STATISTICA, anno LXVI, n. 2, 2006

ATTRACTION POLES IN THE UNIVERSITY OF BOLOGNA: A SOCIAL NETWORK ANALYSIS

L. Stracqualursi, P. Monari

1. INTRODUCTION

The aim of our analysis is to identify which faculties attract students – or, in other terms, which faculties the students prefer – in order to improve academic recruiting and orientation.

With the aim to infer the dynamic and structural mechanisms that lead the evolution of the students' movements between the faculties in the University of Bologna, we built a complex evolving network.

The idea is to see the movements of students between faculties within the University of Bologna as a network which can be represented as a graph.

In mathematical terms, a graph is a pair of sets $\boldsymbol{S}\{\boldsymbol{A},\boldsymbol{E}\}$, where $\boldsymbol{A}$ is a set of N nodes (or vertices) $A_{1,}A_2,...,A_N$ and $\boldsymbol{E}$ is a set of edges (or links) who each connecting two elements of $\boldsymbol{A}$. Graphs are usually represented as a set of dots, each corresponding to a node, two of these dots being joined by a line if the corresponding nodes are connected (see fig. 1).

Graph theory has its origins in the eighteenth century in the work of Leonhard Euler, the early work concentrating on small graphs with a high degree of regularity (Ballobás, 1985).

In the twentieth century graph theory has become more statistical and algorithmic. A particularly rich source of ideas has been the study of random graphs, that is graphs in which the edges are distributed randomly. Networks with a complex topology and unknown organizing principles often appear random; thus random-graph theory is regularly used in the study of complex networks (Barabási and Albert, 2002).

In this paper we are interested to identify the group of "*attraction*" nodes, that is, faculties (or degree courses) that tend to attract students from other faculties. On the other hand, "*escape*" nodes would identify faculties (or degree courses) which have a high out-transfer rate.

A top to bottom approach and graphical visualisation are used for the analysis. Rather than considering the full university network and then simplifying it, we start from a basic network of 23 faculties. Subsequently, we extend the network by analysing the 365 degree courses offered by the faculties.

In section 2 we built the faculties network in several academic years with some general measures, to verify the existence of "attraction"" faculties. In section 3 we see the degree courses network, with particularly attention to degree courses belonging to "attraction" and "escape" faculties: the aim is to learn which degree courses define and characterize the faculty situation. Finally we propose a network model for explaining the topology of observed network.

2. THE FACULTIES NETWORK

As a first data exploration, we analyse the network of the 23 faculties within the University of Bologna. In this network, the nodes are the faculties, while the links are the students' inter-faculty transfers. In this context, the weight of each link is given by the number of students that transfer. Such network is called a valuated (or weighted) network.

We utilise data from the electronic database "datawarehouse" provided by the University of Bologna, which contains all relevant information regarding the students' university career, for three academic years (2004-2007). The software Ucinet 6.158, and Netdraw 2.055 are used for processing data.

Since the aim is to identify "attraction" nodes in the network of faculties, the first measures that we calculate are global network centralization indices, which express the degree of inequality or variance in our network.

For non-symmetric data the in-degree of a vertex v is the number of ties received by v and the *out*-degree is the number of ties initiated by v. In addition, if the data is valued then the degrees (in and out) consist of the sums of the values of the ties.

For a given binary network with vertices $v_1,...,v_N$ and maximum degree centrality $c_{\max}$, the network centralization measure is:

$$C_N = \frac{\sum_{i=1}^{N}(c_{\max} - c_i)}{\max\left\{\sum_{i=1}^{N}(c_{\max} - c_i)\right\}} \tag{1}$$

- where c_i is the degree centrality of vertex v_i.

This because it would be useful to express the degree of variability in the degrees of actors in our observed network as a percentage of that in a star network of the same size (maximums). The star network is the most centralized or most unequal possible network. In this network all the actors but one have degree of one, and the "star" has degree of the number of actors, less one (Freeman, 1979).

TABLE 1

Network centralization measures

	Out-degree	*In*-degree
2004-05	9.69%	12.54%
2005-06	13.66%	14.78%
2006-07	10.55%	8.37%

In the three academic years, the values of the *out*-degree and *in*-degree network centralization measures are included in the range 8.37 – 14.78 per cent (table 1).

We would arrive at the conclusion that there is only a little amount of concentration or centralization in the whole network. That is, positional advantages (the faculties the students prefer, in our case) are rather equally distributed in this network. Furthermore this index temporary decrease underlying trend to reduce further the few positional advantages.

In order to identify and visualize "classes", we apply multi-dimensional scaling (MDS) of Geodesic Distance, that is display graphs network methods based on proximity among actors. MDS represents the patterns of similarity or dissimilarity in the tie profiles among the actors as a "map" in multi-dimensional space. This map lets us see how "close" actors are, whether they "cluster" in multi-dimensional space, and how much variation there is along each dimension. Figure 1 displays the graph nodes in a two-dimensional space, where actors having big between-interaction are placed near, while actors having small interaction are placed in a more distant space (Kruskal, 1978). Different colours signify different geographic locations: faculties outside Bologna are represented by a white small circle, while faculties within Bologna by a grey small circle. In the graph below, we can see a group of center faculties having more between-interaction, and are all located in Bologna. A possible reading is that students studying in Bologna prefer to move in other faculties located in Bologna too (fig. 1).

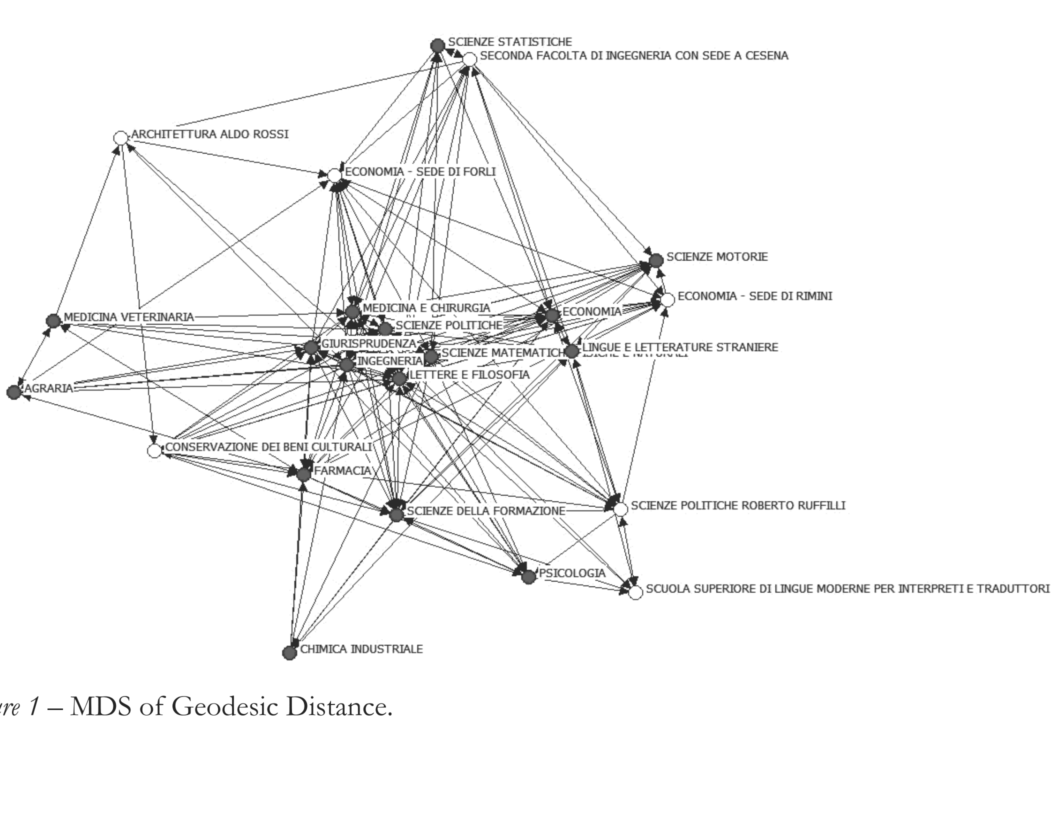


Figure 1 – MDS of Geodesic Distance.

2.1. *The identification of "attraction" and "escape" nodes*

First empirical measurements which allow us to discover "attraction" and "escape" nodes are *in* and *out*-degree. Particularly, bigger is the number of node enter links (*in*-degree), bigger is its power to recall students, vice versa bigger is the number of node out links (*out*-degree), smaller is the faculty attitude to keep students.

Since these quantities depend by different faculties dimensions, we have normalized the degrees by using for each faculty relative students' number. After this normalization, the results change: other faculties, displayed in grey in table 2, are now included as "escapes" or "attractions" nodes. Besides potential "escape" and "attraction" nodes – that is faculties which a high *out*-degree and *in*-degree – are displayed in bold.

TABLE 2

Degree and normalized degree

	Degree						Normalized Degree					
	2004-05		2005-06		2006-07		2004-05		2005-06		2006-07	
	Out	In	Out	In	Out	In	Out	In	Out	In	Out	In
Agricolture	15	15	26	13	24	17	3,8%	3,8%	6,9%	3,4%	8,0%	5,7%
Architecture	7	3	1	7	1	11	5,1%	2,2%	0,7%	5,1%	0,9%	9,6%
Industrial Chemestry	10	3	2	3	9	4	12,5%	3,8%	2,0%	3,0%	9,3%	4,1%
Preservation of Cultural Heritage	17	17	14	9	10	15	6,5%	6,5%	5,7%	3,7%	5,5%	8,2%
Economics	60	72	66	64	54	66	4,6%	5,6%	5,4%	5,2%	5,7%	7,0%
Economics - Fo	16	26	25	37	19	24	3,1%	5,1%	6,9%	10,2%	4,4%	5,5%
Economics -Rn	12	24	24	21	15	21	2,2%	4,3%	5,6%	4,9%	3,9%	5,4%
Pharmacy	73	26	92	21	80	23	12,3%	4,4%	24,7%	5,6%	10,6%	3,1%
Law	86	65	79	75	82	42	5,4%	4,1%	5,2%	5,0%	6,0%	3,1%
Engineering	**145**	29	**106**	28	**120**	31	9,2%	1,8%	6,8%	1,8%	8,3%	2,1%
Arts and Humanities	94	**175**	89	**190**	89	**134**	2,9%	5,5%	3,1%	6,6%	3,5%	5,3%
Foreign Languages ..	80	27	77	35	66	50	10,5%	3,5%	9,8%	4,5%	8,6%	6,5%
Medicine	18	**131**	22	**138**	14	**149**	1,9%	13,8%	2,1%	12,9%	1,4%	14,5%
Veterinary Medicine	9	21	17	27	13	13	5,1%	12,0%	9,4%	15,0%	6,8%	6,8%
Psychology	3	20	9	25	6	30	1,2%	8,2%	3,6%	10,0%	2,5%	12,5%
Education Sciences	39	72	29	82	40	80	3,4%	6,2%	2,5%	7,1%	4,2%	8,4%
Mathematical, Physical ... Sciences	**145**	77	**179**	74	**177**	68	9,9%	5,2%	12,0%	5,0%	14,0%	5,4%
Exercise and Sport Sciences	7	25	14	22	10	22	2,5%	8,9%	4,7%	7,5%	3,4%	7,4%
Political Sciences	84	**100**	77	64	71	76	6,3%	7,5%	6,3%	5,2%	6,9%	7,4%
Political Sciences - Fo	34	20	30	27	17	27	6,6%	3,9%	6,2%	5,6%	3,6%	5,7%
Statistical Sciences	5	5	5	16	12	13	4,0%	4,0%	2,4%	7,6%	8,3%	9,0%
Advantage School Of Modern Languages ..	2	11	3	5	2	8	1,3%	6,9%	1,9%	3,1%	1,3%	5,2%
Engineering II Faculty	30	27	23	26	26	33	7,7%	6,9%	5,9%	6,7%	7,7%	9,8%

The "escape" faculties in addition to "Engineering" and "Mathematical, Physical and Natural Sciences" are "Pharmacy", "Industrial Chemistry" and "Foreign Languages and Literature", because they have a high out-degree related to their number of matriculate students in the expected academic year. At the same way, additional "attraction" faculties are "Veterinary medicine" and "Psychology". Figure 2 displays graphically the network, where nodes with high *in*-degree are in white while nodes with high out-degree are in grey.

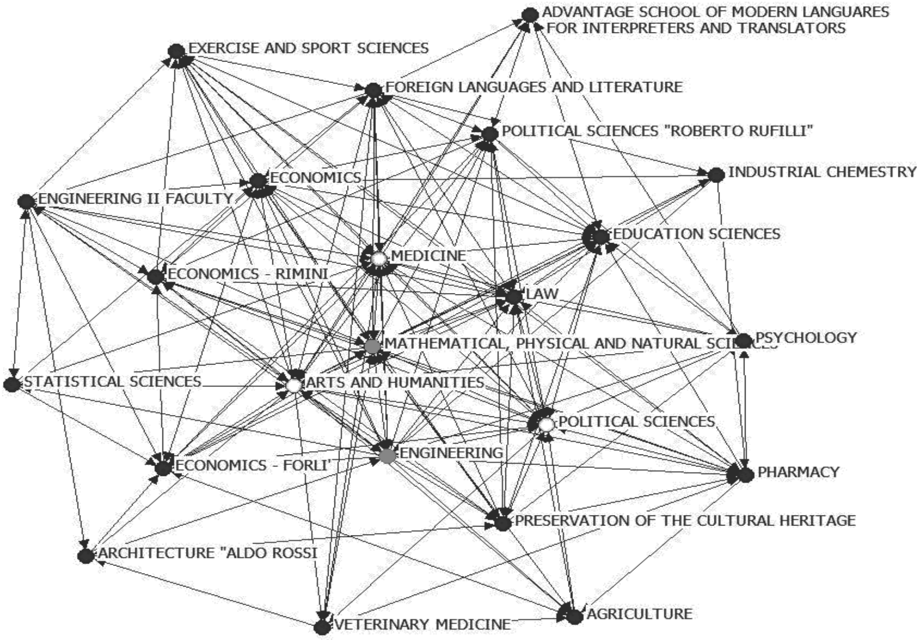


Figure 2 – Faculties network of University of Bologna.

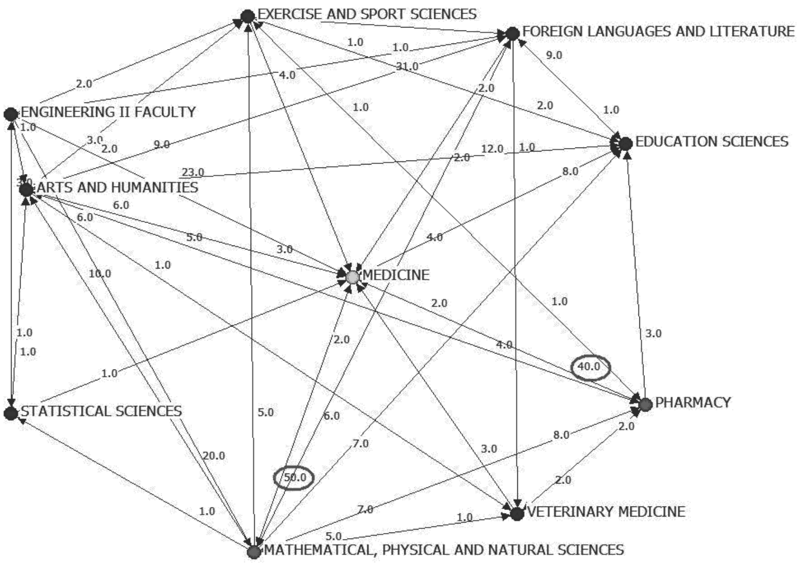


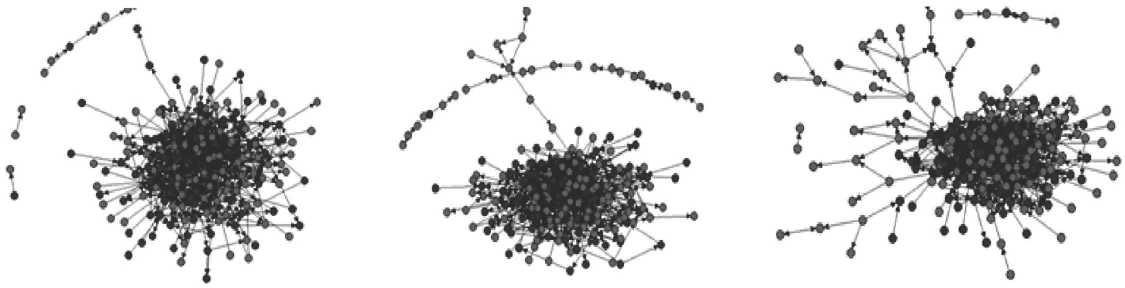
Figure 3 – Egonet of Medicine.

In order to discover the great *in*-degree of "Medicine" we built its weighted egonet (fig. 3). Only two faculties, "Mathematical, Physical and Natural Sciences" and "Pharmacy", concur remarkably to its *in*-degree. This because "Medicine" has a close number: students do not pass the admittance test matriculate in similar examinations faculty and then, after one year, ask for transfer. The second step analysis on degree courses will show that the specific degree course used by students to pass the drawback of failure admittance test in "Medicine" are "Biological Sciences" and "Pharmacy".

3. THE DEGREE COURSES NETWORK

After to have identify "attraction" and "escape" faculties, we analyse the network of the 365 degree courses into faculties within the University of Bologna. In this new network, the nodes are the degree courses, while the links are always the students' inter-degree courses transfers. We analyze this network for three academic years (2004-2007).

The first consideration is that the evolution is not in terms of growth, but of change in its linkage system. The graphs in figure 4 display the changing linkage structure in the last three academic years, particularly in the year 2006-07 isolated nodes respect to the pass are very few.



(a) 2004-05 (b) 2005-06 (c) 2006-07

Figure 4 – Whole degree courses networks.

The simplest and most common method to simplify a network is binarizing data with a specific cut-off.

In deciding what cut-offs to use to create images, after empirical simulation we have chosen cut-off=3: it means that links between degree courses which have less then 3 students movements are not considered.

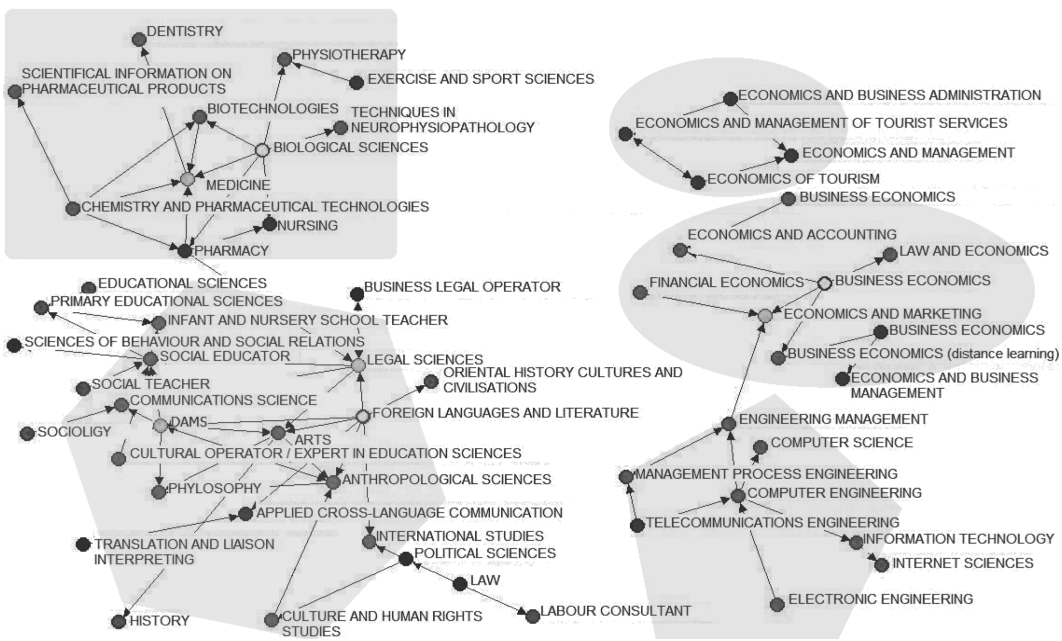


Figure 5 – Semplified network with cut-off (academic year 2004-05).

The results for the academic year 2004-05 are shown in figure 5. The different colours indicate: black – degree courses located outside Bologna; grey - degree courses located in Bologna; white – degree courses with high *out*-degree; soft grey – degree courses with high *in*-degree.

Particularly, we can see 4 areas:

- "medic and paramedic" area,
- "economic" area,
- "humanistic and juristic" area
- "engineering" area.

It's useful to observe that the "economic" area is split in two clusters corresponding to two different geographic locations and having not "between" exchanges. It means that students who change economic course prefer to stay in the same geographic location.

Only one bridge links "economic" and "engineering" area, between "Engineering management" and "Economics and marketing". Those degree courses are felt by students rather similar even if they belong from different areas.

The nodes with only *in*-degree or "leaf" nodes represent the "final" courses, namely students choose those courses like target, and not as passage.

In 2005-06 the network evolves in the link structures: the "engineering" area is split into "informatics" and "traditional" engineering, therefore students that choose "informatics" courses change only to come to another "informatics" course, but in 2006-07 the structure of "engineering" area is bound again.

Until this moment we have spoke only of "escape" and "attraction" degree courses, but is necessary to rename some "escape" degree courses in "transit" degree courses. This is the case of "Biological Sciences" and "Pharmacy", used by students to pass to "Medicine".

Even if the linkage structure changes, the "leaf" nodes remain about the same: for example, in all three years, in "medicine" area, "Dentistry", "Physiotherapy" and "Scientifical information on pharmaceutical products" are degree courses in which students only arrive.

3.1. *Some centrality measures in degree courses network*

In the degree courses network, with the aim to discover the "attraction" and "escape" nodes, we have used *in* and *out*-degree, their normalized percentage value, the matriculate students' number, and finally the "degree centrality" of each node. In fact, while in faculties network, it is uncommon that students move from one node to arrive all to another one, in degree courses network this event often occurs. Therefore it needs considering "attraction" a degree course that has an high normalized *in*-degree and a degree centrality different from 1. The same is for "escape" courses on the base of the *out*-degree. The number of students matriculate in the degree course in the current year needs in order to identify closure courses, that are courses in which the year matriculations are zero.

The potential "escape" and "attraction" degree courses in several years are shown in tables 3 and 4 respectively.

Particularly, in table 3 we can see some degree courses deleted, like "Social safety and control operator", "Economics and management of tourist services" and "Computational Informatics-Mathematics". The first two are deleted because, even if their *out*-degree is high, the year matriculations are zero: these are concluded courses. The last one has degree centrality equal to 1, that is all the students that leave this course go to a particular new activated course.

We point out that "Business Economics" is a course split in two cities: Forlì and Bologna. So the high *out*-degree is due to closure of the degree course in Forlì. The same for "Legal Sciences", which is split in a course in Ravenna and another one in Bologna: the closure of degree course in Ravenna in 2006-07 has created the high *out*-degree.

As we said before, "Biological Sciences" and "Pharmacy" are "transit" nodes, used by students to go to "Medicine". All other actors in table 3 that have a high normalize *out*-degree can be considered "escape".

TABLE 3
Potential "Escape" degree courses

Potential "Escape" Nodes	*Out*-degree	*In*-degree	# matric.	Norm. *Out*-deg	Degree centrality
YEAR 2004-05					
Social Safety and Control Operator	**161**	0	0	-	2
Business Economics (**)	**150**	6	198	**75.76%**	28
Biological Sciences (*)	**90**	19	432	20.83%	39
Foreign Languages and Literature	**80**	27	761	10.51%	41
Computer Engineering	48	24	154	**31.17%**	29
Telecommunications Engineering	25	3	49	**51.02%**	15
Computational-Informatics Mathematics	7	1	14	**50.00%**	1
YEAR 2005-06					
Economics and management of tourist Services	**250**	0	0	-	9
Biological Sciences (*)	**112**	30	527	21.25%	46
Foreign Languages and Literature	**95**	34	729	13.03%	41
Pharmacy (*)	56	18	141	**39.72%**	28
Chemistry and Pharmaceutical Technologies	44	13	121	**36.36%**	21
Telecommunications Engineering	12	1	33	**36.36%**	6
Cultural Operator	11	17	11	**33.33%**	17
Territory Technician	9	0	18	**81.82%**	2
Scientifical Information on Pharmaceutical Products	6	5		**33.33%**	7
YEAR 2006-07					
Biological Sciences (*)	**106**	20	334	**31.74%**	42
Foreign Languages and Literature	**75**	47	709	10.58%	41
Legal Sciences (**)	**72**	6	25	**288.00%**	35
Management Process Engineering	22	21	40	**55.00%**	10
Telecommunications Engineering	17	4	30	**56.67%**	9
Animal Production Science and Technologies	13	5	33	**39.39%**	9
Cultural Operator	8	6	21	**38.10%**	10

(*) 'Transit' (not 'escape') nodes, use by students to go to 'Medicine'.
(**) Courses split in two cities: their high *out*-degree is due to closure of one of them.

Similar observations regard "attraction" degree courses. In table 4 two degree courses are deleted because their high in-degree is due to a particular closure degree course. Particularly the high *in*-degree of "Sociology and criminology science for safety" is due to closure degree course "Social safety and control operator; the high *in*-degree of "Economics and markets of tourist system" is due to "Economics and management of tourist services" and the great enter link of "Economics and marketing" is due to closure course of "Business economics".

TABLE 4

Potential "Attraction" degree courses

Potential "Attraction" Nodes	*Out*-degree	*In*-degree	# matric.	Norm. *In*-deg	Degree centrality
YEAR 2004-05					
Sociology and Criminology Science for Safety	6	**174**	200	**87.00%**	14
Economics and Marketing	12	**134**	219	**61.19%**	23
DAMS	45	**68**	1114	6.10%	37
Legal Sciences	68	**65**	1235	5.26%	53
Medicine	24	**65**	259	25.10%	28
Physiotherapy	1	18	36	**50.00%**	11
Dentistry	1	13	15	**86.67%**	6
YEAR 2005-06					
Economics and market of tourist System	2	**243**	145	**167.59%**	5
Legal Sciences	76	**76**	1181	6.44%	56
Medicine	20	**61**	266	22.93%	29
DAMS	42	**54**	993	5.44%	38
History	7	**48**	238	20.17%	28
Physiotherapy	1	20	41	**48.78%**	12
Cultural Operator	11	17	33	**51.52%**	17
Dietetics	0	6	5	**120%**	3
YEAR 2006-07					
Medicine	13	**79**	251	**31.47%**	31
Foreign Languages and Literature	75	**47**	709	6.63%	41
Business Economics	4	**46**	130	**35.38%**	21
History	21	**44**	218	20.18%	30
Infant and Nursery School Teacher	20	**42**	320	13.13%	27
Law	20	**42**	1009	4.16%	38
DAMS	31	**41**	743	5.52%	29
Social Educator	33	**40**	218	18.35%	36
Arts	41	**36**	397	9.07%	29
Physiotherapy	1	22	31	**70.97%**	14
Dentistry	1	9	21	**42.86%**	4
Techniques in medical radiology, imaging ...	1	9	17	**52.94%**	9
Techniques in the Biomedical Laboratory	2	7	11	**63.64%**	5
Techniques in Neurophysiopathology	1	4	8	**50.00%**	3

(--) Deleted because their high *in*-degree is due to a particular closure degree course.

All other actors in table 4 that have an high normalized *in*-degree can be considered "attraction" courses, but in the time series visualization the trend of student prefereces is towards "medicine" area, while other potential "attraction" courses tend to balance.

4. ARCHITECTURAL FEATURES OF UNIVERSITY OF BOLOGNA NETWORK

The random graphs theory introduced in the late 1950s, based on Erdős-Rényi (ER) model, assumes that a fixed number of nodes are connected randomly to each others. This network starts with N nodes and connects every pair of nodes with probability p, creating a graph with approximately $pN(N-1)/2$ edges. Therefore most nodes have roughly the same number of links, approximately equal to the network's average degree $\bar{k}$, while nodes that have significanly more or less links than $\bar{k}$ are absent or very few. In probabilistic terms, in ER networks the nodes follow a Poisson distribution with a peak at $\bar{k}$.

Usually when we describe a real phenomena the large events are rare, but small ones quite common. Therefore the random network model cannot explain the topological properties of real networks (Barabási *et al.*, 2003). Paticularly in contrast to the Poisson degree distribution, for many real networks the number of nodes with a given degree follows a power law: the probability that a node has exactly k links is

$$P(k) \sim k^{-\gamma} \tag{2}$$

where γ is the degree exponent (its value is between 2 and 3 for most networks).

That's implies that nodes with only a few links are numerous, but a few nodes (called "hubs") have a very large number of links. Networks with a power-law degree distribution are called scale-free (Barabási and Albert, 1999).

Two mechanisms characterize the scale-free model: *incremental grow*, where the networks grow by the addition of new nodes, which join nodes already present in the system; and *preferential attachment*, characterized by a higher probability to join a node with a large number of connections.

The Bologna University network approximates a scale free topology but, because of "institutional" constraints, has not "incremental grow" and "preferential attachment". The first evidence comes from the analysis of degree courses (see section 3), in which the nodes are the degree courses, while the links are the students' inter-degree courses transfers. As the network is directed, for each degree course an "in" and "out" degree can be assigned that denotes the number of students that "enter" or "leave" it, respectively.

The analysis of 365 different degree courses indicates that this network has a scale-free topology, in which most degree courses have few "enter" or "leave" of students and only a few courses have most students that "enter" or "leave". The degree distribution of this network follows a power law with exponent $\gamma \simeq 1.7$ (see fig. 6).

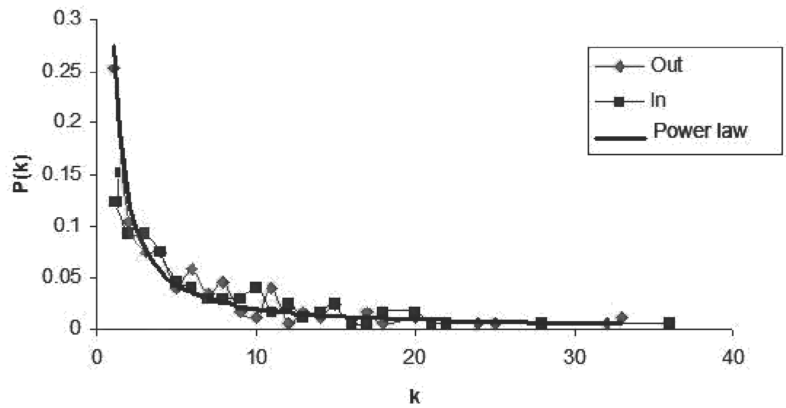


Figure 6 – Degree distribution of University degree courses network.

5. CONCLUSIONS

The structural mechanisms that govern the topology of Bologna University network is a constraint scale-free, where institutional regulation inhibits the free growth and the preferential attachment. Even if there are "hubs", their rule are strongly influenced by different institutional decisions and choices applied to degree courses: such as geographical location, admittance tests with close number, closure of some courses and so on.

Therefore, the evolution of network is not in terms of growth, but in its linkage structure. In time series data, the decrease of isolated nodes and the increase of the number of "attraction" degree courses underline the ability of academic system to fit the courses in order to meet students' interest.

ACKNOWLEDGEMENT

The authors thank Prof. Aura Reggiani for her useful remarks and Dr. Filippo Capriotti for his assistance and suggestions in processing data.

Dipartimento di Scienze Statistiche "Paolo Fortunati" LUISA STRACQUALURSI
Università di Bologna PAOLA MONARI

RIFERIMENTI BIBLIOGRAFICI

A.L. BARABÁSI, and R. ALBERT (1999), *Emergence of Scaling in Random Network*, "Science", 286, pp. 509-512.

A.L. BARABÀSI (2001), *The physics of the web*, "Physics World", July, 33.

A.L. BARABÁSI, and R. ALBERT (2002), *Statistical mechanics of complex networks*, "Reviews of Modern Physics", 74, pp. 47-97.

A.L. BARABÁSI et al. (2003), *Scale-Free and hierarchical structures in complex network*s, "Modeling of Complex Systems: Seventh Granada Lectures", AIP, Melville New York.

B. BOLLOBÁS (1985), *Random Graphs*, Academic Press, London.

L.C. FREEMAN (1979), *Centrality in Social Networks: Conceptual clarification*, "Social Networks", 1, pp. 215-239.

J.B. KRUSKAL, and M. WISH (1978), *Multidimensional Scaling*, Sage Publications, Newbury Park.

J. B. KRUSKAL (1964), *Multidimensional Scaling by optimizing goodness-of-fit to a non-metric hypothesis*, "Psychometrika" 29, pp. 1-27.

A. REGGIANI, and P. NIJKAMP (2006), *Spatial Dynamics, Networks, and Modelling*, Edward Elgar Publishing Ltd.

RIASSUNTO

Identificazione di poli di "attrazione" nell'Ateneo di Bologna attraverso un'analisi delle reti sociali

I trasferimenti degli studenti tra le diverse facoltà e i diversi corsi di laurea dell'Ateneo Bolognese possono essere rappresentati come una rete complessa. L'analisi condotta su tre anni accademici (2004-05, 2005-06, 2006-07) si pone come obiettivi di individuare la struttura e i meccanismi che conducono all'evoluzione di tale rete complessa. Servendosi di analisi grafiche e di alcun misure di centralità opportunamente normalizzate, si è arrivati a individuare alcuni poli di attrazione all'interno dell'ateneo e a descrivere i meccanismi e le preferenze che direzionano la dinamica dei trasferimenti.

SUMMARY

Attraction poles in the University of Bologna: a social network analysis

The movements of students between the faculties of the University of Bologna represent a complex evolving network. By mapping the electronic database containing all relevant data about the students, named "datawarehouse", for three academic years (2004-2007), we aim to infer on the dynamics and the structural mechanisms that lead the evolution and topology of this complex system. Empirical measurements allow us to discover the topological characteristics of the network at a given moment, as well as their evolution in time. The results highlight stable 'attraction' and 'escape' nodes within the University network, and provide a simple description of the student movements and preferences.