Social Network Analysis of an innovative Narrative-Evidence Based model of learning/teaching

Umberto Giani

Department of Preventive Medical Sciences, Faculty of Medicine,
University of Naples Federico II, Italy

Corresponding Author:
Umberto Giani, Department of Preventive Medical Sciences, Faculty of Medicine,
University of Naples Federico II
Via S. Pansini 5, 80131 Naples, Italy
E mail: ugiani@unina.it

Summary
Objectives. The present paper describes an attempt to analyze the dynamics of social contacts among students in order to develop innovative models of learning/teaching medical statistics, aimed at stimulating both creative and critical thinking. The rationale is that learning is usually conceived of as a psychological mechanism concerning the mind of a single individual, disconnected from the social environment in which he or she is embedded. The diffusion of an innovative learning format within the classroom was modeled as a contagion spreading in a structured micro-society.
Materials and Methods. University students starting a course of medical statistics and informatics were given the choice between two alternative assessment formats: i) a traditional assessment based on a written test; ii) an innovative assessment based on a written test plus a working group report designed to integrate narrative- and evidence-based medicine approaches (innovation).
The students were also allowed to use an e-learning system. Two models of the diffusion of the innovation were tested: i) direct influence and ii) indirect influence of structurally equivalent individuals.
A logistic auto-regressive model was used to explore the balance of individual and social factors affecting the spread of the contagion.
Results. Gender, general culture, and indirect social influence were statistically associated with adoption of the innovation. The results are discussed in relation to the development of innovative network-based educational models in medical statistics.
Conclusions. Learning should be conceived of as the outcome of the dynamics of a social network of interacting people. Therefore, analysis of the properties of the structure of the relations among learners could be the basis for the development of an innovative learning/teaching paradigm focused on network-based intervention rather than exclusively on single individuals. The adoption of this paradigm could enhance, in students, the creative and critical abilities that today's innovation-based society absolutely demands.

KEY WORDS: *Social Network Analysis, graph theory, diffusion of innovations, Narrative- and Evidence-Based Medicine, e-learning, creative and critical thinking.*

Introduction

The traditional *knowledge acquisition metaphor* (1) likens learning to a process of filling a sort of container, i.e. the learner's mind, with pieces of information. Hence, learning is, implicitly or explicitly, represented as an individual cognitive mechanism which enables people to construct their knowledge as though they were atomized units disconnected from the social system. Therefore individual cognition remains central to this approach, and learning is considered as a psychological phenomenon concerning a single individual. The metaphor is at root of classic summative assessments based on the assignment of scores to individual students.

By contrast, according to the *participation metaphor,* learning is a process of "becoming a member of a community" through participation in cultural events and social activities (2). Thus, the social community plays a crucial role in the processes of learning and knowledge construction. This suggests that there is a need to develop suitable models for analyzing how the direct and indirect interactions among the members of these communities affect behaviors, opinions, beliefs, knowledge, and learning.

The concept of *evolutionary epistemology* stemmed from the germinal idea proposed by Popper (3), who analyzed the way in which culture spreads and evolves in human societies.

Evolutionary epistemologists (4) highlight two complementary Darwinian processes: the apparently random emergence of cultural innovations (blind variation), and selection (selective fixation) of the fittest ideas. From this perspective, creative thinking (5) is aimed at exploring new ideas (innovation) whilst critical thinking (6) makes it possible to organize, evaluate and select (selective fixation) the (maybe provisional) "best knowledge" within a given domain and a given historical period.

An approach that set out to explore this paradigm scientifically stemmed from Cavalli-Sforza's intuition that there is an analogy between the transmission of genetic traits and the transmission of cultural traits between generations (7). Cavalli-Sforza attempted to apply the mathematical models of genetic transmission to the inter-generational transmission of ideas and culture.

In his book *The Selfish Gene*, Dawkins (8) proposed a gene-centered representation of evolution and explicitly formulated the concept of *memes*. Even though Dawkins became involved in broad epistemological and theological debates, his seminal idea nevertheless gave rise to so-called *memetics* (9). According to this approach, *memes* are the cultural analogues of genes, i.e. elementary cultural units that are transmitted horizontally and vertically in society. The diffusion of memes was likened to the diffusion of rumors and gossip within human populations.

However, although Cavalli-Sforza recognized the role of reciprocal interaction between like-minded individuals in the adoption of cultural traits or behaviors, he did not analyze further the role of social structure in cultural transmission.

Instead, the relational nature of human societies was soon recognized by social psychologists. For example, Moreno (10) conceived of psychological well-being itself as related to "balanced" social configurations of interpersonal choices, attractions, repulsions, friendships, and so on.

A very productive idea, from the perspective of the mathematical analysis of the knowledge transmission process, was the *epidemiological metaphor* according to which the diffusion of innovations is like the spread of an infectious disease among susceptible members of a population (11).

However, the classic SEIR (susceptible-exposed-infectious-recovered) epidemiological models were based on an elementary classification of the members of a population into a handful of categories, e.g. susceptible, infectious, recovered. Therefore, these models did not take into account the way in which contacts among a population's members are structured.

A series of seminal papers by Rapoport, published in the Bulletin of Mathematical Biophysics (12, 13), described the statistics of disease and information spreading through populations structured to varying degrees. These models were at odds with the classic SEIR models because they took into account the structure of the connections among the members of a network. The main assumption was that infectious diseases spread in the social network following the paths of connections among friends and acquaintances.

Despite some restrictions (e.g. finite subpopulation, strong overlapping of friendship circles), Rapoport developed the analytic solution of the temporal evolution of the spread of infectious diseases.

The concept of contagion was also applied to the diffusion of ideas: a cultural unit was conceived of as a sort of virus, and the cultural evolution as an epidemic propagating within a network of individuals. Contagion was conceived of as one of the key factors sustaining group belief systems.

At the same time, sociologists were developing new concepts in an attempt to grasp the structure of the relationships that exist within human societies. For example, Foster (14) introduced new concepts such as homophily (i.e. people's tendency to associate with people "like" themselves), symmetry (i.e. the reciprocation of relations), and triad closure (the tendency of one's acquaintances also to be acquainted with each other); Skvoretz (15) introduced the concept of social differentiation; Granovetter (16) argued that the stronger the tie between two individuals, the higher the proportion of individuals to whom they are both tied.

The scientific community slowly realized that many social, biological, physical, and psychological phenomena are embedded within webs of interdependences, and that there is a deep similarity among apparently different domains: epidemiology, genetics, epistemology, cultural transmission, psychology, sociology, ecology, biochemistry, biology, web dynamics, and so on.

Thus, it is not surprising that the mathematical graph theory was advocated as a sort of common language for the analysis of these apparently different phenomena.

In fact, the *graph theory*, sometimes called the "new" science of networks (17), is a branch of mathematics whose aim is to study the structure of (complex) systems in the real world which are comprised of a number of units interacting locally in relatively simple ways (18).

Relational data can be represented as entity-relation (ER) graphs, in which nodes represent entities and the binary or valued edges represent the relations between them. Examples of ER graphs include citation networks (19), where the nodes are authors, papers, institutions, and journals connected by meetings, co-citations, and so on; the Internet Movie Database (20), where the nodes are actors interconnected to other entities such as movies and studios; ecosystems (21), where the nodes are animals connected by predator-prey relations; e-mail communication (22), where the nodes are the sources and the destinations connected by messages; biochemical systems (23, 24), where the nodes are the proteins or genes connected by metabolic pathways or other chemical reactions; and web-based asynchronous learning networks (25), where the nodes are students connected by means of web interactions.

In human societies, contacts, ties, and meetings, which relate one agent to another, are examples of relational data. The properties of the ensuing relational structures cannot be reduced to the attributes of the individual agents involved because each agent is influenced by several significant others.

A "social network" can be represented mathematically as a graph (or a multi-graph); wherein each entity, *actor or agent* is a node, and the binary and/or valued relations between actors are the links between the corresponding nodes. Actors can be persons, organizations, groups, and so on; links can represent different types of relation: friendship, esteem, competition, and so on.

Heider showed that in group dynamics (26) people who are close to one another tend to adopt similar attitudes towards other people or events, and tried to develop formal graph-based models in order to clarify how social relations affect the attitudes, opinions, behaviors, and beliefs of the members of a given group or community.

The members of a group are more likely to share interests, and to have more frequent interactions. Hence, a similarity of beliefs and behaviors is, at once, the main factor inducing some individuals to form a group, and the result of the interactions occurring within the group, which in turn reinforce the similarity of behaviors and/or beliefs.

The cohesion of groups can be represented as *cliques* of agents connected internally more than externally. *Cohesion* is a primary network property that contributes to the creation of shared beliefs and behaviors.

Social influence occurs when an agent, the so-called Ego, not only adapts its behavior, attitudes and beliefs to the behaviors, attitudes and beliefs of "significant others", but at the same time modifies the behavior, attitudes and beliefs of the Ego's sub-network.

From the pedagogical point of view, each learner, including the teacher, can be conceived of as an agent in a network of individuals with different backgrounds, learning styles, and cultural interests. Thus, one can assume that in educational settings individual learning depends on a combination of personal attitudes and influences by other agents in the network.

Human relational networks are also self-organizing learning systems, and it is they that should be the units of analysis of the learning process, rather than individuals and their isolated cognitive and learning styles.

In fact, according to the *social learning theory* (27), people learn by observing, adopting and imitating, with more or less marked modifications, the behavior of significant others. The *Social Interdependence The-*

ory of Cooperative Learning (28) suggests that the characteristics of the interaction processes are determined by the way in which *social relations* among members of the group are *embodied* in the learning context.

Joint participation in events provides people with the opportunity to meet and interact: events create ties among actors, and actors create ties among events.

According to the evolutionary epistemology paradigm, within the educational context one should stimulate both creative thinking (which leads to the emergence of new ideas) and critical thinking (which leads to the critical evaluation of these ideas). Therefore, within an educational setting one should develop methods for understanding the factors that lead some actors to create and/or adopt a new idea.

Traditional learning/teaching of medical statistics is preferentially oriented towards the critical reading of medical data and literature (this is true of the evidence-based, or EBM, "movement"). By contrast, exploratory and qualitative data analysis, and narrative-based medicine (NBM) are somewhat neglected, even though they imply a more creative intellectual task.

This paper describes the preliminary results of an application of the contagion theory to the adoption of innovative models of creative and critical learning/teaching of medical statistics.

The main assumption is that students, rather than simply accepting traditional learning formats, are stimulated, to greater or lesser degrees, by the opportunity of becoming involved in innovative group activities in which the aim is to construct new knowledge. Therefore, if students were given the opportunity to adopt, freely, an innovative group-based learning format one could analyze the personal and social factors affecting their decision to adopt this innovative format, i.e. the network self-organization process, and whether an individual's propensity to adopt a new idea increases as the proportion of adopters in his/her personal social network increases. Network-based educational intervention could provide a new type of learning/teaching methodology.

Materials and methods

Students starting the medical statistics and informatics course at the Faculty of Medicine of the University of Naples Federico II are invited to register in Dynamic Virtual Learning Networks (DVLN), an online e-learning system based on a constructivistic approach (29). Registration is compulsory but participation is optional and does not affect the student's final scores.

The course is based on weekly modules. Each module starts with a one-hour face-to-face triggering medical situation proposed by a clinician. In accordance with the problem-based learning approach (30), the students do not have sufficient knowledge to solve the problem, which they must thus try to structure, looking for the knowledge required to solve it.

Then, two two-hour face-to-face interactive lectures introduce the statistical concepts needed in order to structure and solve the problem.

A final one-hour face-to-face multidisciplinary, interactive seminar summarizes the knowledge acquired, and anticipates in a fuzzy way the topics of the next module.

At the end of the week a set of open questions are uploaded into DVLN.

DVLN allows students to participate in different online activities and events that are recorded by the system.

In the present paper the following e-learning student-driven activities were tracked:

a. Forum discussions
b. Registering of interesting websites related to course topics and/or to the specific problem to be solved
c. Literature citations (literary masterpieces, novels, historical or philosophical essays)
d. Registering, in the web-based glossary, of the meanings of newly acquired terms

The individual frequency of each DVLN activity was transformed into a binary variable by dichotomizing at the upper 75th percentile.

The students' scores (for general culture, biology, chemistry, and mathematics) upon entry to the Faculty of Medicine were first orthogonalized by means of factor analysis. Then the standardized factor scores were dichotomized at the value of 1.

Hence, each student was described by a set of eight binary personal attributes plus gender.

The rules of the final examination were negotiated with the students. In particular, they could choose freely between two different assessment procedures:

1. A traditional assessment based on a written, open-questions test (T-students)
2. An innovative assessment format (hereinafter, innovation) based on the same open-questions test plus a written group report based on a qualitative interview of a patient and a group discussion of statistical papers relating to the diseases the interviewed patient suffered from. The working groups were formed freely by the students themselves (I-students).

The written tests were blindly and independently rated by two raters who did not know whether the student had accepted or refused the innovation.

The negotiated maximum score attainable by the T-students was 27/30. The I-students, by contrast, also had to take a risky oral examination (a discussion of their group report), but these students had the possibility of attaining the full score of 30/30 with honours.

In fact, the I-students had to interview different patients (one for each member of the group). Each interview was de-structured into elementary meaning units (MU), according to an online version of content analysis (30). The MU were organized as a sequence of episodes by using an ontology of illness narratives (31). The illness narratives were discussed within the group in order to find their communalities and specificities. Moreover, the students also had to collect EBM literature relevant to the themes emerging from the analysis of the narratives. The oral assessment of the I-students was a group examination in which the aim was to evaluate the knowledge and role of each member of the working group.

A binary vector, y, of the adoption of the innovation was obtained. This vector is the outcome variable of the present study.

The T-students and the I-students who passed the examination were invited freely to mention one or more students with whom they had been in "meaningful contact" during the course. Obviously, the I-students had to exclude the members of their working group.

The directed graph (di-graph) of the meaningful contacts (mentions) was represented as an N×N adjacency binary matrix, X, where $x_{ij}=1$ if the i-th student mentioned the j-th student, and zero otherwise.

For each node, the coefficient of betweenness (Be) was computed: Be_i is the number of paths passing through the node i. Betweenness is therefore a measure of the centrality of each node, and was added to the set of personal attributes describing each student. Thus, an Nx10 matrix, A, represented the personal characteristics of each student.

The mutuality index was computed as

$$MI = \frac{|x_{ij} = 1 \cap x_{ji} = 1|}{|x_{ij} = 1 \cup x_{ji} = 1|}$$

The Katz-Powel coefficient of mutuality was computed as:

$$\rho_{kp} = \frac{2(N-1)^2 M - L^2 + L_2}{L(N-1)^2 - L^2 + L_2}$$

where M is the number of reciprocated relations,

$$L = \sum_{i=1}^{N} x_{i+}$$

and x_{i+} is the number of mentions made by the i-th student, and

$$L_2 = \sum_{i=1}^{N} x_{i+}^2 .$$

This coefficient is equal to 1 when every mention is reciprocated, whereas it is equal to zero when the mentions are independent.

The overall density of the adjacency matrix was computed as:

$$d = \frac{L}{N(N-1)}$$

The adoption of the innovation was modeled as a combination of personal attitudes or attributes and social influence.

Two network influence matrices were defined, i.e. the matrix of direct ties, ω_1, and the matrix of structural similarity, ω_2.

The matrix of direct ties, ω_1, was simply the adjacency matrix, X, excluding self-ties, i.e. the main diagonal of X was set equal to zero.

However, two individuals, i and j, can play the same role in the social network, even if they are not in direct contact with each other.

Two individuals, i and j, are *structurally equivalent* if

they are connected in the same way to the other members of the network. Two nodes are structurally equivalent if they have the same profile of links with the other nodes in the network.
For example, in figure 1 the nodes A and F are structurally equivalent because they are connected in the same way to the nodes B, C, D and E, even though there is no connection between A and F.

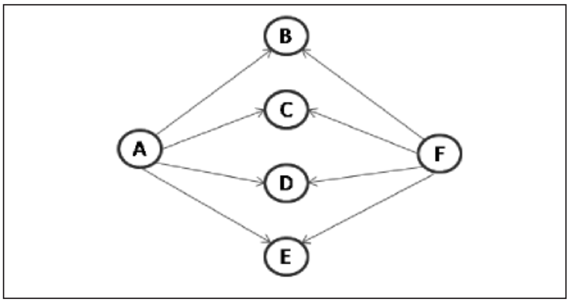

Figure 1. An example of structurally equivalent nodes.

Since B, C, D and E can be connected to other nodes, then A and F could indirectly influence the same subnetwork of nodes. The concept of *structural similarity* weakens the excessively demanding requisite implicit in the definition of structural equivalence. Structural similarity is usually represented as a generalized version of the Pearson correlation coefficient by pooling the correlations between the columns and between the rows of the adjacency matrix. Other measures can be used, too, e.g. the proportion of positive pattern matching. In the majority of cases these measures are practically equivalent.
Two nodes, i and j, are structurally similar if the patterns of their contacts with the other members of the network are similar.
The adjacency matrix, X, was transformed into the matrix of structural similarity, ω_2, as follows:

$$r_{ij} = \left(\frac{\sum_k (x_{ki} - \bar{x}_{\cdot i})(x_{kj} - \bar{x}_{\cdot j}) + \sum_k (x_{ik} - \bar{x}_{i\cdot})(x_{jk} - \bar{x}_{j\cdot})}{\sqrt{\sum_k (x_{ki} - \bar{x}_{\cdot i})^2 \sum_k (x_{kj} - \bar{x}_{\cdot j})^2} \sqrt{\sum_k (x_{ik} - \bar{x}_{i\cdot})^2 \sum_k (x_{jk} - \bar{x}_{j\cdot})^2}} \right)$$

where the sums are over k, and i≠k, j≠k. r_{ij} measures the degree of structural equivalence between pairs of actors: if two actors are structurally equivalent then r_{ij} is equal to 1. The values of the symmetric matrix, R, were ordered from the lowest to the highest, and the value corresponding to a sharp upward inflection was selected as the threshold for dichotomizing the matrix. So, R was dichotomized at the value of 0.25. The final structural equivalence matrix, ω_2, was computed as:

$$\omega_{ij} = r_{ij}\,(1 - x_{ij})$$

where the multiplication by $(1-x_{ij})$ prevents the structural similarity matrix, ω_2, from being blended with adjacency matrix, X.
Two different logistic analyses were performed, taking into account two different socio-matrices

$$\log\left(\frac{p(y=1)}{1-p(y=1)}\right) = \alpha_i + A\beta_i + \lambda_i \omega_i y$$

where the index *i* indicates the influence graph (i=1,2), β_i is the vector of the estimated coefficients of individual attributes, and λ_i the estimate of the coefficient of social influence.

Results

A total of 116 students (50.9% males, 49.1% females) were mentioned by other students, as having been engaged with them in meaningful contact. These mentions were analyzed. The average number of mentions per student was 152/116=1.31.
Forty-nine of the 152 mentions were of students who did not pass the final examination (isolated students). Seventy-one students (61.2%) adopted the innovation. Thirty-two (27.6%) students were active participants in the online activities and 84 (72.4%) were lurkers.
Figure 2 shows the network of direct ties (matrix ω_2). The density of the matrix was equal to 0.0048.
Figure 3 shows the percentage distribution of the out-degree of the non-isolated students.
The average out-degree of the non-isolated students was 1.75.
Figure 4 shows the frequency distribution of the geodesic distance, i.e. the shortest path (in terms of number of edges) connecting two points. The average geodesic distance among reachable pairs was 1.8.

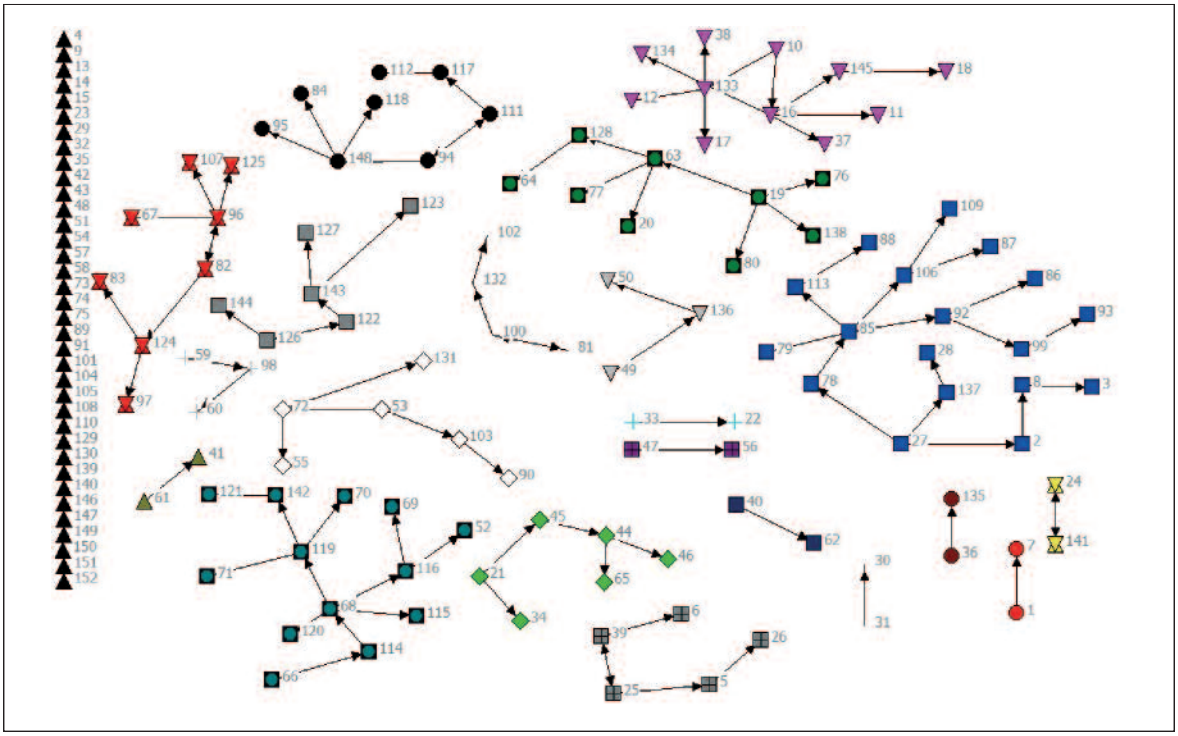

Figure 2. Adjacency matrix. Different shapes indicate different weak components. The isolated students are represented on the left side.

The mean betweenness was equal to 1.13 (SD 2.8). The index of mutuality (MI) was 0.084; the Katz-Powel mutuality coefficient was ρ_{kp}=0.15.

The fragmentation index (i.e. the proportion of nodes that cannot reach each other) was 0.96. The overall index of distance-based cohesion (this index ranges from 0 to 1, with larger values indicating greater cohesiveness) was 0.007.

The adjacency matrix can be decomposed into different weak components. A strong component is a set of nodes in a sub-graph that can reach one another via one or more paths, but that lack any connections outside the sub-graph. The definition of "weak component" is the same, but it disregards the direction of the lines connecting the points. In this paper only weak components are taken into account.

The component sizes ranged from 2 to 18. The component size heterogeneity was 0.95. Thirteen components with three or more members were found.

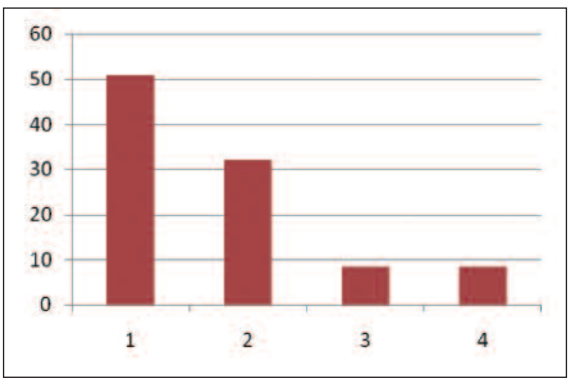

Figure 3. Out-degree distribution.

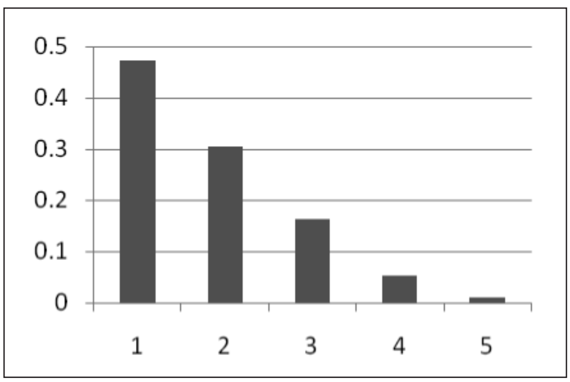

Figure 4. Geodesic distance distribution.

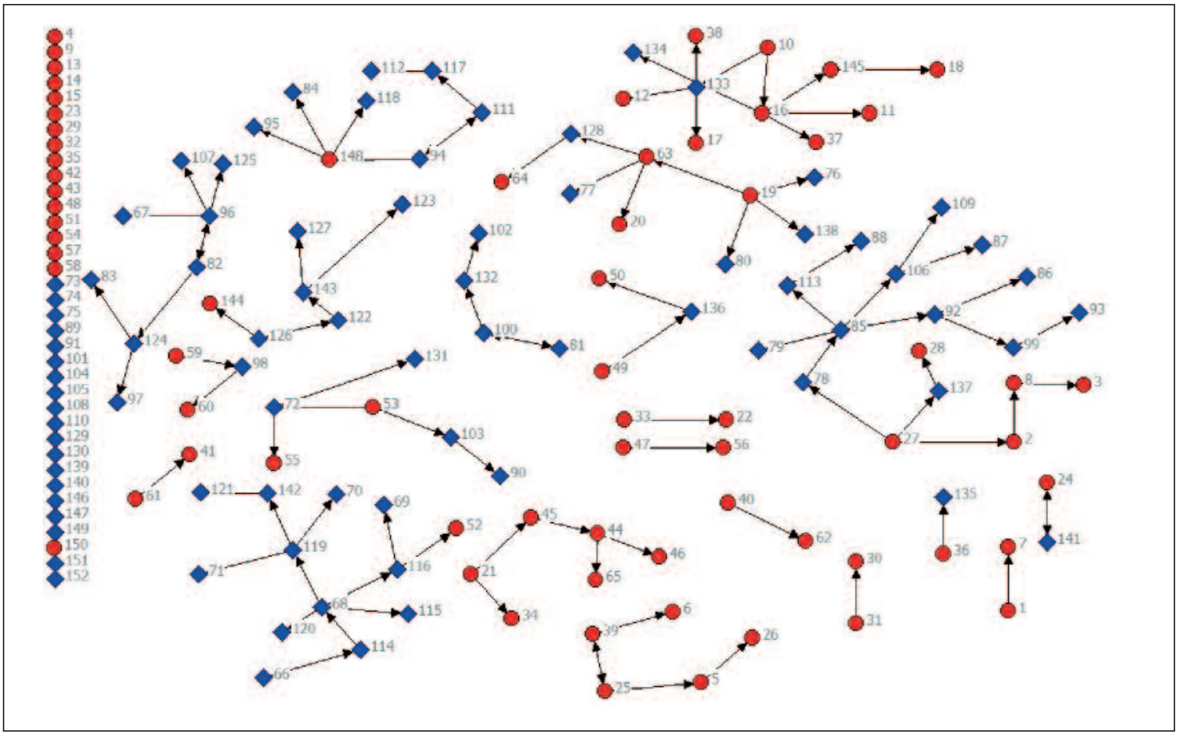

Figure 5. Adjacency graph: diamonds=I-students; circles=T=students.

Figure 5 shows the mapping of the T-students (circles) and the I-students on the adjacency graph.

The number of isolated students is approximately similar in the two groups.

Taking into account only the mentions of non-isolated students, 73% of the adopters mentioned other adopters, while 60% of the non-adopters mentioned other non-adopters (Table 1).

Table 2 shows the parameter estimates of the logistic regressions.

The application of the first model, ω_1, shows that direct influence is not associated with adoption of the innovation, whereas gender, literature citations, culture and degree of betweenness do show statistically significant associations.

The application of the second model, ω_2, shows that indirect influence, gender, literature citations and culture are statistically associated with the adoption of the innovation.

Discussion

The present paper should be regarded as a sort of action-research study rather than a pre-planned experiment, and the results should be interpreted taking this limitation into account.

The rationale of this study is that despite the fact that in educational settings the encounters among students are knowledge-productive in many ways, con-

Table 1. Proportion of mentioned adopters.

		Target		
		Adoption	Non Adoption	
Source	Adoption	46 (73%)	17 (27%)	63 (100%)
	Non Adoption	18 (40%)	27 (60%)	45 (100%)

Table 2. Binary logistic regression.

	ω_1			ω_2		
	OR	Lower 95% C.I.	Upper 95% C.I.	OR	Lower	Upper
Influence	0.90	0.46	1.77	2.74**	1.59	4.74
Gender (M vs F)	3.70**	1.68	8.17	5.85**	2.32	14.79
Forum (High vs Low)	0.91	0.28	2.99	0.88	0.25	3.14
Glossary (High vs Low)	1.37	0.42	4.53	1.00	0.27	3.72
Websites (High vs Low)	1.71	0.49	6.01	2.08	0.51	8.56
Literature (High vs Low)	7.15*	1.40	36.51	5.55*	1.02	30.24
Culture (High vs Low)	4.53*	1.48	13.82	5.37*	1.57	18.32
Biology (High vs Low)	1.50	0.57	3.96	1.14	0.39	3.33
Chemistry (High vs Low)	0.80	0.19	3.46	0.67	0.13	3.43
Mathematics (High vs Low)	2.02	0.69	5.90	2.04	0.68	6.11
Betweenness	1.37*	1.02	1.85	1.06	0.96	1.17
Constant	0.25			0.12		

ω_1: direct influence, ω_2: indirect influence, OR and 95% C.I. *: $p<0.05$, **: $p<0.01$.

sideration is rarely given to the dynamics of the social networks in the learning/teaching process.

One can assume that there are essentially two types of interaction between students: random and systematic encounters.

Random encounters can be conceived of as a sort of background noise which enables students to exchange information, explore and elaborate doubts, modulate learning styles, share didactic materials, and so on.

However, students also tend to form, spontaneously more or less stable and systematic groups, on the basis of attractions, repulsions, common social and knowledge backgrounds, and so on.

According to the evolutionary epistemology paradigm, educational settings should be oriented to promote and support both creative and critical attitudes, and innovations often stem from the interactions within subgroups of "like-minded" people. Thus, it is important to develop pedagogical formats which make it possible to analyze and exploit the mix of personal attitudes and social factors that influence the adoption of innovative ways of learning.

The objective of the present work was to explore whether the adoption of an innovative learning format is influenced by a mix of individual attitudes and social influences, and to compare the relevance of direct and indirect contagion in this adoption.

The main hypothesis was that adoption of an innovation is influenced by the interaction among like-minded students. The innovation was characterized by:

1. Free choice of the members of the working group
2. Free choice of the patients to be interviewed
3. Free choice of the relevant scientific literature
4. Free choice of the EBM topics to be critically appraised
5. A new assessment format.

In this way, in accordance with the evolutionary epistemology paradigm, the groups were involved in a set of creative and critical activities.

In principle, the innovation is the more demanding option due to the additional work involved and the increased risk introduced by the oral examination, which must be taken in addition to the written test. In fact, according to the utility theory people wish to avoid negative consequences, while desiring positive results or effects. If people expect a positive outcome from a given behavior, or think there is a high probability of a positive outcome, then they will be more likely to adopt that behavior.

Seventy-one students (61.2%) decided to adopt the innovation. This figure is somewhat surprising because the utilities of the gamble would be expected to discourage this choice.

Perhaps, the perceived risk is overcome by the prospect of involvement in a more exciting, creative and socialized form of learning, and/or by the possibility of achieving a higher final score. Moreover, the chance to interview a real patient and to apply, side by side, qualitative and quantitative methodologies could be a powerful motivating factor since medical disciplines are usually perceived as practical arts whose aim is to help individuals, and this can conflict with the abstract and population-based characteristics intrinsic to the statistical approach. Finally, it is worth noting that the main course textbook is an interactive e-book (32) which is written as a detective story in which the main characters are confronted with a mysterious medical enigma that can be solved by acquiring statistical knowledge. The continuous interplay between statistical and medical and social concepts could be another motivating factor.

The working group activities were mediated by face-to-face and online interactions.

However, the working groups were not isolated because their members interacted with the members of other working groups as well as with T-students, and it is quite possible that the adoption of a given behavior was also due to influences coming from indirect contacts, e.g. from the friends of friends.

The direct relationships were elicited by inviting each student to mention a list of other students with whom they had shared meaningful ideas and experiences related to the study of medical statistics.

The average number of mentions per student was 1.31; 49 of the 152 mentions were of students who did not pass the examination (isolated students).

However, even though eliciting mentions of social contacts is a standard methodology in social networks analysis, there are various critical issues that should be taken into account.

From the empirical point of view, there are many difficulties in obtaining detailed and reliable accounts of inter-agent relationships. For example, it is well known that many people are poor estimators of their number of friends; the number is highly sensitive to the interpretation of locutions such as "meaningful relationship"; the framing of the question, i.e. the way in which it is formulated, affects the citing of meaningful contacts; also, the number and the quality of contacts change over time. Moreover, in the present work the mentions were elicited at the end of the course and only the students who passed the final examinations were asked to take part in this part of the research.

Thus, one can question the reliability of a student's mentions, and the results must be interpreted cautiously, taking these limitations into account. In this respect, the present paper does not have any unrealistic inferential demands, and should be interpreted as an attempt to arrive at guidelines that may serve as a foundation for the development of a new model of learning/teaching medical statistics.

From the theoretical point of view, it is necessary to define properly the notion of social distance or proximity between students.

Using the classic cases-by-variables rectangular representation of data, similarity/dissimilarity can be described as a proximity/distance measure between two individuals over a set of defining variables. This implies the representation of a set of individuals in a common metric space. This is the option underlying, for example, cluster analysis of cases, multidimensional scaling, and so on.

However, social systems seem to violate the triangle inequality theorem, which states that three points in a given space can always be connected via the three sides of a triangle whose lengths must obey the inequality

$$d(a,c) \leq d(a,b) + d(b,c)$$

where d(.,.) is a metric measure of the distance between the points.

However, this inequality does not necessarily hold true in social systems because it is quite possible for a person to be acquainted with both person j and person k, even if j and k are not even remotely familiar with each other. This applies in the majority of human settings.

Some authors claim that this violates the notion of distance itself, since triangle inequality is one of the basic properties of the metric spaces, upon which traditional statistical analyses are ultimately based.

So, these critics claim that the idea of a metric social space is a somewhat slippery notion, both from the theoretical and the operational point of view. Moreover, social distance is often multi-valued and multiplexed.

This is one of the main reasons underlying the choice of a graph-based approach to the analysis of human societies. In fact, from the theoretical point of view, in the case of the graph, distance is actually defined in terms of network connections, since a network does not necessarily exist in any particular space.

Several graph-based structural measures of inter-agent distances or proximities have been proposed.

In di-graphs, a source node can be connected to a number of target nodes: this number is its out-degree. A target node can receive connections from a number of source nodes: this number is its in-degree. The out-degree describes the propensity of an individual to be in contact with other people, i.e. his expansiveness; the in-degree is a sort of popularity index of a given node.

The relational structure of the students' mentions of other subjects showed that the average out-degree was 1.75, i.e. each student tends to be in "meaningful" direct contact with one or two other students.

The distribution of the geodesic distances confirmed that there is a short-range distance between the students. However, the geodesic distribution shows that even though the majority of the paths have a length of one, some nodes are located at a distance greater than one and can be reached in a maximum of five steps. The average betweenness was equal to 1.13, but the variability was wide.

So, there exist chains of relations that allow the diffusion of knowledge and information between nodes that are not directly connected.

The network is far from the complete graph containing N(N-1) edges, and it is highly fragmented, since 96% of the nodes cannot be reached by each other. This figure also includes isolated nodes.

The structure of the network showed several weak components, i.e. cliques of students who are directly or indirectly connected, but have no connection with the students of other cliques. The sizes of the components range from dyads to large agglomerates: the biggest is composed of eighteen students.

The absence of inter-component links does not mean that there are no contacts between students of different components. Probably these random and volatile encounters are the basis of the rapid diffusion of basic information among the students. However, leaving aside these background, "random" contacts, the components indicate that the class is made up of "factions", each composed of individuals who are more in touch with each other than with people outside the "faction".

The low Katz-Powel coefficient and the low mutuality index show that only a minority of the relations were reciprocated. This could indicate that the majority of the mentions concerned superficial relations, or that the term "meaningful contact" was interpreted as an asymmetric relation with a target student from whom the source student merely received information or help. The somewhat fuzzy framing of the question does not make it possible to disambiguate between these interpretations.

It is interesting to note (Table 2) that the innovators, or adopters, tended to mention other innovators, and the non-innovators other non-innovators. This could be symptomatic of the fact that innovators and non-innovators represent two different classes of cognitive or learning styles, of conceptions of learning, and of attitudes towards knowledge and culture.

The e-learning system could be a factor facilitating innovator-innovator interactions. In fact, the working group project gallery can be freely accessed via the virtual classroom, and it is quite possible that the adopters shared ideas, experiences and doubts with other innovators. Some students could also be induced to adopt the innovation by their perusing of the project gallery.

However, the adoption of the innovation can also be explained on the basis of individual factors.

In order to evaluate the balance of personal and interactive factors in the adoption of the innovation, each student was profiled in relation to different kinds of personal attribute.

a. Gender
b. Degree of betweenness
c. Scores on entering the faculty
    a. General culture
    b. Biology
    c. Chemistry
    d. Mathematics
d. Participation in the online activities
    a. Forum discussions
    b. Registration of interesting websites
    c. Glossary construction

In fact, it is quite possible that there is a gender-based difference in the propensity to adopt the innovation.

Similarly, the position in the network, and in particular the degree of betweenness, can be an indicator of a student's inclination to construct inter-personal relations. Hence, a student who is on the trajectory of different paths can be influenced by and can influence other students.

The students' Faculty of Medicine entrance scores can be considered rough, indirect indicators of the preferential cultural interests and cognitive style of each student. The mean scores were:

Culture:	19.5 (SD 3.6)
Biology:	14.9 (SD 2.6)
Chemistry:	10.2 (SD 2.1)
Mathematics:	5.5 (SD 2.6)

The very low scores in mathematics may explain, at least in part, the difficulties encountered in learning/teaching medical statistics.

The rationale for choosing e-learning activities as indicators of personal attitudes is based upon the consideration that participation in the same online activities could be seen as a means by which like-minded people are brought together, and at the same time as one of the factors leading them to share ideas, and reinforce common beliefs and/or attitudes.

Participation in the DVLN activities/events can be categorized broadly as follows.

- Active actions
    - *Socializing actions*, e.g. registering an interesting website, a definition of a term for the glossary, or interesting literary masterpieces, novels, essays, and so on.
    - *Communicative actions*: putting questions to the teacher, taking part in the forum discussions, sending messages to other colleagues or to the teacher.
- Passive actions
    - *Lurking or predatory actions* e.g. just taking the didactic materials (summaries, open questions, simulations) and/or taking a look at the forum discussions, registered websites, glossary, etc. without posting anything.

Despite the large number of accesses to the system, only 27.6% of the students actively participated in the e-learning activities, whereas the vast majority were lurkers or predators. This fact need not necessarily be interpreted negatively. In fact, it is well known that intelligent, profound and reflective individuals are often shy and not cut out for social communication and interaction.

A further aim of the present analysis was to establish whether the adoption of the innovation was due to the diffusion of the contagion by means of direct or indirect interactions. Several types of influence graph can be defined, and the choice of influence relation is a crucial step for exploring the "mechanism" of the contagion, because the structure of the relationships generates different learning/teaching dynamics. For example, a fully connected di-graph (i.e. a complete graph) would reflect a community of equally contributing cooperators, whereas a star-shaped di-graph would reflect a transmission-receiver style of learning. The latter is implicit in traditional one-to-many lesson-based learning/teaching. Real learning networks lie somewhere between these extremes.

The balance between personal characteristics and social influence was assessed by means of a logistic auto-regression model:

$$\log\left(\frac{p(y=1)}{1-p(y=1)}\right) = \alpha_i + A\beta_i + \lambda_i \omega_i y$$

Two different contagion models were tested. The first model was based on the assumption of transmission of the contagion by direct influence. In this case, the influence matrix is just the adjacency binary matrix of the students' mentions.

The second model was based on the assumption of contagion via indirect influences by structurally similar nodes.

The direct influence, ω_1, was not statistically significant whilst the indirect influence, ω_2, was associated with adoption of the innovation. This means that the social influence or contagion stemmed mainly from indirect contacts among the students. This is confirmed by the fact that betweenness was significant under the structural relation ω_1, whereas under the structural relation ω_2, betweenness was likely absorbed into the effect of structural equivalence.

Gender was statistically associated with adoption of the innovation under both the relational structures, ω_1 and ω_2: males showed a greater proneness to adopt the innovation. Although no statistical association emerged between gender, culture and influence, it can be hypothesized that males have a greater propensity towards exploration and risky situations.

Similarly, general culture was statistically associated with the adoption of the innovation under both the relational structures. While one can question the reliability of the entrance scores as indicators of general culture, one can tentatively explain this statistical association by considering that the innovative learning format in our setting is based on human interactions with real patients and on the analysis of narratives of illness. One can hypothesize that this type of innovation will appeal more to individuals with a more humanistic bent, and/or that high general culture scores are an indirect indicator of a divergent cognitive style.

This interpretation is indirectly confirmed by the fact that, among the e-learning activities, only the citation of literature was statistically associated with the adoption of the innovation.

What practical suggestions, for designing medical statistics courses, can be drawn from this analysis?

The first consideration is that it should be possible to grasp the structure of the relations among the students at the very beginning of a course. This analysis of the relational structure could enable the teacher to plan network-based interventions. For example, in order to facilitate the diffusion of knowledge, favorable paths of influence, key-player students, and so on, might be identified.

Second, it should be possible to identify the personal attitudes of the students that could facilitate or impair their learning of statistical concepts. In the present paper, only indirect indicators, such as the entrance scores, were taken into account. This is not an optimal choice. Instead, it would be interesting to design a standardized tool aimed at eliciting the student's attitudes.

In the present study, the role of the e-learning is somewhat problematic. There were more than 21,000 accesses to the system in four months. This could be taken as an indication that the e-learning system was a great success. However, closer analysis shows that active actions accounted for few of these accesses. This probably means that actual participation in these e-learning activities subsumes different abilities and cognitive styles unrelated to the adoption of the innovation and/or a proneness to risk taking. For example, it is not unusual to find that young people with expertise in the technicalities and use of computers have a more limited cultural background and a rather "narrow-minded" outlook than those who do not have this expertise.

The e-learning system can be considered as general container for exchanging various kinds of information, and most importantly, as the main channel of communication between teacher and students. Moreover, the e-learning system allowed the students to divulge their working projects, given that the project gallery can be accessed via the classroom. This allows students to appreciate the interests of other students who adopted the innovation. A more detailed analysis of the role of the e-learning system is ongoing.

Finally, it is worth mentioning some statistical issues that were not tackled in this paper. In fact, the general statistical problem is the choice between two different models:

$$\text{H1: } y=\lambda_1\omega_1 y + X_1\beta_1$$
$$\text{H2: } y=\lambda_2\omega_2 y + X_2\beta_2$$

where ω_1 and ω_2 are two influence matrices, X_1 and X_2 are (Nxk_1) and (Nxk_2) matrices of explanatory variables, some of which may be included in both X_1 and X_2, and k_1 need not to be equal to k_2.

To the knowledge of this author, this problem has not yet been solved.

In conclusion, the results and the limitations of the present study underline the need for a more in-depth exploration of network-based approaches to learning and teaching.

References

1. Sfard A. On two metaphors for learning and the danger of choosing just one. Educational Researcher 1998; 27: 4-13.
2. Lave J, Wenger E. Situated Learning. Legitimate Peripheral Participation. Cambridge: Cambridge University Press, 1991.
3. Popper KR. Objective Knowledge: An Evolutionary Approach. London: Oxford University Press, 1972.
4. Campbell DT. Evolutionary epistemology. In Radnitzsky G, Barteley WW (Eds) Evolutionary Epistemology, Rationality, and the Sociology of Knowledge. Les La Salle, Illinois: Open Courts 1993; 47-120.
5. Craft A. Creativity Across the Primary Curriculum. London: Routledge, 2000
6. Mayer R & Goodchild F. The Critical Thinker. Chicago: Brown & Benchmark, 1995.
7. Cavalli Sforza LL, Feldman M. Cultural Transmission and Evolution. Princeton: Princeton University Press, 1981.

8. Dawkins R The Selfish Gene (30th Anniversary ed.). Oxford: Oxford University Press, 2006.
9. Blackmore S. Evolution and memes: the human brain as a selective imitation device. Cybernetics and Systems 2001; 32: 225-255.
10. Moreno JL. Sociometry, Experimental Method and the Science of Society. Beacon, New York: Beacon House Inc., 1951.
11. Giani U. An epidemiological model of knowledge diffusion in virtual learning networks. An application to learning statistics in medicine. Cybernetics and Systems 2005; 36: 445-456.
12. Rapoport A. Spread of information through a population with sociostructural bias. I. Assumption of transitivity. Bulletin of Mathematical Biophysics 1953; 15: 523-533.
13. Rapoport A. Spread of information through a population with sociostructural bias. II. Various models with partial transitivity. Assumption of transitivity. Bulletin of Mathematical Biophysics 1953; 15: 535-546.
14. Foster M, Rapoport A, Orwant CJ. A study of large sociogram: elimination of free parameters. Behav Sci 1963; 8: 56-65.
15. Skvoretz J. Random and biased networks: simulations and approximations. Social Networks 1985; 7: 225-261.
16. Granovetter MS. The strength of weak ties. A network theory revisited Sociological Theory 1983; 1: 203-233.
17. Watts DJ, Dodds PS, Newman ME. Identity and search in social networks. Science 2002; 296: 1302-1305.
18. Watts DJ, Strogatz SH. Collective dynamics of ‘small-world’ networks. Nature 1998; 393: 440-442.
19. Zitt M, Bassecoulard E. Development of a method for detection and trend analysis of research fronts built by lexical or co-citation analysis. Scientometrics 1994; 30: 333-351.
20. Bradley M, Airoldi E, Carley KE. A network analysis model for disambiguation of names in lists. Computational & Mathematical Organization Theory 2005; 11: 119-139.
21. Montoya JM, Pimm S, Solè RV. Ecological networks and their fragility. Nature 2006; 442: 259-264.
22. Broido A, Claffy KC. Internet topology: properties of IP graphs. IEEE SPIE Conference. Denver, Colorado, 2001.
23. Vazquez A, Flammini A, Maritan A, Vespignani A. Modelling of protein interaction networks. Complexus 2003; 1: 38-44.
24. Bahn A, Galas D, Dewey TG. A duplication growth model of gene expression networks. Bioinformatics 2002; 18: 1486-1493.
25. http://www.sloan-c.org/publications/jaln/index.asp
26. Gawroński P, Gronek P, Kułakowski K. The Heider balance and social distance. Acta Physica Polonica 2005; 36: 2649-2558.
27. Ormrod JE. Human Learning (3rd ed.). Upper Saddle River, NJ: Prentice-Hall, 1999.
28. Skon L, Johnson DW, Johnson R. Cooperative peer interaction versus individual competition and individual efforts: effects on the acquisition of cognitive reasoning strategies. Journal of Educational Psychology 1981; 73: 83-92.
29. http://elearning.medicina.unina.it/dvln
30. Giani U, Martone P. Distance learning, problem based learning and dynamic knowledge networks. Int J Med Inform 1998; 50: 273-278.
31. Giani U. Collaborative Multi-professional Narrative Based e-learning. An application to teaching Medical Statistics. Proceedings of International Conference on Interactive Computer Aided Learning. Villach (Austria), 2007.
32. http://www.scriptaweb.it/pensiero_critico_nelle_scienze_ della_salute