WSB University in Wroclaw
Research Journal
ISSN 1643-7772 I eISSN 2392-1153
Vol. 16 I No. 3

Zeszyty Naukowe Wyższej Szkoły
Bankowej we Wrocławiu
ISSN 1643-7772 I eISSN 2392-1153
R. 16 I Nr 3

Wyższa Szkoła Bankowa
we Wrocławiu

WSB UNIVERSITY
RESEARCH
JOURNAL

Application of Selected Methods of Grouping Data in the Study on Spatial Differentiation of Socio-Economic Phenomena on the Example of Outward Mobility in Poland in 1989-2002

Author: Katarzyna Maruszewska

Abstract

The aim of the paper is to demonstrate the usefulness of some methods of grouping data in a spatial analysis of socio-economic phenomena on the example of the analysis of temporary emigration from Poland in the years 1989-2002. The study aimed at assessing spatial differentiation at a high level of detail, i.e. accounting for local differences. For temporary emigration, this kind of assessment was made possible by the results gathered by the National Census of Population in 2002.

The paper shows examples of analyses of mobility territorial differentiation, where selected optimization-iterative and hierarchical methods of grouping were applied. The elements in the grouping space were units of territorial division covering the following breakdowns: gmina (municipality), poviat (district) and sub-region.

During the socio-economic transition period following the year 1989 temporary emigration in Poland was a phenomenon which varied considerably in regional and local dimension, as has been confirmed by the findings of the studies conducted. A pronounced spatial dispersion was found not only in terms of its level and intensity, but also in terms of diversification of the demographic and social structure of the phenomenon. The application of the adopted methods of data classification made it possible to assess in detail the spatial differentiation of temporary emigration in Poland according to the breakdowns covering the minor, yet numerous, territorial units, as well as to visualize the results using map charts.

The temporal scope of the study was limited to the years 1989-2002 because of the nature of the source material used and a lack of comparable statistical data on the post-accession and contemporary emigration.

Keywords: spatial differentiation, cluster analysis, international migration, emigration
JEL: F22, C38, C49

History: Otrzymano 2015-11-29, poprawiono 2016-06-08, zaakceptowano 2016-06-10

Katarzyna Maruszewska
Wyższa Szkoła Bankowa w Poznaniu
katarzyna.maruszewska@wsb.poznan.pl

Introduction

Gaining insight into the spatial differentiation of economic phenomena is necessary, for it allows explaining their determinants. Moreover, this knowledge can be applied while shaping economic and social policy as well as regional policies, e.g. while implementing measures designed to improve the efficiency of local labour markets or to better regional development.

In Poland, the phenomenon which is particularly relevant not only in the socio-economic but also in the historical context is emigration. Poland is one of a few European countries where a prevalence of emigration over immigration has been observed over decades, with Poles occupying an important place in the international flow of labour force. The migration traditions stem from the country's past, the partitioning of its territory between three neighbouring powers, and, following the country's independence, from the different paths towards socio-economic development of individual regions. Owing to the specificity of Polish migration processes, this phenomenon has existed with varied intensity across the nation's diverse regions for a number of decades. After 1989, different pasts and differentiated conditions of development have shaped the migration communities participating in the various forms of migration activity, differing in terms of their demographic, social and economic characteristics. Poland's accession to the European Union in 2004, which made foreign labour markets gradually accessible, produced a spectacular increase in the mobility scale. The short-term and partial migration, which was prevalent before the accession, was in many instances replaced by medium and long-term migration. Those shifts were accompanied by the changes in regional distribution of the Polish migration activity.

The aim of the paper is to demonstrate the application possibilities of selected methods of data clustering as a tool for the assessment of the spatial differentiation of demographic, social and economic phenomena. Temporary emigration exemplifies such a phenomenon. The paper presents partial findings of the study on Polish temporary emigration – the analysis was concerned with departures abroad for temporary stay between 1989-2002. The results of the National Census of Population from 2002 (NCP 2002) provided the data source on the emigration population.

The paper shows the results of the territorial differentiation analysis using various methods and data breakdowns on the basis of several examples. The level of the phenomenon was examined as well as its intensity, while taking into consideration selected demographic characteristics of migrants: their residential environment and education structure. The temporal scope of the study covers the years 1989 -2002, i.e. from the beginning of the political system transformation and opening of borders for migration until the census study. Thus, the most recent migration flow – after Poland's accession to the European Union – has not been included in the analysis. The choice of the temporal scope was determined by the availability of data – the type of migration under discussion was subject to extensive investigation during the national census in 2002. The study was carried out using a complete method and covered the departures between 1989 and 2002, including the unregistered ones, which clearly dominated. Moreover, the in-depth studies on international migration carried out during the NCP of 2002 were the reason as to why the analysis was undertaken, as the availability of data made it possible to assess the migration diversification with a high level of detail (at a local level).

The results of the national census from 2011 do not provide information within such scope because of a different methodology employed in the study. Furthermore, the data coming from the current records do not constitute a reliable source of data on factual temporary emigration, for people who leave the country do not usually register their departure to stay temporary abroad.

Research Methods

The assessment of the territorial units similarity was carried out using a cluster analysis – both hierarchical (agglomerative) and non-hierarchical methods (iterative). The choice of techniques was conditioned by the analysis dimension and was preceded by a preliminary assessment and the comparison of the results obtained through the application of different grouping procedures. While examining the structural similarities, the results of the hierarchical analysis were applied for further analyses (such as determining the number of classes) and the verification of the adopted division using a non-hierarchical approach.

Grouping territorial units (objects) on the basis of emigrants' education structure was done using a complete linkages clustering and the classic k-means method. The idea behind the k-means method was already developed in the 1950s, being then extended to cover multidimensional cases in the 1960s. In crude terms, the method consists in partitioning a set of data according to the principle of maximizing the within-group variance, while at the same time minimizing the variance within groups.

A generalised k-means method implemented in STATISTICA Data Miner software was employed in the study on the differentiation of the overall level and intensity of emigration and emigration from cities. This method belongs to indirected methods of data mining. The adopted approach involves searching for a model which fits best the data, regardless of what mechanism generated those data (compare Sokołowski, Pasztyła 2004, p. 92). The procedures applied in a generalised cluster analysis are similar to the k-means method used in a typical cluster analysis. The algorithm of the generalised method makes it possible to define the subsets on the basis of the value of one variable or a set of quantitative variables describing objects, and additionally it offers the possibility to include qualitative variables in the analysis.

In order to determine the optimum number of groups for the analysis using the generalised method, v-fold cross validation was deployed. This validation involves partitioning the entire data set into v subsets. The same analysis is then applied for the observations from v-1 sets (training sets), and the resulting model is applied to subset v (the subset which was not used to determine clusters, i.e. testing set); also, the precision of the current result is determined. The results of subsequent v rounds are averaged, giving one measure of the stability of the model. The resulting number of clusters is determined on the basis of changes of the error function (average distance from the cluster centres). In the analysis, the minimum decrease of the error function is assumed at the level of 5%, at which the next result is calculated.

The agglomerative procedure used in the analysis was the complete linkage clustering, also known as the farthest neighbour clustering. Numerous research findings have confirmed this method to be one of the most effective grouping methods (e.g. Młodak, 2006). The method leads to having objects merged into groups, thus creating a tree like structure.

As a method for measuring the objects similarities, the so called city block distance measure (Manhattan, Hamming) was adopted. The city block distance between two points is measured along the measurements defined by the variables. This kind of metric was used in the analysis, for it assigns a lesser weight to the value differences between individual variables in objects than e.g. Euklidean distance, thus allowing for the influence of individual differences (unusual observations) to be suppressed.

Clustering methods are used predominantly when the researcher is at an exploration stage, having no a priori hypotheses at hand – therefore testing statistical significance in the classic sense of that term finds no application here. The analysis is used mainly as a descriptive technique and the result obtained (partitioning into clusters) is not unequivocal, being conditional on the research techniques adopted (Peczkowski, 2009, p. 139). In statistical practice, in order to assess the classification accuracy, the variation within clusters (low if the classification is sound) is compared with the variation between clusters (high if the classification is sound). For the purpose of the study, suitable parametric or non-parametric procedures were applied as a supportive measure.

Spatial Differentiation of Temporary Emigration Intensity

The level of temporary emigration, i.e. the absolute level of outflow from a given area has been adopted as one of the basic categories describing the mobility of the Polish people. As the preliminary assessment of the country's distribution of temporary emigration has showed, in half of the gminas the number of people who emigrated did not exceed 91 persons, whereas the median absolute deviation was equal to ±65 persons (71,43%). What particularly influenced the clearly differentiated level of the phenomenon were the varied levels of migration between gminas with a low total population number and between highly urbanised areas from which lots of people migrated. Moreover, in such cities as Warsaw, Cracow, Wrocław, Białystok, Zabrze, Gdańsk and Bytom the scale of the phenomenon differed so significantly from the remaining areas that they had to be excluded from the analysis (outliers) in order to improve the grouping effectiveness.

The classification resulted in dividing gminas into six classes corresponding to the outcomes of grouping (Fig. 1 and Table 1). Since the analysis was concerned with one-dimensional population, the grouping was performed using the generalized k-means method. The findings revealed that nearly 1/3 of gminas (768) were characterized by a very low level of emigration, that is not exceeding 48 persons (cluster 1). Those gminas were classified into class I – the most numerous, while showing a considerable degree of internal differentiation (45%). The country's central and north-west regions were characterised by the lowest level of temporary emigration, with the exception of voivodship capitals (class I) – including mainly the gminas located in: Wielkopolskie, Łódzkie and Mazowieckie voivodships; there were only 6 gminas where there were no people registered as temporary emigrants (Chodów, Goszczyn, Góra Św. Małgorzaty, Parysów, Pniewy, Powidz, Radzanów k. Radomia).

The remaining groups were characterised by a low within-variation (coefficients below 31%). Drawing conclusions based on the distances between the cluster centres, the gminas classified as class II (751)

appeared to be considerably similar to gminas in class I in terms of the number of persons who emigrated. Class II comprised most of the gminas located within the borders of Świętokrzyskie voivodship, but also a number of gminas in Lubelskie and Lubuskie voivodships, especially in the area bordering with Wielkopolskie and Zachodniopomorskie voivodships. Classes I and II in total vaccounted for over 60% of gminas, which implies that the number of emigrants did not exceed 135 persons in the majority of the units under discussion.

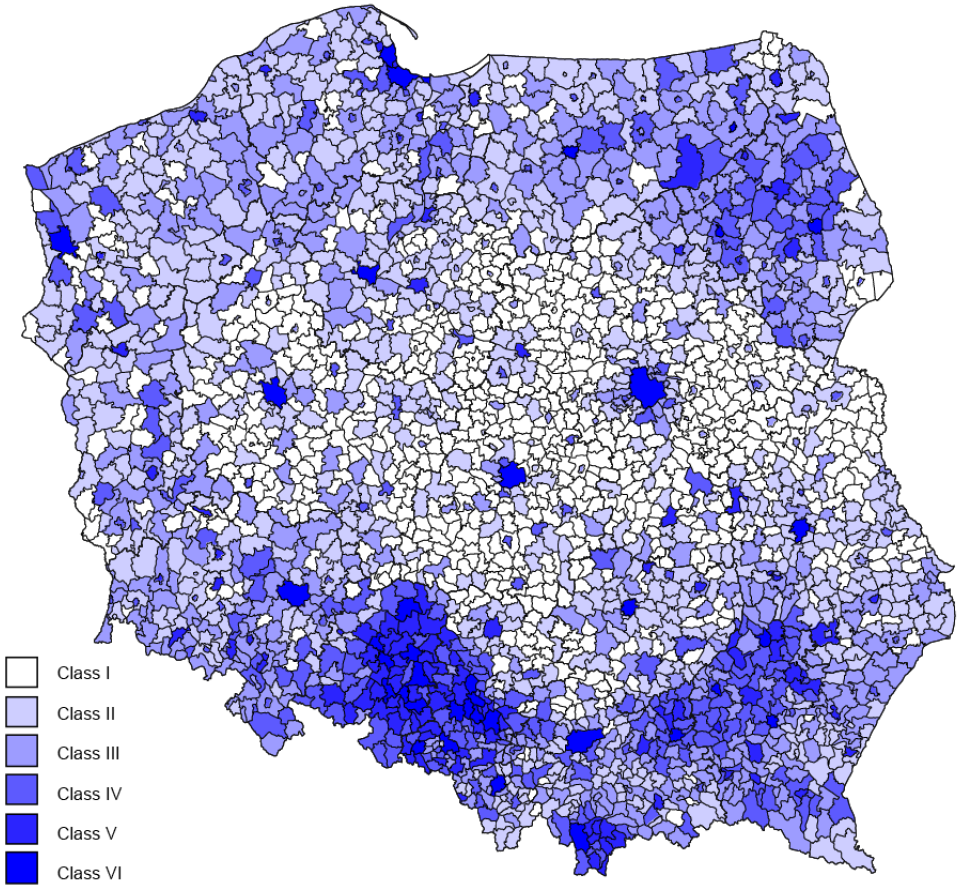

Fig. 1. Spatial differentiation of temporary emigration level in gminas across Poland in 2002.

Source: Author's own study.

Table 1. Grouping of gminas according to temporary migration level

Classes	Gminas (*N*=2471)		Average number of emigrants	V_x	H Kruskal-Wallis (*p*-value)
	n_i	%		%	
Class I	768	31,08	26,60	45,75	2297,653 (0,000)
Class II	751	30,39	84,04	29,17	
Class III	538	21,77	226,93	29,12	
Class IV	268	10,85	609,28	30,97	
Class V	116	4,69	1703,31	25,08	
Class VI*	30	1,21	4243,43	24,24	

* excluding outliers

Source: Author's own calculations.

A moderate level of emigration (136-377) was observed in the country's border regions, in particular, in gminas of Dolnośląskie and Warmińsko-Mazurskie voivodships, as well as in many gminas of Zachodniopomorskie, Pomorskie, Lubuskie and Świętokrzyskie voivodships (class III). A higher level (379-1098) was observed mainly in the areas located within the borders of: Podkarpackie, Podlaskie and Małopolskie voivodships (class IV). A high level of temporary outflow at national level (1140-2858) was noted in 116 gminas (class V). The greatest number of Poles emigrating included those living in Górny Śląsk (Opolskie voivodship and central and southern part of Śląskie voivodship), and also in Małopolskie, Podkarpackie, Warmińsko-mazurskie and Podlaskie voivodships.

The level of the phenomenon deviating from the values observed in the remaining areas (2931 and more) was recorded for 37 gminas of class VI. This class included gminas with the farthest clustering 6 and the outliers which were disregarded in the grouping. It was found that by far the highest levels of migration occurred in highly urbanized areas, that is in cities and metropolitan areas, among which Warsaw dominated (26,7 thousand of emigrants), followed by Cracow (13,8 thousand), Wrocław (11,4 thousand), Białystok (10,6 thousand), Zabrze (10,4 thousand); high values (over 5,0 thousand) were observed for the following cities: Gdańsk, Bytom, Łódź, Opole, Kędzierzyn-Koźle, Szczecin, Tarnów, Lublin and Gdynia.

The results verification which was performed using Kruskal-Wallis test confirmed the emigration level differentiation to be statistically significant (H=2297,653; p=0,000) in the generated groups of gminas.

The emigration level, although providing important data on the numbers of migrant communities, does not provide information about a relative intensity of the phenomenon. Hence, in further parts of the analysis, relative indicators have been applied, relating the migration scale to the population size of the individual gminas. The analysis of the temporary emigration intensity (per thousand population) found that there were compact areas with differentiated levels of the migration activity (Fig, 2 and Table 2). As the distribution basic characteristics show, in half of the spatial units under study the intensity indicator did not exceed 11,77 with the median absolute deviation of ±6,66 (56,57%).

The grouping yielded 7 clusters of gminas whose internal differentiation

was relatively low (not over 32,27%). Not unlike for the emigration level, the lowest intensity of the phenomenon (0,00-4,53) was observed in the gminas located mainly in the country's central regions (class I), including the following voivodships: Wielkopolskie, Łódzkie and Mazowieckie, as well as in the several gminas situated within the borders of Lubelskie and Świętokrzyskie voivodships – in total, 415 units (16,75%) were classified into class I.

The next classes II, III and IV proved to be considerably similar in terms of the emigration intensity, which was confirmed by small distances from the cluster centers, and by numbers (21-23% of all gminas per cluster) and their internal diversification (15,33%-23,30%). In those classes, gminas with a moderate intensity were classified (between 4,55 and 28,89 emigrants per thousand inhabitants), located primarily in the regions along the country's western border, in Zachodniopomorskie and Lubuskie voivodships, as well as along the eastern border, that is the area of Lubelskie voivodship and in numerous gminas of Świętokrzyskie and Kujawsko-pomorskie voivodships.

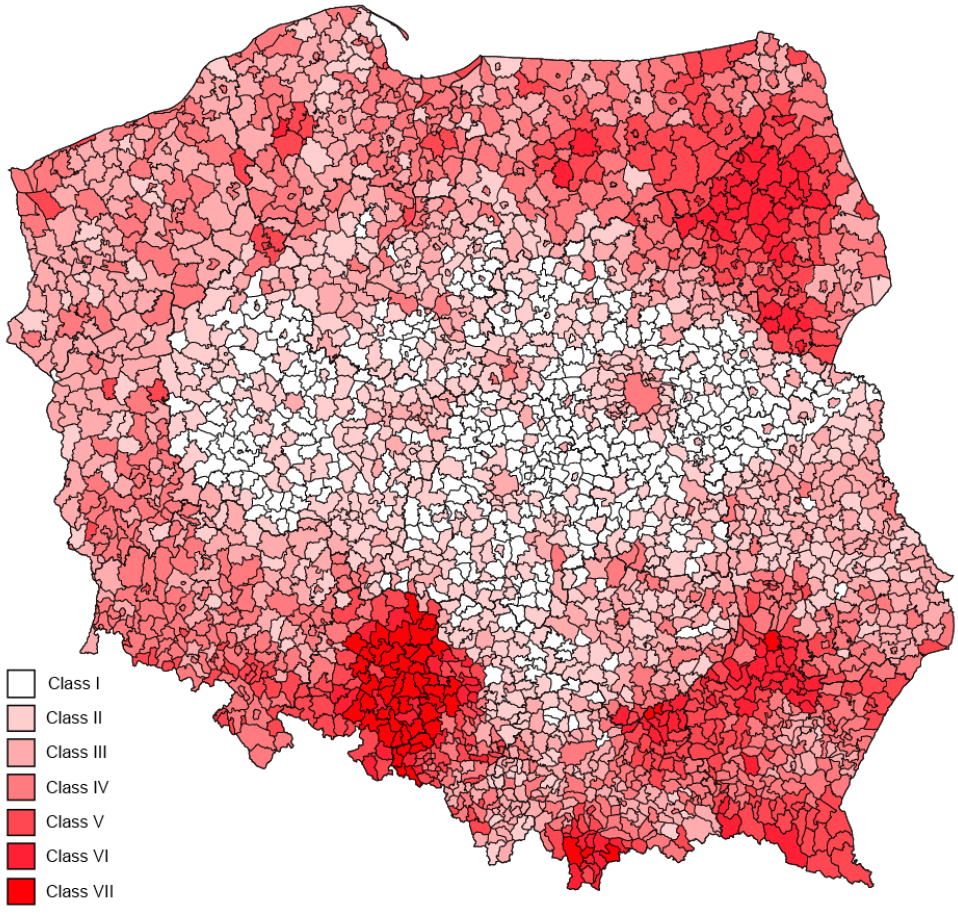

Fig. 2 Spatial differentiation of temporary emigration intensity in gminas across Poland in 2002

Source: Author's own study

Table 2. Grouping of gminas according to temporary emigration intensity (per thousand inhabitants)

Classes	Gminas (*N*=2478)		mean indicator of emigration intensity	V_x	H Krus-kal-Wallis (*p*-value)
	n_i	%		%	
Class I	415	16,75	3,07	32,27	2381,476 (0,000)
Class II	532	21,47	6,51	18,20	
Class III	574	23,16	11,78	15,33	
Class IV	539	21,75	20,38	18,03	
Class V	261	10,53	39,77	19,99	
Class VI	104	4,20	78,99	23,30	
Class VII	53	2,14	181,03	17,95	

Source: Author's own calculation

A high migration activity (23,93-126,59 per thousand inhabitants) featured among the country's characteristic regions, where the high absolute level of the phenomenon was also observed – these are compact clusters consisting of gminas (class V and VI) located predominantly in Podlaskie voivodship, but also in Podkarpackie and in the western part of Warmińsko-mazurskie voivodship and in the south of Dolny Śląsk, as well as in a number of gminas of Śląskie and Małopolskie voivodships. A special region on the spatial differentiation map of the country's emigration is Śląsk Opolski –nearly all gminas in this region were classified as class VII, i.e. a cluster consisting of 53 gminas (2,14%) experiencing extremely intense migration (132,13-253,01 per one thousand inhabitants).

It was shown that the clusters obtained through grouping differ significantly from one another (H=2381,476; p=0,000). Analyzing the differences between the groups, it can be observed that there is a pronounced distance between the centres of clusters 6 and 7 and the rest of the groups, which allows for conclusion that the regions marked by the gminas included in class VI and VII are atypical regions, within the scope of the variable, against the background of other regions – i.e. the regions with a very high migration activity. Such strong activity induces one to advance the hypothesis that there exist specific determinants for migration in those regions.

Based on the findings of the spatial analysis, one could conclude that, in local dimension, (according to the gminas breakdown) both the level and intensity of the emigration in Poland show features characteristic of a non-uniform and highly differentiated phenomenon. Not only the regions, but also individual gminas were observed as going through extremely intensive emigration. The analysis made it possible to distinguish five regions with a very high intensity of migration:

- Górny Śląsk – the highest emigration activity was characteristic for the following gminas:

Komprachcice, Kolonowskie, Cisek, Popielów, Murów, Łubiany, Żębowice, Walce, Jemielnica, Tarnów Opolski, Krzanowice;
- Podlasie – with the dominance of the following gminas: Stawiski, Radziłów, Mońki, Trzcianne;
- Podkarpacie – mainly the area of Lesk and Sanok, and numerous gminas at the right bank of the Vistula River, between the Dunajec River and the San River (including, Dzikowiec, Ulanów, Bojanów, Jarocin k. Tarnobrzegu, Pysznica);
- Małopolska – the Tatra Mountains region and the Dunajec River mouth (gminas: Bukowina Tatrzańska, Czarny Dunajec, Nowy Targ);
- Mazury (Masuria), including predominantly the area around Olsztyn.

Apart from the regions mentioned above, individual gminas, or small clusters of migration gminas dispersed across the country were identified, where a high migration activity occurred locally. Such clusters were located in, e.g: Pomorze – (Tuchomie, Bytów, Lipnica, Studzienice), in northern Wielkopolska (gmina and the town of Złotów, Zakrzewo), in Zachodniopomorskie voivodship (Rewal and Stępnica), and also in Lubuskie voivodship (Lubniewice and Pszczew).

Migration Intensity in Rural Areas

The studies conducted in the country's areas typically regarded as migration regions show that since the early 1970s most people migrating abroad came from the cities, mainly from Warsaw (see more in: Frejka et al., 1998; Okólski 2001, pp. 50-53). Wealthy and educated population usually concentrate in the cities, and is more willing to take the risk of embarking on the "unknown". Yet, already in the early 1970' the travellers from big cities were joined by the pioneers of international migration who lived in the back country. As the result, the proportions between the population from highly urbanized and industrialized regions, and that from the peripheral regions started to shift gradually. In the early 1990', one could observe a relative predominance of migration amongst the population from the rural areas.

In the context of those shifts, Marek Okólski, using the elements of Zelinsky's concept of spatial mobility transition[1] , put forward a hypothesis on the so called incomplete migration (unfulfilled exodus from rural to urban areas). The changes in the economic strategy in the late 1970', including, among other things, a departure from the idea, which was at the time advocated, of heavy industry, caused a drop in demand for work. This had the effect that the population functioning in the space between the city and village, and thus having some sort of experience in mobility, began to search abroad for new sources of income. The mobility of people from small towns and rural areas was taken over by international labour circulation, in this way generating the so called partial migration, which usually lasted less than two years, showing no clear intention of a long-term migration (Okólski, 2001, p. 52-53). The traditions of migration from rural areas of the industrialized regions became permanently established over subsequent years. A detailed assessment of the scale and differentiation of the migration from rural areas is done in further stages of the study. It was carried out according to the breakdown covering poviats.

Grouping poviats exclusively on the basis of the temporary emigration

[1] The concept of spatial mobility transition W. Zelinsky outlines in: *The hypothesis of the mobility transition*, in: „The Geographical Review", No. 61/1971, pp. 222-245.

intensity in rural areas did not yield satisfactory results, for the generated clusters proved to be non-uniform (Vx>100%). Therefore an attempt was made to include an additional qualitative variable in the analysis which would identify the type of the poviat (rural or a township) – while assuming that this variable evinces grouping properties in the set under discussion. As the result, an optimized partition into seven clusters was obtained, which were statistically significant (H=364,693; p=0,000) and internally uniform.

The analysis revealed that after 1989 the intensity of the emigration from the rural areas was considerably lower than from the cities. On the other hand, the differentiation of the phenomenon appeared slightly higher in the rural areas. In half of the poviats, the emigration intensity from the cities was 17,44 or less (per thousand urban inhabitants) with the median deviation of ±7,71 (44,23%), while for the rural areas, in half of the poviats the temporary emigration intensity did not exceed 8,61 ±5,82 (67,59%). The results of the poviat classification based on the migration activity in rural areas, taking into account the variable indicating the "type of the poviat", are illustrated in Table 3.

In the process of grouping poviats, a cluster was produced which was significantly different from other groups within the scope of the qualitative feature (χ =378,000; p=0,000) describing the poviat type (class I), in which only township poviats were classified (66, i.e. 17,41%). Low indicators of the migration from rural areas (1,43-8,59, class II and III) were obtained for the poviats located in the central part of Poland (Fig. 3): Wielkopolskie, Mazowieckie, Łodzkie voivodships and, e.g. in the southern part of Kujawsko-pomorskie voivodship, the northern part of Świętokrzyskie and Lubelskie voivodships, and also in some poviats of Opolskie and Śląskie voivodships (Gliwicki, Będziński, Kłobucki, Częstochowski).

A moderate intensity (8,63-40,67, i.e. classes IV and V) was recorded mainly in the poviats of the country's outermost voivodships, along Poland's western border – particularly in the region of Zachodniopomorskie, Lubuskie and Dolnośląskie voivodships, as well as Pomorskie and Warmińsko-mazurskie, in the eastern part of Podlaskie voivodship and with several poviats located in the central region (including the following: Średzki wielkopolski, Nowodworski mazowiecki, Tomaszowski mazowiecki, Ostrołęcki, Makowski). A relatively high migration intensity (41,73-93,80, i.e. class VI) was characteristic for poviats in the western part of Podlaskie voivodship (e.g. Łomżyński, Grajewski, Moniecki, Kolneński, Zambrowski, Siemiatycki), Podkarpackie voivodship (e.g. Leski, Kolbuszowski, Mielecki, Leżajski, Nizański, Stalowowolski), Małopolskie voivodship (Tatrzański and Nowatorski) and Opolskie and Śląskie voivodships ((Nyski, Głubczycki, Wodzisławski).

Into class VII – with the highest indicators of temporary emigration from rural areas, 11 poviats were classified; they were located predominantly close to highly industrialized regions (the area of Górnośląski Okręg Przemysłowy (Upper Silesia's Industrial District). These include poviats from cluster VII (111,33-197,37) – situated in Opolskie voivodship (Kluczborski, Oleski, Strzelecki Opolski, Krapkowicki, Kędzierzyńsko-kozielski, Prudnicki, Raciborski), Lubelskie voivodship (Opolski lubelski), and Śląskie voivodship (Raciborski, Lubliniecki), and also two Śląskie poviats with extremely strong migration activity of the rural areas population (outliers): Mikołowski (346,01) and Bieruńsko-Lędziński (423,11) poviats.

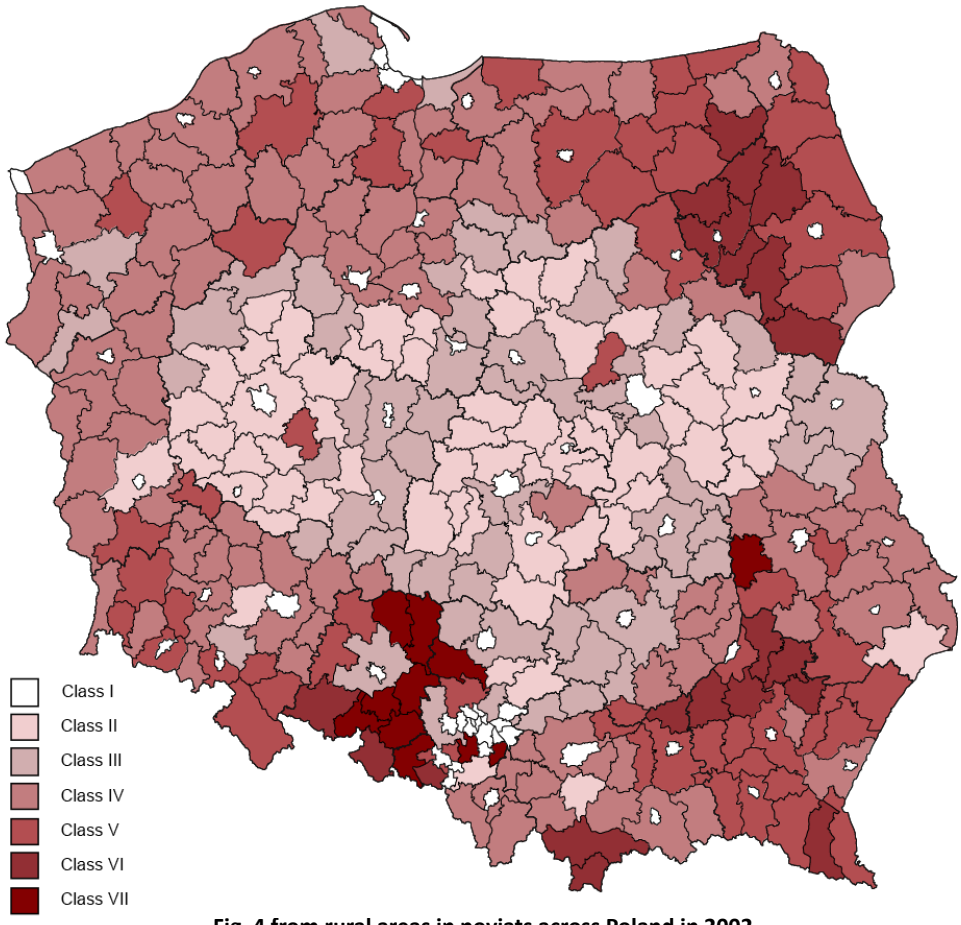

Fig. 4 from rural areas in poviats across Poland in 2002

Source: Author's own study.

Table 3. Grouping of poviats by emigration intensity from rural areas (per thousand) and by poviat type

- **statistics for the quantitative variable: emigration intensity from rural areas**

Classes	Poviats (*N*=378)		mean indicator of emigration intensity	V_x	H Kruskal-Wallis (*p*-value)
	n_i	%		%	
Class I	66	17,46	-	-	364,693 (0,000)
Class II	61	16,14	3,35	19,25	
Class III	63	16,67	6,24	18,97	
Class IV	103	27,25	12,39	21,23	
Class V	56	14,81	26,91	26,77	
Class VI	20	5,29	63,37	26,42	
Class VII*	9	2,38	156,36	23,88	

- **statistics for categorised variable: poviat type**

Classes	categorised centres of clusters	Poviats (*N*=378)		χ^2 (*p*-value)
		township n_i	rural n_i	
Class I	towship	66	-	378,000 (0,000)
Class II	rural	-	61	
Class III	rural	-	63	
Class IV	rural	-	103	
Class V	rural	-	56	
Class VI	rural	-	20	
Class VII*	rural	-	9	

* excluding outliers

Source: Author's own calculations.

Classification of Subregions Based on Emigrants' Education Structure

Education is one of the basic markers of the position of an individual in the society; it determines to a large extent the chances of being successful economically and professionally, facilitating or hindering one's functioning in the labour market. From a macro-economic point of view, this characteristic decides about human capital of a given country, and in particular about the quality of labour resources. Intensive migration phenomena can affect the education structure of a country's population, and, depending on the trends observed – in the case of outflow of well educated, or specialized in specific fields, population - they can lead to local shortages; the loss of intellectual potential and deficiencies in specific qualifications. This characteristic is extremely relevant not only from the perspective of the emigration consequences, but also in the context of its causes. In light of migration theories explaining the reasons behind migration, we can assume that the education level, besides other socio-demographic characteristics, has an impact on the migration decision by individuals. Having professional skills that are searched for abroad and foreign language skills may constitute factors which induce one to emigrate, especially when this is accompanied by considerable differences in wages between the country of origin and that of the planned migration.

In the assessment of the subregions differentiation in terms of the migration activity according to education, the following categories of the characteristic were examined: higher education, post-secondary education, general secondary education, secondary vocational, basic vocational, completed primary education and incomplete primary education, and a lack of school education. Accounting for the methodology of the study on emigrants' education during the NCP 2002, the data on the population aged 13 and above were used in the analysis. The study was carried out according to the breakdown covering subregions.

The dendrogram of grouping, obtained as the result of the agglomerative grouping using a complete linkage clustering (Fig. 4) indicates that there is a clear differentiation between

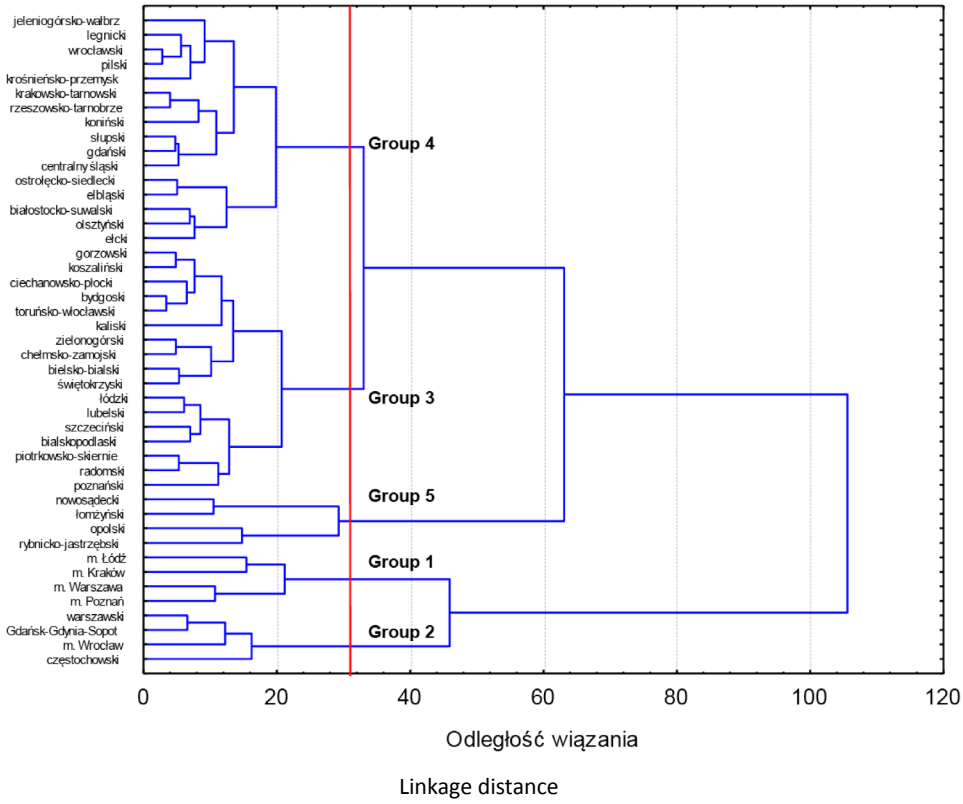

Fig. 4. Dendrogram of grouping subregions according to emigrants' education structure – complete linkage clustering with city **block** distance (Manhattan)

Source: Author's own study

Table 4. Subregion classification groups according to emigrants' education structure (percent)

Education	Group 1	Group 2	Group 3	Group 4	Group 5	F (*p*-value)
	mean (V_x)	mean (V_x)	mean (V_x)	mean (V_x)	mean (V_x)	
higher	37,6 (11,8)	27,2 (12,5)	16,8 (10,7)	11,7 (13,5)	6,7 (11,5)	154,268 (0,000)
post-secondary	6,4 (12,9)	6,2 (14,3)	5,4 (15,0)	4,8 (12,2)	3,5 (15,6)	12,433 (0,000)
general secondary	21,6 (9,3)	16,9 (9,7)	15,7 (8,1)	12,8 (8,3)	9,4 (17,0)	52,934 (0,000)
vocational secondary	17,8 (10,6)	23,7 (7,4)	23,9 (7,6)	25,5 (5,4)	20,8 (11,5)	17,152 (0,000)

basic vocational	8,2 (28,6)	15,7 (19,3)	24,8 (7,5)	28,5 (9,7)	35,6 (19,2)	53,621 (0,000)
completed primary	7,9 (13,8)	9,5 (14,5)	12,4 (10,8)	15,7 (13,5)	22,7 (14,8)	41,430 (0,000)
incomplete primary and lack of education	0,6 (20,8)	0,7 (21,5)	1,0 (29,4)	1,0 (25,8)	1,3 (12,6)	5,057 (0,002)

Source: Author's own calculation.

Tab. 5. Euclidean distances of classification groups

Groups	Group 1	Group 2	Group 3	Group 4	Group 5
Group 1	0,0	**32,1a**	**115,0**	**184,0**	**299,2**
Group 2	5,7	0,0	**28,8**	**66,5**	**151,8**
Group 3	10,7	5,4	0,0	**8,9**	**54,0**
Group 4	13,6	8,2	3,0	0,0	**22,7**
Group 5	17,3	12,3	7,4	4,8	0,0

* a marked in bold, denote the squares of Euclidean distances between groups

poviats in terms of emigrants' education structure. On the basis of the agglomeration tree (dendrogram), the distance 26 was assumed as the cutting point, thus producing five groups of subregions. Two of them: cluster 1 and 2 comprised four subregions each and in terms of the education structure they showed some similarities, while being very distinctive from other clusters (the linkage with them occurred at a distance of over 112) – these subregions are made up mainly of large urban centres and their agglomerations. Cluster 3 and 4 were also similar in terms of the structure and more numerous – they comprised: 15 and 18 subregions, respectively localized across the entire country. Moreover, cluster 5, comprising four subregions, showed its distinctiveness from all the remaining classes – resembling only slightly clusters 3 and 4. Regions with the highest migration intensity were included in cluster 5. The resemblance structure thus formed of the areas under study was nearly fully confirmed by the classification using the non-hierarchical k-means method – both through Euclidean distances and the composition of the individual clusters (table 6).

The analysis results showed that the following subregions: Łódź, Warsaw and Poznań (group 1), were characterized, against other units, by a clearly

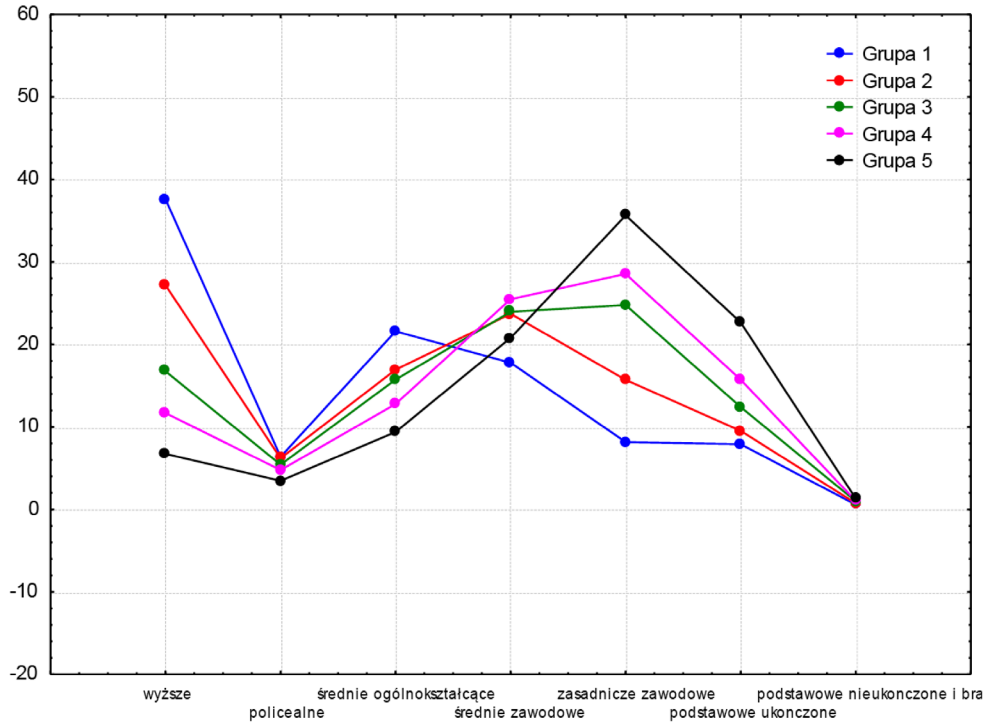

Fig. 5 Subregion classification groups according to emigrants' education structure (percent)

Source: Author's own study.

distinctive education structure of emigrants – these differences are demonstrated by the diagram of the centres of the individual clusters (Fig. 5), and they are confirmed by the Euclidean distances (Table 5). A clear majority, that is over 80% of people migrating from those regions, have skills at least at the level of secondary education (compare Table 4), including: higher education 37,61%, post-secondary education 6,38%, general secondary education 21,59% and secondary vocational 17,78%. At the same time, the percentage of emigrants coming from those subregions with basic vocational education was the lowest in the country (8,15%). Cities with such a characteristic education structure are part of the country's largest university centres and well developed business centres which are open to international cooperation.

Similar patterns were characteristic for the following subregions: Warszawski, Częstochowski, Cracow, Wrocław and Gdańsk-Gdynia-Sopot (group 2). In those units, nearly 75% of emigrants had at least secondary education; however, the percentage of emigrants with higher education was slightly lower, with 27,19% of emigrants holding university degrees, post-secondary – 6,20%, general secondary – 16,93%; correspondingly higher was the percentage of persons less educated, in

particular with secondary vocational (23,71%) and basic vocational education (15,70%).

The temporary emigration from the subregions included in group 1 and 2, characterized by a high education level, certainly results largely from the overall high education level of the population living in the largest urban areas. The emigration activity of the cities which evince traits typical for business, science or cultural centres must have had various bases. However, we should point out that the beginning of the first decade of the 21st saw the highest unemployment during the transition period, resulting not only from recession, but also from structural mismatches in the labour market. Young and highly qualified graduates were often unable to find work upon entering the country's labour market. People who lost their jobs found themselves in a similar situation. The alternative then was to look for employment abroad. The emigration was also fostered by easier access to some institutional initiatives, such as for example: employment services and recruitment agencies, the possibility of doing internship abroad or participation in academic international cooperation programmes, access to the Internet.

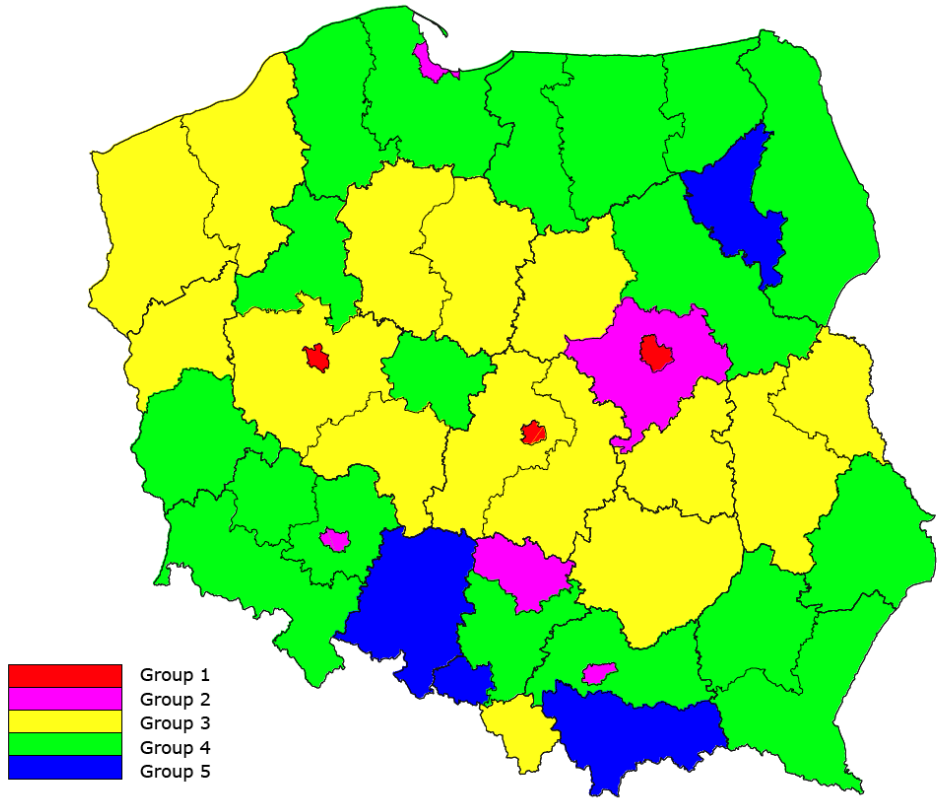

Fig. 6 Subregions classification according to emigrants' education structure

Source: Author's own study based on Tables 4 and 6

Table 6. Subregions classified according to emigrants' education structure

Complete linkage clustering (Manhattan)	***k***-means method
Group 1 (n = 4)	Group 1 (n = 3)
Kraków, Łódź, Warszawa, Poznań	Łódź, Warszawa, Poznań
Group 2 (n = 4)	Group 2 (n = 5)
Częstochowski, Gdańsk-Gdynia-Sopot, Warszawski, Wrocław	częstochowski, Gdańsk-Gdynia-Sopot, *Kraków*, Warszawski, Wrocław
Group 3 (n = 15)	Group 3 (n = 15)
Bialskopodlaski, Bielsko-bialski, Bydgoski, *Chełmsko-zamojski*, Ciechanowsko-płocki, Gorzowski, Kaliski, Koszaliński, Lubelski, Łódzki, Piotrkowsko-skierniewicki, Poznański, Radomski, Szczeciński, Świętokrzyski, Toruńsko-włocławski, *Zielonogórski*	Bialskopodlaski, Bielsko-bialski, Bydgoski, Ciechanowsko-płocki, Gorzowski, Kaliski, Koszaliński, Lubelski, Łódzki, Piotrkowsko-skierniewicki, Poznański, Radomski, Szczeciński, Świętokrzyski, Toruńsko-włocławski
Group 4 (n = 18)	Group 4 (n = 18)
Białostocko-olsztyński, Centralny śląski, Elbląski, Ełcki, Jeleniogórsko- gdański, Koniński, Krakowsko-tarnowski, Krośnieńsko-przemyski, Legnicki, Ostrołęcko-siedlecki, Pilski, Rzeszowsko-tarnobrzeski, Słupski, Suwalski, Wałbrzyski, Wrocławski	Białostocko-olsztyński, Centralny śląski, *Chełmsko-zamojski*, Elbląski, Ełcki, Jeleniogórsko- gdański, Koniński, Krakowsko-tarnowski, Krośnieńsko-przemyski, Legnicki, Ostrołęcko-siedlecki, Pilski, Rzeszowsko-tarnobrzeski, Słupski, Suwalski, Wałbrzyski, Wrocławski, Zielonogórski
Group 5 (n = 4)	Group 5 (n = 4)
Łomżyński, Nowosądecki, Opolski, Rybnicko-jastrzębski	Łomżyński, Nowosądecki, Opolski, Rybnicko-jastrzębski

Source: Author's own study.

Further groups: 3 and 4 covered the subregions with a structure typical for the majority of regions in the country (in total 33 subregions). In the north-western subregions: Gorzowski, Poznański, Szczeciński, Koszaliński, Bydgoski, Toruńsko-Włocławski, Bielsko-bialski, Świętokrzyski, Bialskopodlaski, Lubelski, and in the central subregions: Łódzki, Kaliski,

Piotrkowsko-Skierniewicki, Ciechanowsko-Płocki and Radomski (group 3) people with basic vocational education were prevalent – 24,77%, secondary vocational – 23,94%, with only 16,75% of emigrants having higher education. A similar structure, although showing a slightly lower education levels, was characteristic for subregions located in the south-western and southern part of the country: Zielonogórski, Jeleniogórsko-Wałbrzyski, Legnicki, Wrocławski, Krakowsko-Tarnowski, Centralny Śląski, Rzeszowsko-Tarnobrzeski, Krośnieńsko-Przemyski, Chełmsko-Zamojski, and in the northern and north-eastern part: Słupski, Gdański, Elbląski, Olsztyński, Ełcki, Ostrołęcko-Siedlecki, Białostocko-Suwalski, and also part of Wielkopolska: Pilski, Koniński (group4) – here over half of emigrants had vocational qualifications: basic – 28,53% or secondary vocational – 25,49%; a significantly lower percentage was observed for the population with university degree – 11,70% (Tables 4 and 6).

Special attention should be drawn to the education structure of emigrants from the following subregions: Nowosądecki, Opolski, Rybnicko-Jstrzębski and Łomżyński (group 5) – as many as 60% of people migrating from those areas did not have secondary education; most of them reported basic vocational education – 35,65%, followed by completed primary education – 22,66%; a little over 20% showed secondary vocational education. At the same time, in the subregions of cluster 5, the number of university graduates was the smallest – 6,74% (compare Fig. 5). As already shown, those are the areas which are very active in terms of migration nationally, with the outflow being observed especially in the rural areas. Low education level (or practically none) implies a low position on the labour market, which certainly was a factor supporting the job migration.

The classification was assessed using a variance analysis (Table 4) suggesting that all the education categories which were distinguished had a significant impact on the outcome of the process of clustering the subregions in terms of the similarity of the structure of the characteristic under discussion. The biggest contribution was obtained for the following category: higher and basic vocational education, while the least significant trait proved to be incomplete primary education or no education at all.

Summary.

The aim of the paper was to present the application possibilities of selected methods of grouping data as a tool for the assessment of the spatial differentiation of the phenomena based on the analysis of temporary emigration from Poland in 1989-2002. The study has sought to demonstrate findings at possibly low level of data agglomeration – the chosen breakdown covered gminas and poviats. The investigation of emigrants' education structure differentiation was carried out according to subregions, which was, on the one hand, dictated by the availability of the data, and on the other, provided the opportunity to demonstrate, by using a dendrogram, the similarities and differences (agglomerative distance hierarchy) between the regions in a way that is easy and clearly legible for interpretation, without having to resort to a map. However, it should be noted that the aggregation of data at the level of larger territorial units results in averaging the measurements of the phenomenon, and thus in lowering the spatial differentiation metrics.

The application of the selected techniques of the cluster analysis made it possible to carry out an effective grouping of territorial units, and thus the identification of regional patterns

through distinguishing the areas (a group of units) similar in terms of the properties examined. The spatial visualization of the results of the grouping was illustrated on the charts. Grouping data using an exploratory and generalised k-means method may constitute an alternative to other methods designed to determine the number of groups and to classify individual objects into them. This method appears to be a particularly effective procedure for the pronounced differentiation between one-dimensional populations, when the traditional partitioning based on mean values and standard deviations, applied in regional statistics, does not yield satisfactory results. The explorative approach enables one to include qualitative variables in the process of grouping, which can improve considerably the partitioning effects.

In literature there is a scarcity of works attempting to investigate the territorial differentiation of temporary emigration in Poland before and after the EU accession (e.g. Kaczmarczyk, 2008; Lusińska-Grabowska, Okólski, 2009). The available papers present the analysis according to the breakdown covering voivodships, or limiting the analysis to selected regions. However, there are no studies addressing this issue comprehensively and locally (NUTS-4 and NUTS-5). This paper, which examines temporary emigration in the years 1989-2002, i.e. from a historical perspective, provides a basis for this kind of research. This represents a huge challenge to scholars because temporary migrations are a phenomenon which is very difficult to capture in statistical sense[2], where movements are largely unregistered at the time of a free flow of people.

Bibliography

Cluster Analysis (2008), in: Statistica Electronic Manual – dokumentacja pakietu STATISTICA PL, Statsoft Inc., USA

Frejka T., Okólski M., Sword K. (1998), In-depth Studies on Migration in Central and Eastern Europe: The Case of Poland, United Nations, New York-Geneva

Kaczmarczyk P. (2005), Migracje zarobkowe Polaków w dobie przemian, Wyd. Uniwersytetu Warszawskiego, Warsaw

Kaczmarczyk P. (2008), Współczesne migracje zagraniczne Polaków. Aspekty lokalne i regionalne, Ośrodek Badań nad Migracjami UW, Warsaw

Kaczmarczyk P., Łukowski W. (ed.) (2004), Polscy pracownicy na rynku pracy Unii Europejskiej, Scholar, Warsaw

[2] Temporary emigration data gathered by the Central Statistical Office (GUS) refer to the registration of one's departure abroad for a temporary stay; those statistics cover only a small portion of actual migration; researchers believe that the best source of migration data is national census. Because of having abandoned traditional census method and replacing them with a representative method while investigating emigration, the results of the National Census of Population 2011 cannot be used in the analyses of spatial differentiation at the gmina level. Estimates of the temporary emigration in 2011 are made available by GUS at the level of voivodships (NUTS-2).

Lusińska-Grabowska I., Okólski M. (2009), Emigracja ostatnia?, Scholar, Warszawa

Marek E. (2001), Zatrudnienie Polaków za granicą. Praca cudzoziemców w Polsce, IPiSS, Warszawa

Młodak A. (2006), Analiza taksonomiczna w statystyce regionalnej, Difin, Warszawa

Metodologia badania migracji zagranicznych w narodowym spisie powszechnym ludności i mieszkań w 2011 roku (NSP 2011) (2014), GUS, Warsaw

Metodologia badania migracji z Polski (2007), Ośrodek Badań nad Migracjami UW, Warsaw

Nowak E. (2004), Metody klasyfikacji w badaniach geograficznych, Wyd. Nauk. Bogucki, Kielce-Poznań

Okólski M. (2001), Mobilność przestrzenna z perspektywy migracji niepełnej, in: Ludzie na huśtawce. Mobilność między peryferiami Polski i Zachodu, E. Jaźwińska, M. Okólski (ed.), Scholar, Warsaw

Panek T. (2009), Statystyczne metody wielowymiarowej analizy porównawczej, Oficyna Wydawnicza SGH, Warsaw

Pęczkowski M. (2009), Analiza skupień, in: Wielowymiarowa analiza statystyczna, E. Frątczak (ed.), Oficyna Wydawnicza SGH, Warsaw

Horodeński R., Sadowska-Snarska C. (ed.) (2003), Rynek pracy w Polsce na progu XXI. Aspekty makroekonomiczne i regionalne, IPiSS, Białystok-Warszawa

Sokołowski A., Pasztyła A. (2004), Data Mining w prognozowaniu, Statsoft Polska, Kraków

Strzelecki Z. (ed.) (1984), Teoretyczne i metodologiczne problemy badania różnych form migracji ludności Polski, Zakład Demografii Społeczno-Ekonomicznej Szkoły Głównej Planowania i Statystyki w Warszawie, Koszalin

Śleszyński P. (2006), Zmiany emigracji z Polski według oficjalnych statystyk w ujęciu przestrzennym po przystąpieniu Polski do Unii Europejskiej, in: ,,Biuletyn Migracyjny", Dodatek Nr 10, grudzień

Zamojski J.E. (2006), Migracje – dzieje i aktualia. Refleksje, w: Migracje: dzieje, typologia, definicje, A. Furdal, W. Wysoczański (ed.), Wyd. Uniwersytetu Wrocławskiego, Wrocław

Zelinsky W. (1971), The hypothesis of the mobility transition, w: ,,The Geographical Review", No. 61

Zastosowanie wybranych metod grupowania danych w badaniu przestrzennego zróżnicowania zjawisk społeczno-ekonomicznych na przykładzie mobilności zewnętrznej w Polsce w latach 1989-2002

Abstrakt

Celem artykułu jest ukazanie przydatności wybranych metod grupowania danych w analizie przestrzennej zjawisk społeczno-ekonomicznych na przykładzie analizy okresowej emigracji z Polski w latach 1989-2002. Badanie zmierzało do oceny przestrzennego zróżnicowania w wysokim stopniu szczegółowości, tj. z uwzględnieniem różnic występujących lokalnie. W przypadku emigracji okresowej taką ocenę umożliwiły zgromadzone w 2002 roku wyniki Narodowego Spisu Powszechnego.

W artykule zaprezentowano przykłady analiz terytorialnego zróżnicowania mobilności, w których zastosowano wybrane optymalizacyjno-iteracyjne i hierarchiczne metody grupowania. Elementami w przestrzeni grupowania były jednostki podziału terytorialnego w przekrojach: gmin, powiatów oraz podregionów.
Po 1989 roku, w okresie transformacji społeczno-ekonomicznej, czasowa emigracja w Polsce była zjawiskiem niezwykle silnie zróżnicowanym w wymiarze regionalnym oraz lokalnym, co potwierdziły wyniki przeprowadzonych badań. Wyraźną dyspersję przestrzenną wykazano pod względem poziomu, natężenia, ale także zróżnicowania demograficzno-społecznej struktury zjawiska. Zastosowanie przyjętych metod klasyfikacji danych umożliwiło szczegółową ocenę przestrzennego zróżnicowania okresowej emigracji w Polsce w przekroju niewielkich, a zarazem licznych jednostek terytorialnych oraz wizualizację wyników za pomocą wykresów mapowych.
Zakres czasowy badania został zawężony do lat 1989-2002 ze względu na charakter wykorzystanego materiału źródłowego oraz brak porównywalnych danych statystycznych o emigracji poakcesyjnej i współczesnej.

Słowa kluczowe: przestrzenne zróżnicowanie, analiza skupień, migracje zagraniczne, emigracja