

Results of a Cluster Analysis of the ICT Sector*

Case study from the Slovak republic

Emília Madudová

Department of Communications
University of Žilina, FPEDAS
Slovak Republic
Emilia.Madudova@fpedas.uniza.sk

Peter Kolarovszki

Department of Communications
University of Žilina, FPEDAS
Slovak Republic
Peter.Kolarovszki@fpedas.uniza.sk

Abstract—Even when the ICT industry in the Slovak Republic has been one of few industries in which the employment has grown last years, the socio and economic regional disparities in Slovak regions influence an economic development in Slovak Republic (further SR). The purpose of this study is to analyze clusters from the 8 regions in the Slovak republic in terms of the ICT sector. The paper deals with the problem of socio-economic disparities in the Slovak Republic and their impact on the development of the ICT sector. Through hierarchical clustering and Cophenetic Correlation Coefficient the socio-economic regional disparities and their impact on the ICT sector in SR have been researched. Based on the research results and findings, the impact of regional disparities on the ICT sector has been evaluated.

Keywords—ICT sector, Regional Economy, Regional Disparities, Cluster analysis, Cophenetic Correlation Coefficient

I. INTRODUCTION

Many authors have already few years been interested in concepts that include a variety of phenomena that can be evaluated through quantitative and efficacious indicators [1]. The development of a region is affected by concepts linked to the ability to lagging economic conditions. This paper primary focuses on clustering in the sector of Information and Communication Services (further ICT).

Automotive and ICT industries are the driving force of the Slovak Economy. The Slovak Republic is a country that in the world rankings has ranked the first place in a number of cars produces per capita in last few years. From this point of view, the automotive industry is the key industry for the GDP in the Slovak Republic. The ICT industry in the Slovak Republic has been one of few industries in which the employment has continuously grow. Great disparities in the socio-economic levels of individual regions may have a significant impact on the country's overall economic development.

Recent studies of European countries indicate that the contribution of the ICT sector to the regional economy is weakening and slowing economic growth [2]. On the other hand, some economists have long recognized that technological

change is one of the most important forces driving economic growth, together with human capital and knowledge accumulation. This conclusion can be found, for instance [3].

II. METHODOLOGY

The goal of the primary research was to draw up the cluster analysis based on regional disparities economic and development in the ICT sector. Data were used from the SK NACE classification of economic branches, information from INFOSTAT and from the Statistical Office of the SR. The firm database included 7 452 firms from all the regions of the Slovak republic, operating in ICT sector in the Slovak republic.

The ICT sector according to the Statistical Office of the SR (SK NACE rev.2) includes: *Wired telecommunications activities, Wireless telecommunications activities, Satellite telecommunications activities, Other telecommunications activities, Computer programming activities, Computer consultancy activities, Computer facilities management activities, Other information technology and computer service activities, Data processing, hosting and related activities, Web portals, News agency activities, Other information service activities.*

A. Methodology Specifications

The cluster analysis was used to find out if there is any existence of clusters in this sector. Once the different clusters have been identified, they could then be examined in greater detail to establish what the differences are.

Each data point has been defined in the agglomerative and hierarchical approach consequently existing clusters were combined at each step. Every object was considered as a separate cluster of objects. Subsequently, the objects and clusters have been associated based on the distance between them. Finally, all the objects formed a cluster. The criterion for determining similarity and distance between cases, criterion for determining which clusters are merged at successive steps and the number of cluster needed were took into account. In a process of clustering, we linked to clusters of the most similar clusters (we refer the m as q,p), where d_{ir} mean a distance between clusters.

In Single linkage (1), the distance between two clusters was defined. There is the minimum distance between any

This paper is supported by the APVV agency No. APVV-010-10 and the EUREKA AUTOEPCIS, No. E! 7592 RFID



single data point in the first cluster and any single data point in the second cluster in the single linkage.

$$d_{t,r} = \min (d_{t,r}, d_{t,r}) \quad (1)$$

Average linkage (2), defines the distance between two clusters as the average of the distances between all pairs of regions in which one member of the pair is from each of the clusters.

$$d_{t,r} = \frac{n_p \cdot d_{t,r} + n_q \cdot d_{t,r}}{n_p + n_q} \quad (2)$$

For each cluster, the means for all variables were calculated in Ward's method (3). Then, for each case, the squared Euclidean distance to the cluster means was calculated.

$$d_{t,r} = \frac{(n_p + n_q) d_{t,p} + (n_r + n_s) d_{t,q} - n_p \cdot d_{p,q}}{n_t + n_r} \quad (3)$$

Centroid method (4) calculates the distance between two clusters as the sum of distances between cluster means for all of the variables. In the centroid method, the centroid of a merged cluster is a weighted combination of the centroids of the two individual clusters, where the weights are proportional to the sizes of the clusters.

$$(4) \quad d_{t,r} = \frac{n_p}{n_p + n_q} \cdot d_{t,r} + \frac{n_q}{n_p + n_q} \cdot d_{t,r} - \frac{n_p \cdot n_q}{(n_p + n_q)^2} \cdot d_{p,q}$$

Median method (5) was used because of the existence of the disadvantage in the centroid method, because if the sizes of two groups are very different, then the centroid of the new group will be very close to that of the larger group and may remain within that group. Median method takes into consideration the size of a cluster, rather than a simple mean.

$$d_{t,r} = \frac{d_{t,r} + d_{t,r} - d_{p,q}}{2} \quad (5)$$

In Complete linkage (6), the maximum distance between two clusters was defined, the maximum distance between any single data point in the first cluster and any single data point in the second cluster.

$$d_{t,r} = \max (d_{t,r}, d_{t,r}) \quad (6)$$

For all this linkages, the Euclidean metric was applied.

It was appropriate to express the degree of correlation between the properties of objects and the resulting process of aggregating by exact indicator, even when the properties of clustering algorithms are known as well. For the purposes of this cluster analysis, the Cophenetic Correlation Coefficient – CPCC has been chosen. This is a correlation coefficient between the primary elements of the distance matrix D objects and elements of the cophenetic matrix C. The Cophenetic matrix is the triangular matrix, which elements make up the distance between the object clustering at the time when they were first added to the cluster. The cophenetic coefficient CPCC is calculated on the basis of the relationship (7).

$$CPCC = \frac{cov_{c,c}}{S_d \cdot S_c} \quad (7)$$

Where d is an element of primary matrix and c is an element of cophenetic matrix. This coefficient has been chosen for the reason that, higher is the value of cophenetic correlation coefficient, lower is the information loss, generated in the process of clustering.

All the calculus were calculated in Matlab and graphically drawn in Statistica program by dendrogram.

B. Composite indicators: a multivariate analysis

In this paper, we apply a quantitative approach based on a multivariate analysis of the principal components that influence regional disparities. We have grouped the selected variables into three main groups (Table 1) Demography (D), Economics general (EG), Economics ICT (EICT).

TABLE I. CLUSTER ANALYSIS INDICATORS

Main Group	Code	Definition	Year
Demography	D	Population density	2005-2015
		Population in rural areas	2005-2015
Economics general	EG	The unemployment rate	2005-2015
		The regional gross value added (RGVA)	2005-2015
Economics ICT	EICT	Share of ICT to RGVA	2005-2015
		Average nominal monthly wage in ICT sector	2005-2015
		Number of ICT firms (regional indicator)	2005-2015
		Number of employees (regional indicator)	2005-2015

III. CASE STUDY OF THE ICT SECTOR

A positive aspect of the ICT sector is higher wages than in other sectors. Among the experts are continuing discussions about the importance of the ICT sector for the Slovak economy. Even when the number of employees in the sector is below the overall employment of the Slovak republic (2,5 %), the share of the total tax levy is 4,7 %. This results from the high-wage sector. The share of ICT firms income tax is 9,6 %.

The main contribution of the ICT sector across the economy is promotion of economic growth of the country. However, if the country wants these effects strengthen, it is needed to create better conditions in this sector. The Slovak ITC sector should propose in two areas: the business environment and human capital.

Before 1989, the former Czechoslovakia showed one of the smallest regional disparities in Europe. But, it was only at the cost of reducing the competitiveness in "stronger" regions. After the transformation to a market economy (1989), the

disparities has started grow rapidly.

Even when there are large disparities in unemployment rate in individual regions, the regional gross value added disparities in ICT are not as significant.

The best results achieves Bratislava region. The problem not only in the ICT sector is that ICT firms, even though they operate in different regions of Slovakia, have headquarters in Bratislava region (which the capital city Bratislava) and all the indicators are reported in the Bratislava Region. Table II presents, how the number of the ICT sector firms has evolved since 2005.

TABLE II. NUMBER OF ICT FIRMS IN TIME PERIOD 2005-2015

Year	Number of ICT firms		
	EU27	Slovak Republic	Quotient
2005	716 000	2 051	0,28
2006	145000	2 448	0,32
2007	180 487	2 805	0,35
2008	802 333	2980	0,37
2009	796 242	935	0,11
2010	827 839	12 120	1,38
2013	906 343	14 512	1,60
2015	904560	14 760	1,406

IV. THE RESULTS

The methods of cluster analysis were used to cluster ICT sector in the Slovak republic. The Squared Euclidean distance was used to measure the distance. Degree of correlation between the properties of objects and the resulting process of aggregating by exact indicator, was evaluated by the Cophenetic Correlation Coefficient – CPCC. As can be seen in table III.

TABLE III. COPHENETIC CORRELATION COEFFICIENT SCORE

Metric	Cluster analysis	
	Method/linkage	Cophenetic Correlation Coefficient
Euclidean Metric	Average linkage	0,9715
	Centroid Method	<i>not appropriate</i>
	Complete Linkage	0,9676
	Median Method	0,9686
	Single Linkage	0,9579
	Weighted	0,9705
	Ward's method	0,9489

The most appropriate according to CPCC is the average linkage, this linkage defines the distance between two clusters as the average of the distances between all pairs of regions in which one member of the pair is from each of the clusters.

It can be stated, that the ICT sector consist of two different clusters in terms of spatial disparities Fig.1. The Bratislava Region (BA) and the other regions (Level 1, Table IV.). Bratislava Region (BA) is the smallest Slovak region but with the capital city Bratislava. Most economic indicators are at the highest level in this region, this “indicator score” is influenced by the statistics organization, when as written above, most of the ICT firms have headquarter in Bratislava region, or are registered in Bratislava region, but truly operate in different Slovak region. The advantage of Bratislava Region is an excellent transportation position as well.

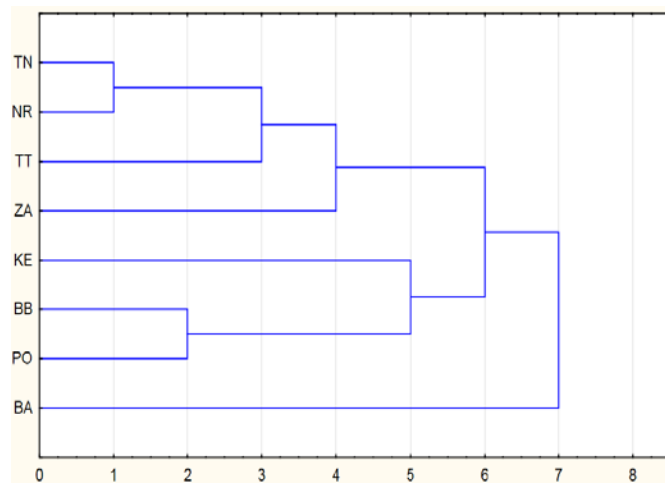


Fig. 1. Dendrogram of Hierarchical clustering

The current situation in the ICT sector the best presents the third level, with the first cluster – Bratislava Region (BA), the second one with Banská Bystrica Region (BB) and Prešov Region (PO), the third one Košice Region (KE), the fourth one Žilina region (ZA) and the fifth one the Trnava, Trenčín and Nitra regions (TT, TN, NR). Žilina Region (ZA) - The third largest region on 6th place in population density rate. Economic performance can be assessed as moderate. Žilina region has an excellent international transport position. Košice Region (KE) - The fourth largest region with a positive natural growth but negative migration tendency. According to GDP, Košice region is the most important region of SR.

TABLE IV. DEFINING OF THE ICT CLUSTERS

Cluster	Cluster analysis			
	Level 1	Level 2	Level 3	Level 4
1	BA	BA	BA	BA
2	PO, BB, KE, ZA, TT, NR, TN	KE, PO, BB	BB, PO	BB, PO
3		TN, NR, TT, ZA	KE	KE
4			ZA	ZA
5			TT, NR, TN	TT
6				TN, NR

Banská Bystrica Region (BB) is the largest region with the lowest density (69,5 inhabitants per km²), since 1997 unsuccessfully struggling with population decline. Fairly rugged terrain with massive mountains of the north complicates the development of various forms of transport. Pillars of the economy are tourism (spa...), production of metals and metal products. Nitra region (NR), where the natural increase is negative in the long term. The region has a significant industrial sector and the utilized agricultural area. Prešov Region (PO) with the largest population is well known as a tourism industry. Trenčín Region (TN) is small region with uneven population. The natural decreases is accompanied by an aging population. Nature of the economy is industrial and agricultural region with important rail routes. Trnava Region (TT) is the second smallest region. The development of natural increase rate is negative as well. On the other hand, Trnava region is one of the most productive agricultural region, agricultural land takes up 69,7 %.

V. CONCLUSION AND DISCUSSION

The existence of disparities between regions is indisputable. The greatest differences are not only in regional GDP but also in spatially differentiation. The most concentrated is the sector in the capital city Bratislava, the least in regions of Prešov and Trenčín. Surprisingly, the disparities in the regions in the middle of the order are not so significant. Cluster analysis confirmed main regions, where the ICT sector should be supported the most, Fig. 2. The advantage is not only the location of these regions, but also the existence of the largest slovak universities, educating ICT professionals.



Fig. 2. ICT spatial clusters

Quantification of relative values showed, that disparities are worthy of attention. Even when Slovak Republic has got quite good position in ICT rates, finding showed, that it is important to blur these regional disparities. As findings showed, the nature of the ICT sector allows to start a business wherever, what can be also seen at the relatively uniform deployment of ICT firm in individual regions. The sector is not tied to physical location. With sensible intervention and selective location support, the state could achieve much faster removal of the regional disparities.

It is difficult to answer the question, how quickly will the Slovak ICT sector progress. It depends mainly upon what direction will develop conditions, particularly in term of

quality of business environment and the level of skilled labor. The aim should be to promote the education in the sector and improve conditions in the business environment. In some issues of environmental quality, such as last enforcement, is the Slovak Republic in the last ten rankings nearly 150 countries, including several developing countries. For all these reasons mentioned above, authors think, that such a large disparities in ICT sector are the problem. As mentioned before, these disparities “inflated” strong concentration large, international enterprises in capital city, but even apart from this fact, it is not appropriate, when one region concentrate half amount of firms and more than 60 % of employees in the ICT sector. The sector do not expect active financial support, only the better business conditions and better quality in education system in the Slovak Republic in the area of the ICT sector, reduce the administrative burden (especially e-government), reduced rate of income tax burden at least for start-ups, or create a new simple legal forms of business. The finding confirmed the importance of the ICT sector in the Slovak Republic as well as that the ICT sector is the driving force of the Slovak economy and boots economic development. All these are reasons why to pay attention to regional disparities in ICT sector in the Slovak Republic. The automotive industry has not been mentioned in introduction accidentally. Not only the ICT sector together with the automobile industry are the drivers of the Slovak economy, but they are connected to each other. Further research will therefore focus on automotive industry clusters and subsequent linking of the ICT and automotive industry in the Slovak republic.

REFERENCES

- [1] I. Rohman, “On the weightless economy: Evaluating ICT sectors in the European, Asian and African regions”, [Dissertation]. Chalmers University of Technology, Sweden, 2012
- [2] A. Maddison, “Dynamic forces in capitalist development: A long-run comparative view”, Oxford, New York: Oxford University Press, 1991
- [3] Romer, P.M. “Increasing return and long run growth”, *The Journal of Political Economy*, 94(5), 1002–1037. 1986
- [4] P.M. Romer, “Endogenous technological change”, *The Journal of Political Economy*, 98(5), 71–102 Y. 1990
- [5] M. Buček, Š. Reháč and J. Tvrdoň, “Regionálna ekonómia a politika”, *Regional economics and policy* 1. vyd. Bratislava: Iura Edition, 2010. 269 s. ISBN 978-80-8078-362-4.
- [6] I. K. Rohman, “The globalization and stagnation of the ICT sectors in European countries: An input-output analysis”, *Telecommunications Policy*, Volume 37, Issues 4–5, May–June 2013, Pages 387–399, ISSN 0308-5961, <http://dx.doi.org/10.1016/j.telpol.2012.05.001>.
- [7] I. Holúbek, et al. “Exploitation of quantitative methods for the assessment of regional performance of the Slovak economy”, [online]. [cit. 2015-01-27]. Dostupné na internete: <<http://www.sciencedirect.com/science/article/pii/S1877042813055043>>.
- [8] WEF, “Index of Country/Economy profiles. In *The Global Information Technology Report 2014 – Rewards and Risks of Big Data*”. [online]. 2014. Ženeva: World Economic Forum, 2014. [cit. 2015-02-11]. Dostupné na internete: <<http://www.weforum.org/reports/global-information-technology-report-2014>>. ISBN 978-92-95044-63-0.
- [9] OECD. “Glossary of statistical terms: Information, communication technology” (ICT) sector. [online]. [cit. 2015-01-20]. Dostupné na internete: <<http://stats.oecd.org/glossary/detail.asp?ID=3038>>.