European Regional Science Association

UNIVERSITY OF JYVÄSKYLÄ

ERSA 2003 Congress

THE CLADISTIC TECHNIQUE FOR TERRITORIAL STUDIES

Giovanni RABINO, Francesco SCARLATTI*

(*) Di.A.P. - Dept. of Architecture and Planning - Polytechnic of Milan – Piazza Leonardo da Vinci 32 - 20133 Milan Italy - giovanni.rabino@polimi.it, francesco.scarlatti@polimi.it

ABSTRACT

The cladistic is a technique born in the field of biology. Starting from the physics characteristics of some elements, the cladistic method try to find the hierarchical links between them in the bases of their similarity and of the research of “evidences” of relationship.

The goal of this paper is to verify the potentialities of this method in the studies of the territory, in order to find correlation among different spatial areas. Considering social, economic, environmental indicators that characterise the areas of a given territory, we analyse the links between different areas, organising these links according to a hierarchical tree.

The result is then compared with ones obtained using more traditional methods as multicriterial analysis, statistics and clustering

.

1 THE CLADISTIC (THE METHOD)

The cladistic is born in the field of biology. Every living shape comes down from few types, of far common ancestors. The variety that is generated is the effect of the Darwinian processes of natural selection. An evolution can be seen also from a territorial point of view, but also economic, finding “relation links” “hierarchical links” between several the elements.

We notice first that it is very difficult to perceive the presence of evolutionary mechanisms or links "of relationship" between the elements. We need a hard classification system (or "Systematic") that allow subdividing the elements in definable categories.

The cladistic analysis (or more better the cladistic method of investigation) is born with the goal to express "the philogenetic" relations (like saying of "natural history", of "evolution") of the several living shapes. A total scheme from the cladistic point of view, is a tree (klàdos, in Greek mean" branch"), that start from a point, and end in many branches, one for every single population of organisms. From this point of view the fact that the systematic cladistic reproduces the more faithfully picture offered variety of the environment and does not offer sure advantages of practical order, of availability. The Systematic Cladistic start from some presupposed: to define cluster on the base of the many physical, measurable, characteristics of an element, joining the elements "in a similitude order".

The cladistic analysis is a particular method to formulate hypothesis of correlation between organisms in evolutionary sense. Like the other methods, it is constituted from a cluster of tasks, procedures and limits. The cladistic analysis is recognised like the better method today available for the philogenetica analysis, since supplies explicit and verifiable hypotheses of the correlation between organisms. The base idea that is behind such method is that the members of a group share a common evolutionary history and are correlate you more, between them, than not with other organisms members of other groups. These characteristics are calls sinapomorphisms. The sinapomorphisms are the bases for the analysis. The choice of the right characteristics is the more important step of the cladistic analysis. The tasks in the cladistic analysis:

- every group of organisms is in relation with the descendants of common ancestors;

- the model of clado genesis is always branched off;
- The birth of new characteristic happens for extraordinary descendants (speciation).

First of all is necessary to choose the taxa, that is the elements whose relations are analysed. The taxa chosen must be "branch", for being able to turns out right results: in the case the territorial cladistic the taxa they can be homogenous cities, regions, states, etc. Once chosen the elements, objects of the analysis, it will be necessary to determine the "characters" (characteristic of the elements) and to examine every taxon in order to determine the state of the characters (to decide if every taxon must or must not have every character). All the taxa must be unique.

At this point is possible to build groups of taxa for sinapomorphisms (characteristic derivatives and shared) trying to level the conflicts, that happen in the method applied, using "parsimony" (diminishing is the number of conflicts).

The organisation of all these groups in a tree will concur to create the cladogram (than it is not an evolutionary tree) following those rules:

- all the taxa go on the final points of the cladogram, never on the nodes;
- all the nodes of the cladogram must have a list of sinapomorphisms and must be in common with all the taxa of that node;
- all the sinapomorphisms appear in cladogram one single time,

The result is the cladogram that is a branched diagram that can be read like a family tree.

The cladistic analysis is used in order to create classification systems and is currently the more common method used to carry out classifications. Moreover the cladistic, as every other scientific system widely used, represents a model that is not only used in order to describe what is observed, but in order to foretell what has not been still observed. The cladistic supplies hypothesis on the correlation between organisms or elements, in such way to foretell the particular properties. Those characteristics can be searched only basing on evolutionary theories and the cladistic analysis can exactly be used like instrument for these researches.

The Cladistic has some method errors. In order to avoid influencing the results with such errors, the scrupulous student would have to be adhered to the line of conduct described here:

- It must choose with careful reflection the character set to consider, finding an equilibrium between “general” characters (that regard larger groups of organisms, increasing to the phantom of our analysis) and "particular" characters (that shrink such spectrum).
- We must avoid the not indispensable complications, remembering that the probability to incur in error increases with the dimensions of the matrix, and therefore, omitting the taxa and the characters that are not indispensable.
- The characters considered must be observable with no doubt. But it is not right to fall in the opposite excess, excluding interesting taxa only because a part of the characters is not observable in the set.
- Possibly the input matrix must codified the modality "a/p", only with characters that can be described as "absent" (0) or "present" (1).
- Once obtained the matrix, we have to analyse it with statistics methods in order to determine its congruence, confronting it with matrices generated in accidental way. If we have chosen well the group of taxa, however, this step is superfluous.

2 THE PROBLEM

The various aspects that characterise every single municipality, every single territory, from that we could define geomorphologic to the human conditions (the density, the industrialisation, the standard of life, etc.), constitute a kind of territorial “genetic code”. Those that environmental, territorial, economic, industrial, social and cultural aspects are recorded this DNA: code and constitutes the identity of a place. Although every territorial entity is different from all the others, inside its genetic code can find again some common elements to other cities, countries and province.

In this kind of vision, that is very similar to the Biology one, we can think to carry out some analyses of taxonomic kind habitually used in biological field to construct a family trees. Obviously in our case the similarity of the codes will not presuppose a biological descendancy but one territorial typological affinity. In particular, in order to identify the degree of affinity or similarity between the

municipalities of the of Bergamo province, we will use the cladistic analysis because it supplies explicit and verifiable hypotheses of the correlation between social, economic and territorial elements.

In order to carry out this analysis have been chosen, as characteristic elements the 36 indicators uses from the financial daily paper “Il sole 24 ore” on the year’s quality of the life report of the Italian provinces. The elements examined will be:

1) STANDARD OF LIFE

THE PRODUCED WEALTH (the added value pro capite)
THE SAVINGS (banking for inhabitant)
FAMILIAR EXPENSES (dynamic indicator of the consumption of the families)
PENSIONS (perceived annual medium amount from pensions)
WHO WORKS and WHO RESTS (number of pensions every 1000 occupied)
THE HUOSE (average price for m^2 for an apartment in semicentral zone)

2) BUSINESS AND JOB

ENTREPRENEURIAL SPIRIT (recorded enterprises every 100 inhabitants)
THE FAILURES (bankrupt enterprises every 1000 recorded)
THE SEARCH OF A PLACE (% of persons searching job in connection with the workforce)
WHO BET ON EXPORT (% of exports on the added value)
THE COLLECTED CREDITS (medium amount of the protests for inhabitant)
THE PROCESSES NOT WITHDRAW (civil procedures pendants every 1000 inhabitants)

3) SERVICES AND ENVIRONMENT

THE INFRASTRUCTURE (Tagliacarne index of the infrastructure equipment)
THE ECOLOGICAL REPORT (synthetic index Legambiente on the city ecosystem) BEAUTIFUL STABLE (difference between the medium temperatures of the warmer month and the month more cold)
MORE SPACE TO STUDENTS (medium number of students for class in the schools)
THE FATAL ILLNESS (the dead men for cancer on the total for deaths)
RISK ON THE ROADS (street accidents every 100000 inhabitants)

4) CRIME

ALARM HOLDUPS (number of denounced hold-ups every 100000 inhabitants)
THE CARS IN THE VIEW-FINDER (number of thefts of cars every 100000 inhabitants)
BURGLED APARTMENTS (number of thefts every 100000 inhabitants)
SMALL IS UGLY (bag-snatching denounced every 100000 inhabitants)
TREND (variation of trend on the total of the crimes denounced from the police)
THE DEFENSE Of THE PUBLIC ORDER (index of the performances and equipment of the police)

5) POPULATION

DEMOGRAPHIC DENSITY (number of inhabitants for km^2)
IN THE CRIBS (number of born alive every 1000 inhabitants)
THE DEATH RATE (number of dead every 1000 inhabitants)
ARRIVALS AND DEPARTURES (new registrations for transfers from others province every 100 cancellations)
REFUSED LIFE (every number of suicides and tried you of suicides every 100000 inhabitants)
END OF A LOVE (number of divorces and separations every10000 families)

6) FREE TIME

THE PLEASURE TO BE ASSOCIATED (number of artistic, cultural and recreational associations every 100000 inhabitants)
CULTURE IN DISPLAY WINDOW (number of bookcases every 100000 inhabitants)
FITNESS (number of gyms 100000 inhabitants)
THE PASSION FOR THE CINEMA (number of cinemas every 100000 inhabitants)
THE AUDIENCE Of THE STAGE (medium expense for inhabitant assisting to theatre and musical shows)
AT THE STADIUM (medium expense pro capite to assist to sport shows)

3 THE DATA

The base point of our research has been the matrix of the data constructed on the 36 indicators use by “Il sole 24 ore”. The data puts in the matrix have been gain with a complex research that has involved the municipalities offices, newspapers, agencies and other structures, associations. To fill the data gaps we have done some statistical evaluations and simplifications.

	STANDARD OF LIFE											
MUNICIPALITIES	THE PRODUCED WEALTH (the added value pro capite)	Normalizated data	THE SAVINGS (banking for inhabitant)	Normalizated data	FAMILIAR EXPENSES (dynamic indicator of the consumption of the families)	Normalizated data	PENSIONS (perceived annual medium amount from pensions)	Normalizated data	WHO WORKS and WHO RESTS (number of pensions every 1000 occupied)	Normalizated data	THE HUOSE (average price for m2 for an apartment in semicentral zone)	Normalizated data
Adrara San Martino	34546,00	5	12737,31	5	600,63	7	11525,63	9	222,36	0	1518,75	5
Adrara San Rocco	9731,54	1	5889,82	2	637,81	8	12162,14	9	759,80	2	1579,50	5
Albano Sant'Alessandro	21513,98	3	7977,60	3	644,78	8	10738,57	8	293,42	1	1637,82	6
Albino	19443,46	3	9092,59	4	614,15	8	10175,35	8	432,56	1	2430,00	8
Algua	7442,76	1	7616,49	3	588,09	7	6241,87	5	1583,83	3	1579,50	5
Almè	24025,10	3	12194,66	5	662,68	8	8781,69	7	356,35	1	2673,00	9
Almenno San Bartolomeo	17182,89	2	9533,89	4	637,05	8	7890,06	6	321,76	1	2138,40	7
Almenno San Salvatore	10589,73	1	7430,66	3	650,27	8	6934,43	5	973,48	2	2430,00	8
Alzano Lombardo	14198,79	2	7589,04	3	598,30	7	9645,01	8	679,71	1	2187,00	8
Ambivere	12177,31	2	6554,01	3	657,21	8	10602,64	8	572,28	1	2393,55	8
Antegnate	19654,58	3	10056,68	4	637,20	8	10984,31	9	484,01	1	1336,50	5
Arcene	12100,60	2	7232,45	3	660,72	8	9097,11	7	734,06	1	2344,95	8
Ardesio	13311,65	2	7696,52	3	634,13	8	6422,56	5	711,91	1	1883,25	6
Arzago d'Adda	10605,80	1	6665,12	3	627,71	8	10145,75	8	751,90	1	2029,05	7
Averara	26303,28	3	8393,93	3	452,57	6	9000,00	7	1392,56	3	1397,25	5
Aviatico	10628,67	1	13002,03	5	490,00	6	3858,75	3	1220,87	2	2138,40	7
Azzano San Paolo	29376,32	4	11616,92	5	649,89	8	10657,75	8	276,63	1	2162,70	7
Azzone	10431,95	1	6180,52	2	585,74	7	4120,25	3	1170,65	2	1701,00	6
Bagnatica	34300,14	4	9610,66	4	656,94	8	10309,15	8	194,35	0	2332,80	8
Barbata	20901,87	3	7874,58	3	676,98	8	11319,29	9	638,38	1	1822,50	6
Bariano	9668,73	1	7470,52	3	656,58	8	7150,59	6	778,81	2	2004,75	7
Barzana	20761,59	3	9030,66	4	655,66	8	10943,18	9	298,77	1	1968,30	7
Bedulita	13147,95	2	7187,23	3	660,69	8	8915,96	7	748,45	1	1640,25	6
Berbenno	14139,13	2	10320,06	4	647,86	8	7696,32	6	655,59	1	1731,38	6
Bergamo	33831,92	4	13224,23	5	549,57	7	10550,00	8	535,25	1	2916,00	9
Berzo San Fermo	7214,32	1	5116,09	2	640,56	8	9185,44	7	1201,34	2	1822,50	6

Figure 1: Part of the input matrix

Is necessary to tell that the data must be normalised between 0 to 9

4. THE RESULTS

Starting from the input matrix, containing normalised data, we have done an elaboration using the freeware software "WINCLADA®"

This elaboration consists in a comparison of all the string two by two to characterise of differences and the similarities to find the "affinity degree" between them. Because of the particular architecture of the software it has been necessary to choose term of comparison to which link the affinity degree of every municipality. Our choice has been the capital of the province "Bergamo".

The elaboration's result is a tree of all the municipalities of the province that, starting from the root represented from the city of Bergamo, branches out following a hierarchical logic. Covering the several ramifications we can read the differences between every single municipality and Bergamo basing on the distance in the cladogram. In our analysis we have decided to elaborate various cladograms on the base of different kind of indicators

The first tree consider all the indicators:

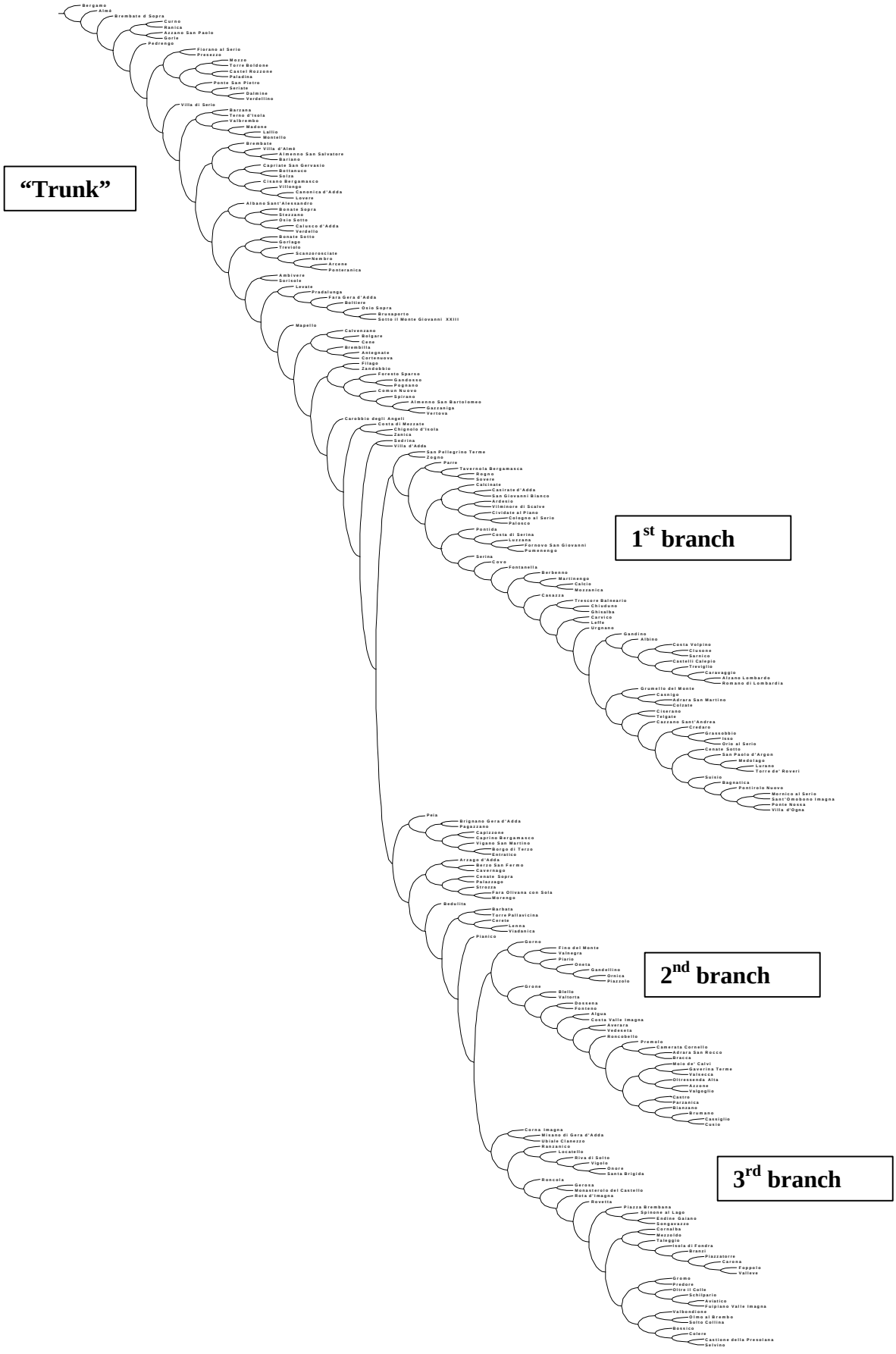

Figure 2: generals tree (all the indicators)

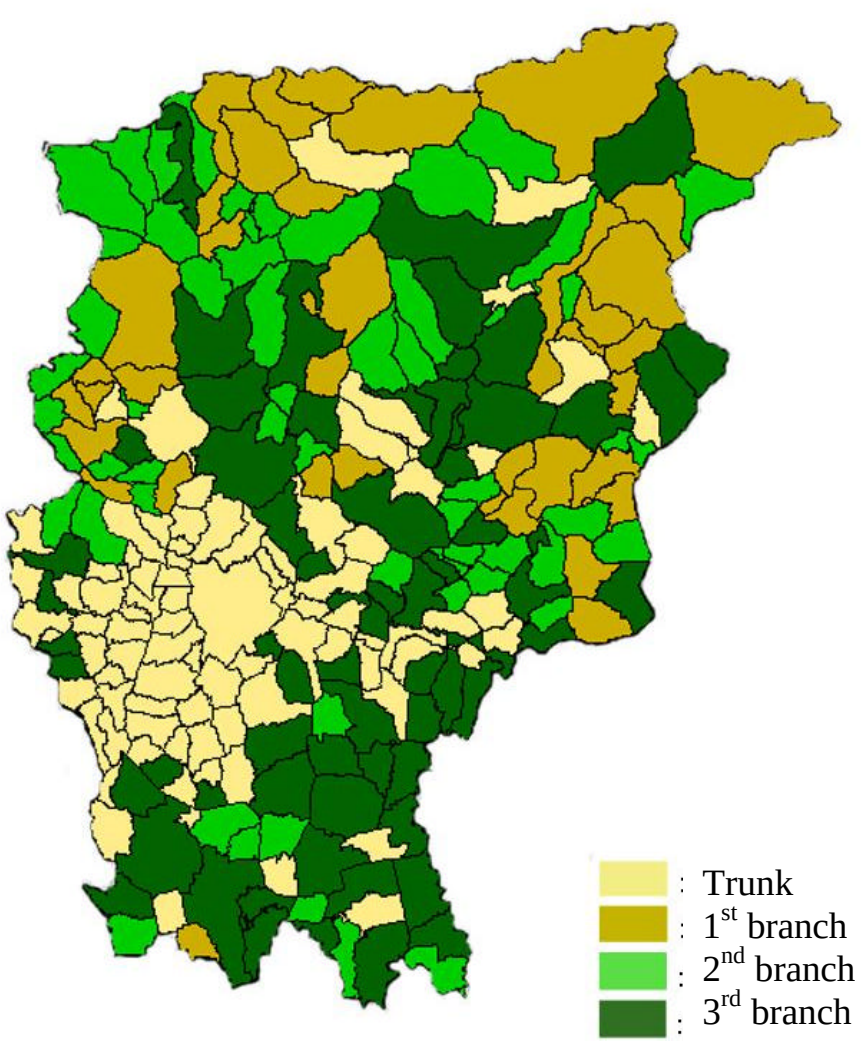

Figure 3: location of branches on the map

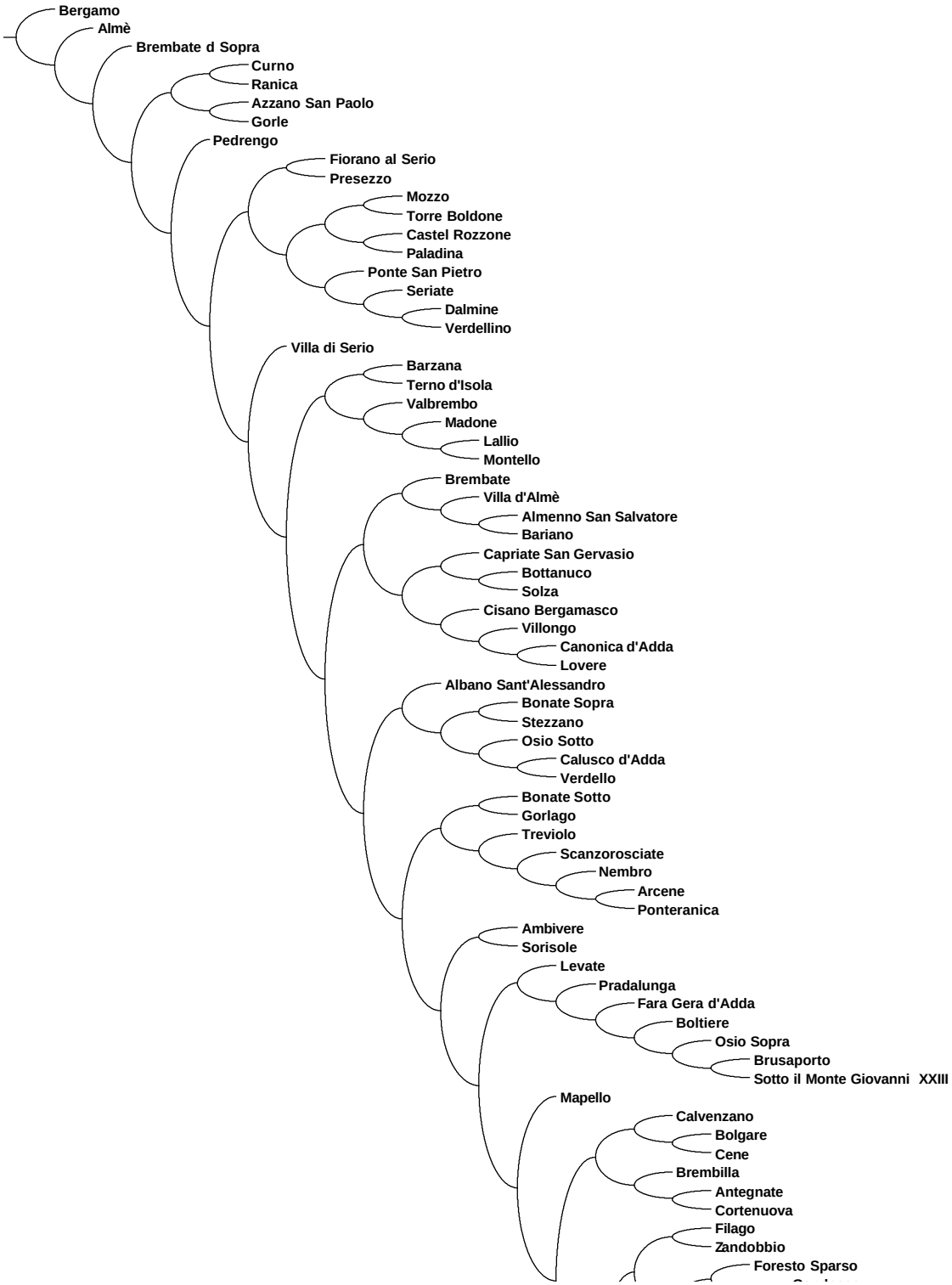

Figure 4: The trunk. Municipalities that has affinity with Bergamo

Observing the tree we can notice that, leaving from trunk, it divides itself in three main branches. As obvious consequence we can presume that every branch differs from the trunk in qualitatively different way. This means that the diagram has to be read in two dimensions: in every ramification, leaving from the departure point we will

obtain a quantitative differentiation; choosing to covering a branch rather than an other the differentiation will be of qualitative type.

In order to analyse more easily the elaborated tree we have mapping (figure 3) the territory. Every colour corresponds to a branch of the tree.

This zoning divides the municipalities in four main groups that have inside some affinities. In particular we notice that the municipalities that have more affinity with Bergamo are not only the once nearest but also those that assume function of territorial capital for the valleys of the Bergamasca region. Brembilla in Val Brembilla and Nembro in Val Seriana represent some examples of this affinity with Bergamo.

STANDARD OF LIFE

Analysing the quality of life the municipalities that are more similar to Bergamo are those municipalities of industrial kind that supply numerous services and are characterise from a high density. We could define “similar” to Bergamo all those municipalities that in the so-called “Bassa Bergamasca”. A plan geographic area characterised by a high density of industries.

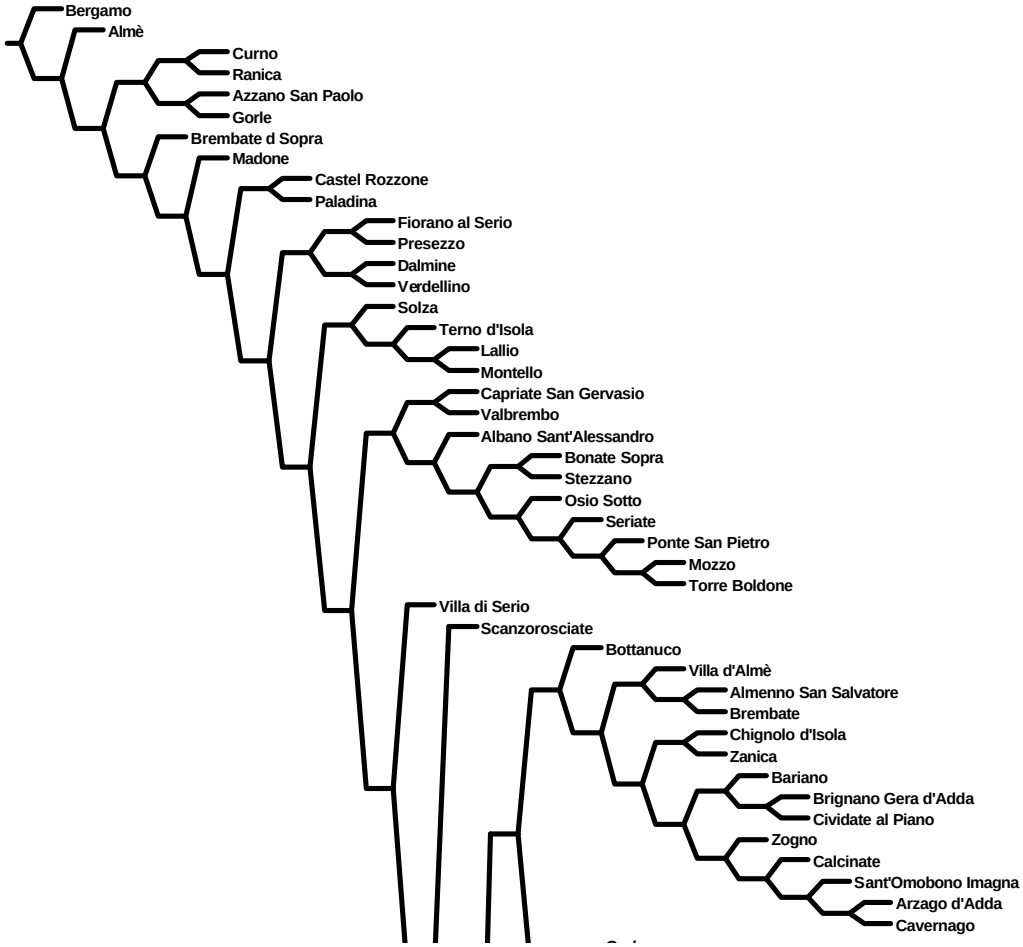

Figure 5: Municipalities similar to Bergamo for standard of life

BUSINESS AND WORK

Field as industrialisation, accessibility, developments of the tertiary sector characterise the municipalities that have affinity with Bergamo in this section.

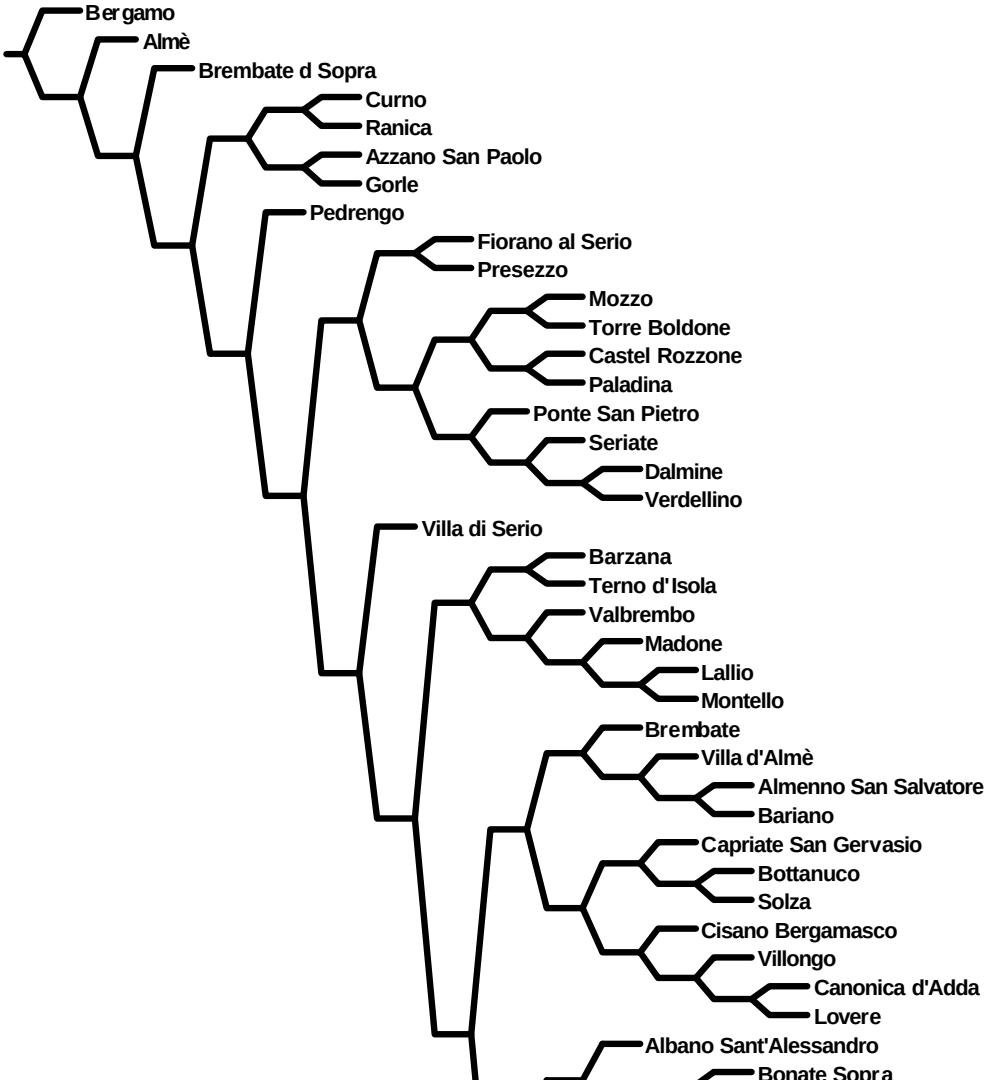

Figure 6: Part of the tree “Business and work

SERVICES AND ENVIRONMENT

In this case the mountains area seems to be the further from Bergamo. This is an obvious conclusion; in fact the quality of environment here is very different by the Bergamo one.

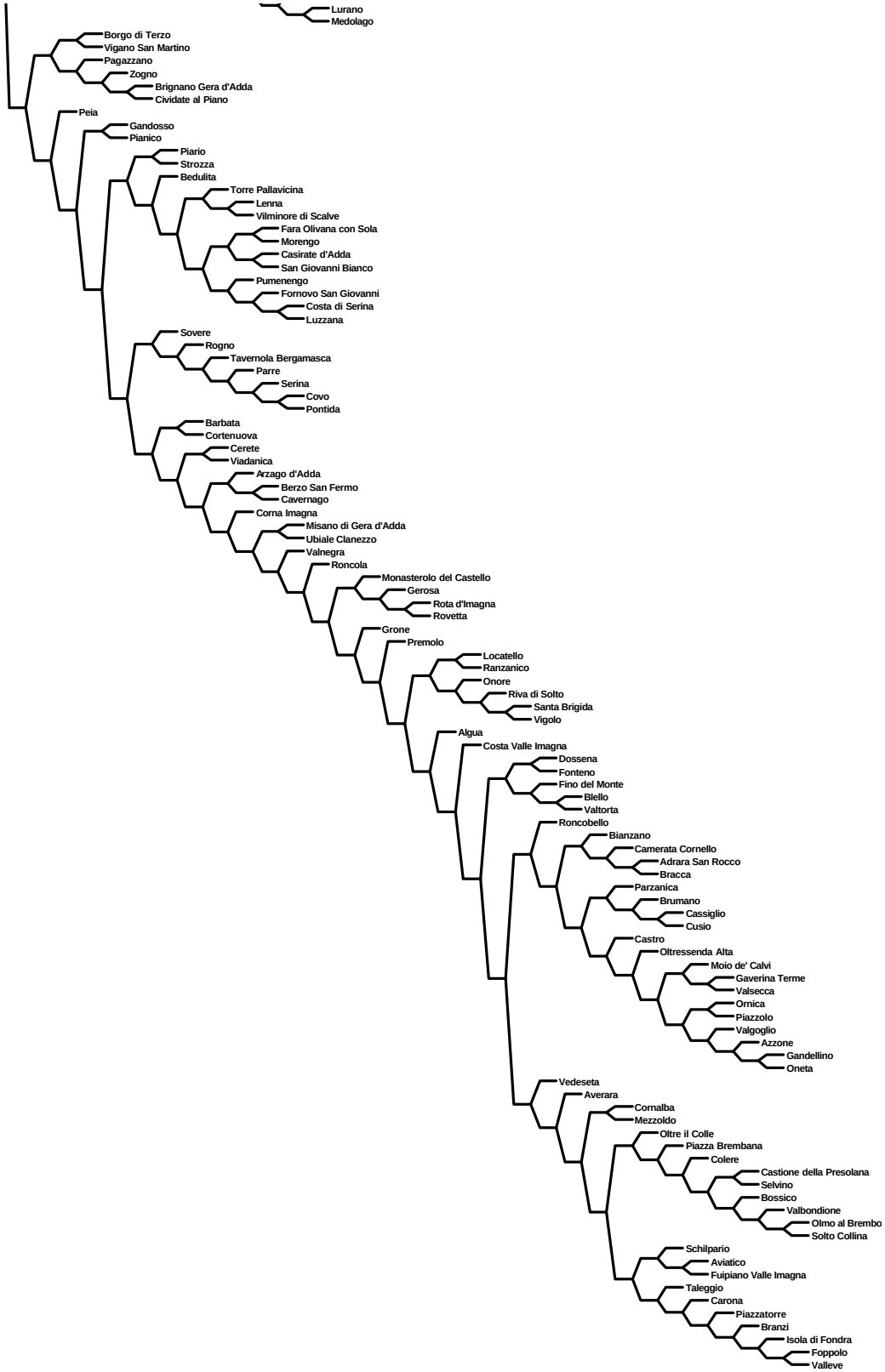

Figure 7: Services and environment

We have done similar tree for the other argument: crime, population and free time. We omit to show them, because they are not interesting in this context.

5 CONCLUSIONS

This research is a first attempt to apply the cladistic to the territorial studies. The idea is that some social, economic and territorial indicators constitute a kind of genetic code of the territory. This approach has some useful characteristic for planning as:

- Find similarities and dissimilarities between different areas;
- Give local indications for investments;
- Use data easy to find;
- Help to find a hierarchical organisation of the territory;
- Is different from the traditional cluster analysis because, using an evolutionary theory, give an idea of the development of the territory;
- Can be used to describe a territory and to do previsions. In fact knowing the trend of the indicators is possible to formulate an hypothesis on the future organisation of the territory;
- Give an interpretation of the mutual relations of indicators.

There are also some problems to solve:

- The results depend by the first string of indicators (individuation of the root);
- The individuation of the “genome” (the right indicators);
- The results have no explicit justification.

The next step of the research is to try to solve these problems by applying the method to one well-known territory and compare the results with some other, obtained with traditional analysis.

References

McCarthy and Fernandez "E-business and the Implications for Manufacturing Companies", in 2000/01 Edition of the Institute of Management's Skills & Training Handbook.

Gillies and Leseure "Adaptable networks: perspectives from a business context", in Managed Care Networks - Principles and Practice, Editors James R. and Miles A., pp 65 - 91, Aesculapius Medical Press, ISBN 1 903044 27 8, 2002.

McCarthy "Organisational Diversity, Configurations and Evolution", in Complex Systems Book, Editor: Eve Mitleton-Kelly, Due for publication in 2001.

McCarthy, Rakotobe-Joel, and Frizelle "Complexity and Complex Systems in Manufacturing", Chapter in the Manufacturing Engineers Handbook, Editor: Prof. Pham, Springer Verlag, Due for publication in 2001.

McCarthy, and Rakotobe-Joel "Complex Systems and Complexity in Industry - Proceedings of the International Conference on Complexity and Complex Systems in Industry", ISBN 0 902683 50 0, 19th -20th Sept, 2000, University of Warwick..

McCarthy, Ridgway, and Mort N "Appropriate Technologies for Mature Industries" in Wiley Encyclopaedia of Electrical and Electronics Engineering, Editor: Professor John Webster, John Wiley and Sons, Jan 1999.