Regional productivity and knowledge transfer through patent co-inventorship – the role for network structure

Tamás Sebestyén
University of Pécs
Faculty of Business and Economics
Department of Economics and Regional Studies
Rákóczi Street 80., Pécs, Hungary, 7632

e-mail: sebestyent@ktk.pte.hu
tel: +36-72-501-599 / 3150

Abstract

The role and specific characteristics of knowledge transfer are proven to be an important factor in regional economic growth and continuous innovativeness. One the other hand, the role of personal interactions in this knowledge transfer is also emphasized by several studies, which draws the attention on the network of these interactions. At the same time this focus of interest got an impulse from the advances in network theory which achieved interesting and important results in the last decades. The integration of the two fields is starting to gain increasing interest among researchers of innovation, especially among those concerned with regional aspects. In this paper I follow this line of thinking by linking the network of patent inventors to the productivity of regions. The paper builds on a database of patent co-inventorship in the high-tech sector of three European countries and analyzes the effect of knowledge transfer, happening through these networks, on regional productivity. The results reveal a special role for the structure of networks in the analysis while showing that knowledge transfer through these networks indeed has a positive effect on regional productivity. The role for network structure is shown by the fact that the positive effect is observed if the specific scale-free structure of the underlying network is accounted for.

Keywords

Knowledge transfer, patent inventor network, scale-free network, regional productivity

1. Introduction

In recent years network-based approach to natural and social phenomena has become more popular. The distinguished attention can be devoted partly to the fact that quite a lot of phenomena can be described by the abstract concept of networks and partly to the research results showing that observed network structures display substantive similarities. (*Barabási* [2003], *Barabási et al.* [2000], *Csermely* [2005]). In the field of economics literature on innovation started to use network approaches in the first place. Studies on innovation diffusion rely on explicit modeling of networks more widely in contrast to previous approaches which were based either on a macro-scale or micro-based perspective (*Jackson and Wolinsky* [1996], *Abrahamson and Rosenkopf* [1997], *Bala and Goyal* [2000], *Cowan and Jonard* [2004], *Cowan et al.* [2006], *Carayol and Roux* [2009], *Sebestyén* [2010]).

Besides diffusion theory network analysis is found in empirical approaches to innovation. This line of research derives basically from the literature on knowledge spillovers (*Griliches* [1979], *Jaffe* [1986], *Griliches* [1992], *Jaffe et al.* [1993], *Feldman* [1994], *Anselin et al.* [1997]). These studies emphasize the role of geography in the diffusion of knowledge fertilizing a whole branch of literature on the relationship between locality and innovation. However, *Breschi and Lissoni* [2003] draw the attention on the fact that personal contacts have an important role in knowledge transfer and thus locality requires a more detailed approach for the analysis of knowledge spillovers and agglomeration effects. They emphasize that spatial proximity can be regarded as a proxy for social proximity: the former is important as long as it contributes to the establishment of social relationships and the development of trust embedded in these relationships. As spatial proximity largely eases the formation of these connections, social relationships will be locally dense and the agglomeration of innovative (or in a wider sense economic) activity will appear as an important medium for knowledge transfer. This hides the real situation where spillovers expound their effects through personal contacts and social networks and are local to the extent to which these networks themselves are local. On this line of thought several studies show that local effects of knowledge spillovers are based solely on labor mobility (*Zucker et al.* [1994], *Almeida and Kogut* [1999], *Balconi et al.* [2004]).

The role for social contact drew the attention of innovation researchers on the importance of social networks. In this sense mapping innovation-related networks proves to be a nontrivial task. Applications primarily rely on patent citations and patent cooperation to unfold these networks (*Ellis et al.* [1993], *Jaffe and Trajtenberg* [2002], *Verspagen* [2005], *Maggioni and Uberti* [2006], *Li et al.* [2007], *Maggioni et al.* [2010], *Gress* [2010], *Maggioni and Uberti* [2010], *Sebestyén and Parag* [2010]). However, these network based analyses do not pay much attention on the global structure of the networks, but concentrate on spatial dimensions. On the other hand, network theory emphasize that the global structure of the networks are an important aspect for explaining the overall performance of the network (*Jackson and Wolinsky* [1996], *Barabási* [2003], *Cowan et al.* [2006]).

In this paper the role of network structure is analyzed from a special perspective. Primarily, the role of network connections is examined in regional productivity: to what extent does the knowledge stock of other regions, available through network connections, contribute to the productivity of regions in addition to the own knowledge stocks of the regions. In order to do this, a database is presented which is available for the representation of interregional knowledge networks through patent co-inventorship data then the effect of these connections on regional productivity is analyzed. On the other hand, contradictory results on the role of network connections are resolved with the consideration of the role of global network structure: I show that it is the special scale-free structure of the network which accounts for the counter-intuitive results.

The second section of the paper the database is described with special interest on the data describing the knowledge networks. Then the third section shortly present the results obtained by the inferential analysis of the database and the seemingly paradoxical results are explained by a simple network model. In section four the empirical analysis of the intuitive explanation is given while the last section draws some conclusions.

The data

In this paper two data sources are exploited. The first is the regional database of the Eurostat which contains information (among other things) on GDP, employment and patenting activity. The other data source is a database developed at the University of Pécs which contains information on patent co-inventorship between NUTS2 regions of three European Countries (Germany, France and the United Kingdom). Some methodological issues are discussed below with respect the network data.

Network database

In contrast to economic indicators there are no directly available data on knowledge network thus such a database must be developed on the basis of other data sources. Using patent citations for this purpose is widely accepted as these citations show the trace of knowledge transfer to a certain extent (*Karki* [1997], *Oppenheim* [2000], *Chakrabarti et al.* [1993], *Chen and Hicks* [2004], *Singh* [2003]). Some studies examine networks of patent citations. *Gress* [2010] analyze relationships between technological fields in US patents with this methodology. The study draws conclusions on the originality and effectiveness of several patents and technological categories. *Li et al.* [2007] examine networks of patents, organizations, technological fields and countries in the nanotechnology sector. Their interesting result is that a central cluster can be shown for all types of these networks covering the majority of the connections and the scale-free structures can unambiguously detected.

Ejermo and Karlsson [2004] suggest that cooperation between patent inventors should be used instead of citations as these connections serve as a more robust proxy for knowledge transfer. The authors examine the role of inventor networks in spatiality for Sweden. Further studies based on inventorship networks can be found in *Maggioni and Uberti* [2006] or in *Maggioni et al.* [2010]. *Griliches* [1990] gives an overview on the advantages and drawbacks of using patent databases in spillover analysis.

In this study knowledge networks are built on a regional basis that is, the nodes of the networks are regions while the edges represent the intensity of knowledge transfer between regions. On the other hand, the network is developed on the basis of information on inventor cooperation meaning that two regions are said to be connected if inventors from these regions have cooperated on a patent. The more such cooperation is observed, the higher intensity of the connections is assumed.

According to these considerations a co-inventorship based regional network database was started to be developed at the University of Pécs in order to gain a representation on knowledge networks.[1] The database, at its present state, allows for the analysis of knowledge transfer between NUTS2 regions of three European countries (Germany, France and the United Kingdom). The data covers patents applied for at the European Patent Office and are also filtered on a sectoral basis: the database used here contains information on patents related to the high-tech sector. This means six areas in the sector: aviation, computer and automated business equipments, communication technology, laser, semiconductors, micro-organisms and genetic engineering.[2]

The network obtained is a weighted one as the intensity of cooperation may vary over time and across connections but the weighting can be done according to different criteria. The difference is the method how the intensity of knowledge transfer is derived (computed) from inventor cooperation. In this paper the intensity of knowledge transfer is determined such a way that the weight of a given connection is increased by one if inventors from the two regions connected by this link cooperated on one patent.

It is assumed that co-inventing means connection and knowledge transfer among all inventors or in other words the sub-network of the inventors of one specific patent is meant to be fully connected. Of course it may happen that the links between inventors are structured in a different way but the data source (patent statistics) does not provide information for the derivation of these ‘internal’ structures. At the same time the assumption of fully connected sub-networks does not mean any significant bias as typically a small number of inventors are matched to a patent, so the fully connected sub-network can not be very far from reality (assumption of a specific structure would be relevant for a larger number of co-inventors). The resulting network can contain loops so intra-regional patenting activity (knowledge transfer) can also be taken into account. However, this opportunity is only partially exploited in this paper.

An important extension for the above considerations is the situation when the same inventor contributes to different patents. This situation could be handled easily if a unique identifier would be allocated to each inventor. However, the data source provides inventor names in strings thus the cleaning of this information would require further efforts especially for the present case with several million records. On the other hand, the resulting bias would be significant only in that situation when the inventor changes his residence. In the present case, however, this change can also be regarded as interregional knowledge transfer although not in that interpersonal perspective on which the development of the network is based.

According to the methodology above, the annual network of inventor cooperation is available among the NUTS2 regions of the three countries considered. This is described by an $\mathbf{R}_t$ adjacency matrix where the r_{ijt} general element represents the strength (intensity) of connection between regions i and j in year t. As the development of the network is based on the assumption of symmetry, the matrix is also symmetric ($r_{ijt} = r_{jit}$).

---

[1] For the technical details on the development of the database see *Kruzslicz et al.* [2010].

[2] The IPC codes of these fields are the following. Aviation: B64B, B64C, B64D, B64F, B64G; computer and automated business equipments: B41J, G06C, G06D, G06E, G06F, G06G, G06J, G06K, G06M, G06N, G06T, G11C; communication technology: H04B, H04H, H04J, H04K, H04L, H04M, H04N, H04Q, H04R, H04S; laser: H01S; micro-organisms and genetic engineering: C12M, C12N, C12P, C12Q; semiconductors: H01L.

The determination of regional productivity

According to the aims formulated in the introduction the effect of external knowledge, available through network connections, on regional productivity is to be analyzed. For this the estimation of regional productivity and knowledge stocks is required in addition to the empirical assessment of network connections. The later part was tackled previously so the estimation of regional productivity and knowledge stocks is described in what follows.

The explained variable in the analysis below is the labor productivity of the regions under consideration. This can be computed from the publicly available data from the Eurostat with some minor corrections. First, nominal GDP values must be transformed into real terms. This can be done by using the data on the growth rate of real gross value added also available from the Eurostat. As the focus of this analysis is on the high-tech sector, labor productivity must be estimated for this sector as well which requires some additional corrections.[3]

High-tech GDP is computed with the help of two simple correction factors. In the first case the corrections factor is the ratio of patent related to the high-tech sector and the number of total patents. Patent data for the high-tech sector (and total patent number) are available from the regional database of the Eurostat. As the Eurostat classification was used for the development of inventor networks, the two different datasets are consistent. This kind of correction has the disadvantage that regional high-tech GDP values generated on the basis of patent number correlate with patent number per definition which can be problematic in what follows because patent number is used as an independent variable during the estimation of the relationship between productivity and knowledge transfer. For this reason, as a control factor, a second correction factor is also applied. In this second case it is employment rather than patent number which serves as the basis for the correction. The regional database of the Eurostat contains the employment of the high-tech sector so this information is used for the second correction factor which is simply the ration of high-tech employment and total employment in the regions.

Employment data, required for the computation of productivity, is available from the public Eurostat database thus labor productivity can be calculated for the 96 NUTS2 regions of the three countries as the ratio of adjusted real GDP values and high-tech employment data.

The determination of regional knowledge stock

In addition to labor productivity the knowledge stock of the regions must also be estimated. The statistical measurement of knowledge faces a very important problem, namely that it can not be measured directly. Already *Krugman* [1991] emphasized this fact focusing specially on the measurement of knowledge transfer, stating that the process itself does not leave a paper trail. However, *Jaffe et al.* [1993] confuted this thesis referring to patent citations as a suitable proxy for measuring knowledge transfer. There are several possible indicators which can serve as a proxy for the knowledge stock of a country, a region or a smaller economic unit. The most important and widely used ones out of these are based on patent statistics. The patent-based approach assumes that patents are reliable paper trails of new knowledge thus for example a simple counting of them allows for a first approximation of this knowledge stock. Using patents naturally raises some problems. First, it is hardly certain that the results of innovation (and the new knowledge embedded in them) are patented. Moreover, it can be

[3] Although Eurostat provides sectoral GDP values but for the widest NACE categories which is not suitable for the filtering of the high-tech fields.

shown that the propensity to patenting shows sectoral differences.[4] Another problem with simple counting is that it does not differentiate between patents according to their significance. This problem can be resolved using patent citations but then problems of citations arise (who adds the citation to the patent: does it reflect any real knowledge transfer?). A further drawback is that patents typically reflect technological knowledge thus patent-based indicators can not cover the whole spectrum of the knowledge base. This latter problem can be handled by using information from academic publication which represents a different (the academic) segment of the knowledge base. Another tool for measuring knowledge can be the estimation of human capital that is, knowledge embedded in people.

In the present case only patent statistics are used as indicators of knowledge stocks. The main reason for this is that the Eurostat provides information only on this variable at the relevant level of detail. On the other hand, patent cooperation is used in the network database thus using patents as proxies for knowledge stock seems to be a reasonable choice.

An important question is the very method of computation of knowledge stocks from the available information. The Eurostat database contains information both on the total number of patents and the number of patents related to the high-tech sector applied for at the EPO. This information reflects the growth in knowledge stock so these flow variables must be transformed into relevant stock variables. Knowledge stocks are derived two different ways from the flow variables. The first method defines knowledge stock as a simple cumulative sum of patent counts – the Eurostat database contains this information since 1978. The second method allows for the depreciation of knowledge. Of course the details of this amortization (its nature, extent and time span) could be the topic of another study, therefore an ad-hoc method is used here: knowledge stock in a given year is defined as the patent count of that year plus the patent counts of the four preceding years. Thus depreciation is implemented with a simple rule-of-thumb: the life-span of each patent is taken to be five years. Although this method has an ad-hoc character, the results show that there are no significant differences between the depreciated and the non-depreciated knowledge stock.

Network connections and external knowledge

Finally the external knowledge, available through network connections, must be estimated. For the computation of this element all relevant information is available: the estimation of regional knowledge stocks is described as well as the database reflecting network connections. Let us denote the vector of regional knowledge stocks in year t by $\mathbf{k}_t$ and the adjacency matrix of network connections by $\mathbf{R}_t$. Disregarding intra-regional knowledge flows that is, setting the diagonal elements of $\mathbf{R}_t$ to zeros, the $\mathbf{s}_t$ vector of external knowledge available for regions can be computed simply by the $\mathbf{s}_t = \mathbf{R}_t\mathbf{k}_t$ product.

Regression results

In the previous section those methodological issues were discussed which are necessary for the evaluation of the effect of internal and external knowledge stocks on regional productivity. Data on regional productivity, internal and external knowledge stocks is at hand according to the principles discussed so far, but there are several methods for the specific computation of data series. First, it was shown that regional knowledge stocks can be

[4] This propensity is typically higher in high-tech sectors that is, in those areas which are the basis of this analysis. Thus these differences do not lead to unavoidable bias.

calculated on the basis of simple cumulation and the depreciation method. Second, when computing external knowledge the original $\mathbf{R}_t$ matrix can be used on the one hand which weights interregional connections but the matrix can be rewritten in a binary form when the emphasis is on the existence of a connection and not on its weight.[5] Thus there are altogether four different ways of constructing the data series: all of them are used for the analysis right below.

It is also important to note that the time-span of the analysis is three years which stems from the fact that productivity data are available only from 2000 and the network and patent data can be used up to 2002 for the reasons already mentioned. According to all these, the following panel regression model is built and analyzed:

$$(1)\ \ln(PROD_{it}) = \beta_0 + \beta_1 \times \ln(PAT\_STOCK_{it}) + \beta_2 \times \ln(NETWORK_{it}) + \upsilon_i + \varepsilon_{it}$$

where variable $PROD_{it}$ shows the labor productivity of region i in year t, variable PAT_STOCK_{it} is the knowledge stock of region i in year t whereas $NETWORK_{it}$ stands for externally available knowledge. Variable υ_i reflects region-specific but time-invariant effects and ε_{it} is the white noise. As it was emphasized before, the PAT_STOCK_{it} and $NETWORK_{it}$ variables in the regression can be computed according to different methods and in different combinations. The results of the panel regression estimated with random effects are summarized in Table 1. The table contains only the most important results for sake better presentation: results for the regression coefficients and their significance level (asterisks indicate significance levels as usual). Detailed results can be found in the appendix.

Knowledge stock *Adjacency matrix*	Depreciation Weighted		Simple cumulation Weighted	
Constant	5,344	***	5,597	***
PAT_STOCK	0,180	***	0,163	***
NETWORK	-0,178	***	-0,205	***
Knowledge stock *Adjacency matrix*	Depreciation Binary		Simple cumulation Binary	
Constant	3,514	***	3,868	***
PAT_STOCK	0,160	***	0,127	***
NETWORK	0,094		0,069	

Table 1 – Regression results for the relationship between regional productivity and internal and external knowledge, dependent variable: regional labor productivity

There are several simple observations from the table. First, there is a strongly significant positive correlation between the internal knowledge stocks and the productivity of the NUTS2 regions of the three countries under consideration. This positive relationship can be found irrespective of the method of computing patent stocks and the weighting of the adjacency matrix – also the values of the coefficients are close to each other. Second, the coefficients for variable $NETWORK_{it}$ reflecting the role of external knowledge show a much more complex picture. The coefficient is significant and negative if the adjacency matrix is weighted and

[5] Formally this means that the elements of the adjacency matrix is set to one if the given element of the original $\mathbf{R}_t$ matrix is higher than zero and to zero otherwise.

non-significant positive if the matrix is binary. The difference between the two cases is that in the binary case the knowledge stocks of partner regions are meant to be completely available while in the weighted case the value of external knowledge also depends on the intensity of the connections.

The negative effect obtained for the role of external knowledge seems contradictory: it would be expected that external knowledge available through network connections have a positive effect on regional productivity as well as internal knowledge stock. Although the direction of this effect is positive for the binary adjacency matrix, for the weighted (lit. the more relevant) case the direction is negative and the effect is highly significant. On the other hand this result does not refer to a casual relationship but only to a co-movement. The negative relationship reveals only that those regions which are characterized by higher productivity typically connect to regions with lower knowledge level and/or possess fewer connections. This hypothesis also casts some light on the question why the relationship is more significant for the weighted adjacency matrix. In this case also the weights differentiate in the sense that external knowledge reflects the differences in connections weights while for the binary matrix only the knowledge levels of the partner regions are relevant.

There can be two specific ways to be sketched for the investigation of this hypothesis. According to the first one the data series used in the analysis must be adjusted so that these effects are accounted for. However, this method runs into the problem that data used for the adjustment are derived from the same information set as the variables which are to be adjusted and/or strongly correlate. Therefore, a suitable adjustment derives favorable results but the observed positive correlations stem from the definitional cointegration of the adjusted variables and the variables to be adjusted. However, a simple analysis is presented below showing that the hypothesis posed earlier really has an empirical relevance, namely the relationship between productivity and the role of external connections is examined. There are two suggested measures for the evaluation of the role of external connections for which the original adjacency matrix is used. The network database introduced previously contains information not only on the external connections of regions but on the intra-regional connections as well. Practically this means that the diagonal elements of the raw adjacency matrix $\mathbf{R}_t$ are not exclusively zeros but positive numbers with some of them being zeros. This additional information has not been used so far and the diagonal elements of $\mathbf{R}_t$ have been converted to zero. However, this additional information is used here in order to obtain two measures accounting for the role of external links in a region. The first measure simply shows the share of external connections in total connections:

$$(2)\ INTER\_SHARE_{it} = \frac{\sum_{j=1}^{N} r_{ijt} - r_{iit}}{\sum_{j=1}^{N} r_{ijt}} = 1 - \frac{r_{iit}}{\sum_{j=1}^{N} r_{ijt}}$$

The second measure relates the number of external connections to the number of internal connections:

$$(3)\ REL\_LINK_{it} = \frac{\sum_{j=1}^{N} r_{ijt} - r_{iit}}{r_{iit}}$$

With the two indicators (2) and (3) there is an opportunity to analyze the relationship between the role of external connections and productivity. The method of the analysis is again a panel-regression on the basis of the following simple equation:

(4) $ln(PROD_{it}) = \beta_0 + \beta_1 \times ln(INTER_{it}) + \upsilon_i + \varepsilon_{it}$

where the variables $INTER\_SHARE_{it}$ and REL_LINK_{it} (calculated on the basis of equations (2) and (3) respectively) are to be substituted for $INTER_{it}$. Table 2 contains the main results,[6] the details can be found in the Appendix.

Depreciation method and weighted matrix					
Constant	5,61587	***	Constant	5,3868	***
REL_LINK	-0,0995	**	INTER_SHARE	-0,327	**

Table 2 – Regression results for the role of external links, dependent variable: regional labor productivity

As it can be seen from the table, there is a negative and significant relationship between productivity and external connections that is, regions with higher productivity tend to rely less on network (more precisely external) connections (note that the indicators used here are relative indicators). On the basis of these it can be concluded that regions with lower productivity typically rely more on external connections. It is important to note that the negative relationship revealed here does not reflect the available knowledge through networks but the number (intensity) of connections thus the negative results here are not the simple rediscovery of the negative relationship revealed in the previous section with alternative indicators.

However, as noted above, using such corrections for the original variables and re-estimating regression (1) can be misleading as the adjustment variables such as (2) and (3) contain the same information as those to be adjusted (for example the indicators (2) and (3) rely on patent counts as well as knowledge stocks which are to be corrected). For these reasons the second method of testing the hypotheses for negative coefficients is discussed in more details in what follows.

The hypothesis was that regions with higher productivity typically connect to regions with lower productivity and/or possess relatively fewer links. However, this hypothesis is based on the specific structure of the network: it is supposed to be scale-free meaning that some central actors possess a large number of connections while the majority has few links (*Barabási* [2002]). As a result of this, the simple adjustment of the data series can not be successful because this specific structure of the network is not accounted for as long as data are interpreted at the level of the nodes of the network. The second method for testing the hypothesis is based on this recognition and a simple network model is used in order to capture the role of global network characteristics.

A simple model of network formation

[6] Only the results for weighted adjacency matrix and amortization method are presented here.

Adjacency matrix $\mathbf{R}_t$ used so far has been only one specific realization of possible adjacency matrices filled with empirical values. One realization represents one specific structure of the network thus differences between network structures can be dealt with only if different possible realizations of $\mathbf{R}_t$ are compared. However, it is obvious that the number of possible realizations is too large for individually handle these specific structure even for quite small networks. For this reason the literature on networks suggests different network-generating models with the help of which this large number of different structures can be classified according to some principles. For example the model of *Watts and Strogatz* [1998] is widely used for the identification of 'small-world' network structures. In what follows, a network model is used which is a modified version of the model of *Barabási and Albert* [1999] but has some features similar to the Watts-Strogatz model.

The main principles of the model of *Barabási and Albert* [1999] are network growth and preferential attachment. The former means that new nodes are added to the network continuously while the latter refers to the way how these new nodes form their links in the network, precisely that links are formed with nodes of higher degree with higher probabilities. These two principles lead to the emergence of scale-free characteristics in the network. Preferential attachment provides more connections to nodes which already have more links while the very fact of growth also generates scale-free structures: the more links will be possessed by the 'older' nodes (*Barabási* [2003], *Sebestyén and Parag* [2010]).

In the present paper a simple modification of the original model of Barabási and Albert is introduced which allows for the determination of different levels of scale-free characteristics with the change of a single parameter. Let us have an initial network with M nodes and with average degree k. Then the size of the network is increased to N in $N-M$ steps in a way that a new node is added to the network in each step and the new node establishes k number of links in that step. In contrast to the original version of the preferential attachment model, the formation of new links can happen according to different scenarios. The specific scenario can be selected with the determination of a parameter, denoted by r. The link formation algorithm is the following. The new link of the new node forms with the node with the highest degree with probability r, while with probability $1-r$ the new link forms randomly with one potential partner. Thus such a model is obtained which results in a network with average degree k and is more or less characterized by scale-free structures according to the value of r. It is clear that r must take values between zero and one allowing for the two extremes.

Stochastic features are implemented into the model through parameter r. If $r=1$ then an extremely centralized network is obtained where randomness is limited to the initial core network and the nodes of this core network possessing lots of links while newly added nodes have only k connections. If $r=0$ then links are formed randomly thus randomness play a role outside the core network as well. The steps of the algorithm are summarized below.

- First a random network is generated with M nodes and k average degree. Therefore the probability parameter of the internal algorithm generating the random graph is $k/(M-1)$ and the $k \leq M-1$ condition must hold.
- A new node is defined in the network which forms k connections to already existing nodes. The following rule determines this link formation:
    - With probability r the new node connects to the potential partner with the highest degree. Potential partners are those nodes with which no connection

exists and loops are excluded. If more than one partner can be chosen according to this rule, a random choice decides among them.
  - With probability $1-r$ the new link forms randomly with one of the potential partners.
- The previous step is repeated until the number of nodes in the network reaches N.

Two important notes must be made. First, the network can be the most extremely centralized (the star network) under very specific conditions as the randomness of the initial network allows for the star topology only for $M=2$ and $k=1$. For any other cases with $r=1$ a peripheral set of nodes with low link numbers emerges around a densely connected core. Second, the resulting network is not a pure random network even for $r=0$, because in spite of the maximum weight on randomness the fact of network growth inherent in the model leads to a scale-free structure where older nodes typically have more connections than younger ones. There could be several methods to correct for this bias but the network model in the present form is useful here as it really shows increasing degree of scale-freeness as r increases.

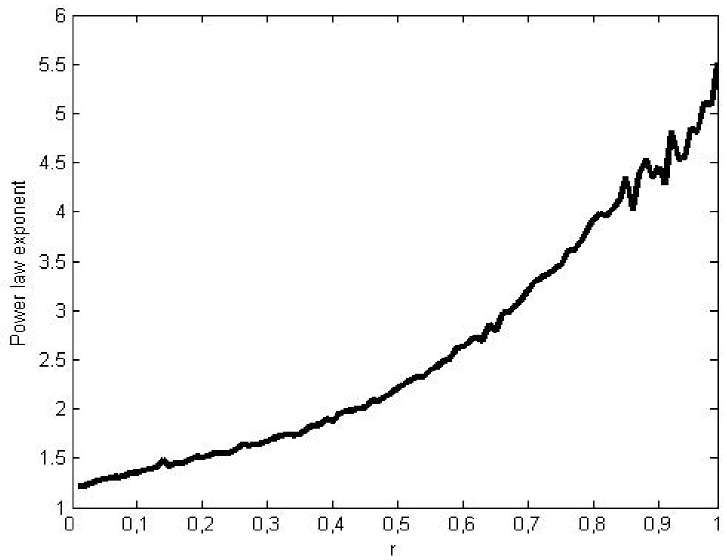

Figure 1 – Scale-freeness as measured with the power law exponent in the modified Barabási-Albert model

Figure 1 nicely illustrates this tendency where parameter r changes between 0 and 1 on the horizontal axis while the exponent of the power law distribution fitted to the degree distribution is depicted on the vertical axis. It can be easily seen that this exponent increases with r increasing so the network structure really becomes more scale-free for higher values of r.

Network structure and knowledge transfer – a simulation approach

With the help of the network model presented above the former regression analysis on the relationship between internal and external knowledge and productivity can be rewritten for simulated networks with alternative structural characteristics. This is carried out in what follows according to the process discussed below. First, a simulated network is generated with the modified preferential attachment model, then several knowledge levels are assigned to the nodes of the networks (regions in the empirical counterpart). The productivity of the nodes is calculated according to the following rule:

(5) $p_i = k_i + \sum_{j=1}^{N} a_{ij}(\theta k_j)$

where k_i is the (simulated) knowledge level of node (region) i, a_{ij} is the general element of the adjacency matrix of the simulated network,[7] θ is a spillover parameter which corresponds to the estimated coefficients for the $NETWORK_{it}$ variable in regression (1), p_i stands for productivity in region i and N is the number of nodes. Equation (5) means that there is a positive relationship *assumed* between external knowledge and productivity. With the help of simulations the question is tried to be answered whether such a situation could exist when the *observed* relationship between external knowledge and productivity are negative.

First of all, the parameters of the network are determined. In order that the simulations results to be comparable with the empirical results the size of the network is set to 96, and average degree to 15 (the latter value is obtained as the temporal average of the average degree of the (binary) empirical adjacency matrices. After setting these basic parameters for the network model, it is used to generate a specific network structure. Thus the $\mathbf{A}$ adjacency matrix is at hand the elements of which are to be substituted into equation (5).

In addition to the value of parameter θ the k_i knowledge levels of the nodes must also be determined at the outset in order to compute the p_i values. These values are generated according to the following method, on an empirical basis. As the logarithm of the empirical knowledge stock of regions ($\ln(PAT_{it})$) follows a normal distribution a vector with 96 element can simply generated in which the elements follow a normal distribution with the same parameters. Then a simple exponential transformation gives the k_i values necessary for the equation (5), thus the p_i productivity values can be computed.

Then the relationship between these p_i values and the $\mathbf{Ak}$ product must be analyzed. The analysis means a simple regression analysis similar to regression (1) but the time-dimension is disregarded for this case. The following regression is estimated on the simulated data for the network generated with the modified preferential attachment model, knowledge stocks given as above and productivity values calculated according to (5):

(6) $p_i = \beta_0 + \beta_1 k_i + \beta_2 s_i + \varepsilon_i$

where $\mathbf{s} = \mathbf{Ak}$. The estimated β_2 coefficients are relevant from this regression. For each value of r 100 independent simulations runs were executed (with the given fixed network parameters) and the estimated β_2 coefficients recorded, then the results were averaged over these 100 runs. These averages are presented in Figure 2.

[7] Binary adjacency matrix is used during the simulations. Allowing for a weighted matrix ($0 \le a_{ij} \le 1$) does not qualitatively alter the results but makes the observed tendencies more pronounced.

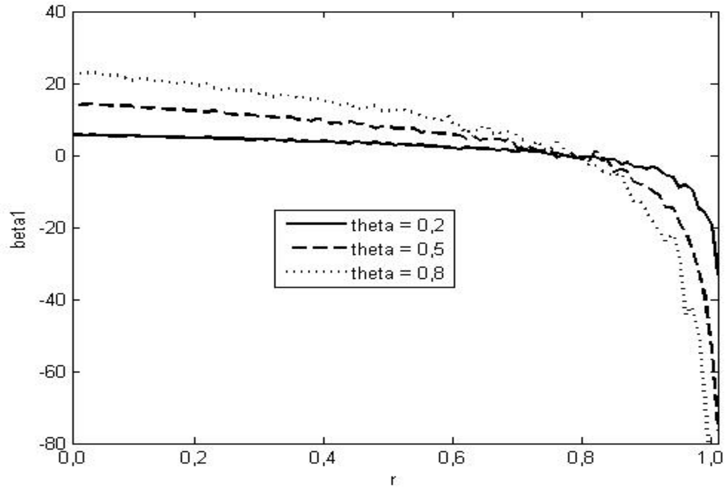

Figure 1 – The effect of scale-free structures on the β_2 regression coefficients

There are three different cases shown in the figure, each corresponding to a different value of θ. The results show that the positive relationship between external knowledge and productivity, observed for lower values of r, gets weaker as r increases then at a certain point the relationship turns into negative. For high levels of r that is, for strong scale-free structures there is a strong negative relationship between the two variables. These results strongly support our hypothesis that the knowledge networks of European regions (based on patent co-inventorship) show considerable scale-free characteristics and this is why negative relationship is observed between external knowledge and productivity. In other words the empirical relationship is found to be negative not because external knowledge affects productivity negatively but because the specific network structure hides the fundamental positive relationship between the two variables.

Scale-free structures in patent cooperation networks

In order to close the argumentation presented so far the very fact that the empirical regional network of patent inventors really shows scale-free characteristics. As already mentioned, the degree of scale-freeness can be approximated by the exponent of the power law function fitted on the degree distribution of a given network. However, in the presented model of modified preferential attachment the degree of scale-freeness can be determined by the value of parameter r. This parameter can not directly measured for an empirical network, but it can be derived indirectly.

The power-law exponent of the given empirical network can be computed as described before. Then a network is simulated with all parameters corresponding to that of the empirical network (size, average degree) except for its degree of scale-freeness which is set to be $r = 1$ and the power-law exponent for the simulated network is also computed. The same is done for the other extreme that is, for $r = 0$. Finally, the ratio of the power law exponents for the empirical and the simulated networks provides a simple approximation for the r parameter of the model. However, the ratio is normalized to the interval between the $r = 0$ and the $r = 1$ cases. If the empirical power law exponent is denoted by δ^e, the exponent for the $r = 0$ and the $r = 1$ cases are denoted by δ^0 and δ^1 respectively, then the normalized network parameter for scale-freeness is written as

$$(7)\ \bar{r} = \frac{\delta^e - \delta^0}{\delta^1 - \delta^0}$$

It is easy to see that the value of $\bar{r}$ is one if $\delta^e = \delta^1$ and zero if $\delta^e = \delta^0$. In other words this ratio corresponds to that value of r at which the modified preferential attachment model reproduces the empirically observed network structure. The empirical and simulated values of the power-law exponents are presented in Table 3.

	2000	2001	2002	Average
Empirical values	1.949	1.946	1.949	1.948
Simulated values r = 0	1.964	1.968	1.964	1.965
Simulated values r = 1	1.753	1.767	1.721	1.747
$\bar{r}$	0.929	0.888	0.941	0.920

Table 3 – Empirical and simulated values of the power-law exponents of degree distribution

The results clearly show that the examined regional network of patent inventors is really scale-free. First, the power-law exponent itself is around 2 which indicates a considerable degree of scale-freeness. Second, the calculated $\bar{r}$ coefficient is well over 0.9 which again shows that the structure is close to the extreme scale-freeness in the context of the modified preferential attachment model. The results from the table clearly confirm our hypothesis that the empirical network structure displays scale-free characteristics, moreover, to the extent for which negative coefficients for the regression between external knowledge and productivity are likely to arise. That is, the observed negative relationship between externally available network-mediated knowledge stocks and productivity can be a result of the strongly scale-free network structure and not a fundamentally negative relationship.

Conclusion

In the paper a simple question was posed which led to interesting fields: does the knowledge available through (patenting) network connections contribute to the productivity of regions? A regression analysis is conducted where regional productivity is the dependent variable whereas regional knowledge stock as approximated by the patent stock and knowledge transferred through the networks (also approximated by patent stocks of the partner regions) are among the independent variables. The analysis is conducted for several variants of these variables in the sense that both the weight of network relationships and knowledge stocks are calculated according to different principles.

The results for these regressions seem contradictory at the first sight. Regional knowledge has an (obvious) positive significant effect on productivity but the effect of knowledge obtained through network connections is negative and significant for several variable combinations used. However the hypothesis is suggested that this negative effect is not the sign of a real negative relationship between the two variables but the result of the specific network structure.

The structure of the analyzed co-inventorship network is shown to be scale-free that is, there are dominant central regions with many links but the majority of the regions have much less connections. However, central regions in the network tend to be those with high productivity and knowledge stocks. Thus the negative effect observed is a result of the tendency that peripheral regions with lower productivity tend to connect to central regions with high productivities and knowledge stocks.

The hypothesis is tried to be proved two different ways: first by the plausible corrections of the empirical variables and second by simulations focusing on the consequences of different network structures. However, the first attempt is limited by some problems in the data structure: adequate correction with the available data leads to correlations between the original and the corrected variables. On the other hand, simulation experiments show that network structure indeed has a significant effect not just on the strength but also on the direction of the relationship between knowledge coming from partners and productivity.

This way two important and interrelated conclusions can be drawn from the analysis presented in the paper. First, tacit knowledge transferred through interregional patent inventor networks have a positive effect on regional productivity but second, this effect is detectable only if one accounts for the specific structure of the underlying network, which is scale-free. An implication of these results is that when the statistical analysis of any kind of flows comes to the picture, the specific structure of the network serving as a medium for these flows plays an important role which can not be disregarded.

References

Abrahamson, E., Rosenkopf, L. (1997): Social Network Effects on the Extent of Innovation Diffusion: A Computer Simulation. *Organization Science*, 8(3) pp. 289-309.

Almeida P, Kogut B (1999): Localization of knowledge and the mobility of engineers. *Management Science*, 45 pp. 905–917.

Anselin, L., Varga, A., Acs, Z.J. (1997): Local Geographic Spillovers between University Research and High Technology Innovations. *Journal of Urban Economics*, 42(3) pp. 422-448.

Bala, V., Goyal, S. (2000): A Noncooperative Model of Network Formation. *Econometrica*, vol. 68(5), pp. 1181-1230.

Balconi M., Breschi S., Lissoni, F. (2004): Networks of inventors and the role of academia: An exploration of Italian Patent data. *Research Policy*, 33 pp. 127–145.

Balconi M., Breschi S., Lissoni, F. (2004): Networks of inventors and the role of academia: An exploration of Italian Patent data. *Research Policy*, 33 pp. 127–145.

Barabási, A.L. (2003): *Linked: How Everything is Connected to Everything Else what It Means for Business, Science and Everyday Life*. Penguing Group, USA.

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

Barabási, A.L., Albert, R., Jeong, H. (2000): Scale-free characteristics of random networks: The topology of the world wide web. *Physica A*, 281, pp. 69-77.

Breschi, S., Lissoni, F. (2003): Mobility and social networks: localised knowledge spillovers revisited. CESPRI, Working Paper nr. 142.

Carayol, N., Roux, P. (2009): Knowledge flows and the geography of networks: A strategic model of small world formation. *Journal of Economic Behavior & Organization*, 71(2), pp. 414-427.

Chakrabarti A.K., Dror, I., Eakabuse, N. (1993): Interorganizational Transfer of Knowledge an Analysis of Patent Citations of a Defense Firm. *IEEE T. Eng. Manage*. 40 pp. 91–94.

Chen C.M., Hicks, D. (2004): Tracing knowledge diffusion. *Scientometrics*, 59 pp. 199–211.

Cowan, R., Jonard, N. (2004): Network structure and the diffusion of knowledge. *Journal of Economic Dynamics and Control*, 28(8) pp. 1557-1575.

Cowan, R., Jonard, N., Zimmermann, J.B. (2006): Evolving networks of inventors. *Journal of Evolutionary Economics*, 16(1) pp. 155-174.

Csermely, P. 2005. *A rejtett hálózatok ereje*. Budapest: Vince.

Ejermo, O., Karlsson, C., (2004): Spatial Inventor Networks As Studied by Patent Coinventorship. Working Paper Series in Economics and Institutions of Innovation 17, Royal Institute of Technology, CESIS - Centre of Excellence for Science and Innovation Studies.

Ellis, P., Hepburn, G., Oppenheim, C. (1993): Studies on Patent Citation Networks. *Journal of Documentation*, 34(1) pp.12-20.

Feldman, M.P. (1994): *The Geography of Innovation*. Boston, Kluwer Academic Publisher.

Gress, B., (2010): Properties of the USPTO patent citation network: 1963-2002. *World Patent Information*, 32(1) pp. 3-21.

Griliches Z. (1990): Patent statistics as economic indicators: a survey. *Journal of Economic Literature*, 20(4) pp. 1661.

Griliches, Z. (1979). Issue in Assessing the Contribution of R&D to Productivity Growth. *Bell Journal of Economics*, 40(3-4) pp. 251-258.

Griliches, Z. (1992): The Search for R&D Spillovers. *Scandinavian Journal of Economics*, 94(0) pp. S29-47.

Jackson, M.O., Wolinsky, A. (1996): A Strategic Model of Social and Economic Networks. *Journal of Economic Theory*, 71(1) pp. 44-74.

Jaffe, A.B., Trajtenberg, M. (2002): *Patents, Citations and Innovations: A Window on the Knowledge Economy*. MIT Press, Cambridge, MA.

Jaffe, A.B., Trajtenberg, M., Henderson, R. (1993): Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations. *The Quarterly Journal of Economics*, 108(3) pp. 577–598.

Karki M.M. (1997): Patent citation analysis: a policy analysis tool. *World Patent Inf.* 19 pp. 269-272.

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

Li, X., Chen, H.. Huang, Z., Roco, M. (2007): Patent citation network in nanotechnology (1976-2004). *Journal of Nanoparticle Research*. 9(3) pp. 337-352.

Maggioni, M.A, Uberti, T.E., (2010): Networks and geography in the economics of knowledge flows. Cognitive Science and Communication Research Centre Working Paper, CSCC WP 01/10.

Maggioni, M.A., Uberti, T.E. (2006): International networks of knowledge flows: an econometric analysis. *Papers on Economics and Evolution*, 2005-19, Max Planck Institute of Economics, Evolutionary Economics Group.

Maggioni, M.A., Uberti, T.E., Usai, S. (2010): Treating patent as relational data: Knowledge transfers and spillovers across Italian provinces. *Industry and Innovation* forthcoming. A previous version is available: DISEIS Working Paper 0802, Vita e Pensiero, Milano (2008)

Oppenheim C. (2000): Do Patent Citations Count? In: Cromin, B., Atkins, H.B (eds.): *The Web of Knowledge*. Information Today, Inc, Medford, pp. 405–432.

Sebestyén, T. (2010): Innovation and diversity in a dynamic knowledge network. KRTI Working papers, 2010/1.

Sebestyén, T., Parag, A. (2010): The Dynamics of Link Formation in Patent Innovation Networks. *Perspectives of Innovation, Economics and Business*, 4(1) pp. 21-25.

Singh J. (2003): Social Networks as Drivers of Knowledge Diffusion. Elérhető: SSRN (Social Science Research Network) DOI: 10.2139/ssrn.431872.

Verspagen, B. (2005): Mapping Technological Trajectories as Patent Citation Networks. A Study on the History of Fuel Cell Research. MERIT Research Memoranda 019.

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

Zucker, L., Darby, M., Armstrong, J. (1994): Intellectual capital and the firm: The technology of geographically localized knowledge spillovers. NBER Working Paper Series, Working Paper no. 4946, NBER, Cambridge

Appendix

	Coefficient	*Std. Error*	*t-ratio*	
Const	5.344	0.418	12.784	***
PATSTOCK	0.180	0.040	4.511	***
NETWORK	-0.178	0.059	-3.007	***

Mean dependent var	5.108	S.D. dependent var	0.637
Sum squared resid	95.133	S.E. of regression	0.588
Log-likelihood	-245.025	Akaike criterion	496.049
Schwarz criterion	506.921	Hannan-Quinn	500.411

Table A1 – Results of panel regression (1) with weighted adjacency matrix and depreciated knowledge stock

	Coefficient	*Std. Error*	*t-ratio*	
Const	5.597	0.381	14.677	***
PATSTOCK	0.163	0.040	4.092	***
NETWORK	-0.205	0.055	-3.723	***

Mean dependent var	5.108	S.D. dependent var	0.637
Sum squared resid	97.792	S.E. of regression	0.596
Log-likelihood	-248.843	Akaike criterion	503.687
Schwarz criterion	514.559	Hannan-Quinn	508.049

Table A2 – Results of panel regression (1) with weighted adjacency matrix and non-depreciated knowledge stock

	Coefficient	*Std. Error*	*t-ratio*	
Const	3.514	0.492	7.137	***
PATSTOCK	0.160	0.041	3.903	***
NETWORK	0.0945	0.058	1.618	

Mean dependent var	5.108	S.D. dependent var	0.637
Sum squared resid	93.927	S.E. of regression	0.584
Log-likelihood	-243.258	Akaike criterion	492.517
Schwarz criterion	503.389	Hannan-Quinn	496.879

Table A3 – Results of panel regression (1) with binary adjacency matrix and depreciated knowledge stock

	Coefficient	*Std. Error*	*t-ratio*	
Const	3.868	0.495	7.814	***
PATSTOCK	0.127	0.041	3.052	***
NETWORK	0.0686	0.059	1.170	

Mean dependent var	5.108	S.D. dependent var	0.637
Sum squared resid	97.400	S.E. of regression	0.595
Log-likelihood	-248.286	Akaike criterion	502.573
Schwarz criterion	513.445	Hannan-Quinn	506.935

Table A4 – Results of panel regression (1) with binary adjacency matrix and non-depreciated knowledge stock

	Coefficient	*Std. Error*	*t-ratio*	
Const	5.616	0.081	69.043	***
INTER	-0.099	0.049	-2.019	**

Mean dependent var	5.543	S.D. dependent var	0.664
Sum squared resid	107.567	S.E. of regression	0.641
Log-likelihood	-255.615	Akaike criterion	515.230
Schwarz criterion	522.374	Hannan-Quinn	518.101

Table A5 – Results of panel regression (4) with REL_LINK as independent variable

	Coefficient	*Std. Error*	*t-ratio*	
Const	5.387	0.087	61.963	***
INTER	-0.327	0.164	-1.998	**

Mean dependent var	5.514	S.D. dependent var	0.670
Sum squared resid	115.818	S.E. of regression	0.644
Log-likelihood	-273.715	Akaike criterion	551.431
Schwarz criterion	558.700	Hannan-Quinn	554.347

Table A6 – Results of panel regression (4) with INTER_SHARE as independent variable