A Study of Comparative Clustering of EU Countries Using the DBSCAN and K-means Techniques within the Theoretical Framework of Systemic Geopolitical Analysis

Ilias K. Savvas

Department of Computer Science and Engineering,
T.E.I of Thessaly, Larissa, Greece
E-mail: savvas@teilar.gr

Alekos Stogiannos

Faculty of Turkish and Modern Asian Studies,
National and Kapodistrian University of Athens, Greece
E-mail: astogiannos@turkmas.uoa.gr

Ioannis Th. Mazis

Faculty of Turkish and Modern Asian Studies,
National and Kapodistrian University of Athens, Greece
E-mail: yianmazis@turkmas.uoa.gr

Abstract:

As a geographical method of analyzing power redistribution, Systemic Geopolitical Analysis (according to Ioannis Th. Mazis theoretical basis) proposes a multi-dimensional, interdisciplinary research pattern, which embraces economic, cultural, political and defensive facts. The amount of data produced combining these attributes is extremely large and complex. One of the solutions to explore and analyze this data is clustering it. In this work, two clustering algorithms were used, namely DBSCAN and the k-means techniques which both of them cluster data according to its characteristics. While DBSCAN groups data based on the minimum size of participating objects per cluster and the minimum required distance between them, k-means clusters the data objects according the pre-desired number of groups. Thus, since the two methods use different roads to group the data objects, they form different clusters but each one has its importance depending on the characteristics of the applied method. As a result, in this work a comparative study is presented.

Keywords: Systemic Geopolitics; Data Mining; MPI, Parallel K-means; DBSCAN.

Reference to this paper should be made as follows: Ilias K. Savvas, Alekos Stogiannos, and Ioannis Th. Mazis (2016) 'A Study of Comparative Clustering of EU Countries Using the DBSCAN and K-means Techniques within the Theoretical Framework of Systemic Geopolitical Analysis', *International Journal of Grid and Utility Computing*, Vol. x, No. x, pp.xxx–xxx.

Biographical notes: Ilias K. Savvas received his PhD in Computer Science at the University College of Dublin in Ireland. He is an Associate Professor at the Department of Computer Science and Engineering, T.E.I of Thessaly, Greece. His research interests include issues related to Parallel and Distributed Computing, Data Mining and High Performance Computing. He is author of a great deal of research studies published at international journals, conference proceedings as well as a book chapter.

Alekos Stogiannos received his PhD in Geopolitics at the National and Kapodistrian University of Athens. He is director the sector of Applied Geopolitics Laboratory, at the Laboratory of Geo-cultural Analyses of Greater Middle East and Turkey. His research focuses on Geopolitical Theory, German and European Geopolitics.

Ioannis Mazis is Professor of Economic Geography and Geopolitics of Middle East and Turkey at the School of Economics and Political Sciences and Chair of the Dpt of Modern Turkish and Modern Asian Studies, National Kapodistrian University of Athens. He is Chevalier de l' Ordre des Palmes Acadmiques de la Rpublique Franaise (2014). He was awarded with the High Distinction of the Egyptial Left (Tagamoo) in 2011. He has been honored by the Central Board of Jewish Communities and by the Federico II University of Naples.

1 Introduction

As a resultant of many cited definitions, Geopolitics can be minimally understood as a geographical tool of analyzing power redistribution, in the frame "power" has been defined by P. Kondylis in contradiction to "violence" [1].

Mazis, founder of Systemic Geopolitical Analysis, reflects power as resultant of four ontological discernible pillars of geopolitical influence, namely Defense, Economy, Politics and Culture, which is an analytical framework following the classical geopolitical approaches of Friedrich Ratzel, Rudolf Kjellen or Nikolas Spykman (Figure 1) [2, 3].

	Ratzel	Kjellen	Spykman	Mazis
Physical Geography	Physical Geography	Geopolitics	Physical Geography	Geographical space
Human Geography	Human Ethnography	Demographic policy	Demographic density / national Composition of population	Cultural / defensive pillar
	Economy	Economic policy	Economic structures	Economic pillar
	Society	Social policy	Ideally and values of population	Cultural pillar
	Culture	State policy	Forms of governance	Political / defensive pillar
			Complexes and biases of foreign ministers	Defensive pillar

Figure 1 Primary comparative compositional elements of Geopolitical approaches

In this framework, Mazis defines the Geopolitical analysis of a geographical System characterized by an uneven distribution of power as *the geographical method that studies, describes and predicts the attitudes and the consequences ensuing from relations between the opposing and distinct political practices for the redistribution of power as well as their ideological metaphysics, within the framework of the geographical complexes where these practices apply* [4]. The basic characteristics of the Systemic Geopolitics can be reported as follows [2, 5, 6, 7]:

- the use of geographic analysis tools in order to explore power,
- the perception of Geography as a human - centered science,
- the distinction between Geopolitics, a strict neutral and rational analysis, and Geostrategy, the biased implementation stage of the geopolitical conclusions,
- the application of strict scientific methods.

Data Mining is the process for extracting useful information from large data-sets. One of the most important techniques of data mining clustering is the k-means algorithm [8]. K-means takes as inputs the desired number of clusters k, and a dataset. It assigns data objects to the clusters according to a certain similarity measure which could vary from Euclidean distance to several others [9]. The mean value, or centroid, is a summary measure of the similarity of data objects within the same cluster. As the datasets are very large, the operation of assigning and/or reassigning data objects to their nearest centroids is very time consuming. One common method to overcome the k-means complexity is to reduce the initial dataset size by using a representative sample and then use this small sample to form the k clusters. The challenge here lies in identifying the representative sample, as the choice of the sample impacts directly on the final cluster centers. Another method is to distribute the dataset among a set of processing nodes and perform the calculation of centroids in parallel. This method follows the Single Instruction Multiple Data (SIMD)and can be implemented by using MPI.

The Message Passing Interface - MPI provides the tools to send and receive data from one node to other nodes (send and receive operations), synchronization mechanisms (barrier operation) and combining results obtained by the worker nodes (reduce operation) which is similar to the MapReduce corresponding operation [10]. In addition, MPI includes many functions to obtain network's information like the number of participating nodes, the name of each node an so on. Finally, MPI can support various virtual network topologies and peer-to-peer operations can work on both synchronous and asynchronous ways. The number of the parallel working nodes is unlimited. Therefore, MPI is able to operate on terrabytes of data and as large number of worker nodes is needed.

Another data clustering technique is the DBSCAN algorithm [11]. DBSCAN is a density based clustering technique which means that its goal is to form clusters with data object with many neighbors. Unlike k-means, DBSCAN does not require beforehand the desired number of clusters but two parameters, the minimum number of data points needed in order to form an autonomous cluster and the maximum distance between two points to be considered as neighbors. So, DBSCAN and k-means differ from each other on the way they create the clusters. This is very useful in this study since one could observe the difference of forcing countries to form clusters (k-means) and on the other hand to see how groups of countries are created depending on the density of their appearance on a multidimensional space (DBSCAN).

The main idea behind this work is to apply and compare well known techniques like the DBSCAN and k-means algorithms to Systemic Geopolitical data. Since its high complexity, a parallel version of k-means algorithm is used in this work [12, 13] while the original sequential version of DBSCAN is employed.

The rest of the paper is organized as follows: In Section 2 the theoretical base of Systemic Geopolitics is described in detail. In Section 3, an overview of

Copyright © 2015 Inderscience Enterprises Ltd.

related work and the choice of the indicators is given. The system model is presented in Section 4, while in Section 5, we describe the parallel version of k-means technique used in this study. In Section 6, a concrete example is provided in order to clarify the technique. A brief description of DBSCAN is presented in Section 7. The experimental results are presented in Section 8. Finally, some comments are discussed in Section 9, and Section 10 concludes the paper and highlights the future research directions.

2 Methodological and epistemological foundations of Systemic Geopolitics

For the elaboration of a Systemic Geopolitical Analysis, Mazis [4] suggests the following methodological approach:

Table 1 Systemic Geopolitics: methodological approach

Stages	Description
1	Decoding the title of the topic
2	Identifying the boundaries of the Geopolitical Systems under study
3	Defining the fields of influence of the Geogeopolitical factor,
4	Synthesis
5	Conclusions. (Mazis, 2015: 1063-1068)

From an epistemological point of view, the proposed form of methodological approach adopts a Lacatosian structure, which contains [4]:

- Defining of the fundamental axiomatic assumptions (elements) of the hard core of the geopolitical research programme
- Defining of the auxiliary hypotheses (elements of the protective belt of the geopolitical research project)
- The issue of the positive heuristics of the geopolitical research programme
- The elements of the positive heuristics of the geopolitical research programme.

According to the Lakatosian meta-theoretical approach, the hard core (fundamental assumptions) constitutes the basic premise of a research programme. The hard core is protected by negative heuristics, in short, by the rule that prohibits researchers to contradict the fundamental ideas of a given research programme, as an attempt to address new empirical data which tend to invalidate the theory. On this basis the following axiomatic assumptions are being formulated:

The first fundamental axiomatic assumption (element 1), which constitutes the center of the hard core of the geopolitical research programme, is that all the characteristics of the above-mentioned subspaces of the geographical complex are countable or can be counted, through the countable results which they produce, e.g., the concept of democraticity of a polity (according to western standards, since there are no other). This is a concept identified as a Geopolitical Index within the framework of the secondary causative Political Space, as defined earlier, and can be countable by means of a multitude of specific results, which it produces in the society where this form of political governance is applied. Such are for example the number of printed and electronic media in the specific society, the number of political prisoners or their absence, the level of protection of children of single-parent families, the number of reception areas for immigrants and density of the latter per m^2, etc. These figures are classified, systematized and evaluated according to their specific gravity concerning the function of the figure to be quantified, and constitute the Geopolitical Indices that are going to be present and examined in detail below.

The second fundamental axiomatic assumption (element 2) of the hard core of the systemic geopolitical programme is that, within the framework of the geographical area under study, there exist more than two consistent and homogeneous Poles which are: i) self-determined (as to what they consider gain and loss for themselves), and also in relation to their international environment; ii) hetero-determined, uniformly and identically to their international environment which is determined by the international actors that dwell within them and their common systemic relation is their characteristic, according to the Lakatosian meta-theoretical approach, a research programme has the protective belt of complementary hypotheses, i.e., proposals that are subject to control, adaptation and re-adaptation, and that are replaced when new empirical data come to light. Moreover, given Lakatos dictum that in the positive heuristic of a programme there is, right at the start, a general outline of how to build the protective belts and that a research programme [is defined] as degenerating even if it anticipates novel facts but does so in a patched-up development rather than by a coherent, pre-planned positive heuristic (Lakatos, 1971b: 125), a (provisional) definition of that protective belt for the research programme of Systemic Geopolitic Analysis should be formulaated. Consequently, following the Lakatosian metatheoretical paradigm, the protective belt of the geopolitical research programme should be defined, complemented with the following auxiliary hypotheses-elements:

(element [e1]): First auxiliary hypothesis of the protective belt of the geopolitical research programme: the size of the power is analyzed in four fundamental entities (Defence, Economy, Politics, Culture/Information), which in turn are analyzed in a number of geopolitical indices. These Geopolitical Indices (as already mentioned, are countable or can be counted) are detected and counted in the internal structures of the Poles that each time constitute the

Sub-systems of the Geographical Complexes under geopolitical analysis.

(element [e2]): Second auxiliary hypothesis of the protective belt of the geopolitical research programme: the above Poles constitute fundamental structural components of an international, and always changing, unstable System.

(element [e3]): Third auxiliary hypothesis of the protective belt of the geopolitical research programme: these Poles express social volitions or volitions of the deciding factors that characterize the international attitude of the Pole. Consequently, these poles can be national states, collective international institutions (e.g., international collective security systems, international development institutions, international cultural institutions), economic organizations of an international scope (i.e., multinational companies, bank consortia) or combinations of the above which, however, present uniformity of action within the international framework concerning their systemic functioning.

(element [e4]): Fourth auxiliary hypothesis of the protective belt of the geopolitical research programme: consists of the above-mentioned causal and causative notions of the Primary, Secondary and Tertiary Space, as well as their combinations (Complete and Special Composite Spaces).

(element [e5] Fifth auxiliary hypothesis of the protective belt of the geopolitical research programme is the premise that the international system has a completely unsure, unstable and changing structure.

(element [e6] Sixth auxiliary hypothesis of the protective belt of the geopolitical research programme: systemic geopolitical analysis aims to conclusions of practicology, in short, of some theory of practice (R. Aron), i.e., to the construction of a predictive model of the trends of power redistribution and in no case to guidelines for action under some specific national or polarized perspective. The latter is nothing but the geostrategic biased synthesis, not a geopolitical analysis. This equals the use of the results (of the model of power redistribution) of the geopolitical analysis and follows the stage of geopolitical analysis. It must be noted that the historicity of the elements of the research programme is represented by the cultural formations developing in the context of the fourth geopolitical pillar. Thus, their countability is possible in the same way as is for the rest of the geopolitical pillars that have a qualitative nature, by means of the geopolitical indices of the Cultural pillar. At this stage it should not be forgotten that replacing a set of auxiliary assumptions by another set, is an intra-programme problem shift, since only the protective belt and not the hard core is altered. The intra-programme problem shifts should be made in accordance with the positive heuristics of the problem that is with a set of suggestions or advice that function as guidelines for the development of particular theories within the programme. It should also be emphasized that a key concern of the Geopolitical Research Programme is to describe the suggestions to the researcher that will determine the content of the positive heuristics of the Programme in question. Without them, it is impossible to assess the progressivism of the Geopolitical analysis according to the necessary novel empirical content expected in our analytical spatial paradigm (model). Given these necessary clarifications concerning the elements of the positive heuristics of the geopolitical research programme, following should be defined:

1. The methodology of each theoretical approach should remain stable until a possible detection of continuous degeneration,

2. The requirement of predictive ability and the expansion of the empirical basis of the theoretical approach should be maintained,

3. The empirical facts should constitute the final measure for assessing competitive theoretical approaches of the same set [research programme],

4. The facts that have been used to test a theoretical approach should not be the only ones used for verifying this approach but, with the progress of time of research, the testing of the theoretical approach should be refered also with facts that derive from the expansion of the empirical basis of the given approach.([4]:1068- 1072)

3 Choice of indicators and related work

The diachronic and extended bibliography about both the social and political measurement procedures and the use and choice of indicators is a clear evidence of a complexity, ruling the selection process.

There is growing acceptance among policymakers and many in the social science community that two quite distinctive types of social indicators are appropriate for measuring societal and individual well-being. One type has been generally referred to 'objective' and has been characterized by hard measures describing the environments within which people live and work. Conditions can deal with health, crime, taxes, education, leisure time, voting behavior, housing, and any number of other aspects of peoples' lives. The second type of indicator of societal or individual well-being is commonly referred to as 'subjective' and is intended to describe the ways people perceive and evaluate conditions existing around them [14].

A third type of indicators, namely the Composite Indicators (CI) are getting more and more popular, since many international organizations propose their use in search of evidence based policy [15] (Nardo et al. 2008). From a formal point of view, a composite indicator is an aggregate of all dimensions, objectives, individual indicators and variables used for its construction. This implies that what defines a composite indicator is the set of properties underlying its aggregation convention [15]. The growing interest in composite indicators may be attributed to a variety of reasons, like their ability to summarize complex or multidimensional issues, in view of supporting decision-makers or the facilitation of ranking countries on complex issues. On the other hand, many objections to the use of composite indices are being declared, like the possibility of sending misleading, non-robust policy messages if they are poorly constructed or misinterpreted, or the judgement that has to be made during their construction of stages, concerning the selection of sub-indicators, choice of model, weighting indicators and treatment of missing values, etc [16].

At this early research stage the World Bank's geographical, economic, cultural, defensive[17] and political indicators, (WGI) project [18] have been basically used. Only two geographical indicators (Coastline and Shared boarder) originate from another source [19], while isolated missing indicators were taken by other sources (Table02) and in more details in Section 10 - Appendix [20].

Table 2 Systemic Geopolitics

Category	# of Indices
Geography	17
Culture	51
Defense	11
Economy	35
Politics	6
Total	**120**

With the exception of the political indexes, there have been mostly used objective indicators. Each pillar of geopolitical influence includes basic outlining indicators. In this direction the economical pillar includes 35 indicators referring to GDP and its compositional elements, the national debt, energy production and supply, trade, synthesis of labor force. The cultural pillar embraces data (51 indicators) stating some living conditions in basic sectors, like health, education, technical infrastructure or research productivity. 13 indicators referring to military expenditure, arms imports and exports as well as to migrant and refugee numbers urge the defense pillar, while the Worldwide Governance Indicators (WGI) project incorporates processed indicators regarding Voice and Accountability, Political Stability, Government Effectiveness, Control of Corruption etc. Regarding the indices it should be mentioned that in the cases of missing indicators, the average of the last 5 existing values has been used.

In the present phase the analytical level was national, but future analysis can be conducted in other administrative or political divisions. The transmission of this first attempt to an extended geographical classification, for instance on the basis of the 1342 EU regions at NUTS 3 level [21] or of the 193 state members of the UN [22], in combination with an expansion of the involved indicators, requires a very strong computational capability, which can be achieved using parallel techniques.

Cluster analysis has been so far used in many cases in order to measure various economic, cultural and social asymmetries. The authors in [23] investigated the European economic integration introducing the EU Index, an indicator measuring the extent of economic integration within the European Union. Investigating the economic activity of the EU countries, Anna Blajer-Golebiewska states that classification of economies due to their economic activities is not stable, even thgough without significant changes, while she discovers a high level of similarity of obtained clusters to geographical, historical and political classifications in each of the analyzed years (2001, 2006, 2011) [24]. Exploring the Socioeconomic Diversity of European Regions the authors in [25], implicate indices from more fields, like demography, economy, employment and education, with an equal weight, to classify the European regions into four classes for the sake of comparison with the four clusters solution proposed by the European Commission. It was shown that each of the two major groups of the EC classification - convergence regions and competitiveness and employment regions - comprises at least two significantly different groups of regions, which differ not only in terms of their average income, but also in terms of other indicators. Also, it was revealed that the two other groups - phasing-in regions and phasing-out regions -, beyond their inexpressive denomination, also seem to lack homogeneity, being spread throughout different clusters.

4 System Model

The parallel model used in this work is the Single Instruction Multiple Data (SIMD). Therefore, each computational node of the system performs the same task on different data. Data resides on the master node which is responsible to split it and transfer it to the worker nodes which in turn will independently perform the algorithm. The technique is applied on a computational cluster. A Computational Cluster ($V = \{v_0, v_2, \ldots, v_{P-1}\}$) is a collection of independent and potentially heterogeneous processing-nodes. We assume that each node v_i is autonomous, and has a full information on its own resources.

Finally, the data set D is divided into $P-1$ subsets, d_i, $D = \{d_0, d_1, \ldots, d_{P-2}\}$ of equal or almost equal size (considering that mode v_0 is the master node and no computations are performed on it). Each worker node v_i receives its data set d_i and applies the original sequential k-means. As a result, the worker nodes produce $L = (P-1) \cdot k$ lists of local centroids C_l. In addition, each centroid is calculated and assigned with the number of data points associated with it. Therefore, each node has the information $v_i = (c_l^i, n_l^i)$, where c_l^i stands for the l^{th} centroid from node v_i and n_l^i represents the total number of data points associated with the centroid c_l^i.

5 Parallel k-means

The original k-means algorithm, [8] is described in Algorithm 1 while its parallel version for 1-dimension objects is given in Algorithm 2 and for n-dimensional objects in Algorithm 3 [12, 13].

Algorithm 1 Original k-means Algorithm

Choose k points from the data set. These represent the initial centroids.
repeat
  Assign each point to its closest centroid
  Recalculate the centroids
until no centroid changes

Briefly, the first phase of parallel k-means for 1-d objects is for the master node to discover the available nodes and divide the data to equal or almost equal parts. Then, the master node transfers the data subsets to the workers which in turn, after receiving the data start to apply the original sequential k-means algorithm on it. Then, they transfer the local centroids and the number of data points assigned to each centroid and their job stops. The master node, after collecting this information, calculates the global centroids applying the weighted arithmetic mean. At the beginning, the master nodes sorts the centroids and then splits them into k sub-groups and calculates the global centroids (using the weighted arithmetic mean). The algorithm is described in Algorithm 2 [13].

Algorithm 2 Parallel k-means Algorithm for 1-d Objects

Master node, v_0 collects the number of available worker nodes, $(P-1)$
v_0: splits data set D into $D/(P-1)$ subsets
v_0: transfers data to worker nodes
for all in parallel: Worker nodes $v_i, v_i \in \{v_1, v_2, \ldots, v_{P-1}\}$ **do**
  Receive data set
  Apply k-means algorithm
  Send the local centroids (c_i) and the number of points assigned to each one of them (μ_i) to v_0
end for
Master node, v_0 receives C and M
v_0: sort the centroid list
v_0: Calculate global centroids by applying the weighted arithmetic mean

Finally, the parallel k-means for n-d objects initially applies the parallel version for 1-d objects for each one dimension separately and then the master node combines the results to the global centroids (Algorithm 3) [12].

The calculations to compute the global centroids performed by the master nodes are as follows (Equation 1):

$$C_i = \frac{\sum_{x=1}^{y} c_x n_x}{\sum_{x=1}^{y} n_x} \tag{1}$$

Algorithm 3 Parallel k-means Algorithm for n-d Objects

Master node, v_0 collects the number of available worker nodes, $(P-1)$
v_0: Splits data set D into $D/(P-1)$ subsets
v_0: Transfers data to worker nodes
for all in parallel: Worker nodes $v_i, v_i \in \{v_1, v_2, \ldots, v_{P-1}\}$ **do**
  Receive data set
  for j ← 1 to n **do**
    Apply k-means algorithm for x_j coordinates
    Send the local centroids and the number of points assigned to each one of them to v_0
  end for
end for
Master node, v_0 receives C and R
v_0: Sorts the $(P-1)\cdot n\cdot k$ received centroid lists separately
v_0: Calculates global centroids by applying the weighted arithmetic mean for each dimension
v_0: Combines the $n \cdot k$ centroids to produce the k global centroids

which represents the weighted arithmetic mean of the corresponding centroids.

The computational complexity of the algorithm is depending on the number of the participating worker nodes and the communication overhead needed to transfer the data subsets to them and can be expressed as in Equation 2 [12].

$$T_P = O(\frac{kNM_P}{P-1} + C) \tag{2}$$

where C stands for the communication overhead, and M_P represents the number of iterations for the parallel implementation. Usually, $M_P < M$, since the data set for the parallel implementation is much smaller than the whole data set. On the other hand, the complexity of the original k-means is as follows [8]:

$$T_S = O(kNM) \tag{3}$$

and it has been proven that $T_P << T_S$ [12].

6 Concrete Example

In order to clarify the technique, a simplified concrete example is given. The data set was taken from the index *Agriculture Land* (normalized into the range 0 → 1000, Malta → France) from the category *Geography*. The data represent EU countries in alphabetical order (Austria, Belgium, ...,Sweden, UK). Finally, the data set has been divided into three subsets as in Table 3.

Considering that the number of clusters is $k = 2$, the application of the sequential k-means algorithm produces two clusters for each data subset and a total of six centroids as in Table 4 where the centroids are presented

Table 3 Data subsets.

	Subset 1	Subset 2	Subset 3
1	109.26	577.68	63.53
2	45.78	282.68	503.62
3	177.35	184.81	125.77
4	45..70	156.88	476.01
5	3.98	475.85	66.50
6	146.20	63.50	16.28
7	90.66	98.23	934.82
8	32.80	4.20	105.39
9	78.91	0.00	595.65
10	1000.00		

Table 4 Centroids and number of data points.

	Centroid	# of data points
1	75.00	5
2	80.56	9
3	112.43	7
4	526.00	2
5	627.00	4
6	1000.00	1

(sorted) along with the number of data points assigned to them.

In order to calculate the global centroids which are 2, we divided the centroids into 3 subgroups (following their sorted order). Each one of them is associated with a weight, namely the number of data points assigned to it. Then the sum of the multiplication of each centroid by its corresponding weight over the total number of the data points assigned to all of them is the total centroid (weighted arithmetic mean). In this case we can see that:

$$C_1 = \frac{75 * 5 + 80.56 * 9 + 112.43 * 7}{5 + 9 + 7} = 89.86$$
$$C_2 = \frac{526 * 2 + 627 * 4 + 1000 * 1}{2 + 4 + 1} = 651.43$$

which are identical with the centroids produced applying the sequential k-means technique on the whole data set.

7 The DBSCAN algorithm

The main disadvantage of k-means is that the number of clusters has to be known beforehand, and the technique forms them without taking under consideration the distance between data points but between them and their centroids. To overcome this, another technique was tested on the same data, the DBSCAN algorithm [11]. In contradiction to k-means, DBSCAN takes as input the desired longest distance between two points in order to be considered as neighbors (known as *eps*) and the minimum number of points that can form an autonomous cluster (known as *MinPts*). In addition, DBSCAN characterizes the data points as follows:

1. Core: the points that belong to a cluster and in addition they have a larger number of neighbors than *eps*,
2. Border: the points that belong to a cluster but have a smaller number of neighbors than *MinPts*. These points belong to the cluster because at least one neighbor of them belongs to this cluster and it is a core point to the specific cluster,
3. Noise: the points that do not belong to any cluster.

A brief description of DBSCAN technique is given in Algorithm 4.

Algorithm 4 DBSCAN

for each data point $d_i \in D$ **do**
  if d_i is unvisited **then**
    Mark d_i as visited
    NumPts ← explore the rest of unvisited data points and examine how many and which of them are neighbors to d_i
    if NumPts < *MinPts* **then**
      Mark d_i as Noise
    else
      Generate new cluster containing d_i and its neighbors
      Mark d_i as Core
      Determine the rest data points that belong to the cluster and characterize them as core or border points
    end if
  end if
end for
for each data point d_i characterized as noise **do**
  Examine if d_i belongs to any cluster
  if d_i belongs to a cluster **then**
    Mark d_i as Noise
  end if
end for

In this study, the characterization of data points which represent countries as core, border or noise is extremely interesting. A core country means that its similarities with other countries of the same cluster are more that the desired number of similarities (*MinPts*), while a border country belongs that cluster just because of its similarities with only few core countries of the cluster or maybe even only one core country. On the other hand, a noise country means that there are only few (¡*minpts*) similarities with the rest of the countries.

8 Experimental Results

For the experiment of this work, 25 computational nodes were used (Pentium IV, with XUBUNTU 12.04 operating system and MPICH 3.0.4 and Ethernet 100 Mbit/s). The number of clusters k were in the range

$[2, 3, \ldots, 7]$ while the dimensionality of data was from 7-d (Politics) to 120-d objects (all indices combined together). All indices were normalized within the range $[0, 1000]$ using the following type (different methods of normalization like Principal Method Analysis or developing a composite indicator as presented in [26] are going to be examined in future work):

$$\widehat{x} = 1000\frac{x - x_{min}}{x_{max} - x_{min}}$$

Applying k-means, for each different case (i.e. $[k = 2, d = 7], [k = 2, d = 11] \ldots$) the experiment ran 100 times in order to have the most accurate results. As an outcome, the centroids chosen were those who appeared most of the times. The large number of runs used in this study was because of the importance of the results. On the other hand, DBSCAN applied in such a way to produce the same number of clusters with k-means in order to be comparable to each other. To produce this, several minimum distances and minimum number of data points to form autonomous clusters tested.

From the systemic geopolitical point of view this particular work constitutes a first attempt for the organization of a wider operational data analysis framework with the use of advanced IT and geographical tools.

Since the development of an extended, stable, indicator based analysis system is a part of the aims of this research, the fundamental aim of the present work is the comparison of the two algorithms for the aims of the Systemic Geopolitical Analysis. Parallel to this main scope, even on a basis of preselected indicators, some preliminary observations can be formulated regarding EU countries appearing in the same cluster.

8.1 Economy

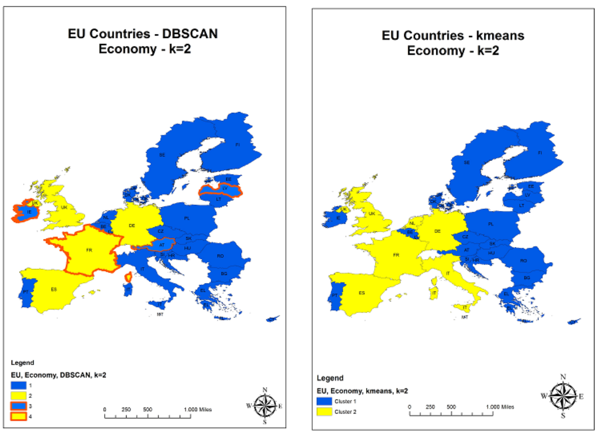

Figure 2 Economy, k=2

Both algorithms, DBSCAN and k-means, indicate on the k=2 level (Figure 2) an East - West economic polarization, with the two western crisis countries, Ireland and Portugal, being clustered with the East countries. A very critical difference appears concerning Italy: According to DBSCAN, Italy belongs to the Eastern economies, while k-means finds more similarities with the western countries.

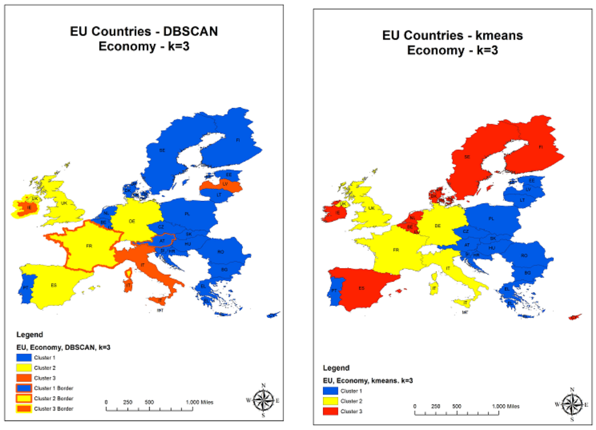

Figure 3 Economy, k=3

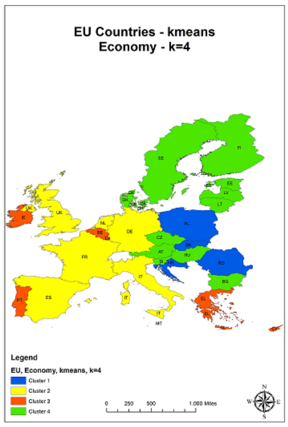

Figure 4 Economy, k=4

The same phenomenon is being observed at k=3 level (Figure 3), where Italy is according to DBSCAN algorithm again not being summarized with the big European economies, German, United Kingdom, France and Spain, but it constitutes a separate cluster with Latvia and Ireland.

Two other remarkable observations are that DBSCAN correspondences crisis country Spain always within the group of the large economies, while France always appears at the border of the same cluster. The fact, finally, that DBSCAN was unable to produce a k=4 clustering (Figure 4) must be taken into consideration for the future research, since it indicates an extended structural economic disparity within the EU.

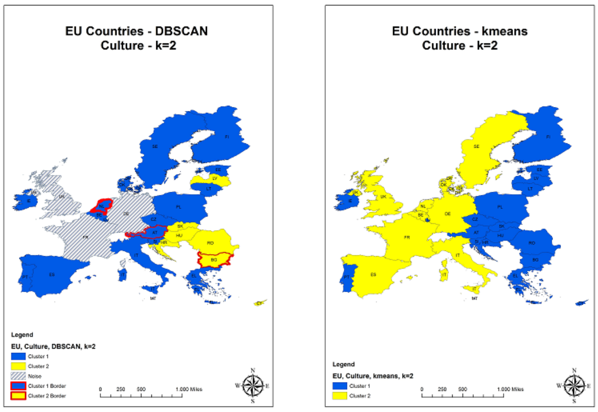

Figure 5 Culture, k=2

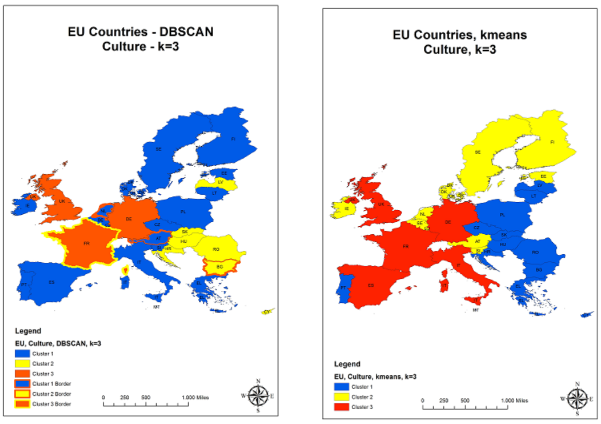

Figure 6 Culture, k=3

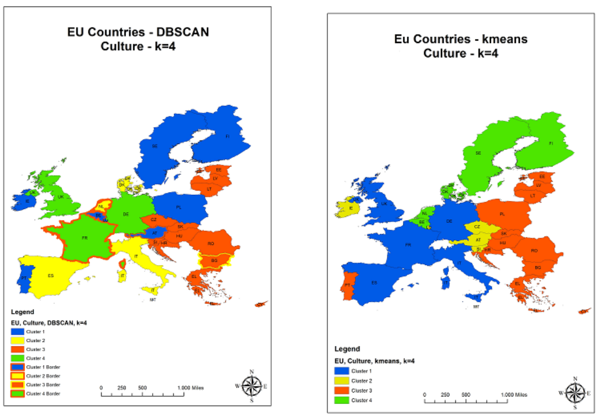

Figure 7 Culture, k=4

8.2 Culture

DBSCAN algorithm creates a small core of countries (UK, DE and FR at the border), which appears in all clusters, even though in k=2 as noise (Figure 5, while k-means creates a larger than the DBSCAN constant appearing group of 5 countries (UK, DE, FR, IT, ES) (Figure 6).

Another group of countries appearing constantly in the same DBSCAN cluster are the Balkan countries (BG, RO, HU, HR and SK). It is although remarkable, that Greece joins the Balkan grouping only at the higher analytical level of k=4 (Figure 7). That fact indicates the higher Greek living standards in comparison to the Balkan neighbors, who joined the EU later.

The East - West diversity, which was observed after the results of the Economic pillar, is also evident in the Cultural pillar clustering. In this framework, a subgroup of Balkan and countries has to be remarked, which appears clearly in every analytical level of the DBSCAN procedure, and in the k=4 clustering of k-means (Figure 7).

8.3 Politics

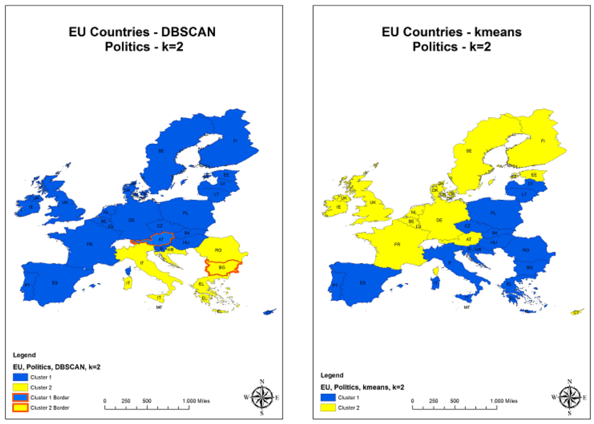

Figure 8 Politics, k=2

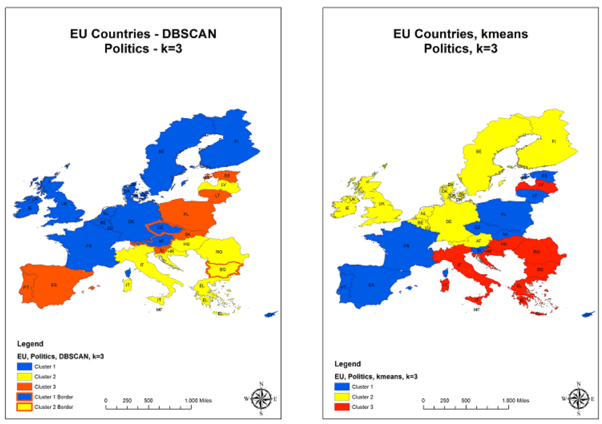

Figure 9 Politics, k=3

The widely assumed and discussed North - South division arises only as a particular clustering result of World Bank's synthetic Worldwide Governance Indicators (WGI), measuring six dimensions of governance (Figures 8,9).

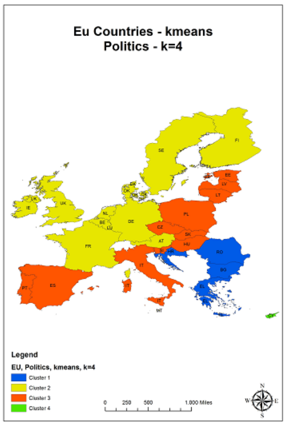

Figure 10 Politics, k=4

It is also remarkable, that in the majority of the clustering results (with the exception of k-means k=2 and k-means k=4) Italy appears as homogeneous with the Balkan countries (Figures 8, 9).

Regarding the performance of the two algorithms, it should be noted, that DBSCAN could not create a k=4 cluster. The larger scale analysis creates in both algorithms clear eyed geographical zones, located in North, South and Southeast EU.

8.4 Defence

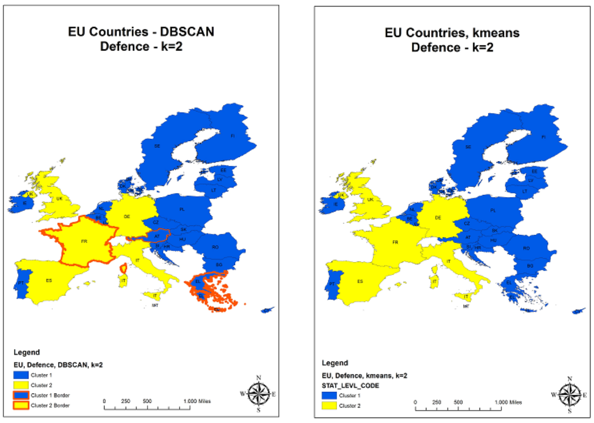

Figure 11 Defence, k=2

The k=2 results are in both algorithms identical (Figure 11), but the border indication of the DBSCAN process, with France, Greece, and Austria being at the border of their clusters, permits a more thorough analysis.

Furthermore, k-means algorithm creates a group of 5 countries, (UK G, FR ES, IT) (Figures 11, 12,13) which appear together in all clustering levels, while the DBSCAN divides this appearing at k=2 level group in the next levels (Figures 12, 13).

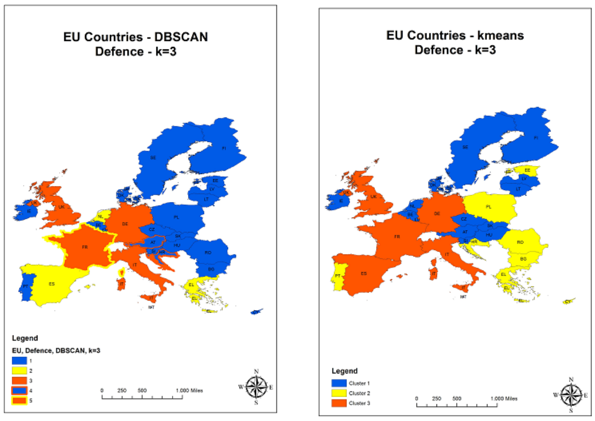

Figure 12 Defence, k=3

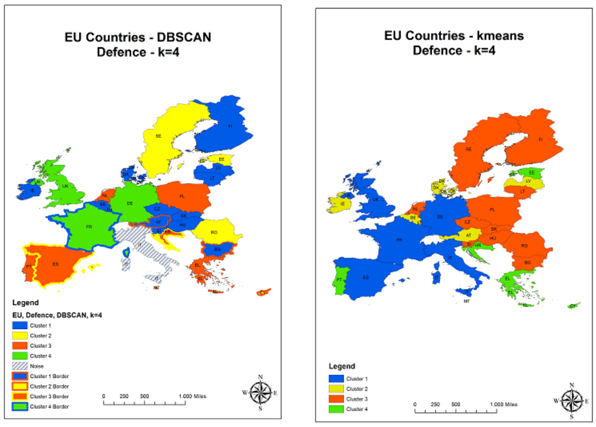

Figure 13 Defence, k=4

In general, he results of the defensive pillar indicate an East West polarization, with the exception of DBSCAN, k=4 clustering.

8.5 All indices

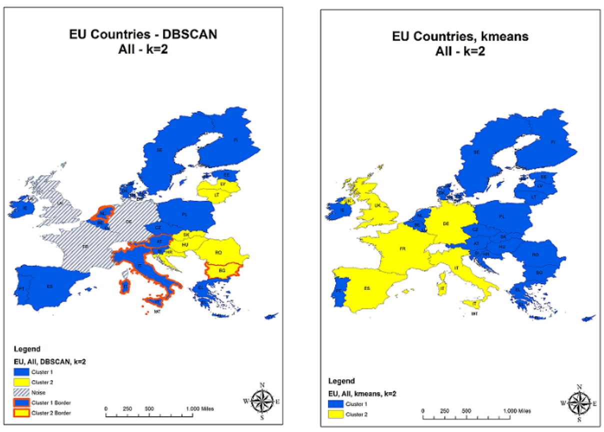

Figure 14 All indices, k=2

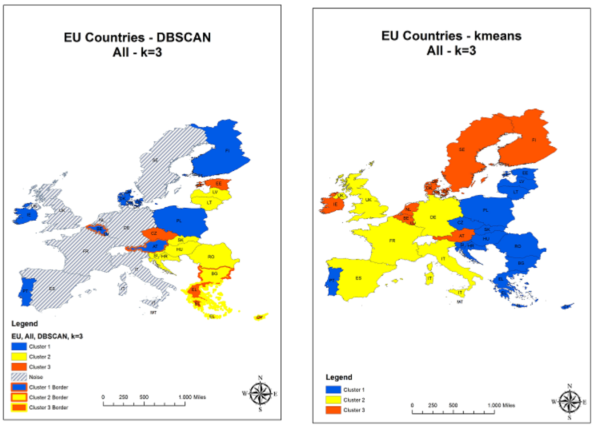

Figure 15 All indices, k=3

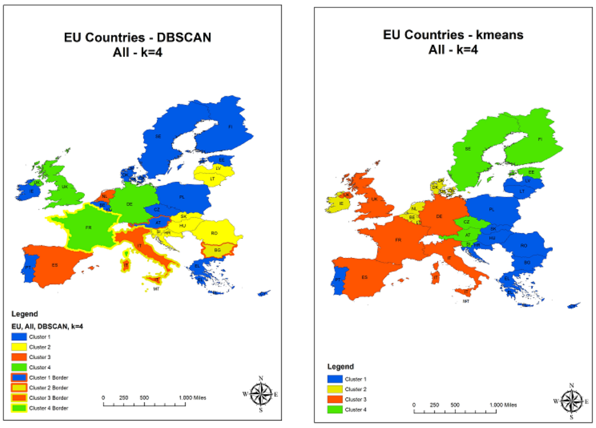

Figure 16 All indices, k=4

This cluster analysis incorporates both the indicators of all geopolitical pillars and also some geographical indicators. At this level the two algorithms provide a different grouping of the countries, which could be perceived as the “core” countries of the EU. DBSCAN algorithm classifies (as “Noise” or cluster) three countries together (United Kingdom, Germany and France at the border of the cluster) (Figures 14, 16), while k-means adds to this “core countries” (Figures 14, 15, 16) grouping constantly Italy and Spain.

In the DBSCAN clustering k=4 analysis, Spain, Italy, Netherland and Luxembourg create a separate cluster between the “core countries” and another group, consisting of countries, which contain western living standards (Sweden,Finland, Poland, Portugal, Ireland etc.) (Figure 16)

Another interesting observation is that, in k=2 and k=4 clustering, DBSCAN separates Greece from cluster consisting of the Balkan and small Baltic countries, and even in the k=3 cluster Greece stands at the border this group. (Figures 14, 15, and 16)

9 Discussion

In this research, two well known clustering algorithms were studied and examined, namely a parallel version of k-means and DBSCAN. Both techniques form clusters of data points according to certain similarities of their characteristics (attributes or else coordinates). The similarity measure used in this work was the Euclidean distance between the data objects. It is extremely interesting in how these algorithms behaved when applied to Geopolitical characteristics of E.U. countries. K-means formed clusters of countries even if their distance was large just because this distance was less than the distance form all the other centroids. This automatically implies that any country belong to a cluster regardless the distance of boundary points-countries. On the other hand, DBSCAN does not necessary place all the countries to clusters but according to their distance or else their similarity, it is possible to produce outlier or noise countries (coutries which do not belong to any cluster) and finally, another interesting property of DBSCAN is the border points. In this case, countries which belong to a cluster but topologically form the borders of the cluster. Precisely, border countries belong to the particular cluster only because their distance from even just one core country is less or equal to the desired one. In addition, the so called core countries of a cluster, represent all the principal characteristics of the cluster which in turn actually can represent and characterize the cluster itself.

From the Systemic Geopolitical point of view, a data mining method, namely clustering, can provide experts with very interesting information on similarities or differences among countries, since it permits the utilization of an extended real time data-set. Furthermore, the cartographic representation of the clustering results can reveal new regionalization processes, resulting from the exploitation of countable and undeniable data.

Finally, with regard to the European Union integration process, the cluster analysis produces already at this preliminary stage interesting results concerning the special role of each country or region.

10 Conclusions And Future Work

In this work, DBSCAN and the k-means algorithms were applied to geopolitical data. Since the large time complexity of k-means, a parallel version of it was used while the original DBSCAN technique wasbapplied. Both algorithms for each case of this study ran multiple times in order to ensure the most precise results because of their importance.

In regard to the key objective of this work, namely the comparison of the two algorithms, the DBSCAN algorithm seems to be more applicable for the targets of the Systemic Geopolitical Analysis, since it offers more thoroughgoing qualitative data, namely the border positioning within a cluster and the possibility that some countries might appear as "Noise". Another fact, which pleads for the use of DBSCAN is the observation, that DBSCAN produces a more rational transition from the lower to the higher clustering levels, meaning a transition, which permits the extraction of logical connections among the cluster levels.

This research constitutes a first attempt for the organization of a wider operational data analysis system, in the framework of the Systemic Geopolitical Analysis theory. Although the development of an extended, stable, indicator based analysis system is a part of the aims of this research, following preliminary remarks, useful as future working assumptions can be formulated:

- the presented results bring evidence of an East - West polarization among the EU Countries, in contradiction to the widespread opinion of a North - South division.
- the aforementioned North - South clustering arises as a particular result of indicators on politics, based on the Worldwide Governance Indicators (WGI) measuring six dimensions of governance. This contradictory result affiliates with the comments in chapter 3, regarding the operative choice of indicators.
- The cluster analysis indicates roughly four main groups of countries, appearing frequently together This grouping could at a later research level lead to a grouping based on a power classification scale from 1-4:
    1. High: including states with representative western living standards and the most decisive role within the EU.
    2. High - medium: including states with representative western living standards and a medium assertiveness within the EU.
    3. Low - medium: including states with representative western living standards, but a low assertiveness within the EU.
    4. Low: including states without western living standards, and a low assertiveness within the EU.

In this framework, future work will focus on the following aspects:

1. expansion and gradual standardization of the indicators urging the 4 pillars of geopolitical influence,
2. assessment of qualitative characteristics of the emerging clusters,
3. development of a dynamic analytical framework through the comparison of long term data,
4. research on the special function of each pillar and every single indicator within a specific geographical complex and time period,
5. regional expansion of the research, and
6. development of an on line clustering system.

References

[1] Panagiotis Kondylis. *Theorie des Kriege (Theory of War)*. Greek Edition, Themelio, 2 edition, 1998.

[2] Ioannis Th. Mazis. *Geopolitics. Theory and Practice.* ELIAMEP and Papazisis Publications. Athens, 2002.

[3] Alexandros Stogiannos. *A modern interpretation of Friedrich Ratzel and the myth of geodeterminism. The case of the Eastern Question.* PhD thesis, University of Athens, November 2012.

[4] Ioannis Th. Mazis. *Dissertationes academicae geopoliticae.* Papazisis Publications, 2015.

[5] Ioannis Th. Mazis. Writing methodology of a geopolitical analysis. structure, concepts and terms. *Geostrategic Review*, 12, 2007.

[6] Ioannis Th. Mazis. *Post - theoretical Review of International Relations and Geopolitics.* Papazisi Publications. Athens, 2012.

[7] Ioannis Th. Mazis. L'analyse geopolitique systemique: Propositions terminologiques et definitions metatheoriques selon l'exigence metatheorique lakatienne. *Geographies, Geopolitiques et Geostrategies Regionales*, 1(1):21–32, 2013.

[8] J. B. MacQueen. Some methods for classification and analysis of multivariate observations. In University of California Press, editor, *5th Berkeley Symposium on Mathematical Statistics and Probability*, pages 281–297, 1967.

[9] S. Garruzzo and D. Rosaci. Agent clustering based on semantic negotiation. *ACM Transactions on Autonomous and Adaptive Systems*, 3(2):1–39, 2008.

[10] MPICH. Message Passing Interface. `http://www.mpich.org/`. [Online; accessed 2014].

[11] Martin Ester, Hans-Peter Kriegel, Jiirg Sander, and Xiaowei Xu. A density-based algorithm for discovering clusters in large spatial databases with noise. In *2nd International Conference on Knowledge Discovery and Data Mining (KDD-96)*, pages 226–231, 1996.

[12] Ilias K. Savvas and Georgia N. Sofianidou. A novel near-parallel version of k-means algorithm for n-dimensional data objects using mpi. *International Journal of Grid and Utility Computing*, accepted, tpb 2015.

[13] Ilias K. Savvas and Georgia N. Sofianidou. Parallelizing k-means algorithm for 1-d data using mpi. In *23rd Int. IEEE Conference WETICE*, pages 179–184, 2014.

[14] Trevor Lee and Robert W. Marans. Objective and subjective indicators: Effects of scale discordance on interrelationships. *Social Indicators Research*, 8(1):47–48, 1980.

[15] Andrea Saltelli. Composite indicators between analysis and advocacy. *Social Indicators Research*, 81(1):69, 2007.

[16] Giuseppe Munda. Choosing aggregation rules for composite indicators. *Social Indicators Research*, 109(3):337–338, 2012.

[17] World Bank. Countries and Economies. `http://data.worldbank.org/country`. [Online; accessed 2015].

[18] World Bank. Worldwide Governance Indicators. `http://info.worldbank.org/governance/wgi/index.aspx#home`. [Online; accessed 2015].

[19] GFP. Total Coastline Coverage by Country. `http://info.worldbank.org/governance/wgi/index.aspx#home`. [Online; accessed 2015].

[20] WMR. World macroeconomic research. `http://kushnirs.org/macroeconomics/index.html`. [Online; accessed 2015].

[21] European Commission. EuroStat. `http://ec.europa.eu/eurostat/web/main`. [Online; accessed 2015].

[22] United Nations. Member States. `http://www.un.org/en/members/growth.shtml`. [Online; accessed 2015].

[23] Jorg Konig and Renate Ohr. Different efforts in european economic integration: Implications of the eu index. *JCMS: Journal of Common Market Studies*, 51(6):1074–1090, 2013.

[24] Anna Blajer-Golebiewska. Economic activity-based cluster analysis of european union countries. *World Review of Business Research*, 4(1):48–61, 2014.

[25] Cristina del Campo, Carlos M. F. Monteiro, and Joao Oliveira Soares. The Socioeconomic Diversity of European Regions (Working Paper, Series 131, Center for European Studies). `http://aei.pitt.edu/9029/1/DelCampo.pdf`. [Online; accessed 2015].

[26] Giuseppe Munda and Michela Nardo. Noncompensatory/nonlinear composite indicators for ranking countries: a defensible setting. *Applied Economics*, (41):1513–1523, 2009.

Appendix

In the following tables, the indicators used are given along with their source and indicator number (first column), their name (second column), and finally their code (last column).

Table 5 Geography

Indicator Nr.	Indicator Name	Indicator Code
WB/5	Agricultural land (sq. km)	AG.LND.AGRI.K2
WB/6	Agricultural land (% of land area)	AG.LND.AGRI.ZS
WB/13	Forest area (sq. km)	AG.LND.FRST.K2
WB/17	Land area (sq. km)	AG.LND.TOTL.K2
WB/23	Surface area (sq. km)	AG.SRF.TOTL.K2
WB/440	Population density (people per sq. km of land area)	EN.POP.DNST
WB/1226	Population ages 0-14 (% of total)	SP.POP.0014.TO.ZS
WB/1227	Population ages 15-64 (% of total)	SP.POP.1564.TO.ZS
WB/1228	Population ages 65 and above (% of total)	SP.POP.65UP.TO.ZS
WB/1232	Population growth (annual %)	SP.POP.GROW
WB/1235	Population, total	SP.POP.TOTL
WB/1240	Rural population	SP.RUR.TOTL
WB/1241	Rural population growth (annual %)	SP.RUR.TOTL.ZG
WB/1243	Urban population growth (annual %)	SP.URB.GROW
WB/1244	Urban population	SP.URB.TOTL
(1)	Coastline (km)	
(1)	Shared border (km)	
(1) →	http://www.globalfirepower.com/coastline-coverage.asp	

Table 6 Politics

Indicator Nr.	Indicator Name	Indicator Code
WB	Control of Corruption: Percentile Rank	CC.PER.RNK
WB	Government Effectiveness: Percentile Rank	GE.PER.RNK
WB	Political Stability and Absence of Violence/Terrorism: Estimate	PV.EST
WB	Political Stability and Absence of Violence/Terrorism: Percentile Rank	PV.PER.RNK
WB	Rule of Law: Percentile Rank	RL.PER.RNK
WB	Voice and Accountability: Percentile Rank	VA.PER.RNK

Table 7 Culture

Indicator Nr.	Indicator Name	Indicator Code
WB/502	Research and development expenditure (% of GDP)	GB.XPD.RSDV.GD.ZS
WB/599	Scientific and technical journal articles	IP.JRN.ARTC.SC
WB/600	Patent applications, nonresidents	IP.PAT.NRES
WB/601	Patent applications, residents	IP.PAT.RESD
WB/629	Quality of port infrastructure, WEF (1=extremely underdeveloped to 7=well developed and efficient by international standards)	IQ.WEF.PORT.XQ
WB/644	Roads, total network (km)	IS.ROD.TOTL.KM
WB/645	Railways, goods transported (million ton-km)	IS.RRS.GOOD.MT.K6
WB/646	Railways, passengers carried (million passenger-km)	IS.RRS.PASG.KM
WB/647	Rail lines (total route-km)	IS.RRS.TOTL.KM
WB/648	Liner shipping connectivity index (maximum value in 2004 = 100)	IS.SHP.GCNW.XQ
WB/649	Container port traffic (TEU: 20 foot equivalent units)	IS.SHP.GOOD.TU
WB/650	Motor vehicles (per 1,000 people)	IS.VEH.NVEH.P3
WB/651	Passenger cars (per 1,000 people)	IS.VEH.PCAR.P3
WB/652	Vehicles (per km of road)	IS.VEH.ROAD.K1
WB/653	Mobile cellular subscriptions	IT.CEL.SETS
WB/654	Mobile cellular subscriptions (per 100 people)	IT.CEL.SETS.P2
WB/655	Telephone lines	IT.MLT.MAIN
WB/656	Telephone lines (per 100 people)	IT.MLT.MAIN.P2
WB/657	Fixed broadband Internet subscribers	IT.NET.BBND
WB/658	Fixed broadband Internet subscribers (per 100 people)	IT.NET.BBND.P2
WB/659	Secure Internet servers	IT.NET.SECR
WB/660	Secure Internet servers (per 1 million people)	IT.NET.SECR.P6
WB/661	Internet users (per 100 people)	IT.NET.USER.P2
WB/897	Primary completion rate, total (% of relevant age group)	SE.PRM.CMPT.ZS
WB/901	Pupil-teacher ratio, primary	SE.PRM.ENRL.TC.ZS
WB/914	School enrollment, primary, private (% of total primary)	SE.PRM.PRIV.ZS
WB/923	Repeaters, primary, total (% of total enrollment)	SE.PRM.REPT.ZS
WB/927	Primary education, teachers	SE.PRM.TCHR
WB/942	Secondary education, general pupils	SE.SEC.ENRL.GC
WB/944	Pupil-teacher ratio, secondary	SE.SEC.ENRL.TC.ZS
WB/945	Secondary education, vocational pupils	SE.SEC.ENRL.VO
WB/959	Repeaters, secondary, total (% of total enrollment)	SE.SEC.REPT.ZS
WB/970	Public spending on education, total (% of government expenditure)	SE.XPD.TOTL.GB.ZS
WB/971	Public spending on education, total (% of GDP)	SE.XPD.TOTL.GD.ZS
WB/973	Proportion of seats held by women in national parliaments (%)	SG.GEN.PARL.ZS
WB/985	Number of infant deaths	SH.DTH.IMRT
WB/987	Number of under-five deaths	SH.DTH.MORT
WB/1005	Hospital beds (per 1,000 people)	SH.MED.BEDS.ZS
WB/1052	Health expenditure per capita (current US$)	SH.XPD.PCAP
WB/1054	Health expenditure, private (% of GDP)	SH.XPD.PRIV.ZS
WB/1055	Health expenditure, public (% of total health expenditure)	SH.XPD.PUBL
WB/1056	Health expenditure, public (% of government expenditure)	SH.XPD.PUBL.GX.ZS
WB/1057	Health expenditure, public (% of GDP)	SH.XPD.PUBL.ZS
WB/1058	Health expenditure, total (% of GDP)	SH.XPD.TOTL.ZS
WB/1211	Birth rate, crude (per 1,000 people)	SP.DYN.CBRT.IN
WB/1212	Death rate, crude (per 1,000 people)	SP.DYN.CDRT.IN
WB/1218	Life expectancy at birth, total (years)	SP.DYN.LE00.IN
WB/1220	Fertility rate, total (births per woman)	SP.DYN.TFRT.IN
WB/1284	Food imports (% of merchandise imports)	TM.VAL.FOOD.ZS.UN
WB/1311	Food exports (% of merchandise exports)	TX.VAL.FOOD.ZS.UN
WB/1340	Intentional homicides (per 100,000 people)	VC.IHR.PSRC.P5

Table 8 Defense

Indicator Nr.	Indicator Name	Indicator Code
WB/674	Arms imports (SIPRI trend indicator values)	MS.MIL.MPRT.KD
WB/675	Armed forces personnel, total	MS.MIL.TOTL.P1
WB/676	Armed forces personnel (% of total labor force)	MS.MIL.TOTL.TF.ZS
WB/678	Military expenditure (% of GDP)	MS.MIL.XPND.GD.ZS
WB/679	Military expenditure (% of central government expenditure)	MS.MIL.XPND.ZS
WB/680	Arms exports (SIPRI trend indicator values)	MS.MIL.XPRT.KD
WB/1197	Net migration	SM.POP.NETM
WB/1198	Refugee population by country or territory of asylum	SM.POP.REFG
WB/1199	Refugee population by country or territory of origin	SM.POP.REFG.OR
WB/1200	International migrant stock, total	SM.POP.TOTL
WB/1201	International migrant stock (% of population)	SM.POP.TOTL.ZS

Table 9 Economy

Indicator Nr.	Indicator Name	Indicator Code
WB/385	Energy imports, net (% of energy use)	EG.IMP.CONS.ZS
WB/386	Alternative and nuclear energy (% of total energy use)	EG.USE.COMM.CL.ZS
WB/506	Central government debt, total (% of GDP)	GC.DOD.TOTL.GD.ZS
WB/713	Gross national expenditure (current US$)	NE.DAB.TOTL.CD
WB/717	Gross national expenditure (% of GDP)	NE.DAB.TOTL.ZS
WB/718	Exports of goods and services (current US$)	NE.EXP.GNFS.CD
WB/740	Gross capital formation (% of GDP)	NE.GDI.TOTL.ZS
WB/741	Imports of goods and services (current US$)	NE.IMP.GNFS.CD
WB/747	External balance on goods and services (current US$)	NE.RSB.GNFS.CD
WB/751	Trade (% of GDP)	NE.TRD.GNFS.ZS
WB/752	Agriculture, value added (current US$)	NV.AGR.TOTL.CD
WB/758	Manufacturing, value added (current US$)	NV.IND.MANF.CD
WB/764	Industry, value added (current US$)	NV.IND.TOTL.CD
WB/775	Services, etc., value added (current US$)	NV.SRV.TETC.CD
WB/819	GDP (current US$)	NY.GDP.MKTP.CD
WB/827	GDP per capita (current US$)	NY.GDP.PCAP.CD
WB/1086	Employment in agriculture (% of total employment)	SL.AGR.EMPL.ZS
WB/1099	Self-employed, total (% of total employed)	SL.EMP.SELF.ZS
WB/1111	Wage and salaried workers, total (% of total employed)	SL.EMP.WORK.ZS
WB/1119	Employment in industry (% of total employment)	SL.IND.EMPL.ZS
WB/1129	Employment in services (% of total employment)	SL.SRV.EMPL.ZS
WB/1147	Labor force participation rate, total (% of total population ages 15-64) (modeled ILO estimate)	SL.TLF.ACTI.ZS
WB/1159	Part time employment, total (% of total employment)	SL.TLF.PART.ZS
WB/1162	Labor force with primary education (% of total)	SL.TLF.PRIM.ZS
WB/1165	Labor force with secondary education (% of total)	SL.TLF.SECO.ZS
WB/1168	Labor force with tertiary education (% of total)	SL.TLF.TERT.ZS
WB/1170	Labor force, total	SL.TLF.TOTL.IN
WB/1173	Unemployment, youth male (% of male labor force ages 15-24) (national estimate)	SL.UEM.1524.MA.NE.ZS
WB/1179	Long-term unemployment (% of total unemployment)	SL.UEM.LTRM.ZS
WB/1193	Unemployment, total (% of total labor force) (national estimate)	SL.UEM.TOTL.NE.ZS
WB/1247	International tourism, number of arrivals	ST.INT.ARVL
WB/1248	International tourism, number of departures	ST.INT.DPRT
WB/1305	Commercial service imports (current US$)	TM.VAL.SERV.CD.WT
WB/1332	Commercial service exports (current US$)	TX.VAL.SERV.CD.WT
WB/1333	High-technology exports (current US$)	TX.VAL.TECH.CD