coefficient (1) (emerging from ranking the variables), as indicated in equation (16). A systematic analysis – based on (15), and under the hypothesis that P_j captures the incoming degree distribution – leads to Table 1, which summarizes our novel findings and provides the following conclusions for various values of q and γ :

- $q = 1$: this expresses Zipf's law – i.e. a 'perfect' hierarchical structure – while we have a value of the degree distribution $\gamma = 2$, which is considered by Barabási and Oltvai (2004) as a hub-and-spoke network, with the largest hub connected with a large fraction of all nodes, as previously indicated. In other words, on the one hand, we can infer that the highest-ranked cities are also the more-connected, behaving as hubs. On the other hand, if we hypothesize that all the centres are – physically or virtually – connected when $q = 1$, we can infer that our urban system follows an SF pattern of a hub-and-spoke type, with also a fractal character. In this perspective, we can state that Zipf may be seen as a predecessor – on the basis of his law – of the socio-economic connectivity features proposed by modern network analysis, or vice-versa, that modern network analysis is revisiting Zipf's law (see also Reggiani, 2009). Fascinatingly, Berliant and Watanabe (2008) came, though with a different methodology, to a similar conclusion, by arguing that an SF network structure explains the city-size distribution.
- $q > 1$: from a spatial-economic viewpoint this situation indicates a high heterogeneity of the variables concerned (in our case, population), as was also indicated by other authors (see Section 2.1). From a social connectivity viewpoint, the corresponding value $\gamma < 2$ highlights the increase of the value of the hubs in the networks, which is in line with the related heterogeneous spatial-economic patterns, i.e. urban agglomeration patterns, and sheds a clear light on the hubs' role.
- $0.5 < q < 1$: from a spatial-economic viewpoint this situation indicates even more varied patterns of the population (in comparison to the previous case), with a trend, however, towards a hierarchy. The emerging network analysis displays for γ the values $2 < \gamma < 3$, which matches a hierarchy of hubs, with the most connected hub being in contact with a small fraction of all nodes. Once again, the results of the two analyses display an interesting correspondence between two sides of the same coin, i.e. spatial-economic and connectivity patterns.
- $q < 0.5$: from a spatial-economic viewpoint this situation indicates a very homogeneous pattern of the population. The network connectivity analysis shows then the value $\gamma > 3$, corresponding to a random network, thus completely in line with the even distribution of urban landscapes.

In summary, Table 1 depicts the 'methodological' correspondence between the rank-size coefficient (q) and the degree distribution coefficient (γ). It shows how Zipf anticipated modern network analysis, as advocated by Barabási about 50 years later. In particular, we may argue that, in general, an increase of the q -coefficient leads to a rise in spatial heterogeneity – which can also be interpreted as an increase of economic development, owing to interurban concentration (Parr, 1985) –, which corresponds to a decrease of the γ -coefficient (indicating an increase of network connectivity/hierarchy of hubs). In particular, for values of $q > 0.5$, we have a rise in spatial heterogeneity (e.g. urban disparities or

inequalities) which corresponds – in our network analysis – to an SF network pattern. In other words, a value of q around 0.5 may act as a critical threshold that makes the transition from spatial homogeneity (in terms of population size) towards spatial inequality, which corresponds to the transition – in the associated connectivity network – from a random network type to an SF network.

Table 1. Correspondence between Rank-Size Coefficient (q) and Degree Distribution Coefficient (γ)

Urban Structure	Urban Network
Rank-size coefficient (q)	(Connectivity) degree distribution coefficient (γ)*
	Scale-Free Network
$q > 1$ High urban heterogeneity	$\gamma < 2$ Increase of the value of the hubs in the networks
$q = 1$ 'Perfect' urban hierarchical structure (Zipf, 1949)	$\gamma = 2$ Hub-and-spoke network, with the largest hub being in contact with a large fraction of all nodes (Barabási and Oltvai, 2004)
$0.5 < q < 1$ Urban heterogeneity	$2 < \gamma < 3$ Hierarchy of hubs, with the most-connected hub being in contact with a small fraction of all nodes
	Random Network
$q < 0.5$ Urban homogeneity	$\gamma > 3$ Random network

* Note here that⁷: $\gamma = 1 + \left(\frac{1}{q}\right)$

Consequently, we can design a 'methodological' synthesis framework (see Table 2) which highlights that economic variables, such as population, GDP, etc., if taken into account in a rank-size framework, may be more appropriately interpreted by jointly exploring the associated connectivity structure (and vice versa).

⁷ See expression (16).

Table 2. Synthesis Framework: Two Sides of the Same Coin (adapted from Reggiani, 2009, p. 279)

Rank-Size Analysis	Network Analysis
<i>(Complex) interactions between nodes</i>	<i>(Complex) interactions between nodes</i>
Focus on the related <i>economic variables</i> and their ranking	Focus on the related <i>links</i> and their probability degree distribution
Focus on the <i>spatial-economic meaning of the functional forms</i>	Focus on the <i>connectivity patterns of the functional forms</i>

Thus far, we have achieved two major goals: (i) a formal correspondence between rank-size rules and entropy, and (ii) a new methodological interpretation of modern network analysis from a rank-size perspective. In the light of these findings, it is important to address the practical relevance of our results. And therefore, we now present an extensive empirical application on spatial-economic connectivity patterns of population in Germany, at different spatial scales.

4 Empirical Application: City-size Distributions in Germany

4.1 Rank-Size Rule Results for Different Spatial Scales

As previously outlined, our objective is to explore how the q -coefficients, emerging from the rank-size rules (12) and (15), as well as the β and λ coefficients emerging from (13) and (14), vary at different spatial scales. This study uses detailed spatio-temporal data on the Federal Republic of Germany.

The choice to use German data was partly instigated by data availability, but also partly by the fact that the Christaller-Lösch system mainly had its origin in the German space-economy. The basic unit is the municipality (Gemeinde); at the end of 2006, Germany consisted of 12,375 municipalities. Administratively, they are grouped into 439 districts (Kreise). In the case of the larger municipalities, some districts consist of only one municipality, and then both these administrative units are identical (117 cases). In these 117 cases, the municipality is called a “district-independent municipality” (“kreisfreie Stadt”). For the sake of analysis, the 439 districts are aggregated to nine spatial types following the classification of the German Federal Institute for Research on Building, Urban Affairs and Regional Development (Bundesinstitut für Bau-, Stadt- und Raumforschung, BBSR), which refers to the degree of urbanization and agglomeration. In addition, we will consider in our analysis a subset of the municipalities, conventionally named ‘cities’, on the basis of two criteria⁸: historical reasons, and size (usually more than 10,000 inhabitants).

In summary, in our study we analysed the following units: aggregation/typologies of districts (9), districts or ‘Kreise’ (439), cities or ‘Staedte’ (about 2100), municipalities (about 12,000), and the city-network (17). The time period of our data covers 15 years, so that we have a full panel sample.

Table 3 displays the mean values of the various coefficients q , β , λ over the years, for the three spatial scale levels (district, city, and municipality), as well as for each sample size of these population centres in Germany. It should be noted that it has been possible to report the mean values, since the

⁸ The authors would like to thank Alfred Garloff and Carsten Pohl (IAB, Germany) for their clarification of the German meaning of a ‘city’.

values of our coefficients (q, β, λ) are rather stable over the years (see Sections 4.2-4.6). In addition, the related mean R^2 is presented. Table 3 shows that the value of q -coefficients, as well as the values of the β and λ coefficients, vary for different spatial scales. In this framework, we have also tested the imposition of the cross-equation coefficient equality restriction⁹. First, we investigated whether there is a difference between the coefficients' values in the first two equations, the first and the third equations, and the second and the third equations. In all three cases the null hypothesis of parameter equality could be rejected at the 5 per cent significance level. These differences in the coefficients might be associated with the diverse economic development profiles of the different population centres concerned, by following the ideas of Berry (1961) on different city-size distributions related to the relative economic development of countries.

Table 3: Rank-Size Coefficients for Different Functional Forms and Spatial Scales in Germany

Rank-size Rule's Functional Form	Coefficient	District Population (439 observations; years: 1985-2006)	City Population (about 2,100 observations; years: 1993-2007)	Municipality Population (about 12,000 observations; years:1993-2007)
Power ^(a)	q	0.6172 (0.00868)	1.0571 (0.00799)	1.3979 (0.001227)
	R^2	0.9203	0.8963	0.8757
Exponential ^(b)	β	0.0047 (0.000078)	0.0018 (0.00001147)	0.000409 (0.0000002)
	R^2	0.890597	0.9210	0.9482
Lognormal ^(c)	λ	0.07244 (0.000865)	0.08961 (0.000733)	0.09581 (0.000064)
	R^2	0.941252	0.8703	0.9248
Tanner ^(d)	q	0.3668 (0.011997)	0.51302 (0.008309)	0.5502 (0.000943)
	β	0.0022 (0.000092)	0.00104 (0.0000138)	0.0003 (0.0000003)
	R^2	0.965178	0.9723	0.9818

Note: Standard errors in parentheses.

^(a)Conventional rank-size with power form: $P = AR^{-q}$;

^(b)Rank-size with exponential form: $P = Be^{-\beta R}$;

^(c)Rank-size with lognormal form: $P = Ce^{-\lambda(\log R)^2}$;

^(d)Rank-size with Tanner form: $P = D(e^{-\beta R} R^{-q})$.

⁹ This is done by means of the SURE test with STATA. A further analysis able to correct biases in small samples might be the rank-1/2 regression advocated by Gabaix and Ibragimov (2011). The authors wish to thank Giovanni Russo for all these suggestions and his valuable assistance in this respect.

Moreover, each function (power, exponential, lognormal and Tanner) seems to fit the city-size distribution well. In fact, the values of R^2 are very high (the smallest is 0.87 and the biggest is 0.98). It means that essentially no particular function is better than another.

If we focus on the conventional power form, it should be noted that the q -coefficient is very close to 1 when considering the city population. In this case, it perfectly matches Zipf's law (hierarchy). In the case of the district population, the q -coefficient is smaller than 1, showing characteristics of some homogeneity between the districts. On the contrary, by analyzing the population of the municipalities, the q -coefficient appears to be greater than 1. This means that there is a clear disparity between the municipalities in Germany, with strong urban agglomeration patterns. From a network viewpoint, the connectivity structures of the three spatial agglomerations seem – at their mean level – oriented towards an SF pattern (see Table 1).

The subsequent sections (4.2-4.6), provide more details concerning the analysis of the 'conventional' rank-size – embedding a power form – with reference to different spatial scale levels in Germany, while Section 4.7 conducts the same analysis by considering the alternative rank-size rule functions described in Section 3.

4.2 Rank-Size Rule Embedding a Power Form: Analysis of Aggregation of Districts in Germany

As a first step of our empirical analysis, formula (2) is applied – for the years 1985-2006 – to the nine typologies of the 439 German districts, according to their degree of urbanization and agglomeration (see Section 4.1, as well as Table A1 and Figure A1 in Annex A).

Table 4 shows the values of the q -coefficients at district level across the 9 types of economic regions in Germany. We can observe some interesting differences in the q -coefficient values; in fact, we see q -coefficient estimates between 0.86 and 0.87 for District-type 1, and between 0.71 and 0.73 for District-Type 2, thus significantly higher than the q values for the other economic types of districts. All the remaining types (3-9) display much lower values of the q -coefficient, below the value of 0.50, except District-types 6 and 8 which show values of around 0.50 in the last six years of the data set (see also Figure 1).

Table 4: Rank-Size Coefficients across Nine Types of Districts in Germany

	District type 1		District type 2		District type 3		District type 4		District type 5		District type 6		District type 7		District type 8		District type 9	
	q	R^2	q	R^2	q	R^2	q	R^2	q	R^2	q	R^2	q	R^2	q	R^2	q	R^2
1985	0.860	0.948	0.711	0.664	0.467	0.916	0.279	0.782	0.422	0.895	0.463	0.745	0.335	0.799	0.464	0.854	0.378	0.811
1990	0.861	0.953	0.716	0.670	0.455	0.926	0.267	0.813	0.407	0.874	0.475	0.753	0.334	0.793	0.470	0.858	0.376	0.795
1995	0.860	0.956	0.717	0.669	0.479	0.920	0.269	0.803	0.388	0.843	0.496	0.761	0.341	0.798	0.493	0.855	0.384	0.818
2000	0.859	0.960	0.721	0.664	0.488	0.902	0.292	0.714	0.379	0.832	0.510	0.757	0.346	0.803	0.513	0.849	0.394	0.820
2003	0.859	0.961	0.724	0.663	0.497	0.892	0.301	0.689	0.384	0.833	0.521	0.757	0.354	0.808	0.523	0.849	0.399	0.830
2006	0.866	0.961	0.726	0.664	0.504	0.883	0.311	0.673	0.385	0.828	0.529	0.760	0.360	0.816	0.530	0.848	0.405	0.841

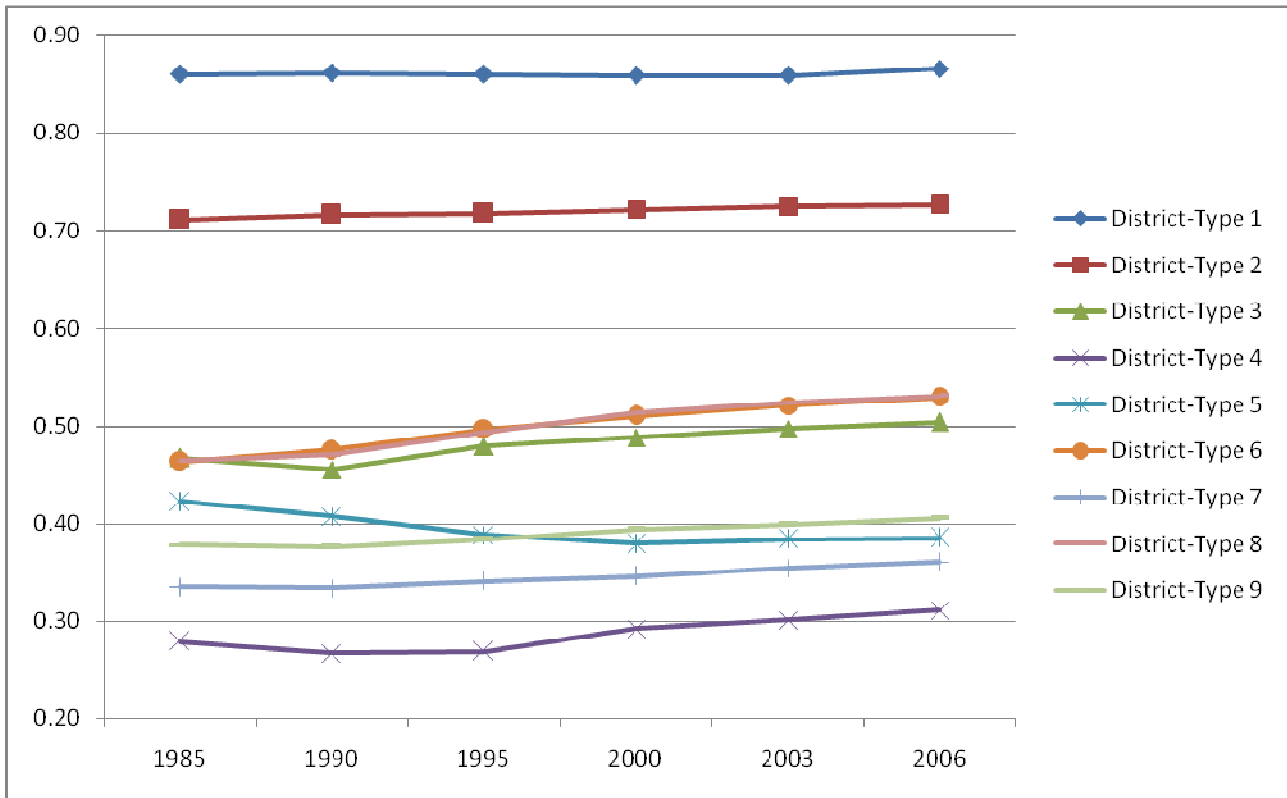


Figure 1: Rank-size Coefficients across Nine Types of Districts in Germany (Table 4).
Legend: x-axis: years (1985-2006); y-axis (q -coefficients).

These results confirm the argument that an increase in the q -coefficient is associated with the economic development of the country/region (see, among others, Berry, 1961; Parr, 1985). It is worth noting that specific types of district (namely, District-types 1 and 2) with higher values of the q -coefficient also have a higher GDP. In addition, Table 1 and Figure 1 show that the q -coefficients increase slightly for every type of district over the years considered (except for the District-type 5). This result is also in line with the rise in economic development in the German regions.

It should be noted that, in this particular typology of areas, the values of the q -coefficients show, in general, values either around 0.8 or around (or lower than) 0.5. This means that District-type 1 and District-type 2, with values of q around 0.7-0.8, indicate a tendency towards urban heterogeneity, and thus towards hierarchical connectivity structures (according to Table 1), while the remaining seven types of districts (District-types 3-9), with values of $q < 0.5$, display rather homogeneous patterns ('club convergence') – which correspond to a random network from a connectivity viewpoint (Table 1). Interestingly, District-types 6 and 8, which show values of around 0.5 in the last six years of the data set, are classified as 'highly urbanized districts' and 'urbanized districts', while District-types 4, 7 and 9, which display very low values of around 0.3, are classified as 'rural districts' (see Table A1). Consequently the values of the q -coefficient – and of its counterpart, the γ -coefficient – seem to perfectly match the classification of the nine typologies of districts.

4.3 Rank-Size Rule Embedding a Power Form: Analysis of German Districts

Next, we apply formula (2) at district level (439 German districts) for the years 1985-2007. Table 5 illustrates the results, where q represents the rank-size coefficient for a particular year. The rank-size regressions of type (2) provide the following q -estimates for total population: a value between 0.60 and 0.63, with an R^2 of about 0.92. The values of the q -coefficients show again a slight increase over the years for the total population. We may thus infer that at district level the related population has rather homogeneous characteristics; however, this pattern is oriented towards urban heterogeneity and hierarchical connectivity structures (Table 1).

4.4 Rank-Size Rule Embedding a Power Form: Analysis of German Cities

We will now apply formula (2) at the city level in Germany. Our study is concerned with the number of inhabitants for about 2,100 cities (Staedte) in Germany. It covers the time period from 1993 to 2007. The results are again given in Table 5. Interestingly, the emerging q -values at the city level, although close to 1, slightly increase from the year 2000 to 2007, thus reflecting an increase in economic development, probably caused by a rise in physical and virtual interurban connectivity. In general, this sample size follows Zipf's law, which states that the value of the q -coefficient is equal or close to 1, thus showing a perfect hub-and-spoke system from a network connectivity viewpoint (Table 1). The R^2 values are rather stable over time, and are always rather high (with an average R^2 of about 0.89). It is worth noting that, by considering a truncated sample of this city data set, Zipf's law no longer holds (see later Subsection 4.9.2).

4.5 Rank-Size Rule Embedding a Power Form: Analysis of German Municipalities

The data set for the study of the rank-size q -coefficients at municipality level in Germany contains the number of inhabitants for about 12,000 municipalities, covering the years 1993-2007. Table 5 shows the results of formula (2) applied at municipality level. In particular, we observe that the values of the q -coefficients are about 1.4 for each year, with R^2 very close to 0.87. Values of the q -coefficient greater than 1 imply that at the municipality level the size increases more than proportionally, while at the district level it increases less than proportionally. This means that at the municipality level there are great inequalities between the units, with an increase of the value of the hubs seen from a network connectivity perspective (Table 1).

Table 5 shows that – analogously to the city level – the values of the q -coefficient slightly decrease until 2000, and then slightly increase from 2000 to 2007. Analogously to the district case, we can then also interpret q as an indicator of economic development for the municipality level. In summary, from our experiments, we may interpret the potential of the q -coefficient as an (aggregate) indicator of urban heterogeneity and economic growth, as well of the associated socio-spatial networks, by means of its 'counterpart', the γ -coefficient (see equation (15) and Table 1). In this perspective, in the absence of flow data, population P_j appears to be a fundamental variable which is able to proxy urban patterns characteristics, as well the linked socio-spatial networks. Clearly, any other variable could be used in order to explore the related patterns and networks (GDP, incoming flows, etc.).

Table 5: Rank-Size Rule Distributions Embedding Different Functional Forms: Analysis of Various Spatial Population Levels in Germany

	District Level								
	Power		Exponential		Lognormal		Tanner		
	q	R^2	β	R^2	λ	R^2	q	β	R^2
1985	0.6042	0.924	0.00453	0.882	0.0707	0.939	0.3749	0.00201	0.963
1990	0.6099	0.926	0.00456	0.879	0.0714	0.941	0.3849	0.00197	0.963
1995	0.6139	0.923	0.00461	0.887	0.0719	0.942	0.3734	0.00211	0.965
2000	0.6181	0.917	0.00469	0.895	0.0726	0.940	0.3579	0.00227	0.965
2003	0.6247	0.916	0.00475	0.899	0.0735	0.941	0.3546	0.00236	0.966
2006	0.6321	0.916	0.00481	0.902	0.0744	0.942	0.3551	0.00242	0.968

	City Level								
	Power		Exponential		Lognormal		Tanner		
	q	R^2	β	R^2	λ	R^2	q	β	R^2
1993	1.0488	0.903	0.00177	0.918	0.0929	0.939	0.5302	0.00099	0.974
1995	1.0540	0.901	0.00177	0.919	0.0933	0.938	0.5275	0.00101	0.973
2000	1.0511	0.896	0.00178	0.921	0.0932	0.935	0.5103	0.00103	0.972
2003	1.0613	0.893	0.00179	0.923	0.0866	0.814	0.5040	0.00106	0.972
2006	1.0616	0.892	0.00179	0.923	0.0848	0.784	0.5034	0.00106	0.971
2007	1.0658	0.892	0.00180	0.923	0.0869	0.811	0.5027	0.00107	0.971

	Municipality Level								
	Power		Exponential		Lognormal		Tanner		
	q	R^2	β	R^2	λ	R^2	q	β	R^2
1993	1.4024	0.881	0.00041	0.947	0.0961	0.929	0.5699	0.0003	0.983
1995	1.3994	0.878	0.00041	0.947	0.0959	0.927	0.5617	0.0003	0.982
2000	1.3935	0.875	0.00041	0.948	0.0955	0.924	0.5456	0.0003	0.981
2003	1.3961	0.875	0.00041	0.949	0.0960	0.924	0.5434	0.0003	0.982
2006	1.3984	0.872	0.00041	0.950	0.0961	0.924	0.5401	0.0003	0.983
2007	1.4045	0.873	0.00041	0.950	0.0963	0.923	0.5401	0.0003	0.982

	City Network Level								
	Power		Exponential		Lognormal		Tanner		
	q	R^2	β	R^2	λ	R^2	q	β	R^2
1985	0.7388	0.958	0.1099	0.844	0.2207	0.887	0.6944	0.00748	0.953
1990	0.7661	0.964	0.1131	0.836	0.2273	0.879	0.7612	0.00081	0.959
1995	0.7807	0.963	0.1147	0.826	0.2309	0.873	0.8026	0.00368	0.958
2000	0.7788	0.967	0.1139	0.821	0.2298	0.872	0.8264	0.00801	0.963
2003	0.7778	0.968	0.1134	0.816	0.2292	0.870	0.8419	0.01079	0.964
2006	0.7785	0.967	0.1136	0.818	0.2298	0.872	0.8349	0.00949	0.963

4.6 Rank-Size Rule Embedding a Power Form: Analysis of the German City Network

The level of analysis in our study refers to the ‘city network’ identified by Reggiani et al. (2011b), on the basis of transport and related economic background factors. Our selected city network covers a

system of 17 German core districts/cities¹⁰ that have been chosen from the 439 Germany districts, according to the following criteria (for each district): (a) high accessibility; (b) high connectivity with speed railway network; (c) relevance for the German economy.

In this particular case, we found a q -coefficient of 0.7, indicating that there is not a ‘perfect’ urban hierarchy in this network constellation (see again Table 5). From a connectivity viewpoint, the γ -coefficient indicates the presence of an SF structure (Table 1). This is plausible, since the districts under analysis, being also the highest ranked ones from an economic viewpoint, are certainly the best interconnected ones from a transport viewpoint,

4.7 Ways Forward

Our analysis has brought to light interesting results on rank-size rules. If we consider the case where the rank-size distribution embeds the power form in the context of different spatial levels of population in Germany, we can report the following important findings: the q -coefficient is, in general, greater than 1 for municipalities, and less than 1 for districts, the city-network and the nine main administrative-economic regions. The q -coefficient is equal to 1 for both the total number of cities, and for a truncated sample (more than half the sample size) of the municipalities. Thus, there is scale and size sensitivity, as was also outlined by others (see, e.g., Guerin-Pace, 1995; Soo, 2005).

In conclusion, our analysis has clearly demonstrated that the general confirmation of the validity of Zipf’s law for the German spatial system, as claimed by Giesen and Suedekum (2011), holds exclusively – and still under specific conditions – for some samples of the municipality level (such as the total city data subset, and a large sample of the municipality data set). Other spatial scales lead to totally different values of the rank-size coefficient, and thus the validity of the findings of Giesen and Suedekum seem to be confined to particular cases. This is in accordance with Brakman et al. (2001, p. 203) who state: “*so an important issue in comparing studies on Zipf’s law is the choice of the sample size*”. These authors mention two strategies which might be used to overcome this problem: a) to use a fixed number of cities; and b) to exclude cities below a certain size threshold. We will now apply both these strategies to our German data set.

Concerning the first approach, we reduced the number of districts and municipalities and analysed the remaining smaller sample size for both these spatial categories. Concerning the second approach, we reduced the sample size by excluding the cities which have a number of inhabitants below a given threshold (see Section 4.9). Before discussing this last step of our analysis, we briefly discuss in the subsequent Section 4.8 the results which emerge – at each spatial level – by considering various rank-size functional forms that are different from the power function.

4.8 Rank-Size Rule Embedding Alternative Functional Forms: Analysis of Different Spatial Levels in Germany

Up to now we have examined the results which emerge from the adoption of a rank-size rule embedding a standard power form, with reference to different spatial population levels in Germany. However, as indicated in Subsection 2.2.3, alternative specifications for city-size distributions can be

¹⁰ This selected city network covers the following 17 interconnected districts in Germany: Berlin, Bonn, Bremen, Dortmund, Dresden, Düsseldorf, Essen, Frankfurt, Hamburg, Hannover, Karlsruhe, Köln, Leipzig, Mannheim, München, Nürnberg and Stuttgart (see for details Reggiani et al., 2011b).

adopted and also justified from a theoretical viewpoint. Along these lines, we adopted here rank-size distributions embedding alternative functions, in particular: the exponential form (12), the lognormal form (13), and the Tanner form (14).

Table 5 shows that each of these functions appears to fit the different urban structures in Germany rather well, with high R^2 values. In addition, we observe that the Tanner function very often performs better. However, since the Tanner function embeds both the exponential and the power form, an F-test should be carried. In general, the lognormal performs better than the exponential and power forms, at least at the district, city and municipality level. However, this does not hold for the city-network level, but this is a particular case of only 17 districts.

Our empirical results do not clearly discriminate between various functional forms, and hence we cannot infer that one function is superior to another. In summary, other functional forms can be equally suitable to represent the rank-size distribution of population, as was also argued in Section 3. A remaining issue is still the role of the sample size in identifying the particular value of the coefficient in Zipf's law, i.e. the value of $q=1$ in the rank-size formulation of type (1). This issue will be further explored in Section 4.9.

4.9 Sensitivity Analysis: The Rank-Size Coefficient vs. Sample Size

4.9.1 Sensitivity Analysis with Reference to the German Districts

In this subsection we examine how the q -coefficients (emerging from the rank-size distributions embedding power forms) can vary with respect to different sample sizes, at different spatial levels of the German data set.

The first step in our sensitivity analysis concerns an investigation of the population patterns of the 439 German districts, by varying the sample size. As indicated in the previous sections, our q -coefficient is rather stable over the years. We present here the results for the year 2006, for the sake of comparison with the years concerning the other spatial scale levels. In our procedure we start from the highest ranked districts, and we gradually increase the number of districts, by gradually reducing the minimum number of inhabitants per district (see Table 6). These results highlight that the value of the q -coefficient never gets closer to 1 if the number of the districts is reduced. In fact, decreasing from 439 to 100 districts, the value of the q -coefficient decreases from 0.63 to 0.48. Subsequently, with less than 100 districts, the value increases to 0.51.

Table 6: Estimated Rank-Size Coefficients (q) vs. Sample Size; German Districts; Year 2006

Number of districts (by highest rank)	Minimum no. of inhabitants per district	q	R^2	SE
25	444,700	0.5135	0.916	0.0346
50	308,331	0.5135	0.948	0.0171
100	224,347	0.4819	0.969	0.0086
300	108,364	0.5345	0.985	0.0038
439	--	0.6321	0.916	0.0091

Note: 439 districts; min. inhabitants for a district: 34,842; max. inhabitants for a district (Berlin): 3,404,037.

From our empirical findings we can infer, therefore, that, at the German district level, a rather homogeneous patterns of population exists – and consequently a random network from a connectivity viewpoint – especially for a reduced number of districts (see also Table 1).

4.9.2 Sensitivity Analysis with Reference to the German Cities

The second step of our sensitivity analysis concerns the investigation of the population patterns of German cities, for different sample sizes, for the year 2006. Table 7 shows the related results. Here we can see that a value of the q -coefficient equal to 1, which occurs when considering the total number of the cities (Section 4.4.), is no longer reached when we reduce the sample size.

Table 7: Estimated Rank-Size Coefficients (q) vs. Sample Size; German cities; Year 2006

Number of cities (by highest rank)	Minimum no. of inhabitants per city	q	R^2	SE
50	159,060	0.7333	0.982	0.0144
100	82,797	0.7850	0.989	0.0085
500	25,084	0.7741	0.997	0.0021
1000	12,583	0.7923	0.996	0.0015
1500	6084	0.8731	0.978	0.0034
2075	0	1.0616	0.892	0.0081

Note: 2,075 cities; min. inhabitants for a city: 363; max. inhabitants for a city (Berlin): 3,404,037.

This specific result does not confirm the general argument in the literature that a q -coefficient equal to 1 seems to appear in truncated samples where small cities are not considered (see, e.g. Giesen and Suedekum, 2011).

4.9.3 Sensitivity Analysis with Reference to the German Municipalities

The third step of our sensitivity analysis concerns the investigation of the population patterns of the German municipalities, for different sample sizes, again for the year 2006. Table 8 shows different values of the q -coefficient with respect to different sample sizes for the German municipalities.

The values of the q -coefficient appear to vary between 0.73 and 1.40. Between 25 and 1000 municipalities, the values of the q -coefficient are rather stable with values around 0.73-0.76. The values of the rank-size coefficients increase from 0.76 (with 1000 municipalities) to 1.40 (when we consider the total sample size). In general, we find a disparity between all municipalities, which means that there are many small municipalities and a few large municipalities (see Table 1). Interestingly, the q -coefficient approaches 1 around 6200-7500 municipalities (more than half the sample size). Figure B1 in Annex B illustrates this point empirically, with reference to the case of 6500 municipalities. In particular, if we exclude the smaller municipalities (with less than 1,447 inhabitants per municipality), we can see a q -value of $0.999 \approx 1$, thus indicating the ‘perfect’ Zipf’s hierarchy among them, which corresponds to a hub-and-spoke system from a network connectivity viewpoint (Table 1). All in all, these results concerning the municipality data set confirm the previous findings concerning the German city data set; in other words, Zipf’s law is obeyed both for the total number of cities (2059) and for a consistent number of municipalities (6200-7500), i.e. for specific subsets of German municipalities.

Table 8: Estimated Rank-Size Coefficients (q) vs. Sample Size; German Municipalities; Year 2006

Number of municipalities (by highest rank)	Minimum no. of inhabitants per municipality	q	R^2	SE
25	264,765	0.7370	0.960	0.0308
50	161,345	0.7333	0.982	0.0144
100	83,028	0.7839	0.989	0.0084
500	25,363	0.7741	0.996	0.0021
1000	14,632	0.7618	0.998	0.0011
5000	2208	0.9359	0.977	0.0020
6200	1447	0.9989	0.968	0.0023
6500	1320	1.0152	0.965	0.0024
7500	976	1.0690	0.957	0.0026
12293	0	1.4018	0.874	0.0048

Note: 12,293 (observed) municipalities; min. inhabitants for a municipality: 7; max. inhabitants for a municipality (Berlin): 3,404,037.

4.9.4 Sensitivity Analysis: Summary

In this final step, we fix different thresholds with respect to the minimum number of inhabitants for the population at the district and the municipality level. Table 9 shows the estimation of the q -coefficient for these two spatial levels (district level and municipal level) for the year 2006. According to the results of Table 5, we can see in Table 9 that, at the district level, the q -coefficient never reaches the value of 1, while, at the municipality level, the q -coefficient starts to approach 1 when considering a threshold greater than 1000 inhabitants. It should be noted that several authors argue that Zipf's law requires data truncation, with a cut-off below 100,000 inhabitants (Giesen and Suedekum, 2011). Our results do not seem to confirm this standard rule. By considering more than 100,000 inhabitants in both cases (districts and municipalities), the q -coefficient becomes much less than 1. This finding is illustrated in Figure B2 in Annex B.

Table 9: Estimated Rank-size Coefficients (q) for Different Spatial Levels (District Level (Year 2006) vs. Municipality Level (Year 2006))

Inhabitants for city	Districts		Municipalities	
	q	no. of districts	q	no. of municip.
> 500,000	0.60712	20	0.72315	14
> 250,000	0.48599	86	0.73539	27
> 100,000	0.54045	321	0.72553	41
> 50,000	0.60542	422	0.80614	188
> 25,000	0.63211	439	0.77357	510
> 10,000		439	0.77033	1566
> 5000		439	0.82687	2874
> 1000		439	1.06510	7427
> 500		439	1.18661	9666
> 5		439	1.40179	12,293

In summary, a Zipf's coefficient equal to 1 is rarely reached in the above experiments; if it happens, it depends on the specific sample size. Mostly, it does not happen in the upper tail of our data set, but only in the presence of a population sample which includes those municipalities with more than 1,000 inhabitants. From the connectivity network perspective, with reference to the cities and municipalities, in general the q -coefficient varies between 0.7 and 1.4, thus confirming a certain presence of hub-hierarchies in these two cases (see the counterpart value of the γ -coefficient in Table 1). In the case of districts, the lower q -value seems to reflect more homogeneous patterns in terms of the districts' population size, and thus more a random structure from a network connectivity viewpoint.

5 Conclusions

In this paper the rank-size rule concept has been critically reviewed and analysed from two complementary perspectives: spatial economics and network analysis. In the spatial-economics context, our theoretical and methodological reflections have demonstrated the following:

- the rank-size rule, and hence, Zipf's law, can be derived from an entropy maximization (or generalized cost minimization) approach and finds thus its roots in welfare theory;
- in the analytical framework outlined above, different specifications for the rank-size distribution may emerge, such as exponential, lognormal or Tanner functions. These alternative forms are plausible and can be justified from a theoretical perspective.

From a network analysis perspective, we have shown that:

- a connectivity interpretation of the rank-size coefficient q (by considering a conventional rank-size expression with the power form) is possible, so that a formal correspondence between the rank-size rule and network connectivity analysis does exist.

As a consequence, the rank-size coefficient can be conceived of in a new spatial-economic framework: on the one hand, as a shadow price for the transaction costs involved in moving from a lower rank towards a higher rank (corresponding to an entropy framework); and, on the other hand, as an indicator for the type of network connectivity structure underlying the interaction between the cities/population units concerned. Thus the rank R and its associated coefficient q may be interpreted as (shadow) costs associated with the existence of a functional-theoretical spatial economic equilibrium.

In summary, the rank-size concept can be conceived as a 'universal' expression of many models traditionally used in spatial economics. It is thus able to: (i) grasp the homogeneity/heterogeneity of the spatial system concerned (at an aggregate level, by means of its rank-size coefficient); and (ii) represent the associated connectivity infrastructure or socio-economic network (or socio-economic constellation) of complex spaces.

It is interesting that Zipf certainly anticipated not only the conventional models adopted in spatial economics, such as spatial interaction models/entropy models, but also the associated socio-economic spatial networks. In addition, since spatial interaction/entropy models are compatible with discrete choice models – and the related micro-economic foundations – the rank-size rule can be also interpreted from this micro-economic perspective. In this framework, the value of the q -coefficient emerging from the rank-size rule – as well as its counterpart, i.e. the γ -coefficient in network analysis – acquires the important connotation of a proxy indicator for the spatial economic development of the region/country under analysis (in case where the associated variable is the population), as conjectured by several authors. As a consequence, it seems that for high values of the rank size coefficient, i.e. in the presence

of increasing economic development, higher urban inequalities emerge, linked to relevant hub networks of an SF structure.

To test various propositions mentioned above, a series of empirical experiments was carried out – with reference to the population of urban centres in Germany – in order to explore the variations of the rank-size coefficients for different spatial scales and population sizes, and for different years. These experiments show – by means of the various rank-size coefficient indicators – a trend towards heterogeneity in Germany’s population network. In particular, the case of the q -coefficient = 1 – which indicates a ‘perfect hierarchy’ between urban centres (i.e. Zipf’s law) – is reached at the city level (by considering the total number of cities) as well as at the municipality level (by excluding the municipalities around 1000-1500 inhabitants). This empirical result on German urban system highlights the economic rationale underlying the value of the q -coefficient equal to 1 (Zipf’s Law). Even though Zipf’s law is only valid for specific subsets of the German municipalities (i.e. the number of municipalities with more than 1000 inhabitants, or alternatively, the total number of cities – which is again a subset of municipalities), the occurrence of this ‘perfect equilibrium’ (Zipf’s law) for these sample cases is remarkable. Clearly, the next step in rank-size research shall be devoted to the process of urban growth underlying this ‘spatial-economic’ equilibrium. In general, we can conjecture a Lotka-Volterra mechanism between the city-populations, since this is strictly related to a dynamic entropy maximization and thus to the general formulation (10) of the rank size-rule (Nijkamp and Reggiani, 1992). Starting from the Lotka-Volterra framework, the underlying dynamic conditions leading to the value of the q -coefficient equal to 1 deserve a more thorough investigation, also in the light of Gibrat’s and Gabaix’s arguments (see, e.g., Gabaix, 1999; Lotti et al., 2009).

Our analysis thus prompts various new research questions. Future research directions should address, from a theoretical viewpoint, the dynamics of the rank-size rule, for instance, by focussing on Gibrat’s law and the related spatial-economic feedback effects. From an empirical viewpoint, more sophisticated estimation procedures may be taken into account, as well as the possibility of examining the rank-size rule at a more micro-level (e.g., firms, etc). Certainly, the city-size distribution reflects many aspects of complex network evolution in interurban interdependencies, such as the emergence of new hubs or new city-networks (with the related decay of ‘old’ hubs), which deserve further attention, also in the light of network resilience issues. More comparative studies in an international context would also be needed in order to test the robustness of our findings.

Acknowledgments

The authors wish to thank Uwe Blien and Anette Haas (IAB, Germany), as well as Kristian Giesen (University of Duisburg-Essen), for kindly providing several data used in our study. In addition, Giovanni Russo (Cedefop, Greece), Franz-Josef Bade (University of Dortmund, Germany), Clifford Wymer (LSE, London, UK) and Kieran Donaghy (Cornell University, Ithaca, USA) are gratefully acknowledged for their valuable suggestions. The authors wish to thank in particular Alfredo Nozzi (University of Bologna, Italy) for his great assistance in carrying out the statistical analysis.

References

- Albert, R. and A. L. Barabási. 2000. Topology of Evolving Networks: Local Events and Universality, *Physical Review Letters*, 85(24), 5234-5237.
- Albert, R. and A. L. Barabási. 2002. Statistical Mechanics of Complex Networks, *Reviews of Modern Physics*, 74, 47-97.
- Adamic, L. A. 2000. *Zipf, Power-Laws, and Pareto – A Ranking Tutorial*, Retrieved 16 April 2007, from www.hpl.hp.com
- Auerbach, F. 1913. Das Gesetz der Bevölkerungskonzentration, *Petermanns Geographische Mitteilungen*, 59, 74-76.
- Bak P. 1996. *How Nature Works: The Science of Self-Organized Criticality*, Copernicus, New York.
- Barabási, A. L. 2002. *Linked: the New Science of Network*, Perseus Publishing, Cambridge Massachusetts.
- Barabási, L.A. and R. Albert. 1999. Emerging of Scaling in Random Networks, *Science*, 286, 509-512.
- Barabási, L.A. and R. Albert. 2000. Topology of Evolving Networks: Local Events and Universality. *Physical Review Letters*, 74, 47-97.
- Barabási, L.A. and Z. N. Oltvai. 2004. Network's Biology: Understanding the Cell's Functional Organization. *Nature Reviews-Genetics*, 5, 101-113.
- Batty, M. 2005. *Cities and Complexity: Understanding Cities with Cellular Automata, Agent-based Models, and Fractals*. MIT Press, Cambridge.
- Benguigui, L. and E. Blumenfeld-Lieberthal. 2010. The End of a Paradigm: Is Zipf's Law Universal?, *Journal of Geographical Systems*, 13(1), 87-100.
- Berliant, M. and H. Watanabe. 2008. A Scale-Free Network Structure Explains the City-Size Distribution, Paper Presented at the NARSC Conference, NY, 19-22 November.
- Berry, B. 1961. City Size Distributions and Economic Development, *Economic Development and Cultural Change*, 9 (4), 573-588.
- Berry, B.T.L. and W.L. Garrison. 1958. Alternate Explanations of Urban-Rank-Size Relationships, *Annals of the Association of American Geographers*, 4(1), 83-91.
- Bos, H.C. 1964. *The Spatial Dispersion of Economic Activity*, Rotterdam University Press, Rotterdam.
- Brakman, S., H. Garretsen and C. van Marrewijk. 2001. *An Introduction to Geographical Economics. Trade, Location and Growth*, Cambridge University Press, Cambridge.
- Brakman, S., H. Garretsen, M. van den Berg. and C. van Marrewijk. 1999. The Return of Zipf: Towards a Further Understanding of the Rank-Size Distribution, *Journal of Regional Science*, 39, 182-213.
- Buchanan, M. 2002. *Nexus: Small Worlds and the New Science of Networks*, W.W. Norton and Co., N.Y.
- Clemente, J., R. González-Val and L. Olloqui, 2011. Zipf's and Gibrat's Laws for Migrations, *The Annals of Regional Science*, 47(1), 235-248.
- Corominas-Murtra B., and R. V. Solé. 2010. Universality of Zipf's Law, *Physical Review E*, 82, 9-17.
- Corominas-Murtra B., J. F. Andreu and R. V. Solé. 2011. Emergence of Zipf's Law in the Evolution of Communication, *Physical Review E*, 83, 32767.
- Duranton, G. 2002. *City-size Distribution as a Consequence of the Growth Process*, London School of Economics and Political Science, Centre of Economic Performance, London.
- Eeckhout, J. 2004. Gibrat's Law for (All) Cities, *American Economic Review*, 94, 1429-1451.
- Einstein, A. 1905. Ist die Trägheit eines Körpers von seinem Energieinhalt abhängig? (Does the Inertia of a Body Depend upon its Energy-Content?), *Annalen der Physik*, 18, 639-643.
- Fotheringham A.S. and M.E. O'Kelly. 1989. *Spatial Interaction Models: Formulations and Applications*, Kluwer Academic, Dordrecht.
- Gabaix, X. 1999. Zipf's Law for Cities: An Explanation, *Quarterly Journal of Economics*, 114(3), 739-767.
- Gabaix, X., and Y.M. Ioannides. 2004. The Evolution of City Size Distributions, in J.V. Henderson and J.-F. Thisse (eds). *Handbook of Regional and Urban Economics*, Vol. 4. Elsevier, Amsterdam, 2341-2378.
- Gabaix, X. and R. Ibragimov, 2011. Rank - $\frac{1}{2}$: A Simple Way to Improve the OLS Estimation of Tail Exponents, *Journal of Business & Economic Statistics*, 29 (1), 24-39.
- Gan, L., D Li, and S. Song. 2006. Is the Zipf's Law Spurious in Explaining City-Size Distributions? *Economic Letters*, 92, 256-262.
- Gibrat, R. 1931. *Les Inégalités Economiques*, Librairie du Recueil Sirey, Paris.

- Giesen, K. and J. Suedekum. 2011. Zipf's Law for Cities in the Regions and the Country, *Journal of Economic Geography*, 11(4), 667-686.
- Giesen, K., A. Zimmermann and J. Suedekum. 2010. The Size Distribution across all Cities – Double Pareto Lognormal Strikes, *Journal of Urban Economics*, 68, 129-137.
- Glaeser, E. L., J. A. Scheinkman and A. Shleifer. 1995. Economic Growth in a Cross-Section of Cities, *Journal of Monetary Economics*, 36, 117-143.
- Gordon, I. 2010. Entropy, Variety, Economics, and Spatial Interaction, *Geographical Analysis*, 42, 446-471.
- Guerin-Pace, F. 1995. Rank-Size Distribution and the Process of Urban Growth, *Urban Studies*, 32, 551-562.
- Hill, B.M. 1975. A Simple General Approach to Inference About the Tail of A Distribution, *Annals of Statistics*, 3, 1163-1174.
- Hinloopen, J. and C. van Marrewijk. 2006. *Comparative Advantage, the Rank-size Rule, and Zipf's Law*, Tinbergen Institute Discussion Papers, 06-100/1, Tinbergen Institute, Amsterdam.
- Ioannides, Y. M. and H. G. Overman. 2004. Spatial Evolution of the US Urban System, *Journal of Economic Geography*, 4(2), 131-156.
- Ioannides, Y. M. and H. G. Overman. 2003. Zipf's Law for Cities: An Empirical Examination, *Regional Science and Urban Economics*, 33(2), 127-137.
- Isard, W. 1956. *Location and Space-Economy*, MIT Press, Cambridge, Mass.
- Isard, W. 1971. On Relativity Theory and Time-Space Models, *Papers of the Regional Science Association*, 26, 7-24.
- Krugman, P. 1996a. Urban Concentration: The Role of Increasing Returns and Transport Costs, *International Science Review*, 19(1&2), 5-30.
- Krugman, P. 1996b. *The Self-organizing Economy*, Blackwell Publisher, Stanford University.
- Levy, M. 2009. Gibrat's Law for (All) Cities: Comment, *American Economic Review*, 99, 1672-1675.
- Lotti, F., E. Santarelli and M. Vivarelli. 2009. Defending Gibrat's Law as a Long-run Regularity, *Small Business Economics*, 32, 31-44.
- Mansury, Y. 2010. *Size Distribution in City Systems with Heterogeneous Agents: A Discrete Markov-Chain Approach*, Cornell University, Ithaca.
- Mansury, Y. and L. Gulyas. 2007. The Emergence of Zipf's Law in a System of Cities: An Agent-based Simulation Approach, *Journal of Economic Dynamics and Control*, 31, 2438-2460.
- Matlaba, V.J., M.J. Holmes, P. McCann and J. Poot. 2011. *A Century of the Evolution of the Urban System in Brazil*, Department of Economics, Working Paper Series, Number 12/11, University of Waikato, Hamilton, New Zealand.
- Mulder, P. and H.L.F. de Groot. 2008. *The Economics of Zipf's Law. A Meta-Analysis*, Paper presented at the ERSA Conference, Lodz, 25-29 August.
- Nijkamp, P. 1975. Reflections on Gravity and Entropy Models, *Regional Science and Urban Economics*, 5, 203-225. Reprinted in A. Reggiani, K. Button and P. Nijkamp (eds) (2006), *Planning Models*, Edward Elgar, Cheltenham, Chapter 7.
- Nijkamp, P. and A. Reggiani. 1992. *Interaction, Evolution and Chaos in Space*, Springer-Verlag, Berlin.
- Nitsch, V. 2005. Zipf Zipped, *Journal of Urban Economics*, 57, 86-100.
- Nota, F. and S. Song. 2007. *Further Analysis of the Zipf's Law: Does the Rank-Size Rule Really Exist?*, Department of Economics, University of Nevada, Reno.
- Parr, J. B. 1985. A Note on the Size Distribution of Cities Over Time. *Journal of Urban Economics*, 18, 199-212.
- Pumain, D., and F. Moriconi-Ebrard. 1997. City Size Distributions and Metropolisation, *Geojournal*, 43(4), 307-314.
- Reggiani, A. 2009. From Complexity to Simplicity: Interdisciplinary Synthesis and Future Perspectives, in A. Reggiani and P. Nijkamp (eds). *Complexity and Spatial Networks*, Springer-Verlag, Berlin, 275-284.
- Reggiani, A. 2012. Complexity and Spatial Networks, in M. Fischer and P. Nijkamp (eds). *Handbook of Regional Science*, Springer-Verlag, Berlin (forthcoming).
- Reggiani A. and P. Nijkamp (eds). 2006. *Spatial Dynamics, Networks and Modelling*, Edward Elgar, Cheltenham, UK.
- Reggiani, A. and P. Nijkamp (eds). 2009. *Complexity and Spatial Networks*, Springer-Verlag, Berlin.
- Reggiani, A., P. Bucci and G. Russo. 2011a. Accessibility and Impedance Forms: Empirical Applications to the German Commuting Networks, *International Regional Science Review*, 34(2), 230-252.

- Reggiani, A., P. Bucci, G. Russo, A. Haas and P. Nijkamp. 2011b. Regional Labour Markets and Job Accessibility in City Network Systems in Germany, *Journal of Transport Geography*, 34, 230-252.
- Richardson H.W., 1969, *Elements of Regional Economics*, Penguin Books, Harmondsworth
- Rosen, K. T., and M. Resnick. 1980. The Size Distribution of Cities: An Examination of the Pareto Law and Primacy, *Journal of Urban Economics*, 8(2), 165-186.
- Simon, H. 1955. On a Class of Skew Distribution Functions, *Biometrika*, 42(3/4), 425-440.
- Soo, K. 2005. Zipf's Law for Cities: A Cross-Country Investigation, *Regional Science and Urban Economics*, 35, 239-263.
- Terra, S. 2009. Zipf's Law for Cities: On a New Testing Procedure, CERDI-CNRS, <http://publi.cerdi.org/ed/2009/2009.20.pdf>
- Tinbergen, J. 1964. Sur un Modèle de la Dispersion Géographique de l'Activité Economique, *Revue d'Economie Politique*, Special Issue, pp.30-44.
- Vega-Redondo, F. 2007. *Complex Social Networks*, Cambridge University Press, Cambridge.
- Weibull, J.W. 1980. On the Numerical Measurement of Accessibility, *Environment Planning A*, 12, 53-67.
- Willigers, J., H. Floor and B. van Wee. 2007. Accessibility Indicators for Location Choices of Offices: an Application to the Intraregional Distributive Effects of High-Speed Rail in the Netherlands, *Environment and Planning A*, 39, 2086-2098.
- Wilson, A. G. 1967. A Statistical Theory of Spatial Distribution Models, *Transportation Research* 1, 253-269.
- Wilson, A. G. 1970. *Entropy in Urban and Regional Modelling*, Pion, London.
- Wilson, A. G. 2009. The 'Thermodynamics' of the City: Evolution and Complexity Science in Urban Modelling, in A. Reggiani and P. Nijkamp (eds). *Complexity and Spatial Networks*, Springer-Verlag, Berlin, 11-31.
- Yook, S., H. Jeong and A. Barabási. 2002. Modeling the Internet's Large-Scale Topology, *Proceedings of the National Academy of Sciences*, 99, 13382-13386.
- Zipf, G. 1949. *Human Behavior and the Principle of Least Effort: An Introduction to Human Ecology*, Cambridge, Addison-Wesley.

Annex A. Classification of German Spatial Districts

Table A1: Nine Types of Districts in Germany

Main Types of Districts	Type of Districts	Number of Districts
A. Districts with urban agglomeration	1. Central cities	43
	2. Highly urbanized districts	45
	3. Urbanized districts	32
	4. Rural districts	27
B. Districts with tendencies towards agglomeration	5. Central cities	29
	6. Highly urbanized districts	81
	7. Rural districts	57
C. Districts with rural features	8. Urbanized districts	57
	9. Rural districts	68

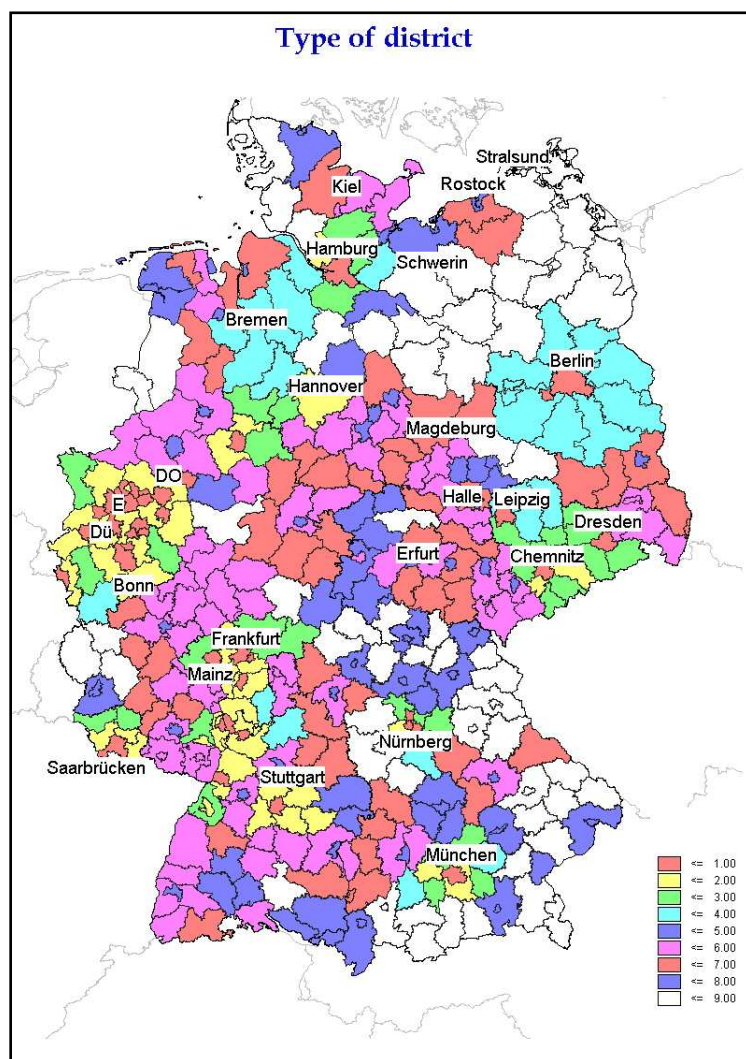


Figure A1: Visualization of the Nine Types of German Districts

Annex B. Zipf's Law and Rank-size Empirics in Germany (Truncated Population Samples)

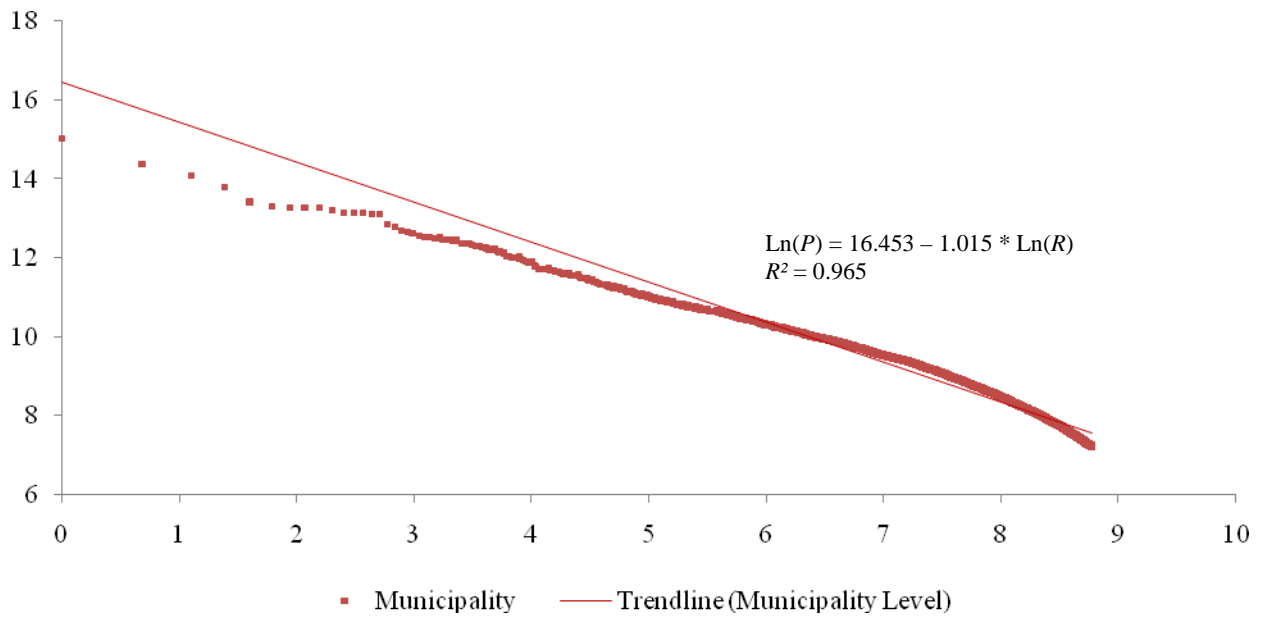


Figure B1: Zipf's Law across 6500 Municipalities (Reference to Table 8).

Legend: x-axis: $\text{Ln}(R)$; y-axis $\text{Ln}(P)$

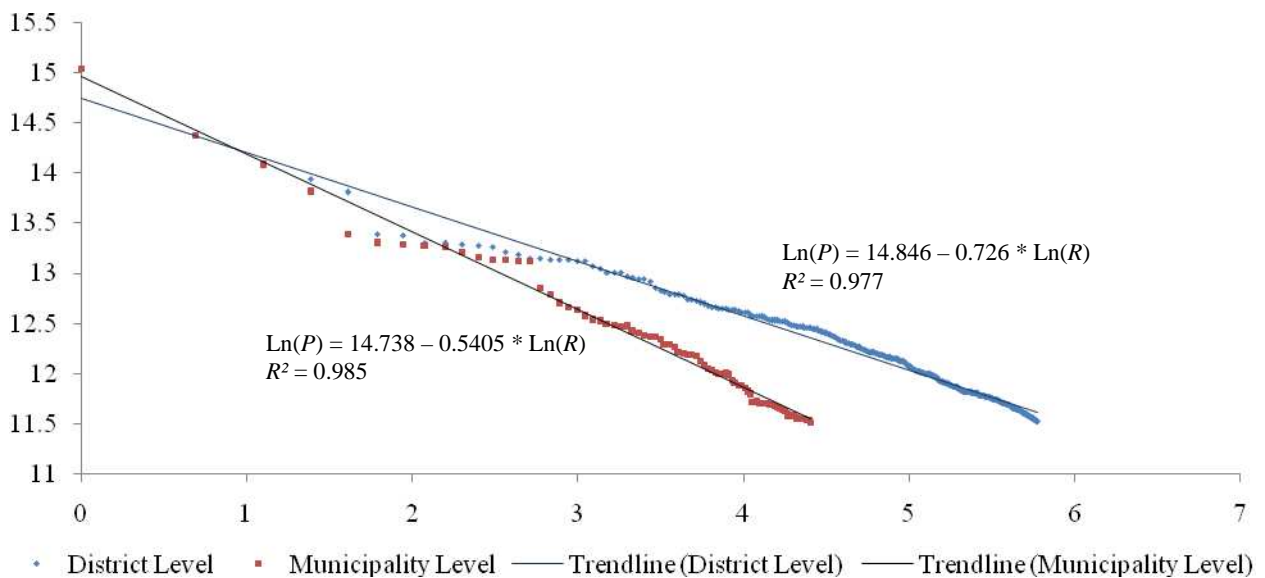


Figure B2: Rank-size Rule across 41 Municipalities and 321 Districts in Germany (Truncated Samples: More than 100,000 Inhabitants) (Reference to Table 9).

Legend: x-axis: $\text{Ln}(R)$; y-axis $\text{Ln}(P)$



Alma Mater Studiorum - Università di Bologna
DEPARTMENT OF ECONOMICS

Strada Maggiore 45
40125 Bologna - Italy
Tel. +39 051 2092604
Fax +39 051 2092664
<http://www.dse.unibo.it>