Perspectives of Innovations, Economics & Business, Volume 4, Issue 1, 2010
www.pieb.cz

THE DYNAMICS OF LINK FORMATION IN PATENT INNOVATOR NETWORKS

TAMAS SEBESTYEN,
ANDREA PARAG

Faculty of Business and Economics
University of Pecs, Hungary

JEL Classifications: O31

Key words: Innovation, network evolution, knowledge-transfer, cluster.

Abstract: The paper presents a simple extension of the Barabasi-Albert model of network evolution. This model is based upon the assumption that new links are formed not only according to the centrality of other nodes in a network but geodetic distance is also important in link formation. Simulation results show that if link formation is based on distance then the resulting network is more clustered than in the case of centrality being the dominant factor in link formation. Our empirical results show that in European regional patent inventor networks distance is a considerably more important factor in link formation than network centrality.

ISSN: 1804-0527 (online) 1804-0519 (print) PP.21-25

Introduction

Networks have attracted considerable attention in the last decades. Works from a broad spectrum of scientific research have revealed that networks in quite diverse areas of life (e.g. the living cells, the World Wide Web, social relationship networks, etc.), although different at the first sight, share some basic common properties, among which the most striking is their invariant scale-free characteristics (Barabasi and Albert, 1999, Barabasi, 2003).

At the same time the literature on innovation has focused on learning and innovation networks, i.e. networks of firms, researchers, specialized institutions etc. It is now clear that innovation can be regarded as an interactive process which requires relationships between different agents of the process (innovators, firms, universities, venture capital, etc.). In this context one can see for example Bathelt et al, 2004; Nonaka, 1994; Lundvall and Johnson, 1994.

The interaction among networks and regional development is also an emergent field of research through the notion of regional clusters which are meant to be the drivers of innovativeness and therefore regional economic success (for a general overview see e.g. Karlsson, 2008).

These two lines of research have been synthetised in the area of innovation networks which tries to reveal the characteristics and dynamic patterns of such networks. The work in this field has proceeded along two different methodological avenues. First, empirical studies made considerable efforts to gain insight into the characteristics of real innovation networks. These studies however, mainly due to the lack of adequate data, grasp only a static view of the networks in question, their structural characteristics and the relationship between these characteristics and their performance. (Varga and Parag, 2009; Ozman, 2006.)

While inventor networks are widely thought to enhance regional innovative capability, there exist few longitudinal studies of formation and evolution over time (Fleming at al, 2007; Ter Wall, 2008). Some focus on examining knowledge transfer from academy to industry showing academic inventors to be more central and better connected than non-academic ones (Balconi at al, 2004). Others investigate the separate effects of inventor agglomeration on metropolitan patenting and the structure of social networks linking inventors within and across metropolitan areas. They find that the structural features of the inventive networks are less important agglomerative features of metropolitan areas than agglomeration. (Lobo and Strumsky; 2008). The paper of Ejermo and Karlsson (2006) explores the structure and the strength of interregional inventor networks as measured by the affinity between inventors to be co-authors in patents across regions. It finds that affinities extend more often to regions which have high patenting, when they have high R&D levels, and to those with more university R&D.

In this paper we focus on the link formation process in evolving networks. To this end we present a simple model setting based on the model of Barabasi and Albert (1999), then we use a newly built database in order to capture basic characteristics of link formation in evolving patent innovation networks. Our database covers patent statistics of European countries through the period between 1978 and 2005. This longitudinal span offers us the possibility to analyse link formation processes.

The paper proceeds as follows. First, we present an extended version of the Barabasi-Albert model of network evolution and analyse some of its main implications. Then we briefly describe our database which gives the basis of our following empirical analysis.

A simple, but extended model of network evolution

Our baseline model, to be extended afterwards, is the model of Barabasi and Albert (1999). This model starts from an initial random network and adds a new node to the network in each step and a new node forms a given number of links to the already existing nodes. Link formation is based on the so called "preferential attachment" which assures that a new node forms a link with a given node with higher probability if that given node have more links, i.e. if it occupies a more central position in the network. In the terminology of network theory the number of links of a given node is called the degree of that node, so we will refer to this kind of preferential attachment as driven by the degree distribution of the network.

Let us denote the degree of node i in the network as D_i. In the model of Barabsi and Albert (1999) the probability of establishing a link with (already existing) node i is simply

$$P_i = \frac{D_i}{\sum_i D_i}$$

Given this simple rule, we can build networks where popular nodes (i.e. those nodes which have more links) become more popular as time passes by. However, there are two important considerations to be mentioned with regards to the model above, which lead us to the extension of the Barabasi-Albert model.

First, preferential attachment is present in the Barabasi-Albert model in the formation of the first link of a new node, i.e. its initial attachment to the network as a whole. However, it is ambiguous to what extent new nodes are informed about the degree distribution of a network. In some cases it might be a reasonable assumption, but in other cases do not.

Second, given that a node has at least one link, it is still questionable if it has perfect information on the degree distribution of the whole network. If information is not perfect, it is reasonable to assume that information is more profound about that local subnetwork which surrounds the given node. This leads us to the conclusion that in addition to the centrality of nodes, their distance in the network (i.e. geodetic distance) is an important factor of link formation as well.

Based upon these considerations we extend the Barabasi-Albert model in two aspects. First, the initial attachment of a given node to the network is totally random, i.e. the first link can be established with any of the existing nodes with equal probability. All other links are then formed upon considering both the centrality (degree) and (geodetic) distance of existing nodes. We define the attractiveness of node i as

$$A_i = \alpha \frac{D_i}{N-1} + \beta \frac{N - L_i}{N-1}$$

where D_i is the degree of node i, L_i is the geodetic distance of node i from the given (newly added) node, N is the number of nodes (the size of the network), whereas α and β are parameters defining the weight of degree and distance in link formation. Given the definition of attractiveness above, the probability of establishing a link with node i is

$$P_i = \frac{1}{N}$$

for the first link and

$$P_i = \frac{A_i}{\sum_i A_i}$$

for the second and further links of a node.

In order to evaluate the difference between the standard preferential attachment model of Barabasi and Albert (1999) and the extended version given above, we run some benchmark simulations and evaluated two different output measures referring to different characteristics of the emerging network, namely clustering and entropy.

Clustering coefficient

With a verbal definition we can think of the clustering coefficient as a measure of how much one's friends are friends of each other (Cowan, 2005). In other words, the clustering coefficient is able to capture local structures in a network: high clustering means dense local structures in a network. The clustering coefficient is calculated basically for one specific node of the network.

Let's denote the neighbourhood of node i by Γ_i. In this case the cardinality of this set, denoted by $\|\Gamma_i\|$ measures the number of neighbours node i has. In this neighbourhood $\|\Gamma_i\|(\|\Gamma_i\|-1)/2$ links can be formed at most. If the number of links in this neighbourhood is the highest possible, then the clustering coefficient is one. If there is no links among i's neighbours, the clustering coefficient is zero. Let $I(j,l)=1$ if node j in the neighbourhood of $l \in \Gamma_i$ is a neighbour of i itself, and $I(j,l)=0$ if this is not true. This way the clustering coefficient of node i can be written as

$$C_i = \sum_{j,l \in \Gamma_i} \frac{I(j,l)}{\|\Gamma_i\|(\|\Gamma_i\|-1)/2}.$$

In order to gain an aggregate measure on the level of the network, we compute the average clustering coefficient, which averages C_i over nodes:

$$\overline{C} = \frac{\sum_i C_i}{N}.$$

Entropy

Using statistical entropy we can measure to what extent a given network is characterised by scale-free properties, i.e. to what extent its degree distribution is exponential meaning that few nodes have lot of links while lot of nodes have few links. In this paper we use the specific relative entropy measure described by Wu et al. (2007). First define I_i as the relative degree frequency of node i:

$$I_i = \frac{D_i}{\sum_i D_i}.$$

Given this, the absolute entropy of the given network is

$$E = -\sum_i I_i \ln I_i .$$

However, there is a maximum and a minimum for this expression. Its value is maximal when all nodes have the same number of links. In this case $I_i = 1/N$ for all i, thus $E_{\max} = \ln N$. The value of entropy is minimal when the network is totally centralised, i.e. one node has $N-1$ links and the other nodes have only one link. In this case $E_{\min} = \ln(4(N-1))/2$. The value of normalised entropy is given by:

$$NE = \frac{E_{\max} - E}{E_{\max} - E_{\min}}.$$

The value of NE is 0 when statistical entropy is maximal, i.e when the network is not centralised (all nodes have the same number of links) and its value is 1 if the entropy is minimal, i.e. in the case of a totally centralised network.

Simulation results

As mentioned above, we run some benchmark simulations with the models above. More specifically, we run two simulations. One in which only the degree of other nodes are considered when establishing a new link (i.e. $\alpha = 1$ and $\beta = 0$ in the equation for attractiveness), and one in which only distance was considered ($\alpha = 0$ and $\beta = 1$). The results show two important insights.

First, it becomes clear that the preferential attachment based on centrality alone is not a necessary condition for the emergence of scale-free networks. A growing network in itself can be a sufficient condition leading to scale-free networks as older nodes per definition posess more links even if link formation is totally random.

Second, and more interesting is the result for the two output measures. Our simulations show that higher role for degree in link formation leads to less clustered networks while the role for distance results in more clustered networks. That is, the underlying link formation process has important effects on the structure of the emerging network. In what follows, we present an empirical analysis of link formation processes with regards to the role of degree and distance in patent inventor networks.

The database

From the data of the European Patent Office (EPO) we built up networks of patent inventors across European regions. The patent data of the EPO contain information about the address of the inventors and obviously the sector to which the given patent belongs. We took this data from 1978 to 2005 and from this information we extracted co-inventorships in the case of each patent and built up the network of patent inventors. In the next step this network was aggregated into European NUTS2 regions. That is, we do not consider network of individuals but network of regions, however behind this network lies the network of individuals. Further, the network among regions is a weighted one, meaning that an edge between two different regions has a weight referring to the number of patents on which inventors of the two regions had worked together.[1] Thus we have a network of regions in which the intensity of interregional relationships is reflected by the number of co-invented patents, nevertheless, the networks are constructed for every year in the period between 1978 and 2005.

The database covers the high-tech sector as used by the Eurostat methodology. In this classification the high-tech sector covers three subsectors as follows: (1) aviation; (2) computer and automated business equipment; (3) communication technology; (4) lasers; (5) micro-organisms and genetic engineering; (6) semi-conductors.[2]

The database is being constructed for all European countries, however only part of the countries' statistics is already available thus we restricted our analysis to the three major countries active in the patent field. According to the patent statistics of the Eurostat, these countries are Germany, France and the United Kingdom, with reference to the total number of patent applications to the EPO.

Link formation in European co-inventorship networks

In order to carry out this analysis we built up a secondary database from the network data. Every time a new link was formed in the networks, two records were taken, according to the two nodes which had a new link. We put the degree of the partner and their geodetic distance into each record. Then, in order to yield comparable results we calculated the ratio of the degree measures to the highest in the actual network and took the inverse of the geodetic distance. This way if we have a value of 1 for degree, this means that the actual node chose the most central node in the network as partner, i.e. the one with the most links. A close to zero value, in contrast, shows that a peripheral node was chosen as partner. Similarly in the case of distance, a value of one represents the closest nodes and a close-to-zero value means a far-away node as partners.[3] In our first analysis we simply cumulated the links year after year, so we disregarded those links which were abandoned (in other words the network of 2005 contains all links formed since 1978). In addition, we considered the choice of only those nodes which were not new in the networks, i.e. they had possessed at least one link in the previous year.

Table 1 contains our main results regarding the calculations described above. The table gives the weights of

[1] Note that the weight of an edge between two regions refers not to the number of personal contacts but the number of co-invented patents between the two regions.

[2] The associated IPC codes are: (1) [B64B, B64C, B64D, B64F, B64G]; (2) [B41J, G06C, G06D, G06E, G06F, G06G, G06J, G06K, G06M, G06N, G06T, G11C]; (3) [H04B, H04H, H04J, H04K, H04L, H04M, H04N, H04Q, H04R, H04S]; (4) [H01S]; (5) [C12M, C12N, C12P, C12Q]; (6) [H01L].

[3] It is important to note that in "ordinary" networks the value of 1 in the case of geodetic distance is meaningless, because these are the neighbours of a node, thus there is no reason to make a new link between them. In our case, however, links are weighted according to the number of cooperations in patent-inventorship. Thus a new link among neighbours is acceptable and interpreted as a more dense cooperation between two regions.

degree and distance in new link formation, according to the schemes given above. We can not observe substantial differences among the different subsectors, however there is a significant difference between the weights of degree and distance, showing that distance is a more important decision variable than degree. This finding supports our remark about the importance of distance. The overall picture shows that geodetic distance is the most important factor in link formation while degree is less important, although not insignificant.

TABLE 1. THE WEIGHT OF DEGREE AND DISTANCE IN LINK FORMATION ON 1 YEAR BASE, DISREGARDING DISSOLVING LINKS

	Degree	Distance
Aviation	0.38	0.70
Computer	0.33	0.79
Communication	0.32	0.83
Laser	0.28	0.69
Semiconductors	0.30	0.76
Micro-Genetics	0.31	0.81
High-tech	0.29	0.86

In order to give a benchmark to the findings above we calculated the same measures but now taking into account dissolving links as well (see Table 2). The basic difference here is that the degree of a node can decrease over time, while the distance between two nodes can increase due to dissolving links. However, the picture is qualitatively the same as before with a slightly higher variance in the data. Differences among subsectors are not significant and distance seems to be the important choice variable.

As a final question, it would be interesting to see if the importance of these decision variables changed over the years. In order to tackle this issue, we calculated these measures for every year. The basic picture is similar to that of presented in Tables 1 and 2, so we only insert here the overall values for the whole high-tech sector.

TABLE 2. THE WEIGHT OF DEGREE AND DISTANCE IN LINK FORMATION ON 4 YEARS BASE, WITH DISSOLVING LINKS

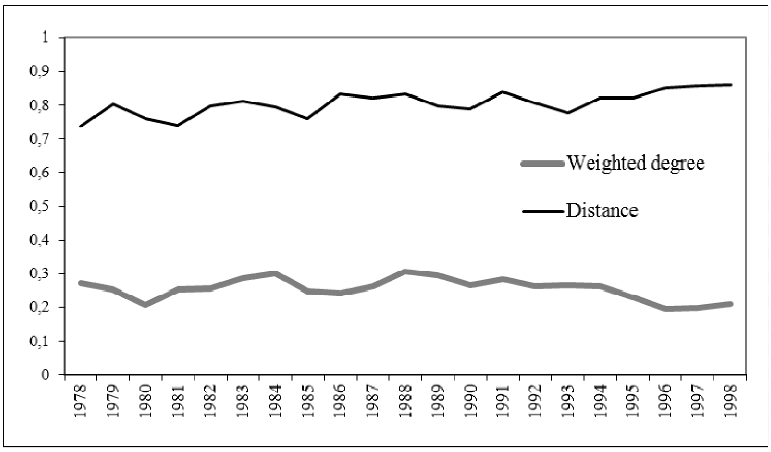
	Degree	Distance
Aviation	0.32	0.57
Computer	0.28	0.71
Communication	0.27	0.76
Laser	0.30	0.51
Semiconductors	0.26	0.68
Micro-Genetics	0.29	0.75
High-tech	0.25	0.81

As it is clear from Figure 1, distance is invariably the more important decision factor during the years without remarkable trend. To conclude, we observe that in our sample of interregional patent inventor networks geodetic distance rather than degree seems to be the most important decision factor when choosing a partner, and this pattern of link formation does not change over our examination period.

Conclusions and further avenues for research

In this paper we presented a straightforward extension of the Barabasi-Albert model of network evolution emphasizing the role of geodetic distance in link formation. Our simulation results show that there are indeed important differences in the resulting network structure depending on the link formation process. If distance is more important in link formation, the emerging networks are more clustered, while if degree is more important, the resulting networks are relatively more centralized, i.e. they are characterised by more significant scale-free properties.

FIGURE 1. THE EVOLUTION OF WEIGHTS FOR DEGREE AND DISTANCE IN LINK FORMATION



Our empirical results, on the other hand, show that in the case of European patent inventor networks distance seems to be the important decision variable in link formation as opposed to centrality. However, there are two important remarks. First, these results do not state that geodetic distance is the dominant factor in link formation in all networks: in other networks different factors may be important. Second, as our database gives regional networks of inventors, the question arises, to what extent geography are present in our model. Although geodetic and geographic distance are not the same, they are obviously interrelated. One of the possible tasks for future research is to specify to what extent link formation depends on geographic and not geodetic distance. A straightforward way would be to measure how much geodetic and geographic distances are correlated. However, the question remains that which is the cause and which is the effect.

References

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., 2003. Linked: How everything is connected to everything else what it means for business, Science and Everyday Life, Penguing Group, USA.

Barabási, A., Albert, R., 1999. “Emergence of scaling in random networks”, Science 286, pp. 509-512.

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

Bathelt, H., Malmberg, A., Maskell, P., 2002. “Clusters and knowledge: Local buzz, global pipelines and the process of knowledge creation”, DRUID Working Papers 02-12, DRUID, Copenhagen Business School, Department of Industrial Economics and Strategy/Aalborg University, Department of Business Studies.

Ejermo, O., Karlsson, C., 2006. “Interregional inventor networks as studied by patent coinventorships”, Research Policy, Elsevier, vol. 35(3), pp. 412-430.

Karlsson, C., 2008. Handbook of research on cluster theory, Edward Elgar, Chelthenham, UK.

Lobo, J., Strumsky, D., 2007. “Metropolitan patenting, inventor agglomeration and social networks: A tale of two effects”, Journal of Urban Economics 63, pp. 871-884.

Lundvall, B., Johnson, B., 1994. “The learning economy, industry & innovation, 1469-8390, Vol. 1, Issue 1, pp. 23-42.

Nonaka, I., 1994. “A Dynamic theory of organizational knowledge creation”, Organization Science, Vol. 5, No. 1, pp. 14-37.

Ozman, M., 2006. “Networks and innovation : A survey of empirical literature”, Working Papers of BETA 2006-07, Bureau d'Economie Théorique et Appliquée, ULP, Strasbourg.

Ter Wal, A., 2008. Cluster emergence and network evolution: a longitudinal analysis of the inventor network in Sophia-Antipolis, Papers in Evolutionary Economic Geography, 0810.

Varga, A., Parag, A., 2009. “Academic knowledge transfers and the structure of international research networks”, in: Attila Varga (ed.), University knowledge transfers and regional development: Geography, entrepreneurship and policy, Edward Elgar Publishers, pp. 138-159.

Wu, J., Tan Y.-J., Deng, H.-Z. and Zhu, D.-Z., 2007. Normalized entropy of rank distribution: A novel measure of heterogeneity of complex networks, Chinese Physics, 16, pp. 1576-1580.