Regional Innovation and Entrepreneurship Research Center

Research Productivity and the Quality of Interregional Knowledge Networks

Working Paper

June 2012

Authors

Tamás Sebestyén

Attila Varga

Közgazdasági és Regionális Tudományok Intézete
Pécsi Tudományegyetem, Közgazdaságtudományi Kar

MŰHELYTANULMÁNYOK

Research Productivity and the Quality of Interregional Knowledge Networks

Tamás Sebestyén

and

Attila Varga

2012/2

2012. június

Szerkesztőbizottság:

Research Productivity and the Quality of Interregional Knowledge Networks

Tamás Sebestyén
Pécsi Tudományegyetem
Közgazdaságtudományi Kar
Közgazdasági és Regionális Tudományok Intézete
sebestyent@ktk.pte.hu
http://www.krti.ktk.pte.hu/index.php?p=contents&cid=27

and

Attila Varga*
Pécsi Tudományegyetem
Közgazdaságtudományi Kar
Közgazdasági és Regionális Tudományok Intézete
vargaa@ktk.pte.hu
http://www.krti.ktk.pte.hu/index.php?p=contents&cid=29

*Corresponding author

Keywords: patents, scientific publications, knowledge networks, R&D productivity, regional knowledge production function, European regions

JEL Codes: O33, R11, R58

Abstract

This paper estimates the impact of interregional knowledge flows on the productivity of research at the regional level. We develop the novel index of 'ego network quality' in order to measure the value of knowledge that can be accessed from a particular region's global knowledge network. Quality of interregional knowledge networks is related to the size of knowledge accumulated by the partners ('knowledge potential'), the extent of collaboration among partners ('local density') and the position of partners in the entire knowledge network ('global embeddedness'). Ego network quality impact on the productivity of research in scientific publications and patenting at the regional level is tested with co-patenting and EU Framework Program collaboration data for 189 European NUTS 2 regions.

Research Productivity and the Quality of Interregional Knowledge Networks[1]

1. Introduction

Scholarly attention towards the spatial dimension of knowledge flows has intensified in the past two decades in economics in general and in regional science and economic geography in particular. Since innovation is significantly related to existing knowledge, understanding the mechanisms of knowledge communication in space is essential for understanding the geography of economic growth. When knowledge flows (e.g., spillovers facilitated by informal relations, learning in research collaborations or knowledge transfers mediated by market transactions) are dominantly local, economic growth will most probably be highly uneven in space. Alternatively, globally flying knowledge may contribute to the spreading of economic growth (Fujita and Thisse 2002).

Most of the research in this field has focused on the role of spatial proximity in innovation. The early papers (Jaffe 1989, Jaffe, Trajtenberg and Henderson 1993, Anselin, Varga and Acs 1997) evidenced that knowledge flows between firms, private and public R&D labs are to a large extent localized geographically. Later on, several studies applying similar research methods in different countries supported this finding (Ghinamo 2010). It became clear soon, that besides pure knowledge spillovers, local labor markets (Breschi and Lissoni 2009), entrepreneurship (Zucker and Darby 1998) or formalized research collaborations (Miguélez and Moreno 2012) contribute importantly to the localized communication of knowledge. Though the proximity of industry-specific (specialized) knowledge could also be important, empirical findings suggest that the diversity of the local knowledge base is associated most frequently with regional growth (Glaeser, Kallar and Scheinkman 1991, Feldman and Audretsch 1999). Deeper analysis reveals that not a simple diversity of industries but their technological relatedness is what nurtures innovation (Frenken, van Oort and Verburg 2007).

Subsequent research clarifies that spatial proximity is neither a sufficient nor a necessary condition for knowledge flows to occur (Boschma 2005). Instead, other forms of proximities (such as cognitive, social, institutional or organizational proximities) are the crucial prerequisites of knowledge

[1] We would like to express our thanks to Dimitris Pontikakis and George Chorafakis for their contribution to the development of the data applied in our analyses.

communication. Geographical proximity of actors in innovation may enhance the flow of knowledge but only if at least one of the other proximities are also in effect. Physically proximate location provides only the opportunities for frequent interactions. These interactions might become instrumental for the speedy flow of both tacit and codified knowledge as well as for the development of trust or the establishment of common codes of communication (Koschatzky 2000). Ponds, van Oort and Frenken (2009) show empirically that spatial proximity helps bridging institutional distances between business, academia and government. Thus, knowledge flows between different types of organizations tend to be localized. On the other hand, knowledge communication could also occur over larger physical distances among actors sharing similar institutional features.

Compared to the high intensity of research on localized knowledge interactions, scientific inquiry on the mechanisms and regional impacts of global knowledge flows is still a relatively less developed and novel phenomenon. Spatial econometric studies brought the first empirical evidence on the existence of interregional knowledge transfers and their positive role in regional innovation (Anselin, Varga and Acs 1997, Varga 1998). Recent efforts in the literature focus on studying various mechanisms of interregional knowledge flows ranging from professional labor mobility (Maier, Kurka and Trippl 2007, Schiller and Diez 2008, Miguélez, Moreno and Suriñach 2009), research collaboration (Maggioni and Uberti 2011) and co-inventorship (Breschi and Lenzi 2011) to the operation of multinational companies (Cantwell and Iammarino 2003).

Network analysis (NA) is an especially promising tool for the study of interregional knowledge flows. Different measures of network structure such as network size, centrality of actors or density of interactions appear particularly powerful for understanding the geography of knowledge production. Much of the scientific inquiry in interregional knowledge interactions applying NA techniques grows out from the spatial econometrics tradition. Researchers realize that weights matrices routinely used in spatial econometrics to represent relations in space can also be applied to characterize relative positions in interregional knowledge networks (Maggioni and Uberti 2011). It is shown in this literature that the number of interregional partners in research or invention and their levels of knowledge (measured by e.g. R&D expenditures or publication stock) are indeed influential factors in regional knowledge production (Maggioni, Nosvelli and Uberti 2007, Hoekman, Frenken and van Oort 2009, Varga, Pontikakis and Chorafakis 2010).

Thus, the literature suggests that the larger the number of interregional partners is and the higher their knowledge levels are, the more effective a region becomes in knowledge generation. Related to this finding at least two important questions arise. First, some complementary relationship might exist between network size and the level of knowledge in the network. A small network with a limited number of highly knowledgeable partners might be as valuable for a region as a large network with partners possessing diverse knowledge levels. A comprehensive measure aggregating different network features could potentially account for such complementarities. Additionally, some other features of interregional networks could also influence research productivity.

In this paper we intend to make a step further in the research addressing the relationship between interregional knowledge flows and regional knowledge creation. We structure the problem by directing attention to the *value* of a particular interregional network for a particular region. This value is reflected by the contribution of knowledge accessed in the network to the production of new knowledge inside the region. In order to make the problem empirically tractable we quantify certain features of interregional networks that could be instrumental for regional knowledge production additional to what have already been accounted for in the spatial literature (i.e., number of partners and their knowledge levels). Building on these measures and continuing the research line in Varga and Parag (2009) we develop a novel comprehensive index to account for the quality of a region's global knowledge network. The higher this index value gets, the larger the amount of knowledge potentially accessed from the network will be. We also provide a test on the hypothesized positive relationship between interregional knowledge network quality and regional research productivity. In pursuit of this we investigate two interregional networks: EU Framework Program collaborations and European co-patenting, both considered at the level of EU NUTS2 regions. The two networks refer to two types of new knowledge generation processes: the production of new scientific knowledge measured by publications and the production of new technological knowledge measured by patent applications.

Our paper is structured as follows. The second section introduces the concept of ego network quality, followed by the section describing the analytical model and the data used in the empirical analysis. The fifth section presents the results on how the quality of a region's interregional network affects the productivity of research in the creation of publications and patents at the regional level. The last section summarizes our findings and highlights potential further research directions.

2. Ego network quality

The theory of innovation emphasizes the role of interactions among different actors in innovation. These interactions follow a system and the characteristics of the system determines to a large extent the efficiency of new knowledge production (Lundvall 1992, Nelson 1993). An extensive survey-based empirical literature evidences that innovation is indeed a collective process where the knowledge and expertise of partners as well as the intensity of collaborations among them largely determines the production of new, economically useful knowledge (e.g., Diez 2002, Fischer and Varga 2002). Representing actors as nodes and their connections as ties, interactions of collaborating agents can be mapped as a network. On the basis of this representation the application of network analysis extends the frontiers of the study of knowledge interactions well beyond the possibilities of traditional innovation surveys.

Research on the interrelationships between the structure of actors' individual network and the performance of actors in knowledge production is a relatively recent phenomenon. Four features of individual (ego) networks are considered in this literature: 1) characteristics of immediate partners and the intensity of the actor's relations with them (number of partners, strength of ties, knowledge of partners); 2) density of interactions between partners; 3) diversity of knowledge accessed through the network (by being in contact with different fields of research); 4) the position of an agent in the entire network (in order to account for the impact of knowledge accessed beyond the ego network).

As shown in the literature, number of partners (Powell et al. 1999, Hopp et al. 2010, Van Der Deijl, Kelchtermans and Veugelers 2011), tie strength (Van Der Deijl, Kelchtermans and Veugelers 2011) and knowledge of partners (Maggioni, Nosvelli and Uberti 2006, Hoekman, Frenken and van Oort 2009, Varga, Pontikakis and Chorafakis 2010) are in a positive relationship with the productivity of research. However, the influence of the intensity of interactions is ambiguous. It positively affects agents' patenting productivity in Salmenkaita (2004) and Cross and Cummings (2004), the impact follows an inverted U-shape in Van Der Deijl, Kelchtermans and Veugelers (2011), while the influence on academic publishing is negative in Rumsey-Wairepo (2006) and Cainelli et al. (2010)[2]. It is also found that both knowledge diversity of partners (Powell et al. 1999, Cainelli et al 2010, Van Der Deijl, Kelchtermans and Veugelers 2011) and a central position in the entire network (Powell et al. 1999, Cainelli et al. 2010, Hopp et al. 2010, Van Der Deijl, Kelchtermans and Veugelers 2011) positively affect performance in knowledge generation.

In this paper we develop and apply the index of Ego Network Quality (ENQ) in empirically testing the impact of interregional knowledge networks on regional research productivity. The higher the value of ENQ is, the higher the level of knowledge to be accessed from the network. The first intuition beyond ENQ is that the level of knowledge accessed from an agent's network is in a positive relationship with the total knowledge of immediate partners. The second intuition is that collaboration among network partners is the source of further growth of knowledge in the network. We also assume that partners in a network not only increase the amount of knowledge accessible, but also contribute to its diversity through building links to different groups in the network.

ENQ integrates ego network features highlighted in the preceding paragraphs into one comprehensive measure. Based on the literature given above, we propose three dimensions for ENQ: Knowledge Potential, Local Density and Global Embeddeddness. Knowledge Potential (KP) measures knowledge accumulated in the immediate neighborhood and it is related to number of partners, strength of ties, and knowledge of individual partners. Local Density (LD) is associated with the intensity of interactions among partners in the ego network, while Global Embeddeddness (GE) measures the extent to which knowledge developed in other networks becomes accessible through immediate partners. In what follows, we propose appropriate measures for these three dimensions and then comprise them into one single index of Ego Network Quality.

---

[2] Referred to as intensity here, the question of cohesion in a network is also tackled from the social capital perspective. Here the debate is on wether cohesive, closed structures (Coleman, 1986) or 'structural holes' (Burt, 1992) provide a better background for performance. Although many of the results in this field show that a position in structural holes contribute to better performance in a diversity of fields (e.g. Hopp et al (2010), Kretschmer (2004), Donckels and Lambrecht (1997), Zaheer and Bell (2005), Powell et al. (1999), Tsai (2001), Burt et al. (2000), Burton et al (2010)), there is still evidence on the opposite (Salmenkaita, 2004), Cross and Cummings (2004). Rumsey-Wairepo (2006) argues that the two structural settings are complementaries rather than substitutes in explaining performance. In our context we also emphasize that different structural dimensions can be important for different networks. When information flows and power is important, structural holes indeed provide better position, however, as in our case, if knowledge production is in the focus, exclusion resulting from structural holes may be harmful and cohesiveness meaning better interaction may have positive contribution.

Knowledge Potential

The concept of KP relates to the amount of knowledge an agent's immediate partners possess. A straightforward way to capture this concept is to simply sum up the knowledge levels of immediate partners, where knowledge level is represented by some exogenous measure contained in a vector $\mathbf{k} = [k_i]$. KP is then measured by the simple sum of the immediate neighbors' knowledge:

$$KP^i = \sum_j a_{ij} k_j, \tag{1}$$

where a_{ij} takes the value of 1 if i and j are partners and 0 otherwise.[3] Note, that in the specific case of evenly distributed knowledge levels, or in other words, if we do not take into account the difference in the weights of the nodes in the network, this measure simplifies to nodal degree, familiar from graph theory and network analysis.[4]

Local Density

We define LD as the intensity of cooperation among the partners of an agent's network. Local clustering coefficient as defined by Watts and Strogatz (1998) measures the density of ties in a node's neighborhood, i.e. the number of existing links among neighbors relative to the maximum possible number of such ties. However, this measure is largely dependent on the size of the neighborhood itself, therefore leading to considerable biases. For example in the case of a node with three neighbors where all three neighbors are linked to each other the local clustering coefficient would give a value of 1, which is the maximum possible. On the other hand, if a node has 10 neighbors among which altogether 15 ties are counted, the clustering coefficient is 0.33, which is much less than unity. However, no one would intuitively conclude that the neighborhood of the first node with three neighbors is denser in the common sense than that of the second node.

In order to avoid this problem, we use a similar measure for local density which is the average number of links in the neighborhood: the total number of ties in the neighborhood divided by the number of neighbors. Formally we can write:

[3] The same notation and methodology can be used in the case of weighted networks. If the weights are normalized to the interval between 0 and 1, than the formulae presented for the binary networks are directly applicable for the weighted counterparts.

[4] If $k_i = 1$ for all i, then this similarity is straightforward. However, if $k_i = \bar{k} \neq 1$, then degree and the above mentioned measure of KP are not equal but proportional.

$$LD^i = \frac{\sum_l a_{il} \sum_j a_{ij} a_{jl}}{2N_i} + 1 \tag{2}$$

where $N_i = \sum_j a_{ij}$ is the number of node i's neighbors. The expression in the nominator counts the number of links among node i's neighbors (j and l are also indices of the nodes). However, as the network is symmetric, every such node is counted twice, hence the number in the denominator. Adding one at the end of the expression serves to include not just ties among node i's partners but the nodes connecting it to the neighbors (dividing this total number of links in the neighborhood with the size of the neighborhood gives the simple unity shift as a result).

It is clear, that the difference between this measure of local density and the local clustering coefficient is minor, and lies in the denominator of the expression and the additional constant. In the case of the local clustering coefficient only the first term is in effect with the denominator being $N_i(N_i - 1)$ instead of N_i. Although it has no additional value in this specific place, the constant is included because in the third measure (GE) not just links among partners at specific distances are relevant but links that tie nodes at different distances. This way, the measure for LD is generalized later and the constant at the end obtains a specific content.

Global Embeddedness

KP and LD are measures that capture the local structure of a node's network. Our concept behind global embeddedness is to take into account the structure of the network behind the immediate neighborhood. In this respect, we are looking for a decent centrality measure.

There are several measures of centrality, which are widely accepted in network analysis. The simplest and most widely used is degree centrality, which only counts one's connections. As it was mentioned, this is included in the KP measure, though in a weighted version. On the other hand, degree is a local centrality measure and we are attempting to look behind local structures. Other centrality measures are for example closeness and betweenness centrality (Freeman, 1979), however, these indices capture specific aspects of centrality, and we are looking for a more generalized view. The 'global' and weighted extension of degree centrality is eigenvector centrality, which gives a recursive definition for the centrality of a node (see Bonacich (1972), Bonacich (2007)). However, it is also redundant for us as it gives the highest weight to the local degree of a node. In order to avoid counting local ties more than once and be as general as possible, we propose a recursive definition for

GE as follows. GE reflects the knowledge-and distance-weighted Knowledge Potential and Local Density of neighborhoods at different distances from the node in question. Formally:

$$GE^i = \sum_d W_d \, KP_d^i LD_d^i \tag{3}$$

where d is the index of distance (starting from 2), KP_d^i is the Knowledge Potential (i.e., the sum of knowledge levels) of the nodes at distance d from node i, LD_d^i is the Local Density (as defined before) of the nodes at distance d from node i. Note that LD_d^i counts the links among nodes at distance d and the links bridging nodes at distances $d-1$ and d. W_d denotes a weight for distance d. Throughout the paper we use linear weights, but any weighting method can be used.[5]

Ego Network Quality

The three measures detailed above are then linked together into a comprehensive index of Ego Network Quality. The three measures are connected as a sum of a local knowledge-weighted cooperation intensity and a global knowledge-weighted cooperation intensity, the latter being a distance-weighted sum of knowledge-weighted cooperation intensities at different distances. In other words, the local part of this measure can be regarded as a special case of the distance-specific knowledge-weighted cooperation intensities. Formally we can write:

$$ENQ^i = KP^i LD^i + GE^i \tag{4}$$

This formula can be converted into a compact form if we define the sequence of node-generated subgraphs S_d^i , where S_d^i stands for a subgarph containing nodes, which are at most distance d from node i, and the edges between all such nodes. Let's denote the matrix of geodesic distances by $\mathbf{G} = [g_{ij}]$, and use the following notation:

$$C_d^i = \frac{\sum_{j,g_{ij}\leq d} \sum_{l,g_{il}\leq d} a_{jl}}{2} \tag{5}$$

In other words, C_d^i gives the number of edges counted in the subgraph containing nodes at distance d or less from node i, i.e., the number of edges in S_d^i. It is easy to see that we can now substitute the previously

[5] By linear weights we mean that weights decrease linearly as distance increases. Specifically, the weight is 1 at distance 1 and it is calibrated to 0 for distance M, where M is the number of nodes in the network thus $M-1$ is the longest possible path. The specific expression employed for the weights is: $W_d = d(1-M) - M/(1-M)$. This specification satisfies the required conditions.

given measure for LD_d^i by $(C_d^i - C_{d-1}^i)/N_d^i$, where N_d^i is the number of nodes at distance d from node i. Using $C_0^i = 0$, by definition,[6] the expression in equation (2) for the case $d = 1$ is written as $LD_1^i = C_1^i/N_1^i$. Using this notation we can reformulate ENQ_i in the following form:

$$ENQ^i = KP_1^i \frac{C_1^i}{N_1^i} + W_2 KP_2^i \frac{C_2^i - C_1^i}{N_2^i} + \cdots + W_{M-1} KP_{M-1}^i \frac{C_{M-1}^i - C_{M-2}^i}{N_{M-1}^i} \tag{6}$$

Again, using $C_0^i = 0$ and $W_1 = 1$, this formula can be written more comprehensively:

$$ENQ^i = \sum_{d=1}^{M-1} W_d KP_d^i \frac{C_d^i - C_{d-1}^i}{N_d^i} \tag{7}$$

It can be demonstrated that if the distance to the partners is 1 and knowledge level is identical across partners ENQ counts the total number of links in the network. On the other hand, if knowledge weights are not considered but distance weights are in effect, ENQ shows common features with the distance-weighted sum of degrees in the network. (The Appendix provides the derivation of these two special cases.)

ENQ in a dynamic context

In this section we provide a descriptive analysis of the previously proposed ENQ measure. The behavior of ENQ in a dynamic context is shown on a sample network. Our goal in this respect is to demonstrate how changes in the network position affect ENQ. In the panels of Figure 1 three snapshots of a network is indicated. The three snapshots are illustrations of a dynamic process of network evolution.[7] Our focus is on one specific node, marked by the black square whereas another reference node is highlighted by the grey square. The grey node starts (in the top panel) in a central position and retains it, while the black node starts in a very peripheral position. The labels at each node indicate the knowledge level, which is attributed to each node by simply calculating the degree for the first snapshot.[8] These knowledge levels are not altered later in order to keep things simple. The size of the squares representing the nodes indicates the ENQ measure, calculated according to the method presented previously.

[6] The node at distance 0 from node i is itself therefore we have a trivial graph where the number of edges is zero by definition.

[7] However, only a specific range of the network changes in order to keep the demonstration as simple as possible.

[8] This way we can reproduce a specific feature of real-world networks, namely that nodes in a central position tend to be nodes with higher knowledge, information or resources in general.

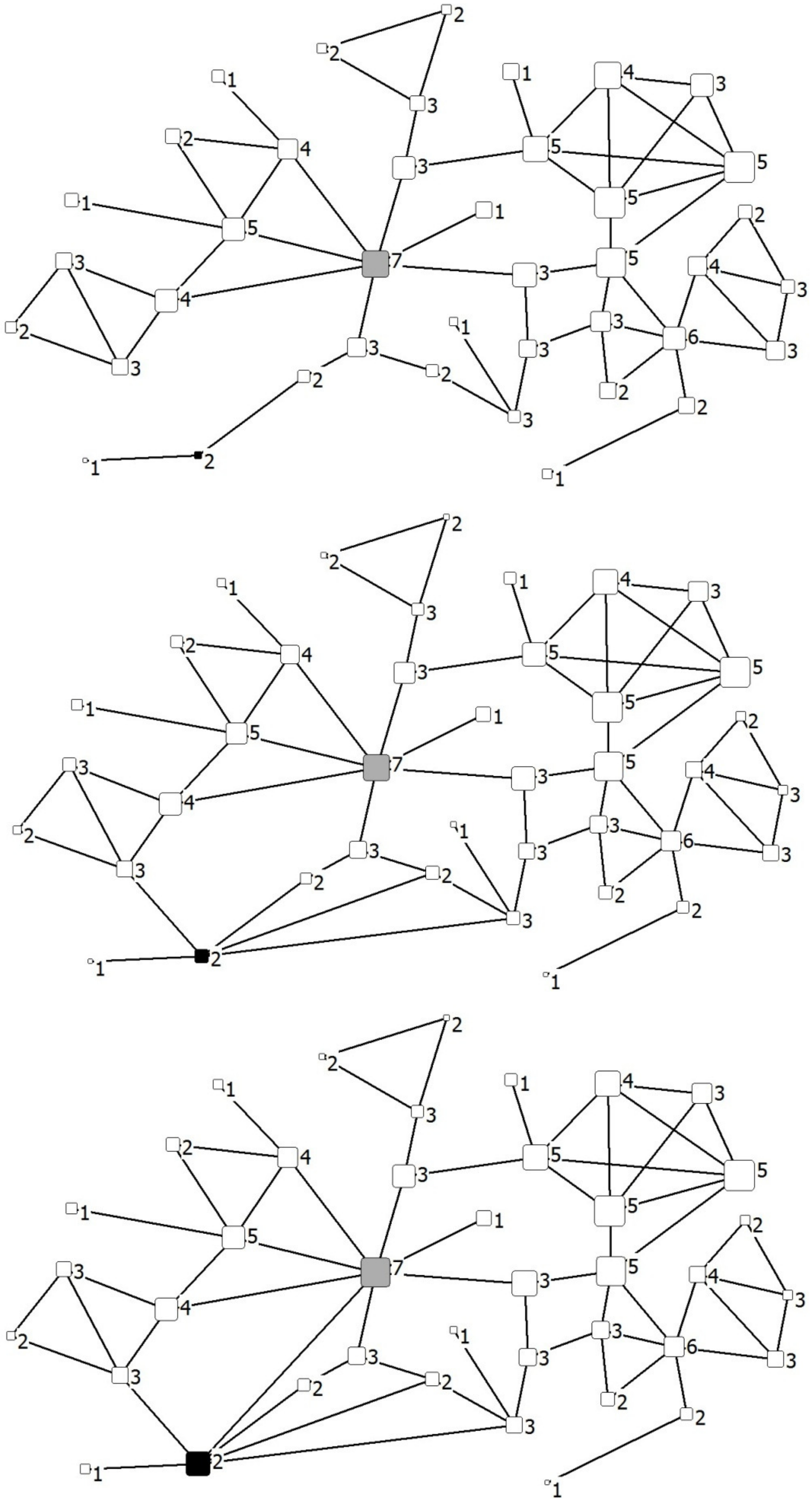

Figure 1. The impact of changing network positions on ENQ

Table 1 presents the KP, LD, GE and ENQ scores of the grey and black nodes in the three snapshots. In the first step the black node is quite peripheral. In the second step it establishes several links, but mainly to other relatively peripheral actors of the network. In step three only one additional link is established, but it links the black node to the grey one, that is, to one of the centers of the network. This special evolution is reflected in the scores for the different measures. At the first stage the black node has a low ranking in all three measures. The second step provides it with considerable new knowledge increasing its local density (LD) and global embeddedness (GE). In the third step, that one link has similar effects as the three links together in the previous step. Although LD decreases (due to the fact that the grey node is not linked to the other neighbors of the black node), KP and GE increase considerably.

Looking at the changes at the central (grey) node, we can see that the first step makes little change in its position. As no new links are established by this node itself and between its neighbors, the KP and LD are unchanged. GE increases a bit as the new links not very far from it provide a better embeddedness (more directly accessible knowledge and a denser structure of the network). In the third step, however, KP also increases with the direct access to a new node. LD decreases for the same reason as it decreases in this step for the black node.

Table 1. The impact of changing network positions on ENQ

BLACK NODE				
	KP	*LD*	*GE*	*ENQ*
Snapshot 1	3,00	1,00	33,49	36,49
Snapshot 2	11,00	1,20	45,23	58,43
Snapshot 3	18,00	1,17	55,15	76,15
GREY NODE				
	KP	*LD*	*GE*	*ENQ*
Snapshot 1	23,00	1,29	47,76	77,33
Snapshot 2	23,00	1,29	50,94	80,51
Snapshot 3	25,00	1,25	53,57	84,82

ENQ in some reference networks

The two sample networks analyzed here were intentionally structured as to serve our goals of exhibition. In the following paragraphs a similar analysis is carried out briefly in some representative reference networks, with attention to the correlation between the three basic measures used in the composite ENQ index. Three specific network structures are usually taken as reference in network theory: (i) the regular network, where each node has the same number of links; (ii) the Erdős-Rényi type random network where links are distributed evenly according to a predefined probability;[9] (iii) the scalefree network where the distribution of degree values follows a power law.[10] The first possibility, the regular network is of no interest here as in such a network the nodes are symmetric thus only the exogenous knowledge levels account for differences in ENQ. The other two specific structures are analyzed below. The networks, for which KP, LD, GE and ENQ measures are calculated, are simulated networks with 300 nodes with 10% density. Figure 2 shows the correlation diagrams between the different measures in question for the random and the scale-free networks.

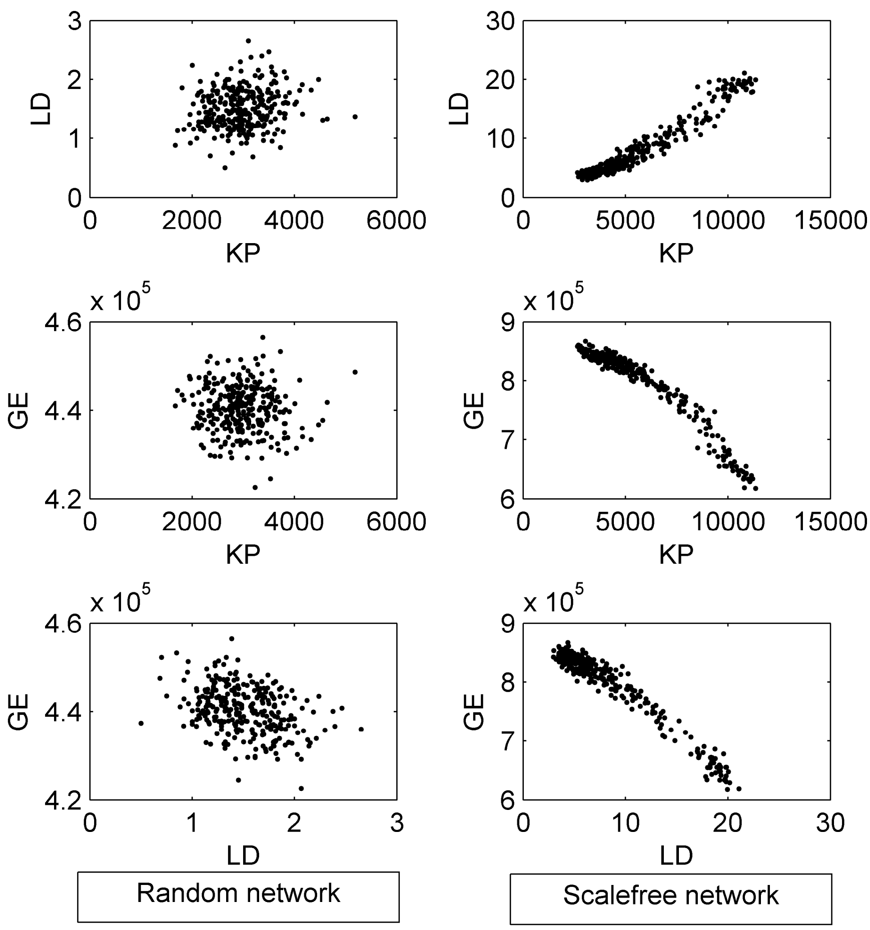

Figure 2. Correlation diagrams – Random and Scalefree network simulations

[9] This probability in turn defines the overall (global) density of the network.

[10] Scalefree networks are generated according to the preferential attachment model (Barabási and Albert, 1999)

As it can be seen from Figure 2, there is no considerable correlation between the three basic measures in a random network. This means that these measures can be regarded as complements: in general, all capture different aspects of the position of a given node. This independence disappears in a scale-free network: there is a positive correlation between KP and LD. For the clear interpretation of these results we must note that the network generating algorithm constructs a one-centered network. This results in the fact that nodes in the center (i.e. with lots of links and surrounding knowledge) tend to be found in the more tightly linked part of the network, which is the center. Similarly, there is a negative correlation between KP and GE, stemming from the fact that nodes in the center have no other considerable cores in the surrounding. The negative correlation between LD and GE results from the correlation between the previous two measures.

These results first prove that the three measures taken into account in our study capture different aspects of network position in general. However, if we consider specific network structures, there may be considerable correlations between these measures. Irrespective of these correlations, however, the ENQ index can be used to reflect individual differences in network position: by comprising three different (and possibly unrelated) aspects of this position, we can catch a more detailed description of it compared to the use of one single measure. This is marked by the fact that there is still variation in the pictures for scale-free networks – and keep in mind that these simulated networks exhibit considerable regularity due to the preferential attachment algorithm.

3. Empirical model and data

Our starting point is the knowledge production function initially specified by Romer (1990) and parameterised by Jones (1995). In the interpretation of the parameters we follow Varga (2006).

$$dA_i / dt = \delta\, H_{A_i}{}^{\lambda} A_i{}^{\varphi} \tag{8}$$

where dA/dt is the temporal change in technological knowledge, H_A refers to research inputs (e.g. number of researchers or research expenditures), A is the total stock of already existing scientific and technological knowledge (knowledge codified into publications, patents etc.) and i stands for the spatial unit. Thus technological change is associated with contemporary R&D efforts and previously accumulated knowledge. The same number of researchers can have a varying impact on technological change depending on the stock of already existing knowledge. Two parameters are particularly important for this paper. The size of φ reflects the impact of the transfer of codified knowledge. Regarding parameter λ the larger its size is, the stronger the impact the same number of researchers will play in technological change. Its value thus reflects the productivity of research in region i. We assume that the size of λ is positively related to the quality of interregional knowledge networks measured by ENQ.

In order to test empirically the hypothesised relationships we follow Varga (2000) and Varga, Pontikakis and Chorafakis (2010) and use the following econometric specifications. Using subscripts *i* and *N* to denote individual regions and nations (in our case EU member states) respectively, the empirical counterpart of the Romer (KPF) is specified as:

$$\log \mathrm{K_i} = \mathrm{a_0} + \mathrm{a_1} \log \mathrm{RD_i} + \mathrm{a_2} \log \mathrm{KSTOCK_N} + \mathrm{Z_i} + \varepsilon_\mathrm{i} \qquad (9)$$

where K stands for new scientific-technological knowledge, RD is expenditure on research and development, KSTOCK represents already existing scientific/technological knowledge at the national level and Z stands for additional regional control variables. We use national patent stock as a proxy for codified technological knowledge reachable with unlimited spatial accessibility within the country.

Equation 10 relates research productivity measured by $a_{1,i}$ the parameter of the research variable in Equation 9 to interregional network quality.

$$\mathrm{a_{1,i}} = \beta_0 + \beta_1 \log \mathrm{ENQ_i} \qquad (10)$$

Substituting equation 10 to equation 9 results in the following equation to be estimated:

$$\mathrm{Log}(\mathrm{K}_{i,\,t}) = \alpha_0 + \beta_0 \mathrm{Log}(\mathrm{RD}_{i,\,t\text{-}k}) + \beta_1 \mathrm{Log}(\mathrm{ENQ}_i) * \mathrm{Log}(\mathrm{RD}_{i,\,t\text{-}k}) + \alpha_2 \mathrm{Log}(\mathrm{KSTCK}_{N,\,t\text{-}k}) + \mathrm{Z}_i + \varepsilon_i. \qquad (11)$$

The empirical analysis in this paper is based on a sample of 189 European regions (a mix of NUTS2 and NUTS1 regions) for which information was complete enough for our purposes. We use a cross-sectional database, though a two-year time lag is employed. The time period under consideration is determined by the duration of the 5th Framework Program of the EU, spanning the years 1998-2002. According to this, dependent variables are given for 2002, the independent variables are given for 2000 and the network variables are based on network connections observed throughout the period between 1998 and 2002.

As to reflect knowledge flows in the production of new scientific results (resulting from Pasteur- type knowledge generation) on the one hand and new technological knowledge (a result of Edison-type knowledge production) on the other, we use different proxies for each, leading to different estimated equations and variables. Dependent variables are patenting activity and publication activity at the regional level as proxied by patent applications to the EPO (PAT_i) and scientific publications in ISI journals (PUB_i) respectively. We do not make any distinction between technological and scientific fields, so patent and publication counts are aggregated measures in this respect. Although using patents as a proxy for technological innovation is far from a perfect solution, there are several reasons why it still remains one of the most widely used and accepted measures (see e.g. Griliches 1990, for a comprehensive study on the issue, or Acs, Anselin and Varga 2002, for an analysis on the links between patent and other innovation counts at the level of regions). Publications are a somewhat stronger proxy for scientific knowledge and also used widely in innovation studies (van Raan 2004). Although the publishing motivation in academic science lends more reliability to publication counts, there are still biases stemming from journal coverage and distortions of evaluating mechanisms. Overall, although such biases may be relevant in an inter-regional comparison, given our central question here there are no strong reasons to think that it could affect European tendencies.

Following Romer (1990), the importance of knowledge stocks (or a 'standing on the shoulders of giants' effect) for knowledge production has been verified empirically (Furman, Porter and Stern, 2002; Zucker et al. 2007). In order to capture this effect we use proxies of national knowledge stocks (available for all regions in the given country) corresponding to the two different knowledge types ($PUBSTCKN_i$, $PATSTCKN_i$). For technological knowledge we use national patent stock and the national publication stock for scientific knowledge.

Table 2. Variable description

Variable Name	Description	Source
PAT_i	Number of patent applications to the European Patents office (EPO) by region of inventor (fractional counts)	Eurostat NewCronos database
PUB_i	Number of publications in scientific journals in the Thomson ISI database (search criteria: article, letter, review)	RFK database (data processed by CWTS, Leiden University)
GRD_i	Gross regional expenditures on R&D, in millions of Purchasing Power Standard (PPS) Euros, 1995 prices	Eurostat NewCronos database
$FPKP_i$ / $PATKPW_i$	Knowledge Potential – the directly available knowledge from a region's partners. FKP is calculated for the binary FP network with accumulated R&D stock as knowledge levels. PATKPW is calculated for the weighted patent network with accumulated patent stocks as knowledge levels.	Authors' elaboration on FP5 administrative database, DG RTD, Dir A and OECD REGPAT database
$FPLD_i$ / $PATLDW_i$	Local Density – the average number of links in a region's neighborhood. FPLD is calculated for the binary FP network, PATLDW is calculated for the weighted patent network.	Authors' elaboration on FP5 administrative database, DG RTD, Dir A and OECD REGPAT database
$FPGE_i$ / $PATGEW_i$	Global Embeddedness – the structure of the network behind a region's immediate neighborhood. FPGE is calculated for the binary FP network, PATGEW is calculated for the weighted patent network.	Authors' elaboration on FP5 administrative database, DG RTD, Dir A and OECD REGPAT database
$FPENQ_i$ / $FPENQW_i$ $PATENQ_i$ / $PATENQW_i$	Ego Network Quality – a comprehensive measure of network position. FPENQ is calculated for the binary FP network, FPENQW is calculated for the weighted FP network, PATENQ is calculated for the binary patent network and PATENQW is calculated for the weighted patent network.	Authors' elaboration on FP5 administrative database, DG RTD, Dir A and OECD REGPAT database
$FPDEG_i$ / $PATDEG_i$	The number of a region's direct partners in the network. FPDEG is calculated for the binary FP network and PATDEG is calculated for the binary patent network.	Authors' elaboration on FP5 administrative database, DG RTD, Dir A and OECD REGPAT database
$PATSTCKN_i$	National patent stock corresponding to the given region	Authors' elaboration on Eurostat NewCronos
$PUBSTCKN_i$	National publication stock corresponding to the given region	Authors' elaboration on Eurostat NewCronos
$AGGL_i$	Index of agglomeration. Size-adjusted location quotient of employment in technology- and knowledge-intensive sectors: high and medium high technology manufacturing, high technology services, knowledge intensive market services, financial services, amenity services – health, education, recreation. For more details see Varga, Pontikakis and Chorafakis (2010)	Authors elaboration of Eurostat NewCronos

Table 3. Variable descriptive statistics

	PUB	*PAT*	*GRD*	*PUBSTCKN*	*PATSTCKN*	
N	189	189	189	189	189	
Mean	2000,28	314,55	730,30	15533,05	30105,69	
Std.dev.	2576,39	519,94	1212,11	14929,01	36317,06	
Min	3	0,01	1	25	11	
Max	21050	3282,27	11314	41111	98481	
	FPDEG	*FPENQ*	*FPENQW*	*FPGE*	*FPLD*	*FPKP*
N	189	189	189	189	189	189
Mean	131,54	7 759 859	132 352	657 315	55,88	120 960
Std.dev.	42,58	1 599 143	18 046	934 644	14,96	23 620
Min	8	3 852 354	58 986	957	4,50	10 789
Max	186	9 300 660	156 383	4 378 189	68,35	137 270
	PATDEG	*PATENQ*	*PATENQW*	*PATGEW*	*PATLDW*	*PATKPW*
N	189	189	189	189	189	189
Mean	52,53	4 995 496	30 350	7 310	0,13	1140,23
Std.dev.	36,16	1 544 242	11 459	6 422	0,07	2440,19
Min	0	0	0	0	0	0
Max	133	8 066 975	48 263	20 607	0,30	16263,33

Patent stock is calculated according to the perpetual inventory method for the 1992–1998 period (see the details in Varga, Pontikakis and Chorafakis 2010) while publication stock is a simple sum of the count of publications in the period of 2000-2002. Variable description is in Table 2, while descriptive statistics of the main variables are presented in Table 3.

4. Cooperation networks in patenting and publication

The two types of knowledge flows between regions are captured by two different cooperation networks. The structure of technological knowledge flows is proxied by an interregional network of patent inventor cooperation while that of scientific knowledge flows is captured by a network of cooperation in the FP5 program.

There are good reasons to expect that participations to the FP can be an appropriate proxy of the relational structure of interregional knowledge diffusion across Europe. The FPs were designed to support 'pre-competitive', collaborative research with no national bias as to the types of technologies promoted and the distribution of funds. The precompetitive character of supported research ensured that Community funding did not clash with the competition principles of the Common

Market and did not function as a form of industrial subsidy; the collaborative character of research and the cost-sharing provisions were seen to guarantee the diffusion of technologies and the involvement of various types of actors from the whole technological knowledge creation spectrum, such as large and small firms, universities and public research institutes. One potential drawback of the FP as a data source is the fact that it is artificial; i.e. collaborating teams will not always coincide with naturally emerging networks of researchers. (Varga, Pontikakis and Chorafakis, 2010)

With regards to technological knowledge flows, patent cooperation patterns can be traced out from a patent database. The REGPAT database of OECD was used for this purpose as a data source (OECD, 2009). This database contains information on patent applications filed to the European Patent Office (EPO), specifically information on the NUTS3 level regions of the inventors of each patent (extracted on the basis of their addresses given on the patent documentation). Certain biases are possible also in this respect, but the nature of co-inventing gives a solid basis for these estimations of knowledge flows as it unlikely to find inventor names in a patent application who do not bear any inventive idea supplied to the inventor's community. This is reassured by the relatively low number of inventors (2-3 on average) appearing in co-invented patents. In contrast to patent citations, tracing inventor cooperation provides insights to the flow of more tacit knowledge elements.

In both cases (FP and patent networks), the regional information (address) of participants in FP projects and inventors of patents, together with the information of the date of cooperation (duration of FP programs and priority years for patents), allows us to construct a simple network where for each project (FP or patent) we assign the regions where the partners/inventors are resident. Then, this two-mode network is converted into a one-mode network where the nodes are regions and the links between the regions refer to the (FP or patent) cooperation between the regions. This conversion is done on the basis of the assumption that each partner/inventor listed for a given FP project/patent are linked to each other. For example, if three actors, A, B and C cooperated in one project/patent, and actors A and B belong to region 1 while actor C belongs to region 2, then we conclude that there is a link between regions 1 and 2. Furthermore, the links in this interregional network is weighted, the link weights corresponding to the number of actor-actor contacts between the regions. In the previous example, we count two links between regions 1 and 2, one for the link between actors A and C and one for the link between actors B and C. This method is then iterated for each patent/FP project and finally, the link weights between 1998 and 2002 are simply summed up to obtain the final weighted adjacency matrix of cooperation between European regions in the two dimensions.

In the empirical analysis both the weighted and the binary versions of these networks are employed. In the binary networks we used the simple rule that a link is existent if there is at least one project/patent in which the two regions cooperated. In the weighted case we use normalized weights, obtained by simply dividing all raw weights by the largest weight in the network.

It is interesting to look at some descriptive measures of these two networks as they show remarkable differences. The densities are 0.694 for the FP network and 0.275 for the patent network. This means that in the FP network almost 70% of all possible ties are present whereas in the patent network only 27.5% exists. This shows that on average the regions in question are better connected through FP partnership programs than through patenting activity.

The differences between the two networks are reflected by their degree distributions, depicted on Figure 3. The horizontal axis contains the possible degree values while

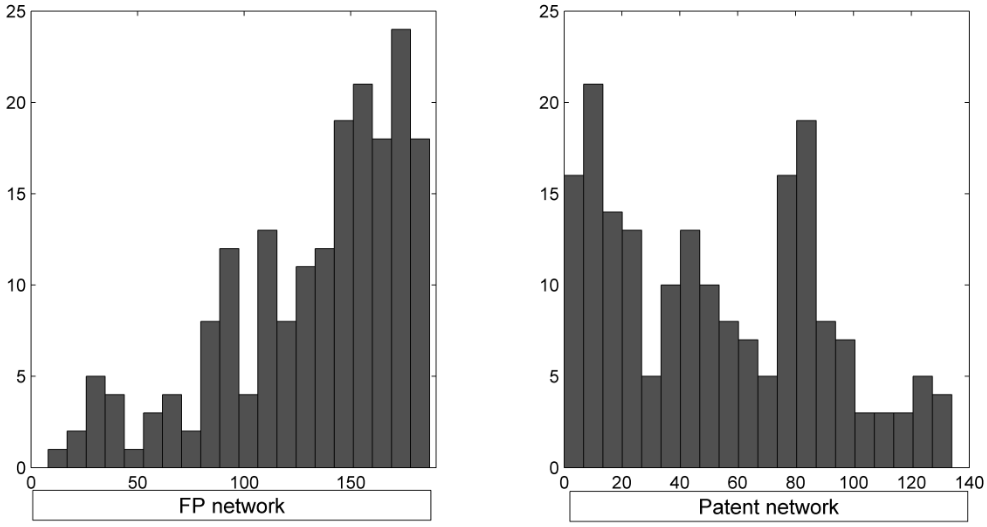

Figure 3. Degree distribution of the FP and patent networks

the vertical axis shows the frequencies of each degree value. The two figures show very different pictures.

The degree distribution is asymmetric in both cases, but the 'direction' of the asymmetry is the opposite. In the FP network the majority of the nodes have lots of links while there are only few regions (nodes) left with few links. The distribution clearly shows the source of the high value for the average degree calculated before. On the other hand, the co-patenting network exhibits an opposite direction of asymmetry: there are relatively few nodes with large number of links while the majority of the nodes have few links. This type of asymmetry is referred to as scale-free property of the network (Barabási and Albert 1999).

The difference between the two networks reveal that in the case of the FP network more or less all regions are involved tightly in the cooperation and only the minority seem to be less connected. In the co-patenting network there are some dominant regions while the majority is less connected. It is important to note that the co-patenting network represents a structure, which has been developing bottom-up, upon the decisions of individual economic actors. On the other hand the FP network contains some top-down elements in its structure in the form of tender evaluation.

5. Interregional knowledge network quality and research productivity of European regions: Empirical analysis

In this section we present estimation results for equation (11). Knowledge production in scientific research (resulting in publications) and technological inventions (measured by patents) will be investigated for 189 EU NUTS 2 regions. Cross-sectional econometric estimations for parameter β_1 will receive a particular attention as this parameter proxies interregional ego network effects on regional research productivity. Additional to core model variables we control for local knowledge flow impacts estimated by the parameter of the variable AGGL measuring agglomeration of knowledge intensive industries in the region[11]. We also test for spatial dependence in order to control for interregional knowledge

[11] Following Varga, Pontikakis and Chorafakis (2010) the index is a size-adjusted (in the spirit of the index developed by Elison and Glaeser 1997) variation of the popular location quotient (LQ) measure and is calculated as:

$AGGL_i = [(EMPKI_i / EMPKI_{EU}) / (EMP_i / EMP_{EU})] / [1 - \sum_j (EMPKI_{i,j} / EMPKI_{j,EU})]*[1 - (EMP_i / EMP_{EU})]$,

where $EMPKI_j$ and EMPKI are employment in knowledge intensive economic sector j and the total of knowledge intensive sectors, EMP is total employment and the subscripts *i* and *EU* stand for region and EU aggregate respectively. A significant and positive parameter of AGGL indicates a positive relation between knowledge output (publications or patents) and the agglomeration of knowledge intensive industries usually found instrumental in innovation such as high and medium technology manufacturing and business services. As common in KPF studies we interpret this result as a sign of influential knowledge flows from the local knowledge intensive industry to the production of new knowledge.

flows communicated by channels different from FP collaborations or co-patenting. In this respect the significant parameter of the spatial lag variable is taken as a sign of the role of such interregional knowledge flows[12]. Due to the presence of the interaction term in equation (11) multicollinearity is a potential problem. We test for its presence by the Multicollinearity Condition Number (MCN)[13].

Table 4. Regression Results for Log (Patents) for 189 EU regions, 2002 (N=189)

Model	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Estimation	OLS	OLS	OLS	OLS	OLS	OLS	OLS	ML-Spatial Lag (NEIGH)
Constant	-1.972***	-1.643***	-1.501***	0.101	-0.769**	-2.071***	-0.837	-0.791
	(0.335)	(0.353)	(0.332)	(0.433)	(0.375)	(0.463)	(0.593)	(0.566)
W_Log(PAT)								0.027***
								(0.008)
Log(GRD(-2))	1.123***	1.044***	0.876***	0.899***	0.602***	0.638***	0.497***	0.526***
	(0.058)	(0.064)	(0.076)	(0.062)	(0.106)	(0.101)	(0.108)	(0.104)
Log(GRD(-2))* PATKPW		0.223***						
		(0.086)						
Log(GRD(-2))* PATLDW			1.094***					
			(0.233)					
Log(GRD(-2))* PATGEW				-0.504***				
				(0.076)				
Log(GRD(-2))* PATENQW					0.456***	0.294***	0.257***	0.195**
					(0.080)	(0.085)	(0.084)	(0.082)
Log(PATSTCKN(-2))						0.192***	0.171***	0.114***
						(0.043)	(0.043)	(0.043)
Log(AGGL(-2))							1.195***	1.149***
							(0.372)	(0.355)
R^2-adj	0.67	0.68	0.70	0.73	0.72	0.74	0.76	0.77
LIK								-269
Multicollinearity Condition Number	7.4	8.8	11.0	11.5	15.7	19.5	22.2	22.2
LM-Err	30.46***	21.36***	14.19***	8.56***	12.41***	9.03***	5.67**	
Neigh	59.21***	34.31***	22.63***	9.13***	12.19***	5.38**	2.33	
INV1	16.60***	9.96***	6.58**	2.75*	4.01**	1.69	0.98	
INV2								
LM-Lag	42.13***	35.01***	28.92***	20.07***	23.39***	12.65***	12.65***	
Neigh	48.06***	39.82***	30.90***	18.51***	21.25***	10.94***	7.09***	
INV1	17.04***	13.01***	9.92***	3.75**	4.64***	3.08*	1.23***	
INV2								
LR-Lag								12.349***
LM-Err								
Neigh								1.456
INV1								0.313
INV2								0364

Notes: Estimated standard errors are in parentheses; spatial weights matrices are row-standardized: Neigh is neighborhood contiguity matrix; INV1 is inverse distance matrix; INV2 is inverse distance squared matrix; W_Log(PAT) is the spatially lagged dependent variable where W stands for the weights matrix INV2. *** indicates significance at $p < 0.01$; ** indicates significance at $p < 0.05$; * indicates $p < 0.1$.

---

[12] The general expression for the spatial lag model is

$$y = \rho Wy + x\beta + \varepsilon,$$

where y is an N by 1 vector of dependent observations, Wy is an N by 1 vector of lagged dependent observations, ρ is a spatial autoregressive parameter, x is an N by K matrix of exogenous explanatory variables, β is a K by 1 vector of respective coefficients, and ε is an N by 1 vector of independent disturbance terms. Because the spatially lagged dependent term is correlated with the errors and as such endogenous, the OLS estimator is biased and inconsistent. Instead of OLS, other estimation methods such as Maximum Likelihood, Instrumental Variables or General Methods of Moments must be applied to the spatial lag model (Anselin 1988).

[13] The value of MCN exceeding 30 suggests a potential problem of specification (Belsley, Kuh, and Welsch 1980).

Table 5. Robustness Specifications for Log (Patents) for 189 EU regions, 2002 (N=189)

Model	(1)	(2)	(3)	(4)
Estimation	ML-Spatial Lag (NEIGH)	ML-Spatial Lag (NEIGH)	ML-Spatial Lag (NEIGH)	ML-Spatial Lag (NEIGH)
Constant	-0.862	-1.140**	-0.539	-1.335**
	(0.585)	(0.550)	(0.695)	(0.525)
W_Log(PAT)	0.028***	0.029***	0.031***	0.029***
	(0.008)	(0.008)	(0.008)	(0.007)
Log(GRD(-2))	0.528***	0.663***	0.493***	0.621***
	(0.121)	(0.081)	(0.105)	(0.098)
Log(GRD(-2))* PATENQ	0.171*			
	(0.096)			
Log(GRD(-2))* PATDEGW		0.080		
		(0.067)		
Log(GRD(-2))* PATENQW			0.242***	
			(0.080)	
Log(GRD(-2))* FPENQW				0.113
				(0.094)
Log(PATSTCKN(-2))	0.130***	0.140***		0.140***
	(0.043)	(0.043)		(0.043)
Log(PUBSTCKN(-2))			0.071	
			(0.065)	
Log(AGGL(-2))	1.204***	1.272***	1.281***	1.277***
	(0.356)	(0.356)	(0.357)	(0.356)
R^2-adj	0.77	0.77	0.76	0.77
LIK	-270	-271	-272	-271
Multicollinearity Condition Number	24.5	17.2	24.1	19.6
LR-Lag	13.40***	14.17***	16.45***	14.69***
LM-Err				
Neigh	1.036	1.173	1.745	1.443
INV1	0.231	0.312	0.479	0.603
INV2	0.306	0.447	0.538	0.537

Notes: Estimated standard errors are in parentheses; spatial weights matrices are row-standardized: Neigh is neighborhood contiguity matrix; INV1 is inverse distance matrix; INV2 is inverse distance squared matrix; W_Log(PAT) is the spatially lagged dependent variable where W stands for the weights matrix INV2. *** indicates significance at $p < 0.01$; ** indicates significance at $p < 0.05$; * indicates $p < 0.1$.

Table 4 presents the results on regional patenting. The highly significant and positive parameter of the research variable and the good regression fit in the second column are usual findings in the KPF literature on patenting. Models 2 to 4 test the separate impacts of the three dimensions of ENQ on research productivity. Impacts with binary and weighted patent networks were both estimated but since estimations for the latter network resulted in better fits to the data, the weighted network results are presented here. Knowledge potential (PATKPW) and the intensity of collaborations (PATLDW) in the ego network are both positively related with research productivity as signaled by highly significant coefficients.

The estimated parameter of global embeddedness (PATGEW) is significant but negative. It is the particular structure of the entire co-patenting network that lies behind this somewhat surprising finding of negative association between global knowledge access and regional R&D productivity. Data clearly tell us that European co-patenting network follows a strong core-periphery structure. High patenting regions in the core tend to collaborate with each other almost exclusively while low patenting regions in the periphery rarely extend their ties to the core[14]. Since most of the knowledge agglomerates in the core, additional knowledge that core regions can learn from the periphery is only marginal. On the other hand, peripheral regions may struggle from absorptive capacity deficits[15]. We decided to leave this result (and the one that echoes it for publications in Table 6) as it is. Our data do not allow more detailed analysis at this point. Information on specific technologies over a longer time frame will certainly provide a deeper knowledge about the role of global embeddedness in regional R&D productivity. With such data at hand, experiments with different weighting methods in GE (equation 3) could also become possible.

Overall considered, it is clear from Table 4 that accounting for the impact of the three sub-indices (that measure the three dimensions of ego network quality) increases regression fit. Additionally, differences in the estimated values of β_0 and β_1 in Models 2 to 4 suggest that the three sub-indices indeed capture different dimensions of ego network quality. ENQ enters the equation with a strongly significant and positive parameter in Model 5 resulting in an equation with R-squared 7.5 percent higher than without considering the effects of interregional networks on R&D productivity in Model 1. Repeating earlier results in the literature (e.g., Varga, Pontikakis an Chorafakis 2010) it is found that both national knowledge stock and agglomeration of knowledge intensive industries affect regional patenting positively. The positive and highly significant parameter of the spatially lagged dependent variable suggests that knowledge flows between neighboring regions are important sources of invention even after controlling for the impact of knowledge communicated through interregional co-patenting networks.

[14] Note the similarity between the structure of the empirical patent network and that of the theoretical scalefree network in Figure 2. Though a somewhat lesser extent but a similar pattern exists in the FP5 network, which explains the comparable findings on GE impacts for publication research productivity (see Table 6 for further details).

[15] As a proximate measure of the relative size of knowledge accessed from outside the ego network we calculated the share of GE over ENQ for each region in the sample. Core regions yield extremely low values (e.g., for Ile de France is is below 1 percent) while on the perihpery the share of globally accessible knowledge above 90 percent is not an exception. This suggests that for several regions in the periphery globally available knowledge can be about nine times higher than the knowledge accessible from their individual networks. We also experimented with different methods to separate core and peripheral regions empirically. Estimation results (not reported here) suggest that the core and the perihpery indeed follow different patterns in utilizing global knowledge while generating new technologies.

Table 6. Regression Results for Log (Publications) for 189 EU regions, 2002 (N=189)

Model	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Estimation	OLS	OLS	OLS	OLS	OLS	OLS	OLS	ML-Spatial Lag (INV1)
Constant	1.402*** (0.229)	2.420*** (0.262)	2.659*** (0.285)	2.554*** (0.280)	2.584*** (0.288)	2.039*** (0.517)	1.980*** (0.402)	1.859*** (0.389)
W_Log(PUB)								-0.004*** (0.001)
Log(GRD(-2))	0.941*** (0.039)	0.206* (0.120)	0.300*** (0.105)	0.794*** (0.043)	0.462*** (0.088)	0.366*** (0.114)	0.377*** (0.095)	0.459*** (0.095)
Log(GRD(-2))* FPKP		0.612*** (0.096)						
Log(GRD(-2))* FPLD			0.490*** (0.076)					
Log(GRD(-2))* FPGE				-0.536*** (0.087)				
Log(GRD(-2))* FPENQ					0.346*** (0.058)	0.380*** (0.060)	0.377*** (0.059)	0.334*** (0.058)
Log(PUBSTCKN(-2))						0.096** (0.045)	0.096** (0.045)	0.140*** (0.045)
Log(AGGL(-2))						0.049 (0.267)		
R^2-adj	0.75	0.80	0.80	0.79	0.79	0.79	0.80	0.81
LIK								-209
Multicollinearity Condition Number	7.4	27.5	23.6	10.5	19.1	27.9	23.3	23.3
LM-Err								
Neigh	0.123	0.118	0.065	0.128	0.117	0.277	0.261	
INV1	1.213	1.078	0.394	0.206	0.359	1.144	1.070	
INV2	0.172	0.253	0.085	1.044	0.031	0.210	0.201	
LM-Lag								
Neigh	11.008***	7.045***	4.944**	6.330**	4.461**	8.139***	8.164***	
INV1	12.988***	6.348**	4.895**	1.528	5.506**	10.833***	9.704***	
INV2	5.106**	1.407	1.288	6.321**	1.977	2.820*	2..523	
LR-Lag								10.50***
LM-Err								
Neigh								0.001
INV1								0.002
INV2								0.267

Notes: Estimated standard errors are in parentheses; spatial weights matrices are row-standardized: Neigh is neighborhood contiguity matrix; INV1 is inverse distance matrix; INV2 is inverse distance squared matrix; W_Log(PAT) is the spatially lagged dependent variable where W stands for the weights matrix INV2. *** indicates significance at $p < 0.01$; ** indicates significance at $p < 0.05$; * indicates $p < 0.1$.

Table 5 shows some robustness results. Model 1 is exhibited to demonstrate the slightly weaker performance of ENQ with binary connections. The message of Model 2 is that R&D productivity in patenting is not associated with the pure number of interregional connections though this aspect of an ego network is frequently applied in spatial studies. Instead, as detailed above, ENQ and its dimensions are the network features that are significantly related to the productivity of research. The last two columns demonstrate that patenting follows a different path than scientific publications: publication stocks (measuring national development in scientific research) are not a substitute for patent stocks (i.e., national technological knowledge) and ENQ of FP 5 research networks is not an alternative of ENQ in co-invention.

Table 7. Robustness Specifications for Log (Publications) for 189 EU regions, 2002, (N=189)

Model	(1)	(2)	(3)	(4)
Estimation	ML- Spatial Lag (INV1)	ML- Spatial Lag (INV1)	OLS	ML- Spatial Lag (INV1)
Constant	1.199*** (0.394)	1.167*** (0.404)	2.443*** (0.108)	1.110** (0.496)
W_Log(PUB)	-0.005*** (0.001)	-0.005*** (0.001)		-0.005*** (0.001)
Log(GRD(-2))	0.835*** (0.063)	0.959*** (0.055)	0.598*** (0.108)	0.973*** (0.092)
Log(GRD(-2))* FPENQW	0.182** (0.071)			
Log(GRD(-2))* FPDEGW		-0.009 (0.077)		
Log(GRD(-2))* FPENQ			0.338*** (0.057)	
Log(GRD(-2))* PATENQ				-0.017 (0.078)
Log(PUBSTCKN(-2))	0085* (0.048)	0.096* (0.049)		0.098** (0.049)
Log(PATSTCKN(-2))			-0.103** (0.049)	
R^2-adj	0.78	0.78	0.79	0.78
LIK	-221	-224		-224
Multicollinearity Condition Number	20.0	18.6	26.0	24.1
LM-Err Neigh INV1 INV2			0.001 0.001 0.001	
LM-Lag Neigh INV1 INV2			2.682 2.590 0.978	
LR-Lag	19.08***	17.94***		14.72***
LM-Err Neigh INV1 INV2	0.010 0.019 0.321	0.254 0.154 0.103		0.251 0.142 0.084

Notes: Estimated standard errors are in parentheses; spatial weights matrices are row-standardized: Neigh is neighborhood contiguity matrix; INV1 is inverse distance matrix; INV2 is inverse distance squared matrix; W_Log(PAT) is the spatially lagged dependent variable where W stands for the weights matrix INV2. *** indicates significance at p < 0.01; ** indicates significance at p < 0.05; * indicates p < 0.1.

Table 6 and 7 exhibit empirical results for regional R&D productivity in scientific publications. Models 1 to 5 in Table 6 follow the patterns already discussed for patent production: components of ENQ (this time with binary network data) enter the equation with significant and varying parameters, and the parameter of ENQ is itself positive and significant while regression fit increases. National level scientific knowledge relates positively to publication output, but regional agglomeration of technology-intensive industries has no significant effect on the production of new scientific knowledge (Models 6 and 7). This finding together with the significant network quality effect suggests that success in scientific research is more related to international and national embeddeddness than to industrial agglomeration. This result echoes what is a frequent observation: high quality research universities are not necessarily located in large cities. They can be equally successful in scientific knowledge production in smaller but well-connected geographical areas (e.g., Varga 2000). The significant but negative coefficient of the spatially lagged dependent variable in Model 8 underlines what is just observed about the role of agglomeration. Regions with high levels of scientific publication do not necessarily locate in spatial proximity, but they tend to scatter geographically. This latter finding underlines what is suggested by Models 6 and 7 in Table 6. Knowledge inputs to scientific research can successfully be transported over large distances due perhaps to the presence of institutional and cognitive proximities (Boschma 2005).

The robustness results in Table 7 are summarized as follows. The presence of network connections seems to be more relevant for R&D productivity in scientific publications than the frequency of interactions as suggested by the lower fit and a less significant ENQ parameter in Model 1 (using weighted network data). Similar to what is found for patenting, the number of collaborations does not affect research productivity. This finding suggests again that research productivity is related to several additional network features captured by the more complex ENQ variable and not to the pure number of connections, which is a commonly applied variable in spatial network studies. Models 3 and 4 again show that regional publication and patenting are governed by different principles. National technological knowledge does not necessarily go hand in hand with outstanding regional scientific production (Model 3). A similar message is communicated by Model 4: the quality of technological knowledge accessed through co-patenting networks is not related to research productivity in publications. Instead, R&D productivity of scientific knowledge creation relies on the quality of research collaboration networks like the ones financed by the EU Framework Programs.

6. Summary

In this paper we contributed to the emerging literature on the role of interregional knowledge flows in the regional production of new (scientific and technological) knowledge. An especially promising tool of research in this area is network analysis, which is applied in our study as well.

To structure the problem of interregional knowledge network effects on research productivity at the regional level we directed attention to the value of knowledge that can be accessed from a particular network. To measure such knowledge value we introduced the index of ego network quality (ENQ). ENQ summarizes three features of networks: the knowledge already accumulated by immediate network partners (knowledge potential – KP), the frequency of collaborations among immediate network partners (local density – LD) and the region's embeddeddness in the entire knowledge network (global embeddeddness – GE).

A systematic spatial econometric analysis were then carried out with European regional data on the role of ENQ in research addressing the development of technological inventions (measured by patents) on the one hand and scientific publications on the other. We found that the quality of interregional networks in both areas of knowledge production is indeed a significant contributor to R&D productivity. We also found that the pure number of collaborations, which is the most frequently used variable in spatial network studies is not a suitable proxy of interregional network effects in R&D productivity contrary to ENQ. Our results show that a more comprehensive approach taking into account several local and global features of the network surrounding the given region provides better insights into the network effects in regional knowledge production.

Our research has a potential relevance for regional development policy as well. Parallel with the worldwide emergence of interest in place-based local development, beginning in 2014 the new European Cohesion Policy will adopt "smart specialization" as a promising alternative to conventional economic development policies supported by the EU for decades (Barca 2009, McCann and Ortega-Argilés 2011). One of the core elements of the smart specialization idea is local development supported by the transfer of knowledge originated in more advanced regions (Foray, David and Hall 2009). However, empirical evidence on the role of interregional knowledge networks on regional development is still limited. The application of the concept of the quality of interregional knowledge networks measured by ENQ may open the possibility of a more systematic research in this area.

We consider this paper as a first step into the research field of network quality impacts on regional knowledge production. Further efforts with more detailed data in the dimensions of time, scientific and technological areas and geography will potentially extend our knowledge in this field.

References

Acs, Z., Anselin, L., and Varga, A. (2002) Patents and innovation counts as measures of regional production of new knowledge. *Research Policy* 31, 1069-1085.

Anselin, L., Varga, A. and Acs, Z. (1997) Local geographic spillovers between university research and high technology innovations. *Journal of Urban Economics* 42, 422–448.

Barabási, A., and Albert, R. (1999) Emergence of scaling in random networks. *Science 286* (5439), 509-512.

Barca, F. (2009) *An agenda for a reformed cohesion policy. A place-based approach to meeting European Union challenges and expectations*. Independent Report prepared at the request of Danuta Hübner, Commissioner for Regional Policy, April 2009. Independent Report prepared at the request of Danuta Hübner, Commissioner for Regional Policy

Belsley, D., Kuh, E., and Welsch, R. (1980): *Regression Diagnostics, Identifying Influential Data and Sources of Collinearity*. New York: Willey.

Bonacich, P. (1972) Factoring and weighting approaches to clique identification. *Journal of Mathematical Sociology* 2, 113-120.

Bonacich, P. (2007) Some unique properties of eigenvector centrality. *Social Networks* 29, 555-564.

Boschma, R. (2005) Proximity and innovation: A critical assessment. *Regional Studies* 39, 61-74.

Breschi, S. and Lenzi, C. (2011) *Net and the city. Co - invention networks and the inventive productivity of US cities*. mimeo.

Breschi, S. and Lissoni, F. (2009). Mobility of inventors and networks of collaboration: An anatomy of localised knowledge flows. *Journal of Economic Geography* 9, 439-468.

Burt, R.S. (1992): *Structural Holes*. Harvard University Press. Cambridge, MA.

Burt, R.S., Hogarth, R.M., Michaud C. (2000): The Social Capital of French and American Managers. *Organization Science* 11:123-147.

Burton, P., Yu W., Prybutok, V. (2010): Social network position and its relationship to performance of IT professionals.(Report). Informing Science: the International Journal of an Emerging Transdiscipline. Informing Science Institute. *HighBeam Research*

Cainelli, G., Maggioni, M., Uberti, E and De Felice, A. (2010) *The strength of strong ties: co-authorship and productivity among Italian economists*. 'Marco Fanno' working papers 125, Department of Economics, University of Padova.

Cantwell, J. and Iammarino, S. (2003) *Multinational corporations and European regional systems of innovation*. Routledge.

Coleman, J.S. (1986): Social Theory, Social Research, and a Theory of Action. *American Journal of Sociology*, 91: 1309-1335.

Cross, R., Cummings, J.N. (2004) Tie and Network Correlates of Individual Performance in Knowledge-Intensive Work. *The Academy of Management Journal*. 47 928-937.

Donckels, R. Lambrecht, J. (1997): The Network Position of Small Businesses: An Explanatory Model. *Journal of Small Business Management*, 35(2): 13-26.

Feldman, M. and Audretsch, D. (1999) Innovation in cities: Science-based diversity, specialization and localized competition. *European Economic Review* 43, 409-429.

Fischer, M. and Varga, A. (2002) Technological innovation and interfirm cooperation. An exploratory analysis using survey data from manufacturing firms in the metropolitan region of Vienna. *International Journal of Technology Management* 24, 724-742.

Foray, D., David, P. and Hall, B. (2009) *Smart Specialisation – The Concept*. Knowledge Economists Policy Brief No 9, June 2009.

Freeman, L.C. (1979) Centrality in social networks: I. Conceptual clarification. *Social Networks*. 1, 215-239.

Frenken, K., van Oort F. and Verburg, T. (2007) Related variety, unrelated variety and regional economic growth. *Regional Studies* 41(5): 685-697.

Fujita M and Thisse J (2002) *Economics of Agglomeration. Cities, Industrial Location, and Regional Growth.* Cambridge University Press Cambridge, MA, London, England.

Furman, J. L., Porter, M. E. and Stern, S. (2002) The determinants of national innovative capacity. *Research Policy*, 31: 899-933

Ghinamo M. (2010) *Explaining the variation in the empirical estimates of academic knowledge spillovers*. Working Paper 2010/2, Universit‡ degli Studi del Piemonte Orientale "A. Avogadro" Department of Economics and Quantitative Methods (SEMeQ).

Glaeser, E.L., Kallal, H.D., Scheinkman, J.A., and Shleifer, A. (1992) Growth in cities. *Journal of Political Economy 100*(6): 1126-1152.

Griliches Z. (1990) Patent statistics as economic indicators: a survey. *Journal of Economic Literature* 20, 1661-1707.

Hoekman, J., Frenken, K. and van Oort, F. (2009) Collaboration networks as carriers of knowledge spillovers: Evidence from EU27 regions. *Annals of Regional Science* 43, 721-738.

Hopp, W.J., Iravani, S., Liu, F., Stringer, M.J. (2010) *The Impact of Discussion, Awareness, and Collaboration Network Position on Research Performance of Engineering School Faculty*. Ross School of Business Paper No. 1164.

Jaffe, A. (1989) Real effects of academic research, *American Economic Review 79*: 957-970.

Jaffe, A., Trajtenberg, M., and Henderson, R. (1993) Geographic localization of knowledge spillovers as evidenced by patent citations, *Quarterly Journal of Economics 108*: 577–598.

Jones, C. (1995) R&D-based models of economic growth, *Journal of Political Economy 103*(4): 759-784.

Koschatzky, K. (2000) *The regionalization of innovation policy in Germany – Theoretical foundations and recent evidence*, Working Papers Firms and Regions No. R1/2000, Fraunhofer Institute for Systems and Innovation Research (ISI), Department “Innovation Services and Regional Development”.

Kretschmer, H. (2004). Author productivity and geodesic distance in bibliographic co-authorship networks , and visibility on the Web. *Scientometrics*, 60(3): 409-420.

Lundvall, B.A. (1992) *National Systems of Innovation*. Pinter Publishers, London.

Maggioni, M. and Uberti, T. (2011) Networks and geography in the economics of knowledge flows. *Quality and Quantity*, 1031-1051.

Maggioni, M., Nosvelli, M. and Uberti, E. (2006) Space Vs. Networks in the Geography of Innovation: A European Analysis. *Papers in Regional Science* 86, 471-493.

Maier, G., Kurka, B. and Trippl, M. (2007) *Knowledge Spillover Agents and Regional Development: Spatial Distribution and Mobility of Star Scientists*. WP 17/2007, Wirtschaftsuniversität Wien.

McCann, P. and Ortega-Argilés, R. (2011) *Smart Specialisation, Regional Growth and Applications to EU Cohesion Policy*. Economic geography working paper 2011: Faculty of Spatial Sciences, University of Groningen.

Miguélez, E. and Moreno, R. (2012) *Skilled Labour Mobility, Networks and Knowledge Creation in Regions: A Panel Data Approach*. Unpublished manuscript.

Miguélez, E., Moreno, R. and Suriñach, J. (2009) *Inventors on the move: tracing invertors' mobility and its spatial distribution*. Research Institute of Applied Economics Working Papers 2009/16, University of Barcelona.

Nelson, R.R. (ed.) (1993) *National Innovation Systems: A Comparative Analysis*. Oxford University Press, Oxford.

OECD (2009) REGPAT Database, October 2009.

Ponds, R. van Oort, F. and Frenken, K. (2009) Internationalization and regional embedding of scientific research in the Netherlands. In Varga, A. (Ed.) *Universities, Knowledge Transfer and Regional Development: Geography, Entrepreneurship and Policy*. Edward Elgar Publishers, 109–137.

Powell, W.W., Koput, K.W., Smith-Doerr, L., Owen-Smith, J. (1999) Network Position and Firm Performance: Organizational Returns to Collaboration in the Biotechnology Industry. In Andrews, S. B., Knoke, D. (eds.): *Networks In and Around Organizations*. JAI Press, Greenwich, CT.

Diez, R. (2002) Metropolitan Innovation Systems - A comparison between Barcelona, Stockholm, and Vienna. *International Regional Science Review* 25, 63-85.

Romer, P. M. (1990) Endogenous technological change, *Journal of Political Economy 5*(98): S71-S102.

Rumsey-Wairepo, A. (2006) *The association between co-authorship network structures and successful academic publishing among higher education scholars*. Brigham Young University.

Salmenkaita, J. P. (2004) Intangible capital in industrial research: Effects of network position on individual inventive productivity. In Bettis, R. (Ed.) *Strategy in transition*. Blackwell Publishing, 220-248, Malden, MA.

Schiller, D. and Diez, J. (2008) *Mobile star scientists as regional knowledge spillover agents*. IAREG Working Paper WP2/7.

Tsai, W. (2001): Knowledge Transfer in Intraorganizational Networks: Effects of Network Position and Absorptive Capacity on Business Unit Innovation and Performance. *The Academy of Management Journal*, 44(5): 996-1004.

Van Der Deijl, H., Kelchtermans, S., Veugelers, R. (2011) *Researcher networks and productivity*. Paper presented at the DIME-DRUID ACADEMY Winter Conference 2011.

Van Raan, A. F. J. (2004) Measuring science. In Moed, H. F., Glänzel, W. and Schmoch, U. (Eds) *Handbook of Quantitative Science and Technology Research: The Use of Publication and Patent Statistics in Studies of S&T Systems*. Kluwer Academic Publishers, 19-50, Dordrecht.

Varga, A. (2000) Local academic knowledge transfers and the concentration of economic activity, *Journal of Regional Science 40*(2): 289-309.

Varga, A. (2006) The spatial dimension of innovation and growth: empirical research methodology and policy analysis, *European Planning Studies 9*: 1171-1186.

Varga, A. and Parag, A. (2009) Academic knowledge transfers and the structure of international research networks. In Varga, A. (Ed.) *Universities, Knowledge Transfer and Regional Development:*

Geography, Entrepreneurship and Policy. Edward Elgar Publishers, 138-159.

Varga, A., Pontikakis, D. and Chorafakis, G. (2010) *Agglomeration and interregional network effects on European R&D productivity*. Working Paper IAREG WP5/25.

Watts, D.J., Strogatz, S.H. (1998) Collective dynamics of 'small-world' networks. *Nature*, 393(6684): 409–410.

Zaheer, A., Bell, G.G. (2005) Benefiting from network position: firm capabilities, structural holes and performance. *Strategic Management Journal*, 26: 809-825.

Zucker L.G. Darby, M.R.,Furner, J., Liu R.C. and Ma, H. (2007) Minerva unbound: Knowledge stocks, knowledge flows and new knowledge production, *Research Policy*, Vol. 36, No. 6, pp. 850-863.

Appendix: Some special cases of ENQ

A.1. ENQ without distance weights

Consider the situation in which distance weights are not considered, i.e. $W_d = 1$ for all d, and knowledge levels are identical across nodes, that is, we can use $K_i = 1$ for all i, therefore (as mentioned before) $KP_d^i = \sum_{j,g_{ij}=d} a_{ij} = N_d^i$. This way the previous formula collapses to

$$ENQ^i = \sum_{d=1}^{S-1} C_d^i - C_{d-1}^i = C_{M-1}^i$$

As S_{M-1}^i is the largest subgraph possible, containing all nodes, in this very special case ENQ_i is simply the number of links in the network, irrespective of the node in question.

A.2. ENQ without knowledge weights

If knowledge weights are not used but distance weights are in effect, we have the following formula for ENQ:

$$ENQ^i = \sum_{d=1}^{S-1} W_d \left(\frac{\sum_{j,g_{ij}=d} \sum_{k,g_{ik}=d} a_{jk}}{2} + \sum\nolimits_{j,g_{ij}=d-1} \sum\nolimits_{k,g_{ik}=d} a_{jk} \right)$$

The first term in the parenthesis gives the number of links between nodes at distance d, whereas the second term counts the number of links connecting distance d with distance $d-1$. If we expand the expression in the parenthesis with the term $\sum_{j,g_{ij}=d} \sum_{k,g_{ik}=d+1} a_{jk}$ (which counts the links between nodes at distance d and $d+1$), we obtain a number which gives the sum of degrees of nodes at distance d. After this expansion we can write

$$DIST^i = \sum_{d=1}^{S-1} W_d \left(\frac{\sum_{j,g_{ij}=d} \sum_k a_{jk}}{2} \right) = \frac{1}{2} \sum_{d=1}^{S-1} W_d \sum\nolimits_{j,g_{ij}=d} DEG_j$$

This last measure is nothing else than the distance-weighted sum of degrees in the network. On the other hand, this last expression bears a close resemblance to the eigenvector centrality, which also reflects a distance-weighted sum of degrees in a network, although it uses a recursive definition with exponential weights leading to an eigenvector problem. This means that our ENQ index, when knowledge levels are homogenous, reflects similar properties to eigenvector centrality, which is a comprehensive measure of network position taking into account the whole structure around a given node from its immediate neighborhood to farther parts of the network.

A KRTI eddig megjelent műhelytanulmányai

Varga Attila: From the geography of innovation to development policy analysis: The GMR-approach (2007/1)

Bessenyei István: Növekedési pólusok a térben és a társadalomban (2007/2)

Darvas Zsolt - Schepp Zoltán: Kelet-közép-európai devizaárfolyamok előrejelzése határidős árfolyamok segítségével (2007/3)

Varga Attila: GMR-Hungary: A Complex Macro-Regional Model for the Analysis of Development Policy Impacts on the Hungarian Economy (2007/4)

Reiff Ádám - Zsibók Zsuzsanna: Az infláció és az árazási magatartás regionális jellemzői Magyarországon, mikroszintű adatok alapján (2008/1)

Varga Attila - Parag Andrea: Egyetemi tudástranszfer és a nemzetközi kutatási hálózatok szerkezete (2008/2)

Schepp Zoltán - Szabó Zoltán: Felsőoktatás-politika és állami finanszírozás: a 2007. évi felvételi tanulságai a gazdaságtudományi alapképzésben (2008/3)

Kaposi Zoltán: Város és agrárrendszer a polgárosodás korában (1850-1914) (a mezőgazdaság változásai Nagykanizsán) (2008/4)

Barancsuk János: Néhány gondolat az „árelfogadó" és „ármeghatározó" fogalmak jelentéséről (2009/1)

Kiss Gy. Kálmán: A szövetkezeti bank megteremtésének kísérlete Magyarországon (2009/2)

Zeller Gyula: Létezik-e a Smith probléma, avagy mennyire egységesek Adam Smith nézetei? (2009/3)

Járosi Péter - Atsushi Koike - Mark Thissen - Varga Attila: Regionális fejlesztéspolitikai hatáselemzés térbeli számítható általános egyensúlyi modellel: a GMR-Magyarország SCGE modellje (2009/4)

Mellár Tamás: Felemás magyar modernizáció (2009/5)

Szabó Zoltán: Az új paternalizmus: a nem-racionális hitelfelvevői magatartás és a túlzott eladósodás néhány gazdasági viselkedéstani összefüggése (2009/6)

Erdős Katalin-Varga Attila: Az egyetemi vállalkozó: legenda vagy valóság az európai regionális fejlődés elősegítésére? (2009/7)

Sebestyén Tamás: Innovation and Diversity in a Dynamic Knowledge Network. (2010/1)

Mellár Tamás: Válaszút előtt a makroökonómia? (2010/2)

Attila Varga- Dimitrios Pontikakis- George Chorafakis: Agglomeration and interregional network effects on European R&D productivity (2010/3)

Attila Varga - Péter Járosi - Tamás Sebestyén: Geographic Macro and Regional Model for EU Policy Impact Analysis of Intangible Assets and Growth (2010/4)

Rappai Gábor – Szerb László: Összetett indexek készítése új módon: a szűk keresztmetszetekért történő büntetés módszere. (2011/1)

Mellár Tamás: Néhány gondolat a makroegyensúly értelmezéséhez (2011/2)

Balatoni András - Mellár Tamás: Rövid távú előrejelző modell Magyarországra (2011/3)

Varga Attila - Járosi Péter - Sebestyén Tamás : A GMR-Európa modell és alkalmazása EU kohéziós politikai reformok előzetes hatásvizsgálata során (2011/4)

Sebestyén Tamás: Hálózati struktúra és egyensúly: a tudás-áramlás szerkezeti jellemzőinek kérdései (2011/5)

Bessenyei István: Az államadósság kezelésének prevenciós és rehabilitációs stratégiái. (2012/1)

Tamás Sebestyén – Attila Varga: Research Productivity and the Quality of Interregional Knowledge Networks (2012/2)