

TI 2012-066/3  
Tinbergen Institute Discussion Paper



# The Death of Distance Revisited: Cyberplace, Physical and Relational Proximities

*Emmanouil Tranos*

*Peter Nijkamp\**

*Faculty of Economics and Business Administration, VU University Amsterdam.*

*\* Tinbergen Institute.*

Tinbergen Institute is the graduate school and research institute in economics of Erasmus University Rotterdam, the University of Amsterdam and VU University Amsterdam.

More TI discussion papers can be downloaded at <http://www.tinbergen.nl>

Tinbergen Institute has two locations:

Tinbergen Institute Amsterdam  
Gustav Mahlerplein 117  
1082 MS Amsterdam  
The Netherlands  
Tel.: +31(0)20 525 1600

Tinbergen Institute Rotterdam  
Burg. Oudlaan 50  
3062 PA Rotterdam  
The Netherlands  
Tel.: +31(0)10 408 8900  
Fax: +31(0)10 408 9031

Duisenberg school of finance is a collaboration of the Dutch financial sector and universities, with the ambition to support innovative research and offer top quality academic education in core areas of finance.

DSF research papers can be downloaded at: <http://www.dsf.nl/>

Duisenberg school of finance  
Gustav Mahlerplein 117  
1082 MS Amsterdam  
The Netherlands  
Tel.: +31(0)20 525 8579

# The death of distance revisited: cyber-place, physical and relational proximities

Emmanouil Tranos, Peter Nijkamp

Dept. of Spatial Economics, VU Amsterdam, Tinbergen Institute , [e.tranos@vu.nl](mailto:e.tranos@vu.nl)

## Abstract

This paper studies the impact of physical distance and different relational proximity types on the formation of the Internet infrastructure. Although there is some anecdotal evidence on the ‘end of geography’ effect of the Internet, the relationship between physical space and the Internet has not been yet scrutinized. In addition, owing to the network nature of the Internet, the structure of the Internet infrastructure (the cyber-place) cannot be approached in a unidimensional way. Our paper builds upon recent studies in economic geography and relational proximities, and aims to study whether physical distance survives in virtual geography even after controlling for relational proximities. In order to do so, a unique and extensive database with geo-coded IP links is utilized. Based on this, a spatial interaction model with panel data specifications is constructed to study the impact of different types of proximity on the formation of cyber-place. The above analysis is framed by a complex network analysis exercise, which enhances our understanding of the complexity of the Internet infrastructure from a spatial network perspective. Our results indicate that physical distance, but also different relational proximities, have a significant impact on the intensity of the Internet infrastructure, highlighting the spatiality of the Internet.

Keywords: Internet geography, Internet infrastructure, distance, proximities, spatial interaction models.

JEL codes: C23, H54, L96

## 1. Introduction

This paper aims to study the impact of physical and relational proximity on the formation of the Internet infrastructure, which is defined as the *cyber-place* (CP), combining network analysis techniques with econometric modeling. CP is an integral part of Batty's (1997) concept of *virtual geography*, and is defined as the projection of the infrastructural layer of *cyberspace* on traditional space. Thus, CP is defined in accordance with cyberspace<sup>i</sup>, the function of which is supported by CP. Just like other Internet terms including the Internet *superhighways*, *virtual communities*, *web-surfing*, *telecommuting*, etc., CP and cyberspace are geographic metaphors. Apart from being a way to understand the complexity of the Internet, such metaphors reveal its strong spatial underpinning (Graham 1998). Similarly to any other social and economic activity, which is "*inscribed in space and takes place*" (Swyngedouw 1993, p. 305), the Internet, as a platform for *virtual interactions* among individuals and organizations, has necessarily a geographical component. Following Batty's conceptualization, the virtual geography's element which is mostly responsible for the Internet's spatiality is the CP, because of its tangible geographic signature (Malecki 2002).

Despite the interpretation of early commentators that the Internet has an anti-spatial nature (Mitchell 1995), the above discussion highlights the need to further explore the spatial dimension of such a system. Indeed, the rapid Internet penetration has generated research which approaches the impact of information and communication technologies (ICTs) in a rather deterministic way, declaring the emergence of telecottages (Toffler 1980), the rising of a 'borderless world' (Ohmae 1995), the death of cities (Gilder 1995; Drucker 1998; Kolko 1999), and in general the 'end of geography' (O'Brien 1992) and the 'death of distance' (Cairncross 2001). However, such narratives, with the most recent one being Friedman's best seller book "*The World is Flat*" (2005), have not been accompanied by empirical investigation and hard evidence. Although we know that "*cities are well and alive*" (Malecki 2002, p. 419), and that ICTs did not generate such one-way dramatic impacts, there is not yet sufficient empirical

knowledge about the relation between the Internet and physical space, and most importantly on whether distance performs a meaningful role in virtual geography.

Despite the death of distance thesis, geographical proximity, as depicted by Euclidean distance, performs important economic roles: it triggers agglomeration economies, underpins regional economics and provides the foundation for vital theoretical contributions including Central Place Theory and New Economic Geography (Partridge et al. 2008). Rietveld and Vickerman (2004, p. 241) characterized the 'death of distance' discussion, which is based on assertions regarding the dominance of footloose economic activities, the vanishing of agglomeration economies and the annihilation of communication cost, as "unmistakably premature". Polèse and Shearmur (2004) found that factors such as physical distance and city size remain good predictors for most industrial activity, both in manufacturing and services. Disdier and Head (2008) suggested that despite the digital revolution, distance's explanatory power as an international trade determinant increases over time. Similarly, Gaspar and Glaeser (1998) concluded that telecommunication improvements will result in increased demand for face-to-face interactions (see also Brakman and Marrewijk 2008) and therefore, the importance of cities in spatial structure and economic activity, as centers of interaction, will also increase. Simai and Waldfogel (2004) raised the question whether the Internet is a substitute or a compliment of cities. Based on information regarding the geographic scope (local/non local) of online information, they concluded that the complementarity of local websites with local agglomeration offsets the Internet's substitution effect. Moreover, Forman et al. (2005), using data from 2000, concluded that although Internet adoption for firms with more than 100 employees was faster in smaller cities in 2000, the adoption of more sophisticated Internet based applications was positively related with urban size.

Drawing upon the above, the first research question of this paper is whether physical distance survives in the frame of the digital economy. Although we know that spatial configuration and the importance of

agglomeration for social and economic activities is valid in the context of the digital economy, we still do not know if and how the Internet itself is affected by the tyranny of distance. While it is well established that the Internet is a friction-reducing technology (Cohen, Salomon, and Nijkamp 2002; Cohen-Blankshtain and Nijkamp 2004), the effect of distance and proximity on its structure is vague. Even though Wang et al. (2003) presented some first evidence on the distance decay effect on information accessibility and D'Ignazio and Giovannetti (2007) highlighted the importance of proximity in bilateral interconnection decisions at Internet Exchange Points in Europe, we still do not know whether the Internet's infrastructural layer is affected by centripetal or centrifugal forces resulting in spatial clustering. Put simply, we do not know if the cost related with physical distance affects the structure of a digital system as complex as the Internet.

However, our analysis is not limited only to physical distance. Other proximity types may also affect the topology of this digital infrastructural system. Following Massey (1993), we understand the Internet as a relational notion which cannot be approached in a unidimensional way, just as a Cartesian spatial object (Graham 1998). The underlying technological, economic, and social complexities of the formation of the CP force us to explore the impact of other types of proximity on the formation of CP including cognitive/technological, organizational, and institutional distance. Just like physical distance, the above types of relational proximities are approached here as different costs for the formation of CP. To study these different types of costs on the formation of CP, we create a multi-dimensional space where physical and relational proximities co-exist. In other words, we empirically test here whether Tobler's (1970) First Law of Geography<sup>ii</sup> is valid in the formation of CP, but we expand the notion of distance to include relational proximities as well.

The novelty of our approach lies in the fact that although the geographic analysis of the Internet already has a short history of almost 15 years (Moss and Townsend 1997; Wheeler and O'Kelly 1999; Malecki

and Gorman 2001), the impact of distance on the formation of the Internet and the CP has not yet been empirically tested. In addition, research has largely ignored the role that relational types of proximity play in the topology of CP. These remarks reflect the overall disregard of the Internet by spatial sciences because of its intangible, elusive, and complex technical nature (Bakis 1981; Hepworth 1989; Kellerman 1993). After all, telecommunications infrastructure only becomes visible when it stops working (Star 1999).

The second research question of this paper is related with shedding more light on the complex nature of the CP from a spatial network perspective. The infrastructural layer of the Internet has been developed as a network of numerous different networks. Although the complexity of the Internet has been extensively studied from a complex system perspective (e.g. Faloutsos, Faloutsos, and Faloutsos 1999; Pastor-Satorras and Vespignani 2004; Adamic and Huberman 2002) and was one of the main test-beds for scale free network models (Barabási and Albert 1999; Albert and Barabási 2002), to date insufficient effort has been made to approach this complexity from a spatial perspective<sup>iii</sup>. Apart from bridging this gap, gaining knowledge of the spatial network attributes of CP is essential to interpret the impact of distance and relational proximities on the formation of the digital infrastructure.

To empirically test the above research questions, we utilize here an extensive aggregated data set for the European CP, which, as far as we are aware, has never been used before in a spatial context. Complex network analysis and spatial interaction models with panel data specifications, will be employed to quantitatively approach the research questions.

The structure of the paper is as follows. Next, in Section 2, the CP database is described. Then, in Section 3, the network structure of CP is explored using complex network analysis methods. Section 4 presents the modeling exercise on the impact of different types of proximity on the formation of CP. Finally, the paper ends in Section 5 with some concluding remarks.

## 2. Database Description

The main data used for this study is the output of the DIMES project. This is “a distributed scientific research project, aimed to study the structure and topology of the Internet, with the help of a volunteer community” (DIMES 2010). It is based on 3–6 million traceroute<sup>iv</sup> measurements made daily by a global network of more than 10,000 agents, who are voluntarily participating in this research project (Carmi et al. 2007; for a description of the DIMES project, see also Shavitt and Shir 2005). The final outcome of the DIMES project is derived after the triangulation and geo-location of the IP (Internet Protocol) links discovered by the DIMES volunteers, and it contains all the IP links between any two cities discovered by the DIMES agents. Although overlapping connections between any two regions are included in the database, there is no indication of the capacity of these links. However, this is still an infrastructural measure, as the IP links represent physical (overlapping) data links between cities, which follow the IP protocol<sup>v</sup>.

Nonetheless, a few important points need to be highlighted here. Firstly, this is only a very small fraction of the Internet. Indeed, the DIMES project only includes the IP links which have been captured by the volunteers who acted as the DIMES agents. By sending data packets from the agents’ locations to known destinations, DIMES researchers record the different IP links used by its agents, completing in this way the largest available data source for geo-coded IP links.

In addition, an inherent limitation of the geographic analysis of the Internet is its topological rather than geographic basis. Indeed, the Internet was designed as a logical network, the links of which are defined in topological terms, not using geographic coordinates. Thus, the Internet architecture of topological destinations (IP addresses) has little to do with geographical locations (Dodge and Zook 2009). In order to understand the above structure from a geographic perspective, an indirect approach is adopted and effort is spent to geo-code the different IP addresses using IP registration tables. This task is part of the

DIMES project. We need to highlight here a potential accuracy issue due to the geo-coding process. It is common that IP addresses are owned by specific firms, which lease these IP addresses to the content providers (Dodge and Zook 2009). The outcome of this process is that usually the physical location of the IP address, which is derived by the geo-coding process, does not match with the location of the content. However, this does not create any bias here, as the focus of this paper is the CP and the physical infrastructure of the Internet.

For the purposes of our analysis, an aggregation process has been carried out. Initially, the IP links provided by the DIMES project, were geo-coded at the city level. In order to homogenize and standardize the data, the IP links among European cities were aggregated at the regional NUTS-3<sup>vi</sup> level in such a way that the city-to-city links were transformed to region-to-region links. The intra-regional links derived from the aggregation process were also included in the analysis. In total, the database follows panel specifications including IP links among almost all NUTS-3 regions for the period 2005-2008 at an annual base. Table 1 below provides the details.

Table 1: IP links data  
description

<b>Year</b>	<b>Inter-region links</b>
2005	22259
2006	21277
2007	21756
2008	18408
Total	83700
Unique	44518

In general, although specific limitations exist in studying the Internet from a geographic perspective, the DIMES data set appears to be the richest available data source with a geographic projection of the Internet infrastructure. Despite the above-mentioned limitations, the size of the DIMES experiment and the scattered locations of the DIMES agents enable us to safely use this data set, especially considering the general lack of geographic data on the Internet and the Internet infrastructure.

### 3. The Network Structure of CP

The first step of our analysis explores the network topology of the IP links using concepts and methods from the complex network analysis (CNA) field. To provide a brief introduction, the ideas which underpin this section are derived from the *new science of networks* (Barabási 2002; Buchanan 2002; Watts 2003, 2004), an analytical field which focuses on large-scale real-world networks and their universal, structural, and statistical properties (Newman 2003). Despite the fact that the CNA starting point was governed by statistical physics, strong parallels exist between CNA and regional science, as the latter traditionally has a strong interest in networks and interregional systems (Cornell University 2010): while regional science focuses on spatial structure, network analysis focuses on topological structure; and, while the former emphasizes the economic meaning of functional forms, the latter stresses the connectivity patterns of functional forms (Reggiani 2009; Reggiani and Nijkamp 2009). Drawing upon this conceptual parallel, CNA is utilized here as a tool to explore connectivity patterns in the topological configuration of the CP. The understanding of the latter is an essential step in order to move on to the second part of our analysis, where the impact of distance and different relational proximity types on the CP's topology will be explored.

Table 2 presents some basic network statistics for the CP for the years 2005 and 2008. Although it seems that the size of the CP has decreased over time, this refers to the part of the IP space captured by the

DIMES project and not to the overall Internet. Thus, we cannot draw immediate conclusions about the change of the size of the CP. Nonetheless, a change in the topology of the network can be observed. Although less IP links are included in the analysis for 2008, an increase in the average and maximum *degree centrality* is observed, with the latter being almost doubled during the study period. Degree centrality is a connectivity measure and, in this case, is defined as the number of the accumulated IP links in each region discovered by the DIMES agents during the course of one year<sup>vii</sup>. The average and maximum value of this measure refer accordingly to the average and maximum degree centralities among all network nodes. Such measures reflect the topological attributes of the network and in this case the difference between average and maximum reflects the existence of some very well connected nodes, which perform hub roles in the network. To further illustrate the connectivity inequality among network nodes, the Gini coefficient is introduced. According to this measure, connectivity appears to follow a highly uneven distribution in both years. Notable is also the almost stable over time high inequality.

Table 2: Network statistics

Year	# of European nodes	# of intra-European IP links	av. degree <sup>a</sup>	max. degree*	Gini coef.	density <sup>a</sup>	av. dist. <sup>a</sup>	av. dist. RN	CC <sup>a</sup>	CC RN
<b>2005</b>	1376	23352	1084	44313	0.727	0.024	2.295	2.831	0.71	0.012
<b>2008</b>	1276	19521	1490	77692	0.741	0.023	2.176	2.891	0.69	0.012

<sup>a</sup> for these metrics, links between Europe and the rest of the world were also included in the analysis.

The outcome of the uneven distribution of the IP links among the European regions is an *efficient* CP. Indeed, despite the very low density of the CP, the average network distance is exceptionally short. In the CNA framework, distance does not refer to Euclidean distance, but to the number of nodes that separate any two nodes<sup>viii</sup>. For the case of CP, any two regions are separated on average by one intermediate node, which results in a network distance a little higher than 2. The latter is an indication of efficiency, as it reflects the ability of the network to transfer data flows with minimal routing.

The above qualities and the efficiency of the network can be attributed to the *small world* (SW) characteristics of the CP. The latter refers to a widely used network model, whose main characteristic is the existence of highly-connected cliques<sup>ix</sup>, which gain global connectivity via a few links that span the entire network, linking distant clusters (Watts and Strogatz 1998). This theoretical network model became popular because of its real-world applications. The CP resembles SW networks because of the short average distance – shorter than that observed in same size random networks (RN) – and the high clustering coefficient<sup>x</sup> – higher than that observed in same size RN.

Apart from the latter, an essential element of the SW networks is the distribution of nodes' degree centrality, which distinguishes this network type from another widely used type of network model known as *scale free* (SF). SF networks share the above characteristics with SW networks, but the degree distribution of their nodes follows power laws, contrary to the exponential laws which characterize SW networks. The different distributions reflect the difference between these two types of networks in terms of the nodes' heterogeneity: while the power-law degree distribution of the SF networks reflects the existence of a very few super-connected hubs and a vast majority of less-connected vertices (Barabási and Albert 1999), the exponential-degree distribution of SW networks resembles highly-connected cliques and less heterogeneous nodes<sup>xi</sup>. Following Newman (2005), the estimation of the

degree distribution curve is based on the *cumulative degree function* (CDF) derived from an inverse rank-plot graph. The CDFs for the years 2005 and 2008 are presented in Figure 1.

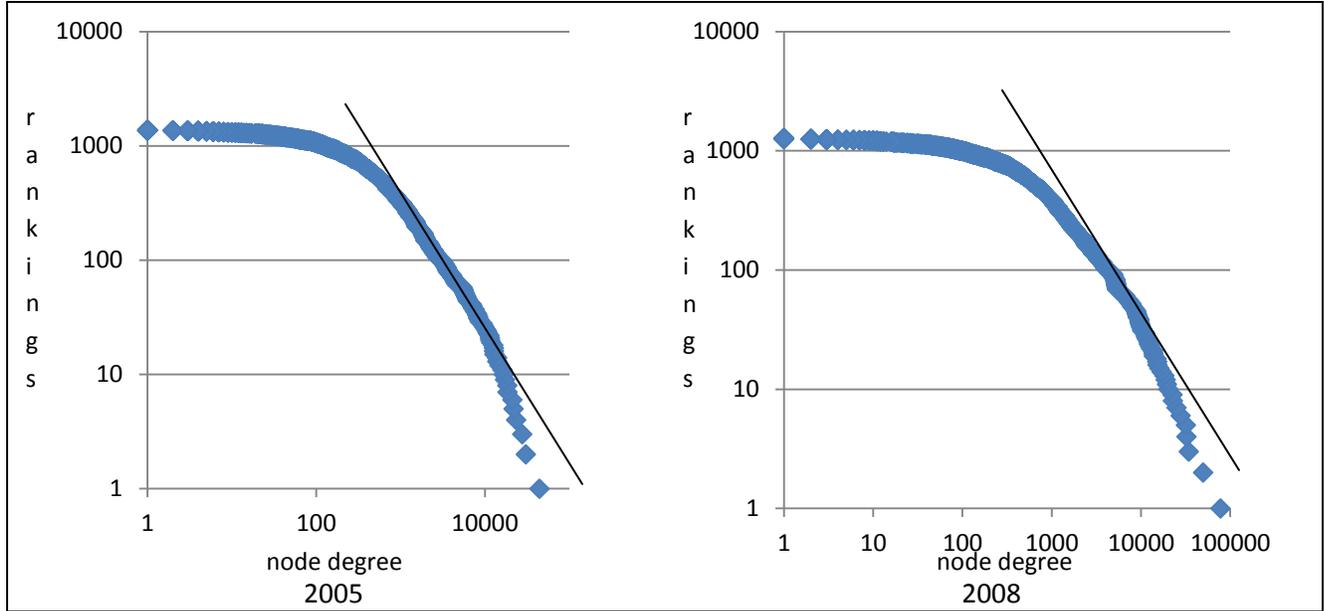


Figure 1: Cumulative degree distribution of NUTS-3 regions based on IP links

The scatter plots reveal the existence of two different curves for both years: a straight line indicating a power function for the most-connected nodes of the IP network and a curve suggesting an exponential function for the least-connected nodes. This ‘dual’ character of the CDF suggests a *power law with a cutoff*, since the power function does not fit the overall distribution, but only the most-connected nodes. The above visual observations are supported statistically by curve estimations based on OLS and the relevant log-log transformations (Faloutsos, Faloutsos, and Faloutsos 1999; Gorman and Kulkarni 2004; Patuelli et al. 2007; Schintler et al. 2004; Tranos 2011; Reggiani, Bucci, and Russo 2010). The results of the OLS are presented in Table 3, where three different forms are tested: exponential, power, and power with cutoff (Tanner function), accordingly:

$$p(x) \propto e^{-ax}, \tag{1}$$

$$p(x) \propto x^{-a}, \tag{2}$$

$$p(x) \propto x^{-a} e^{-\lambda x}. \quad (3)$$

Table 3: Degree distribution fit

	N	Exponential		Power		Tanner function		
		R <sup>2</sup>	Coef.	R <sup>2</sup>	Coef.	R <sup>2</sup>	Power Coef.	Exp. Coef.
<b>2005</b>	1376	0.679	0.0003	0.733	-0.481	<b>0.909</b>	-0.323	-0.0002
<b>2008</b>	1276	0.632	0.0002	0.712	-0.435	<b>0.889</b>	-0.305	-0.0001

Indeed, the OLS results confirm the visual observation that Tanner functions better fit the overall distribution for both years<sup>xii</sup>. So, at an aggregated NUTS-3 level, the European IP network fails to form a clear SF structure. In spatial terms, this can be interpreted as an agglomeration effect of IP connectivity in a limited number of regions which act as hubs. At the same time, the exponential tail reflects the existence of a cluster of less-connected regions, which is more homogeneous in terms of IP connectivity than if a hierarchical and clear SF topology were present. Although the above analysis provides interesting insights into the nature of this infrastructural system at a global (i.e. overall network) level, it does not provide insights on the mechanisms behind the creation of these links. Moving a step forward, the next section will explore the impact of physical distance and relational proximities on the formation of this complex network. The knowledge gained in this section will be used as a tool to interpret the effect different proximities on this complex network.

#### 4. CP and Proximities

The main focus of this section is to test the impact of distance on the formation of CP. In addition, the empirical analysis is enriched by exploring the impact of other – relational – proximities on CP. The latter apart from offering fruitful results, enables us to test the sensitivity of distance (or in other words geographical proximity) against relational proximities. Drawing upon the economic geography literature, relational proximity is a multidimensional notion including notions such as cognitive, organizational, and institutional proximity. The main hypothesis is that distance as well as relational proximities affect the structure of the CP and the intensity of the IP connectivity between places. In addition, distance decay effect remains significant (and negative) even after controlling for relational proximities. The conceptual model of this analysis is formulated in the below generalized version of a spatial interaction model (SIM), according to which the number of IP installed links between  $i$  and  $j$  ( $IP_{ij}$ ) is affected by the characteristics of  $i$  ( $A_i$ ) and  $j$  ( $B_j$ ) and various proximity types between  $i$  and  $j$  ( $F(d_{ij})$ ).

$$IP_{ij} = A_i B_j F(d_{ij}), \quad (4)$$

Before presenting the econometric specifications and the results of the above simple unconstrained SIM, the different relational proximities are introduced, as well as their quantification in the CP framework.

##### *Different proximities*

The starting point for defining the different dimensions of proximity lies in the *French School of Proximity*, which studied the spatial dimension of enterprises and organizations. The main objective of this group of industrial economists was to endogenize space in economic analysis, and, more specifically, to incorporate space and other territorial proximity elements in a research framework, which aims to better understand the dynamics of innovation (for a review of the French School of proximity, see Torre and Gilly 2000). A second development in further decomposing and analyzing the different components

of proximity was research related to innovation and territorial learning in the broader framework of evolutionary economic geography. In recent years we have experienced an increased interest in the factors which explain how firms and regions interact as part of a 'collective learning process', since learning and knowledge creation are an essential component of the firms' and regions' competitive advantage (Boschma 2005). The notion of proximity and its different components are juxtaposed with ideas about knowledge transfer and creation, tacit knowledge, and learning regions (Boschma 2004).

The common basis of these approaches is the importance of non-spatial types of proximity in innovation creation. Although the latter is not the subject of the present paper, we can 'borrow' the conceptual work on the different proximity dimensions, redefine and use them in a new empirical framework to understand the impact of different types of proximity on the formation of the CP. After all, proximity and distance are just other facets of cost which needs to be incorporated in the connectivity decisions taken by Internet Service Providers (ISPs). Starting now from the French School, two different types of proximity can be identified: *geographical* and *organized* (Torre and Rallet 2005). The former type is more straightforward and reflects physical distance. Nonetheless, different conceptualizations of geographical proximity could also be utilized, as it might be affected not only by Euclidean distance, but also by the transportation cost between two places and their accessibility. In our case, geographical proximity is represented by the Euclidean distance between the centroids of two NUTS-3 regions and it is expected to negatively affect IP connectivity.

Unlike geographical proximity, organized proximity is a relational notion, and refers to the ability of an organization to enhance interaction between its members. The main point behind this concept is that members of the same organization will interact together more easily than actors outside the organization. This is based on two different logics (Torre and Gilly 2000): (a) adherence logic, according to which actors, who are close in organizational terms, such as a firm, network etc., are part of the same

relational space; and (b) similarity logic, according to which the organizationally close actors tend to be alike. In a nutshell, while geographical proximity reflects separation in space regarding physical distance, organized proximity is considered as the overall framework in which the different actors interact. In the same vein, Boschma defined *organizational proximity* “as the extent to which relations are shared in an organizational arrangement, either within or between organizations” (Boschma 2005, p. 65). Since digital infrastructure follows the expected demand for IP communications, we would expect that IP connectivity would be higher among places which are organizationally proximate.

The main difficulty in quantifying and transferring organizational proximity to our CP context is the aggregate level of our analysis. Indeed, this relational proximity type was initially conceptualized at the micro-firm level. Nonetheless, we argue here that such a concept is also applicable to the aggregated regional level, as firms do not operate in a vacuum. On the contrary, agglomeration forces perform a significant role in the firm’s location factors. The summation of organizational links in a region can provide an overview of the organizational profile of that region. In order to build a proxy for this type of proximity, the world city ranking by Peter Taylor is utilized (GaWC 2008). The latter is a well-known research approach in the world cities research field, according to which urban hierarchies are derived based on the links that advanced producer service firms share with the rest of the world. In more detail, Taylor (e.g. 2004), instead of directly using cities as the nodal level of his network, created a third sub-nodal level in order to include in his analysis the agents which “taken together, are primarily responsible for shaping the world city network”; these are service firms, city governments, service-sector institutions, and nation states (Taylor 2004, p. 58). From these four, he identified service firms and more specifically Sassen’s (1991) advanced producer services as the main agent for world city formation. His latest version of the world city network (Taylor et al. 2010) is based on relational data from about 175 multinational firms, which can be identified as advanced producer services. Using the intra-firm connections, he created a roster of 525 cities. Drawing upon this work, here we propose a dummy

variable which takes the value 1 when both ends of an IP link are included in Taylor's world city network. This is an aggregated measure of organizational proximity, as virtual interaction, and, consequently, the digital infrastructure between two cities with extensive global urban organizational links, is expected to be higher than interaction between less-connected cities.

Building upon the French School of proximity findings, Boschma (2005) approached proximity as a five-dimension notion. *Cognitive* proximity is the point of departure for his conceptual framework and is defined as the level of similarity of the knowledge base of different organizations (Nooteboom 2000; Balland 2011). Organizations collaborate and form links and networks using as criteria for their choices the knowledge background of the potential partners, as people and organizations, which share the same knowledge background and expertise may learn from each other (Boschma 2005). The complexity of the learning process is reflected in the non-linear effect of cognitive proximity in learning. In order for knowledge to be transferred, gained or created, there is a need for optimal cognitive proximity, since too high cognitive proximity will eliminate any novelty from the interaction, while, vice versa, too high cognitive distance will result in communication difficulties (Nooteboom 2000).

Despite the effort spent in the relevant literature, cognitive proximity is still a rather fuzzy concept, and it is difficult to quantify. Strong links can also be identified between cognitive proximity and technological similarity. While some authors distinguish these two (e.g. Knobens and Oerlemans 2006), more often than not these notions are used interchangeably in an empirical context (e.g. Marrocu, Paci, and Usai 2011; Walukiewicz 2007). While cognitive proximity represents the similarity of the knowledge bases of two organizations or regions, technological proximity reflects the similarity between the technological knowledge among economic actors (Dangelico, Garavelli, and Petruzzelli 2010). Examples of the empirical specification of cognitive and technological proximity include, among other things, the use of human capital skills in an organization (Crisuolo, Salter, and Wal 2010), the economic activity and

technological classification of firms (Broekel and Boschma 2011), and the use of patents (Dangelico, Garavelli, and Petruzzelli 2010). In the view of this, and because of the strong technological nature of the main research subject of this paper, here we introduce two dummy variables in order to capture technological proximity in the context of digital infrastructure. The first variable is equal to 1 for those IP links which connect regions which are part of the topological core of the IP network. The second dummy variable is equal to 1 when the IP link connects a peripheral with a core region. The core and the periphery of the IP network are defined here as those regions which are placed, respectively, in the first and the fourth quartile of the IP degree distribution. These variables also reflect the discussion in the previous section regarding the node degree distribution and the few super-connected hubs. In order to avoid endogeneity issues, the degree distribution, just as in the previous section, is derived after the inclusion in the analysis of the IP links between European and non-European destinations. With regard to the expected sign in the regression, a positive impact is expected for the core-to-core dummy variable, since core IP hubs need to intensively connect to each other as their spokes gain global connectivity through them. On the other hand, the links between the spokes and the hubs are of less structural importance for the overall network, so that the impact of the core-to-periphery variable is expected to be negative.

In addition, institutional proximity is a concept also proposed by the French School (Kirat and Lung 1999). Following North's (1993) definition, institutions are the amalgamation of formal rules and informal constraints including behavioral and social standards, while organizations can be approached as a group of agents performing the same activity. Put simply, organizations define agents' practices and strategies in the overall context provided by the institutional ecosystem in which they are positioned (Kirat and Lung 1999). In our context, following Hoekman et al. (2009), institutional proximity is defined on the basis of whether or not two regions are part of the same country. The underlying assumption is that two regions, which share the same institutional-country characteristics, will be characterized by a

higher level of virtual interaction than two regions from different countries. Thus, the installed digital infrastructure between nearby places in institutional terms is expected to be more intensive than between distant places. For this reason, a dummy variable is built with 0 denoting an IP link between two regions from a different country and 1 indicating an intra-country IP link. In addition, the adopted level of analysis enables us to capture even more detailed institutional characteristics, as the adopted city-region spatial unit is characterised by a degree of functional integration. On that basis, it may be assumed that the location of two cities in the same region will positively affect the intensity of the digital infrastructure between them. For this reason, a second dummy variable is built with 1 indicating intra-regional IP links, and 0 vice versa.

In total, six variables denoting four different types of proximity (geographical (1), cognitive (2), organizational (1), and institutional (2)) will be used in order to explore the impact of proximity on the formation of CP. In addition, a variable indicating the population difference between the linked regions is also included in our analysis in order to capture size effects. Due to the lack of previous research in this area, it is difficult to predict if there is a cost in digitally connecting dissimilar agglomerations. The expected signs and a brief description are presented in Table 4.

Table 4: Proximity types and related variables

Proximity type	Variable	Data source	Variable name	Expected sign
Geographic	Physical distance in km (natural logarithm)	Own calculations	$dist_{ln_{ij}}$	-
Cognitive	Core-to-core (IP)	Own calculations	$c2c_{ijt}$	+

	Core-to-periphery (IP)	Own calculations	$c2p_{ijt}$	-
Organizational	World cities	GaWC, own calculations	$gawc_{ij}$	+
Institutional	Intra-country virtual interaction	Own calculations	$cntr_{ij}$	+
	Intra-region virtual interaction	Own calculations	$inter_{ij}$	+
Population	Absolute population distance (natural logarithm + 1)	Eurostat, own calculations	$pop\_diff\_ln_{ijt}$	?

---

### *Empirical results*

In order to investigate the impact of physical distance and other types of proximity on the intensity of CP links, equation (4) has been expanded to the following simple unconstrained SIM:

$$\ln(IP_{ijt}) = a_0 \ln k + a_1 IP\_ln_{it} + a_2 IP\_ln_{jt} + a_3 t_2 + a_4 t_3 + a_5 t_4 + b_1 dist\_ln_{ij} + b_2 c2c_{ijt} + b_3 c2p_{ijt} + b_4 gawc_{ij} + b_5 cntr_{ij} + b_6 inter_{ij} + b_7 pop\_diff\_ln_{ijt} + b_7 \sum_{s=1}^N cntr_{ij}^s + \varepsilon_{ijt}.$$

(5)

The dependent variable here is the natural logarithm of the number of IP links between any  $i$  and  $j$  during the period 2005-2008. The right-hand side variables reflect the different types of proximity, and are described in Table 2. In addition,  $\sum_{s=1}^N cntr_{ij}^s$  represents an array of dummy variables controlling for unobserved country-to-country effects. In total, 34 European countries are included in the analysis. Out of the 561 potential country-to-country combinations<sup>xiii</sup>, 508 different country-to-country pairs have been derived from the NUTS 3 to NUTS 3 region links. Furthermore,  $IP\_ln_{it}$  and  $IP\_ln_{jt}$  are the masses of

the Newtonian formula, and in this case they represent the natural logarithm of the weighted degree centrality of the IP links. Simply put, these variables present the total number of IP links originating or terminating in  $i$  and  $j$ . It should be noted here that for the calculation of these variables, apart from the intra-European city-to-city links, the IP links of European cities with the rest of the world were also included in the analysis. This choice was made in order to better reflect the overall importance of  $i$  and  $j$  in CP. Finally, yearly effects  $t_2-t_4$  are also included in the regressions,  $\alpha_0$  is the effect common to all years and pairs of regions, and  $\varepsilon_{ijt}$  is the error term.

Instead of estimating the above model cross-sectionally, a panel data specification is preferred. Firstly, panel data improves the researchers' ability to control for missing or unobserved variables (Hsiao 2003). Such an omitted-variable bias as a result of unobserved heterogeneity is a common problem in cross-section models. In addition, potential selection bias in IP links due to the traceroute process can be addressed more efficiently with panel data. In a nutshell, a panel data specification reduces the risk of obtaining biased estimators (Baltagi 2001).

While panel data introduces substantial gains, there are also methodological shortcomings that need to be addressed. According to the literature (Wooldridge 2003), the most widely used panel data models are the *fixed effects* (FE) and *random effects* (RE). As the main aim of this paper is to estimate the impact of the different proximity measures on the creation of CP, a RE model would have been preferred at a first instance because the first differentiation process of the within estimator (FE) would have resulted in the elimination of most of the time-invariant proximity measures (e.g. Brun et al. 2005; Etzo 2011). Nonetheless, the efficiency of the RE model does not come for free: in order for the RE estimates to be consistent, there is a need for the unobserved random effects to be uncorrelated with the regressors. For instance, the proximity variables might be endogenous by being correlated with omitted variables which affect the installation of IP links between regions (Baier and Bergstrand 2001). If this is the case,

the instrumentation of the endogenous variables would be necessary in order to obtain unbiased estimators. However, such instrumentation is not be an easy task given the complexity of the CP and the lack of prior empirical research in this area. Therefore, a two-way fixed-effects estimation is adopted to overcome the estimation problem. This specification is differentiated by the usual FE because it addresses unobserved effects at two dimensions (Baltagi 1995). Thus, the error term  $\varepsilon_{ijt}$  from equation (5) can be analyzed as following:  $\varepsilon_{ijt} = \mu_{it} + \lambda_{jt} + v_{ij}$ . In this case,  $\mu_{it}$  and  $\lambda_{jt}$  are the  $i$  and  $j$  as well as time-specific effects and  $v_{ij}$  the remainder stochastic disturbance term. The two-way FE will result in omitting the  $i$  and  $j$  time-specific effects (i.e. the masses of the Newtonian equation which are time-variant) as well as the yearly effects and enable the estimation the effect of distance and the other relational proximities. However, it should be highlighted here that because of the high dimensionality of the  $it$  and  $jt$  effects (around 5000 effects for each case), the estimation of this model needs more sophisticated methods than a simple introduction of dummy variables and within transformation. To overcome this computational problem the Stata algorithm `felsdsvreg` (Cornelissen 2008) is utilized here. This algorithm estimates a linear regression model with two high-dimensional fixed effects: one effect is eliminated by the within transformation and the other is included as dummy variables. This algorithm omits the creation of dummy variables and based on the information provided by the group identifiers, it creates the cross-product matrices for least-squares equations. Elsewhere (Andrews, Schank, and Upward 2006), this method has been named as FEILSDVj owing to the combination of FE and least-squares dummy variable model. In addition to the two-way FE, pooled OLS, RE and Poisson estimations are also presented as a sensitivity test.

Table 5 presents the results of the different specifications of the spatial interaction model. Starting in the first column with a regression including only the effect of physical distance and of course the fixed effects which have not been estimated, the first indication of the negative effect of distance on the intensity of IP connectivity can be observed. Thus, 1% increase in distance between two locations will

result in 1.2% decrease of the installed IP infrastructure between these two locations. To investigate the combined effect of distance with relative proximities on the formation of CP, the relational proximity variables are incrementally introduced into the regressions. Firstly, the effect of cognitive proximity is explored. The analysis reveals the cost in establishing digital links between regions with a different level of IP infrastructure. Indeed, while links between core and peripheral regions are less favored in terms of connectivity intensity, the exact opposite applies to the core-to-core IP links (column 2). This is not surprising as it reflects the partial hierarchical hub-and-spoke topology reflected in the Tanner function, which was discussed in the previous section. IP connectivity between nodal locations is expected to be higher than connectivity between nodal and peripheral ones on account of the structural roles that former links perform for the overall network function. At the same time, the cost of establishing IP links in highly connected locations is expected to be lower because of the local knowledge and the associated human capital expertise. It needs to be noted here that the *c2c* variable loses its significance when the organization proximity variable reflecting intra-regional IP connections is included in the model (see for instance columns 7 and 8). Interestingly enough, although technological proximity is a significant predictor of the intensity of IP connectivity, the elasticity of distant is only slightly decreased after the inclusion of the variables reflecting the former.

Similarly, organizational proximity has a positive impact on the formation of the CP. Indeed, the intensity of IP links appears to be higher between regions with cities which are part of the world city network. The rationale behind this lies in the increased demand for IP communications among cities which are intensively inter-linked with the global economy. The positive impact of this variable can also be approached as a justification for the importance of the digital economy in the formation of the world city network, since the digital infrastructure is an underpinning layer of global urban interdependencies (Tranos 2011). Again, the inclusion of the organizational proximity variable does not affect the importance of physical distance in the formation of CP (column 3).

A different effect is captured by institutional proximity, which has a positive impact on the formation of the CP. Column 4 indicates the increased demand for intra-country IP links, or otherwise, an increased cost for country-to-country IP links, which represents a *border effect*. Despite the importance of transnational IP links in gaining global connectivity, it seems that the number of IP links is positively affected by institutional proximity. In the same vein, an institutional effect also emerges at the regional level suggesting localization effects on the structure of the IP network. The latter result confirms the argumentation provided in Section 3 about the SW attributes of the CP. Nonetheless, it needs to be highlighted here that the focus of the analysis is on the IP connectivity and the number of IP addresses, and not the actual capacity of these links. If capacity data were available at such a fine scale, then we would expect to identify a relatively small number of international links, which are characterized by very high capacity (Tranos and Gillespie 2009, 2011). Owing to data constraints, such conclusions cannot be drawn here. After the inclusion of the country-pair effect (columns 6, 7 and 8) it becomes apparent that intra-country effect is stronger than the intra-regional effect. Regarding physical distance, the high coefficients of these two dummy variables result in a significant decrease of the distant effect. This is not surprising as distance is related with intra-country and intra-region links. Nonetheless, what is important here is that physical distance remains significant even after the inclusion of these two variables.

Moreover, population difference between linked regions has a negative effect, which becomes significant only after the inclusion of country-to-country effects (column 6) and highly significant after the exclusion of the regional institutional proximity variable (column 7). This result indicates the cost of establishing infrastructure and the lack of demand between dissimilar agglomerations. Despite its significance, it only affects marginally the distance elasticity.

In addition, column 8 in Table 5 also includes variables to control for selection bias. As the panel is unbalanced, the lack of IP connectivity in some regions is due to the DIMES IP capturing process. In order to correct for potential selection bias due to this process and following Nijman and Verbeek (1992), we introduce three additional explanatory variables including: the number of years that an IP link has been present in the panel (*help\_sel\_b1* in Table 5); a dummy variable that takes the value 1 if an IP link is observed during the entire four-year period (*help\_sel\_b2* in Table 5); and a dummy variable equal to 1 if an IP link is present in  $t-1$  (*help\_sel\_b3* in Table 5). Although all three variables are significant, the results remain unchanged. Not surprisingly, the inclusion of these variables results in a significant decrease of the effect of distance. In total, the distance decay is decreased by 76% after the inclusion of the relative proximity and control variables. Nonetheless, the importance of our findings is that despite the different relative proximities addressed by the models presented in Table 5, physical distance maintains its significant negative character in the formation of CP.

Finally, to increase the robustness of the analysis, Table 6 presents the results of alternative estimators for model 5. Column 1 depicts a simple OLS without any fixed effects, column 2 presents the main effects ( $i$ ,  $j$  and  $t$ ) using a LSDV estimation following Egger and Pfaffermayr (2003) and column 3 estimates 5 using RE. Moreover, column 4 addresses the count nature of the dependent variable using Poisson regression, which is estimated by means of maximum likelihood estimation techniques (see among others Fischer and Wang 2011). The latter addresses a potential drawback of the OLS based estimations, that the data might not meet the general assumption that counts of the dependent variable are log-normally distributed around their mean value with a constant variance (Hoekman, Frenken, and Tijssenc 2010). Not surprisingly, all above different specifications validate the previous outcome, since they result in a negative and of same magnitude (in absolute and relative terms) coefficient for the effect of distance on installed Internet backbone capacity between two regions.

Table 5: SIM results on the intensity of IP links (natural logarithm)

Dep. Var.: IP_In <sub>ijt</sub>	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
dist_ln <sub>ij</sub>	-1.204 (131.00)***	-1.169 (126.88)***	-1.164 (126.36)***	-0.576 (54.59)***	-0.56 (51.15)***	-0.463 (39.15)***	-0.59 (50.23)***	-0.294 (25.80)***
IP_ln <sub>it</sub>				Omitted				
IP_ln <sub>jt</sub>				Omitted				
c2p <sub>ijt</sub>		-0.451 (16.96)***	-0.398 (14.87)***	-0.526 (20.83)***	-0.5 (19.08)***	-0.528 (20.42)***	-0.419 (15.98)***	-0.423 (17.18)***
c2c <sub>ijt</sub>		0.662 (15.63)***	0.652 (15.43)***	-0.025 (0.60)	0 (0.00)	0.057 (1.34)	0.409 (9.67)***	-0.048 (1.2)
gawc <sub>ij</sub>			0.599 (14.62)***	0.297 (7.68)***	0.272 (6.51)***	0.36 (8.62)***	0.133 (3.15)***	0.304 (7.66)***
cntr <sub>ij</sub>				2.489 (74.81)***	2.422 (69.02)***	5.224 (2.09)**	6.881 (2.70)***	2.688 (0.95)
inter <sub>ij</sub>				2.322 (66.99)***	2.337 (48.67)***	2.339 (49.14)***		1.753 (38.31)***
pop_diff_ln <sub>ijt</sub>					-0.003 (0.50)	-0.012 (1.74)*	-0.234 (46.21)***	-0.013 (2.01)**
help_sel_b1 <sub>ij</sub>								0.391 (34.10)***
help_sel_b2 <sub>ij</sub>								0.503 (21.27)***
help_sel_b3 <sub>ijt</sub>								0.323 (15.53)***
Country-pair effects	no	No	no	no	no	yes	yes	Yes
Observations	83,700	83,700	83,700	83,700	77,553	77,553	77,553	77,553
Unique links	44,518	44,518	44,518	44,518	42,396	42,396	42,396	42,396

\*  $p < 0.1$ ; \*\*  $p < 0.05$ ; and \*\*\*  $p < 0.01$ ; t tests in parentheses.

Table 6: SIM alternative specifications

Dep. Var.: IP <sub>ln<sub>ijt</sub></sub>	OLS	OLS and	RE	Poisson
	main effects			
	(1)	(2)	(3)	(4)
dist <sub>ln<sub>ij</sub></sub>	-0.362 (45.76)***	-0.572 (54.14)***	-0.344 (35.59)***	-0.13 (31.18)***
IP <sub>ln<sub>it</sub></sub>	0.575 (113.43)***	0.637 (41.97)***	0.571 (100.66)***	0.214 (81.34)***
IP <sub>ln<sub>jt</sub></sub>	0.572 (115.07)***	0.56 (37.29)***	0.568 (102.39)***	0.214 (83.00)***
c2p <sub>ijt</sub>	-0.149 (7.99)***	-0.373 (17.01)***	-0.153 (8.14)***	-0.057 (6.46)***
c2c <sub>ijt</sub>	0.386 (12.42)***	0.147 (4.21)***	0.368 (12.15)***	0.211 (15.08)***
gawc <sub>ij</sub>	0.491 (14.19)***	0.326 (7.86)***	0.303 (6.79)***	0.076 (4.14)***
cntr <sub>ij</sub>	1.885 (96.62)***	2.379 (70.52)***	1.823 (79.77)***	0.74 (71.58)***
inter <sub>ij</sub>	3.017 (73.52)***	2.269 (48.68)***	2.967 (56.05)***	0.784 (36.18)***
pop_diff <sub>ln<sub>ijt</sub></sub>	0.079 (16.01)***	-0.002 -0.25	0.043 (7.32)***	0.022 (8.34)***
t2		-0.175	-0.211	-0.083

		(10.36)***	(15.87)***	(13.58)***
t3		-0.25	-0.265	-0.101
		(14.40)***	(19.46)***	(16.44)***
t4		-0.308	-0.318	-0.132
		(15.07)***	(19.90)***	(18.24)***
<i>i and j effects</i>	no	yes	no	No
constant	-5.661	-4.428	-5.442	-1.604
	(64.61)***	(5.36)***	(54.31)***	(88.17)***
R <sup>2</sup>	0.36	0.47		
Observations	77,553	77,553	77,553	77,553

---

\*  $p < 0.1$ ; \*\*  $p < 0.05$ ; and \*\*\*  $p < 0.01$ ; t tests in parentheses.

To summarize the results, IP connectivity appears to be higher between neighboring regions in terms of technological, organizational, and institutional proximity. Border and localization effects become significant, even for the digital infrastructure. In addition, costs are also observed in terms of linking dissimilar agglomerations. But most importantly, physical distance continues to play a significant negative role in the formation of CP event after addressing different relative proximities and controlling for selection bias.

## 5. Conclusions

Although spatiality is usually not an issue in the discussion around the Internet, our study addresses this viewpoint by approaching the Internet infrastructure from a spatial perspective, despite the technical and conceptual difficulties that such an exercise involves. The novelty of this paper lies not only in the spatial perspective adopted in the analysis, but also in the effort to quantify these issues through an evidence-based modeling approach. In addition and from a methodology point of view, this paper also

highlights the importance of network analysis in a mainstream econometric framework and the value added of bridging these two analytical strands.

Our analysis reveals that Tobler's First Law of Geography is valid in the CP. The intensity of IP connectivity is higher between neighboring regions indicating the role of physical distance in the formation of the CP. The latter can be approached both as an indication of cost in physical connectivity and also as an indication of higher demand in IP communications between nearer destinations. Our results also support previous evidence on the relation between physical distance and the digital infrastructure (Waxman 1988; Wang, Lai, and Sui 2003; D'Ignazio and Giovannetti 2007). Most importantly, physical distance survives in the virtual geography even after controlling for relational proximities. CP is not only affected by costs deriving from physical distance, but also by costs deriving from other relational proximities. The complex topology of the CP is the result of the impact of different costs projected on the different types of proximities analyzed here and the prospects of ISPs for the location of demand for IP communications.

Drawing on the results of our analysis, specific spatial processes can be identified. Firstly, centripetal forces agglomerate IP links in specific locations, which act as the hubs of this digital infrastructure. During the four-year study period, the uneven distribution of IP connectivity has – marginally – increased. But, overall, CP appears to be strongly curved by agglomeration forces. On the other hand, centrifugal forces 'protect' the less-connected regions, securing a level of connectivity which would have not been observed if a clear SF structure was present. Other forces, including the – limited – provision of IP connectivity from non-private ISPs, ensure that less-connected regions are not as thinly connected as an SF structure would indicate.

In addition, a core-periphery pattern can be identified at a global level. Regions which are strongly integrated with the global economy enjoy higher levels of connectivity. Such a pattern overcomes the

importance of national borders, and illustrates the significance of global urban interdependencies. At the same time, a counterbalancing force was also revealed from our analysis, as border and even local effects have a strong impact on IP connectivity reflecting both cost constraints and prospects for local communications.

In total, this paper highlights the hidden spatiality of the so-called 'placeless' Internet. Its complexity decreases the explanatory value of (over-)simplistic approaches such as the cartoonish 'shrinking world' metaphor, and the subsequent Internet shrinking rather than recasting spatial impacts (Kirsch 1995).

## Acknowledges

The authors would like to thank Jos van Ommeren for his model estimation suggestions.

## References

- Adamic, Lada A. and Bernardo A. Huberman. 2002. "Zipf's law and the Internet." *Glottometrics* 3, 143-150.
- Albert, Reka and Albert-Laszlo Barabási. 2002. "Statistical mechanics of complex networks." *Reviews of modern physics* 74, 47-97.
- Andrews, M., T. Schank and R. Upward. 2006. "Practical fixed-effects estimation methods for the three-way error-components model." *The Stata Journal* 6, 461-481.
- Baier, Scott L. and Jeffrey H. Bergstrand. 2001. "The growth of world trade: Tariffs, transport costs, and income similarity." *Journal of International Economics* 53, 1-27.
- Bakis, Henry. 1981. "Elements for a geography of telecommunication." *Geographic Research Forum* 4, 31-45.
- Balland, Pierre-Alexandre. 2011. "Proximity and the evolution of collaboration networks: evidence from research and development projects within the global navigation satellite system (GNSS) industry." *Regional Studies* forthcoming.
- Baltagi, Badi H. 1995. *Econometric analysis of panel data*. Chichester: John Wiley & Sons.
- . 2001. *Econometric analysis of panel data*. Chichester: John Wiley & Sons.
- Barabási, Albert-Laszlo. 2002. *Linked: The new science of networks*. Cambridge, MA: Perseus.
- Barabási, Albert-Laszlo and Reka Albert. 1999. "Emergence of Scaling in Random Networks." *SCIENCE* 286, 509-512.
- Batty, Michael. 1993. "The geography of cyberspace." *Environment and Planning B* 20, 615-616.
- . 1997. "Virtual Geography." *Futures* 29, 337-352.
- Boschma, Ron. 2004. "Competitiveness of regions from an evolutionary perspective." *Regional Studies* 38, 1001-1014.
- . 2005. "Proximity and innovation: A critical assessment." *Regional Studies* 39, 61-74.

- Brakman, Steven and Charles van Marrewijk. 2008. "It's a big world after all: on the economic impact of location and distance." *Cambridge Journal of Regions, Economy and Society* 1, 411–437.
- Broekel, Tom and Ron Boschma. 2011. "Knowledge networks in the Dutch aviation industry: the proximity paradox." *Journal of Economic Geography* forthcoming.
- Brun, Jean-Francois, Céline Carrere, Patrick Guillaumont and Jaime de Melo. 2005. "Has distance died? Evidence from a panel gravity model." *The World Bank Economic Review* 19, 99-120.
- Buchanan, Mark. 2002. *Nexus: Small Worlds and the Groundbreaking Science of Networks*. New York: W. W. Norton.
- Cairncross, Frances. 2001. *The death of distance 2.0*. London: Texere Publishing Limited.
- Carmi, Shai, Shlomo Havlin, Scott Kirkpatrick, Yuval Shavitt and Eran Shir. 2007. "A model of Internet topology using k-shell decomposition." *PNAS* 104, 11150-11154.
- Cohen-Blankshtain, Galit and Peter Nijkamp. 2004. "The Appreciative System of Urban ICT Policies: An Analysis of Perceptions of Urban Policy Makers." *Growth and Change* 35, 166-97.
- Cohen, Galit, Ilan Salomon and Peter Nijkamp. 2002. "Information-communications technologies (ICT) and transport: does knowledge underpin policy?" *Telecommunication Policy* 26, 31-52.
- Cornelissen, Thomas. 2008. "The Stata command felsdreg to fit a linear model with two high-dimensional fixed effects." *The Stata Journal* 8, 170-189.
- Cornell University. 2010. "About regional science," in Ithaca, NY.
- Criscuolo, Paola, Ammon Salter and Anne Ter Wal. 2010. "The role of proximity in shaping knowledge sharing in professional service firms," paper presented at Opening up Innovation: strategy, organization and technology London.
- D'Ignazio, Alessio and Emanuele Giovannetti. 2007. "Spatial Dispersion of Interconnection Clusters in the European Internet." *Spatial Economic Analysis* 2, 219-36.
- Dangelico, Rosa Maria, Achille Claudio Garavelli and Antonio Messeni Petruzzelli. 2010. "A system dynamics model to analyze technology districts' evolution in a knowledge-based perspective." *Technovation* 30, 142–153.
- DIMES. 2010. "DIMES Project," in Tel Aviv.
- Disdier, A. C. and K. Head. 2008. "The puzzling persistence of the distance effect on bilateral trade." *The Review of Economics and Statistics* 90, 37-48.
- Dodge, Martin and Matthew A. Zook. 2009. "Internet Based Measurement," in R. Kitchin and N. Thrift (eds), *The International Encyclopedia of Human Geography*: Elsevier.
- Drucker, Peter F. 1998. "From Capitalism to Knowledge Society," in D. Neef (ed.), *The knowledge economy* Woburn, MA: Butterworth-Heinemann. pp. 15-34.
- Egger, Peter and Michael Pfaffermayr. 2003. "The proper panel econometric specification of the gravity equation: A three-way model with bilateral interaction effects." *Empirical Economics* 28, 571-580.
- Etzo, Ivan. 2011. "The determinants of the recent interregional migration flows in Italy: a panel data analysis." *Journal of Regional Science* forthcoming.
- Faloutsos, Michalis, Petros Faloutsos and Christos Faloutsos. 1999. "On Power-Law Relationships of the Internet Topology." *Computer Communication Review* 29, 251-262.
- Fischer, M. M. and Jinfeng Wang. 2011. *Spatial Data Analysis. Models, Methods and Techniques*. Heidelberg, Dordrecht, London, New York: Springer.
- Forman, Chris, Avi Goldfarb and Shane Greenstein. 2005. "How did location affect adoption of the commercial Internet? Global village vs. urban leadership." *Journal of Urban Economics* 58, 389-420.
- Friedman, T. L. 2005. *The world is flat*.
- Gaspar, Jess and Edward L. Glaeser. 1998. "Information Technology and the Future of Cities." *Journal of Urban Economics* 43, 136-156.

- GaWC. 2008. "The World According to GaWC 2008," in Loughborough.
- Gilder, George. 1995. *Forbes ASAP*. pp. 56.
- Gorman, Sean P. and Rajendra Kulkarni. 2004. "Spatial small worlds: new geographic patterns for an information economy." *Environment and Planning B* 31, 273-296.
- Graham, Steve. 1998. "The end of geography or the explosion of place? Conceptualizing space, place and information technology." *Progress in Human Geography* 22, 165-185.
- Hanneman, Robert A. and Mark Riddle. 2005. "Introduction to social network methods." *Riverside, CA: University of California, Riverside* (published in digital form at <http://faculty.ucr.edu/~hanneman/>).
- Hepworth, Mark. 1989. *Geography of the Information Economy*. London: Belhaven Press.
- Hoekman, J., K. Frenken and F. van Oort. 2009. "The geography of collaborative knowledge production in Europe." *The Annals of Regional Science* 43, 721-738.
- Hoekman, Jarno, Koen Frenken and Robert J.W. Tijssenc. 2010. "Research collaboration at a distance: Changing spatial patterns of scientific collaboration within Europe." *Research Policy* 39, 662-673.
- Hsiao, Cheng. 2003. *Analysis of panel data*. Cambridge: Cambridge University Press.
- Kellerman, Aharon. 1993. *Telecommunications Geography*. London: Belhaven Press.
- Kirat, Thierry and Yannick Lung. 1999. "Innovation and Proximity : Territories as Loci of Collective Learning Processes." *European Urban and Regional Studies* 6, 27-38.
- Kirsch, Scott. 1995. "The incredible shrinking world? Technology and the production of space." *Environment and Planning D: Society and Space* 13, 529-555.
- Knoben, Jorris and Leon. A.G. Oerlemans. 2006. "Proximity and inter-organizational collaboration: a literature review." *International Journal of Management Review* 8, 71-89.
- Kolko, Jed. 1999. "The Death of Cities? The Death of Distance? Evidence from the Geography of Commercial Internet Usage," paper presented at Selected Papers from the Telecommunications Policy Research Conference 1999 Newcastle.
- Latora, Vito and Massimo Marchiori. 2001. "Efficient Behavior of Small-World Networks." *Physical Review Letters* 87.
- Malecki, Edward J. 2002. "The economic geography of the Internet's infrastructure." *Economic Geography* 78, 399-424.
- Malecki, Edward J. and Sean P. Gorman. 2001. "Maybe the death of distance, but not the end of geography: the Internet as a network," in T. R. Leinbach and S. D. Brunn (eds), *Worlds of E-Commerce: Economic, Geographical and Social Dimensions* West Sussex: Wiley. pp. 87-105.
- Marrocu, Emanuela, Raffaele Paci and Stefano Usai. 2011. "Proximity, Networks and Knowledge Production in Europe," Working Paper, CRENOS.
- Massey, Doreen. 1993. "Power-geometry and a progressive sense of place," in J. Bird, B. Curtis, T. Putnam, G. Robertson and L. Tickner (eds), *Mapping the futures: local cultures, global change* London: Routledge. pp. 59-69.
- Mitchell, William J. 1995. *City of bits: space, place and the infobahn*. Cambridge, MA: MIT Press.
- Moss, Mitchell L. and A. M. Townsend. 1997. "Tracking the net: Using domain names to measure the growth of the Internet in US cities." *Journal of Urban Technology* 4, 47-60.
- Newman, Mark E. J. 2003. "The structure and function of complex networks." *SIAM Review* 45, 167-256.
- . 2005. "Power laws, Pareto distributions and Zipf's law." *arXiv:cond-mat/0412004v3*.
- Nijman, Theo and Marno Verbee. 1992. "Incomplete Panels and Selection Bias," in L. Matyas and P. Sevestre (eds), *The Econometrics of Panel Data* Dordrecht, Netherland: Kluwer.
- Nooteboom, Bart. 2000. "Learning by Interaction: Absorptive Capacity, Cognitive Distance and Governance." *Journal of Management and Governance* 4, 69-92.
- Nooy, Wouter de, Andrej Mrvar and Vladimir Batagelj. 2005. *Exploratory Social Network Analysis with Pajek*: Cambridge University Press.

- North, Douglass C. 1993. "Institutional change: a framework of analysis," in S. E. Sjöstrand (ed.), *Institutional Change: Theory and Empirical Findings* New York: M.E. Sharpe.
- O'Brien, R. 1992. *Global Financial Integration: The End of Geography*. London: Pinter.
- Ohmae, K. 1995. *The Borderless World: Power and Strategy in an Interdependent Economy*. New York: Harper Business.
- Partridge, Mark D., Dan S. Rickman, Kamar Ali and M. Rose Olfert. 2008. "Employment Growth in the American Urban Hierarchy: Long Live Distance." *The B.E. Journal of Macroeconomics* 8, Article 10.
- Pastor-Satorras, Romualdo and Alessandro Vespignani. 2004. *Evolution and Structure of the Internet*. Cambridge: Cambridge University Press.
- Patuelli, Roberto, Aura Reggiani, Sean P. Gorman, Peter Nijkamp and Franz-Josef Bade. 2007. "Network Analysis of Commuting Flows: A Comparative Static Approach to German Data." *Networks and Spatial Economics* 7, 315-331.
- Polèse, Mario and Richard Shearmur. 2004. "Is Distance Really Dead? Comparing Industrial Location Patterns over Time in Canada." *International Regional Science Review* 27, 431-457.
- Reggiani, A. and P. Nijkamp. 2012. "Did Zipf anticipate socio-economic spatial networks?" *Mimeographed paper* Dept. of Spatial Economics, VU Amsterdam.
- Reggiani, Aura. 2009. "From Complexity to Simplicity," in A. Reggiani and P. Nijkamp (eds), *Complexity and Spatial Networks* Berlin: Springer-Verlag.
- Reggiani, Aura, Pietro Bucci and Giovanni Russo. 2010. "Accessibility and Network Structures in the German Commuting." *Networks and Spatial Economics*, 1-21.
- Reggiani, Aura and Peter Nijkamp. 2009. *Complexity and Spatial Networks*. Berlin: Springer-Verlag.
- Reggiani, Aura and Sandra Vinciguerra. 2007. "Network Connectivity Models: An Overview and Applications to the Space-Economy," in T. Friesz (ed.), *Network Science, Nonlinear Science and Infrastructure Systems* New York: Springer-Verlag.
- Rietveld, Piet and R. W. Vickerman. 2004. "Transport in regional science: The "death of distance" is premature." *Papers in Regional Science* 83, 229-248.
- Sassen, Saskia. 1991. *The global city. New York, London, Tokyo*. Princeton, New Jersey: Princeton University Press.
- Schintler, Laurie, S. P. Gorman, A. Reggiani, R. Patuelli, A. Gillespie, Peter Nijkamp and Jonathan Rutherford. 2004. "Complex Network Phenomena in Telecommunication Systems." *Networks and Spatial Economics* 5, 351-370.
- Shavitt, Yuval and Eran Shir. 2005. "DIMES: Let the Internet Measure Itself." *ACM SIGCOMM Computer Communication Review* 35, 71-74.
- Sinai, Todd and Joel Waldfogel. 2004. "Geography and the Internet: is the Internet a substitute or a complement for cities?" *Journal of Urban Economics* 56, 1-24.
- Star, Susan L. 1999. "The ethnography of infrastructure." *American Behavioral Scientists* 43, 377-91.
- Swyngedouw, Erik. 1993. "Communication, mobility and the struggle for power over space," in G. Giannopoulos and A. Gillespie (eds), *Transport and communications in the new Europe* London: Belhaven. pp. 305-325.
- Tassey, Gregory. 1992. *Technology infrastructure and competitive position*. Boston, MA: Kluwer Academic Press.
- . 2008. "Modeling and measuring the economic roles of the technology infrastructure." *Economics of Innovation and New Technology* 17, 615-29.
- Taylor, Peter J. 2004. *World city network*. London and New York: Routledge.
- Taylor, Peter J., Pengfei Ni, Ben Derudder, Michael Hoyler, Jin Huang and Frank Witlox. 2010. "Global Urban Analysis: a survey of cities in globalization," in London: Earthscan.

- Tobler, Waldo. 1970. "A computer movie simulating urban growth in the Detroit region." *Journal of Economic Geography* 46, 234-40.
- Toffler, Alvin. 1980. *Third way*. New York: William Morrow.
- Torre, André and Jean-Pierre Gilly. 2000. "On the analytical dimension of proximity dynamics." *Regional Studies* 34, 169-180.
- Torre, André and Alain Rallet. 2005. "Proximity and localization." *Regional Studies* 39, 47-59.
- Tranos, Emmanouil. 2011. "The topology and the emerging urban geographies of the Internet backbone and aviation networks in Europe: a comparative study " *Environment and Planning A* 43, 378-392.
- . forthcoming. *The Geography of the Internet: Cities, Regions and the Internet Infrastructure in Europe*: Edward Elgar.
- Tranos, Emmanouil and Andrew Gillespie. 2009. "The spatial distribution of Internet backbone networks in Europe: a metropolitan knowledge economy perspective." *European Urban and Regional Studies* 16, 423-37.
- . 2011. "The urban geography of Internet backbone networks in Europe: roles and relations." *Journal of Urban Technology* 18, 35– 49.
- Vinciguerra, Sandra, Koen Frenken and Marco Valente. 2010. "The geography of Internet infrastructure: an evolutionary simulation approach based on preferential attachment." *Urban Studies* 47, 1969–1984.
- Walukiewicz, Stanisław. 2007. "The Dimensionality of Capital and Proximity," Working Paper, Systems Research Institute.
- Wang, Yong, Phillip Lai and Daniel Sui. 2003. "Mapping the Internet using GIS: The death of distance hypothesis revisited." *Journal of Geographical Systems* 5, 381-405.
- Watts, Duncan J. and Steven H. Strogatz. 1998. "Collective Dynamics of Small-World networks." *NATURE* 393, 440-442.
- Watts, Duncan J. 2003. *Six Degrees: The Science of a Connected Age*. New York: W. W. Norton & Company.
- . 2004. "The 'New' Science of Networks." *Annual Review of Sociology* 30, 243-270.
- Waxman, Bernard M. 1988. "Routing of multipoint connections." *IEEE Journal on Selected Areas in Communications* 6, 1617 - 1622.
- Wheeler, David C. and Morton E. O'Kelly. 1999. "Network topology and city accessibility of the commercial Internet." *Professional Geographer* 51, 327-339.
- Wooldridge, Jeffrey M. 2003. *Introductory econometrics: a modern approach*. Mason, Ohio: South-Western.

---

<sup>i</sup> Cyberspace can be defined as a "new kind of space, invisible to our direct senses, a space which might become more important than physical space itself [and which is] layered on top of, within and between the fabric of traditional geographical space" (Batty 1993, p. 615-616)

<sup>ii</sup> "Everything is related to everything else, but near things are more related than distant things" (Tobler 1970, p. 236)

<sup>iii</sup> Exceptions include among other the work of Gorman and Kulkarni (2004), Schintler et al. (2004), Vinciguerra et al. (2010), and Tranos (2011).

<sup>iv</sup> Traceroutes are specific programs, which map the route that a data packet follows through different nodes in order to reach its final destination (Dodge and Zook 2009).

---

<sup>v</sup> These links function at level 3 of the OSI model. As noted elsewhere (Tranos forthcoming), the first three layers of the OSI model represent physical infrastructural capital, while the four highest layers reflect ‘infratechnologies’ (Tassey 1992, 2008).

<sup>vi</sup> NUTS is the French acronym for the Nomenclature for Territorial Units of Statistics, and NUTS-3 is the most detailed level usually representing a province.

<sup>vii</sup> This is a ‘weighted’ degree centrality measure in the sense that if two regions  $i$  and  $j$  are connected by multiple links, all of these links will be added in the degree centrality of  $i$  and  $j$ . If it had been a ‘binary’ centrality measure, then the multiplicity of the links between  $i$  and  $j$  would have been neglected.

<sup>viii</sup> Because there are usually numerous different ways to connect any two given nodes (known as *walks*), research commonly focuses on the shortest path, known as distance (Nooy, Mrvar, and Batagelj 2005).

<sup>ix</sup> A clique is a “sub-set of a network in which the actors are more closely and intensely tied to one another than they are to other members of the network” (Hanneman and Riddle 2005).

<sup>x</sup> The clustering coefficient  $C_i$  of node  $i$  is the ratio between the number of edges  $E_i$  that exist among its nearest neighbours (nodes which are directly connected with node  $i$ ) and the maximum number of these edges, where  $k_i$  is the number of nodes in clique:  $C_i = 2E_i / k_i(k_i - 1)$  (Latora and Marchiori 2001).

<sup>xi</sup> For a review of the new science of networks from a spatial economics perspective, the reader is referred to Reggiani and Vinciguerra (2007), and, for an application of CNA on the Internet infrastructure, to Tranos (2011).

<sup>xii</sup> In order to validate this result and because Tanner function is a nested function of the other two forms tested here, a cross-equation coefficient equality restriction was tested with the relevant F-tests, the null hypotheses of which (i.e. parameter equality) were rejected at the 5 per cent significance level (see also Reggiani and Nijkamp 2012).

<sup>xiii</sup> The maximum number of edges in a network is  $n(n-1)/2$ .