MAPPING THE DISTRIBUTION OF BUSINESS KNOWLEDGE WITH THE HELP OF AGGREGATED STATISTICAL DATA

Zoltán Bartha, PhD.
University of Miskolc, Institute of Economic Theory
H-3515 Miskolc-Egyetemváros
Hungary
zolib@hu.inter.net

ABSTRACT

It is a fundamental assumption of the literature dealing with business knowledge that valuable knowledge is highly concentrated. Knowledge hot spots are formed in different regions, and these regions are the ones that generate most of a nation's economic growth. Many case studies have been prepared of such hot spots, however the distribution of knowledge across different regions of a country is extremely difficult to measure. Not least because knowledge itself is a phenomenon that seemingly resists all efforts of researchers to be precisely quantified.
This paper first presents a convenient typology of business knowledge, and then an effort is made to measure certain elements of this knowledge with the help of factors which are surveyed by traditional statistics. In the paper data of Hungarian industrial parks is used, where the number of companies operating within an industrial park, and the number of workers employed in an industrial parks are used as possible indicators of business knowledge. The distribution of these indicators resembles the pattern of a scale-free network.
Scale-free networks have been observed in various fields of human and natural sciences. A basic characteristic of these networks is that strong network centres are present. The presence of these network centres provides a high level of connectivity in such scale-free networks. The paper suggests that the scale-free distribution of the indicator values can be regarded as a proof to business knowledge's concentration.

JEL CLASSIFICATION: O, R

KEY WORDS
business knowledge, scale-free networks

INTRODUCTION

Business knowledge or some variant of it is more and more often called a strategic resource in the managerial literature. Getting access to this strategic resource has therefore become a central point in a company's supply strategy. Arguments based on theoretical considerations point our attention to the assumption that business knowledge is geographically centralised, and therefore is only accessible in certain regions. When faced with the above situations, companies are forced to establish subsidiaries in regions where the knowledge in centralised.
With use of data of the Hungarian industrial parks the article tries to prove that business knowledge is centralised, exactly as assumed by arguments based on theoretical considerations.

CHARACTERISTICS OF BUSINESS KNOWLEDGE

In the article business knowledge is called all action related skills or recorded messages that contribute to the effective combination of resources, and to the profitable operation o the business [2]. Business knowledge is a scarce factor, the acquisition of it is costly. It consists of tacit and explicit elements [6], however tacit knowledge is a must in the creation and understanding of explicit knowledge.
One very important element of business knowledge is the tacit knowledge of the employees, in other words the employee's competence and contact base [2]. According to assumptions of the literature, these elements of business knowledge are geographically centralised, immobile. There are several explanations to the immobility. First of all, the creation of the knowledge is centralised, partly because of the characteristics of the educational structure, partly because of the cooperation between universities and companies that enables learning by doing. Secondly, the formal and informal contact systems formed in innovation systems [3], that makes learning quicker, is also a main immobilising factor. Quicker learning is made possible by mutual assistance based on trust. Trust is formed by regular personal interactions, and the need for personal interactions for

trust to be built, makes employees carrying valuable business knowledge immobile.
Because tacit business knowledge is partly immobile companies can only get access to it if they choose the indirect way, and ensure the supply by capital investment. However companies settling in regional innovation systems not only gain by acquiring valuable business knowledge. They also become part of such a network characterised by intensive contacts that enables quick information interchanges, and so the learning process is quickened in all firms taking part in the network.

CONCENTRATION AND SCALE-FREE NETWORKS

The theory of networks addresses the problem of systems with many connections. When analysing the concentration of knowledge, a system with many connections can be used as a model, where the dots connected are represented by carriers of knowledge. If business knowledge is concentrated, we can expect in the model network the dots are concentrated near certain centres; these are the network centres.
Scale-free networks are characterised by exactly those kind of network centres. The theory was developed by Barabási [1]. The main characteristics of these networks is that the large majority of the dots posses only a small amount of the connections, while some of the dots are hugely interconnected, possessing the majority of the connections. The distribution of these networks is described by the power-function: $N(k) = k^{-\gamma}$ showing that the number of dots having exactly k connections is a function of the k connectivity number, where the γ scale index is equal to the steepness of the trend line fitted to the log-log interpretation of the data. Networks described by power-functions are called scale-free networks. If business knowledge is characterised by scale-free distribution, then we can expect that certain region are very rich in business knowledge, while most of the regions are very scarcely endowed with it.
As a model of tacit knowledge the connections between employees and the stakeholders of the company will be used in this article. These connections represent the so called 'know who' of employees [5] that can be divided into two categories:

- Official or formal connections: permanent interactions that are registered by the companies in form of contracts or other written agreements.
- Informal connections: permanent contacts formed among employees working for different companies; these connections are based on trust.

Be it a formal or an informal connection, it is obvious that the more firms operate within a region, the greater the chance of forming such relations. Therefore the concentration of the firms may refer to the formal connections. This nexus provides a nice verification opportunity. If business knowledge (formal and informal connections in this article) is indeed concentrated, the following two characteristics should be found in the geometrical concentration of firms:

- first, it should be noted that companies tend to settle in regions where there is already a large concentration of companies;
- secondly, there should be a hierarchy within these regions too, in other words it is expected that a large amount of firms operates in certain regions, while others are relatively scarcely settled.

The first one should not be surprising as economists have been analysing industrial districts and industrial agglomerations since the very early times. Many reasons have been mentioned therefore. The most important one of these is still the presence or lack of different infrastructural conditions: a firm can only settle in a region if the infrastructure needed for the operation is present, and the labour force is available too. Logistical considerations are important as well.
The factors listed above however cannot explain the hierarchy among industrial districts as regards to the number of firms settled in them. The infrastructure needed is present in many regions, and logistical conditions are favourable in many places as well. The model developed by Porter [7] may better explain the hierarchy, as the proximity of suppliers or potential buyers may explain the forming of large industrial centres, just as the availability of skilled labour. It should noted however, that all these factors are connected to the business knowledge available in a region. Potential suppliers and buyers add to the number of formal connections of a firm, while skilled labour extends the potential informal connections of the employees. If empirical data showed that there are considerable difference among regions (in terms of the number of firms settled) that posses the same infrastructural conditions, it would prove that business knowledge is indeed concentrated.

DATA OF INDUSTRIAL PARKS

To test the concentration of business knowledge the data of Hungarian industrial parks are used. Although industrial parks are a unique group of industrial districts, their data can be very useful for several reasons.
All industrial parks have well developed infrastructure. The original goal of the industrial park project was to create new jobs by providing additional sources for infrastructural development. As the basic infrastructure is available in all parks, the difference among the numbers of firms settled in the parks can only be explained by other factors, e.g. the quantity and quality of business knowledge available.
It is a further advantage that the area of industrial parks is precisely given, so the firms settled can be counted without any problems what so ever.
As of 2007, there are 165 industrial parks in Hungary. The number is substantially big to enable statistical analysis , whilst the distribution of the industrial parks is also favourable, as all regions of the country are well

represented. But the industrial parks are representative also if the number of settled firms, or the value added by them is considered. Around 25% of Hungary's total industrial output is created within the premises of the industrial parks, while the employment rate is close to 15% (compared to the total number of industrial employees). In recent years the number of SMEs settled in industrial parks has been growing the fastest, so the output is not only created by a few big multinational companies [4].
Finally, a practical reason of using industrial park data is the fact that these statistics have already been prepared by the Hungarian Ministry of Economy and Transport, therefore no survey had to be directed to gather the information. When the contract is signed by the industrial parks and the ministry, all parks agree to provide detailed information concerning the number of firms settled, the value of direct investments, the number of people employed, the value of total sales etc. As the information are provided by the industrial parks themselves, they are extremely accurate.
The only downside of using pre-gathered statistics is that the information are confidential, and so the data of the individual industrial parks could only be used anonymously, without any codes or registration marks that could make the identification possible. Without identification however, the geographical dimension is completely lost – although we might suspect that the differences may partly be explained by pure geographical conditions, the effects of these cannot be considered. All the result below should be interpreted with the notion of this deficiency.

SCALE-FREE NETWORK BY THE NUMBER OF FIRMS

In the year 2005 a total of 2615 firms operated in the 165 Hungarian industrial parks; that gives an average of 16. Interestingly enough only 4 parks have exactly 16 settled firms, and what is even more interesting, 70% of the 165 parks have less settled firms than the average number. The most frequent number of firms is 0, followed by 5, and all the others within the 0-10 range. There are big holes on the other end of the scale, the top 10 industrial parks have settled firms in the 40-106 range, and the top 4 – far exceeding all the others – possesses around 15% of all the firms.

Parameters	Regression	Degree of freedom	F-probe
β_0=2,769651	ss_{rez}=9,530913	ν_1=1	F_{limitr}=4,08
β_1=-0,61107	ss_{total}=30,44959	ν_2=42	$F_{calculatedt}$= 42,00103
	r^2=0,686994		

Table 1.
Parameters for the power-function regression of firm number

Such a distribution of the data suggests that it could be described by a power-function. Therefore the logarithm is calculated of all the data, and a linear trend line is fitted to the dots, using the method of ordinary least squares. The results of the regression analysis can be found in Table 1. The quadratic value of the correlation coefficient is 69% (with an original value of 0,83), signalling a strong correlation.
Having seen that it is indeed reasonable to use regression analysis, the distribution of the aggregation is examined. The information required for the F-probe can be found in the last column of Table 1. The F-probe is a simple hypothesis analysis used to determine whether the assumption of linear regression was a valid. The sample consists of 44 different firm numbers in the 165 industrial parks, therefore the critical value of the F-probe (with degrees of freedom ν1=1 ν2=42, at a 5% level of significance) is Flimit=4,08. The Fcalculated obtained from the analysis is considerably higher, so it can be safely said that the use of linear regression is valid at all significance levels.
Despite the F-probe, a sample was tested for exponential regression as well. According to the calculations made, the quadratic correlation coefficient is 50% this time, considerably lower than in the previous case (69%).
Finally, the validity of the chosen regression model can also be tested with the analysis of the residual deviation of the model. If the difference of real and trend values is calculated for all the individual firm numbers, we will get a diagram like the one showed by Figure 1. It is clear to be seen that the residuals are randomly scattered in the ±1 range of the real values (in fact, within the logarithm of the real values). The distribution of the residuals does not show any clear trends, which is a good sign, as a clear pattern would indicate that the model set up is mistaken.
In order to receive even more convincing parameters, the data used is somewhat transformed. As mentioned earlier, although there are 165 industrial parks, if they are structured according to the number of firms settled, we only get a sample of 44, as there are only 44 different firm numbers in those 165 parks. If Hungary had thousands of industrial parks a much bigger sample could be analysed, one that has very few holes, missing values, and so the power-function would be much clearer. Because such a sample is unavailable, an inverse approach is taken. According to the number of firms settled, parks are put into different categories. Within each category the parks are very similar to each other, i.e. the number of firms settled in them is very similar. Industrial parks with 0-9 firms fall into the category of small parks, the next category is made up of parks with 10-19 firms etc. 11 distinct, relatively homogenous groups can be set up this way, the distribution of which is shown by Figure 2.

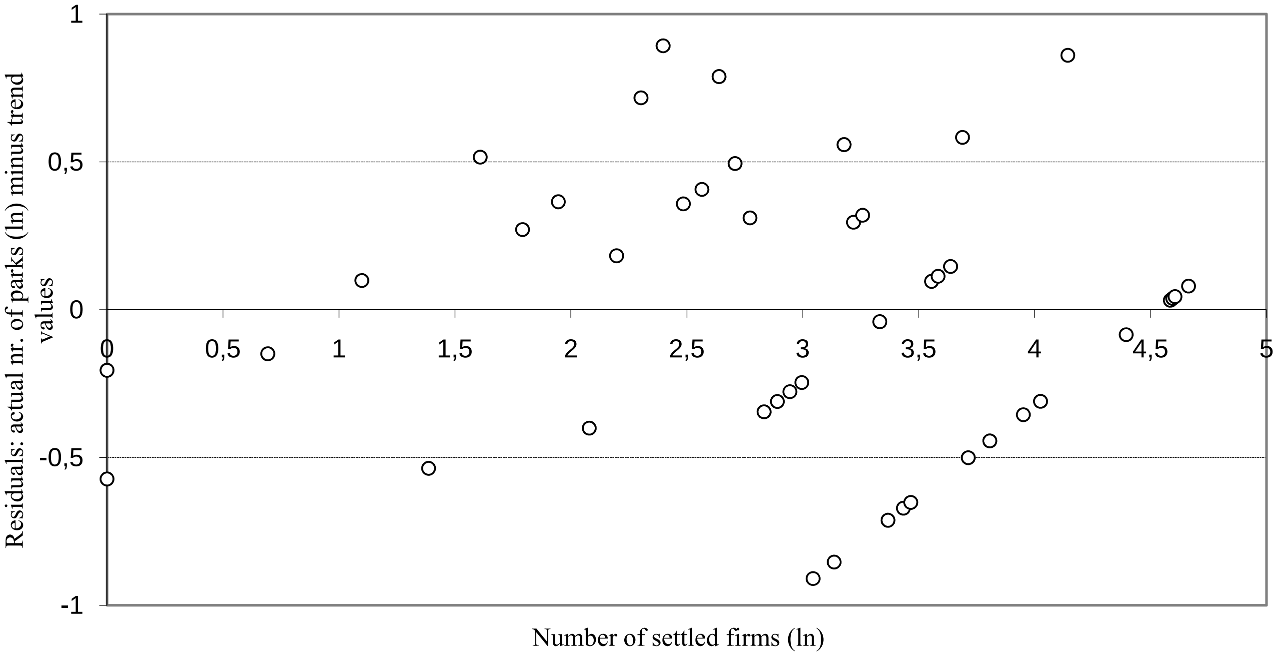

Figure 1.
Residual deviation values belonging to the number of firms

Figure 2. shows nicely all the main characteristics of the scale-free networks: a lot of dots (industrial parks) belong to the lowly connected category (number of firms settled), while the highly connected ones only have a few dots, the so called network centres. What is suggested by Figure 2., is proved by Table 2., where the most important parameters are listed.

With the adjustments the quadratic value of the correlation coefficient has increased to 87%, showing a very close correlation. In the scale-free network of industrial parks the value of the scale index previously denoted by γ is 1,6.

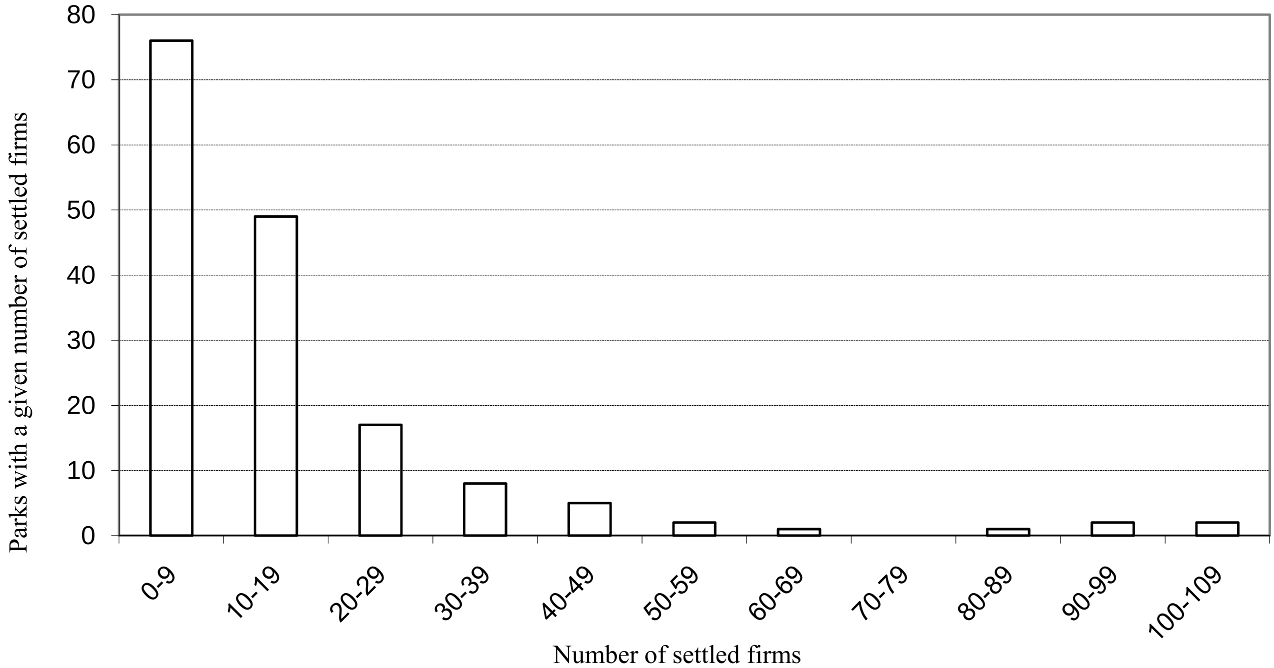

Figure 2.
Distribution according to the number of firms - categories

Parameters	Regression	Degree of freedom	F-probe
β_0=7,445124	ss_{rez}=3,194738	ν_1=1	F_{limitr}=5,12
β_1=-1,58527	ss_{total}=24,55253	ν_2=9	$F_{calculatedt}$= 60,16773
	r^2=0,869882		

Table 2.
Parameters for the power-function regression of the firm numbers – categories

Parameters	Regression	Degree of freedom	F-probe
β_0=12,2662	ss_{rez}=3,045771	ν_1=1	F_{limitr}=4,54
β_1=-1,39185	ss_{total}=30,56089	ν_2=15	$F_{calculatedt}$= 135,5081
	r^2=0,900338		

Table 3.
Power-function regression of the number of employees (scaling of 500)

NUMBER OF EMPLOYED AND SCALE-FREE NETWORKS

The distribution of industrial parks cannot only be examined by the number of settled firms, but also by the number of workers employed by settled firms. While the number of firms was important in analysing the formal connections, the number of employees is connected to the informal connections. There can be parks where there are only a small number of settled firms, those few however may employ a lot of workers. Because there is no close correlation between the number of firms and the number of employees, the scale-free distribution showed in the previous chapter will have to be tested again.

Just as in the previous chapter, homogenous park-categories are created. More versions of the transformation were tested, with scalings of 250, 500 and 1000. As the highest number of employees in an industrial park is above 8.000, with the first scaling a total of 33 groups can be created, but 13 of these are empty categories, with no industrial parks in them. With the 1000 scaling no empty groups are created, however we only get 9 of them, which is too small a number for comfort. This is why the scaling of 500 seems to be the best, as this way we get a total of 17 groups with 3 of them empty. 17 groups are just about enough to conduct a regression analysis.

Table 3. shows all the relevant parameters characterising the regression model calculated from the scaling of 500 type sample. The quadratic value of the correlation coefficient is 90%, higher than in any other case before (scaling-250: 78%, scaling-1000: 90%), while the value of the Fcalculated is yet again much higher than the critical value. As in the previous chapter, we can again conclude that the distribution of industrial parks when the number of employees is considered can be characterised by a power-function. The γ scale index of the scale-free network is 1,4 this time, as opposed to 1,6. Both these indices are very close to each other.

Finally, the possibility of an exponential regression is tested again. Exponential regression analysis can only be performed if none of the dependent variables has a value of 0, therefore the regression is tested on the sample with the scaling of 1000. The quadratic value of the correlation coefficient is much lower than before, a mere 50%, so we can conclude that the use of linear regression was again justified.

CONCLUSION

All the calculations presented above prove that the Hungarian industrial parks show a pattern of scale-free distribution. The scale-free distribution tells us that there are strong network centres that have a lot of firms settled on their premises, which firms employ a lot of workers. It also tells us that the majority of the industrial parks barely attract any investors at all. With the use of industrial park statistics our original suggestion that business knowledge represented in the form of formal and informal connections of firms and employees, is geographically concentrated, proved to be true. Certain elements of business knowledge are concentrated in big industrial centres, the so called network centres. These network centres are heavily connected, the large number of connections lead to an increased pace of learning. That is why the centres provide the best opportunities for investors, so it can be expected that they get bigger and bigger, i.e. more connected as time goes on.

The findings also warn policy makers. One of the main consequences that can be drawn from the study is that when industrial development is the goal, it is not enough to only concentrate on the physical infrastructure. Network centres are not formed by regions where all the physical infrastructure is available, but rather regions that posses valuable business knowledge. Business knowledge can be developed if the cooperation among firms, universities and research institutes is encouraged. the cooperation should primarily concentrate on learning by doing.

It is also important to be noted that regional development should concentrate on network centres first. If investments are directed towards areas that are not so connected, the gain from them will not be too significant. On the other hand, development sources spent network centre development have a much better chance of creating spill over effects.

ACKNOWLEDGEMENT

My acknowledgments go to the Hungarian Ministry of Economy and Transport and Promei Kht. for letting me use the database of Hungarian industrial parks.

REFERENCES

[1] Barabási, A-L., *Behálózva - A hálózatok új tudománya* (Budapest, Magyar Könyvklub, 2003).
[2] Bartha, Z.: A tudáselőállítás kockázatai. *A II. Országos Közgazdaságtudományi Doktorandusz Konferencia előadásai*, 2003, 19-25.
[3] Freeman, C., *Technology Policy and Economic Performance: Lessons from Japan* (London, Pinter Publishers, 1987).
[4] Ministry of Economy and Transport, *Az ipari parkok szerepe a magyar gazdaságban* (Budapest, Ministry of Economy and Transport, 2004).
[5] Lundvall, B. A., Johnson, B., The Learning Economy, *Journal of Industry Studies, 1 (2)*, 1994, 23-42.
[6] Polányi, M., *The Tacit Dimension* (London, Routledge and Kegan, 1966).
[7] Porter, M. E.: *The Competitive Advantage of Nations* (New York, Free Press. 1994).