Knowledge Flows and Regional Disparities in Europe: geographic, functional and sectoral distance*

Mario A. Maggioni
mario.maggioni@unicatt.it

T. Erika Uberti
erika.uberti@unicatt.it

DISEIS (Department of International Economics, Development and Institutions)
and Faculty of Political Science
Catholic University, Milan

Paper prepared for the Conference on
Agglomeration Economies and Regional Growth
Cagliari, 20-21 May 2005

This version: May 2005

ABSTRACT

In this paper we address the manifold nature of knowledge through the analysis of four distinct but complementary phenomena (Internet hyperlinks, European research networks, EPO co-patent applications, Erasmus students mobility) which characterize knowledge as an intrinsic relational structure (directly) connecting people, institutions and (indirectly) regions across five European countries.

We firstly study the structure (in terms of density, clustering and centralisation) of these networks flows through network analysis techniques and we test the influence of geographical distance as opposed to sectoral (based on the industrial distribution of the innovative activity) and functional (based on the value of the RSII European technological leadership index) distances in shaping the strength of knowledge relations though a gravitational model.

The empirical analysis shows the existence of a polarized centre-periphery hierarchy of European regions which is reflected in the structure of knowledge flows

By using a "gravitational" model we demonstrated that, far from the claim of the "death of distance", geographic distance is still relevant for determining the structure of inter-regional knowledge flows.

Functional and sectoral distance play also a crucial role suggesting that knowledge flows easily between similar (according to their scientific, technological and sectoral characteristics) regions.

If the EU intends to build a "truly European" Research Area in which the networking of "centres of excellence" acts as "catalysts for backward areas" this target may still be far away

JEL Classification: R12, F14, D85, O31

* The authors would like to thank S. Beretta, B. Dettori, M. Nosvelli, M. Riggi, G. Turati, S. Usai for fruitful discussions and M. Gioè for research assistantship. The following people and institutions have been helpful in the process of unpublished data gathering: N. Poupard and P. Vareschi (UK Socrates-Erasmus Council, London), M. Preda (Istituto di studi su popolazione e territorio, Università Cattolica del Sacro Cuore, Milano); S. Usai and B. Dettori (CRENoS, Università degli Studi di Cagliari), M. Thelwall, (School of Computing and Information Technologies, University of Wolverhampton). We acknowledge the financial support of MIUR under the Research Project «Dinamica strutturale: tecnologie, reti, istituzioni» (2003131274_001).

1. Introduction

Aim of this paper is to analyse the impact of knowledge on regional economic development and, consequently on regional disparities across five major European countries: France, Germany, Italy, Spain and United Kingdom.

In particular our focus is on the nature of knowledge not only as a fixed cost in the production process (leading to scale economies), an investment good (influenced by accumulation and depreciation dynamics), and an experience good (whose quality attributes can be detected only upon using, or consuming, the good), but also a "relational" good displaying relevant network externalities.

In this paper we address the manifold nature of knowledge through the analysis of four distinct but complementary phenomena (Internet hyperlinks, European research networks, EPO co-patent applications, Erasmus students mobility) which characterize knowledge as an intrinsic relational structure (directly) connecting people, institutions and (indirectly) regions across five European countries.

Two are the main research questions addressed in the paper: the first deals with the notion of regional disparities, the second refers to the different concepts of distance, namely geographical, functional and sectoral.

Regional disparities can no longer be defined only in terms of statistical differences in the values of standard macroeconomic indicators. Knowledge does matter more and more in defining both the level and the growth rate of a given region GDP (Sapir et al., 2004). For this reason new relational indicators have to be built and compared in order to develop a new kind of (relational) analysis able to complement the traditional "attributional" one.

Traditionally, regional economic disparities have been ascribed to peripherality – measured by the distance from the main centres of population and economic activity – and/or to a high level of dependence on declining sectors (mainly "mature industries").

The scale of regional and other disparities, as well as the political approach and the specific policy instruments used at the European level to deal with this problem, have changed very much over the years. Europe is lagging behind USA in terms of growth and investments in knowledge infrastructures, but this general statement, while true, hides huge variance across European regions and nations (European Commission and DG Enterprises, 2004).

In the last fifteen years income differences among European Member States have been strongly narrowing even if the process has been matched with a widening of the inter-regional variance within single countries (Martin, 1998). All that, evidently, rises a shadow on the whole season of European regional policies, explicitly designed to reduce geographical imbalances and strengthen regional cohesion, and questions on the consequences of the future Europe enlargement as the gap is expected to widen.

A very peculiar and worrying aspect of the European context is that the productive capacity agglomeration process– as a consequence of market forces – may become too strong and risky to be socially unacceptable.

Furthermore the unification process may cause the excessive specialization of European regions, exposing them to a high risk of idiosyncratic supply and demand shocks. It is not by chance that around some 60% of Structural Funds have been used to finance physical infrastructures, which seems to be a 'necessary even if not sufficient' condition to counterbalance existing development gaps.

In addition, at the Lisbon 2000 European Council, the EU set itself the ambitious goal to become "the most competitive and dynamic knowledge-based economy in the world, capable of sustainable economic growth with more and better jobs and greater social cohesion" (European Commission, 2000) and the Council requested the Commission to annually report on the structural indicators of progress in member states towards the Union's strategic goal[1]. All these calls for robust evidence and rigorous monitoring of outcomes, originated the development of a set of comprehensive structural indicators to underpin further analyses.

The diffusion of ICTs and the increase in the knowledge content of goods and services have been often seen as the driving forces behind a major paradigm change in human history: the "death of distance" (see Cairncross, 1997 and Coyle, 1997). After the burst of the "New Economy bubble" many claims of radical changes and revolution have been silenced and economists and geographers have shown that the concepts of distance, space and clustering are still relevant (if not more relevant) in the "Internet era" (Feldman, 2001; Leamer and Storper M., 2001; Redding and Venables, 2004).

However, what may be interesting to study are the different effects of geographical, functional and sectoral distance on the relational activity of different territories. This is exactly the object of the present analysis which looks, within a "gravitational" framework, at four different relational variables (Internet hyperlinks, European research networks, EPO co-patenting applications and Erasmus students flows) between 110

[1] Despite the recent downward revision of the targets set in Lisbon, the EU policy is still informed by the same principles.

European NUTS2 regions located in five European countries: Germany (40 regions), Spain (16), France (22), Italy (20) and United Kingdom (12).

Gravitational models usually include geographical distance (based on geodesic path or road distance) between two areas to capture a series of distance related phenomena which are difficult to measure (such as: transport costs, time elapsed during shipment, synchronisation costs, communication costs, transaction costs, cultural distance).

In this paper we used "geographical" distance, calculated as the shortest road distance existing between two NUTS2 "capitals", but we add two concepts of distance, the "functional" distance, calculated as the difference (in absolute value), between the level of innovative performance of different regions (based on the RSII index contained in the European Innovation Scoreboard) and the "sectoral" distance (based on the sectoral distribution of the patenting activity).

The paper is organised as follow. Sections 2 and 3 describe the variables used in the different analyses; Section 4 presents some descriptive statistics and different types of correlations (Pearson, Spearman and Quadratic Assignment Procedure); Section 5 illustrates the use of Social Network Analysis to detect structural properties of different knowledge exchange flows; Section 6 is devoted to the econometric analysis of two "gravitational" models; Section 7 concludes the paper.

2. Four types of knowledge flows

Krugman, in his *Geography and trade*, stated that "knowledge flows (...) are invisible; they leave no paper trail by which they may be measured and tracked" (Krugman, 1991, p. 53).

Jaffe et al. (1993) reacted to the previous statement by suggesting that "knowledge flows do sometimes leave a paper trail, in the form of citations in patents. Because patents contain detailed geographical information about their inventors, we can examine where these trails actually lead" (ibid, p. 578).

We tried to move the approach a little further by focussing on four knowledge-based relational phenomena: digital information exchange (transmitted through Internet hyperlinks), participation in the same research networks (funded by the EU Fifth Framework Programme), EPO co-patent applications and Erasmus students' exchange flows. Through these variable we attempt to measure the intrinsic relational structure of knowledge flows which directly connect people, institutions and, indirectly, regions across five European countries.

These four variables capture different types of knowledge (spanning from "pure tacit" to "pure codified" knowledge) and different stages of the knowledge creation process. Although ICTs reduce communication and transmission costs, the nature of knowledge and its creation process are very complex and require social processes involving different modalities of interactions. Even in the Internet era face-to-face relations remain crucial (Feldman, 2002)[2].

It is worth nothing that the relational variables considered in the analysis span the entire spectrum of "relational" aspects of knowledge creation, suggesting alternative ways to detect knowledge trail: from new and immaterial way of information exchange (i.e. Internet hyperlinks), to physical and virtual institution-based interactions built to improve knowledge creation (i.e. Research Networks) by exchanging mostly codified knowledge, to physical and virtual individual-based relationships aimed at developing marketable innovations (i.e. co-patents applications) by exchanging mostly tacit knowledge and know-how, to physical movement of people leaving their own region in order to acquire a part of their university education in a foreign institution (Erasmus students exchange).

2.1 Digital information exchange (through Internet hyperlinks)

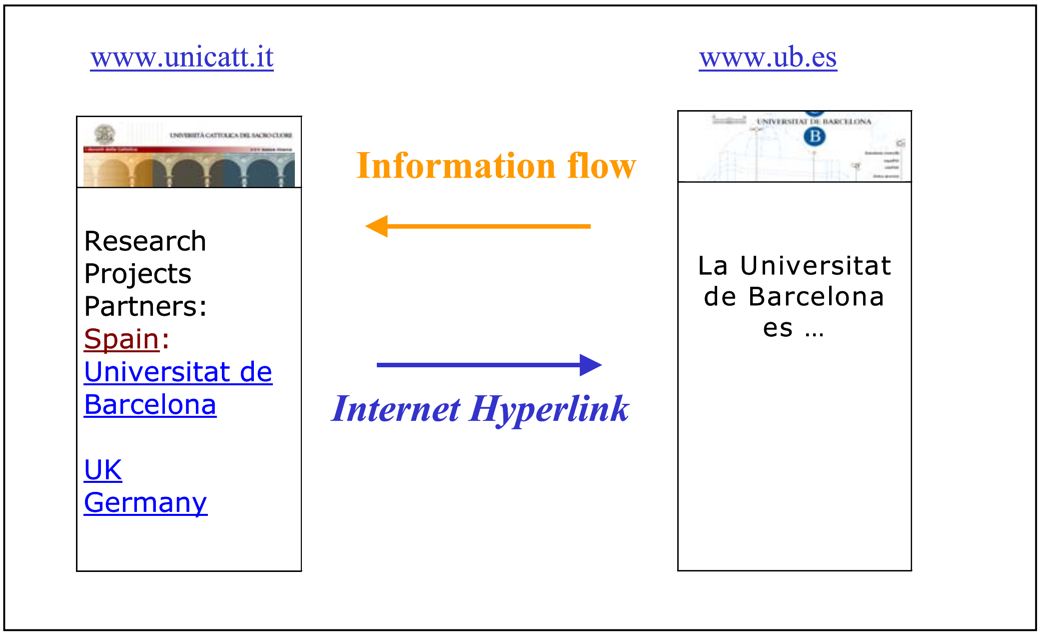
The recent spreading of ICTs, and in particular of the Internet, stimulated several analyses to measure the diffusion of such a phenomenon across countries, regions, cities, ethnic groups, social classes in order to map the current state and to detect the presence of "digital divide".

Different indicators may be used to detect the diffusion of ICTs. The simplest way is to measure the "endowment" of the ICT equipment (i.e. number of Internet hosts, personal computers, broadband connections) and, more in general, all telecom infrastructures allowing efficient connections. A second way concerns the measurement of the "access" conditions to ICTs services, in terms of the market structures (and prevailing pricing strategies) of the relevant markets (telecoms, ISPs etc.). Another way is related to the "use" of ICT, which may be detected by measuring the number of people on-line, to time spent on-line, to size of different on-line activities (e-commerce, e-government etc.).

[2] The value added of scientific workshops and conferences is the transmission of tacit knowledge which is not conveyed in the papers.

A further way concerns the relational nature of the physical infrastructure of the Internet, (made of cables, routers, satellite and radio connections), and of the the www (world wide web), the Internet virtual interface and service platform that allows to visualise and exchange the information. Since the www is a network of web pages linked through Internet hyperlinks, it can be used to map the inner structure of communication channels and to detect the producers and consumers of digital information. When an Internet hyperlink button is clicked, the content of the target web page is transferred to the clicking computer. One may thus think that the web page containing the hyperlink button acts as an importer of digital information and the "target" web page represents the information "exporter" (Figure 1) or, more precisely, think at an Internet hyperlink as an index of revealed comparative advantages in the production of specific types of digital information.

Figure 1: Internet Hyperlinks as imports/exports of digital information and contents



Internet hyperlinks therefore may be used as an indicator of "potential use" [3], since the existence of an hyperlink from a web page to another signals the willingness of the

[3] While it is possible for a single web site to count and map all access, it is extremely difficult to collect this information at a wider scale since it would involve the cooperation of the web masters of all web sites.

"owner" of one web page to import digital information from another web page and increases the probability that the targeted web page is actually accessed[4].

One may argue that the number of Internet hyperlinks in a web page is uninformative since the inclusion of a new hyperlink button is not constraint by a monetary budget since it does not cost a cent, However the presence of buttons within a page is subject to a harder "graphical" budget constraint. Web design handbooks (see, for example, Lynch and Horton, 2002) show that while the number of hyperlinks is a key element in determining the attractiveness of a web page, such attractiveness is a non monotonic function of the number of hyperlinks: it is good to have a few buttons but not too many.

The www has been thoroughly analysed by mathematicians, physicists, information scientists, engineers[5], in order to detect its structure and development laws. Barabasi (2002) argues that the www has a scale-free topology in which a small number of "central" web pages are very popular (are targeted by a huge number of hyperlinks), while the rest of the web is made of peripheral page which are almost unconnected and virtually unknown[6]. Other studies show that different typologies of web pages have different organisational structures. According to Pennock et al. (2002) some types of web pages (i.e. universities' and newspapers' web pages) display a random networks structure in which there are few extremely central and extremely peripheral nodes, while most of the nodes is targeted by a number of hyperlinks around the average[7].

Uberti and Maggioni (2004) analysed the connectivity of different institutions web pages[8] at the regional level and showed that universities are the most active "traders" of digital information.

In this paper we therefore included the number of Internet hyperlinks between 308 Universities[9] web pages located in German, Spanish, French, Italian and UK regions at NUTS2 level[10].

---

[4] A further confirmation of the informational content of Internet hyperlinks relates to the fact that several search engines (and in particular, the popular Google) use hyperlinks counting as ranking criteria of web pages since they consider it a good proxy of the quality and relevance of the web page.

[5] A new discipline (webometrics) devoted to the analysis of the *www* using bibliometric procedures has been created (see Almind and Ingwersen, 1997; Björneborn and Ingwersen, 2001; Russeau, 1997; Thelwall and Smith, 2002), as well as a scientific association (i.e. International Society for Scientometrics and Informetrics, ISSI).

[6] The nodes distribution, according to the number of hyperlinks, follows a power-law distribution, with a degree comprise between 2.1 and 2.72 (Albert and Barabasi, 2001).

[7] The nodes distribution, according to the number of hyperlinks, follows a Poisson distribution (Albert and Barabasi, 2001).

[8] The sample included Internet hyperlinks from different kinds of web sites (i.e. Universities, Local Authorities and Chamber of Commerce).

Since our analysis is devoted to the analysis of information and knowledge flows, we transposed the matrix of the Internet hyperlinks (i.e. the presence of Internet hyperlink from region *j* to country *i*, is analysed as the presence of an information channel flowing in the opposite direction, from region *i* to region *j*).

2.2 Research networks

Since the early 1980s, the EU promoted the creation of research consortia (between firms, universities, research centres and public agencies) in order to increase the competitiveness of the European industry (versus USA and Japan) and to foster intra-European cohesion through the exchange and diffusion of scientific and technological knowledge. In both the Single European Act and the Maastricht Treaty (article 130G) the European institutions have been given competence in the area of science and technology and have developed several actions in order to promote Research and Development (R&D), create European networks, coordinate R&D and stimulate the European mobility of researchers (Breschi and Cusumano, 2002).

Framework Programmes always constituted the main planning instrument and funding source for R&D policies in the European Union, but as time passed, priorities changed: "… the latest programmes have shifted the emphasis from supply-side factors, central in the design of the first policies, to diffusion-oriented projects and the increase of central skills and knowledge among Europeans" (ibidem, p. 6).

In 1998 the European Council and the European Parliament approved the Fifth Framework Programme (5FP), a programme with a different structure from the previous ones, valid for five years (from 1998 to 2002), and financed with about 13 millions euros. This 5FP is divided in 10 thematic and horizontal programmes, and provides for 12 different types of contracts[11].

---

[9] We chose all 308 universities, members of EUA (European University Association), located in Germany, Spain, France, Italy and United Kingdom. The selected sample accounted for 51 university in France, 61 in Germany, 53 in Italy, 45 in Spain, 95 in the UK. However its representation of the total population of European universities largely differs form country to country because of the exclusion of *Hochschulen* and *Ecoles Superiores* which are not members of EUA.

[10] The retrieval of Internet hyperlinks – following Thelwall and Smith (2002) – in 2003 using a public search engine, Altavista.

[11] Cooperative research contracts, coordination of research actions, cost-sharing contracts, demonstration contracts, explanatory awards, explanatory awards (demonstration), explanatory awards (thematic network), preparatory, accompanying and support measures, research grants (individual fellowship), research networks contracts, study contracts, assessment contracts and thematic network contracts. Some of these contracts are assigned to single applicants (i.e. research grants), while some others require the creation of research networks among the participants (i.e. research networks contracts and thematic network contracts).

In this analysis we focus on two contracts – explicitly dedicated to the establishment and use of scientific networks, namely: thematic networks contracts and research network contracts – whose coordinator is a University located in one the 110 regions of the sample and we included all participants located in these regions, irrespective to their typology (i.e. universities, research centres or business organisations).

2.3 Co-patents

Patents (and patent applications) are one of the most established output indicators of innovative activities. Since the seminal contribution of Scherer (1965), patents have been used in the economic literature[12] (Grilliches, 1981, 1990), in order to measure knowledge spillovers and other spatial externality effects which, in contrast to what argued by Krugman (1991), "do leave a paper trail" (Jaffe et al., 1993).

The constitution of the European Patent Office in Munich in 1977 allowed researchers to use a common dataset to analyse the innovative performance of different European countries and regions. In particular Paci and Usai (2000, 2004) and Breschi and Lissoni (2004) have developed systematic analyses of patenting activity throughout Europe at different NUTS levels, showing the existence of significant clustering phenomena (whose agglomeration indexes are even higher than those registered by high-tech manufacturing) within a core-periphery geographical pattern.

Later studies analyse patent data as relational variables. Unlike Breschi and Lissoni (2004) analysis on patent citations, Maggioni and Usai (2005) look at patents as a relation between inventors and applicants at NUTS 2 level and study the distributions of these relationships within different European countries searching for industry-specific patterns and testing the hypotheses of a diffused "brain-drain" dynamics by which peripheral regions do host inventors, but do not exploit the economic outcomes of their scientific and technological creativity since applicants (mostly firms) are located in the core regions.

In this paper we consider another relational aspect of patents: the co-invention process. Out of a total of more than 170900 patent applications belonging to every IPC sections (coming from inventors located in the above mentioned 5 countries in the period 1998-2002) – extracted by the CRENOS files based on the original EPO database – we selected only those patents whose applications were recorded by more than one inventor. Next we split each patent into equal shares attributed to each

[12] Not to forget the wide economic geography and regional science literature.

inventor. We then added these data for each NUTS2 regions in order to built a matrix in which a generic cell *ij* represents the share of patents[13] recorded jointly by inventors located in region *i* and region *j* (where region *i* and region *j* could belong to different nations), to a total amount of around 30000 co-patents.

Co-invention (and thus co-patenting) is a process involving both tacit and codified knowledge exchanges. For this reason it implies a series of both "face to face" and "over the distance" relationships between inventors. That is why it is interesting to analyse the relative importance of "geographic" versus "functional" distance as forces shaping the interregional (international) structure of this knowledge flows network.

2.4 Erasmus students exchange

Erasmus students' exchange represents another relevant part of the spectrum of relational activities involving knowledge flows among European regions: the mobility of tertiary education students, that represents the basic channel for international training and education[14].

The Erasmus programme, introduced in 1988 – and, since 1995, part of Socrates Programme[15] – is a European programme devoted to foster higher education and to create a "European dimension" of education. Its popularity, in terms of students participation, is constantly increasing, and has definitely improved after the Lisbon Council emphasised the enforcement of education and training, and students' mobility as important goals to be achieved.

Erasmus students exchange reflects several important features, equally contributing to "[strengthen] the whole fabric of relations existing between the peoples of Europe" (European Commission, 2005): the "institutional" integration among European countries; the "openness" of national tertiary systems and the "relative attractiveness" of a country, either in terms of its culture and in terms of reputation of its tertiary education system.

[13] A patent registered by three inventors located in three distinct regions *i*, *j*, and *z*, would be split in $n^*(n-1)$ cells and respectively in *i* with *j* and *z*, *j* with *i* and *z*, and *z* with *i* and *j*. Hence a invention co-patented by three individuals in three different regions and is registered with a value of 0.1666 in the cells corresponding to 6 different couplets.

[14] The EU devoted a new programme to post-graduate students exchange (Erasmus-Mundus), started only two years ago and not adequately monitored yet, hence in this analysis we focus on the Erasmus students' flows.

[15] Socrates Programme is the European programme for education, includes eight actions and was developed "to promote the European dimension and to improve the quality of learning by encouraging cooperation between the participants countries" (European Commission, 2005).

As in the case of digital information flowing through hyperlinks, we are interested in the flow of knowledge; hence we consider the region in which the hosting university is localised as the "emitting" region (region *i*) of the knowledge flows embedded in the "learned" students returning to their original "receiving" region (region *j*) after the studies[16].

3. Explicatory (attributional and relational) variables

In this paper we use some attributional variables to detect differences in the knowledge-based characteristics of 110 European regions.

The attributional variables include: GDP, R&D intensity (the ratio between total R&D expenditure and GDP), and three measures of distance (or dissimilarity): the geographical distance (based on road distance between "capitals"), the functional distance (based on the Regional Summary Innovation Index contained in the European Innovation Scoreboard), and the sectoral distance (based on the 2-digit sectoral composition of regional patent application[17]).

Here follows a brief description of variables, their transformations and data sources. Note that throughout the paper subscript *i* refers to "emitting" region and *j* to "receiving" region, while *I* and *J* refer to countries.

GDP_i and **GDP_j**: Gross Domestic Product of region *i* and *j* expressed in purchasing power standards (pps). GDP data, expressed in million of euros refer to year 2000. Source: Eurostat, (2005).

RD_i and **RD_j**: Research and development intensity of region *i* and *j*. It is calculated as the ratio between the regional levels of Gross Expenditure on Research and Development (GERD) and GDP and refers to various years. Source: European Commission and Enterprise DG, (2003).

$GDist_{ij}$: Geographical distances among 110 European regions are calculated according to the shortest road distance (in kilometres) between regional "capitals". The notion of "regional capital" implied the use of a certain degree of arbitrariness since NUTS2 level are administrative meaningful entities in Italy, Germany, Spain and France, but not in the UK. In this last case we used population as a the selecting criteria

[16] The data are a courtesy of the UK Socrates-Erasmus Council, the UK National Agency responsible for the administration of the Erasmus programme in the UK.
[17] Data kindly provided by CRENoS.

to identify the most relevant city (which we called "capital"), irrespective to the presence of an administrative capital[18]. Source: Mapping-tools (2005).

FDist$_{ij}$: Functional distance is measured as the difference (taken in absolute value) of the values registered by the two regions on the Regional Summary Innovation Index (RSII) contained in the EIS, European Innovation Scoreboard. The RSII measures the "European technological leadership" and rank the absolute innovative performance of European regions. RSII is calculated by re-scaling the regional values of the 13 available indicators[19] according to the following formula and then taking the un-weighted average of the re-scaled values per each region:

$$RSII_{jJ} = \sum_{f=1}^{m} \left[\frac{X_{fjJ} - \min(X_{fjJ})}{\max(X_{fjJ}) - \min(X_{fjJ})} \right] \qquad (1)$$

where X_{fjJ} is the value of an indicator *f* for region *j* in country *J*, and *m* is the number of available indicators for the *j* region (European Commission and Enterprise DG, 2003).

This composite index is based on data recorded on different years but is officially referred to 2003. Source: EIS, European Innovation Scoreboard (European Commission and DG Enterprise, 2003).

SDIST$_{ij}$: Sectoral distance is measured as the inverse of the technological nearness index (Moreno, Paci, Usai, 2005) calculated as a correlation coefficient between the sectoral composition of patent application registered by region *i* and by region *j* at EPO in the period 1997-2000. Source: CRENoS.

CONTIG$_{ij}$: Contiguity, or adjacency, is a dummy variables which takes value 1 for contiguous regions (i.e. which share a border), 0 elsewhere[20].

COUNTRY$_{IJ}$: is a dummy variable which is used to control for fixed national effects both on the "emitting" and the "receiving" regions.

[18] For example for Scotland we selected Glasgow instead of Edinburgh.

[19] The RSII indicators are: population with tertiary education, participation in life-long learning, employment in medium-high and high-tech employment; employment in high-tech services; public R&D; business R&D; EPO high-tech patent applications; EPO patent applications; share of innovative enterprises in manufacturing sector and service sector; innovation expenditures in manufacturing and in services; sales of 'new to the firm but not new to market' products.

[20] A regional border in our sample may sometimes be also a national border, and borders are a significant variable in many empirical papers based on the gravitational model. However we did not distinguish these two cases since, with the joint use of the contiguity and country dummies, we are able to identify these cases.

4. Descriptive statistical analysis

4.1 Simple statistics

Table 1 presents some simple statistics on the variables which will be used as regressors in the econometric analysis for the 110 regions of the sample.

Table 1: Some descriptive statistics on attributional data

Variable	Year	min	MAX	Average	Coeff. of Variation	N.
$GDPpps_i$	2000	3210 Valle D'Aosta (IT12)	390162 Ile De France (FR10)	58562.97	0.98	110
$GERD_i$	Various years	0.25 Illes Balears (ES53)	6.21 Braunschweig (DE91)	1.48	0.73	110
$RSII_i$	2003	0.01 Illes Balears (ES53)	0.6 Oberbayen (DE21)	0.27	0.37	110

Source: Eurostat (2005), European Commission and DG Enterprise (2003).

It is interesting to note that total GDP shows, as expected, the largest coefficient of variation (0.98)[21] given the different size (and economic structure) of NUTS2 regions, followed by GERD, with a coefficient of variation equal to 0.73, and finally by the RSII (0.37).

Focussing on the geographical side of the table, Illes Balears records the largest number of minimal values in our sample, followed by Valle d'Aosta. This rank probably due to the smaller size of these regions and to the tourist nature of their economies, together with a high concentration of knowledge based activities in both nations. On the opposite position Germany records the maximal values (Oberbayern and Braunsweig), followed by the French capital region, Ile de France.

[21] Although the value for per capita GDP is much lower (0.27). with a minimal value is registered in Extremadura (9838) and a maximum registered Hamburg (42148).

Table 2: Some descriptive statistics on relational data

Variable	Year	Number of relationships*	Average	Coeff. of Variation	N. of couplets
$Diginfo_{ij}$	2003	548394	53.76	21.35	11990
RN_{ij}	1998-2002	244	0.20	3.65	11990
Pat_{ij}	1998-2002	29712	193.42	15.53	11990
$Eras_{ij**}$	2002	53093	5.76	2.40	9216

* Diginfo relationships refer to number of Internet hyperlinks; RN relationships refer to number of research network and thematic network contracts; PAT relationships refer to the patent applications equivalent number recorded by at least two inventors located in two different regions; Eras relationships refer to the number of students.

** Excluding intra-regional values.

Source: Altavista (2003), Cordis (2005), CRENoS (2005), UK Socrates-Erasmus Council (2004).

Table 2 shows similar statistics for the relational dependent variables. The international digital information flows between universities records the higher coefficient of variation (21.35), followed by international co-patenting (15.53), research networks (3.65) and Erasmus students' flows (2.40). Hence the greatest variation is registered for the "softest" variable, showing the grater diversity in creating digital information and contents among European regions.

These results are confirmed looking at the raw matrices. In the Internet hyperlinks matrix, the most active regions in creating digital information are the British ones, having a natural absolute advantage in the creation and diffusion of information in English (the www *lingua franca*). On the opposite side of the spectrum, many Italian, French and German regions are much more peripheral in the ability to create and export digital information. Spanish regions (in particular Cataluna and Madrid) show interesting features because their centrality in importing digital information is greater than the centrality in one in exporting, confirming a certain degree if openness in creating their digital space.

The geography of research networks shows interesting features: Ile de France is the region most involved in these knowledge flows, followed by all English regions and Scotland. France, Spain, Italy and Germany confirm their traditional intra-national disparities being Rhone-Alpes and Alsace; Cataluna, Madrid and Andalusia; Lombardia, Toscana, Lazio and Piemonte; Oberbayem, Tubingen, and Berlin the most involved regions in research networks.

In the co-patents matrix the most active regions are German (Darmstadt, Oberbayem Düsseldorf, Stuttgart, Karlsruhe, Köln, Rheinhessen-Pfalz), followed by Ile de France and South East; while some peripheral Spanish regions (La Rioja, Extremadura, Galicia, Illes Balers) show minimal figures in co-patenting.

Finally interesting geographical and socio-cultural features characterise the Erasmus matrix: countries with the highest number of received and sent students are Ile de France and Spanish regions (Madrid, Andalucia, Cataluna and Comunidad Valenciana), while the minimal numbers are registered in Southern Italy regions (Basilicata and Molise), Dessen and Corse. These first data may suggest "tourist" and cultural distortions of Erasmus flows (more on this issue in section 6), rather than the attractiveness of education system. It is remarkable to note the asymmetries that exist in UK regions that (except Northern Ireland and the East) are the favourite destinations of students' flows, but record small number of students willing to study abroad.

4.2 Correlations

This section shows some results on the correlation existing between the four dependent variables we used to describes knowledge flows (Internet hyperlinks, research networks, co-patents and Erasmus students flows) used in the econometric exercise. In particular we will focus our attention on simple correlation(Pearson), rank correlation (Spearman) and QAP correlation coefficients.

Table 3: Pearson and Spearman correlations between knowledge flows variables

	Pearson's correlation					Spearman's rank correlation			
	$Diginfo_{ij}$	RN_{ij}	Pat_{ij}	$Eras_{ij}$		$Diginfo_{ij}$	RN_{ij}	Pat_{ij}	$Eras_{ij}$
$Diginfo_{ij}$	1.000				$Diginfo_{ij}$	1.000			
RN_{ij}	0.458	1.000			RN_{ij}	0.313	1.000		
Pat_{ij}	0.167	0.264	1.000		Pat_{ij}	0.276	0.058	1.000	
$Eras_{ij}$	0.331	0.339	0.212	1.000	$Eras_{ij}$	0.322	0.241	0.196	1.000

Table 3 presents both Pearson's simple correlation coefficients and the Spearman's rank correlation coefficients for our relational dependent variables. All coefficients are positive and significant showing that the four variables selected for the analysis measure different sides of the same phenomenon: information and knowledge flows.

Digital information and research networks, on one side, and digital information and Erasmus exchange, on the other side, show the highest correlation coefficients[22]. This may be interpreted as a sign of complementarity between virtual and physical interactions among European universities (and regions). One may also note that the high correlation coefficient between research networks and Erasmus exchange programme shows the existence of hysteresis and lock-in phenomena in the university inter-regional (and international) relationships. Once a relationship is established, both professors and students exploit it

EU attempts to build research networks aimed at producing not only "pure research", but also applied research and marketable innovations seems to be partially successful: in fact the correlation between research networks and co-patenting is quite high, 0.264[23].

We further analysed the relationship among these knowledge flows by using the Quadratic Assignment Procedure (QAP) correlation, a bootstrap method that computes correlation indexes between entries of two square matrices and assesses the frequency of random measures as large as actually observed.

The QAP algorithm proceeds in two steps. In the first step it computes Pearson's correlation coefficients between corresponding cells of the two data matrices. In the second step, it randomly (synchronously) permutes rows and columns of one matrix and re-computes the correlation to the other matrix. The second steps is carried out hundreds of times (in our case: 5000 times) in order to compute the proportion of times that a random measure is larger than or equal to the observed measure calculated in step 1. A low proportion (smaller than 0.05) suggests a strong relationship between the two matrices that is unlikely to have occurred by chance (Borgatti, Everett, Freeman, 2002).

Table 4: QAP correlation between knowledge flows variables

	Pearson's correlation			
	$Diginfo_{ij}$	RN_{ij}	Pat_{ij}	$Eras_{ij}$
$Diginfo_{ij}$	1.000			
RN_{ij}	0.302**	1.000		
	(0.031)			
Pat_{ij}	0.220**	0.062	1.000	
	(0.043)	(0.040)		
$Eras_{ij}$	0.270**	0.245**	0.103	1.000
	(0.038)	(0.036)	(0.048)	

** significant at 5%; standard error in parenthesis

[22] With the exception of Spearman correlation between co-patents and digital information.
[23] The lower value for Spearman correlation could be partially explained by the typology of the research networks considered in this study, which excludes those coordinators that are not university organisation.

Table 4 shows the results of such a procedure[24]. The highest correlation is registered for research networks and Internet hyperlinks (0.302), followed by Erasmus students' flows and Internet hyperlinks (0.270) confirming the simple correlation results and showing the high complementarities between these flows of knowledge.

QAP procedure shows that the correlation between co-patenting and research networks and co-patenting and Erasmus exchange flows (which registered low Spearman and Pearson correlation coefficients) is not significant, indicating the persistence of frictions between different worlds (i.e. business and academic environment). Although the sample of research networks included in the analysis has been heavily selected[25], hence suffers from of some biases, these results may also show that EU programmes seem to fail in connecting different actors, hence these actions need to be redefined to be really effective across different institutions (and in particular between profit and non-profit organizations).

5. Network Analysis of knowledge flows

Network Analysis (henceforth NA) uses quantitative techniques, derived from graph theory, to study and describe the structure of interactions (edges) between given entities (nodes) (Scott, 1991; Wasserman and Faust, 1994). Initially used by sociologists, ethnologists (Nadel, 1957; Coleman, 1964; Mitchell, 1969) to study complex personal interactions, NA has recently been used by economists (Snyder and Kick, 1979; Maggioni, 1993 and 2000; Leoncini, Maggioni, Montresor, 1997; Uberti, 2001; Breschi and Lissoni Cusmano, 2004) to analyse institutional, technological and commercial relationships between agents, industries, regions and countries.

Therefore, in this paper, 110 NUTS2 European regions are treated as nodes, while their different knowledge flows are treated as edges.

Orthodox approaches describe the innovation process through an atomistic principle that assumes the existence of individual utility maximisation procedures and does not take into account the wider social, economic and institutional framework. By contrast, NA highlights some relevant structural features. "The 'behaviour' of a node (in terms of strategy and performance) has to be interpreted in terms of both structural limits and internal features. Internodal relationships must be examined from two complementary

[24] These correlations were calculated using binary matrices dichotomised according to the average of raw matrices: the cell *ij* value above the mean would be registered 1 and 0 otherwise.
[25] See section 2.2.

perspectives: the single node's and the whole system's perspective. Neither a single node nor a pair of nodes can be meaningfully analysed, when isolated from the system framework (holistic principle). Systems display a surprisingly intrinsic fractal nature: both the macro level (whole system) and the micro level (nodes) are composed by a plurality of structurally interrelated elements. The interdependence of observations does not hinder NA techniques, allowing a wider use of this methodology even when more traditional statistical and econometric techniques based on pure attributional variables suffer" (Bramanti and Maggioni, 1997, p. 327).

In the following analysis we will use, as index of systemic connection, the density and the clustering coefficients. Density is defined as the ratio between the actual number of edges e and the maximum number of directed edges in a network composed by n nodes[26] or, in formula:

$$D = \frac{e}{n(n-1)} \qquad (2)$$

The clustering coefficient of node i characterises the extent to which nodes adjacent to it are adjacent to each other. (Watts, 1999).

$$C_i = \frac{\nu_i}{\Lambda}$$

where ν_i is the number of nodes connected to i and Λ is the total number of possible edges in i's neighbourhood. The clustering coefficient for the whole network is obtained by averaging the clustering coefficient of all nodes in the network.

The networks analysed in the paper describe different knowledge flows: co-pantenting and research networks are symmetric, and Internet hyperlinks and Erasmus exchange flows are a-symmetric. In the latter case, for each node an outdegree (number of outward connections) and an indegree (number of inward connections) have been calculated.

Furthermore, NA indexes have been calculated from a dichotomized version[27] of the original intersectoral innovation flow matrices. The customary procedure implies the choice of an 'appropriate' (often ad hoc) threshold; however it must be considered that

---

[26] For symmetric networks with undirected edges, the desity is calculated as follows: $D = 2e/n(n-1)$

[27] A value equal to 1 is substituted to the actual value of the edge when it is greater than or equal to the cut-off; a 0 when the actual value is smaller than the cut-off. The use of valued vs. unvalued networks is widely discussed in the literature (Wasserman and Faust, 1994). In the econometric analysis performed in section 6 networks have been used in their valued (i.e. containing all different numerical values) version, while in this section networks are dicotomized according to their average.

the choice of a given threshold is strategic because different values produce different dicothomised networks. In this analysis we choose the network average as the threshold value.

In order to detect the most central sector within the system and the definition of a scale of hierarchy (inequality), centrality and centralisation indexes have been designed[28]. Formally, the degree centralisation of a network (system) of n nodes (regions) ($\overline{\overline{C}}{}^{i}_{g}$) can be defined as follows:

$$\overline{\overline{C}}{}^{i}_{g} = \frac{\sum_{i}(C^{*}_{g} - C^{i}_{g})}{(n-1)(n-2)} \tag{3}$$

where C^{*}_{g} is the centrality value of the most central region in the system and the denominator reflects the maximum level of centrality obtainable in a system of *n* regions. The centralisation indices (which lies between 0 and 1) measure the difference in centrality levels between the most central region and the other ones. A high centralisation index identifies a very hierarchic system where differences in positions are maximised, and a pivotal node exists. A low centralisation index identifies a structure where most of the positions are similar and interchangeable.

Table 5: Network Analysis indexes of knowledge flows structures

	Density	Clustering	Isolated nodes	Centralization	
				Outdegree	Indegree
Diginfo$_{ij}$	0.111	0.741	33	0.425	0.379
RN$_{ij}$	0.240	0.542	15	0.361	0.361
Pat$_{ij}$	0.140	0.727	22	0.281	0. 281
Eras$_{ij}$	0.199	0.464	3	0.586	0.410

Density indexes of dichotomised networks show that the digital information and the co-patenting networks are the less dense, while the research network is the most dense. The data shows that knowledge flows do not spread evenly between European regions and that traditional face-to-face interactions do remain one of the most active phenomenon of knowledge creation, although virtual ones are cheaper.

The ranking based on clustering coefficient is almost the opposite than the density-based one. This can be explained by referring to the number of isolated nodes. The most

[28] If both degree centrality (for the single node) and centralisation (for the whole system) indexes are used on a directed network, then it must be stressed that inward and outward measures (relative to the inward and outward links of a node) are, in general, not equal. In the paper, therefore, centrality and centralisation indexes - without any further specification - identify the outward measure of the indexes.

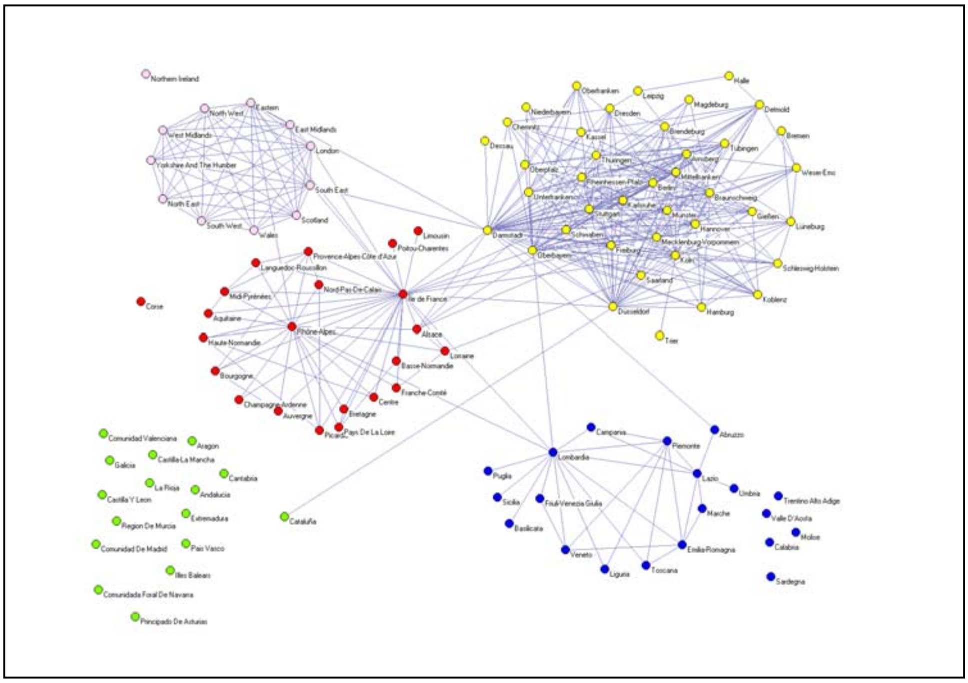
clustered networks (digital information and co-patenting) have lots of isolated nodes (private club structure): meaning that if a region is connected to another one, then it is very likely that the same region is also connected to the original node neighbours.

Centralisation indexes shows that, in general, the Erasmus students' network are the most centralised, while co-patenting is the less centralised network. However – since Internet hyperlinks and Erasmus are asymmetric, while research networks and co-patenting are symmetric – it is more useful to consider each couplets in isolation.

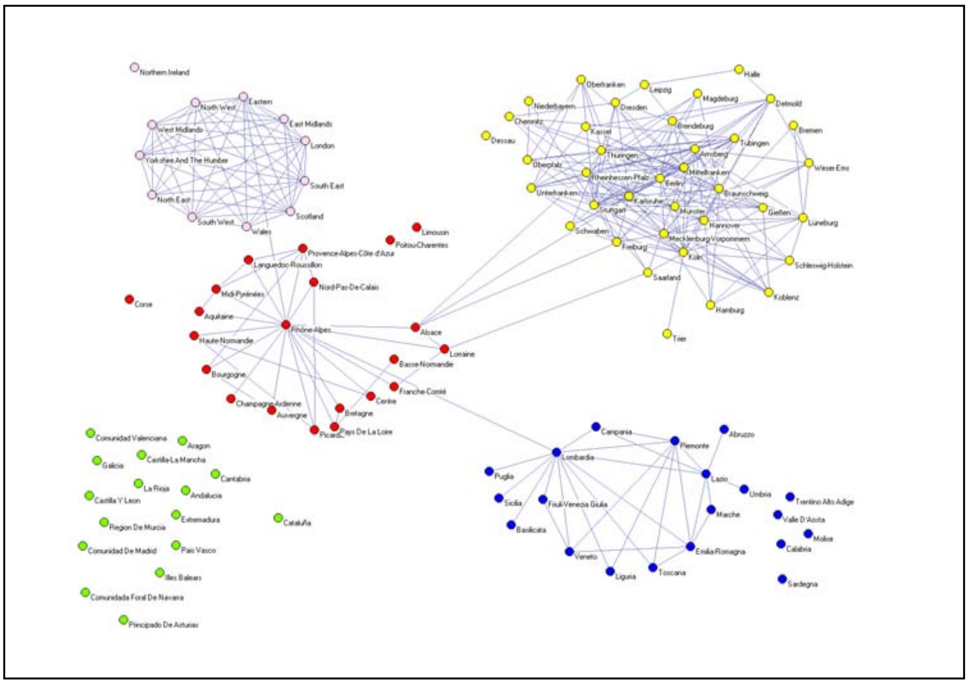
As far as symmetric relationships are concerned: research networks exhibit a more hierarchical regional structure than co-patenting, suggesting that educational institutions are tied for better and worse to the region's performance, while individual inventors are more evenly diffused and they interactions follows a more uniform pattern.

The co-patenting network has a rather non hierarchic structure due to the presence of some very central regions (Oberbayern, Darmstadt, Düsseldorf and Ile de France) and to a series of other regions that are connected not exclusively to the most central ones but also with their national neighbourhood. In fact, by removing the most central nodes from the network, highly connected national "islands" emerge.

Figure 2: Co-patents network 1998-2002



including Oberbayern, Darmstadt, Düsseldorf and Ile de France



excluding Oberbayern, Darmstadt, Düsseldorf and Ile de France

As far as a-symmetric relationships are concerned: Erasmus students flows display a more hierarchical structure (some European regions are highly engaged in the Erasmus Programme either as source or as destination of student flows, while others are almost not involved) than the digital information one (differences in the number of hyperlinks are not so relevant).

The difference in centralisation values (referred to outdegree and indegree) of the Erasmus Programme may be interpreted as the existence of a larger difference in the participation to the Erasmus programme of different European regions as recipient of students than as sender (a greater number of regions send their students abroad, but their destination is concentrated in a smaller number of regions).

A similar (although smaller) difference is shown by the centralisation indexes of the digital information networks. European regions show a greater difference in their information exports than in their information import. In other words while Universities (and regions) are more similar in the number of hyperlinks buttons inserted in their web pages, few universities (and regions) records a larger share of total Internet hyperlinks destinations.

6. Knowledge flows and gravity equations: an econometric exercise

Looking for the sources of regional disparities we thought about using a gravity equation in order to assess whether "geographic distance" was responsible for such a phenomenon (i.e. peripherality exogenously causes poor performances of regions, therefore determining the polarization of a rich core and a depressed periphery) or whether "functional distance" (i.e. difference in the scientific and technological levels) endogenously plays a major role in determining the existence of a much dense core (the network of more advanced regions) and a residual sparse set of relations within the periphery. Finally, we tested the influence of the similarity/dissimilarity of the productive structure of different regions by detecting the effects of "sectoral distance" (measured through patent activity) on knowledge flows.

The gravity equation model is an extremely successful tool of empirical analysis to explain social interactions (for example international trade, foreign direct investment, migration, tourism) according to the existence of "attractive" and "impeding" forces.

This range of models is derived from the "Law of universal gravitation" proposed by Newton in 1687 stating that "gravitational force between masses decreases with the distance between them, according to an inverse-square law. … [T]he theory notes that the greater an object's mass, the greater its gravitational force on another mass" (Wikipedia, 2005).

In the economic literature these models are commonly used to explain international trade: bilateral trade between two countries is proportional to their economic mass (i.e.

GDP or population) and inversely related to their geographical distance. These models are a successful tool for empirical analysis since the '60s: the signs of parameters of importing and exporting countries' GDPs are positive, roughly equal to unity and significant, and the sign of geographical distance is negative and significant (Tinbergen, 1962; Poyhonen, 1963). Recently this empirical success has been theoretically demonstrated (Anderson, 1979; Bergstrand, 1985; Helpman, 1988; Deardorff, 1998; Feenstra, 2002; Dalgin, Mitra, Trindade, 2004).

In order to do so we built a gravitational model which explain the level of a particular type of knowledge flows between two generic regions *i* and *j* as a function of a series of relational and attributional variables. All variables are taken in logs in order to interpret the estimated coefficients as elasticities.

The generic dependent variable, KF_{ij}, stands four different typologies of knowledge flows: digital information ($KF_{ij} = Diginfo_{ij}$), research networks ($KF_{ij} = RN_{ij}$), co-patenting ($KF_{ij} = Pat_{ij}$), or Erasmus students exchange ($KF_{ij} = Eras_{ij}$); the independent variables are as defined in section 3. Table 6 shows the results of 8 OLS regressions[29], where we considered alternatively functional and sectoral distance in the regressors.

$$\ln(KF_{ij}) = \alpha_0 + \alpha_1 \ln(GDP_i) + \alpha_2 \ln(GPD_j) + \alpha_3 \ln(RD_i) + \alpha_4 \ln(RD_j) + \alpha_5 \ln(GDist_{ij}) + \alpha_6 Contig_{ij} + \alpha_7 \ln(FDist_{ij}) + \delta_I + \varepsilon_{ij} \qquad \text{(regression a)}$$

$$\ln(KF_{ij}) = \beta_0 + \beta_1 \ln(GDP_i) + \beta_2 \ln(GPD_j) + \beta_3 \ln(RD_i) + \beta_4 (RD_j) + \beta_5 \ln(GDist_{ij}) + \beta_6 Contig_{ij} + \beta_7 \ln(SDist_{ij}) + \delta_I + \omega_{ij} \qquad \text{(regression b)}$$

where δ_I indicates country dummies variables and ε_{ij} ω_{ij} are standard error terms
Table 6 present the results of the econometric analysis.

[29] Since we are mainly interested in the significance and signs of the coefficient, simple OLS estimation provides valid results. Alternative estimation procedure (either count data models or OLS with box-cox transformation) would allow detailed analysis of the coefficient values.

Table 6 Gravity equation for knowledge flows

	Digital Information Flows $Diginfo_{ij}$		Research Networks RN_{ij}		Co-Patents Pat_{ij}		Erasmus students $Eras_{ij}$	
Independent Variable	(Ia) Functional distance	(Ib) Sectoral Distance	(IIa) Functional Distance	(IIb) Sectoral distance	(IIIa) Functional Distance	(IIIb) Sectoral distance	(IVa) Functional distance	(IV Sect dista
GDP_i	0.761*** (0.021)	0.758*** (0.021)	0.076*** (0.004)	0.075*** (0.004)	0.748*** (0.025)	0.733*** (0.025)	0.621*** (0.018)	0.604 (0.0
GDP_j	0.600*** (0.022)	0.594*** (0.021)	0.076*** (0.004)	0.075*** (0.004)	0.747*** (0.026)	0.733*** (0.025)	0.461*** (0.018)	0.441 (0.0
RD_i	0.240*** (0.026)	0.236*** (0.026)	0.045*** (0.005)	0.041*** (0.005)	0.391*** (0.030)	0.394*** (0.030)	0.065* (0.025)	0.05 (0.0
RD_j	0.035 (0.027)	0.033 (0.026)	0.045*** (0.005)	0.041*** (0.005)	0.392*** (0. 030)	0.394*** (0.030)	0.140*** (0.024)	0.137 (0.0
$GDist_{ij}$	-0.449*** (0.038)	-0.445*** (0.038)	-0.020** (0.006)	-0.024*** (0.007)	-0.525*** (0.043)	-0.580*** (0.039)	-0.008 (0.038)	-0.0 (0.0
$Contig_{ij}$	-0.168** (0.100)	-0.139 (0.098)	-0.011 (0.017)	-0.015 (0.017)	1.098*** (0.074)	0.975*** (0.072)	1.220*** (0.158)	1.16 (0.1
$FDist_{ij}$	-0.051** (0.016)		-0.020*** (0.003)		0.004 (0.018)		-0.033** (0.014)	
$SDist_{ij}$		-0.403** (0.118)		-0.052** (0.017)		-0.520*** (0.135)		-0.67 (0.1
Constant	-7.678*** (0.411)	-7.400*** (0.397)	-1.454*** (0.067)	-1.346*** (0.066)	-10.215*** (0.477)	-9.539*** (0.465)	-8.960*** (0.446)	-9.52 (0.4
Number of observations	6513	6709	11643	11772	4623	4752	5100	51
R-squared	0.539	0.541	0.218	0.217	0.642	0.643	0.359	0.3
F-test	608.92	636.37	109.79	110.37	679.38	695.77	264.54	276

*** significant at 1%; ** significant at 5%; * significant at 10%; robust standard error in parenthesis; country dummies are included in all regressions but not reported.

Regression I describes the structure of information flows running through Internet-hyperlinks established between European Universities. These flows are positively influenced by both regions' GDP, confirming the existence of a positive relation between the "economic size" of a region and its involvement in ICT (in terms of endowments, access and use) which may well lead to "digital divide" phenomenon. Note also the coefficient of the emitting region is slightly larger that that of the receiving one.

A positive and significant coefficient is registered by the R&D intensity of the emitting region, while the coefficient of the "receiving" region is not significant. This suggests that the level of intensity of innovation inputs of a region determines the "visibility" of the local Universities web sites (perhaps via a relationships between public funding of R&D and University relevance[30]), while the positioning of hyperlinks buttons follows a different logic.

Geographical distance is negative and significant suggesting that, at least for our sample of university web sites, the advent of the Internet did not caused the "death of distance"[31]. Digital relationships are considered in academia as complement to physical ones and face-to-face contacts are still crucial. It is however worth noting that the coefficient of the contiguity variable is also negative and significant. Such a result may be explained in terms of a limited use of Internet–based information flows between neighbouring university[32].

Functional distance bears a significant and negative coefficient, thus signalling that university networks of relations as measured by Internet hyperlinks tend to develop between similar regions. A similar result is shown by the measure of sectoral distance, perhaps suggesting the existence of a deep relation between a region industrial structure and the characteristics of its universities.

Regression II analyse the joint participation of research institution belonging to different regions to two different types of research networks under the Fifth EU research framework programme. These flows are positively influenced by both regions' GDP, confirming the existence of a positive relation between the "economic size" of a region, its research potential and its scientific networking activitiy. The R&D intensity

---

[30] In terms of international ranking of its research output.

[31] This is reinforced by the fact that the coefficient of the geographical distance is larger than the coefficient on the functional distance and (slightly) the coefficient of the sectoral distance.

[32] The low level of digital interaction at short-range may be also explained – at least for some European countries, such as Italy and Spain – to some hidden forms of spatial competition on the local pool of perspective students.

coefficients of the emitting and receiving regions are positive and very similar, suggesting that the propensity to be involved in the network is positively correlated with the “scientific and technological level” of both regions.

The coefficient of geographical distance is negative and significant; however its size is very small suggesting a limited influence of spatial effects in this activity whose aim is explicitly to link research units from different places all over Europe. This is confirmed by the insignificance of the coefficient on contiguity.

The coefficients on both functional and sectoral distance are negative and significant: meaning that both the scientific and technological level and the sectoral specialization of a region play a positive role in determining the probability of joining the same research network. In other words research networks have a “club” structure in which similar agents match with similar. If this is the case, than research networks cannot be used as policy tools to support cohesion and inclusiveness since their structure is a “segmented” one in which stronger regions cooperate with stronger, and weaker with weaker. It is also worth noting that the coefficient on the sectoral distance has a higher value than the geographic distance.

Regression III describes the structure of scientific relationships which derives from the exchange of knowledge and know-how between European inventors. Co-patenting relationships are strongly and positively influenced by both regions’ GDP and R&D intensity; the coefficients of both variables are very similar. This confirms that both size (i.e. larger and richer regions have a greater number of patentable inventions) and technological level play an important role in determining the amount of knowledge exchange needed to develop a patentable innovation.

Geographic distances has a significant and negative coefficients. This could be explained in terms of the need for face-to-face contacts in the R&D activity (based on tacit knowledge exchange) leading to a patent application. Since the coefficient of functional distance is not significant, we focus the attention on model b in order to test whether the sectoral distance plays a more relevant role. This is exactly the case: the negative and significant coefficient shows that a common sectoral specialisation of the technological activity of the two regions is important to determine the level of scientific collaboration between inventors.

The positive and significant coefficient on contiguity registered in both specifications (*a* and *b*) confirms Jaffe et al. (1993) results and shows that the innovation process is deeply rooted in a given territory and that knowledge spillovers easily overcome regional borders. The coefficient of the geographic distance is not only larger than the coefficient of the functional distance (not significantly different from zero) but also than the coefficient of the sectoral distance: In the innovation process space does matter.

Regression IV looks at students flows within the Erasmus programme. As already explained, since in this paper we are focussing on knowledge flows, we consider the region in which the "hosting" university is localised as the "emitting" region of the knowledge flows embedded in the "learned" student returning to its original and "receiving" region after the studies.

Regional GDP and R&D intensity coefficients are significant and positive. Larger, richer and technologically advanced regions are more aware of the advantages of an international education process and more involved in this Programme. The coefficients on both functional and sectoral distance are negative and significant, while the coefficient of geographical distance is insignificant. Taken all together this may be interpreted as showing that Erasmus programme does foster the geographical mobility of European students but not as much cohesion and convergence of the scientific level of European regions. Geographical distance does not influence the flows of Erasmus students; however, students from top regions (in terms of their respective RSII) tend to study abroad in "better" foreign regions than their counterparts, coming from bottom regions.

In every model shown in table 6 country dummies – included in the estimation to take into account institutional factors of emitting and the receiving regions which may be determined by national characteristics – record significant coefficients.

The regression constant – in the gravitational models literature - refers to a regional fixed effect which is sometimes interpreted as an indirect measure of remoteness (i.e. the distance of one region to every other regions). If one believes this story, than our results support the common wisdom that peripheral (in geographical sense) regions are also peripheral in a functional sense and that knowledge and information flows have an hierarchically segmented structure with limited evidence of filtering down process.

7. Conclusion

The reduction of regional disparities has been one of the main target of EU policies since its very beginning. However the digital revolution has given new flavour to this concept. Per capita GDP and unemployment rates are still relevant economic indicators but so are knowledge and ICT indicators (in terms of endowments, access and use).

This paper – which focuses on the structure of knowledge flows as measured by four distinct but complementary variables (Internet hyperlinks, research networks, EPO co-patent applications and Erasmus students mobility) – has underlined the intrinsic relational nature of knowledge; it has shown that there exists a positive correlation between knowledge exchange flows and that these flows are influenced by different types of distance: the geographical, the functional and the sectoral one.

NA techniques showed that Erasmus students flows and Internet hyperlinks have a more hierarchical structure in their outdegree than in their indegree. These results confirms the existence of a polarized centre-periphery hierarchy of European regions which is reflected in the structure of knowledge flows[33].

The NA perspective showed that although the co-patents network displays some international relations connecting European regions, co-patenting still remains a mainly intra-national activity, mostly connecting regions in the same nation.

By using a "gravitational" model we demonstrated that, far from the claim of the "death of distance", geographic distance is still relevant for determining the structure of inter-regional knowledge flows.

Functional and sectoral distance play also a crucial role suggesting that knowledge flows easily between similar (according to their scientific, technological and sectoral characteristics) regions.

If the EU intends to build a "truly European" Research Area in which the networking of "centres of excellence" acts as "catalysts for backward areas" this target may still be far away.

[33] A larger number of regions send their students abroad, but their destination is concentrated in a small number of regions; it is easier to be the origin of an hyperlink that to be the target.

References

Adams J. D. (2001). Comparative Localization of Academic and Industrial Spillovers, *NBER Working Paper Series* 8292, Cambridge (Mass).

Albert R., Jeong H., Barabasi A. L. (1999). Diameter of the World Wide Web, *Nature*, vol. 401, pp. 130-131.

Almind T. C., Ingwersen P. (1997). Informetric analyses on the world wide web: Methodological Approaches to «Webometrics», *Journal of Documentation*, vol. 53, n. 4, pp. 404-426.

Anderson J. E. (1979). A theoretical Foundation for the Gravity equations, *American Economic Review*, vol. 69, pp. 106-116.

Anderson J. E., van Wincoop E. (2001). Gravity with gravitas: a solution to the border puzzle. A Theoretical Foundation for the Gravity Equation, *NBER Working Paper Series* 8079, Cambridge (Mass).

Barabasi A. L. (2001). The physics of the Web, *physicsweb*, online http://physicsweb.org/article/world/14/7/9# .

Batagelj V., MrvarA. (2005). *Pajek – Program for Large Network Analysis*, http://vlado.fmf.uni-lj.si/pub/networks/pajek/.

Bellini E., Ottaviano G. I. P., Pinelli D. (2003). The ICT Revolution: Opportunities and Risks for the Mezzogiorno, *FEEM Working Paper 86.2003*, Milano.

Bergtrand J. H. (1985). The Gravity Equation in International Trade: Some Microeconomic Foundations and Empirical evidence, *The Review of Economics and Statistics*, vol. 71, pp. 143-143.

Björneborn L., Ingwersen P. (2001). Perspectives of webometrics, *Scientometrics*, vol. 50, n.1, pp. 65-82.

Borgatti S. P., Everett M. G., Freeman L. C. (2002). *UCINET 6 for Windows: software for Social Network Analyis*, Harvard: Analytic Technologies.

Bramanti A., Maggioni M. A. (1997). The Dynamics of Mileux: the Network Analysis approach, in Ratti R., Bramanti A., Gordon R. (eds.), *The Dynamics of Innovative Regions*, London: Ashgate, pp. 321-343.

Bröcker J., Capello R., Lundquist L., Pütz T., Rouwendal J., Schneekloth N., Spairani A., Spangenberg M., Spiekermann K., Vickerman R. and Wegener M., (2003). *Territorial Impact of EU Transport and TEN Policies: Third Interim Report of Action 2.1.1 of the European Spatial Planning Observation Network,* version August 2003, online http://www.espon.lu/.

Breschi S.,. Lissoni F. (2002). Knowledge Networks from Patent Data: Methodological Issues and Research Targets, in Moed H.F., Glänzel W., Schmoch U. (eds) *Handbook of Quantitative Science and Technology Research,* Dordrecht: Kluwer.

Breschi S., Lissoni F. (2004). Knowledge networks from patent data: Methodological issues and research targets, *CESPRI Working Papers 150*, CESPRI, Milano.

Breschi S., Cusmano L. (2004). Unveiling the texture of a European Research Area: emergence of oligarchic networks under EU Framework Programmes, *International Journal of Technology Management*, vol. 27, pp.747-772.

Cairncross F. (1997). *The Death of Distance: How the Communications Revolution Will Change Our Lives*, Cambridge (Mass): Harvard Business School.

Cambridge Econometrics (1997), *European Regional Prospects 1997*, Cambridge, UK.

Coleman J. S. (1964). *Introduction to mathematical sociology,* New York: The Free Press.

Coyle D. (1997). *The Weightless World: Strategies for Managing the Digital Economy*, Cambridge (Mass): MIT Press.

Christie I., Hepworth M. (2001). Towards the sustainable e-region, in Wilsdon J.(ed.), *Digital Futures: Living in a Dot-ComWorld*, Earthscan, London.

Cooke P. (2002). *Knowledge Economies. Clusters, Learning and Cooperative Advantage*, London: Routledge.

Dalgin M., Mitra D., Trindade V. (2004). Inequality, Nonhomothetic Preferences, and Trade: A Gravity Approach, *NBER Working Paper Series* 10800, Cambridge (Mass).

Deardorff A. (1998). Determinants of bilateral trade: does the gravity work in a Neoclassical world?, in Frankel J. A. (ed.), *The Regionalization Of The World Economy*, Chicago: University of Chicago Press, pp. 7-22.

European Commission (2000). *Lisbon European Council,* March 2000.

European Commission (2005). http://europa.eu.int/comm/education/index_en.html

European Commission, Enterprise DG (2003). *2003 European Innovation Scoreboard: Technical Paper No 3, Regional innovation performances*, online http://trendchart.cordis.lu/.

Eurostat (2005). *Series: Regions*, online http://epp.eurostat.cec.eu.int/.

Feenstra R. C. (2002). Border Effects and the Gravity Equation: Consistent Methods for Estimation, *Scottish Journal of Political Economy,* vol. 49, pp. 491-506.

Feldman M. P. (2002). The Internet Revolution and the Geography of Innovation, *International Social Sciences Journal*, vol. 54, pp. 47-56.

Gillespie A, Richardson R., Cornford J. (2001). Regional Development and the New Economy, *European Investment Bank Papers*, vol. 6, pp. 109-131.

Griliches Z. (1981). Market Value, R&D and Patents, *Economic Letters*, vol. 7, pp. 183-187.

Griliches Z. (1990). Patent Statistics as Economic Indicators: A Survey, *Journal of Economic Literature*, vol. 28, pp. 1661-1707.

Helpman E. (1988). Imperfect competition and international trade: evidence from fourteen industrial countries, in Spence A. M. and Hazard H. A. (eds.), *International competitiveness*, Cambridge: MIT Press, pp. 197-220.

Jaffe A. B., Henderson R., Trajtenberg M. (1993). Geographic Localization Of Knowledge Spillovers As Evidenced By Patent Citations, *Quarterly Journal of Economics*, vol. 108, pp. 557-598.

Krugman P. (1991). *Geography and Trade*, Cambridge (Mass): MIT Press.

Leamer E. E., Storper M. (2001). The economic geography of the Internet age, *NBER Working Paper Series 8450*, Cambridge (Mass.).

Lynch P. J., Horton S. (2002). *Web Style Guide: Basic Design Principles for Creating Web Sites* – 2nd ed., online http://www.webstyleguide.com/.

Leoncini R., Maggioni M. A., Montresor S. (1996). Intersectoral Innovation Flows and National Technological Systems: Network Analysis for Comparing Italy and Germany, *Research Policy*, vol. 25, pp. 415-430.

Maggioni M. A. (1993), *Network Analysis of Regional Industrial Dynamics and Local Economic Policies*, paper presented at the V SASE Conference, New York, 25-28 March..

Maggioni M. A. (2000). "Intersectoral Innovation Flows within and between Nations and Regions: Network Analysis and Systems of Innovation", in Punzo L., Farina F., Fabel O. (eds) *European Economies in Transition*, McMillan, London, pp. 148-173.

Maggioni M. A., Usai S. (2005). Patents as relations: the organisation and the evolution of the innovative activity in two European countries, paper prepared for the *Open Conference on Knowledge and Regional Economic Development*, Barcelona, 9-11 June 2005.

Maignan C., Ottaviano G.I.P., Pinelli D. (2003). ICT, Clusters and Regional Cohesion: a Summary of Theoretical and Empirical Research, *FEEM Working Paper 58.2003*, Milano.

Martin P. (1998). Can Regional Policies Affect Growth and Geography in Europe?, World Economy, vol. 21, pp. 757-774.

McCallum J. (1995). National Borders Matter: Canada-U.S. Regional Trade Patterns, *American Economic Review*, vol. 85, pp. 615-623.

Mitchell J.C. (1969), *Social Networks in Urban Situations*, Manchester University Press, Manchester.

Moreno R., Paci R., Usai S. (2004). Geographical and sectoral clusters of innovation in Europe, *Contributi di Ricerca CRENoS 04/15*, Cagliari.

Moreno R., Paci R., Usai (2005). Spatial spillovers and innovation activity in European regions, *Contributi di Ricerca CRENoS 03/10 (2003)*, forthcoming in *Environment and Planning A*, special issue (2005).

Nadel S. F. (1957). *The Theory of Social Structure,* London: Cohen & West.

Paci R., Battateta E. (2003). Innovation Networks and Knowledge Flows across the European Regions, *Contributi di Ricerca CRENoS 03/13*, Cagliari.

Paci R., Usai S. (2000). Technological Enclaves and Industrial Districts. An Analysis of the Regional Distribution of Innovative Activity in Europe, *Regional Studies*, vol. 34, pp. 97-104.

Pennock D. M., Flake G.W., Lawrence S. et al. (2002). Winners don't take all: Characterizing the competition for links on the web, *Proceedings of the National Academy of Scinces*, vol. 99, pp. 5207-5211.

Poyhonen P. (1963). A tentative model for the volume of trade between countries, *Weltwirtschafliches Archiv* 90, vol. XC, pp. 93-99.

Quah D.T. (2001). ICT clusters in development: Theory and evidence, *EIB Papers*, vol. 6, pp. 85-100.

Redding S., Venables A.J. (2004). Economic geography and international inequality, *Journal of International Economics*, vol. 62, pp. 53-82.

Russeau R. (1997). Sitations: an explanatory study, *Cybermetrics*, vol. 5, on line http://www.cindoc.csic.es/cybermetrics/articles/v1i1p1.html.

Sapir A., Aghion P., Bertola G., Hellwig M., Pisani-Ferry J., Rosati D., Viñals J., Wallace H. (2004). *An Agenda for a Growing Europe. The Sapir Report*, Oxford: Oxford University Press.

Scherer F. M. (1965). Firm Size, Market Structure, Opportunity and the output of patenting inventions, *American Economic Review*, vol. 55, pp. 1097-1125.

Scott J. (1991), *Social Network Analysis,* Sage Publications, London.

Snyder D., Kick E. L. (1979). Structural Position in the World System and Economic Growth, 1995-1970: A Multiple Network Analysis of Transnational Interactions", *American Journal of Sociology*, vol. 84, pp. 1096-1126.

Thelwall M., Smith A. (2002). Interlinking between Asia-Pacific University Web sites, *Scientometrics*, vol. 55, pp. 335–348.

Tingerben J. (1962). *Shaping the World Economy: Suggestions for an International Economic Policy*, New York: The 20th Century Fund.

Uberti T.E. (2002). Flussi commerciali, tecnologici ed informativi: elementi per un'analisi strutturale del processo di globalizzazione, Tesi di Dottorato, Milano.

Uberti T. E., Maggioni M. A. (2004). Infrastrutture ICT e relazionalità potenziale. Un esercizio di "*hyperlinks counting*" a livello sub-nazionale, *Quaderno Diseis n. 0402*, Università Cattolica del Sacro Cuore, Milano: Vita e Pensiero.

Wasserman S., Faust K., (1994). *Social Network Analysis: Methods and Applications,* Cambridge (Mass.): Cambridge University Press.

Wikipedia (2005). "Law of universal gravitation", *Wikipedia*, http://en.wikipedia.org/wiki/Law_of_universal_gravitation.

Appendix: List of NUTS2 regions

Code	NUTS2 region	Capital	N.
DE11	STUTTGART	Stuttgart	1
DE12	KARLSRUHE	Karlsruhe	2
DE13	FREIBURG	Freiburg im Breisgau	3
DE14	TUBINGEN	Tübingen	4
DE21	OBERBAYERN	München	5
DE22	NIEDERBAYERN	Landshut	6
DE23	OBERPFALZ	Regensburg	7
DE24	OBERFRANKEN	Bayreuth	8
DE25	MITTELFRANKEN	Ansbach	9
DE26	UNTERFRANKEN	Würzburg	10
DE27	SCHWABEN	Augsburg	11
DE30	BERLIN	Berlin	12
DE40	BRENDEBURG	Brandenburg	13
DE50	BREMEN	Bremen	14
DE60	HAMBURG	Hamburg	15
DE71	DARMSTADT	Darmstadt	16
DE72	GIESSEN	Gießen	17
DE73	KASSEL	Kassel	18
DE80	MECKLENBURG-VORPOMMERN	Schwerin	19
DE91	BRAUNSCHWEIG	Braunschweig	20
DE92	HANNOVER	Hannover	21
DE93	LUNEBURG	Lüneburg	22
DE94	WESER-EMS	Oldenburg	23
DEA1	DUSSELDORF	Düsseldorf	24
DEA2	KOLN	Köln	25
DEA3	MUNSTER	Münster	26
DEA4	DETMOLD	Detmold	27
DEA5	ARNSBERG	Arnsberg	28
DEB1	KOBLENZ	Koblenz	29
DEB2	TRIER	Trier	30
DEB3	RHEINHESSEN-PFALZ	Neustadt an der Weinstraße	31
DEC0	SAARLAND	Saarbrücken	32
DED1	CHEMNITZ	Chemnitz	33
DED2	DRESDEN	Dresden	34
DED3	LEIPZIG	Leipzig	35
DEE1	DESSAU	Dessau	36
DEE2	HALLE	Halle (Saale)	37
DEE3	MAGDEBURG	Magdeburg	38
DEF0	SCHLESWIG-HOLSTEIN	Kiel	39
DEG0	THURINGEN	Erfurt	40
ES11	GALICIA	Santiago de Compostela	41
ES12	PRINCIPADO DE ASTURIAS	Oviedo	42
ES13	CANTABRIA	Santander	43
ES21	PAIS VASCO	Vitoria-Gasteiz	44
ES22	COMUNIDADA FORAL DE NAVARRA	Pamplona	45
ES23	LA RIOJA	Logrono	46
ES24	ARAGON	Zaragoza	47
ES30	COMUNIDAD DE MADRID	Madrid	48
ES41	CASTILLA Y LEON	Valladolid	49

ES42	CASTILLA-LA MANCHA	Toledo	50
ES43	EXTREMADURA	Merida	51
ES51	CATALUNA	Barcelona	52
ES52	COMUNIDAD VALENCIANA	Valencia	53
ES53	ILLES BALEARS	Palma de Mallorca	54
ES61	ANDALUCIA	Sevilla	55
ES62	REGION DE MURCIA	Murcia	56
FR10	ILE DE FRANCE	Paris	57
FR21	CHAMPAGNE-ARDENNE	Chalons en Champagne	58
FR22	PICARDIE	Amiens	59
FR23	HAUTE-NORMANDIE	Rouen	60
FR24	CENTRE	Orléans	61
FR25	BASSE-NORMANDIE	Caen	62
FR26	BOURGOGNE	Dijon	63
FR30	NORD-PAS-DE-CALAIS	Lille	64
FR41	LORRAINE	Metz	65
FR42	ALSACE	Strasbourg	66
FR43	FRANCHE-COMTÈ	Besançon	67
FR51	PAYS DE LA LOIRE	Nantes	68
FR52	BRETAGNE	Rennes	69
FR53	POITOU-CHARENTES	Poitter	70
FR61	AQUITAINE	Bordeaux	71
FR62	MIDI-PYRÈNÈES	Toulouse	72
FR63	LIMOUSIN	Limoges	73
FR71	RHONE-ALPES	Lyon	74
FR72	AUVERGNE	Clermont-Ferrand	75
FR81	LANGUEDOC-ROUSSILLON	Montpellier	76
FR82	PROVENCE-ALPES-COTE D'AZUR	Marseille	77
FR83	CORSE	Ajaccio	78
IT11	PIEMONTE	Torino	79
IT12	VALLE D'AOSTA	Aosta	80
IT13	LIGURIA	Genova	81
IT20	LOMBARDIA	Milano	82
IT31	TRENTINO ALTO ADIGE	Trento	83
IT32	VENETO	Venezia	84
IT33	FRIULI-VENEZIA GIULIA	Trieste	85
IT40	EMILIA-ROMAGNA	Bologna	86
IT51	TOSCANA	Firenze	87
IT52	UMBRIA	Perugia	88
IT53	MARCHE	Ancona	89
IT60	LAZIO	Roma	90
IT71	ABRUZZO	L'Aquila	91
IT72	MOLISE	Campobasso	92
IT80	CAMPANIA	Napoli	93
IT91	PUGLIA	Bari	94
IT92	BASILICATA	Potenza	95
IT93	CALABRIA	Catanzaro	96
ITA0	SICILIA	Palermo	97
ITB0	SARDEGNA	Cagliari	98
UKC	NORTH EAST	Newcastle	99
UKD	NORTH WEST	Manchester	100
UKE	YORKSHIRE AND THE HUMBER	Leeds	101

UKF	EAST MIDLANDS	Nottingham	102
UKG	WEST MIDLANDS	Birmingham	103
UKH	EASTERN	Norwich	104
UKI	LONDON	London	105
UKJ	SOUTH EAST	Southampton	106
UKK	SOUTH WEST	Bristol	107
UKL	WALES	Cardiff	108
UKM	SCOTLAND	Glasgow	109
UKN	NORTHERN IRELAND	Belfast	110