



Munich Personal RePEc Archive

# **Graph Regionalization with Clustering and Partitioning: an Application for Daily Commuting Flows in Albania**

BENASSI, FEDERICO and DEVA, MIRELA and  
ZINDATO, DONATELLA

Department for Censuses, Statistical and Administrative Archives,  
Italian National Institute of Statistics, Cartography and GIS sector,  
National Institute of Statistics, ) Department for Censuses,  
Statistical and Administrative Archives, Italian National Institute of  
Statistics

July 2015

Online at <https://mpra.ub.uni-muenchen.de/73946/>  
MPRA Paper No. 73946, posted 23 Sep 2016 11:25 UTC

FEDERICO BENASSI<sup>a)</sup> – MIRELA DEVA<sup>b)</sup> – DONATELLA ZINDATO<sup>c)</sup>

## **Graph Regionalization with Clustering and Partitioning: an Application for Daily Commuting Flows in Albania\***

### **Abstract**

The paper presents an original application of the recently proposed spatial data mining method named GraphRECAP on daily commuting flows using 2011 Albanian census data. Its aim is to identify several clusters of Albanian municipalities/communes; propose a classification of the Albanian territory based on daily commuting flows among municipalities/communes. Starting from 373 local units, we first applied a spatial clustering technique without imposing any constraining strategy. Based on the input variables, we obtained 16 clusters. In the second step of our analysis, we impose a set of constraining parameters to identify intermediate areas between the local level (municipality/commune) and the national one. We have defined 12 derived regions (same number as the actual Albanian prefectures but with different geographies). These derived regions are quite different from the traditional ones in terms of both geographical dimensions and boundaries.

*Keywords:* GraphRECAP, regionalization, daily commuting flows, census data, Albania, territorial imbalances.

### **Introduction**

In the last decade, there has been a growing interest in modelling and understanding commuting behaviours (Eliasson-Lindgren-Westerlun 2003, Schwanen-Dieleman-Dijst 2004, Champion 2009) and the derived urban spatial structure (Ding 2007, Knox–McCarthy 2005). Particular interest has been devoted to the use of work commuting flows as a base to construct functional areas and to propose more efficient delimitations of the territory (Cövers–Hensen-Bongaerts 2009, Landré–Håkansson 2010, Rain 1999).

For the first time in Albanian history, the 2011 Population and Housing Census has collected data on commuting from home to work. Based on the use of these data at the municipality/commune level of analysis, this paper aims to i) identify several clusters of municipalities/communes, similar in terms of commuting profiles; ii) propose a

a) Department for Censuses, Statistical and Administrative Archives, Italian National Institute of Statistics, Rome, 00144, Italy. E-mail: benassi@istat.it

b) Cartography and GIS sector, National Institute of Statistics, Tirana, 1004, Albania. E-mail: mdeva@instat.gov.al

c) Department for Censuses, Statistical and Administrative Archives, Italian National Institute of Statistics, Rome, 00144, Italy. E-mail: zindato@istat.it

\* The authors would like to thank Professor Nicola Salvati (University of Pisa) for reading a preliminary version of the paper and for providing us with many useful considerations.

classification of the Albanian territory based on daily commuting flows among municipalities/communes.

To achieve our research objectives (i and ii), we applied a recently proposed spatial data mining method, named GraphRECAP (Guo 2010). GraphRECAP – Graph Regionalization with Clustering and Partitioning – is a toolkit for partitioning spatially embedded graphs (such as country-to-country migrations or commuting flows) and deriving spatially contiguous regions based on graph connections (Guo 2010). Readers are referred to Guo’s publication for computational, technical and methodological details (Guo 2008, 2009, 2009a, 2010, 2010a).

### **Graph Regionalization with Clustering and Partitioning: the first exploratory analysis without imposing a spatial constraining strategy**

We computed ten indicators for each of the 373 Albanian local units (municipalities/communes). These indicators plus the spatial attributes of each municipality/commune are the input variables of GraphRECAP. The ten indicators are: 1) surface (square meters); 2) daily outflow degree: the number of destinations that have daily flows from each municipality; 3) daily outflow: the total daily volume (i.e. total daily commuters) going outside each municipality/commune; 4) adjusted daily outflow ratio: the ratio of the daily outflow and the usually resident population; 5) adjusted daily net flow ratio: the ratio of the daily net flow (daily inflow-daily outflow) and the usually resident population; 6) adjusted daily inflow ratio: the ratio between daily inflow and the usually resident population; 7) daily inflow degree: the number of destinations that have daily flows to each municipality/commune; 8) usually resident population; 9) daily inflow: the total daily volume (i.e. the total daily commuters) coming to each municipality/commune; 10) daily net flow: daily inflow – daily outflow.

On the basis of the standardized input variables, and thanks to the use of the SOM (Self-Organizing Map) technique, GraphRECAP identified 16 clusters of municipalities/communes. This number is due to the dimension of the SOM (4\*4), which was chosen out of the various alternatives, since it proved to ensure the best results in terms of cluster differentiation (each cluster has the highest possible inner homogeneity while heterogeneity among clusters is the highest possible). As mentioned previously, SOM stands for Self-Organizing Map and was first described by Kohonen (1995). A SOM is an artificial neural network that is trained by using an unsupervised learning process. The dimension of the SOM ( $X \times Y$ ) is normally defined by the user; we just recall that mathematically a 2D-SOM is a 3D-matrix of dimension with  $n$  the dimension of the input space. The SOM is composed of a neuron that is a cell characterized by its position in the plane of the SOM. The dimension of the neuron is  $n$ . The SOM is initialized randomly with a uniform distribution within the range of the input data. For each iteration, the algorithm randomly selects an input vector from the input space. It then computes the Euclidean distance between the input vector and each neuron of the map. The neuron with the minimal distance, which is called the Best Matching Unit (BMU), is therefore selected. After these two steps, the map is modified with the following formula:

$$M(t+1) = M(t) + \alpha(t) \times \theta(t, \beta_1, \beta_2) \times (V - W_{ij}(t)) \quad 1 \leq i \leq X, 1 \leq j \leq Y \quad (1)$$

where  $M(t)$  and  $M(t+1)$  are, respectively, the maps at iteration  $t$  and at iteration  $t+1$ ;  $\alpha(t)$  is the learning rate that weights the effect of the input vector during the training process;  $\theta(t, \beta_1, \beta_2)$  is the neighbourhood function that regulates the influence of Best Matching Unit  $(\beta_1, \beta_2)$  on the neighbouring neurons; and finally,  $(V - W_{ij}(t))$   $1 \leq i \leq X, 1 \leq j \leq Y$  is a linear adjustment of the weights. In particular, the neighbourhood function is:

$$\theta(t, \beta_1, \beta_2) = \exp\left(-\frac{(i - \beta_1)^2 + (j - \beta_2)^2}{2\sigma^2(t)}\right) \quad (2)$$

In (2),  $\sigma$  is the radius of the neighbourhood function, and it is equal to:

$$\sigma(t) = (\sigma(t_i) - \sigma(t_f)) \times \exp\left(-\frac{t}{\lambda}\right) + \sigma(t_f) \quad (3)$$

while  $\alpha$ , the learning rate, is equal to:

$$\alpha(t) = (\alpha(t_i) - \alpha(t_f)) \times \exp\left(-\frac{t}{\lambda}\right) + \alpha(t_f) \quad (4)$$

The iterative procedure continues since each elementary unit is assigned to a certain cluster. The results are finally presented in a U-matrix (Unified distance matrix). This matrix is the mean Euclidean distance of each neuron with their  $n$  neighbours. That is to say:

$$U = \frac{1}{n} \sum_{\mu \in N(v)} d(v, \mu) \quad (5)$$

From a graphical perspective, each SOM node (cluster) is represented by a circle, whose size (area) is proportional to the number of elementary units (municipalities/communes) that it contains. As was mentioned, SOM uses Euclidean distance to assess the multivariate similarity between spatial objects. Therefore, nearby clusters are more similar to each other than those far away (Guo–Gahegan–MacEachren–Zou 2005, Kohonen 1995). Behind the SOM nodes, there is a U-Matrix layer where hexagons are shaded to show the multivariate dissimilarity between neighbouring nodes, with darker tones representing greater dissimilarity (Guo 2010a).

Furthermore, a Parallel Coordinate Plot (PCP) is used to reveal the meaning of each municipality/commune assigned to each cluster by SOM, since on this plot we can observe the clusters' statistical profile related to the statistical indicators used in the analysis, and their level of dissimilarity. Finally, the results are related to each other and visualized on a multivariate/interactive map.

In Figure 1, we can see the general result of this first step of the analysis. In the multivariate mapping (a), municipalities/communes with the same colours belong to the same cluster while, in the clustering with SOM (b), node hexagons (clusters of municipalities/communes) with similar colours present a lower level of dissimilarity. This difference can also be seen in the PCP (c), where lines of colours very different from each other refer to clusters that present divergent values of the input indicators (that is to say a

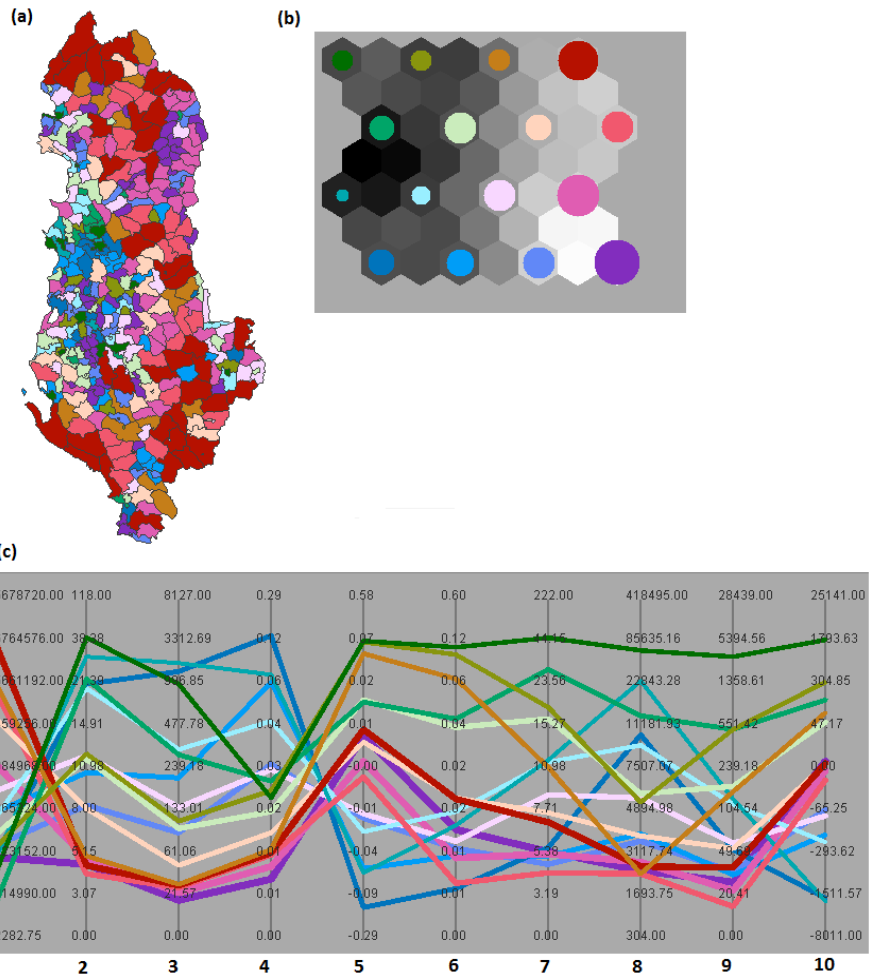
divergent “profile”). It should be noted that the thickness of the lines in (c), as the circumference of the node hexagons in (b), is proportional to the size of clusters.

PCP is made up of as many parallel axes as the indicators used in the analysis. Every axis is scaled using the nested means method, which puts the mean value at the centre of the axis, thus making comparable axes defined by different units and different data ranges (Guo 2010a). This scaling method can alleviate overlapping problems in PCP for skewed data distributions. More specifically, a nested means method is a non-linear scaling that recursively calculates a number of mean values (and sub means) and uses these values as break point to divide each axis into equal-length segments.

An explorative analysis was carried out based on these first general results, with the aim of obtaining a further classification of the 16 clusters (and therefore of the basic units that belong to each cluster) in order to identify some primary groups. Based on the properties of the SOM (i.e. clusters with similar colours present low levels of dissimilarity among each other), the 16 clusters have been further grouped into 6 groups. We will now describe and discuss the main features of each of them. The first, Group 1, is characterized by clusters of municipalities/communes with a comparatively high value of indicator 1; low value of indicators 2, 3, 4, 8 and 9; medium values (except for the brown cluster) of indicators 5, 6 and 10. This group is therefore composed of clusters of municipalities/communes with a large territory and a small usually resident population, characterized by a low level of daily spatial interactions (the levels of daily inflow and daily outflow are comparatively low). Looking at the spatial distribution of the municipalities/communes belonging to this first group, we can clearly see that almost all of them are located in the rural and mountain areas of Albania and on the coastal areas of the southwestern part of the country. It should be noted that the municipalities/communes belonging to this group are often territorially contiguous, thus showing the existence of spatial patterns. We can define this group as “Big/Peripheral” since it is composed of clusters of municipalities/communes that are big in terms of surface, yet peripheral with regard to their territorial location but also with regard to the role played in the daily interactions spatial system of Albania (Figure 2).

Figure 1

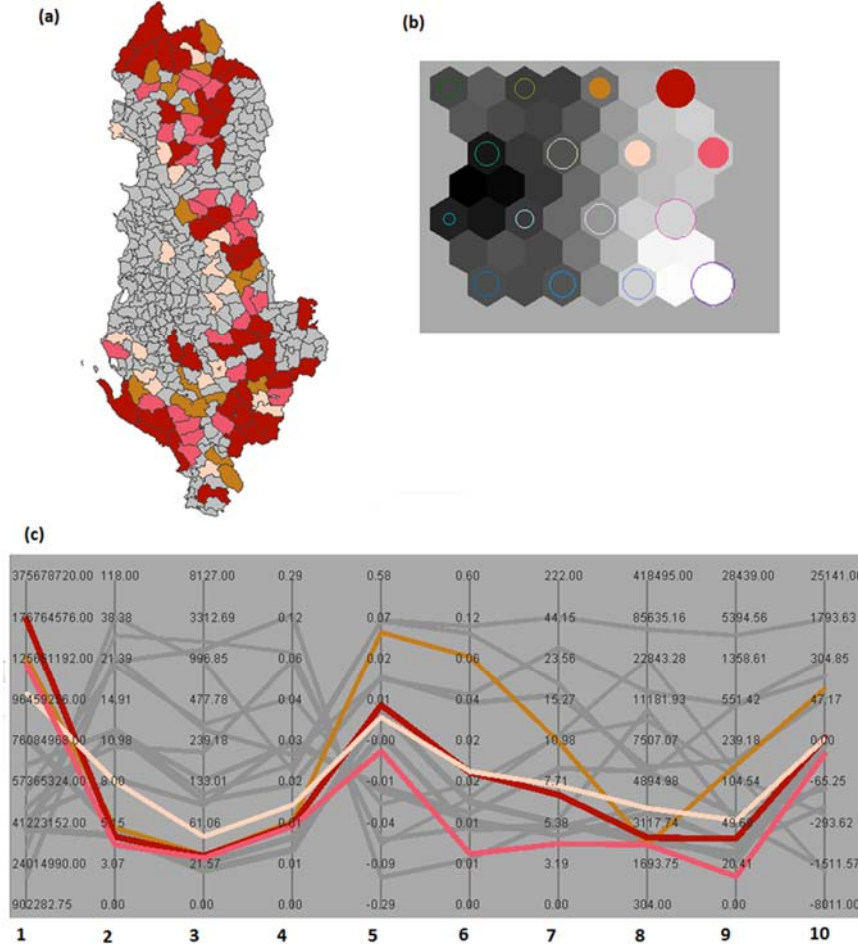
General results. (a) Multivariate mapping. (b) Clustering with SOM. (c) Multivariate visualization of clusters (Parallel Coordinate Plot)



Source: own processing on Instat data, 2011 Population Census.

Figure 2

Group 1, “Big/Peripheral”. (a) Multivariate mapping. (b) Clustering with SOM. (c) Multivariate visualization of clusters (Parallel Coordinate Plot)



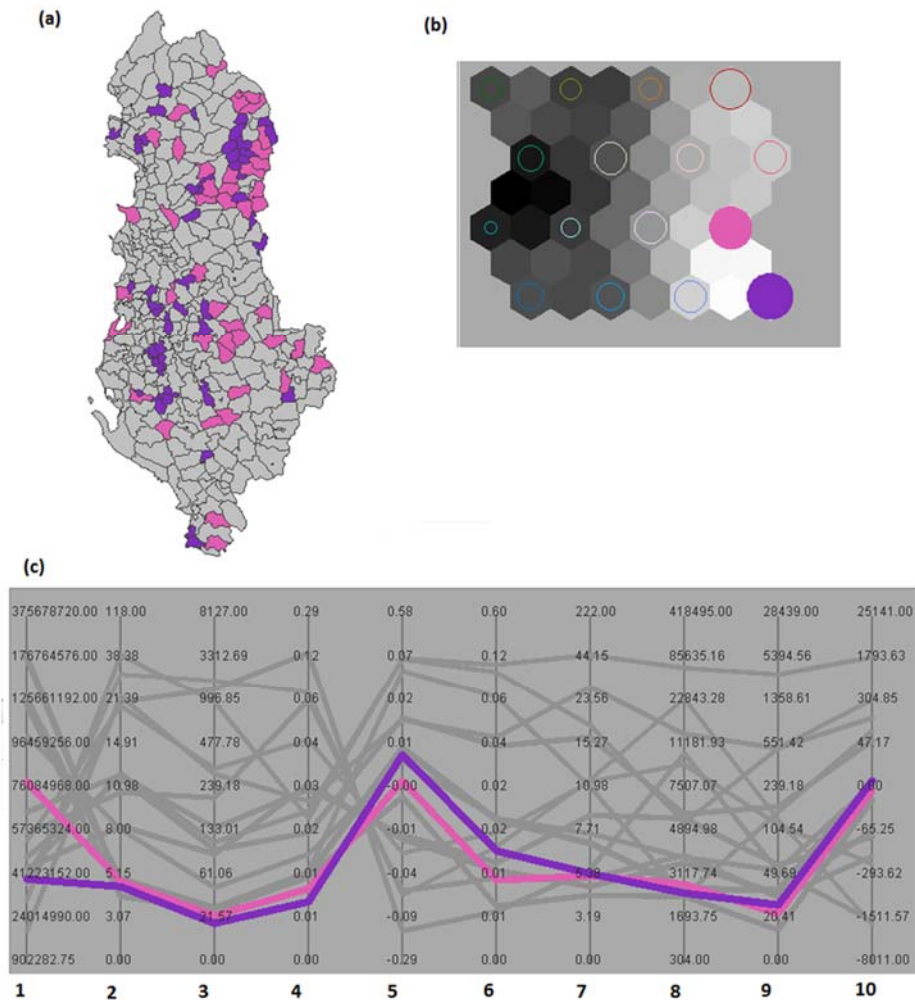
Source: own processing on Instat data, 2011 Population Census.

The second group, Group 2, is quite similar to the first one, except the values recorded for indicator 1 (surface). Municipalities/communes belonging to this second group are characterized by a small territorial area, by comparatively low levels of indicators 2, 3, 4, 6, 7, 8 and 9, and by medium values of indicators 5 and 10. We can define this group as “Small/Peripheral” since it is composed of municipalities/communes with a low level of daily spatial interactions, and which are comparatively small in terms of usually resident population and territory. The level of spatial contiguity among municipalities/communes belonging to this group is lower compared to that of municipalities/communes belonging

to group 1, but even in this group some municipalities/communes are territorially contiguous (namely, municipalities/communes located in the north-eastern part of the country). In terms of localization, this group may be divided into two categories; the first category is composed of municipalities/communes located in the north-eastern part of the country, with a certain level of spatial concentration and spatial contiguity among them. The second category, on the contrary, is composed of municipalities and communes, which are quite scattered and located mainly in the central and in the southern part of the country (Figure 3).

Figure 3

Group 2, “Small/Peripheral”. (a) Multivariate mapping. (b) Clustering with SOM. (c) Multivariate visualization of clusters (Parallel Coordinate Plot)



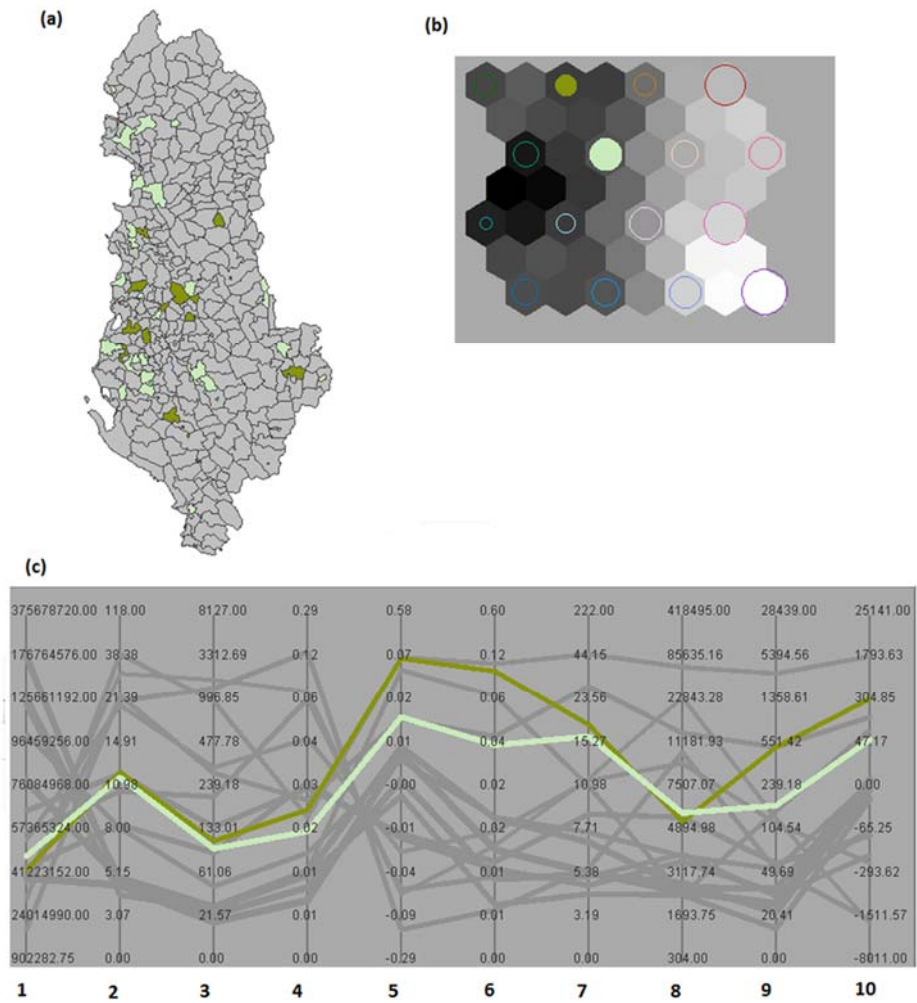
Source: own processing on Instat data, 2011 Population Census.



The third group, Group 3, is quite different from groups 1 and 2. We have defined this group as “Medium/Semi Central” since the municipalities/communes that belong to this group are characterized by almost medium values of indicators 1 and 3, medium values of indicators 2, 4 and 8, and by comparatively high values of indicators 5, 6, 7, 9 and 10. In other words, the municipalities/communes that belong to this group are characterized by a system of spatial daily interactions where the level of daily inflows is higher than the level of daily outflow (for this reason, the values of daily net flow and adjusted net flow ratio are quite high). The number of municipalities/communes from where the flows originate is quite high, while almost all other indicators reveal a medium situation, especially in terms of surface and usually resident population. The municipalities/communes of this group are therefore medium in terms of these two dimensions (surface and usually resident population) and play a semi-central role in the system of spatially daily interactions of Albania (Figure 4).

Group 4 is very different from those described until now. It is composed of municipalities/communes with a very low level of indicator 1, a medium value of indicator 4, and very high level of indicators 2, 3, 5, 6, 7, 8, 9 and 10. Consequently, the municipalities/communes belonging to this group are characterized by a very dynamic system of daily spatial interactions, where the level of the daily inflow is higher than the level of daily outflow. The number of municipalities/communes both originating and providing a destination for daily movements is very high, as well as the size of the usually resident population. On the contrary, the municipalities of this group are very small in terms of surface. We have defined this group as “Small/Central (Prey)”. The municipalities of this group are, in fact, small in terms of surface, but play a crucial role in the system of daily spatial interactions of Albania. The term “prey” is adopted taking into account in a broad sense the logic and definitions of the prey-predator model elaborated by Lotka (1925) and Volterra (1926). This is because the municipalities/communes of this group (as we will see when describing the profile and characteristics of the next two groups) are predated by a number of other municipalities/communes that are very close to them in terms of spatial location, but, at the same time, they present divergent profiles compared to the them. In terms of spatial location, we can clearly see that, in this case, the condition of spatial contiguity is not confirmed. Municipalities belonging to this group are quite scattered in terms of spatial location; finally, it should be noted that the main Albanian municipalities (Tirana, Durrës, Vlorë, Elbasan and Shkodër) belong to this group (Figure 5).

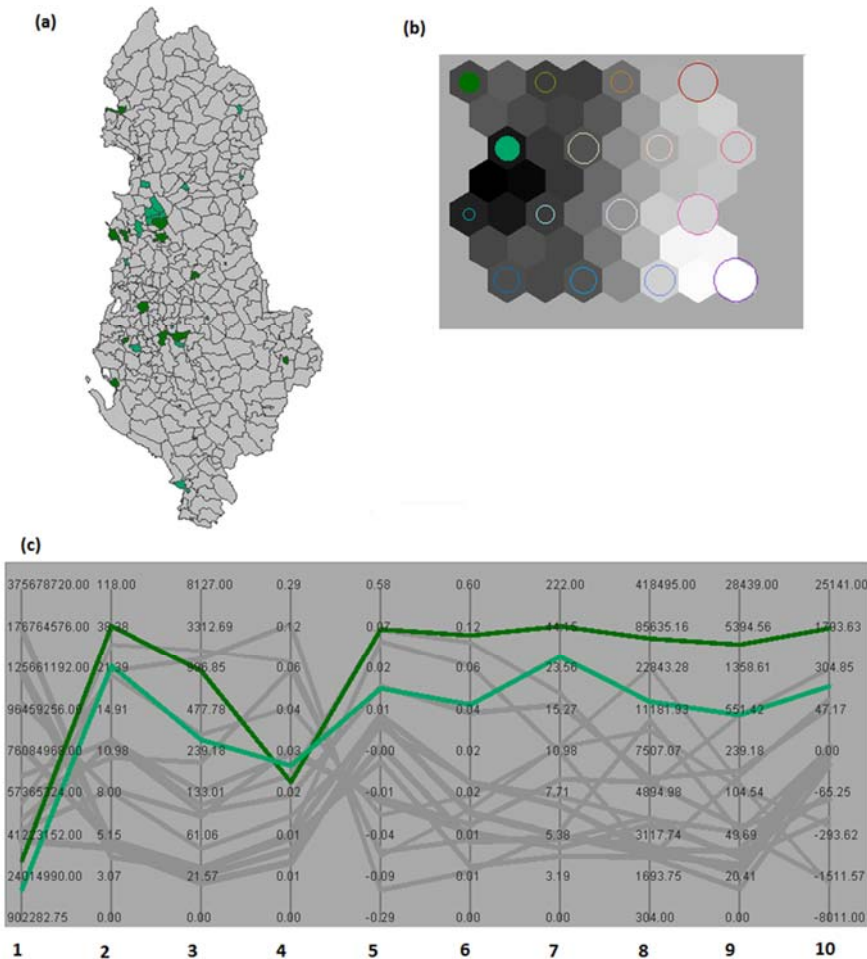
Figure 4  
 Group 3, “Medium/Semi Central”. (a) Multivariate mapping. (b) Clustering with SOM.  
 (c) Multivariate visualization of clusters (Parallel Coordinate Plot)



Source: own processing on Instat data, 2011 Population Census.

Figure 5

Group 4, “Small/Central (prey)”. (a) Multivariate mapping. (b) Clustering with SOM. (c) Multivariate visualization of clusters (Parallel Coordinate Plot)



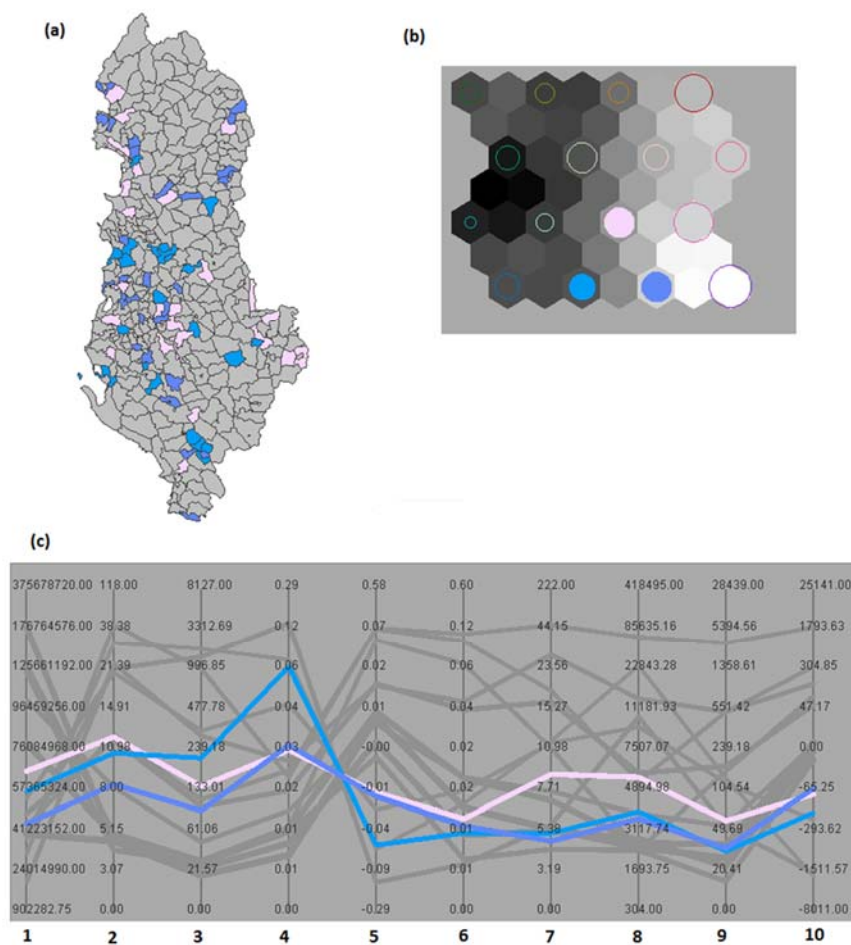
Source: own processing on Instat data, 2011 Population Census.

Group 5, “Medium/Central (semi-predators)”, is composed of municipalities/communes that have medium values of the indicators 1 and 2; medium/high values of the indicators 3 and 4 and, finally, comparatively low values of indicators 5, 6, 7, 8, 9 and 10 (with the exception, in some cases, of the pink cluster of municipalities/communes). That is to say that this group is characterized by a low level of daily inflow and a comparatively high level of daily outflow. This is also why the level of daily net flow and adjusted daily net flow ratio is comparatively low. Looking at the spatial location of the municipalities/communes that belong to this group, we can see from the map that they are often close to the municipalities belonging to group 4 (prey). For such a reason, we defined

this group as “Medium/Central (semi predators)”. The municipalities belonging to this group have a medium-small dimension in terms of usually resident population and in terms of surface. They play a central role in the Albanian system of daily spatial interactions. A role that can be defined as that of “semi-predators” in that they present a quite low level of daily inflows and a relatively high level of outflows together with a spatial distribution that underlines that they are usually not so far from the prey (Figure 6).

Figure 6

Group 5, “Medium/Central (semi predators)”. (a) Multivariate mapping. (b) Clustering with SOM. (c) Multivariate visualization of clusters (Parallel Coordinate Plot)



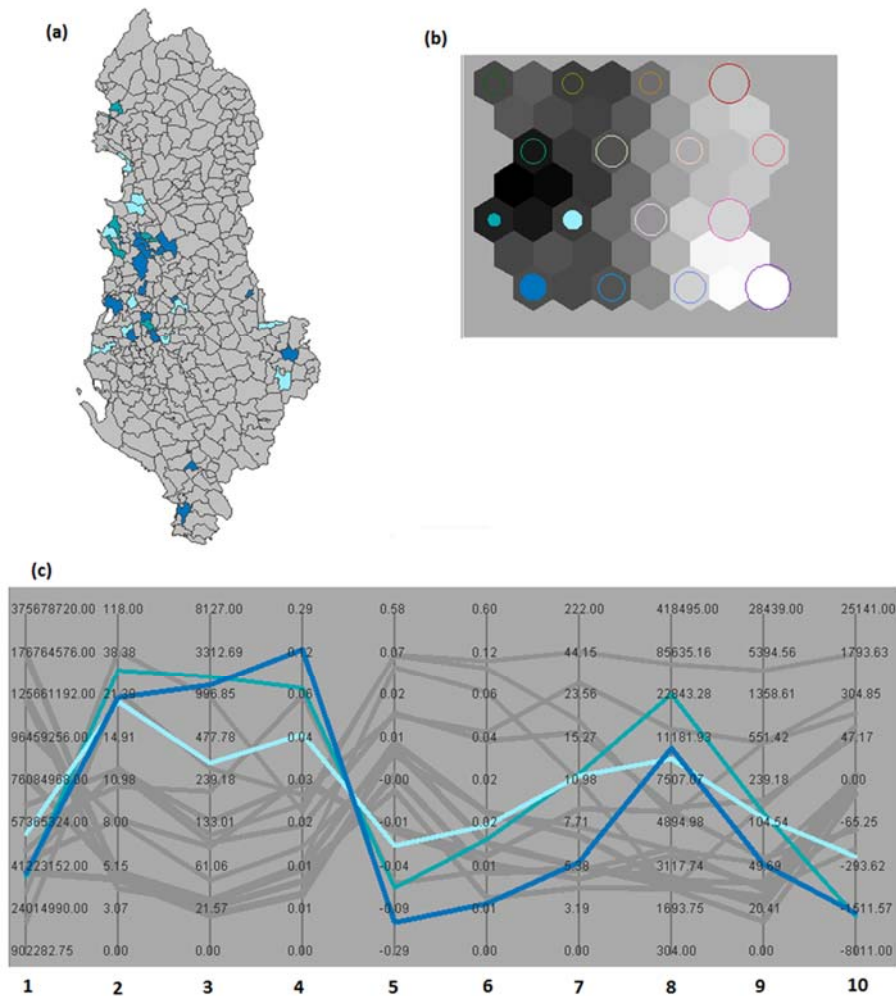
Source: own processing on Instat data, 2011 Population Census.

The last group, “Small/Central (predators)”, clarifies the previously mentioned concepts of prey and predators. This group is characterized by very high levels of indicators 2, 3, 4 and 8; by a low level of indicators 1, 5, 6, 9, 10 and by medium/low levels of

indicators 7. That is to say that, in the municipalities belonging to the clusters that constitute this group, the level of total daily flow (daily inflow + daily outflow) is relatively high, but with a clear prevalence of daily outflows. The level of daily inflow is, in fact, lower compared to the level of daily outflow. Municipalities of this group are relatively big in terms of usually resident population and relatively small in terms of territory. Looking at the spatial location of the municipalities belonging to this group, we can clearly see that they are mainly located close to the prey municipalities, with a high level of spatial contiguity. It is especially the case of Durrës, Tirana and Shkodër (Figure 7).

Figure 7

*Group 6, “Small/Central (predators)”.* (a) Multivariate mapping. (b) Clustering with SOM. (c) Multivariate visualization of clusters (Parallel Coordinate Plot)



Source: own processing on Instat data, 2011 Population Census.

### **Graph Regionalization with Clustering and Partitioning: Identifying Natural Regions By Applying a Spatial Constraining Strategy**

In the second step of our analysis, we used a spatially constrained graph partitioning technique to identify a hierarchy of natural regions defined by spatial interactions. The natural (derived) regions so identified, under a set of constrained parameters, are composed of clusters of municipalities/communes that are both spatially contiguous and homogeneous in terms of the characteristics of their daily spatial interactions. At a territorial level, the natural regions are intermediate areas between the local level (municipalities/communes) and the national one.

This kind of spatial statistical analysis belongs to the class of regionalization methods. As defined in Guo (2008), regionalization is a process that divides a large set of spatial objects into a number of spatially contiguous regions, while optimizing an objective function, typically a homogeneity (or heterogeneity) measure of the identified regions. Therefore, regionalization is a special kind of spatial clustering where the condition of spatial contiguity between spatial objects plays a key role.

As recalled by Bernetti–Ciampi–Sacell–Marinelli (2011), regionalization processes play an important part in many research sectors, finding applications in areas like climatic zoning (Fovell–Fovell 1993, Wang–Zhang–Li–Song 2010), environmental analysis (Henderson 2006, Romano–Balzanella–Verde 2010), landscape analysis (Long–Nelson–Wulder 2010), the interpretation and organization of Census data (Openshaw–Rao 1995) and public health data (Haining–Wise–Blake 1994, Osnes 1999), the analysis of socio-economic phenomena (Assunção–Neves–Câmara–Da Costa Freitas 2006), the analysis and interpretation of demographic and urban/regional dynamics (Behnisch–Ulsch 2010, Benassi–Bocci–Petrucci 2013) and the analysis of migration flows (Guo 2009). The concept of regionalization hypothesized and applied to socio-economic entities by Openshaw (1977) results in the creation of geographic objects formed by combining contiguous elements sharing one or more characteristics and it is closely connected with spatial statistics (Bernetti–Ciampi–Saccelli–Marinelli 2011).

The starting point is that, following Guo (2010, 2010a), spatial interactions naturally form a network/graph, where each node is a location (or area) and each link is an interaction between two nodes (location). Such spatial interaction networks (e.g. municipalities/communes to municipalities/communes daily commuting flows) normally consist of:  $S$ , a set of locations (nodes), in our case the municipalities/communes of Albania;  $F$ , a set of flows (links) between locations, in our case the daily spatial commuting flows (direct) among Albanian municipalities/communes; and  $V_f$ , a set of variables for each flow. From this perspective, regionalization can reduce spurious data variations caused by uneven sizes or small base populations, and generalize (i.e. find general rules in) large spatial interactions data to discover general flow patterns. The key requirement is that the regionalization process should allow major patterns in the network to be preserved while suppressing details (Guo 2010, 2010a).

Coming back to our study, our aim is to identify  $n$  spatial areas (natural regions), intermediate between the local and the national level, that, under a constrained strategy, will minimize inner heterogeneity (within regions) and maximize external heterogeneity (between regions) with regard to daily spatial interactions.

The key challenge of this operation is to identify regions based on commuting flows (this is why we call them natural or derived regions) instead of using pre-defined political or administrative boundaries.

After computing a contiguity matrix, which specifies which items (municipalities/communes) are neighbours in space, in order to complete the regionalization process, we have to define a constraining strategy and a set of parameters. Referring to the work of Guo for in-depth methodological details (2008, 2010, 2010a), we will now describe the constraining strategy and parameters adopted. As a regionalization method, the Full Order-ALK method has been chosen, which is a combination of the agglomerative clustering method named ALK (Average Linkage Clustering) and the spatial constraining strategy named Full Order. The ALK method derives natural regions in two steps. It first constructs a hierarchy of clusters from the bottom by iteratively merging the most connected clusters. Therefore, the method needs a contiguity matrix as input. The output is a spatially contiguous tree, where each edge connects two geographic neighbours and the entire tree is consistent with the cluster hierarchy. Second, the spatially contiguous tree is partitioned from the top by finding the best edge to remove. By repeating this step for each new region, a hierarchy of regions is constructed. During this partitioning process, additional constraints may be enforced; for example, we may want to impose a minimum population size for each region (Guo 2010a). Guo proves in his work (2008) that this method derives regions of significantly better quality (in terms of the objective function value) than other existing methods.

As a flow expectation model, we chose the Expectation SI\_FLOW; this model calculates an expected flow value for each pair of spatial objects based on the total in and out flows of each object (in our case the total in and out daily flows of each Albanian municipality/commune).

Finally, to derive regions from the spatial interactions flows, a measure of similarity of the strength of connection has to be defined for each pair of locations (or regions). Following the work of Guo (2010a), in this paper we adopt the concept of modularity measure, which is defined by the following equation:

$$\text{Modularity} = \text{Actual Flows} - \text{Expected Flows} \quad (6)$$

Different statistical models can be used to calculate expectation flows. In this paper, the simplest model is used, which assumes that interactions among locations are random and proportional to the origin and destination populations (Guo 2010a). In our case, we assume that each individual has the same probability to commute and the choice of the destination is proportional to the population of the destination place:

$$\text{Expected Flows} (A, B) = P_A P_B F / P_S^2 \quad (7)$$

where  $P_S$  is the total population for all locations  $S$ ,  $P_A$  is the population of the region  $A \subseteq S$ ,  $P_B$  is the population of  $B \subseteq S$ ,  $A \cap B = \emptyset$ , and  $F$  is the total flow among all locations (including flows within the same location). In this way, we ensure that the total expected flows are the same as the total actual flows (Guo 2010a).

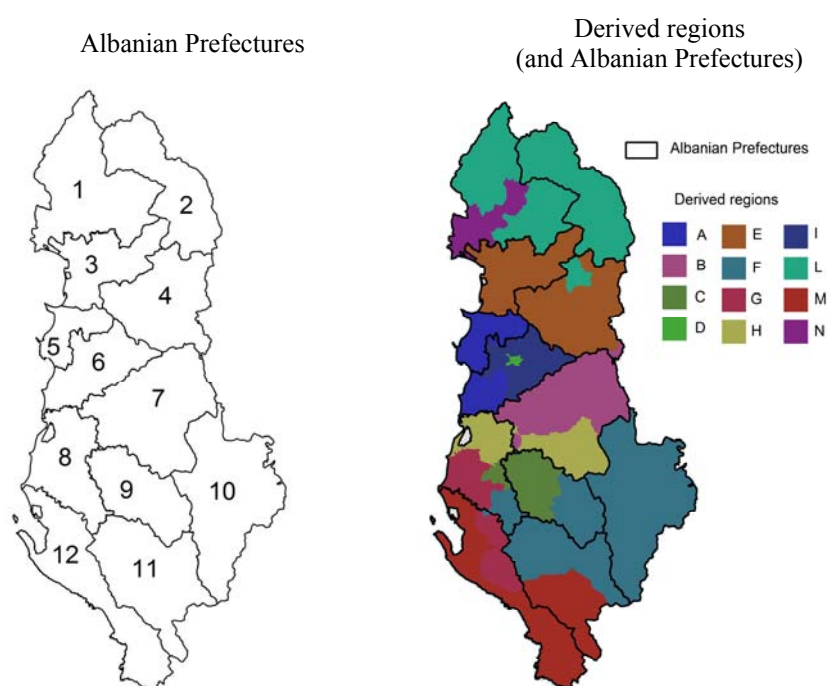
In addition to these parameters, we impose two additional constraints during the partitioning process. We have fixed a maximum number of regions equal to 12 (the actual number of Albanian prefectures) and a minimum size of the usually resident population

per region equal to 150,000 units. With these parameters and under the constraint of spatial contiguity, we identified 12 natural regions (Figure 8).

As can clearly be observed from Figure 8, the derived regions are quite different from the traditionally used Albanian prefectures, both in terms of geographical dimension (some are smaller, others are bigger) and, obviously, in terms of territorial boundaries. Combining the geography of the derived regions with the commuting profiles of the elementary units of which they are composed (municipalities/communes), we can define the nature of each derived region inside the system of daily spatial interactions of Albania.

Figure 8

*Albanian prefectures and derived regions*



*Note:* Albanian prefectures: 1 Shkodër; 2 Kukës; 3 Lezhë; 4 Dibër; 5 Durrës; 6 Tirana; 7 Elbasan; 8 Fiër; 9 Berat; 10 Korçë; 11 Gjirokastër; 12 Vlorë.

*Source:* own processing on Instat data, 2011 Population Census.

From this perspective, we found 5 (E, F, G, L, M) derived regions characterized by a peripheral commuting profile; 4 derived regions (B, C, H, N) characterized by a semi-central commuting profile and finally, 3 derived regions (A, D, I) characterized by a central profile (Table 1). The derived regions with a peripheral commuting profile are almost all located on the mountains and the border areas of the country and, in general, in the less urbanized part of Albania. They are composed of medium/big municipalities in terms of surface, with comparatively small populations, which play a marginal role in terms of attraction and repulsion of commuting flows. The derived regions with a semi-central commuting profile are, on the contrary, located in areas with a high level of urbanization



in the coastal part but also on the western, central and northern part of Albania. Finally, the derived regions with a central commuting profile are located in areas with a comparatively high level of urbanization. We refer in particular to Durrës and its surrounding area, and to Tirana and its surrounding area. These derived regions are characterized by high levels of daily commuting inflows and outflows and qualify themselves as primary players in the Albanian system of daily spatial interactions.

Table 1

*Synoptic table: derived regions, traditional components of the derived regions, commuting profiles of the derived regions, geographical locations of the derived regions*

Derived regions	Traditional components of the derived regions	Commuting profile of the derived regions	Geographical locations of the derived regions
A	Durrës (5) + part of Tirana (6)	Central profile	Western/coastal areas in the central part of Albania
B	Part of Elbasan (7)	Semi central profile	Eastern/ central areas of Albania
C	Part of Fier (8) + part of Berat (9)	Semi central profile	Western coastal and central part of Albania
D	Part of Tirana (6)	Central profile	Western/central part of Albania (municipality of Tirana)
E	Lezhë (3) + part of Dibër (4)	Peripheral profile	Eastern and Western part of Albania/Central-North areas
F	Korçë (10) + part of Fier (8) + part of Berat (9) + part of Gjirokastër (11)	Peripheral profile	Southern/east and Southern/central part of Albania
G	Part of Vlorë (12) + part of Fier (8)	Peripheral profile	Southern/west coastal area in south part of Albania
H	Part of Fier (8) + part of Berat (9)	Semi central profile	Western coastal/central part of Albania
I	Part of Tirana (6)	Central profile	Central part of Albania (surrounding areas of Municipality of Tirana)
L	Kukës (2)+ part of Shkodër (1)	Peripheral profile	Eastern and Western North part of Albania/mountain areas
M	Part of Vlorë (12) + Part of Gjirokastrër (11)	Peripheral profile	South/western coastal part of Albania
N	Part of Shkodër (1)	Semi central profile	North Western coastal part

## Conclusions

The application of the recently proposed method GraphRECAP on daily commuting flows of Albania has obtained interesting results.

Starting from 373 local units (municipalities/communes of Albania) and, therefore, from a 373\*373 square matrix of daily commuting flows, we first applied a spatial clustering technique, without imposing any constraining strategy, and obtained 16 clusters.

These clusters were then further classified into sub-groups defined by a number of demographic and territorial dimensions and by the role played in the commuting system of Albania. We thus identified 6 primary sub-groups: 1) Big/Peripheral; 2) Small/Peripheral; 3) Medium/Semi central; 4) Small/Central (Prey); 5) Medium/Central (semi-predators); 6) Small/Central (predators).

In the second step of our analysis, we imposed a set of constraining parameters to identify intermediate areas between the local level (municipality/commune) and the national one. We have defined 12 derived regions, the same number as Albanian prefectures. Although, these derived regions are quite different from the administrative ones both in terms of geographical dimensions and boundaries.

In our opinion, the derived regions (as well as the 6 subgroups) effectively represent the territorial and demographic imbalances that characterize today's Albania. Here, mountainous and less developed areas are composed of a comparatively large number of municipalities/communes that are quite vast in terms of surface but with a very low demographic density and with a marginal role in the commuting system. These municipalities/communes are characterized by a depopulation process where the population migrates to other areas and in particular, to the bigger urban centres (Tirana and Durrës) and their surrounding areas. This is why these areas (Tirana and Durrës) have been identified as specifically derived regions with a high level of commuting activities. The surrounding areas of these are characterized by a semi-central commuting profile and by an intermediate situation between the two described.

The largest municipalities (Tirana, Durrës, Vlorë, Elbasan, Shkodër, Fier, Korçë) act both as attraction poles and as poles from which a daily redistribution of commuting workers takes place. Altogether, large municipalities and municipalities/communes in their surrounding areas form complex systems of daily mobility, which in some cases are also linked to each other (Tirana/Durrës; Tirana/Elbasan; Tirana/Fier).

The leading role of large municipalities is confirmed and further clarified by the results of the multivariate spatial analysis of daily commuting flows. More precisely, they appear as playing a crucial central role, but also as “prey” predated by a number of other municipalities/communes (that we could define “predators”), which are very close to them in terms of spatial location and that present quite divergent profiles in terms of the indicators chosen for the analysis (e.g. they are characterized by a low level of the daily inflow and by a comparatively high level of the daily outflow; they are also medium/small size in terms of usually resident population and in terms of surface). From a commuting perspective of analysis, it seems reasonable to define Albania as a dual system. A system in which one part plays a primary role (A, B, C, D, H, I, N derived regions) and the other part, which seems to play a very marginal role (E, F, G, L, M, derived regions). In the author's opinion, these dynamics could reinforce territorial imbalances and have negative repercussions on Albania's sustainable development.

## Summary

This paper presents an original application of GraphRECAP, a recently proposed method of Graph Regionalization with Clustering and Partitioning elaborated by Guo (2010). The application concerns the study of daily spatial interactions (work commuting flows) among

the Albanian municipalities/communes, based on the use of Population Census data (2011), which for the first time investigated this phenomenon.

The study firstly defines clusters of municipalities/communes similar in terms of commuting profiles and then, through the imposition of a set of constrained strategies, proposes a new kind of regionalization of the Albanian territory. The results clearly show the actual territorial and demographic imbalances of Albanian society. These imbalances seem to be reinforced by the spatial patterns of the commuting flows. A dual territorial space comes to light, in which largest municipalities act both as attraction poles and as poles from which a daily redistribution of commuting workers takes place, while the smallest and peripheral municipalities/communes play the role of origin areas of commuting flows.

#### REFERENCES

- Assunção, R. M.–Neves, M. C.–Câmara, G.–Da Costa Freitas, C. (2006): Efficient regionalization techniques for socio-economic geographical units using minimum spanning trees *International Journal of Geographical Information Sciences* 20 (7): 797–811.
- Behnisch, M.–Ultsch, A. (2010): *Are there clusters of communities with the same dynamic behavior?*, Classification as a tool for research, Springer, Verlag.
- Benassi, F.–Bocci, C.–Petrucci, A. (2013): *Spatial data mining for clustering: an application to the Florentine Metropolitan Area using RedCap*, Classification and Data Mining, Springer, Berlin/Heidelberg/New York.
- Bernetti, I.–Ciampi C.–Sacelli, S.–Marinelli, A. (2011): The planning of agro-energetics districts. An Analysis model for Tuscany Region *Italian Journal of Forest and Mountains Environments* 66 (4): 305–320.
- Champion, T. (2009): Urban-Rural Differences in Commuting in England: A Challenge to the Rural Sustainability Agenda? *Planning Practice & Research* 24 (2): 161–183.
- Cövers, F.–Hensen, M.–Bongaerts, D. (2009): Delimitation and Coherence of Functional Administrative Regions *Regional Studies* 24 (2): 161–183.
- Ding, C. (2007): *Urban Spatial Planning: Theory, Method and Practice* Higher Education Press, Beijing.
- Eliasson, K.–Lindgren, U.–Westerlund, O. (2003): Geographical Labour Mobility: Migration or Commuting *Regional Studies* 37 (8): 827–837.
- Fovell, R.G.–Fovell, M.Y.C. (1993): Climate zones of the conterminous United States defined using cluster analysis *Journal of Climate* 6 (11): 2103–2135.
- Guo, D.–Gahegan, M.–MacEachren, A. M.–Zhou, B. (2005): Multivariate Analysis and Geovisualization with an Integrated Geographic Knowledge Discovery Approach *Cartography and Geographic Information Science* 32 (2): 113–132.
- Guo, D. (2008): Regionalization with Dynamically Constrained Agglomerative Clustering and Partitioning (REDCAP) *International Journal of Geographical Informative Sciences* 22 (7): 801–823.
- Guo, D. (2009): Flow Mapping and Multivariate Visualization of Large Spatial Interaction Data *Transaction on Visualization and Computer Graphics* 15 (6): 1041–1048.
- Guo, D. (2009a): *Greedy Optimization for Contiguity-Constrained Hierarchical Clustering*, Proc., 4<sup>th</sup> International Workshop on Spatial and Spatio Temporal Data Mining, Miami, Florida.
- Guo, D. (2010): *GraphRECAP: A Toolkit for Spatially Constrained Graph Partition* User Manual [Online]. Available: [www.SpatialDataMining.net](http://www.SpatialDataMining.net)
- Guo, D. (2010a): Flow Mapping with Graph Partitioning and Regionalization. User Manual [Online]. Available: [www.SpatialDataMining.net](http://www.SpatialDataMining.net)
- Haining, R. P.–Wise, S. M.–Blake, M. (1994): Constructing regions for small area analysis: material deprivation and colorectal cancer *Journal of Public Health Medicine* 16 (4): 429–438.
- Henderson, B. (2006): Exploring between site in water quality trends: a functional data analysis approach *Environmetrics* 17 (1): 65–80.

- Knox, P. L.–McCarthy, L. (2005): *Urbanization: An Introduction to Urban Geography* 2nd Edition, Pearson, Toronto.
- Kohonen, T. (1995): *Self-Organizing Maps* Springer, New York.
- Landré, M.–Håkansson J. (2010): Rule versus interaction Function: Evaluating Regional Aggregations of Commuting Flows in Sweden *EJTIR* 13 (1): 1–19.
- Long J.–Nelson T.–Wulder, M. (2010): Regionalization of landscape pattern indices using multivariate cluster analysis *Environmental Management* 46 (1): 134–142.
- Lotka, A.J. (1925): *Elements of Physical Biology* Williams & Willikins, Baltimore.
- Openshaw, S. (1977): A geographical solution to scale and aggregation problems in region building, partitioning and spatial modelling *Transaction of the Institute of British Geographers* 2 (4): 259–472.
- Openshaw, S.–Rao, L. (1995): Algorithms for reengineering 1991 Census Geography *Environmental and Planning* 27: 425–446.
- Osnes, K. (1999): Iterative random aggregation of small units regional measures of spatial autocorrelation of cluster localization *Statistics in Medicine* 18 (6): 707–725.
- Rain, D. R. (1999): Commuting Directionality, a Functional Measure For Metropolitan and Non Metropolitan Areas Standards *Urban Geography* 20 (8): 749–767.
- Romano, E.–Balzanella, A.–Verde, R. (2010): *A new regionalization method of spatially dependent functional data based on local variogram models: an application on environmental data* Proc., 45<sup>th</sup> Scientific Meeting of the Italian Statistical Society, Padua, Italy.
- Schwanen, T.–Dieleman, F. M.–Dijst, M. (2004): The Impact of Metropolitan Structure on Commute Behavior in the Nederland's: A Multilevel Approach *Growth and Change* 35 (3): 304–333.
- Volterra, V. (1926): *Variazioni e Fluttuazioni del Numero d'Individui in Specie Animali Conviventi* Regia Accademia dei Lincei, Roma.
- Wang, H.–Zhang, X.–Li, S.–Song, X. (2010): Spatial clustering for the regionalization of maize cultivation in China and its outlier analysis *Transactions on Informative Science and Applications* 7 (6): 860–890.