

# HOMOGENEOUS CLUSTERING OF BRAZILIAN MUNICIPALITIES BASED ON STRUCTURING HEALTH CARE THROUGH CONCEPT MAPS

## ABSTRACT

This work aims to apply a new approach to support the formulation and structuring of health care in Brazilian municipalities using concept maps and data mining. The analysis of the results and conclusions can provide support for public policies through the determination of indicators of municipal performance in the future. Homogeneous clustering generates adequate sets for the application of benchmarking methods that compare data from several decision-making units.

**KEYWORDS:** Data Mining, Concept Map and Health Care

## INTRODUCTION

According to the Federal Constitution of 1988, health is a responsibility of the government and the right of everyone, guaranteed by social and economic policies aimed at universal and equal access. It is therefore incumbent upon the public authorities to control and regulate health care as well as enforce the rules on the health care provided by public and private entities, particularly through the Unified Health System (SUS). Examples of national scope are the implementation of actions and programs such as UPA24H (emergency care units), Dengue Control, Smile Brazil (better oral health for all), Popular Pharmacy (low-cost drugs), SAMU 192 (mobile urgent care service) and Family Health, among others.

The Unified Health System (SUS), which serves the great majority of the Brazilian population, is financed with funds from the social security budget of the federal, state, Federal District and municipal governments, as well as other sources. SUS has the following

guidelines: community participation, comprehensive care, with priority given to preventive activities, without prejudice to assistance services, as well as decentralization, with single direction in each sphere of government.

The tasks of SUS are defined by the Basic Health Law (Law 8,080/90), at the three levels of government, which consolidates decentralization. Each government level has its attributions regarding system planning for municipalities, public consortium formation, implementation of health services of various natures and supervision of private health services.

The Federal Constitution of 1988 states the attributions of the Unified Health System (SUS) as:

- Ordering the training of human resources in the health area;
- Controlling and monitoring procedures, products and substances of interest to health;
- Supervising and inspecting food, including the control of their nutritional content, as well as drinks and water for human consumption;
- Participating in the formulation of policies and implementation of basic sanitation actions;
- Participating in the production of medicines, equipment, blood products and other inputs;
- Performing the actions of health surveillance and epidemiology, as well as protecting workers' health;
- Encouraging scientific and technological developments;
- Collaborating in the protection of the environment, including a healthy workplace; and
- Participating in the control and monitoring of the production, transport, custody and use of psychoactive substances and radioactive and toxic products.

However, there is a blatant disregard of the Federal Constitution because the resources devoted to health are diverted by corruption and lack of information on financial matters. The control structure of SUS funds is minimal and the delay in controlling shares of

the money spent on health by the CGU (Office of the Comptroller General), TCU (Federal Audit Tribunal) and MPF (Federal Prosecutor's Office) makes it difficult to recover misappropriated or misspent money in cases disclosed by the press, such as the overpricing of ambulance acquisition and rigging of public tenders. Added to these problems is widespread waste of resources, such as failure to use drugs before their expiry dates.

There are proposals to promote further investments in public health, such as new taxes for health expenses, but the tax burden in Brazil is already very high. An alternative solution is the regulation of Constitutional Amendment 29, which defines how municipal, state and the federal governments must share responsibility for investments in health. All public policies aimed at further investments in health care should tackle inefficiency, promote professional management using performance targets and promote good business-like management.

This work presents a strategic approach to the formulation and structuring of health care systems, taking advantage of the large database comprising 5,565 Brazilian municipalities. First we use concept maps to represent the many factors that characterize the complex issues involved in health care management in Brazil. Then we investigate heterogeneity and apply a data mining technique to find appropriate clusters.

This work is organized in four sections, as follows: introduction, methods, results and discussions. Section 1 outlines the case study to give an overview of public health in Brazil. Section 2 presents an overview about the formulation and structuring of the case study through the concept map method and data mining technique. Then, section 3 presents the results obtained with concept map and the implementation of a data mining clustering technique and section 4 presents the conclusion.

## METHODS

According to Rosenhead and Mingers<sup>1</sup>, it is important to make explicit the search for proposals to deal with complex problems and not just solve a simplified part of a problem under a particular perspective. In this improved approach, the structuring of matters, issues and situations is one of the stages of the modeling at the very beginning of the decision-making process.

As defined by Okada et al. <sup>2</sup>, concept maps are graphical tools used for knowledge representation, so that two concepts can be connected via a linking phrase, generating a proposition. Novak<sup>3</sup> presents concept mapping as a useful tool for student learning at different levels, as well as to assist in troubleshooting problems in organizations.

A graphical representation can be more effective than text for the communication of complex contents because the mental processing of images can be less cognitively demanding than the verbal processing of a text Vekiri<sup>4</sup>. Many guidelines on the formulation of concept maps can be found in the literature, as in Moreira<sup>5</sup> and Ruiz-Moreno et al.<sup>6</sup>.

We developed the formulation and structuring of public health in Brazilian municipalities using information from public health experts reported in legal documents and press interviews with representatives government entities – Federal Prosecutor’s Office; National Health Council; National Congress; Ministry of Science, Technology and Innovation; Office of the Comptroller General; Policy Council and Chamber for Management, Performance and Competitiveness (both linked to the Office of the President); and SUS Audit Department – as well as the National SUS Auditors Union and of nongovernmental organizations and health industry associations, such as the Brazilian Collective Health Association.

We used a computational tool called CmapTools, version 4.11, developed by the Institute for Human and Machine Cognition (IHMC) of the University of West Florida,

which allows users to construct, navigate, share and criticize knowledge models represented as concept maps.

The conceptual map depicts a picture of all intervening factors that can be used in a formal model. Here these factors are qualitative concepts of a legal and organizational nature, which can characterize differences in municipal health management.

Several analytic methods are based on the identification and comparison of units operating under similar conditions, thus requiring a previous method for obtaining homogeneous clusters according to pre-specified criteria. Therefore, the main focus of this paper is to apply data mining in connection with the mapping technique.

According to Weis and Indurkha<sup>7</sup>, data mining is the search for valuable information on large databases. It is a cooperative effort between humans and computers. People design databases, describe problems and define goals. Computers evaluate data and look for patterns that match the goals established by people.

To implement the clustering technique, we employed the Profile of Brazilian Municipalities Database (MUNIC), from the Brazilian Institute of Geography and Statistics (IBGE), covering the year 2009. It is a survey of basic municipal information published annually by IBGE with information about 16 themes of 5,565 Brazilian municipalities, such as human resources, legislation and municipal planning tools, education, culture, sports, housing, transportation, health and environment, among others.

The goal of the MUNIC 2009 is to build a comprehensive profile of the country's municipalities, from the management of their public administrations, which provides information with national coverage to support municipal planning and efficient implementation of policies.

The software used for clustering was the Waikato Environment for Knowledge Analysis (WEKA), a set of algorithms and programs produced by the University of Waikato

in New Zealand, implemented in the Java language, which has as main characteristic its portability. This way it can be used in different operating systems. WEKA is available at <http://www.cs.waikato.ac.nz/ml/weka> under GPL license. For the purposes of this work, we restricted the qualitative variables to the health theme of the MUNIC 2009, namely: existence of Municipal Council Fund and Municipal Health Plan; Health Units (maternity care center, maternity care center with civil registry of birth, emergency unit, clinical analysis laboratory, Popular Pharmacy Program and community health agent) and Family Health Program. A particular kind of variable accounts for the institutional integration among the three governmental spheres and the private sector, comprising public consortiums between municipalities or with the state and/or federal governments, agreements for partnership with the private sector and support from the private sector and communities.

Some concepts of the conceptual map can be associated to quantitative variables and also suggest using a management tool that can be used in hard operations research methods, such as data envelopment analysis (DEA).

Data envelopment analysis has been used in the calculation of performance indicators and to establish benchmarks for regulation of public sectors. The method lends itself to use in multidisciplinary regulatory issues in the estimation of production frontier functions. It can also incorporate expert judgments, like a multicriteria method.

Scientific methods for measurement and follow-up of regulated health firms, in particular DEA, have been developed and are being used by regulatory agencies in several countries (Chilingerian and Sherman<sup>8</sup>; Sherman and Zhu<sup>9</sup>; Ozcan<sup>10</sup>). The first practical works using data envelopment analysis for health assessment developed by Sherman<sup>11</sup> and Nunamaker<sup>12</sup> studying the performance of American hospitals and nursing homes respectively.

In DEA, homogeneous productive units are called decision-making units (DMUs) and must be comparable and act under the same conditions, so differentiation occurs only in the intensity. Therefore, it is very important to ensure the homogeneity of the data. According to Ceretta and Niedearauer<sup>13</sup> and Badin<sup>14</sup>, different orders of magnitude in DEA can be handled through a grouping technique, also known as cluster analysis, which we use here.

The development of new technologies and methodological tools to guide decision-making on the allocation of public health resources is particularly meaningful when combining methods from "soft" and "hard" operational research.

## RESULTS

The resulting formulation and structuring of public health in Brazilian municipalities is consolidated in the concept maps shown in Figures 1 through 6. Note that the original concept map is divided into six parts in order to permit visualization and interpretation of the concepts, linking phrases and existing connectors.

The concepts used in data mining are depicted with gray background. They correspond to available qualitative variables selected for use in the data mining technique. The concepts with a black background correspond to quantitative variables and also suggest using a management tool that can be used in hard operations research methods, such as data envelopment analysis (DEA).

It is worth noting that the map gathers concepts, linking phrases and connectors that belong to four categories of phenomena, namely:

- Systemic dysfunction: characterized by high volume of diverted public money;
- Legal: represented by the legislation on healthcare in Brazil;
- Management: indicates the urgent need of professionalization of Brazilian health management; and
- Consequences: shows failure to satisfy the fundamental right to decent health care.



Figure 1- Concept Map of Brazilian Public Health (Part 1)



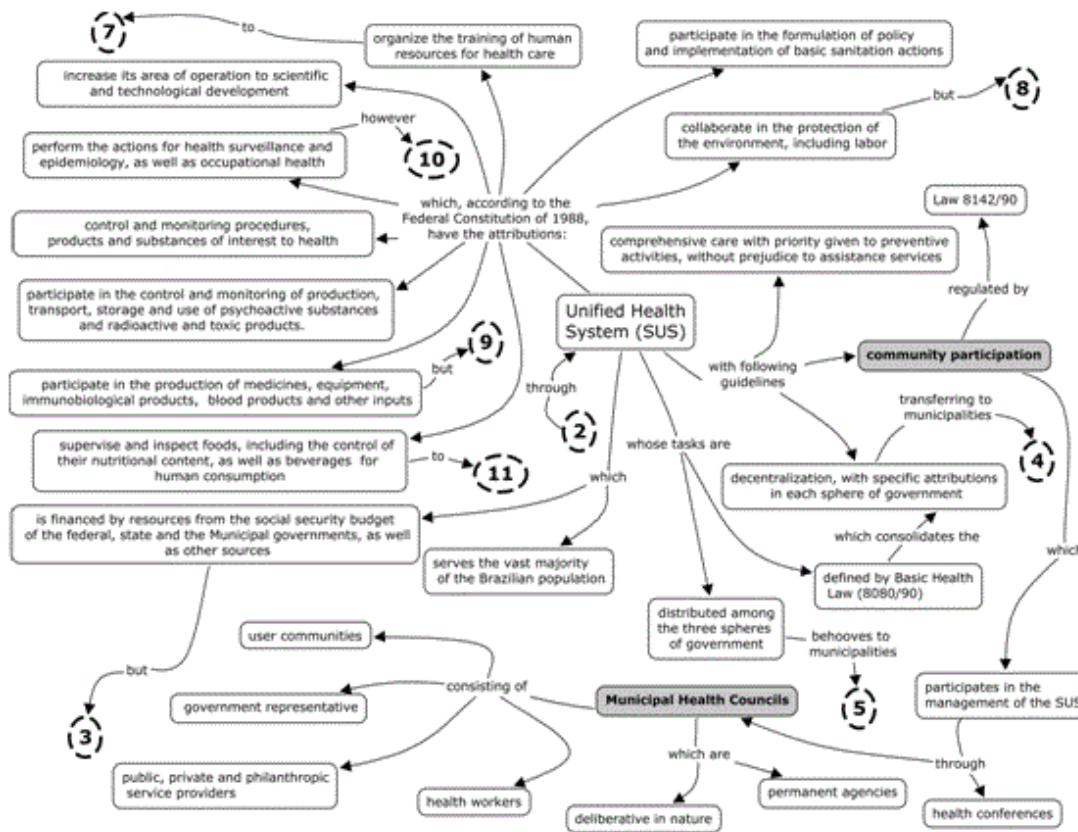


Figure 2- Concept Map of Brazilian Public Health (Part 2)

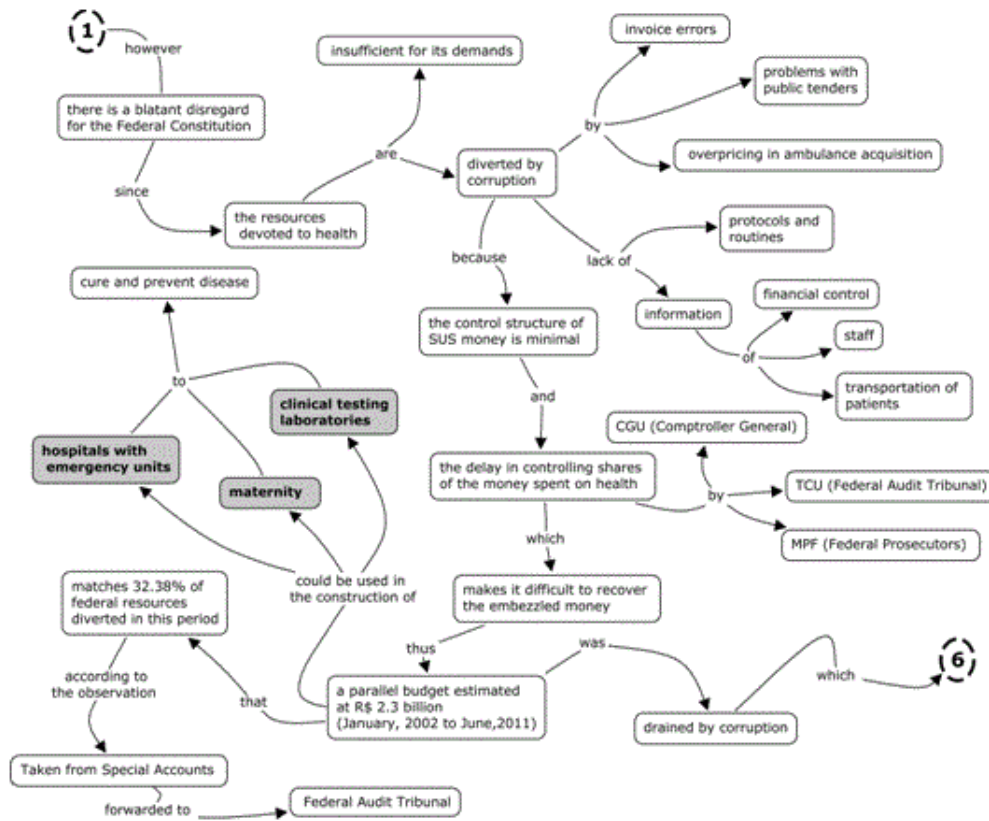


Figure 3- Concept Map of Brazilian Public Health (Part 3)

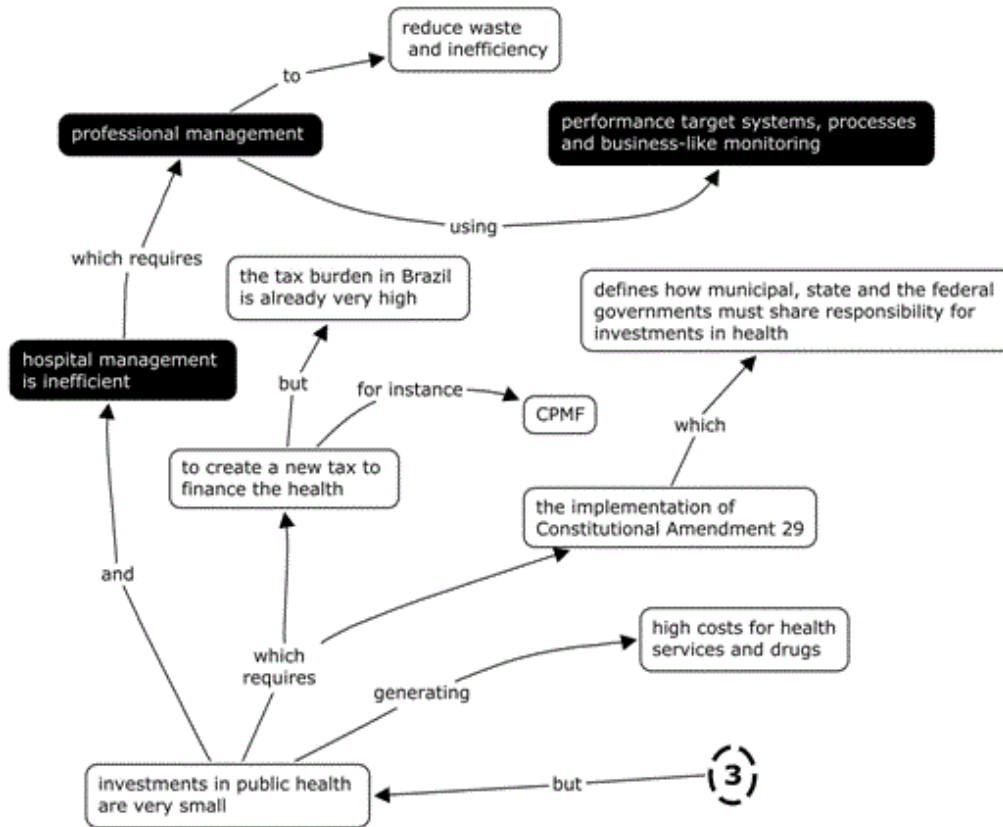


Figure 4- Concept Map of Brazilian Public Health (Part 4)

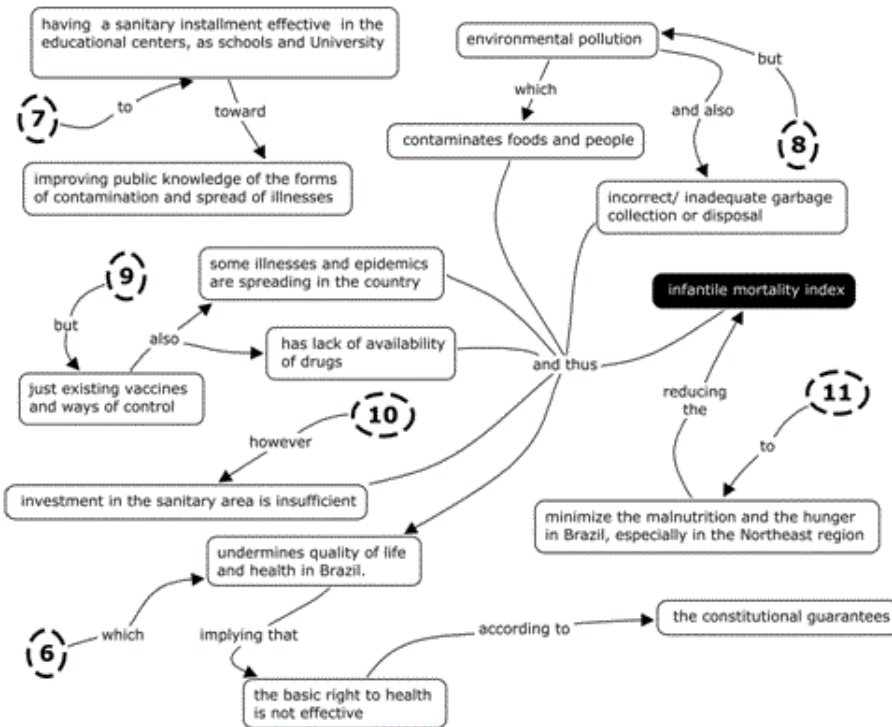


Figure 5- Concept Map of Brazilian Public Health (Part 5)

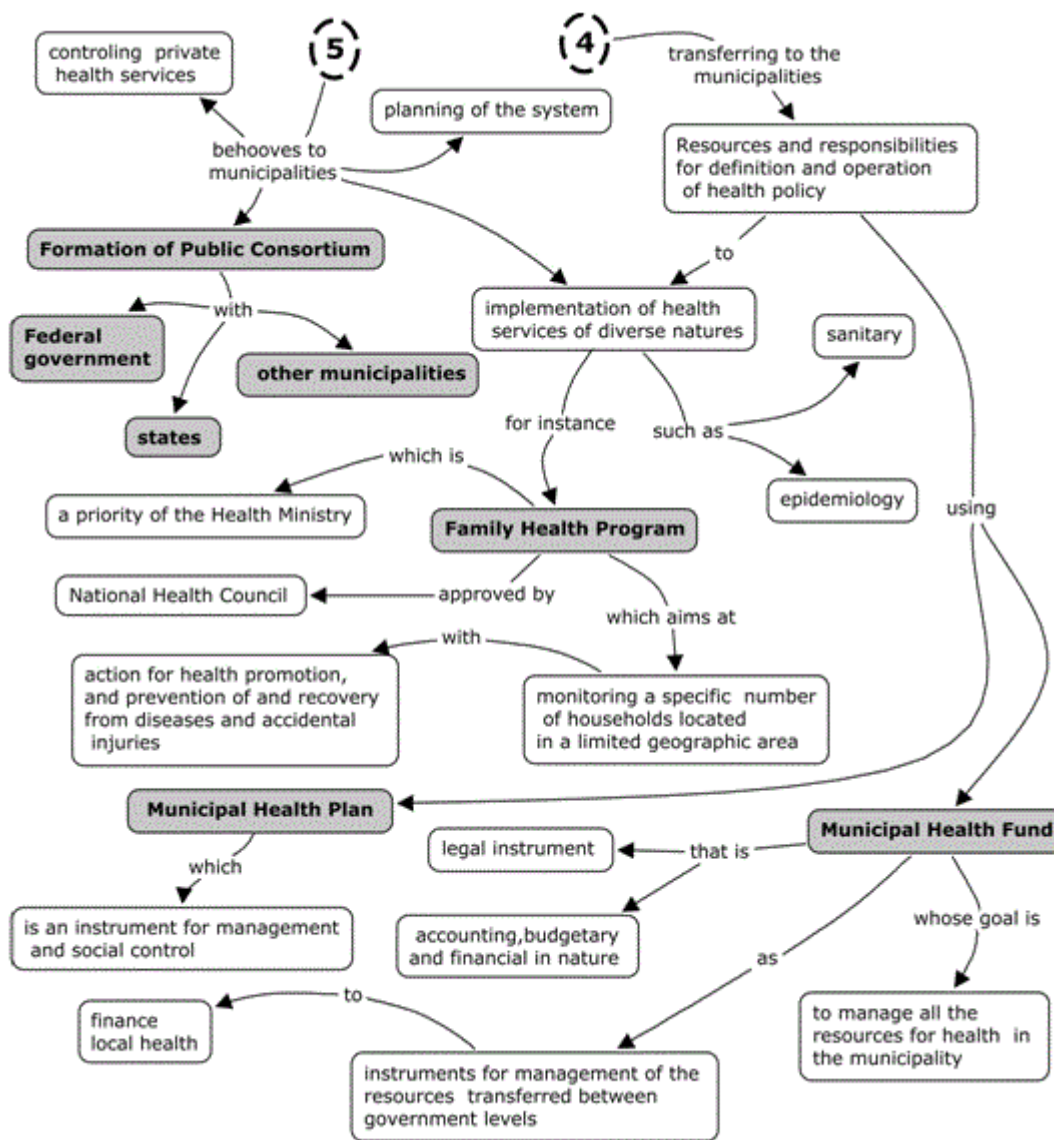


Figure 6- Concept Map of Brazilian Public Health (Part 6)

In Data Mining technique, in order to investigate a possible relationship in distribution of municipalities by groupings with the internal divisions in Brazilian regions (North, South, Southeast, Midwest and Northeast), we first imposed the number of  $k=5$  to the simple K-means algorithm. Then we arbitrarily tested for a second alternative of  $k=27$ , observing the division into 26 states and the Federal District.

Table 1 presents the results of the experiment considering k=5 and k=27. The best fit was given by the greatest value of the log-likelihood parameter, as specified in Table 1, corresponding to the number of selected clusters equal to five.

Table 1- Determination of weka.clust values for different groupings

Number of Clusters (k)	Weka.clust Log_likelihood
5	-7.33
27	-8.87

Table 2 shows that the distribution of municipalities by clusters keeps no relationship with the Brazilian political-administrative and regional division.

Table 2: Distribution of municipalities in the clusters (%)

	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5
Midwest	11.69	11.52	7.80	8.55	4.35
Northeast	34.29	33.04	33.13	31.58	30.53
North	10.15	14.30	3.86	5.93	7.10
Southeast	26.25	25.97	35.06	28.28	32.60
South	17.62	15.17	20.16	25.66	25.42

Figure 7 shows the percentage of municipalities with attributes for the whole set of municipalities (5,565) and for each of the five clusters. A high coefficient of variation (CV) denotes great dispersion among the average group attributes.

Attribute	Attribute description	Full Data 5565	Cluster					Coefficient of variation (%)	
			1	2	3	4	5		
A391	Public Consortium with the Federal Government	0.1119	0.9828	0.0326	0.0237	0.0135	0.0183	200.63973	High dispersion, Heterogeneous
A393	Public Consortium with the State	0.1558	0.9943	0.0723	0.0684	0.072	0.0634	162.86648	
A394	Maternity	0.4171	0.5019	0	0.9921	0.6302	0.0664	93.96926	
A395	Maternity with Civil Registry of birth	0.0712	0.1015	0.0572	0.1656	0.0510	0.0107	76.39671	
A396	Popular Pharmacy	0.4097	0.4674	0.5242	0.8896	0	0.2756	76.02962	
A397	Support of the private sector and communities	0.0609	0.1092	0.0349	0.1271	0.0443	0.026	67.93266	
A398	Laboratory of clinical analyses	0.724	0.7299	1	0.9623	0.9685	0	57.81242	
A399	Covenant of partnership with the private sector	0.2358	0.2874	0.1755	0.4382	0.1845	0.1489	48.27742	
A401	Emergency unit	0.5754	0.6494	0.556	0.8826	0.6152	0.2565	37.97658	
A403	Public Consortium between municipalities	0.4174	0.4387	0.2478	0.369	0.6227	0.4053	32.60836	
A404	Municipal Health Plan - existence	0.7766	0.7701	0.7363	0.8747	0.7929	0.7160	7.92815	Low dispersion, Homogeneous
A405	Characteristic of the Council: Normative	0.4681	0.4808	0.4694	0.5127	0.4411	0.4504	5.96504	
A406	Characteristic of the Council: Advisory	0.6135	0.6533	0.6267	0.6337	0.5859	0.5954	4.50061	
A407	Program Community Health Agent	0.9181	0.9368	0.9182	0.9325	0.9407	0.8748	2.93059	
A408	Characteristic of the Council: Deliberative	0.9048	0.9195	0.8928	0.9474	0.8987	0.8794	2.92518	
A409	Characteristic of the Council: Supervisory	0.7856	0.8065	0.7990	0.8054	0.7562	0.7771	2.75742	
A414	Municipal Fund of Health - Existence	0.9351	0.9157	0.9293	0.9720	0.9437	0.9076	2.72617	
A415	Family Health Programme - Existence	0.9506	0.9579	0.9571	0.9606	0.9602	0.9229	1.70112	
A416	Nature of Council joint	0.9916	0.9981	0.9929	0.993	0.9925	0.9855	0.45266	
A417	The Council meeting was held in the last 12 months	1.0153	1.0134	1.0175	1.0184	1.0143	1.0122	0.26327	
A418	Municipal Council of Health	0.9734	0.9732	0.9698	0.9737	0.9767	0.9733	0.25151	

Figure 7– Percentage of municipalities with attributes in each cluster

## DISCUSSION

We can summarize the results emphasizing some predominant characteristics in each cluster. The first cluster presented 98.28% occurrence of public consortiums with the federal government and 99.43% with the state government. Cluster 2 presents a high percentage of with the Popular Pharmacy program and no presence of maternity care units, in contrast to cluster 4, which has the second highest presence of maternity units and no Popular Pharmacy program.

The second cluster exhibits low incidence of municipalities with public consortiums with the federal and state governments, with 3.26% and 7.23%, respectively. It is noteworthy that 24.78% of municipalities with public consortiums have the smallest participation of this variable compared with other groupings. There is also another important feature in Cluster 2, which is the low incidence of support from the private sector and communities (3.49%).

Cluster 3 has high incidence of maternity units (99.21%), Popular Pharmacy (88.96%), clinical analysis laboratory (96.23%) and emergency care units (88.26%), contrasting with the insignificant occurrence of public consortiums with the federal (2.37%) and state government (6.84%) compared to the first cluster. Regarding agreements for partnership with the private sector, this grouping has the strongest presence among all the clusters analyzed (43.82%). The same holds for the support from the private sector. It turns out that most Brazilian capitals belong to this group, except Rio Branco (AC), Macapá (AP), Manaus (AM), João Pessoa (PB), Maceió (AL), and there is a higher incidence of municipalities in the Southeast region in that group (35.06%)

Cluster 4 contains 1,333 municipalities, no one with a Popular Pharmacy program. There is no Brazilian capital belonging to this group, and a higher incidence of municipalities from the Northeast region (31.58%). This group presents a considerable participation of

public consortiums between municipalities but negligible participation of the federal and state governments.

Cluster 5 is characterized by very low values for all attributes, below the average for the full data set (5,565 municipalities), as shown in Figure 2. In particular, the value of the percentage of clinical analysis laboratories is nil. João Pessoa (PB) is the only Brazilian capital belonging to this group, which also presents a higher incidence of municipalities in the Southeast region (32.60%).

Finally, with regard to the variables of Municipal Health Plan, Community Health Agent Program, Municipal Health Fund, Family Health Program and the Municipal Health Council, the groupings are quite homogeneous. Table 3 presents a summary with the main features of each group by the clustering technique.

Table 3: Main feature of each cluster

Cluster	Main feature
1	High public consortium participation with the federal and state government
2	Participation of variables near the average of the municipalities
3	Municipalities have an infrastructure in the area of health better than those of other groups, but not necessarily ideal
4	Considerable participation of public consortiums between municipalities
5	Participation of the variables below average of the municipalities

The goal of the work was to develop a multi methodological approach to investigate the health profile theme of Brazilian municipalities, based on the integration of concept maps and data mining. The interface consists of some concepts of the map in the grey background, in order to provide context for the data mining modeling. The resulting clusters are necessary for subsequent determination of municipal benchmarking and performance indicators, although these are not within the scope of the present work.

Both qualitative and quantitative information is depicted inside the concept maps as inputs for different formal models with specific scopes, either to provide adequate subsets of

data or to yield performance indicators. Therefore, maps are a semiformal way to structure and model complex societal problems.

There are important differences among the five clusters, evidenced by the 10 variables elicited in the concept map, which presented coefficient of variation (%) greater than 30 (see Figure 7), showing a high degree of dispersion between the groups analyzed. This is decisive for the success of results of the cluster technique, when it is followed by a quantitative model that requires homogeneity in the database.

## REFERENCES

1. Rosenhead, J; Mingers J. Rational analysis for a problematic world: Problem-structuring methods for complexity, uncertainty and conflict. 2nd ed. West Sussex: John Wiley and Sons. 2001. 375p.
2. Okada A, Buckingham Shum S and Sherborne T. Knowledge Cartography: Software Tools and Mapping Techniques. (Eds.). Springer: Advanced Information and Knowledge Processing Series. 2008; pp. 25-46.
3. Novak JD. Learning, Creating, and using Knowledge: Concept Maps as Facilitative Tools in Schools and Corporations, Lawrence Erlbaum Associates, NJ. 1998.
4. Vekiri I. What Is the Value of Graphical Displays in Learning? Ed. Psychol. Rev., 14, 261. 2002.
5. Moreira MA. A teoria da aprendizagem significativa e suas implementações em sala de aula. Brasília: Editora Universidade de Brasília. 2006.
6. Ruiz-Moreno L; Sonzogno MC; Batista SHS and Batista NA. Mapa conceitual: ensaiando critérios de análise. Ciência Educação. 2007. 13, 453-463.
7. Weis SM and Indurkha N. Predictive Data Mining; Morgan Kaufmann Publishers, Inc. 1999.

8. Chilingirian JA, Sherman D. Health Care Applications from Hospitals to Physicians; from Productive Efficiency to Quality Frontiers. In: Cooper WW, Seiford LM, Zhu J, editors. Handbook on Data Envelopment Analysis. Boston: Kluwer Academic Publishers. 2004.
9. Sherman HD and Zhu J. Applying DEA to Health Care Organizations. In: Service Productivity Management: Improving Service Performance using Data Envelopment Analysis. New York, Springer. 2006.
10. Ozcan YA. Health Care Benchmarking and Performance Evaluation: An Assessment using Data Envelopment Analysis (DEA). Massachusetts: Springer. 2008.
11. Sherman HD. Hospital Efficiency Measurement and Evaluation. Medical Care.1984. 22(10): 922-939.
12. Nunamaker TR. Measuring Routine Nursing Service Efficiency: A comparison of cost per patient day and data envelopment analysis models. Health Serv Res. 18(2 Pt 1): 1983. 183-208.
13. Ceretta PS ; Niederauer CAP. Rentabilidade e eficiência no setor bancário brasileiro. RAC, Set-Dez, 2001, v.5, n.3, pp. 7-26.
14. Badin, NT. Avaliação da produtividade de supermercados e seu benchmarking. [Dissertação]. Florianópolis(SC): Programa de Pós-Graduação em Engenharia de Produção, Universidade Federal de Santa Catarina, 1997.