

Spatial Dispersion of Peering Clusters in the European Internet

Alessio D'Ignazio and Emanuele Giovannetti

January 2006

CWPE 0601

Not to be quoted without permission

Spatial Dispersion of Peering Clusters in the European Internet

Alessio D'Ignazio, University of Cambridge

Emanuele Giovannetti, University of Cambridge, and University of Rome "La
Sapienza"¹.

JEL C21, C25, D85, L86, R12, Z13

Keywords: Internet, Peering, clustering, agglomeration, Networks, IXP

Abstract

We study the role played by geographical distance in the peering decisions between Internet Service Providers. Firstly, we assess whether or not the Internet industry shows clustering in peering; we then concentrate on the dynamics of the agglomeration process by studying the effects of bilateral distance in changing the morphology of existing peering patterns.

Our results show a dominance of random spatial patterns in peering agreements. The sign of the effect of distance on the peering decision, driving the agglomeration/dispersion process, depends, however, on the initial level of clustering. We show that clustered patterns will disperse in the long run.

¹ * We would like to thank Joerge Lepler, Stefanie Sbrillon, Enrico Santarelli, Giancarlo Spagnolo, John Souter (LINX), Chris Fletcher (LINX), Valeria Rossi (MIX) and the participants at the 7th Euro-IX Forum in Prague September 2005 for their help, useful comments and data. The usual disclaimer applies. The authors would also like to acknowledge EU financing through the 6th Framework Project IST-2004-2012 and from the Isaac Newton Trust, Trinity College, University of Cambridge

1 Introduction

The dramatic reduction in communication costs taking place in the last decade induces a necessary reconsideration of both the notion and the role played by geographical distance in the process of agglomeration of the production activities. This is particularly true for digital goods, whose transport costs converged to zero.

In this paper we study the different effects played by geographical distance between providers within the Internet industry. We focus on the interconnection decisions between operators participating at Internet Exchange Points² (IXP) in Europe. The agreements can be classified into two main categories: peering and transit. Transit is a unilateral provider-to-customer relationship, where the customer pays the provider in order to have its traffic packets sent anywhere in the Internet; instead, peering is a bilateral peer-to-peer relationship, where the networks exchange for free traffic directed to their own customers.

Our objective is twofold. Firstly, we assess whether or not the IXP-based Internet industry shows agglomeration in peering. Indeed, early studies show the presence of some form of agglomeration, inherited by previously existing industrial agglomeration for IT products. In order to do so we measure clustering indexes for peering agreements realized by different operators.

These clustering measures depict however static configurations only, a snapshot of spatial autocorrelation of the distribution of peering agreements (we also provide maps of these snapshots as a visual tool to assess clustering).

Our second objective focuses on the dynamics of the agglomeration process by studying the role played by distance in the formation of a peering relation. Indeed, the relationship between the industry structure and the interconnection outcome is bi-directional: the network structure (clustering,

² Internet Exchange Points are specialised organisations where ISPs can connect to exchange their Internet traffic.

dispersion, randomness in the distribution of peering agreements) affects the role of geographical proximity on the formation of peering agreements, which, in turn, once established will redesign the network structure.

From a static point of view, our results consistently show the presence of spatial random pattern in the peering agreements distributions studied, i.e. no agglomeration. This result could be seen as an emergent equilibrium pattern from an underlying network formation game³. Given this result, we next assess the possibility for distance to still play a role in interconnection decisions, and hence to have a dynamic effect on the agglomeration structure. In particular, we do this by estimating a binary model for both the IXPs displaying a random morphology and for the few exceptions showing on the contrary, either a clustered or a dispersed spatial configuration of their peering agreements.

Our analysis shows that the role of distance in affecting the probability of bilateral peering between ISPs is different depending on whether we consider an initially random 'equilibrium' pattern, a clustered or a dispersed one. In the first and third cases the distance is consistently negatively related to peering, suggesting the existence of localized positive externalities expressed in the form of mutual knowledge and reputation effects. On the contrary, in the case of a clustered pattern, the econometric analysis shows a positive relationship between the geographical distance and the likelihood of peering. We interpret this sign change as showing a dominant centrifugal force, the prevalence of the positive effect of geographic differentiation, over the centripetal one due to localized externalities. In the long run, presumably this effect will lead to cluster breaking and the process will converge to the dominant random pattern.

The rest of the paper is organized as follows. Section 2 introduces and motivates the role of agglomeration in the Internet, while Section 3 provides

³ For a survey of this literature see D'Ignazio and Giovannetti (2006).

an overview about the Internet structure and the interconnection agreements. In Section 4 we analyze the morphology of peering within European IXPs; Section 5 complements the results obtained in the previous section by analyzing the determinants of bilateral peering decisions. Finally, section 5 concludes.

2 Agglomeration in the Internet

In one of the earliest contributions of the renewed debate on the geographical agglomeration of economic activity Krugman (1991) identified concentration as 'the most striking feature of the geography of economic activity'. More recently Fujita and Thisse (2002) described agglomeration as the interplay between two forces: localised positive externalities, which have a centripetal effect, and transport costs, which act as centrifugal force.

Since Marshall (1920), agglomeration has been attributed to three forces: a pooled labour market, greater provision of non-traded inputs, and knowledge spillovers. Glaeser, Kallal, Scheinkman and Schleifer (1992) stressed the importance of geographic proximity in defining the extent of knowledge spillovers within firms of a given industry to explain the agglomeration in cities. Following this approach proximity matters since a basic input for firms' activities, tacit knowledge, is assumed to be only transferable through face to face interaction: 'the transfer of information through modern transmission devices requires its organization according to some pre-specified patterns, and only formal information can be codified in this way' (Fujita and Thisse, 2002, p. 172).

This paper focuses on the analysis of the relevance of geography for the Internet industry itself. Is there evidence of agglomeration in peering decisions? Is there any emerging equilibrium morphology?

Previous studies showing some evidence of agglomeration in the Internet are linked to both the Internet usage and previously existing industrial agglomeration. Forman *et al.* (2002) found that Internet usage and access in the U.S. vary across regions and industries. Although Internet use is widespread, not all industries adopt the Internet to enhance computing processes in order to have a competitive advantage (e.g. electronic commerce). Furthermore, rural and smaller urban areas often lag somewhat behind.⁴ Isaksen (2004), and Power and Lundmark (2004) come to the conclusion that there is evidence for industry clusters in ICT related branches. To a certain degree, the location of such clusters may be explained by previously existing industrial agglomerations that have a high demand for ICT and related services. For example, financial services have a strong need for fast and secure network connections, hence important network infrastructure.

As we argued before, the static analysis of the Internet morphology needs to be complemented by a dynamic study of the elements behind it. We achieve this by studying the effects of proximity for strategic decisions by Internet operators. Indeed, the Internet morphology is likely to affect interconnection decisions, which, in turn, affect the morphology itself. The key question hence becomes: does distance play a role in determining the interconnection decisions?

The role of proximity and face-to-face contact becomes more important as production processes become more fragmented and as firms have to rely on incomplete contracts, thus highlighting the importance of mutual trust (Spagnolo, 1999). For example, Learner and Storper (2001) stress the relevance of face-to-face relations for the establishment of trust. This is particularly important for the Internet, where there is dominant asymmetric

⁴ This so called "digital divide" refers to the fact that Internet access and use is distributed unevenly over social groups and geographic regions (see e.g. Warf, 2001).

information involving the Operators performances and commitment. As discussed below, in the peering decision between ISPs there are substantial aspects of the transaction that are impossible to measure or monitor, so that the peering decision may require substantial trust and informal cooperation between peering partners. This may activate the centripetal force discussed above: face-to-face meetings and social connections may facilitate the governance of peering agreements, and the former may benefit from geographical proximity.

It is important to notice, however, that the face-to-face interaction is not the only, and not necessarily the best, way to transmit tacit knowledge or to enforce exchanges. Interaction can instead be developing in places, not spaces, whereby a place is a relational structure providing identification for the individuals belonging to it, hence it is characterized by 'insiderness' (Relph 1976, Place and Placeness from Dodge and Kitchin, 2000). A place, following this view is not linked to a specific geographical location since the insiderness of, for instance, an online community will in fact define its borders, through affinity and cultural identification, dimensions, these, not often drawn in geographic space, particularly so if the place considered lies in cyberspace.

To become a sustainable alternative to geographically defined districts, the virtual ones will have to develop the ability to establish, maintain and verify reputation and trust. Hence the relevant issue in understanding the possible emergence of virtual districts becomes: 'do new technologies provide the means for the emergence of conventions necessary to facilitate trust in cyber mediated exchanges?' Online places have been historically characterized by behavioral codes also called netiquette, the breaking of which has often disruptive consequences on the deviant's reputation within the community. If there is competition between geographical and virtual districts, their relative competitive advantage will depend on whether the monitoring of these codes is easier through geographical proximity or via

online interaction, and on whether the ensuing necessity of a credible retaliation of a deviant's behavior is more easily implemented within an online connected community or in a geographically clustered one (Giovannetti, Neuhoff, Spagnolo 2006). These elements taken together should drive the agglomeration/polarization dynamics in the specific industry under study, defining the shape and borders, if existing, of the geographical or virtual distribution of the industry⁵.

3 The Internet

The Internet is composed of many independent networks of very different sizes, located around the globe, all directly or indirectly interconnected with each other. This last feature guarantees the Internet's most important property: universal exchange of traffic between all end users (universal connectivity). The industry is still mainly unregulated, and networks are left completely free to decide where, how and with whom to interconnect. Lacking a really dominant network, competitive forces and positive network externalities have been sufficient until now to keep all the networks interconnected.

Small Internet Service Providers (ISPs) rely on interconnections both among themselves and to larger networks for the delivery of their customers' data packets to their destinations outside the range of the ISP's own subscribers. The largest networks are called Backbones. These own or lease national or international high-speed fiber optic networks and deliver packets around the world for the many smaller networks connected to them.

⁵ The issue of the survival of geographical agglomeration when ICT becomes an efficient substitute of face-to-face dealings has been addressed by Santarelli (2004). In a panel data analysis of the long term evolution of Emilia Romagna's industrial districts, he found that spatial concentration is no longer the most crucial factor in agglomeration and the term "multi-located" district describes recent forms of industrial agglomeration in a better way.

3.1 Interconnection Agreements

Two simple types of interconnection agreements have emerged to regulate traffic at exchange points between networks: transit agreements and peering agreements. In a transit agreement, a large network – the transit provider - offers access to the entire Internet to a smaller customer network against the payment of a fee often related to the capacity of the connection link.

Under a peering agreement two networks exchange the traffic directed to each other's end users only. Peering can be seen as a reciprocal, non-monetary exchange relationship that often implies various forms of cooperation. Peering, when taking place privately, implies establishing direct exchange points between the two networks, with the costs of creating and maintaining the exchange points typically shared evenly. Peering agreements may also take place at Internet Exchange Points (IXP), specialized facilities and organizations where ISPs can connect to each other to exchange Internet traffic. To peer at an IXP, an ISP usually has to establish a connection and pay a membership fee, then it can use the circuit to exchange traffic with all other members of the IXP willing to peer with it. This multiple peering possibility at a single location makes peering at an IXP cheaper than establishing multiple direct bilateral peering exchanges each requiring a single physical connection. Being a member of an IXP also offers further advantages to an ISP: sharing of information and a free mutual technical help forum, possible elements towards forming insiderness of the virtual community of the Exchange members. However, it is important to recall that, for an ISP, being a member of an IXP does not imply also being able to peer with all other members. Often, many ISPs are refused bilateral peering by other members of the same Exchange. This refusal to peer also implies that only a partial bilateral connectivity structure takes place at an IXP. This is represented with a peering matrix, displaying a value equal to one when two members peer and zero when they do not.

In the following we focus on the specific bilateral peering decisions between ISPs at IXPs, trying to understand their main drivers and focussing on the question of whether or not geographical proximity among the ISPs is affecting their peering choices at the Internet Exchange Points.

3.2 The Peering Decision

Earlier work has identified several factors and problems that may affect networks' decision whether and with whom to peer. A first, rather obvious factor is size. Peering requires establishing bilateral traffic exchange points, or peering points, which entail fixed and variable technological costs. It follows that a sufficiently intense traffic flow between the end users of the two networks is a necessary precondition for peering to be economically viable.

The larger two networks are, the more intense will be the traffic between their end users, and therefore networks' size is necessarily a determinant of the peering decision. In fact, almost all large backbone networks peer with each other, the traffic being exchanged at several interconnection points homogeneously distributed on the relevant geographical areas. Somewhat smaller networks also peer with networks of comparable size, but typically have to supplement their interconnection with transit agreements with backbone networks. Since the costs of setting up and maintaining peering points are usually shared equally by peering networks, unbalanced traffic implies an unbalanced distribution of gains from peering against a balanced distribution of costs. Such unbalanced situations have developed in some cases, and have led to the discontinuation of the peering arrangement and to its replacement with a transit one. Finally when two networks are peering and one of them is congested, the perceived speed of connection would not improve were the non-congested network to upgrade its infrastructure. And if the congested network chooses not to upgrade its infrastructure, it enjoys

the full cost savings while it shares the reduced performance with all the networks it is peering with. This problem may of course induce caution in networks' decision whether and with whom to peer.

3.3 Agglomeration in Peering?

Little is known about the potential effects of ISPs' geographical location on their peering decision, the focus of our empirical analysis. Should we expect the geographical location of different ISPs to influence their peering decision? Of course, if two ISPs are very far away building a connection from scratch would be very costly; hence one would expect that very far ISPs would not peer. However, consider a situation where there is an IXP where peering is cheaper, and that there is a number of ISPs all of which are already connected to this IXP. Should we then expect the geographical location of these ISPs to matter in their choice of peering partners? Should agglomeration patterns be observed in the peering decision? The centrifugal force discussed before, softening competition through local differentiation, would not be active in this case, since the decision to peer at the IXP is independent from the location choice of the ISPs with respect to end users.

Some centripetal forces considered in the literature, such as knowledge spillovers obtained through interactions with peers, may be moderately active; while transport costs do not matter here since we consider a population of ISPs that are already connected to a given IXP and their only decision is whether or not to switch on an interface at an IXP were they are already located. As discussed above, many features of a peering agreement are not directly monitorable, not to say verifiable/contractible. Hence peering agreements may require a great deal of trust and informal cooperation, in which case face-to-face can be important.

4 Clustering in the European Internet

In this section we analyze the strength of agglomerating forces on the peering decision in Europe. In particular, for each European IXP we study the distribution of the number of peering agreements signed by its members. The analysis, which involves the estimation of spatial autocorrelation indexes, is complemented with a series of maps.

4.1 Spatial Autocorrelation and the Moran's I Index

The Moran's I index (Moran, 1948) is often used to test the hypothesis of no-clustering for spatially distributed variables. This index, actually measuring spatial autocorrelation, is calculated by taking into account the value assumed by the variable under analysis at different locations. In particular, let N denote the total number of observations, let z_i be the value that the variable takes at location i , let μ be its average and let w_{ij} be elements of a spatial weights matrix. Then the Moran's index is given by

$$I = (N/S_0) \sum_i \sum_j w_{ij} (x_i - \mu)(x_j - \mu) / \sum_j (x_j - \mu)^2$$

where S_0 is a normalizing factor given by $S_0 = \sum_i \sum_j w_{ij}$

The inference is based upon the analysis of the standardized z -value* of the Moran's index: this is obtained by subtracting its expected value under the hypothesis of no spatial autocorrelation (in this case the expectation of I is $1/(N-1)$, see Upton and Fingleton⁶ (1985) and Anselin (1992)) from the I statistic and dividing the result by the observed standard deviation. The z statistic is employed to test the hypothesis of no spatial correlation. Moreover the z test also indicates the sign of the geographical clustering, if any, (a

positive value for the z statistic suggests positive spatial autocorrelation - clustered outcome - while a negative value for the z statistic suggests a dispersed one).

The following table 1, reports the spatial autocorrelation statistics (Observed Moran's index, Theoretical Moran's index under the hypothesis of no spatial autocorrelation, Variance and Standardised Moran's index) for each of the IXP members of the European Internet Exchange Association (Euro-IX, <http://www.euro-ix.net/>). Table 1 shows a very strong regularity across IXPs: the variable "number of peering agreements for each ISP" indeed follows almost everywhere a random geographical distribution. This, result can be interpreted as an equilibrium morphology. A noticeable exception is represented by the case of Vienna Internet eXchange (VIX), whose distribution results strongly clustered. Other exceptions are represented by the cases of España Internet Exchange (ESPANIX) and Romanian Network for Internet eXchange (RONIX), both showing some degree of dispersion.

⁶ We would like to thank Dr Fingleton for helpful suggestions on this topic.

Table 1: Spatial autocorrelation statistics for peering agreements within the
Euro-IX Members

IXP	Moran's <i>I</i> index	Location	Theoretical <i>I</i> index under the hypothesis of no spatial autocorrelation			Outcome
			Variance	<i>z</i> -value		
AIX	-0.0912	Athens, Greece	-0.07692	0.01087	-0.1369	random
AMS-IX*	-0.008	Amsterdam, Netherlands	-0.00625	0.00012	-0.1569	random
BIX	-0.0306	Budapest, Hungary	-0.02222	0.00147	-0.2196	random
BCIX	-0.0564	Berlin, Germany	-0.07143	0.00501	0.21302	random
BNIX	-0.0425	Brussels, Belgium	-0.02632	0.00209	-0.3553	random
CATNIX	-0.1197	Barcelona, Spain	-0.09091	0.01248	-0.2579	random
CIXP	-0.1005	Geneva, Switzerland	-0.05556	0.00971	-0.4559	random
DE-CIX*	-0.007	Frankfurt, Germany	-0.00794	0.00022	0.06605	random
ESPANIX	-0.1777	Madrid, Spain	-0.04	0.00485	-1.9773	dispersed
FICIX	-0.0673	Helsinki, Finland	-0.07692	0.02584	0.05993	random
GIGAPIX	0.01981	Lisbon, Portugal	-0.07143	0.01928	0.65701	random
GN-IX	-0.1898	Groningen, Netherlands	-0.125	0.03914	-0.3274	random
INEX	-0.0681	Dublin, Ireland	-0.2	0.06038	0.53685	random
LIPEX	-0.0626	London, UK	-0.025	0.00108	-1.1438	random
LIX	-0.1216	Luxembourg	-0.11111	0.01544	-0.0847	random
LINX*	0.00556	London, UK	-0.0068	0.00019	0.89601	random
LONAP	-0.0251	London, UK	-0.02857	0.00163	0.08683	random
MADIX	-0.2096	Manchester, UK	-0.25	0.08067	0.14229	random
MIX	0.01394	Milan, Italy	-0.01754	0.00087	1.06691	random
MSK-IX	-0.0202	Moscow, Russia	-0.00952	0.00025	-0.6725	random
NAMEX	-0.1527	Rome, Italy	-0.07692	0.00816	-0.839	random
NDIX	-0.1956	Enschede, Netherlands	-0.25	0.03283	0.30044	random
NETNOD	-0.043	Stockholm, Sweden	-0.02326	0.00138	-0.5323	random
NIX-CZ	-0.0292	Prague, Czech Republic	-0.02857	0.00253	-0.0131	random
NIX	-0.0329	Oslo, Norway	-0.02439	0.00154	-0.2162	random
PARIX	-0.0136	Paris, France	-0.03448	0.00762	0.23892	random
RONIX	-0.22	Bucharest, Romania	-0.05	0.00881	-1.8113	dispersed
TIX	-0.0267	Zurich, Switzerland	-0.02222	0.00157	-0.1141	random
TOPIX	-0.077	Torino, Italy	-0.09091	0.00894	0.1466	random
VIX	0.07147	Vienna, Austria	-0.01389	0.00061	3.45367	clustered
XCHANGEPOINT	-0.0407	London, UK	-0.01266	0.00053	-1.2253	random

In the following we focus on a subset of IXPs. Among those characterized by a random pattern of peering agreements, we selected three particularly relevant IXPs looking at both their geographical location and the number of participants for each: they are the London Internet Exchange Point (LINX), the Amsterdam Internet Exchange (AMS-IX), and the Deutsche Commercial Internet Exchange (DE-CIX), located in Frankfurt. We also focus on the few IXPs that exhibit ‘non equilibrium’ spatial distribution in peering: these are the VIX (Vienna), showing a clustered pattern of agreements, and the ESPANIX (Madrid) and the RONIX (Bucharest), displaying, on the contrary a dispersed peering morphology. For these IXPs we provide the relevant map of peering agreements distributions.

AMSIX is one of the largest IXPs in Europe with its 225 members. Figure 1 below synthesizes the characteristics of AMSIX members (geographical position and number of peering agreements reached) and the features of the AMSIX peering matrix. Each ISP is represented by a bar, geographically positioned at the location of its headquarter, whose height directly depends upon the number of peering agreements signed by that ISP.

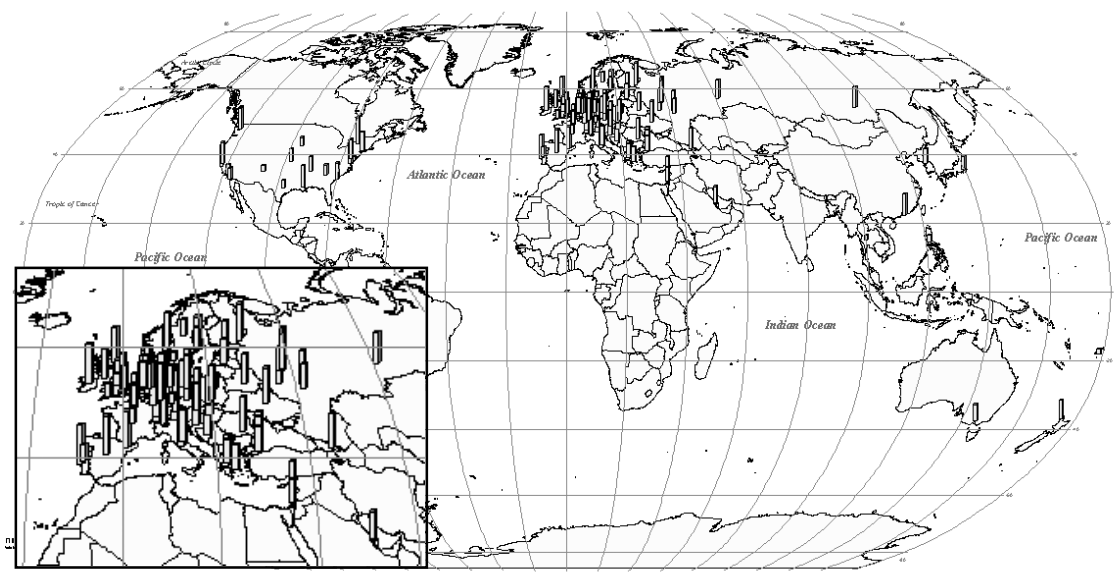


Figure 1: Peering geographical distribution at the AMS-IX (random)

No clustering characterizes the number of peering agreements reached; in fact, the heights of the bars seem to be random and not depending upon the geographical position within the area of interest. This conclusion is also supported by the standardized Moran's 'I' statistic of spatial autocorrelation⁷ of -0.157 (see Table 1 above). There is thus strong evidence in support of the claim that the distribution of peering agreements follows a random geographical pattern.

With 182 and 165 members respectively, both LINX and DE-CIX are among the largest IXPs in Europe, with the LINX also being the largest in terms of Internet traffic exchanged. As for the AMS-IX, also at both LINX and DE-CIX the geographical distribution of peering agreements seems not to be influenced by geographical location of the ISPs. The randomness characterizing the geographical distribution of peering agreements is supported also by the standardized Moran's Index, given by 0.89 and -0.066 , respectively.

A very different pattern characterizes the VIX, (the main Austria's IXP, having 87 members) resulting in a clustered structure, and both the EXPANIX (28 members) and the RONIX, (25 members) both having a dispersed structure.

⁷ As described in paragraph 4.1, the significance of spatial autocorrelation is tested through a z-test on the standardised values of the Moran's *I* Index. The sign of the standardised coefficient also indicates the direction of the autocorrelation (dispersed, if negative, and clustered, if positive).

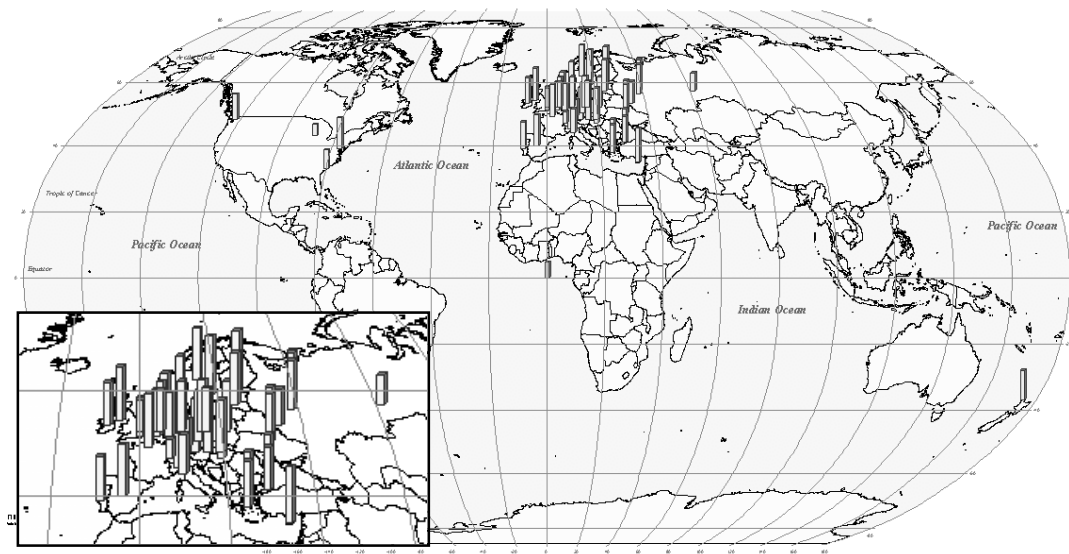


Figure 2: Peering geographical distribution at the VIX (clustered)

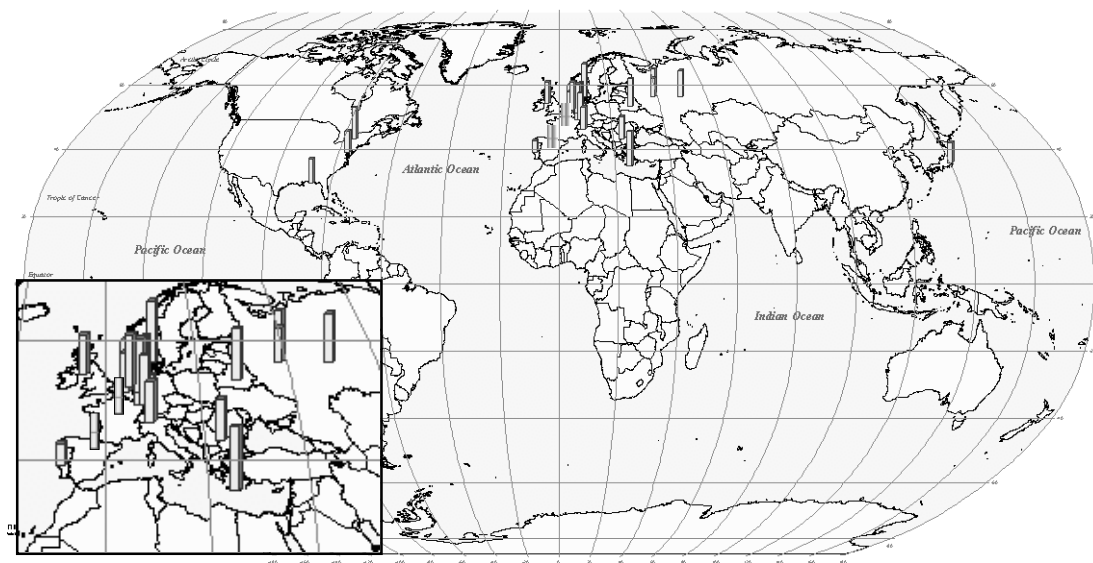


Figure 3: Peering geographical distribution at the ESPANIX (dispersed)

5 An econometric analysis of the role of proximity effects on peering

In the previous section we showed that the European peering within IXPs consistently follow a random distribution, characterising 27 IXPs over a total of 31. The randomness in the peering agreements distribution can be interpreted as an equilibrium state, where no agglomeration occurs, with the only exception given by the VIX (clustered) and the ESPANIX and RONIX (dispersed). In this section we extend the analysis to study whether or not the geographical proximity still plays a role in explaining bilateral peering choices. We differentiate between IXPs in the equilibrium state (again, we focus on AMSIX, DE-CIX and LINX) from IXPs out of the equilibrium state (VIX, ESPANIX and RONIX). This analysis is relevant also from a dynamic point of view. If a clear relationship between proximity and peering decisions exists, then the relationship between the morphology and the peering decisions becomes bilateral: the morphology affects the interconnection decision, which, in turn, affects the morphology.

The pictures below represent the actual peering interconnection within LINX and DE-CIX: the providers are again represented by bars whose height depends on the number of peering agreements signed. The peers are joined by a line.

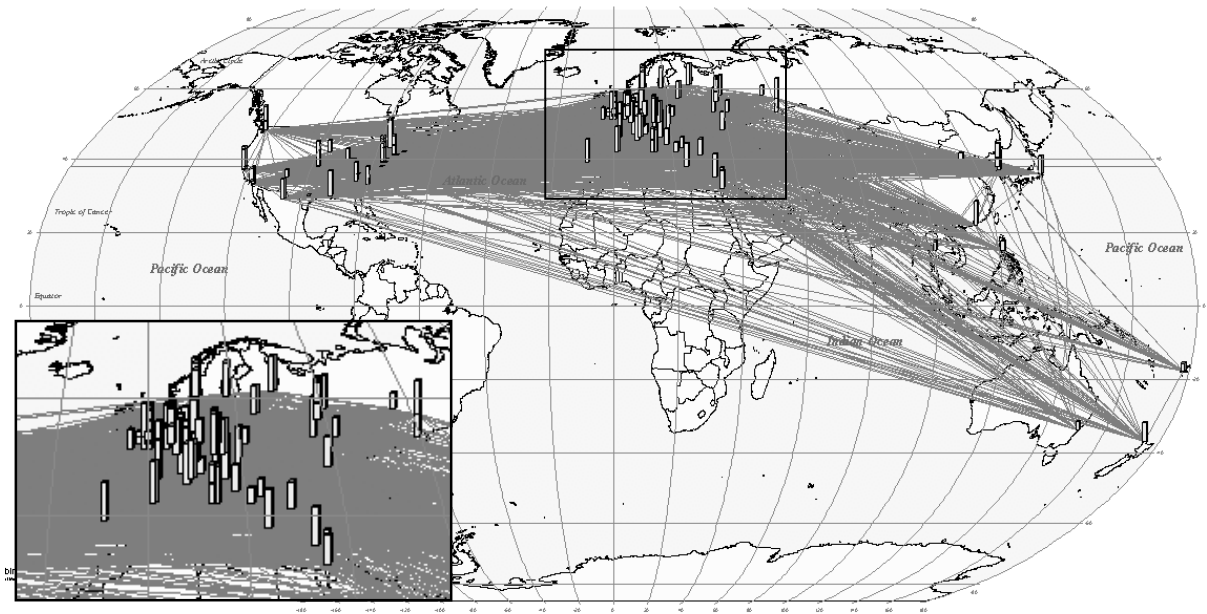


Figure 4: Peering geographical distribution at the LINX

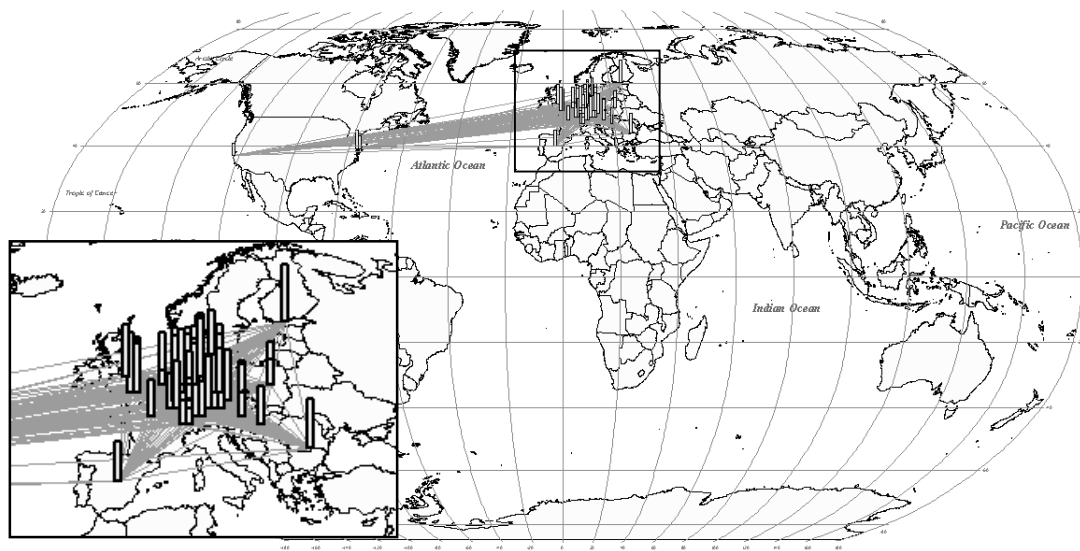


Figure 5: Peering geographical distribution at the DE-CIX

Four possible configurations may emerge from joint consideration of the IXP clustering and their dynamic agglomeration through bilateral; peering choices:

1. An IXP is clustered and distance has negative sign on peering then we can say that yes, the econometric analysis confirms that it is more likely to observe a clustered structure since further away ISPs do not tend to peer.
2. An IXP is clustered and distance has positive sign we can say that since the distributions of peering agreements per locality/headquarters is concentrated around a locality, headquarter distance has a positive added value for bilateral peering.
3. An IXP has spatially dispersed peering agreements and the sign of distance is negative, implies that while headquarters prefer localised peering agreements probably these are among neighbouring ISPs, so these may be scattered in peering groups.
4. Finally an IXP has spatially dispersed peering agreements and the sign for the distance in the regression is positive, this shows an industry where activity is dispersed but connected through valuable peering links across long distance

Each of these 4 possibilities might suggest a graph representation: connected dispersed clusters, unconnected dispersed clusters connected clustered clusters, unconnected connected clusters.

5.1 Empirical Specification of the Probit Model

We now explore these possible scenarios by specifying a binary probit model. The dependent variable, the peering decision, is obtained from the IXP peering matrix, showing, in correspondence of each pair of providers, if either they are in a peering relationship (peering=1) or not (peering=0).

We introduce several independent variables to explain the peering decision. The influence of the geographic location on the peering decision, involving each pair of providers, is modeled by the geographic distance

between them⁸. Moreover for each Internet provider we derived the set of Euro-IX members at which it participates; this information was used to devise a second set of variables. Firstly, the possibility of reputation effects in peering decisions (Titley, 1997) and the technical element of the hot potato routing are expressed by a variable indicating, for each pair of providers, how many IXPs they are both members of⁹. Secondly, we introduce a variable to model the asymmetry in the providers' size by considering the difference in the number of Euro-IX IXPs memberships every ISP has

Finally, a variable to model traffic imbalances is devised. Although traffic flows are kept confidential, it is possible to determine a proxy for traffic imbalances by looking at the publicly available Border Gateway Protocol (BGP) routing tables¹⁰. In particular, we calculated the following measure of betweenness centrality (Shimbel, 1953) for each Internet operator \underline{v} :

$$B_s(\underline{v}) = \sum_{s \neq v \neq t \in V} \sigma_{st}(\underline{v})$$

where $\sigma_{st}(\underline{v}) = \sigma_{ts}(\underline{v})$ is the number of shortest BGP paths from the Internet Operator s to the Operator t on which the \underline{v} lies on. High betweenness for \underline{v} indicates that presumably a relevant quantity of traffic flows among that node.

⁸ We followed a two stage process to calculate the distance: we first individuated latitude and longitude for each ISP's headquarter, then we used the "Great Circle Distance Formula using decimal degrees" to calculate the distance between any couple of headquarters. The formula is given by $\text{dist}(P1, P2) = 3963.0 * \arccos[\sin(\text{lat1}/57.2958) * \sin(\text{lat2}/57.2958) + \cos(\text{lat1}/57.2958) * \cos(\text{lat2}/57.2958) * \cos(\text{lon2}/57.2958 - \text{lon1}/57.2958)]$

⁹ In order to obtain the relevant variables from the initial set of data, given by the peering matrixes for the Euro-IX members, we created several visual basic routines.

¹⁰ The Border Gateway Protocol (BGP) is a series of "instructions" that govern the transmission of packets over the Internet. The BGP establishes the paths that data packets will take through connected networks. The BGP is itself data, and by design nearly always take the same paths; this method (in-band

Table 2: Probit Regression Model Variables

dependent variable	
PEERING (dummy)	Assumes value 1 in case of peering between providers, 0 otherwise.
independent variables	
DIST_ISPS	distance between each couple of ISPs (in thousands of miles)
COMMON_IXPS	number of European IXPs in which the ISPs are both present
DIFF_IXP_MEMBER	difference in the number of memberships in European IXPs
DIFF_RELEVANCE	difference in the betweenness value (in hundreds thousands of units)

transmission) avoids the introduction of new false positive routing information: a non-existent link can not be traversed by routing data (Woodcock, 2002).

The Probit Model

Let y be a binary variable, \mathbf{x} be a $(1 \times K)$ vector of explanatory variables and $\boldsymbol{\beta}$ be a $(K \times 1)$ vector of unknown parameters. The probit model is derived from the following underlying latent variable model:

$$y^* = \mathbf{x}\boldsymbol{\beta} + e, \quad y = 1[y^* > 0]$$

where e has a standard normal distribution¹ and is independent of the explanatory variables \mathbf{x} . Let Φ be its cumulative density function (cdf); then, it can be shown that

$$P(y = 1 | \mathbf{x}) = p(\mathbf{x}) = \Phi(\mathbf{x}\boldsymbol{\beta})$$

We are interested in the effect of the explanatory variable x_k on the above response probability. In such a model, this effect is not entirely determined by β_j : it also depends on the values assumed by the explanatory variables \mathbf{x} . In particular, if x_j is continuous,

$$\frac{\partial p(\mathbf{x})}{\partial x_j} = \phi(\mathbf{x}\boldsymbol{\beta})\beta_j$$

where $\phi(z) = \frac{d\Phi}{dz}(z)$ is the standard normal density function. Instead, if x_k is discrete the partial effect of x_k going from c_k to $c_k + 1$ is given by

$$\Phi[\beta_1 + \dots + \beta_{k-1}x_{k-1} + \beta_k(c_k + 1)] - \Phi[\beta_1 + \dots + \beta_{k-1}x_{k-1} + \beta_k c_k]$$

Hence, the partial effect of x_j on $p(\mathbf{x})$ depends on \mathbf{x} through $\phi(\mathbf{x}\boldsymbol{\beta})$. However, since Φ is a strictly increasing cdf, the sign of the partial effect is determined by the sign of β_j . It is worth to notice, finally, that since the latent variable does not have a unit of measurement, the magnitude of the β_j are not meaningful.

5.2 Estimation Results

In the following we introduce the probit model estimation results for AMS-IX, LINX, DE-CIX, VIX, ESPANIX and RONIX.

Table 3: Probit Estimated Model Results

	AMS-IX [♣]	DE-CIX [♣]	LINX [♣]	VIX [♣]	ESPANIX [♣]	RONIX [♣]
DIST ISPs	-.043 (10.23)	-.041 (2.78)	-.009 (1.90)	.055 (3.77)	.078 (1.46)	-.115 (1.11)
COMMON_IXP	.454 (30.59)	.338 (20.83)	.394 (26.27)	.215 (6.46)	.326 (2.98)	- -
DIFF_IXP_MEMBER	-.022 (6.23)	-.024 (5.24)	.011 (2.62)	-.065 (7.80)	-.021 (1.00)	.034 (1.39)
DIFF_RELEVANCE	-.016 (22.32)	-.016 (18.15)	-.014 (17.76)	-.247 (10.36)	-.347 (3.25)	1.308 (3.99)
CONSTANT	.296 (19.18)	.337 (16.44)	.144 (7.11)	.578 (10.92)	.122 (0.53)	-1.25 (5.16)
Observations	18145	9591	12880	2415	210	153
Pseudo R-Square	0.0681	0.0579	0.0583	0.0869	0.1053	0.1225
Log-likelihood	-11538.3	-6009.1	-8214.43	-1314.20	-118.31	-66.45
Significance test statistic: LR chi2(6)	1685.40	738.82	1017.30	250.17	27.85	18.54
Absolute value of z statistics in parentheses						
♣ = random peering distribution; ♠ = clustered peering distribution; ♦ = dispersed peering distribution						

The above table shows that three variables seem to consistently affect peering decision between a pair of providers. These are the hot potato and reputation effect, expressed by the number of IXPs that they are both member of, having a positive effect, and the difference in both size and traffic flows, respectively modeled through difference in the number of memberships at the Euro-IX IXPs and the difference in traffic flows (captured through the betweenness).

One of our most interesting findings is that the geographical distance seems to play a role that depends on the nature of the IXP morphology. Indeed, if we consider IXPs exhibiting a random pattern about the peering

agreement distribution (AMSIX, DECIX, LINX), the distance seems to have a negative effect on peering. The significance level is very high for both AMSIX and DECIX, while lower for LINX. This result also matches the result of the Moran's z value for LINX being much higher than those characterizing both AMSIX and DECIX.

While turning on the VIX, characterized by a clustered distribution, the geographical distance is still significant but with positive sign. Distance plays the opposite role on the likelihood of establishing bilateral peering agreements depending on the existing degree of clustering of the IXP. Finally, IXPs characterized by dispersed patterns of the peering agreements distributions distance does not seem to play a role.

The table below provides the partial effects calculated in correspondence of the mean values vector of the explanatory variables.

Table 4: Probit Estimated Partial Effects

Variable dy/dx	AMS-IX [♣]	DE-CIX [♣]	LINX [♣]	VIX [♣]	ESPANIX [♠]	RONIX [♠]
DIST ISPs	-.017 (10.23)	-.015 (2.78)	-.004 (1.90)	.018 (3.77)	.027 (1.46)	-.029 (1.12)
COMMON_IXP	.177 (30.75)	.128 (20.97)	.152 (26.48)	.071 (6.48)	.111 (3.16)	- -
DIFF_IXP_MEMBER	-.009 (6.23)	-.009 (5.24)	.004 (2.62)	-.021 (7.82)	-.007 (1.00)	.008 (1.40)
DIFF_RELEVANCE	-.006 (22.30)	-.006 (18.12)	-.005 (17.77)	-.081 (10.30)	-.119 (3.23)	.332 (3.94)

Absolute value of z statistics in parentheses

♣ = random peering distribution; ♠ = clustered peering distribution; ♦ = dispersed peering distribution

From a dynamic point of view, the results obtained are very interesting. By combining the results obtained in section 4 and section 5 it comes out that the European peering morphology at IXPs is converging through the randomness

equilibrium distribution. Indeed, the positive role played by the distance in peering within clustered IXPs is likely to be cluster breaking in the long run.

6 Conclusions

In this paper we investigated the relevance of geography for the Internet Industry by analyzing the peering decisions between Internet Service Providers participating at different European Internet Exchange Points.

Firstly we focused on a static analysis, providing the picture of the emerging morphology in peering. In particular we computed spatial autocorrelation statistics for the number of peering agreements realized within the different IXPs. Results showed a dominant presence of random pattern, suggesting a non-agglomerated equilibrium. This result is in line with the effects of the ‘death of distance’.

We then investigated the possibility that distance still plays a role in the strategic interconnection decisions between providers. Indeed, the possibility of mutual control is argued to be important for the governance of relationships, such peering, characterized by highly asymmetric information. Interestingly, geographical distance is significant in explaining peering within the equilibrium ‘random’ IXPs and the IXPs resulting instead clustered. However, it seems to play a very different role depending on the pre-existing morphology. Within IXPs characterized by a random morphology, geographical distance has a negative effect on peering: this may indicate relevance for mutual knowledge and face-to-face contact in peering decisions. When considering clustered IXPs, instead, the likelihood of peering seems to increase with the distance within the providers. From a dynamic point of view, this effect is likely to be cluster breaking in the long run, with the clustered morphology converging to the random equilibrium.

Other interesting results in the analysis of bilateral peering show consistently, across both random, clustered and dispersed morphologies, that hot potato and reputation effects positively affect peering, while difference in both size and traffic flows has a negative effect.

REFERENCES

- Anselin, L. (1992), 'Spatial Data Analysis with GIS: an Introduction to Application in the Social Sciences', University of California, Technical Report 92-10.
- D'Ignazio, A. and Giovannetti, E. *forthcoming*, 'From Exogenous to Endogenous Networks: Internet Applications', Journal of Economic Surveys.
- Dodge, M. and Kitchin, R. (2001), *Mapping Cyberspace*, Routledge, London and New York.
- Forman, C., Goldfarb, A. and Greenstein, S. (2002), 'Digital Dispersion: An Industrial and Geographic Census of Commercial Internet Use', NBER Working Paper No. 9287.
- Fujita, M. and Thisse, J. F. (2002), Economics of Agglomeration: Cities, Industrial Location, and Regional Growth, Cambridge: Cambridge University Press.
- Giovannetti, E., Neuhoff, K. and Spagnolo, G. (2005) "Agglomeration in Internet Co-operation Peering Agreements" Cambridge Working Paper in Economics n. 505
- Glaeser, E., Kallal, H., Scheinkman, J. and Schleifer, A. (1992), 'Growth of cities', Journal of Political Economy, 100, 1126-1152.
- Isaksen, A. (2004), 'Knowledge-Based Clusters and Urban Location: The Clustering of Software Consultancy in Oslo', Urban Studies, 14 (5-6), 1157-74.
- Krugman, P. (1991), Geography and Trade, Cambridge, MA: MIT Press
- Learner, E. and Storper, M. (2001), 'The Economic Geography of the Internet Age', NBER Working Paper No. 8450.
- Marshall, A. (1890), Principles of Economics, London: Macmillan.
- Moran, P. (1948), 'The interpretation of statistical maps', Journal of the Royal Statistical Society, Series B, 10, 243-251.

- Power, D. and Lundmark, M. (2004), 'Working through Knowledge Pools: Labour Market Dynamics, the Transference of Knowledge and Ideas, and Industrial Clusters', Urban Studies, 41 (5-6), 1025-44.
- Santarelli, E. (2004), 'Patents and the Technological Performance of District Firms Evidence for the Emilia-Romagna Region of Italy', Max Planck Institute for Research into Economic Systems Papers on Entrepreneurship, Growth and Public Policy, n, 29-2004.
- Shimbel, A. (1953), 'Structural parameters of communication networks', Bulletin of Mathematical Biophysics, 15, 501-507.
- Spagnolo, G. (1999), 'Social Relations and Cooperation in Organizations', Journal of Economic Behavior and Organization, 38 (1), 1-26.
- Titley, N. (1997), 'An Analytical Model of Peering between Internet Service Providers', Release 1.
- Upton, G. J. G., and Fingleton, B. (1985), Spatial data analysis by example. Volume 1: Point pattern and quantitative data, Chichester, England: Wiley.
- Warf, B. (2001), 'Segueways into cyberspace: multiple geographies of the digital divide', Environment and Planning B, 28, 3-19.
- Woodcock, B. (2002), 'BGP for Bankers (White Paper on Transactions and Valuation Associated with Inter-Carrier Routing of Internet Protocol Traffic)', Packet Clearing House.