Regional Innovation and Entrepeneurship Research Center

WORKING PAPER SERIES

2021-02

The R-EntInn index
An attempt towards an integrated measurement of regional entrepreneurial ecosystems and innovation systems

Tamás Sebestyén
MTA-PTE Innovation and Economic Growth Research Group
University of Pécs, Faculty of Business and Economics

László Szerb
University of Pécs, Faculty of Business and Economics

Attila Varga
MTA-PTE Innovation and Economic Growth Research Group
University of Pécs, Faculty of Business and Economics

Regional Innovation and Entrepreneurship Research Center
Faculty of Business and Economics
University of Pécs

H-7622, Pécs Rákóczi str. 80.

Phone: +36-72-501-599/23121
E-mail: rierc.center@ktk.pte.hu
Web: http://hu.rierc.ktk.pte.hu/

The R-EntInn index

An attempt towards an integrated measurement of regional entrepreneurial ecosystems and innovation systems

Tamás Sebestyén
Regional Innovation and Entrepreneurship Research Center
MTA-PTE Innovation and Econonomic Growth Resarch Group
University of Pécs, Faculty of Business and Economics

László Szerb
Regional Innovation and Entrepreneurship Research Center
University of Pécs, Faculty of Business and Economics

Attila Varga
Regional Innovation and Entrepreneurship Research Center
MTA-PTE Innovation and Econonomic Growth Resarch Group
University of Pécs, Faculty of Business and Economics

Abstract

In this paper we establish a methodological tool which is able to comprehensively take into account the characteristics of local entprereneurial ecosystems and innovation systems while also reflecting the interconnectedness of these local systems shaping the availability of local resources. In this attempt, we draw on previous methodological developments carried out at the Regional Innovation and Entrepreneurship Research Center. First, we use the KAPIS index which is designed to capture position in knowledge networks. Second, we build on the REDI which describes regional entrepreneurial ecosystems in a comprehensive way. By integrating the two tools, we are able to set up a framework for jointly analyzing the role of local conditions and also interregional embeddedness in shaping the entrepreneurial environment.

Keywords

Innovation systems, entrepreneurial ecosystem, innovation networks, network structure

1. Introduction

Recognizing that innovation is inherently a collaborative process, recent research has shed light on the prominent role of cooperation in innovative activities (e.g. Lundvall, 2010). Through innovation, different networks of cooperation can thus contribute to the development and growth of regions, showing that policies targeting network formation can be effective tools in promoting regional development. In addition to the general understanding that collaborative ties can positively contribute to innovation and growth, results on this field specifically call attention to the importance of interregional cooperation (e.g. Hoekman et al. 2008, Varga et al. 2014, Sebestyén and Varga 2013). Moreover, these more distant ties of knowledge flows can significantly improve innovation performance in those lagging regions where the local supply of resources used in innovation is scarce, because the networks provide access to similar resources accumulated elsewhere (Varga and Sebestyén, 2017).

Also, literature on entrepreneurship emphasize that an ecosystem of different actors, institutional conditions can largely contribute to the entrepreneurial process and the development of a given location (e.g. Acs et al. 2017; Stam 2015; Alvedalen and Boschma 2017). However, these local entrepreneurial ecosystems are not isolated as well, they may interact through interregional networks, thus the latter may significantly contribute to the functioning of the local ecosystems. Literature on innovation ecosystems also draw attention to the fact that while conceptual descriptions are widely available, there is a scarcity of methodological developments, which aim to measure and assess these ecosystems in a rigorous, systematic manner (Alvedalen and Boschma 2017; Neumeyer and Santos 2018; De Hoyos-Ruperto et al. 2013; Motoyama and Watkins 2014; Roundy et al. 2018; Kurato et al. 2017).

In this paper we establish a methodological tool which is able to comprehensively take into account the characteristics of local entprereneurial ecosystems and their interconnectedness, while it provides an evaluation of how this connectedness contributes to the functioning of the ecosystems. In this attempt, we draw on previous methodological developments carried out at the Regional Innovation and Entrepreneurship Research Center. First, we use the KAPIS index which is designed to capture position in knowledge networks. Second, we build on the REDI which describes regional entrepreneurial ecosystems in a comprehensive way. By integrating the two tools, we are able to set up a framework for jointly analyzing the role of local conditions and also interregional embeddedness in shaping the entrepreneurial environment. The REDI captures the quality of local entrepreneurial ecosystems, while it also reflects some aspects of the innovation system within a given region, through one of its pillars. The KAPIS index then adds extra-regional innovation relationships to the REDI. It shows how innovation networks (innovation connections, knowledge flows, collaborations) shape the available resources for innovation and entrepreneurship. As a result, we end up with a comprehensive measurement tool which is able to reflect the local endowments and extra-local available resources for innovation and entrepreneurship in a relatively comprehensive way.

The paper is structured as follows. In section 2, we present the methodology behind the KAPIS index, which describes knowledge network positions. Section 3 provides a brief introduction of the REDI index. Section 4 presents the data used in the analysis and provides a set of preliminary results based on our analysis and the developed framework. Conclusion closes the paper with an outlook on further research areas.

2. The KAPIS network index

The Knowledge Access Position in Innovation Systems (KAPIS) index was developed in order to establish a measurement tool which evaluates the position of a node in a network by taking into account the level and quality of knowledge it can access from its direct and indirect network. This measurement tool takes into account

- the level of knowledge available at the direct and indirect partners,
- the embeddedness of direct and indirect partners within the network,
- the distances within the network i.e. closer knowledge and dense interactions are valued more.

In what follows, we provide some definitions and develop the index in detail.

2.1 Definitions

Let $G(N, E, p)$ denote a weighted graph (network), defined by the nodes $N: i, j, k \in \{1, \dots, n\}$, edges $: f \in \{1, \dots, e\}$ and the projection $: E \to \mathbb{R}$ interpreted on the edges. We use the subindices i, j, k and l to denote the nodes. All information on the graph G is condensed in the adjacency matrix $\mathbf{A} = [a_{ij}]$ where the general element $a_{ij} \geq 0$ gives the weight of the connection between nodes i and j.

In what follows, we normalize the connection intensities a_{ij} by the sum of these intensities within the whole network:

$$w_{ij} = \frac{a_{ij}}{\sum_i \sum_j a_{ij}} \tag{1}$$

The normalized connection weight w_{ij} thus shows how intense the connection between i and j is, relative to the total intensity of connections (flows) in the network. In case of binary connections where $a_{ij} \in \{0; 1\}$ this is practically the reciprocal of the total number of edges: $w_{ij} = 1/e$ for all $a_{ij} > 0$. The reason of using this normalization is twofold. First, it creates a common basis for comparing different networks with different units of measurement in $\mathbf{A}$. Second, as our method requires matrix-inversion, the computation of this matrix inverse is eased by this normalization.

Using the matrices $\mathbf{A}$ or $\mathbf{W}$, which describe connection weights, we can equivalently define matrix $\mathbf{B}$ describing direct partnerships between the nodes of the network. The general element of this matrix is defined as

$$b_{ij} = \begin{cases} 1, & \text{ha } a_{ij} > 0 \\ 0, & \text{ha } a_{ij} = 0 \end{cases} \tag{2}$$

Matrix $\mathbf{B}$ can be regarded as an indicator function which takes the value of 1 if i and j are direct neighbors in the network and it is 0 if they are not.

The matrices $\mathbf{A}$, $\mathbf{W}$ and $\mathbf{B}$ describe network structure and connection intensities. Another crucial ingredient in our approach is the knowledge level possessed by each node in the network. These knowledge levels are contained in the vector $\mathbf{q} = [q_i]$, where q_i is an appropriate measure of knowledge level at node i. Without losing generality, we normalize these knowledge levels with the average knowledge level:

$$k_i = \frac{q_i}{\sum_j q_j / n} \tag{3}$$

The k_i normalized knowledge levels are groups in vector $\mathbf{k}$, and reflect the extent to which a given node's knowledge is higher/lower than average in the network.

Finally, we define the direct neighborhood of node i as $N_1^i = \{j|b_{ij} = 1\}$, i.e the set of those nodes which are in direct connection with i.

2.2 Knowledge access position for the direct neighborhood

Using the definitions set forth above, we define the knowledge access position of node i with respect to its direct neighborhood as

$$z_i^1 = \sum_{j \in N_1^i} w_{ij} k_j + \sum_{j,l \in N_1^i} w_{jl} \tag{4}$$

The first term on the right hand side of equation (4) sums the knowledge levels of direct partners, weighted by the strength of connection, pointing to the given partner. The second term sums the connection intensities between the direct partners. These direct-neighborhood scores can be written in a compact matrix form as follows:

$$\mathbf{z}^1 = \mathbf{Wk} + (\mathbf{B} \circ \mathbf{WB})^{\mathrm{T}} \mathbf{1} \tag{5}$$

where $\circ$ denotes element-wise matrix multiplication and $\mathbf{1}$ is a vector full of ones.

The main novelty of this version compared to that of Sebestyén and Varga (2013a, 2013b) is that here connections weights w_{ij} and knowledge levels are interacting directly at the node/connection level, and then these interaction terms are summed over the direct neighborhood. In contrast, our previous approach aggregated knowledge and connection intensities at the neighborhood level first and the interaction was implemented at this aggregated level.

2.3 Knowledge access position for the whole network

The scores calculated for the direct neighborhood in (4) and (5) can be extended to include the knowledge which is accessible from farther parts of the network, through the direct partners. In order to do this, we use a recursive approach. Once the scores for the direct neighborhoods are calculated for all nodes, we assume that the knowledge access positions of the direct partners also contribute to the value of network position, weighted by the connection intensity through which they connect to the node in question. Formally, we write

$$z_i = z_i^1 + \sum_j w_{ij} z_j \tag{6}$$

With matrix notation, we can comprise (6) as

$$\mathbf{z} = \mathbf{z}^1 + \mathbf{Wz} \tag{7}$$

and we can express the final z_i scores in vector $\mathbf{z}$ as

$$\mathbf{z} = (\mathbf{I} - \mathbf{W})^{-1} \mathbf{z}^1$$

(8)

Substituting (5) into (8) shows that these knowledge access position scores (z_i) can be calculated from matrix $\mathbf{W}$, $\mathbf{B}$ and vector $\mathbf{k}$, i.e. information on the network structure and knowledge levels.

This approach is again somewhat different from that of Sebestyén and Varga (2013a, 2013b). While the latter solution explicitly goes through all neighborhoods at different distances, using an exogenous parameter to discount neighborhoods farther away, in this new solution the recursive definition implicitly considers all neighborhoods and gets rid of using an exogenous distance decay parameter.

2.3 Understanding the KAPIS index

The previous section has provided the definition of the Knowedge Access Position in Innovation Systems (KAPIS) index. In this part, we are going to use some illustrations in order to interpret the scores provided by this method. First, we derive the scores under a complete and symmetric network structure. Then, two sample networks are used to illuminate the working of the method in incomplete networks. Finally, we provide some comparison with standard measures of network centrality.

2.3.1 The KAPIS index in a complete and symmetric network

It is useful to establish how the KAPIS indices defined in (8) look like if the underlying network structure is complete and symmetric, i.e., all connections exist and have equal weight. Also, the knowledge levels are equal across nodes. In this case we have zeros on the main diagonal of matrix $\mathbf{A}$ and ones everywhere else. Given that the size of the network is n, the normalized weight matrix $\mathbf{W}$ contains zeros on its main diagonal and $w_{ij} = 1/[n(n-1)]$ everywhere else. If knowledge levels are identical, then $k_i = 1$ for all i. As all nodes are direct neighbors to all other nodes, the scores in (4) simplify to

$$z_i^1 = (n-1)\frac{1}{n(n-1)} + (n-1)(n-2)\frac{1}{n(n-1)} = \frac{n-1}{n}$$

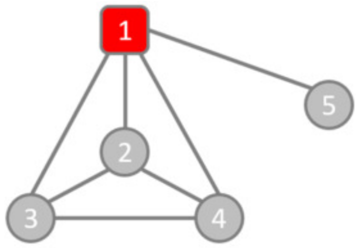
(9)

where the first part marks that $(n-1)$ direct partners enter with weight $1/[n(n-1)]$ and knowledge level 1, while the second part shows that there are altogether $(n-1)(n-2)$ connections among the neighbors with the given weight. It can be shown (see the Appendix) that in this case the full-network knowledge access position indices (z_i scores) equal to 1. As a result, the complete network serves as a natural reference point for the KAPIS index which becomes unity in this special case for all nodes. If the network becomes sparser (connections are deleted), the value of the index becomes smaller. It is easy to verify that in the empty network where $w_{ij} = 0$ everywhere, the KPAIS index becomes 0.

While being size-independent is appealing on the one hand, it makes it difficult to compare the resulting scores across different networks on the other. However, the raw score can be scaled according to some natural quantity of interest. For example, we may use the total knowledge level or the total number of connections as a scaling factor. In this case the resulting scores will show that how much of the total knowledge is accessible from a given position in the network. In the next section we present a simple example in this fashion.

2.3.2 The KAPIS index in a simple reference networks

Figure 1 below shows a simple reference network with $n = 5$ nodes. The highlighted node 1 is the most central actor, while nodes 2, 3 and 4 have symmetric, relatively well-connected positions and node 5 is a peripheral one with only one connection. We assume the connection weights to be of the same size (unity) and knowledge levels to be identical (unity).



1. Figure – A simple reference network

In table 1, we summarize the calculated KAPIS scores as in (8). We indicate the z_i^1 scores for the direct partners and the final z_i scores for the whole network. In addition to the absolute scores, we present the relative scores compared to that of node 1 in both cases.

Node	A: KAPIS for direct partners		B: KAPIS for whole network		Difference between B and A
	Absolute	Node1 = 100%	Absolute	Node1 = 100%	
1	0.7143	100	0.9008	100	26.11%
2	0.6429	90	0.8251	91.59	28.34%
3	0.6429	90	0.8251	91.59	28.34%
4	0.6429	90	0.8251	91.59	28.34%
5	0.0714	10	0.1358	15.07	90.08%

1. Table – KAPIS scores for the sample network in Figure 1

The numbers of the direct neighborhood reflect that the most valuable position is that of node 1 which is connected to all other nodes and has a dense neighborhood with respect to nodes 2, 3 and 4. The latter nodes has a 10% lower score, identically, reflecting their identical position in the network. While node 1 has direct access to all 5 units of knowledge in the network (1 at each node), its neighborhood lacks 3 connections (connections between node 5 and the other three). This means that the direct neighborhood score without normalization would be 4 + 6 = 10. Now if we look at node 2, it has direct access to 3 units of knowledge, while its direct neighborhood also counts 3 connections, so its non-normalized score for the direct neighborhood would be 3 + 6 = 9. This explains the 10% difference between the direct scores of node 1 and node 2. Similarly, as node 5 is very peripheral with only one direct partner, its score is much lower.

Looking at the absolute scores obtained for the whole network (i.e. according to (8) reveals that these scores are higher than the direct scores. This shows that the KAPIS index positively evaluates the role of indirect connections and the knowledge which is channeled to certain nodes through these indirect connections. In order to understand the nature of these differences, we should recall that the matric inverse in (8) can be written as the infinite sum

$$(\mathbf{I} - \mathbf{W})^{-1} = \mathbf{I} + \mathbf{W} + \mathbf{W}^2 + \mathbf{W}^3 + \cdots \tag{10}$$

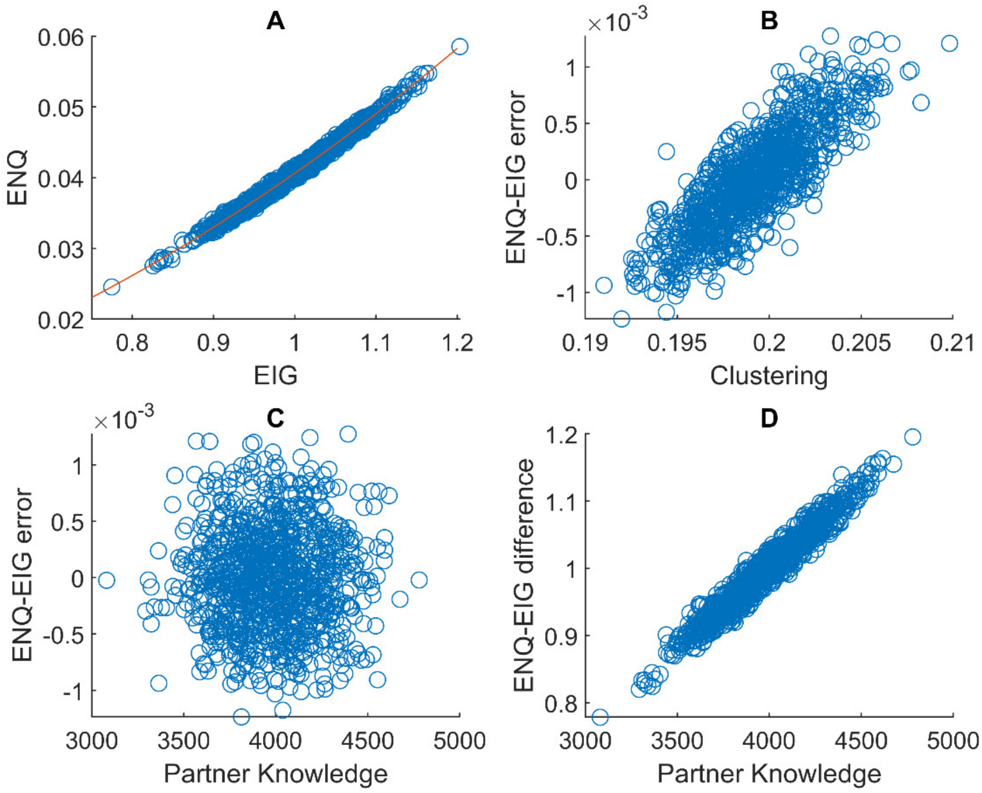
It is known that the powers of the adjacency matrix reflect the number of walks between any two nodes with length corresponding to the given power. As a result, the elements of this matrix inverse reflect the overall strength of direct and indirect connections between any two nodes in the network. In other terms, a given row of this inverse matrix contain weights for the node corresponding to the row reflecting its connection strength to all other nodes in the columns. Now, recalling (8) it can be seen that the final z_i scores are weighted sums of the z_i^1 scores where the weights reflect direct/indirec connection intensity and the z_i^1 scores reflect the embeddedness of the nides in their direct neighborhoods. The identity matrix on the right hand side of (10) ensures that own direct neighborhoods get the highest weight in all cases.

The last column of Table 1 shows that the additional value given by the indirect connections is roughly one quarter for nodes 1, 2, 3 and 4 while the score almost doubles for node 5. This is explained by the fact that for the first 4 nodes there is a minor part of the network outside of the direct neighborhood (for node 1 this addition only reflects recursive walks/connections within its direct neighborhood), while for node 5 the majority of knowledge and density in the network is found outside its direct neighborhood. This is also reflected by the relative scores in the whole network case: taking into account the whole network, the z_i scores are somewhat compressed than the z_i^1 scores on the direct neighborhood. This is because the knowledge and density accessible beyond the direct neighborhoods is larger for node 5 in the periphery compared to nodes 1, 2, 3 and 4 which are in the center.

2.3.3 The KAPIS index relative to existing positional measures

The recursive definition in (8) raises the question if our approach has something in common with the eigenvector centrality measure which is standard in network analysis. The eigenvector centrality is also based on a recursive definition and assumes that a node in a network is more central if it is surrounded by other central nodes.

Figure 2 summarizes the results of a small simulation exercise, where we generated a random network of the Erdős-Rényi type with 1000 nodes. The knowledge levels were generated randomly from a normal distribution, but correlated with the degree of nodes (more central nodes possess more knowledge). Then, we calculated the KAPIS for all nodes, given these knowledge levels and also calculated the eigenvector centralities for the nodes. Panel A in Figure 1 shows the correlation between eigenvector centrality and KAPIS, which proves to be quite strong. Then, we calculated the clustering coefficient for all nodes (the density of the direct neighborhood) and also calculated the difference between KAPIS and the estimated deterministic relationship between eigenvector centrality and KAPIS (marked by the red line in panel A). Panel B shows how clustering of nodes correlate with the over- or underestimation of eigenvector centrality by KAPIS. The significant positive correlation reflects that in the KAPIS index we not only take into account centrality (which is mostly captured by eigenvector centrality in a similar way), but also the density of neighborhoods – which is reflected by clustering. KAPIS is going to be systematically higher than eigenvector centrality for those nodes which are more strongly (and directly) embedded in intensive collaboration structures.



2. Figure – KAPIS versus eigenvector centrality

Panel C in Figure 2 shows a similar analysis with respect to knowledge levels. Here, we calculate the knowledge level of direct partners and this is correlated with the difference between observed KAPIS and estimated eigenvector centrality. The picture shows that there is no correlation between the two factors. The main conclusion is that while the difference between eigenvector centrality and KAPIS is explained by clustering, the knowledge level of partners do not affect this difference. Panel D uses an alternative approach where the relative difference between the KAPIS index and the eigenvector centrality of a node is measured on the horizontal axis. In this case there is a strong positive correlation.

To sum up, our KAPIS index is a close relative to the eigenvector centrality which captures centrality in a complex manner. However, in addition to centrality, the KAPIS index is also sensitive to the clustering of nodes' neighborhoods (strong collaboration density) and to the level of knowledge available at partners.

3. The REDI

In this section we briefly refer to the Regional Entrepreneurship and Development Index as a measure for entrepreneurship at the regional level. This description is based on the report in Varga et al. (2018).

3.1 The structure of REDI

The Regional Entrepreneurship and Development Index (REDI) has been constructed for capturing the contextual features of entrepreneurship across EU regions. The REDI method builds on the National Systems of Entrepreneurship Theory and provides a way to profile Regional Systems of Entrepreneurship. Important aspects of the REDI method including the Penalty for Bottleneck (PFB) analysis, which helps identifying constraining factors in the Regional Systems of Entrepreneurship. The novelty of this method that it portrays the entrepreneurial disparities amongst EU regions and provides country and regional level, tailor-made public policy suggestions to improve the level of entrepreneurship and optimize resource allocation over the different pillars of entrepreneurship.

A six level index-building methodology is followed while creating the REDI index: (1) sub-indicators (2) indicators (3) variables, (4) pillars, (5) sub-indices, and finally (6) the REDI super-index. The three sub-indices of attitudes (ATT), abilities (AB), and aspiration (ASP) constitute the entrepreneurship super-index, which is called REDI. All three sub-indices contain four or five pillars, which can be interpreted as quasi-independent building blocks of this entrepreneurship index. Each of the 14 pillars is the result of the multiplication of an individual variable and an associated institutional variable. In this case, institutional variables can be viewed as particular (regional-level) weights of the individual variables.

Structure of the GEDI 3 Sub-indexes 14 Pillars		National and regional institution variables	Regional level individual variables
Entrepreneurial Aspiration Sub-index	Financing	FINANCIAL INSTITUTIONS	INFORMAL INVESTMENT
	Globalization	CONNECTIVITY	EXPORT
	High growth	CLUSTERING	GAZELLE
	Process innovation	TECHNOLOGY DEVELOPMENT	NEW TECHNOLOGY
	Product innovation	TECHNOLOGY TRANSFER	NEW PRODUCT
Entrepreneurial Ability Sub-index	Competition	BUSINESS STRATEGY	COMPETITORS
	Human capital	EDUCATION & TRAINING	EDUCATION LEVEL
	Technology sector	ABSORPTIVE CAPACITY	TECHNOLOGY LEVEL
	Opportunity start-up	BUSINESS ENVIRONMENT	OPPORTUNITY MOTIVATION
Entrepreneurial Attitudes Sub-index	Cultural support	OPEN SOCIETY	CARRIER STATUS
	Networking	SOCIAL CAPITAL	KNOW ENTREPRENEURS
	Risk acceptance	BUSINESS RISK	BUSINESS ACCEPTANCE
	Startup skills	QUALITY OF EDUCATION	SKILL PERCEPTION
	Opportunity perception	MARKET AGGLOMERATION	OPPORTUNITY RECOGNITION

3. Figure – The structure of the Regional Entrepreneurship Development Index

3.2 The creation of the Regional Entrepreneurship and Development Index

All pillars from the variables are calculated using the interaction variable method; that is, by multiplying the individual variable with the proper institutional variable:

$$z_{i,j} = IND_{i,j} * INS_{i,j}$$

(10)

for all j= 1 ... k, the number of individual and institutional variables. $IND_{i,j}$ is the original score value for region i and variable j individual variable, $INS_{i,j}$ is the original score value for region i and variable j institutional variable, $z_{i,j}$ is the original pillar value for region i and pillar j.

All index building is based on a benchmarking principle. The selection of the proper benchmarking considerably influences the index points and also the rank of the regions. However, the existence of outliers could lead to set up inappropriate benchmarks. Hence, it is needed to handle extreme value outliers. Capping is also frequently used to handle outliers. The question relates to the value of the cap. The 95-percentile score adjustment is selected meaning that any observed values higher than the 95 percentile is lowered to the 95 percentile. It also means that at least five percent of different regions reach the maximum value in all of the 14 pillars. Like other composite index components, the pillars are in different magnitudes. In order to be in exactly the same range, the normalization of the pillars is necessary. After handling the outliers the pillar values are normalized, where the distance normalization technique was used that preserves the distance (relative differences) amongst the regions:

$$x_{i,j} = \frac{z_{i,j}}{\max z_{i,j}} \tag{11}$$

for all j= 1,..m. m=14 is the number of pillars, $x_{i,j}$ is the normalized score value for region i and pillar j, $z_{i,j}$ is the original pillar value for region i and pillar j, $\max z_{i,j}$ is the maximum value for pillar j.

Applying the distance methodology the pillar values are all in the range [0,1], however the lowest pillar value is not necessary equal to 0. In this case all regions' efforts are evaluated in relation to the benchmarking region but the worst region is not set to zero per se.

The different averages of the normalized values of the 14 pillars imply that reaching the same performance requires different effort and consequently resources. Higher average values - e.g. Opportunity startup – could mean that it is easier to reach better scores as compared to lower average value – e.g. Financing. Since the aim is to apply REDI for public policy purposes, the additional resources for the same marginal improvement of the pillar values should be the same for all of the 14 pillars, on the average. So improving by 0.1 unit Opportunity startup should require the same additional resource as compared to all the other 13 pillars. As a consequence, we need a transformation to equate the average values of the 14 pillars.

Practically we have calculated the average values of the 14 pillars after the capping adjustment and the normalization and made the following average adjustment. Let's x_i to be the normalized score for region i for a particular pillar j. The arithmetic average of pillar j for region n regions is:

$$\bar{x}_j = \frac{\sum_{i=1}^{n} x_{i,j}}{n} \qquad \forall\, j \tag{12}$$

We want to transform the x_(i,j) values such that the potential values lay in the [0,1] range.

$$y_{i,j} = x_{i,j}^{k} \tag{13}$$

where k is the "strength of adjustment", the kth moment of $\bar{x}_j$ is exactly the needed average, $\bar{y}_j$. We have to find the root of the following equation for k:

$$\sum_{i=1}^{n} x_{i,j}^{k} - n\bar{y}_j = 0 \tag{14}$$

It is easy to see based on previous conditions and derivatives that the function is decreasing and convex which means it can be quickly solved using the well-known Newton – Raphson method with an initial guess of 0. After obtaining k, the computations are straightforward. Note that if

$$\bar{x}_j < \bar{y}_j \quad k < 1$$

$$\bar{x}_j = \bar{y}_j \quad k = 1$$

$$\bar{x}_j > \bar{y}_j \quad k > 1$$

(15)

that is k be thought of as the strength (and direction) of adjustment.

We have defined entrepreneurship as the dynamic interaction of entrepreneurial attitudes, abilities, and aspirations and developed the Penalty for Bottleneck (PFB) methodology for measuring and quantifying these interactions (Acs et al., 2013a; Rappai and Szerb 2011). Bottleneck is defined as the worst performing weakest link, or binding constraint in the system. With respect to entrepreneurship, by bottleneck we mean a shortage or the lowest level of a particular entrepreneurial indicator as compared to other indicators of the sub-index. This notion of bottleneck is important for policy purposes. Our model suggests that attitudes, ability and aspiration interact, and if they are out of balance, entrepreneurship is inhibited.

The sub-indices are composed of four or five components, defined as indicators that should be adjusted in a way that takes this notion of balance into account. After normalizing the scores of all the indicators, the value of each indicator of a sub-index in a region is penalized by linking it to the score of the indicator with the weakest performance in that region. This simulates the notion of a bottleneck, and if the weakest indicator were improved, the particular sub-index and ultimately the whole REDI would show a significant improvement. To the contrary, improving a relatively high pillar value will presumably enhance only the value of the pillar itself, and in this case a much smaller increase of the whole REDI index can be anticipated. Moreover, the penalty should be higher if differences are higher. Looking from either the configuration or the weakest link perspective it implies that stable and efficient sub-index configurations are those that are balanced (have about the same level) in all indicators. Mathematically, we model the penalty for bottlenecks by modifying Casado-Tarabusi and Palazzi (2004) original function for our purposes. The penalty function is defined as:

$$h_{(i),j} = min\, y_{(i),j} + \left[1 - e^{-(y_{(i)j} - min y_{(i),j})}\right]$$

(16)

where $h_{(i),j}$ is the modified, post-penalty value of pillar j in region i, $y_{(i)j}$ is the normalized value of index component j in region i, $min\, y_{(i),j}$ is the lowest value of $y_{(i)j}$ for region i. i = 1,2,……n is the number of regions, j= 1,2,…m is the number of pillars.

Definitely, the advantage of this method that it is an analytical method, therefore it is not sensitive to the size of the sample. There are two potential drawbacks of the PFB method. One is the arbitrary selection of the magnitude of the penalty. The other problem is that we cannot exclude fully the potential that a particularly good feature can have a positive effect on the weaker performing features. While this could also happen, most of the entrepreneurship policy experts hold that policy should focus on improving the weakest link in the system. On the other hand, both theories emphasize the importance of balanced performance and characteristics. Altogether, we claim that the PFB methodology is theoretically better than the arithmetic average calculation. However, the PFB adjusted REDI is not necessary an optimal solution since the magnitude of the penalty is unknown. The most important message for economic development policy is that improvement can only be achieved by abolishing the weakest link of the system, which has a constraining effect on other pillars.

Due to the average pillar adjustment the marginal rate of substitution becomes the same for all indicators. However, the real substitution rate of the pillar values of a particular region depends on the weakest pillar's relative ratio compared to other pillars. Most importantly, the penalty function should reflect to the magnitude of the penalty, lower difference implies lower penalty while higher unbalance implies higher penalty. The penalty function also reflects to the compensation of the loss of one pillar for a gain in another pillar.

The value of a sub-index for any region was then calculated as the arithmetic average of its PFB-adjusted indicators for that sub-index multiplied by 100 to get a 100 point scale:

$$ATT_i = 100 \sum_{j=1}^{5} h_{i,j}$$

$$ABT_i = 100 \sum_{j=6}^{10} h_{i,j}$$

$$ASP_i = 100 \sum_{j=11}^{14} h_{i,j}$$

(17)

where h_(i,j) is the modified, post-penalty value of pillar j in region i, i = 1,2,……n is the number of regions, j= 1,2,…m is the number of pillars.

The REDI super-index is simply the arithmetic average of the three sub-indices:

$$REDI_i = \frac{1}{3}(ATT_i + ABT_i + ASP_i)$$

(18)

where i = 1,2,……,n is the number of regions.

4. Integrating the KAPIS and REDI indices – interconnected entrepreneurial ecosystem

In this section we provide a first attempt to integrate the two indices, the KAPIS and the REDI indices. The underlying logic behind this integration is that regional entrepreneurial ecosystems are not isolated from each other. Actors in the local ecosystems may be connected to actors at other places, thus contributing to an interregional network of knowledge flows which can contribute to the performance of these systems.

So what we do is that we utilize properties of the KAPIS index which is able to describe the position within such knowledge networks of the regions under question. However, instead of directly using some proxy for knowledge levels in the KAPIS index, we integrate the REDI as the q_i knowledge levels. This solution can be interpreted as a comprehensive measure of interconnected entrepreneurial ecosystems through innovation connections. The local conditions of entrepreneurship and also some aspects of the local innovation systems are accounted for in the REDI, while the connectedness of these local conditions is taken into account by the KAPIS index. This integration is able to reflect how interregional cooperation contributes to local resources behind entrepreneurship, thus providing a more complex picture of the latter.

However, there is another modification required for the KAPIS index to fully accommodate this framework. The KAPIS index was designed to take into account the external knowledge sources available to a given node, so it does not calculate with the knowledge level of the node under question. As a result, it does not reflect automatically the quality of the local ecosystems. However, a simple modification to the structure of the index solves this issue. Equation (4) has to be modified in the following way:

$$z_i^1 = k_i + \sum_{j \in N_1^i} w_{ij} k_j + \sum_{j,l \in N_1^i} w_{jl} \tag{19}$$

which means the addition of the local (region i) normalized REDI as k_i. This way, we take into account the local entrepreneurial ecosystem as measured by REDI with a unit weight in the KAPSI index.

In the next subsection we describe the data used for this exercise and then we present some preliminary results.

4.1 Data

With respect to the REDI scores, we use the data as described and analyzed in Szerb et al. (2017).

With respect to data on network connections, we use information available on EU-funded Framework Programmes. This information can be retrieved from the Cordis database. For the analysis in this paper, we use information on all projects funded in the three waves of the Framework Programmes: FP5, FP6 and FP7 which means that the data covers the years from 1999 to 2013. The basic unit of this data is a project-participant pair. This means a particular institution (as participant, e.g., university, company) being involved in a funded project. First, we use information on the projects: the contract numbers of the specific projects are used as unique identifiers, and the duration (starting and ending years) of the projects allow us to have a longitudinal approach on the collaboration patterns. Second, information on the participants is used: their location, as the NUTS3 level region they belong to and the type of the institution (e.g., higher education institution, industry actor).

This data had to go through two waves of data cleaning. First, we had to clean regional classification. Although the Cordis dataset provides NUTS3-level categorization of the participants, this is not complete and come with errors in several cases. We did a complete re-classification in this respect based on the information on postal codes, addresses and cities provided in Cordis. If this information was not enough, manual checks were done to assign a clean regional code at the NUTS3 level to all institutions. Second, as the participant identifiers provided by Cordis are similarly problematic, especially to be used across different FP programmes, we did a complete re-identification of institutions. Using information on the name, location (region) and address of the participants we run a string-matching algorithm to reveal the similarity of every institution-pair. The same procedure was done manually as well on a subsample of institutions. The latter provided reference-cases where we were sure about which institutions are the same and which are different. This reference subsample was then confronted with the algorithmic results, in order to establish an ambiguity range. Institution-pairs with a similarity score below this range were assumed to be different, pairs above this range were assumed to be identical. Pairs falling into the ambiguity range were manually checked again to finally arrive at a clean identification of institutions.

In the cleaned dataset we have information about every funded project, the duration of the project, the participants of the project, their location at the NUTS3 level and their type being higher education institution, research institution, industry actor or other. In this analysis we consider only the first three types with merging research institutions and higher education institutions into one category. For simplicity, we will refer to the latter group as research institutions in general.

Before we go into the details about the data manipulation, we have to note that the data we are using reflect an important, but specific aspect of innovation (scientific) networks. First of all, the cooperative connections recorded in this dataset reflect research cooperation: while it is able to reflect how and where the generation of new knowledge is attempted, these records do not show whether these attempts are successful or not (e.g., in the form of scientific publications or patents). Also, the records are selective in the sense, that we have information on funded projects and unsuccessful applications as well as research collaboration without formal infrastructure are out of sight.

Our starting point for data manipulation is the project matrix $\mathbf{P}$ the rows of this matrix correspond to institutions whereas the columns represent projects. A given cell of the matrix is one if institution i was participating in project k. From this project matrix, simple matrix manipulation provides the adjacency matrix $\mathbf{A}$ for all years in our sample: $\mathbf{A} = \mathbf{P}\mathbf{P}^{\mathbf{T}}$, where $\mathbf{P}^{\mathbf{T}}$ is the transpose of$\mathbf{P}^{\mathbf{T}}$. The resulting $\mathbf{A}$ adjacency matrix provides the number of ongoing joint projects between any pair of institutions. Being our starting point for further calculations, this adjacency matrix gives a snapshot of collaboration patterns between institutions with a weighted perspective: we account for the number of joint projects, reflecting the intensity of collaboration.

The abovementioned adjacency matrix $\mathbf{A}$ contains information between all pairs of institutions, regardless of their location (region) and type (research institution or industry actor). In order to account for these features, we use two categorization vectors. $\mathbf{d}^T$ refers to the type of institutions: it has one entry (row) for all institutions and contains 1 of the given institution is a research institution and 2 if it is a company/industry actor. Similarly, $\mathbf{d}^R$ refers to the location of institutions and one entry (row) contains the index of the region the institution belongs to.

In order to ease further exposition, we reshape the adjacency matrix $\mathbf{A}$ into an array $\mathbf{W}$ which structures connections between institutions along their location and type as well. The general element of it is defined as follows:

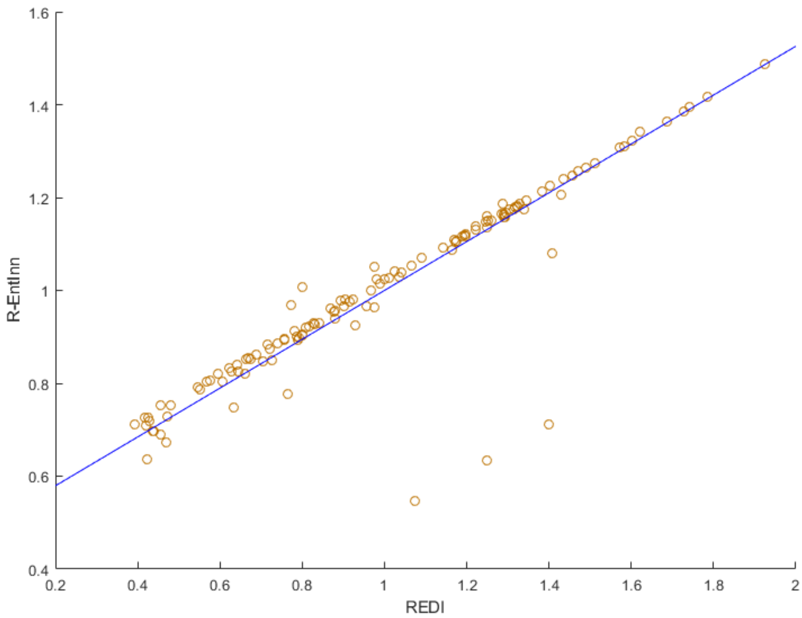
$$w_{rfi,qgj} = a_{l_1 l_2} | d_{l_1}^R = r, d_{l_2}^R = q, d_{l_1}^T = f, d_{l_2}^T = g$$

In other words, $w_{rfi,qgj}$ describes the number of joint collaboration projects between institution i of type f in region r and institution j of type g in region q. Here the indices $f, g = 1,2$, indicating whether institutions are companies (1) or research institutions (2). Then, $r, q = 1,2, \dots, R$ refer to region indices, while $i, j = 1,2, \dots, I_{f,r}$ reflect the indices of institutions. Note that $I_{f,r}$ is different for all region r and institution type f, representing the number of institutions of the given type in the given institution.

4.2 Preliminary results

In what follows, we present the results of a first sample of calculations with the integrated tool. We use the REDI scores for 2007-2011 and the respective average number of FP collaboration ties between regions to calculate the KAPIS scores. In order to take into account the fact that REDI scores are available for a different set of regions (there are NUTS1 and NUTS2 regions in the sample), the collaboration ties were normalized by the population of the given regions.

Once the KAPIS scores are calculated, we get a score for every region, reflecting the overall entrepreneurial ecosystem in a given region together with its interregional embeddedness. Those regions are ranked higher in this scoring which have a good access to extra-regional entrepreneurial ecosystems in addition to their local environment. In other terms, to have a high score, a region needs to have both a good local ecosystem and a rich network of other ecosystems in their reach.



4. Figure – The REDI and R-EntInn scores

In figure 3, we visualized the relationship between the original REDI scores (on the horizontal axis) and the R-EntInn index (the integrated REDI-KAPIS index) (vertical axis). In both cases, the original scores were normalized with the mean of the given dimensions, so the two axes measure the relative deviation of the given score from the mean (the mean equals 1). The blue bubbles represent the regions included in the analysis and the red line shows the least squares fit which crosses the 1-1 point in the middle.

The first observation is that there is a clear positive relationship between the REDI and the integrated index which is not surprising as the former enters the latter on a one-to-one basis. As a result, what is more interesting is the deviation of the observed R-EntInn scores from the one indicated by the red line. Second, the REDI scores are more dispersed: they more-or less uniformly cover the 0.3-2 range. However, the R-EntInn scores are less disperse, much of them is found in the 0.6-1.5 range. This lower dispersion strongly reflects the structure of the FP-collaboration network. This network is relatively dense, which means that many of the regions have quite a few links to others. On the other hand, most of the regions, even if they are not the most developed from an entrepreneurial perspective, are directly or indirectly connected to the best performing regions in the REDI scores. This means that interregional connections are able to compensate for a less-developed local entrepreneurial ecosystem. The three outliers to the bottom of the figure correspond to those regions which are not integrated to the interregional innovation network at all.

Figure 3 shows another important tendency: significant deviations from the average KAPIS score correlate with the REDI: those regions can benefit significantly more from their partners which are relatively worse in the REDI ranking as well. This is line with previous findings in the literature which shows that relatively less developed regions (typically with a low REDI score and weak innovation systems) can benefit more from extra-regional linkages, while for places with rich local resources for innovation and entrepreneurship these relationships are not crucial

Conclusion

In this paper we established integrated the REDI and KAPIS indices in order to set forth a methodological tool which is able to comprehensively account for the characteristics of local entrepreneurial ecosystems and innovation systems while also reflecting the interconnectedness of these local systems shaping the availability of local resources. By integrating the two tools, we are able to set up a framework for jointly analyzing the role of local conditions and also interregional embeddedness in shaping the entrepreneurial environment. After setting up the framework, we used data on REDI scores and data on Framerwork program collaborations to proxy interregional knowledge networks. This data was used to calculate the integrated R-EntInn scores. Our results show that embeddedness in a relatively dense network of FP collaborations can significantly dampen the variability in the final scores by providing access to highly developed regions for less favored ones.

Our further work can focus on primarily two extensions. First, we can use the institutional dimension of the Framework Program data in order to provide a more detailed picture of the role of interregional connectedness of entrepreneurial ecosystems. Second, this framework may be employed to estimate the effect of interregional connectedness on the quality of the local entrepreneurial ecosystems.

Köszönetnyilvánítás

A tanulmány az Európai Unió, Magyarország és az Európai Szociális Alap társfinanszírozása által biztosított forrásból az EFOP-3.6.2-16-2017-00017 azonosítójú "Fenntartható, intelligens és befogadó regionális és városi modellek" című projekt keretében jött létre.

References

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

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

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

Hoekman, J., Frenken, K., van Oort, F. (2008): Collaboration Networks as Carriers of Knowledge Spillovers: Evidence from EU27 Regions. DIME Working Paper in the series on 'Dynamics of Knowledge Accumulation, Competitiveness, Regional Cohesion and Economic Policies, FP7 Project.'

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

Lundvall, B. (ed. 2010): National Systems of Innovation: Toward a Theory of Innovation and Interactive Learning. London, UK: Anthem Press.

Maggioni M., Nosvelli M., Uberti, E. (2007): Space versus networks in the geography of innovation: a European analysis. Papers in Regional Science, 86: 471–493.

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

Ponds, R., van Oort, F., Frenken, K. (2010) Innovation, spillovers and university--industry collaboration: an extended knowledge production function approach. Journal of Economic Geography, 10, 231-255.

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

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

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

Sebestyén, T., Varga, A. (2013a): Research productivity and the quality of interregional knowledge networks. Annals of Regional Science, DOI 10.1007/s00168-012-0545-x

Sebestyén, T., Varga, A. (2013b) A novel comprehensive index of network position and node characteristics in knowledge networks: Ego Network Quality. In Scherngell, Thomas (Ed.) The geography of networks and R&D collaborations. Springer, New York, 71-97.

Sebestyén, T., Varga, A. (2019): Knowledge networks in regional development: an agent-based model and its application. Regional Studies 53 : 9 pp. 1333-1343. , 11 p. (2019)

Szerb, L., Vörös, Zs., Komlósi, É., Acs, Z.J., Páger, B., Ortega-Argilés, R., Abaligeti, G. (2017). The New Regional Entrepreneurship and Development Index: Structure, Data and Description of Methodology. Unpublished manuscript.

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

Varga, A., Pontikakis, D., Chorafakis G. (2014): Metropolitan Edison and Cosmopolitan Pasteur? Agglomeration and Interregional Research Network Effects on European R&D Productivity. Journal of Economic Geography, 14, 229-263.

Varga, A., Sebestyén, T. (2017): Does EU Framework Program Participation Affect Regional Innovation? The Differentiating Role of Economic Development. International Regional Science Review, 40(4) pp. 1-35.

Varga, A., Sebestyén, T., Szabó, N., Szerb, L. (2020): Estimating the economic impacts of knowledge network and entrepreneurship development in smart specialisation policy. Regional Studies 54 : 1 pp. 48-59. , 12 p. (2020)