UNIVERSITY OF NAPLES FEDERICO II

DEPARTMENT OF INDUSTRIAL ENGINEERING
SECTION OF MANAGEMENT ENGINEERING

PHD PROGRAM IN SCIENCE AND TECHNOLOGY MANAGEMENT
XXVI CYCLE

THE EVOLUTION OF RESEARCH NETWORK FOR A UNIVERSITY DEPARTMENT

PHD CANDIDATE: **RAFFAELLA CICALA**

PHD SUPERVISOR: **PROF. GENNARO IMPROTA**

Contents

3. Software for dynamic analysis

4. Collaborative networks in research

5. DIEG case study

6. The case study: co-authorship network

Introduction

Interactions between people are *ubiquitous*: when people make phone calls, speak with other people, connect on the web, send an email, and so on; each and many others of these actions involves people to become members of many different *social networks* as these actions can be collected as relationships between them.

A social network is formally defined as a structure made up of a set of actors (two individuals that make phone calls or that exchange emails), and a set of social ties (phone calls or emails) between them; a social network can be modeled by a *graph* in which the nodes represent the actors while the arcs are ties. The graph representation is universally used to describe a social network as the evident advantage consists in an immediate interface that allows understanding the way with which the actors are connected.

In principle any entity that can be connected to others can be studied as actor, so the range and type of actors and ties can be quite extensive. For instance, the concept of social network can be applied to large-scale phenomena, such as the world trade; in this case, the countries represent the nodes of graph while import/export of goods between countries are arcs.

Thus, on the basis of typology of actors and relations, different kinds of social network can be identified.

The Social Network Analysis (SNA) is an interdisciplinary methodology that conceptualizes social life in terms of relational network existing among actors. At the heart of SNA is the mathematical branch, called *graph theory,* which focuses on the quantitative operations on networks. Thus, SNA can be seen as a specific application of graph theory as it uses its terms, concepts, and algorithms for studying social relations existing among actors, their structures, their properties, and what determines these properties and what consequences they have for the actors or the network as a whole.

Several methods on collecting relational data exist. Relational data can be collected through observations (widely used in field research to study groups of people who have face-to-face interactions), from archives and historical materials (Gould, 1995), or from trace observation of electronic communications (Carley, 2006). Other methods that involve directly actors of network are surveys and interviews but several problems are linked to them. Fundamentally, surveys and interviews collect relational data by asking to respondents (actors of network) to report with whom they share particular relations. But providing information requested often is difficult for respondents as they interpret relation in different ways and, often, they forget the others with whom they have relations. Besides, designing of surveys and interviews present related issues. “Surveys require complicated patterns of skips and loops, with questions not only being asked or skipped based on previous answers, but questions also being created by incorporating previous responses” (Scott and Carrington, 2011). For these and other reasons, surveys and interviews must be designed and realized with great care.

Relational data collected are typically recorded in the form of *sociomatrix* in which the rows and columns represent actors of network and the elements represent the presence or absence of relationship between each pair of actors (adjacency matrix).

On collected relational data, the using of SNA techniques can produce statistics that yield information about the connectedness, distance, and grouping of the network as well as information about the position of single node within the network. Measures for nodes focused on concept of centrality, that allows identifying nodal properties as a function of its position, relating to the structural importance or prominence of a node in the network (Borgatti et al., 2009). In analysis of centrality, three main indices are considered: degree centrality based on the idea that having a large number of ties, makes a central node; closeness centrality, based on idea that being reachable by others at shorter path lengths, what makes a central node; betweenness centrality, based on idea that being in between many other nodes what makes a central node.

At global level, the calculating of network cohesion represents an aspect very important. In general, a high cohesion indicates that a network contains a large number of ties; thus, more ties between actors yield a tighter structure, which is presumably more cohesive. Many measures to detect cohesion such as density, average degree, and centralization exist. For instance, the density, one of the most well known measures for representing cohesion, indicates the percentage of existing versus all possible ties. A closely related measure of structural cohesion is the average degree of the network. It is often a measure of the cohesion more intuitive than density as it indicates the average number of ties for single node.

The growing popularity of SNA due to its applicability to every context has coincided with the growing awareness on importance of knowledge of relational network in any field; in particular, in the last decade, the emphasis on the importance of *networking* in practical management and the proliferation of *social networking* websites have contributed to recent growth of interest in the social approaches.

Precisely for this strong interest in social network field, corresponding literature is very wide and covers many different disciplinary sectors.

In particular, many works have been focused on networks of researchers that represent a typical and interesting example of affiliation network in which a link between two researchers indicates the existence of their scientific collaboration. The distinctive feature of researcher collaboration, respect to collaboration in general, is referred to a model in which the single researcher has the freedom to decide if, with whom, and how to collaborate. So, a research network is characterized by spontaneous relationships that evolve over time depending on dynamic characteristics of researchers and of network.

In this context, the first step of developed studies has been the definition of research collaboration process. Katz and Martin (1997) say that the collaboration in research as a *good thing*, but they do not explain what it is meant.

Regarding the word “research”, Must (2000) suggested some important its features: “science is a collective, creative effort that cannot develop in *isolation*…. The fundamentals for an ample field of scientific research are *openness*, an opportunity to consult, *belief* on the research results of predecessors”. Thus, the *scientific activity implies the collaboration*.

Wikipedia suggests a definition of collaboration according to which it is *a process between two or more persons working together to achieve their goals*. Combining these two considerations, research collaboration occurs when at least two researchers work together, through the sharing their skills and knowledge, in order to achieve common goals (e.g. the production of a scientific paper).

The successive question moves on to define *how closely* researchers have to work together to say that they collaborate (Katz and Martin, 1997); and, then, the difference between research *collaborators* and *co-authors* of a paper. At the most basic level, research collaborators are scientists who work together *over time*, while co-authors can be simply scientists that have their names in a scientific article.

The great diffusion of on-line databases and the wide availability of services provided by digital libraries have favorite the construction of co-authorship networks that have been considered as the most common way to represent the research collaboration and one of the most tangible and well-documented form of social networks for existent databases.

In a co-authorship network, the actors are authors while ties represent papers that they have written together. Thus, in this interpretation the collaboration among researchers is *simply* given by the co-authored in a paper and all co-authors are considered as collaborators.

In the last years, the scientists have started to wonder in their empirical studies how collaborative research is *related* to co-authorship, emphasizing the problem of the *adequacy* to consider co-authorship network as units to identify research collaboration.

In fact, more often, a result of collaboration among researchers may be the writing of one or more papers, but often the collaboration among researchers does not lead to *joint output* like the publication of a paper. So, to be co-authors do not mean that the authors have collaborated. For instance, two researchers work very closely together but they decide to publish separately their findings because they operate in different disciplinary sectors. Thus, two scientists have collaborated intensively but at the end they have to publish in two different sectors.

Another case is when two researchers that have not worked together, decide to link their findings to write them in the same paper. In this case, the scientists have not collaborated but they produce a joint paper. So, in the first case has sense to speak of *research collaboration*, in the second case is reasonable speaking of *co-authors*, although often these two terms are considered as synonymous. In many cases the research collaboration takes place outside of formal relationship that is not recorded in a co-authorship, so the latter represents a *partial* indicator of collaboration (Katz and Martin, 1997).

On the collaboration/co-authorship problem, many studies provide that there is close linkage between collaboration and co-authorship, but the solution of problem is still far.

In a social network study a very important choice concerns which actors to include. In fact, the boundary of the set of actors sometimes may be difficult to determinate (Wasserman, 1999). When study network focus on small collectivities, such as a department, an office, a classroom, actors' set is clearly defined. But, in the cases in which the boundary is unknown, sampling techniques such as *snowball sampling* (Goodman, 1949, 1961) and *random nets* (first proposed by Rapoport 1949a, 1949b, and Fararo and Skvoretz 1984) can be adopted.

Besides, another important choice refers on which ties must be considered. Given a set of actors, ties among them change over time: relational networks are *continuously evolving* because links among actors can be created or destroyed or maintained over time.

Then networks change composition, as their actors may come and go from it, and their relationship.

Thus, the scientists are concerned about ties and actors change over time, and they investigate on ever changing nature and on dynamic structure of social networks.

Dynamic Network Analysis (DNA) draws from and extends concepts, models, and techniques from traditional network analysis area, SNA, taking into account that the structure of the networks is not immutable in time because of the fact that ties among the actors and actors themselves may change over time, finally the characteristics of a network change over time. Modeling network evolution as a dynamic process, the using of longitudinal network data is necessary to address the problem adequately. Longitudinal network data result from the observations of subjects that are measured repeatedly over time, for at least two distinct times. Typically longitudinal data are collected as *panel data.* The studied network is composed by the same set of actors and it is observed at least two points of time (*panel waves*).

In the study of changing networks, the distinction between *dynamics* and *evolution* of networks is essential (Doreian and Stokman, 1997). The two authors describe network dynamics as a more general statement of network evolution over time; they consider the network evolution as having a stricter meaning according to which it is possible to explain network changes via a *process*, that is the *mechanism* that induces network change.

The interval in which the network is observed is a fundamental dimension for catching the changes. Some examples of temporal dimension are the years of publication of papers in co-authorship networks (Newman, 2004), the year of release in the actor–actor collaboration network of movies (Barabasi and Albert, 1999), and so on. These examples of social networks are characterized by relations that change over time, and by temporal dimensions that must be exploited to analyze and understand networks.

Analyzing social networks over time has become increasingly popular. In fact, the literature on network dynamics has generated a large variety of mathematical models and a large range of applications of these models to real contexts.

To study empirically the mechanisms that determine the change in a network, statistical methods represent one of the most productive and recent approaches to study the dynamic nature of social networks.

One of statistical approaches is the actor-oriented model, proposed by Snijders. This model explains network evolution as a function of *endogenous* effects (for example, two individuals socially connected, over time tend to become friends) and individual characteristics, and *exogenous* effects of actors (for example, the formation of relations is based on the similarity between individuals). SIENA software (Snijders, 2005; Snijders et al., 2008) has been designed to model evolution of networks through time as a function of network structure, and individual attributes according to actor-oriented model.

A few studies, produced in very recent years, have treated the application of actor-oriented model to real contexts; these few applications regard mainly friendship networks (van de Bunt et al., 1999), and very few applications concern inter-organizational network (van de Bunt et al., 2007) and scientific communities (Kronegger et al., 2012).

The case study considered in this doctoral thesis concerns research networks emerging by scientific collaboration among researchers that decide to share their skills, knowledge, and interests. In particular, the researchers of DIEG and external people that have collaborated with them form the adopted unit of analysis for which the evolution from 2001 to 2011 has been considered. For its study, static and dynamic methodologies have been adopted. In fact, given dynamic nature of research collaboration, a static study is not able to give information about the network evolution over time.

The thesis is structured by 8 chapters.

The first chapter gives a short introduction to SNA. It is composed by: a brief introduction on its historical development and main models proposed over time; a description of principal elements required for operating the analysis of a social network; some possible contexts in which SNA can be applied.

The second chapter is dedicated to Dynamic Network Analysis that overcomes some limits of SNA (fundamentally its staticity). The models proposed in order to make a dynamic analysis are described, with particular reference to actor-oriented model, which appears to be most suitable for the type of application. The model allows interpreting the changing networks over time as the result of relational choices of actors that decide to create, eliminate or no change their links in the network. Relational choices are defined determining the probabilities with which they can occur and specifying *when* and *what* changes occur.

In third chapter, tools available for static and dynamic analyses are described. In particular, the focus is on three software that have been used to analyze the unit research: UCINET (Borgatti et al., 2009), perhaps the best known and most frequently used for the static analysis of social network, Pajek (Batagelj and Mrvar, 2007) another network analysis and visualization program, and SIENA, (Snijders, 2001, 2004) to perform statistical analysis and estimation of models for the evolution of social networks over time according to actor-oriented model of Snijders.

In the fourth chapter, the literature focused on research collaboration, from its definition until the representation of research collaboration as social network. In research field, interactions among scientists with aim to produce a paper has for long been the essence of scientific practice, in every discipline as well as within and across geographic areas. So, over time the number of papers with more co-authored has recorded a continuous increase. Accordingly, the idea to construct networks in which authors are actors and ties among them are represented by papers, *co-authorship* networks, or journals are actors and ties are citations, *citation* networks, has been very wide. This practice highlighted the problem of adequacy of this kind of network to measure and represent the research collaboration.

In the fifth chapter, methodology, unit of analysis, and hypotheses adopted are presented. Under two hypotheses, a double meaning has been assigned to research collaboration: 1) the scientific production is taken as an expression of the existence of a tie between the authors and, therefore, it is seen as research collaboration between them; 2) it is assumed that isolated papers do not attest a research collaboration between their authors. So two kinds of network have been identified: the *co-authorship network* that includes a set of authors and ties among them represented by all coauthored papers; the *collaborative network* that includes a set of authors and ties only represented by coauthored papers only if the interval between two successive papers is less than five years.

The actors considered in case study are overall 76, including both members of the DIEG and who, belonging to other organizations, has collaborated with them. The experiment has been conducted over 11-years period (from 2001 to 2011), characterized by the entry and exit from the study unit of some actors. To obtain the configuration of the department (people belonging to the DIEG and his/her career level) in each observation time, the official website of the MIUR (Ministry of Education, University and Research) has been used. DIEG's researchers have been described by different attributes, some constant and some varying over time: disciplinary sector, and institutional affiliation (internal or external to department) have been considered constant; professional rank, and scientific production, evaluated by H-index, have been obviously considered changing in the period of observation. Information on researchers and their papers (for each paper: year of publication, title, names of co-authors; for the researcher: the H-index) have been obtained by database *Scopus*, official source for Italian VTR (National Triennial Evaluation of Research).

In the sixth chapter, the co-authorship networks have been analyzed by static and dynamic analyses, and the obtained results have been illustrated. The results of static analysis show that over time the size of co-authorship networks, in terms of number of authors, increases and connections among authors grows too. Longitudinal analysis suggests that the tendency to collaborate in writing a paper is characterized by three different types of behavior: (i) the authors tend mainly to form ties with other authors with whom they share other ties; (ii) the decisions to create ties are a little influenced by H-index so some authors tend to link with others that have the similar H-index; (iii) there are not authors that establish relationships with members exclusively within the same institution and this suggests that the authors have lower probabilities of establishing new ties with others of their same institution. In order to detect the dynamic within department, the co-authorship networks of DIEG have been realized without to consider the external authors. Due to the removal of external authors, the cohesion networks over time shows that there is an increase of aggregation among the author of DIEG but values related them are low enough respect those obtained in whole co-authorship networks. So, only a little part of the components of DIEG is linked with other DIEG components. The results obtained show that the authors tend to collaborate with others that belong to the same disciplinary sector. There are no particular differences in behavior with respect to the role of scientific collaboration and level of carrier.

In the seventh chapter, the same logic followed for co-authorship networks has been kept for collaborative networks, and the results obtained have been shown and compared with those found for co-authorship networks. Over time the cohesion degree of collaborative networks is low, and this is caused by the fact that collaborative ties can be created, eliminated or maintained. Also dynamic analysis presents different results. In collaborative networks the effect that weighs mainly is H-index similarity, so the researchers tend to form ties with others that are characterized by similar H-index values.

In the eighth chapter, findings for the two kinds of network have been compared and discussed. The comparison suggests that there are many differences between co-authorship and collaborative networks, despite the weak hypothesis assumed on collaborative network. In particular, the results of static analysis show that co-authorship networks are characterized by greater cohesion, and this suggests that ties among researchers are representative of occasional collaboration. Besides, longitudinal results suggest that, for co-authorships, forming ties was more likely among researchers that share co-authors, while for collaborative networks, it was more likely among researchers with H-index similar. Finally, the elimination of external people shows that researchers of department prefer to collaborate with externals and that they tend to collaborate with others operating in the same their disciplinary sector.

The principal aspects that make interesting this study are: (i) the originality of unit of analysis chosen, (ii) the utilization of actor-oriented model for an undirected network, (iii) the attempt to clarify relation between co-authorship and collaborative networks, (iv) the individuation of mechanisms that drive the network evolution, (v) the individuation of future lines of research.

1. Social Network Analysis

We live in a world that is paradoxically small and wide: each of us is embedded in a local communities, yet at the same time more and more of us hold contacts that span the globe,………
… each one of us has our own social networks.

Prell Christina, 2012

1.1 Social Network Analysis (SNA) methodology

Every kind of aggregation can be represented in terms of entities that compose it and relations between these entities. This type of representation is called *social network*.

In a social network, the entities, generally called actors, can be individuals, organizations, a company, a country, a blog and so on, included in a social context. The actors are linked together by means of different types of relationships that can represent interactions, collaborations, or influences. The concept of social network can be applied to all phenomena characterized by entities connected among them. Just to think that every individual has his/her personal community (Figure 1).

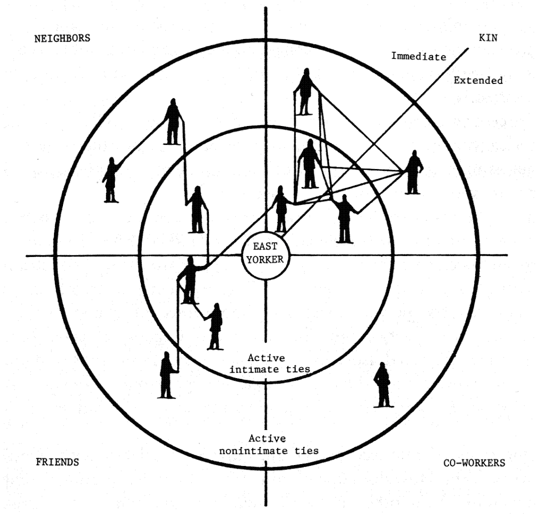


Figure 1 **–** *Typical personal community (Scott et al., 2011).*

Social Network Analysis (in the following indicated as SNA) is an interdisciplinary methodology that seems for a long time to have resisted the integration of empirical research with other branches, such as anthropology, statistic, mathematics, physics and more. SNA is developed with a not linear process due to several persons and multiple academic groups that played a role in its shaping.

"Today, many see SNA as its own paradigms" (Leinhardt, 1977). "This means that SNA is perceived as an unique approach to understanding (primarily) the social world" (Prell, 2011).

SNA starts to identify the properties of network and carries on to understand what determines these properties and what consequences they have for the actors or the network as a whole.

The structure of a network is important in determining what happens inside it as the properties of its individual actors (Borgatti, 2011). In a team, for example, the success of a project depends both on the work done by each component, but also on how all the components work together.

So, SNA tries to understand how actors are related to each other, through the *mapping* and *measuring* of relationships among these actors.

For instance, SNA allows knowing:

- Who knows who;
- Who has a high number of contacts in a group or organizations;
- What are the sub groups that compose a large community;
- How the management are linked in a company;
- So on.

The most general characteristics of social data are that they are rooted in cultural values and symbols. In fact, there are different kinds of data, but only some of them are the most appropriate to SNA.

The main methodological support to SNA is that the network can be modeled by a graph, or digraph; a graph is a pair of sets G = {P, E}, where P is a set of nodes and E is a set of edges that connect two elements of P. 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.

So, a social network is a set of actors, defined formally as nodes, and collection of relations between them, represented by arcs, that specify how these actors are related to one another.

Each tie or relation may be directed (i.e. it originates in source actor and reaches a target actor, e.g. the relation "to be a parent of"), or it may be undirected (i.e. it is a tie that represents co-occurrence, co-presence, or a bonded-tie between the pair of actors, e.g. the relation "to be a sibling of"). Directed ties are represented with arrows, and bonded-tie relations are represented with line segments. Directed ties may be reciprocated (the node *i* links to *j* and *j* links to *i*); such ties can be represented with a double-headed arrow.

In order to represent link direction, it is possible to use a directed graph. In Figure 2, the actor A has a link with actor B and C, but B and C do not reciprocity link.

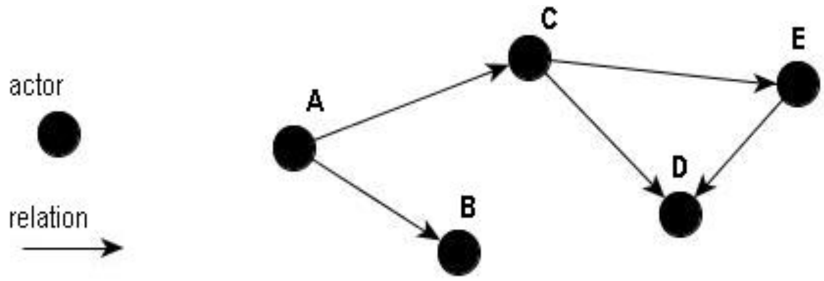


Figure 2 **–** *Social network's representation.*

Ties may have different strengths or weights. These strengths may be binary (representing presence or absence of a tie), signed (representing a negative tie, a positive tie, or no tie); ordinal (representing whether tie is strongest, next strongest, etc.); or numerically valued (measured on an interval or ratio scale).

On a graph, utilizing models and algorithms characteristic of Graph Theory[1], it is possible to lead several analyses to identify some important characteristics of network. In fact, the graph theory assumes a crucial role in order to quantify and measure of some properties of the network, and to represent networks.

The spread of personal computer use from the late 1980s has encouraged much wider use of SNA methods because it made easier to manage large data sets and to visualize social network data in a wide variety of ways. Examples of social networks are online social platforms, like Twitter, Facebook, and more, in which every user can share pictures, music files, create a personal profile page, chat with other users, and comment on other's shared resources.

It is possible to create different types of network, built on the basis of the kind of relationship that users have with other users, like friendship, family, or simple connect. Relationships can be a direction, so they can not be

[1] In the appendix A, a brief description of the origins of graph theory is presented.

symmetric, because of an user A can declare a relationship with other user B (for instance A send an email to B), but B may not declare a relationship with A; in this case, the link has direction from A to B and graph corresponding is labeled (Figure 3).

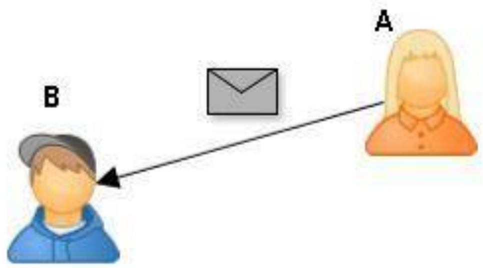


Figure 3 *– Example of Online Social Network.*

In Figure 4, there is an example of a business network in which the companies have selling relationships between them. In this case, it is shown a directed graph in which each arrow points in the direction of the sale.

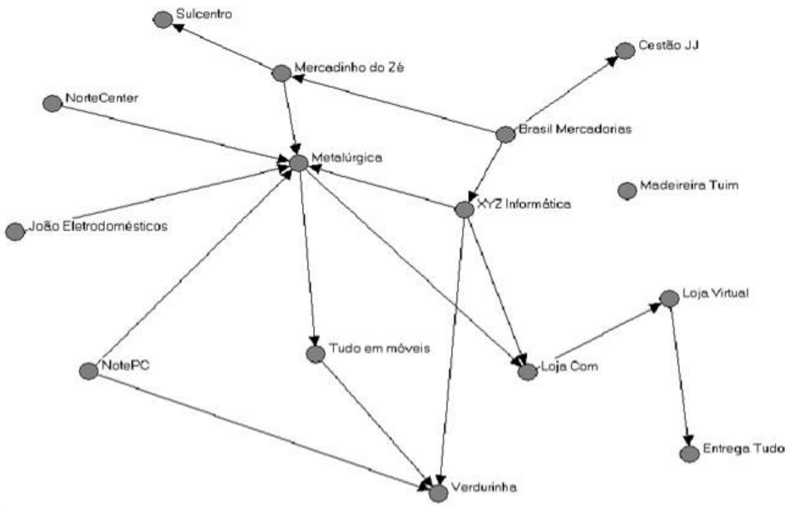


Figure 4 *– Directed graph with sales connections between companies.*

In the end, from the structure of social networks and its key features, it is possible to have paramount information for understanding the spread of knowledge, cultural traits, disease, and many others that can be associated with individuals living in groups or societies.

1.2 A brief history of SNA

Interest in SNA has grown very quickly in the last decade, starting by 1960s.

There is a long history behind this growth of interest and this period has been characterized by publication of many papers.

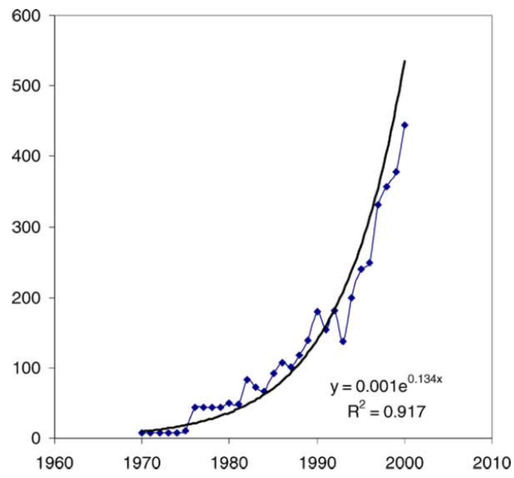


Figure 5 *– Exponential growth of publications indexed by Sociological Abstracts containing "social network" in the abstract or title (Borgatti and Foster, 2003).*

"A number of diverse strands have shaped the development of present-day social network analysis", (Scott, 1991).

Although the scenery is so complex, it was possible to clearly draw a lineage of the most important strands (Figure 5).

So beginning in the 1930s, there are three main and parallel research lines, lead respectively by:

- Sociometric Analysts, who many technical advances realized by using the methods of graph theory;
- Harvard researchers, who studied the patterns of interpersonal relations and the formation of cliques;
- The anthropologist part of the school of Manchester, that, by combining the two previous strands, mainly analyzed the structure of relations within the "community" of tribal societies and village.

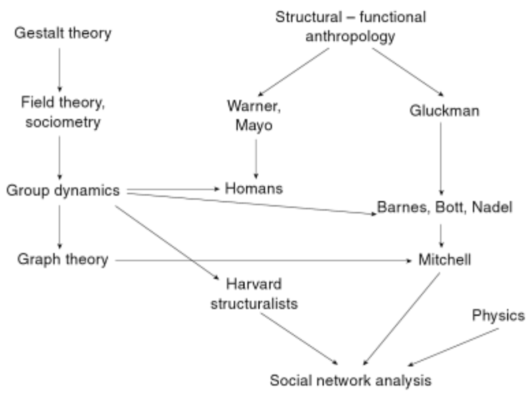


Figure 6 ***–*** *The lineage of SNA (Scott, 1991).*

In the 1930s, the Gestalt theory developed. Moreno, who was likely the most notable, and his collaborator introduced the sociogram, the first tool for the structural analysis of networks, which represents the embryo of network concept.

The sociogram is a diagram in the tradition of spatial geometry and some traditional symbols were the following:

- The triangle was generally referred to as a male member in a group;
- The circle represented a female member in a group;
- The circles and triangles were connected by straight lines called vectors; these vectors represented the type and direction of each person's choice.

In 1932, there was an epidemic of *runways* at Hudson school for girls of upstate of New York: in two weeks, 14 girls had run away from the school (a rate 30 times higher than the norm).

Moreno mapped the social network of runaways at Hudson.

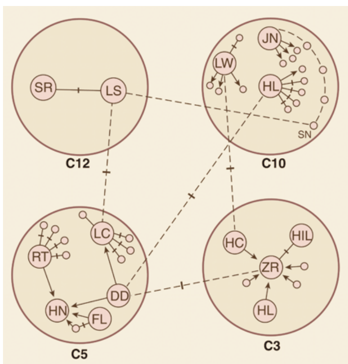


Figure 7 ***–*** *Runaways network (Borgatti et al.).*

In Figure 7, the four largest circles represented cottages in which the girls lived, while the circles within them represented an individual girl. Initials on each small circle identified the runaways.

The all un-directed lines between girls represented the feelings of mutual attraction, while all directed lines were one-way feelings of attractions.

According to Moreno, the links in this social network provided channels for the flow of social influence and ideas among the girls.

In a way that even the girls themselves may not have been conscious of, it was their location in the social network that determined whether and when they ran away.

Before Moreno, others theorists had talked about the plots of connection, the social fabric and networks of relationships, but no one had explained these metaphors with the use of diagrams that meet formal criteria of construction and interpretation.

On the Gestalt theory' s work, in the 1950s, Cartwright and Harary connected the sociogram to mathematical formulas to create graph theory.

This attempt to apply mathematics to the structure of relations group was not a new idea.

Cartwright and Harary had outlined the fundamental idea to represent the groups as a collection of points connected by lines. The resulting sociogram or "graph" represented the interpersonal relationship network among members of a group and they argued that the graph could be analyzed using the mathematical concepts from graph theory.

In Cartwright' and Harary's work, in a graph the points represented the individuals and lines showed the relationships between the one and the other; the lines could be accompanied by the signs + or -, to indicate positive or negative relationships, and could be equipped with arrows, to indicate the relationship orientation.

The construction of graphs with signs and oriented allowed Cartwright and Harary to analyze the structure of the groups from the point of view of each of its members simultaneously, and not only from the point of view of a particular individual focal. It constituted, therefore, an important step forward in the direction strictly sociological.

Between 1927 and 1932 Elton Mayo and other researchers conducted the famous Hawthorne studies, in which they used sociograms to map informal social structures and group behavior in a bank’s wiring room (Fredericks and Durland, 2005).

In the early forties, Elton Mayo started several researches about working conditions and productivity of employed of Chicago’s central plant Hawthorne.

From research result, Mayo came to the conclusion that the increase in productivity due to the fact that the workers reported they felt part of a group, or rather selected for research.

The team Hawthorne’s studies represent the first important research in which the sociograms were used to describe the observed relationships in actual real life situations.

At same time, Warner and Lunt studied a small urban community in New England, that they called Yankee City. They argued that the social configuration consisted of various subgroups such as family, church, classes and associations. To these they added a particular subgroup, called cliques, that indicates an informal association of people among whom there is a feeling of intimacy and of the group and in which there are certain rules of conduct established by the group itself.

Despite its many limitations, the work of "Yankee City" remains attractive for its pioneering attempt to use formal methods of structural analysis.

The importance of network analysis applied to social networks suffered a further turn thanks to the work of some researchers from Manchester University Department of Social Anthropology, in particular Mitchell, Bott, Barnes, who first introduced the term Social network.

Manchester researchers pointed their attention at the effective configuration of relationships deriving from power and conflict between individuals, instead of set up norms and institutions of a society (Scott, 1993).

During the fifties, Parson's theory was the strong dominance on cultural approaches in anthropology and sociology. This helped to direct the work of the Manchester school along a sharply critical tradition.

In opposition to the idea of the sociological classics, who insisted that the actions should be understood in view of their location in a structure of social relations, Parson believed that actions should be explained as expressions of value orientations internalized.

Anthropologists of Manchester managed to combine the techniques of network analysis with sociological concepts nouns, starting to see the structures as *networks* of relationships.

The theoretical conceptions inherited from the past were suitable for understand of simple society, based on kinship, but they were unable to handle these phenomena, and it was a result of the recognition of this inadequacy that they began to look for a systematization of metaphorical notions as *tissue* and *network* of social relations to which actors such as Radcliffe-Brown had made reference. These researchers took them the concept of social network simply as a metaphorical sense, but in the early fifties Barnes began to apply it in a more rigorous and analytical way. His approach greatly influenced the Bott's work, Canadian psychologist who studied anthropology in Chicago with Lloyd Warner, and the two began to explore more closely the work done in the tradition of sociometric.

Mitchell laid the foundations for a systematic framework for the analysis of social networks.

Barnes, having joined the department in Manchester, decided to conduct field research on the environment quite unusual for a fishing village on the southwestern Norway. Despite it was a small village communities, it was not an isolated place and structured only by the kinship relations of the inhabitants. Barnes was strongly attracted to the role played by kinship, friendship and neighborhood in producing community integration. These primary relationships were not directly linked to places territorially defined formal structures or economic or political sphere but they formed distinct and relatively integrated informal and interpersonal relationships. Barnes argued that the totality of social life could be seen as a set of points, some of which are joined by lines in order to form a total network of relationships. The informal sphere of interpersonal relations should be seen as a part of a network part of this total network (Barnes, 1954).

Bott started a research on the lives of a number of British families. The Bott was primarily interested in their family relationships, and employed the concept of network as artifice analysis to investigate the various forms taken by these kinship relations.

Barnes and Bott opened the way for further developments that would consolidate their progress with other contributions by American researchers. Decisive voice in legitimizing this line of theoretical development was Nadel.

Nadel, Austrian psychologist, was passed to anthropological studies in the early thirties. The starting point of Nadel was a definition of structure as an articulation or organization of elements in the form a whole. If you separate the forms of relationship from their content it becomes possible to describe and analyze the general features of the structures with the comparative method. In order to build formal models, Nadel made use of a mathematical approach to the structure.

According to Nadel, the social structure is a total system, a network or pattern of relationships that the analyst abstracts from concrete observable actions of individuals. For network, he means the intersection of relationships for which the interactions implied in a determine those occurring in the other. And in particular, Nadel argued the role should be seen as the central concept in sociological theory. Social structures are structures of roles, and the networks of interdependent activities define roles, together with the complex of roles. Nadel believed that the analysis of the roles should be applied algebraic methods and dies, but he provided little guidance on how this should be done.

Mitchell’s analysis is important because of the reflection about some social indexes (as density, which he sees as the degree of completeness of the network, the extent to which all possible relationships are indeed present) and their meaning in the description of a network too.

With the success of the writings of Mitchell, Barnes and Bott, network analysis was directed towards the study of informal and interpersonal relationships, and intended solely for egocentric networks. This resulted in the lack of development in England of search paths turned to the global properties of social networks.

In this direction, there was the crucial turning point at Harvard.

After a decade from the initial investigations of Homans, a flood of papers began.

The unit element of these researches was to use algebraic techniques to formalize structural relationships and especially to use the network analysis as a method of study.

In the 1973s, Granovetter's work was very important. The purpose of his survey was to know and understand the ways in which people get information on job opportunities due to their social contacts. In particular, he wanted to find out the type of ties underlying the exchange of information, whether these ties were strong or weak and how these were preserved in time.

From the results of his survey, Granovetter formulated his thesis on the *strength of weak ties*: acquaintances are more likely to provide information of the close friends of labor.

From the Harvard group, an international group that acted as a center for the development of the analysis of social networks was created. This group was called INSNA (an acronym for International Network Society of Social Network Analysts).

Among the different kinds of analysis, an original theory, known as *small world phenomenon*.

The small-world experiment comprised several experiments conducted by Milgram (1967) and other researchers examining the average path length for social networks of people in the United States. The research was groundbreaking in that it suggested that human society is a small-world-type network characterized by short path-lengths: two persons meet and find they have acquaintances in common.

In brief, the Milgram experiment was to study the route of letters mailed from Nebraska to direct acquaintance with Pittsburg as final destination. The average number of steps was five with six involved people, and this phenomenon is called *Six Degree of Separation*.

1.3 Network matrices

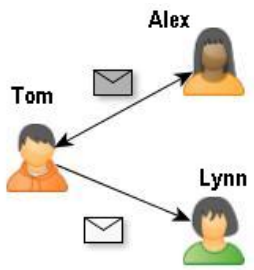
The two most common ways to representing of social network are by using graphs and by using *matrices*.

In the second case, it is possible mapping a social network by different types of matrices (adjacency, affiliation, incidence). This alternative to graphical representation of network is due to several reasons:

- Data matrices are indispensable when network size exceeds the possibilities of visual illustration;
- It is possible to carry out different quantitative analyses so to start picking out the structural features and overriding patterns in the data.

So, a structure of network can be represented by an *adjacency matrix nxn* in which *n* indicates the number of actors. Each row represents an actor in a given sequence, from the first to the last, and this sequence must be the same for a columns. This matrix can be symmetric or asymmetric. An asymmetric matrix records the direction of tie so it represents a directed network, in which the senders are in rows, while the receivers in columns. The diagonal of this matrix represents the sender’s tie to itself; in most situations, a diagonal is ignored because it is considered uninteresting, as actor’s relationships with themselves do not give important information, but only relations that actors have with others. Cells (each cell is given by intersection of a row with a column) indicate the presence or absence of ties. When the values in all cells are 1 or 0, matrix is a binary adjacency matrix so, the generic cell a_{ij} is 1 if there is tie between actor j and otherwise it is 0, if there is not. So the variables A_{ij} represent how actor i is tied to actor j.

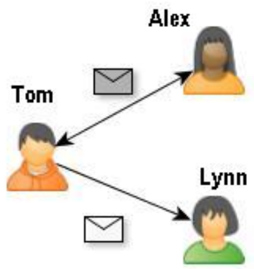
In Figure 8, social network of message exchange is depicted; network is composed by three actors, so the adjacency matrix is *3x3*, and records on who sends a message to whom are represented through use 1 and 0; Tom sends to a message to Lynn, and Alex sends a message to Tom who responds to her message.



	Lynn	*Tom*	*Alex*
Lynn	*0*	*0*	*0*
Tom	*1*	*0*	*1*
Alex	*0*	*1*	*0*

Figure 8 *– Matrix of a message network.*

In addition, a matrix can also convey the intensity of tie by the values contained within the cell. In a previous example, if Tom sends five messages to Lynn, corresponding cell contains value 5 that could indicate that there is a stronger or more intense interest from Tom in Lynn than Alex.



	Lynn	*Tom*	*Alex*
Lynn	*0*	*0*	*0*
Tom	*5*	*0*	*1*
Alex	*0*	*1*	*0*

Figure 9 – *Valued Matrix of a message network.*

1.4 Data for SNA

The kind of data that appears to be more appropriate to SNA can be referred to attribute data and relational data.

Relational data represent the essential requirement of a network and constructing and analyzing of a social network is allowed thanks to this kind of data. Relational data are the contacts, ties and connections, the group attachments and meetings, which relate one actor to another and so cannot be reduced to the properties of the individual actors themselves (Scott, 1991).

Attribute data describe actor's attitudes, opinions, and behaviors; these are regarded as the properties, qualities, and characteristics that belong to them as individuals or groups (Scott, 1991). Attribute data can be used alongside relational data when constructing a social network to provide insight into factors contributing to network structure.

Relational and attribute data are not only types of data used in social science.

A third type comprises ideational data that describe meanings, motives, definitions and typifications themselves (Scott, 1991).

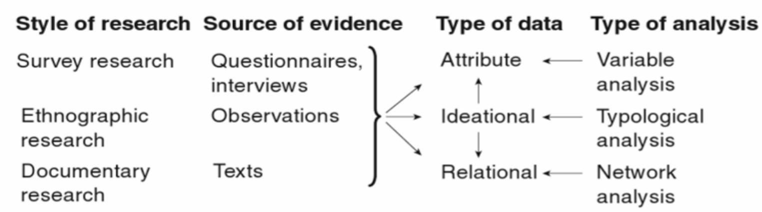


***Figure 10** – Types of data and relative analyses (Scott, 1991).*

Information about these kinds of ties are commonly collected through interviews or surveys, often administered online. For instance, a typical survey might list of t all people working in a team and then to ask each individual to whom among their colleagues they go to when they need client-related information. The result of all responses might show those that have much information, which occupy key roles.

Collecting data about ties is not limited to surveys. This data can also be inferred from a number of existing sources, such as email exchanges (i.e. who writes to whom?), direct observations of group interaction, work hours (i.e., who works on projects with whom?), professional citations (i.e., who publishes with whom?), charitable donations (i.e., who is giving money to whom?), and so on.

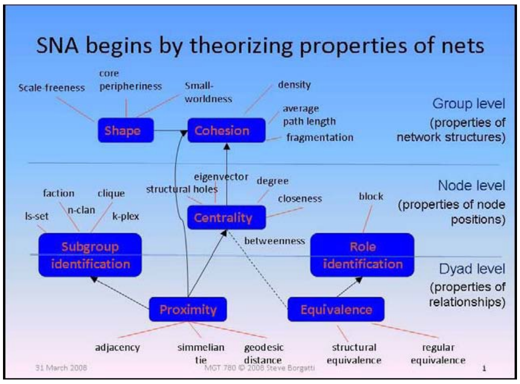
1.5 Main relational metrics

In support to SNA, a set of concepts to describe these network structures and positions of actors within network have been developed.

Structural properties are characterized at three levels of analysis:

- Group;
- Node (actor);
- Dyads[2].

[2] Sub graphs of size 2 consisting of a pair of actors and all ties between them (Wasserman and Faust, 1999).



Level of analysis		Network properties
Individual actor level		• Degree, Indegree , and Outdegree • Betweenness • Closeness • Centrality and prestige • Roles (isolates, liaisons, bridges) • Structural holes (Burt, 1992)
Local level	Dyads level	• Distance and Geodesics • Structural equivalence
	Triad level	• Transitivity •Cyclicity
	Subgroup level	• Component and cliques • Positions
Global level		• Connectedness and diameter • Density • Prestige • Network centralization

Figure 11 *– Levels of analysis (Borgatti et al., 2009).*

At the node level of analysis, the most important concept is centrality that allows identifying nodal properties as a function of node's position, relating to the structural importance or prominence of a node in the network (Borgatti et al., 2009). At this level, many metrics describe the network's cohesion (such as density, average path length, and fragmentation).

At local level, a key aspect consists into identify groups of actors within network (such as *cliques*) on the basis of certain characteristics.

So, structural properties of dyadic relationships are defined both on proximity of nodes (such as adjacency and geodesic distance) and equivalency of nodes (structural and regular equivalence).

According to these three levels of analysis, different metrics can be calculated:

- Cohesion index;
- Identification of the cliques;
- Centrality index.

1.5.1 Cohesion indices

In order to calculate the cohesion network, different metrics can be used:

- Inclusiveness index;
- Density;
- Distance;
- Connectivity

- Clustering coefficient and transitivity.

Inclusiveness refers to the number of actors that are included within various connected parts of the network. In other words, the inclusiveness corresponds to the total number of actors minus the number of isolated actors. The most useful measure of inclusiveness for comparing various networks is the number of connected actors expressed as a proportion of the total number of actors.

A measure much linked to inclusiveness is density that calculates the level of aggregation of actors in a network and then describes the general level of cohesion of the network. In particular, density is the number of ties expressed as a proportion of the number of all possible ties, and it is calculated as the number of actual ties in the network divided by the number of all ties that are present. According to the kind of the network by studying, density could give details on the speed at which information diffuses among the actors, so the high density corresponds to a network characterized by many ties between its components.

For undirected graphs, it is formally defined as:

$$\Delta = 2L/n(n-1)$$

in which L represents the number of present edges, while n is number of nodes.

In Figure 12, three kinds of graph are showed, each of them is characterized by a different density indices.

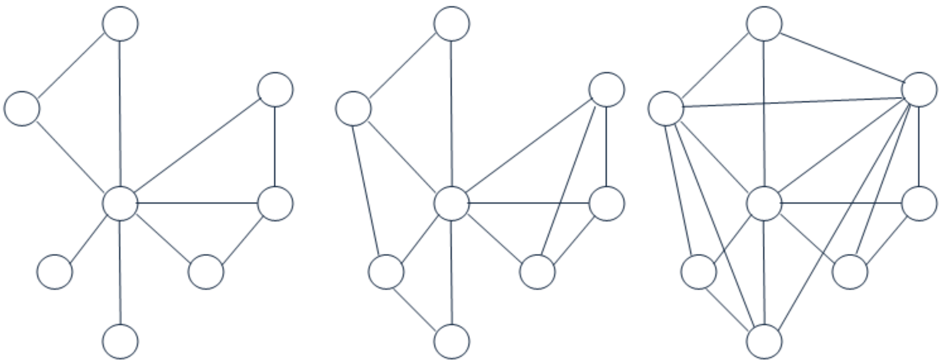
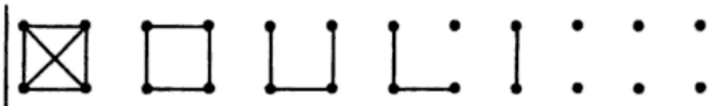


Figure 12 *– Examples of networks with different density (Vargiu A., 2001).*

The first graph is characterized by a density index equal to 0.36, because there are 10 ties and 28 possible ties. In the same way density index of the second graph is calculated and it is equal to 0.46 while the third graph has a density equal to 0.61.

Thus, the more inclusive is a network, the denser will be it. Figure 13 shows *how* density varies with the inclusiveness.



No. of connected points	4	4	4	3	2	0
Inclusiveness	1.0	1.0	1.0	0.7	0.5	0
No. of lines	6	4	3	2	1	0
Density	1.0	0.7	0.5	0.3	0.1	0

Figure 13 – *Examples of calculation of density and inclusiveness (Scott, 1991).*

For weighted undirected graphs, the formula of density takes into account the weights of arcs and becomes:

$$\Delta = 2v_k/n(n-1)$$

in which v_k represents the value of weighted edges. Respect to meaning of density for directed graph, in this case the value could be upper to 1 (when all nodes are linked among them), as it refers to average of weighted edges.

To capture how actors are embedded in a network, one approach is to check how far, in terms of social distance, an actor is from others. The distance between two actors is the minimum number of edges that takes to go from one to another. This is also known as the *geodesic distance*. The actors that are closer to more others may be able to exert more power than actors that are more distant. This index defines how each actor is implicated in a relationship so if the sum of distances is small, then it means that involvement is major.

In Figure 14, two networks in which actors are connected with a different number of relationships are illustrated.

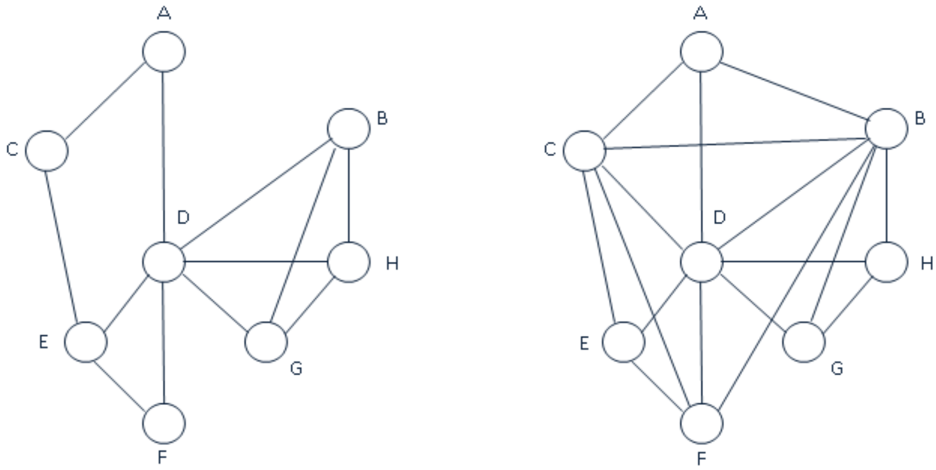


Figure 14 *– Examples of two networks with different distance (Chiesi, 2007).*

First graph is less connected than second one because smaller distances between actors characterize it.

The diameter is another metric to calculate cohesion of a network. It is the largest geodesic distance in the network, and it gives the number of steps that are sufficient to go from any node to any other node so it is sometimes used as a measure of connectivity of a network. If the diameter is high then the informative flow is more difficult because of the steps of information increase.

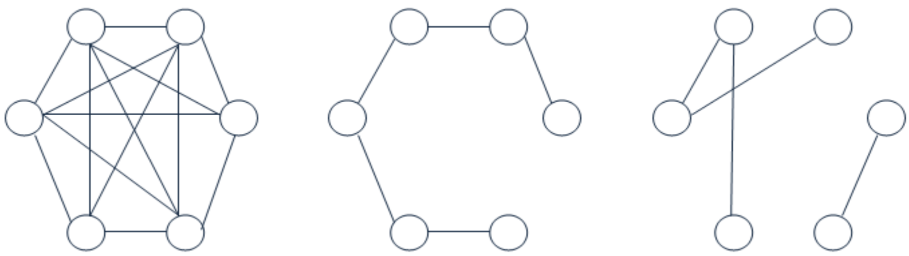


Figure 15 *– Examples of networks with different diameters.*

The *small world* (Wattz and Strogatz, 1998) concept represents an attempt to capture *clustering* idea (friends of friends to be friends). Wattz and Strogatz denoted social ties (edges) among individuals (nodes) as a circular network (Figure 16).

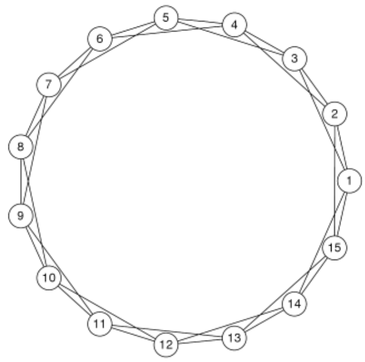


Figure 16 *– Circular network with high clustering (Scott and Carrington, 2011).*

Two authors defined an average clustering coefficient that measures the degree to which each node and its immediate neighbors are directly connected to one another. They defined the clustering coefficient as: "Suppose that a vertex v has k_v neighbors; then at most $k_v(k_v-1)/2$ edges can exist between them. Let C_v denote the fraction of these allowable edges that actually exist. The clustering coefficient for whole network is given as the average of the local clustering coefficients of all nodes of network".

In undirected networks, the clustering coefficient C_n of a node is defined as $C_n=2e_n/(k_n(k_n-1))$, where k_n is the number of neighbors of n and e_n is the number of connected pairs between all neighbors of n. In directed graph, the definition is slightly different $C_n=e_n/(k_n(k_n-1))$. In both case, the clustering coefficient is a ratio N/M, where N is the number of edges between the neighbors of n, and M is the maximum number of edges that could possibly exist between the neighbors of n. the clustering coefficient of a node is always a number between 0 and 1.

The clustering coefficient of whole network is the average of the clustering coefficients for all nodes. Thus, a clustering coefficient is a measure of the degree to which nodes tend to cluster together.

Transitivity was introduced in study of Newman, Wattz, and Strogatz (2002), where it was claimed to be equal to the clustering coefficient. For Newman (2001), transitivity describes symmetry of interaction among trios of actors. It refers to the extent to which the existence of ties between actors A and B and between actors B and C implies a tie between A and C. So transitivity is the fractions of connected triples.

1.5.2 Local structures: cliques

In SNA an important kind of analysis consists in decomposition of a network in sub groups with high density, called *cliques*.

The cliques are composed by set of actors connected in a very *close* way. In a clique every member knows everybody of group. The existence of cliques can evaluate the strength and effectiveness of the entire network. The idea according to relations among some actors form sub graphs denser than whole network and the membership to these cohesive groups conditions significantly the strategies and preferences of the actors, is one of the most important results of Harvard school that for the first time introduced the concept of a clique. In Figure 17, an example of a clique is illustrated, in which all nodes are connected of all other nodes.

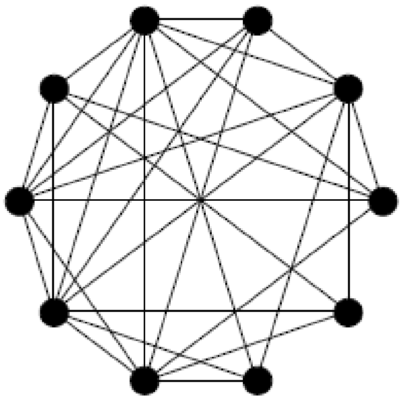


Figure 17 – *Example of a clique (Izquierdo et al., 2006).*

1.5.3 Centrality indices

Over time, the centrality concept has been the subject of numerous disputes. In fact, the idea according to each node has a degree of centrality was introduced by Moreno and in following decades, this idea has been taken by others.

The centrality of a node is a measure of its structural importance; for instance, how important a business individual is within a company, how important information is within a project, or how important a country is within an alliance. There are three approaches to calculate the centrality of an actor based on:

- Degree, based on the idea that actor who has more ties, is more important;
- Betweenness, based on idea that being in between many other actors what makes a central actor;
- Closeness, based on idea that the actors who are more reachable by others at shorter path lengths, are in favoured positions.

Centrality measure	**Interpretation in social networks**
Degree	How many people can this person reach directly?
Betweenness	How likely is this person to be the most direct route between two people in the network?
Closeness	How fast can this person reach everyone in the network?

Table 1 *– Interpretation of measures in social network (Cheliotis).*

According to these approaches, a central actor has a stronger influence on other members of network so centrality measures can be interpreted as measure of power. Obviously the interpretation of these measures depends on the kind of network.

Centrality measure	**Other possible interpretations…**
Degree	In network of music collaborations: how many people has this person collaborated with?
Betweenness	In network of spies: who is the spy though whom most of the confidential information is likely to flow?
Closeness	In network of sexual relations: how fast will an STD spread from this person to the rest of the network?

Table 2 *– Other interpretations of measures in social network (Cheliotis).*

In Figure 19, the network is composed by main two subgroups:

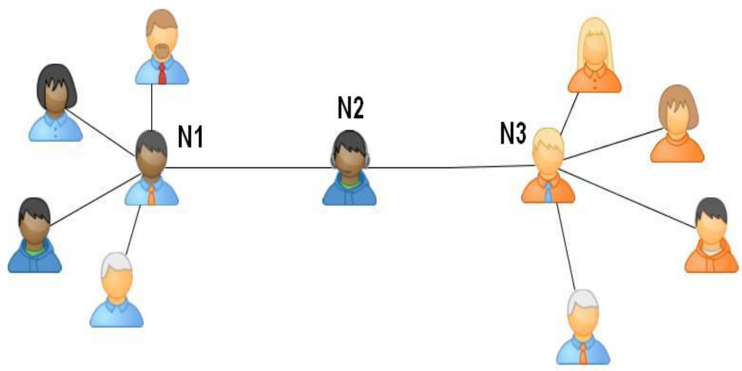


Figure 18 *– Example of calculating of three centrality indexes.*

According to degree centrality, actors N1 and N3 are the most centrals because they have many ties with other people. In order to analyze the network by betweenness centrality, so to consider the whole network, visible exam shows that actor N2 is characterized by highest centrality; in fact, N2 is the intermediary through which two subgroups communicate.

It is obvious that the statement depends on the kind of the ties in question.

If the network represents conversations between friends on a weekend or ties reciprocal affection, one might conclude that N1 and N3 are actually at the center of two circles of friends or acquaintances, while N2 participates in both, but not in the center of them, because what matters in this type of ties is the direct and immediate relationship between the actors. This kind of analysis is based by degree centrality.

If instead it were assumed that the graph illustrates a communication network, the centrality index calculated on only degree centrality would be misleading. In this case, in fact actor N2 occupies a strategic position and represents a point of separation. Without N2, the graph is divided into two separate components; so, for example, the network is composed by persons that exchange figurines, the actor N2 represents the subject, due to its location, better able to finish a collection of figurines, being able to exchange them, unlike the other subjects, on the basis of the information provided by both N1 that from N3. In this case, the nature of the links requires taking into account the overall distances that separate nodes, rather than the simple adjacency between them.

In this case, stated before by Bavelas (1950) and after by Beauchamp and Sabidussi, the centrality of an actor depends by the sum of distances of this actor from the others (the closeness centrality).

A further criteria to calculate centrality based on the position that each actor occupies in the network, initially proposed by Anthonisse (1971) and recovery by Freeman (1977), concerns the betweenness (to be in the middle). According to this criterion, the communication among actors depends on who are located along the paths that connect them. So the most central actor is a broker or agent.

Freeman hypothesizes that the most central actors "can influence the group by withholding or distorting information in transmission".

In Figure 19, if actors N1 and N3 do not know, but both them know actor N2 (who connects two otherwise unconnected N1 and N3), N2 is in a position to manage or "broker" information flow, so N2 is the most central because N1 and N3 can communicate due to N2.

1.6 Some practical applications of SNA

SNA is not linked to specific theory of how, for instance, society or individuals function, and this aspect makes SNA applicable to very different practical uses.

In these years, many practical applications have been proposed. In particular, recent studies focus on using SNA to different aspects of societies, communities, knowledge networks and competitive markets, such as Social Medias through Internet and Telecommunications environments.

Each organizational structure can be considered as a graph because of it composed by nodes and lines; so nodes represent companies, functions or people, while the connections between them are partnerships, informative flows, decisions or so on. So, it is possible to draw a map that shows single nodes and ties between them.

If this map is interpreted in correctly way, it will be able to give some information that can utilize in order to improve the processes.

In reference to a single company, fields in which it is possible to apply SNA are many; for example, in *Knowledge Management* SNA is very useful to analyze the contexts characterized by *collaborative* character.

In these contexts, knowledge transfer occurs by direct contact: a MIT[3] survey has showed that engineers, technicians, and researchers prefer *five* times more often to contact a their colleague than to search in an information system because they trust of information of their colleague.

The efficiency of these collaborative contexts, in word of *productivity* (innovation, creativity, customer satisfaction), is a variable that depends on strength of formal and informal relationships; informal relationships, often more important than these formal, are variables of difficult identification and interpretation, because they are determinate by *intangible* elements, such as collaboration, trust, friendship, and so on. SNA is a tool for identifying and analyzing the structure and strength of informal relations that exist within an organization. In fact, organizational chart (defined as the rational, conscious and institutionalized arrangement of the division of labour) does not consider informal relations and so, in the major cases, it draws realities that are not existent.

In order to understand how a group works in an organization chart, it is possible to find senior people that are empowered to make decisions or to see how work is divided up functionally. Over time, organization evolves and this means that organization chart is no longer an adequate guide to how the group really works.

[3] MIT (Massachusetts Institute of Technology) is a private research university located in Cambridge, Massachusetts, United States. MIT has five schools and one college, containing a total of 32 academic departments, with a strong emphasis on scientific, engineering, and technological education and research.

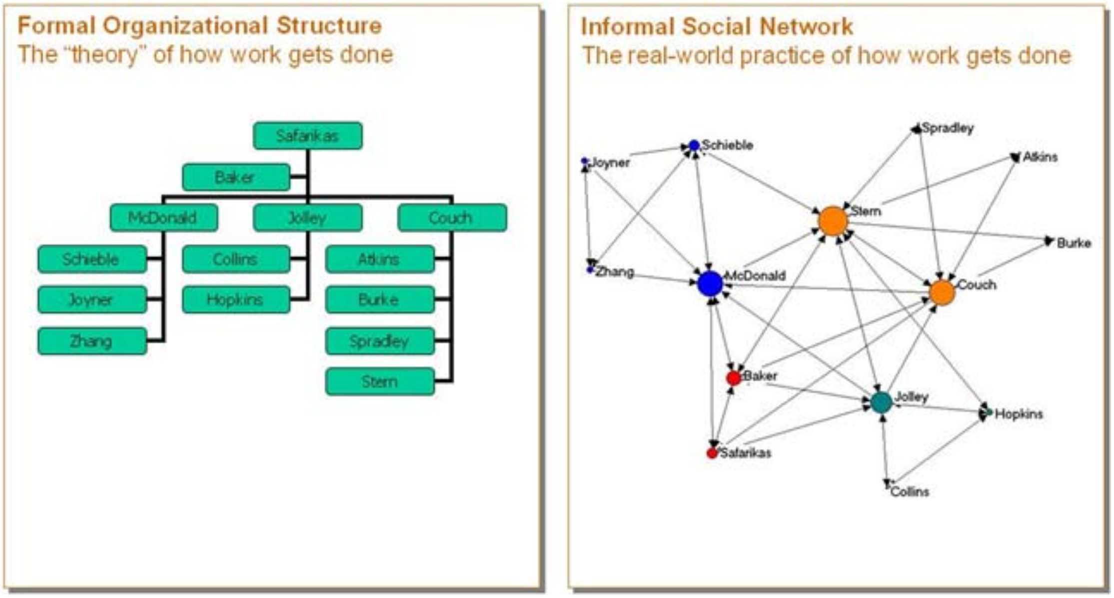


Figure 19 *– Formal vs informal structure (Garcia, 2013).*

In Figure 19, it is illustrated how people work together to solve problems and make decisions in the real world; each node represents each individual whereas the size of nodes is calculated on the individual's centrality in the network. Stern serves in a relatively unimportant position in the organizational chart. The map of informal relationships shows that Stern is a critical player in the organization. Not only was Stern linked to many people, making him very central to the group, but he is also the only link between clusters of people at the top and the rest of the group who were involved in other distinct but critical activities. He plays a central role in facilitating communication across all three teams.

An organization chart provides the theory of how work occurs, while informal social network provides the real-world practice.

So, some advantages resulting by SNA application are:

- Awareness of social network, to be aware of social networks, both internal and external to the organization, it is important for the knowledge management. The social networks represent the base of CoP (Community of Practice);
- Identification and implementation of knowledge maps; besides, the knowledge maps study helps to analyze the strengths and weaknesses of the network. With SNA, managers can have access

to useful data, which support them to improve and to justify strategic decisions on projects for the management of knowledge;

- Retain people, who possess crucial knowledge, this can be get through the increase of the share capital within organization. For example, it is more likely that people who have more connections are satisfied with their job, and then remain with a probability greater;
- Increased of innovation, productivity and the carrying capacity: it can be achieved by reducing the gaps in mutual understanding between the people, including the experiences and skills. Social networks are also important to know where to turn to for support in various situations. In this way, it will decrease the amount of time that people employ to locate and access the knowledge.
- Smarter decisions on formal organizational structure: it is obtained by knowing the structure of existing social networks. The SNA gives a picture of how the work is done in an organization, how decisions are made and on the efficiency of existing organizational structures. An analysis can indicate gaps or overlaps in the information structure of organization, or indicate those who play an important role as intermediaries of knowledge. The organizational changes may be needed to fill some gaps highlighted by the SNA.

Then, tools supporting SNA allow solving concrete problems as:

- Choice of leader: an analysis of the trust and respect of which a person enjoys, can provide information relevant to the selection;
- Choice of operational unit: there are many situations in which managers must form a team of people who are connected in the best possible way within the organization: SNA is very useful for making such decisions. Project management that involving a high number of people;
- Mergers and acquisitions: when an organization plans to incorporate another organization, the SNA is helpful to analyze such situations. In addition to the merger of two corporate cultures, there is also the "union of two separate networks. Therefore, the

SNA helps managers to absorb their networks, putting together the right combination of people in relevant sectors.

- Identification of bottle - holes within organization chart and in information management.
- Regarding research and development, i.e. knowledge creation and transfer of ideas from diverse domains in new application contexts, social works have shown that innovation is not an individual act but a social process by which existing knowledge in different disciplines can cross and merge creating *new knowledge*.

SNA can be considered by companies as a *useful* and *effective* method to get a *snapshot* of the system and what is happening within their organization. In fact, the findings obtained by SNA application give a overview of *what* corrective actions to be taken in order to improve productivity, the efficiency and innovation; these strategic actions include the changing roles and responsibilities to encourage and improve the communication structures, more effective methods to improve the trust, better use of technology to be competitive.

In a business field, SNA finds a wide range of applications and Ehrlich and Carboni suggested some areas:

- **Knowledge management and collaboration.** SNA can help locate expertise, seed new communities of practice, develop cross-functional knowledge sharing, and improve strategic decision-making leadership teams.
- **Team-building.** SNA can contribute to the creation of innovative teams and facilitate post-merger integration. SNA can reveal, for example, which individuals are most likely to be exposed to new ideas.
- **Human Resources**. SNA can identify and monitor the effects of workforce diversity, on-boarding and retention, and leadership development. For instance, an SNA can reveal whether or not mentors are creating relationships between mentees and other employees.

- **Sales and Marketing.** SNA can help track the adoption of new products, technologies, and ideas. They can also suggest communication strategies.
- **Strategy.** SNA can support industry ecosystem analysis as well as partnerships and alliances. They can pinpoint which firms are linked to critical industry players and which are not.

SNA applied to sites like Facebook allows viewing an individual social graph, i.e. the map of relationships between the user and his friends.

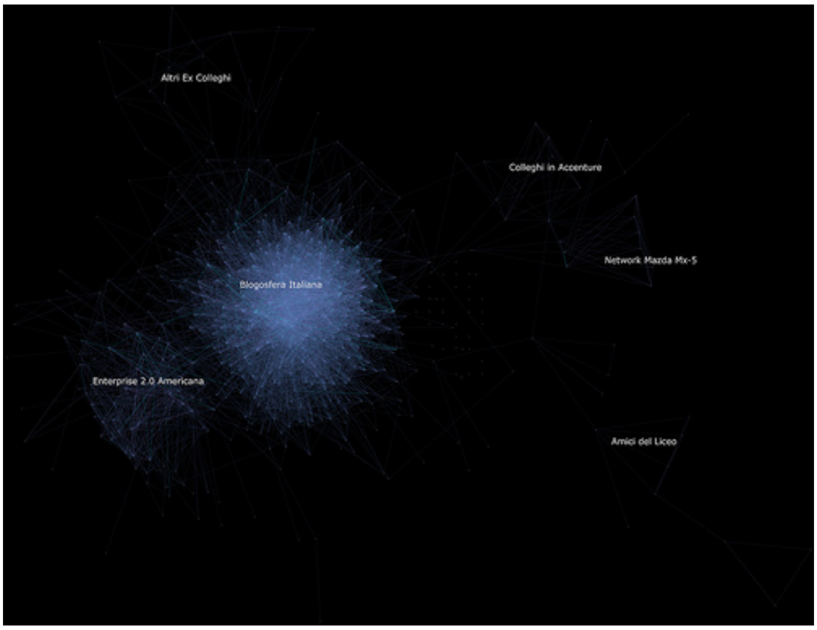


***Figure 20**– A social graph obtained by Nexus.*

On the social graph it is possible to carry out the typical SNA methods, which allows to:

- Analyze level of influence on friends;
- Discover hubs that allows to reach otherwise very distant friends (bridge) with a few steps;
- Identify areas of expertise;
- And so on.

Law enforcement agencies (and the army) use SNA to identify criminal and terrorist networks from traces of communication that they collect; and then identify key players in these networks.

An experiment to launch a data visualization site (Many Eyes) regards the mapping of New Testament social networks.

***Figure 21** – Jesus network according to the New Testament.*

In Figure 21, in addition to social relations of Jesus with respect to the various characters mentioned in the New Testament, connections between the brothers of Jesus are also represented.

1.7 From SNA to Dynamic Network Analysis (DNA)

In social network studies researchers viewed networks, actors and their ties, as static and they did not consider that networks may change over time. Indeed social networks are characterized by *dynamic* nature, so a cross-sectional analysis of networks has a limited capacity to explain the processes that are responsible of change as outcomes observed at one point in time.

In recent years, SNA has shifted more and more to dynamic analysis so researchers started to study toward network dynamics. Dynamic idea has been pursued of SNA.

From 1980s, in order to extend concepts, models, and techniques from a wide range of traditional network analysis areas including SNA, many studies were proposed. These studies on network dynamics identify a field, often called Dynamic Network Analysis (DNA), field able to overcome the limits of SNA.

DNA takes into account that the structure of the networks is not immutable in time because of ties among actors that may change over time, and changes that characteristics of a network suffer over time.

In this field, not only relations among actors are modeled but also the evolution of these relations is considered. So, when the relationships represent some things that are significant at a particular point of time, such as new job opportunities, or the establishment of a new business organization, in which the temporal dimension associated with these events plays a key role to capture important information, then SNA is a *power tool* of analysis, because it allows to take a picture snapshot of the current sample and lead several analyses in order to identify some important characteristics of network, such as centrality and cohesion.

DNA can be useful to model and analyze relationships in several potential scenarios: the informal social relationships of individuals within a family or a group of friends; the structured collaboration of employees in a large enterprise; the widespread connections through social networking services; or the covert activities of small, interconnected terrorist cells (Federico et al., 2011).

DNA brings together traditional *SNA*, *link analysis* and *multi-agent systems* within network science and network theory.

There are two aspects of this field:

- The statistical analysis of DNA data;
- The utilization of simulation to address issues of network dynamics.

Certainly this new way to conceive networks is the main difference between two methodologies, but there are others. In contrast with static analysis, the study of network dynamics requires longitudinal network data, i.e. collected over time. DNA perspective moves (from SNA) to focusing not just on who relates with whom, but also relations of actors to other units such as locations, organizations, information and so on. In order to do this, DNA looks at meta-networks; a meta-network is a multi-mode, i.e. actors can be several types (for example people and locations), multi-link, i.e. there are many types of ties (e.g., friendship and advice), multi-level, that is some actors may be members of other actors, such as a network composed of people and organizations and one of links is who is a member of which organization.

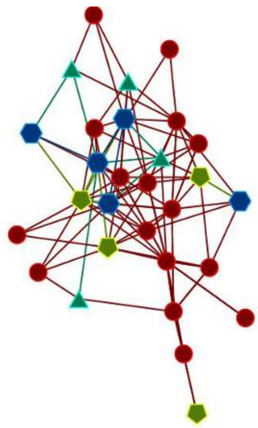


Figure 22 *– An example of a multi-entity, multi-network (Wikipedia).*

The agent-based modeling and other forms of simulations are often used to explore how networks evolve and adapt as well as the impact of interventions on these networks.

Ties in network are not simple binary but they represent the probability that there is a tie.

Analyzing network over time has become increasingly popular. In fact, the literature on network dynamics has generated a large variety of mathematical models and a large range of applications of these models to real contexts. So, it is very hard outline a state of the art of DNA, both because of it is an emergent field, both because of there is a great interest in it.

The most important methodology to study network dynamics are Markov chain, multi-agent systems, and statistical models, linked to main peculiarities of DNA. Continuous time Markov chains were proposed as early as 1977 by Holland and Leinhardt and by Wasserman. Their early work has been significantly improved upon and Markovian methods have even been automated in a popular software package, called SIENA.

A related body of research focuses on evolution of social networks (Dorien and Stokman, 1997) that use multi-agent simulation. Multi-agent system means that social actors are treated as active adaptive agents capable of taking action that can alter the network structure.

Others have focused on statistical models of network change (Sanil, Banks, and Carley, 1995; Van de Bunt et al, 1999; Snijders, and Van Duijn, 1997). Statistical approach now represents one of the most productive and ongoing fields in DNA. Statistical analysis has been used in order to build algebraic models (Pattison and Wasserman, 1995), to know the tendency over reciprocation of choice, or mutuality (Katz and Powell, 1955), to study the effects that drive the network evolution over time (Snijders, 2005), and so on.

	Markov Chain	Multi-Agent	Statistical
Problem Addressed	1. Network evolution based on Markovian assumptions. 2. Determine how underlying social theories affect group dynamics.	1. Network evolution based on node-level behavior. 2. Evaluate the impact of social intervention on group dynamics.	1. Compare the properties of networks at different points in time.
Key Assumptions	1. Future behavior of network is independent of the past. 2. There is no exogenous change in the network.	1. Node level behavior can drive group behavior. 2. Underlying social theories affecting group dynamics are known.	Assumptions vary, but include such things as dyadic independence/ dependence, over-time independence, one node class.
Limitations for change detection	1. Does not account for exogenous change. 2. Markov assumption.	1. Used to model both exogenous and evolutionary change, but not to detect change. 2. Underlying social theories must be known.	1. Does not handle over-time dependence. 2. Not a longitudinal approach.
Strengths	Determining significant social theories affecting group dynamics.	Simulating group dynamics in a social network.	Comparing social networks.

Figure 23 *– Longitudinal Dynamic Network Analysis (McCulloh and Carley, 2009).*

2. Network dynamics: stochastic models for social networks

The analysis of social networks over time has long been recognized as something of a Holy Grail for network researchers.

Wasserman F. et al., 2007.

2.1 Changing networks: network dynamics and network evolution

Interactions between people are *ubiquitous*: when people know other people, make phone calls, send e-mail, and connect on social network sites; these actions can be collected as social network. Social networks are structures composed by *dyadic*[4] ties among actors.

Social networks are *inherently* dynamic, and are subject to change. “Ties are established, they may flourish and perhaps evolve into close relationships, and they can also dissolve quietly, or suddenly turn sour and go with a bag.” (Snijders et al., 2010). For instance, when a group of persons, initially strangers, has the opportunity to interact for a certain period of time, it is very likely that in this period a friendship network will arise.

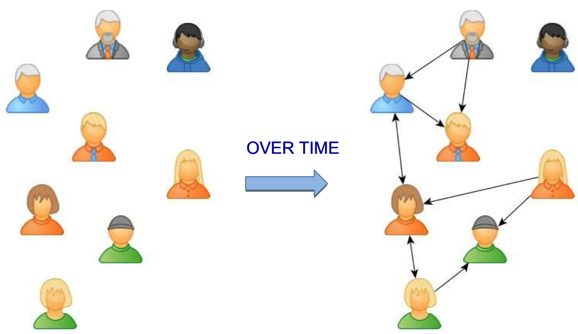


Figure 24 *– An example of social network evolution in a certain period of time.*

The temporal interval in which network is observed represents a fundamental dimension for catching these changes. Some example of changing networks include email network (Diesner et al., 2005), where the time of an email sent, the co-authorship network of scientific publications (Newman, 2004) with the year of publication, and the actor–actor collaboration network of movies (Barabasi and Albert, 1999) with its year of release. All these examples of social networks are characterized by relations that change over time, and by temporal dimensions that must be exploited to analyze and understand these networks.

When changing network is investigated, it is necessary to distinguish between dynamics and evolution of networks.

[4] Dyad consists of a pair of actors and when it is used as an adjective, dyadic describes the tie(s) between them.

Doreian and Stokman (1997) are careful to draw this distinction. Two authors describes network dynamics as a more *general* statement of network over time while they consider the network evolution as having a *stricter* meaning according to which it is possible to understand network change *via* some process, that is the mechanism that induces network change. According to the authors, dynamics is a *broader* concept than evolution; in fact, while dynamics concerns to change and is purely descriptive, evolution includes the explanations of dynamics that is the process that generates dynamics in a social network.

In many studies on social networks evolution, the process that generates the network change is assumed to be *located* in the network structure. This approach follows the line of reasoning according to which empirical social network works show certain network characteristics, while on network evolution these characteristics are taken as tendencies that *drive* the network change. For instance, many studies have demonstrated that some important characteristics of empirical choice networks are the degree of reciprocity and transitivity.

In their studies, Doreian et al. show that reciprocity is a very well above chance level from the beginning, while it does not increase over time. On the contrary, transitivity is not a very well chance level at beginning but it increases over time and remains constant at high level.

More recent, the attention of these studies moves on the evolution process seen as the result of goals-behaviors while tendencies mentioned before are the consequence of actor choices.

2.2 Longitudinal social networks

Networks evolve over time and when change itself is the object of study, the only way to investigate it is by collecting repeated measurement. Much of the interest in longitudinal social networks revolves around understanding how networks develop and change, and in years, longitudinal social networks have represented an important area of study. So, several different approaches have been developed and the literature has generated a large variety of mathematical applications, applications that have been applied to different contexts ranging from the friendship to organizational networks.

Nordlie (1958) and Newcomb (1961) studied changing interactions patterns among a set of undergraduates in two University of Michigan fraternities. These students initially did not know each other. Each student had to rank each of his fellow fraternity members on the basis of positive feeling. Data were collected for a period of fifteen weeks.

Katz and Proctor's (1959) study represents another example, in which observed network is composed by twenty-five boys and girls in an eighth-grade classroom. Data were collected in four times during the school years.

Today, the most well known study is probably The National Longitudinal Study of Adolescent Health (Harris et al., 2003). This work is a longitudinal study of a nationally that explores the causes of health related behaviors of adolescents in grades 7 through 12 in the United States and their outcomes in young adulthood during the 1994-95 school year.

The development of longitudinal network analysis methods is a well-established *problem* in the field of social networks and several methods have been proposed for analyzing repeated observations on social networks.

The dominant models of longitudinal social network analysis include Markov chain models, multi-agent simulation models, and statistical models.

Continuous-time Markov chains were proposed as early as 1977 by Holland and Leihardt and by Wasserman. Their early studies has been significantly improved upon by authors like Leenders (1995), Snijders and van Duijn (1997). In particular, Snijders developed a stochastic actor-oriented models that consists fundamentally into observing network in different points in time and simulating the changes between panels (in the next this model will be described in detail). The latter body of research is included in three special issues (the first was followed by a book version) edited by Doreian and Stokman (1996, 1997, 2001, 2003), in which they shed light on the underlying theoretical micro-mechanisms that induce the evolution of network structure. In particular, the first volume focuses on theory, methods and simulation. The authors underline on necessity of new tools and introduce Snijders' model. The second one moves on contributions in which modeling and empirical analyses are integrated. In the third volume, the authors mention some contributions that try to catch network evolution in completely different contexts. So, the first mentioned contribution is the article of Johnoson et al. in which a description of network evolution is presented; according to Doreian and Stokman, this article represents an excellent starting point. Snijders' stochastic actor-oriented model is another contribution mentioned by Doreian and Stokman; according two authors, it represents a *powerful* method to estimate the mechanisms, also called *effects*, which drive network evolution. Van Duijn et al. (2003) applied this model to the study friendship evolution among sociology freshmen. The authors showed that the friendships evolve in the beginning on the basis of *visible* similarity, but subsequently on the basis of *invisible* similarity (in terms of attitudes and activities). Another contribution mentioned by Doreian and Stokman in our special issues, is represented by Equilibrium-Correction (EC) model for the analysis of dynamic network data.

Evolutionary models often use multi-agent simulation.

Another way to represent changes of network consists to employ highly complex simulations to infer the likelihood of formation or elimination of ties based on structural configurations (Robins and Morris, 2007).

2.3 Longitudinal social data and their representations

In order to catch network changes, it is necessary to choice an appropriate kind of data set, able to describe the states through which network evolves.

Longitudinal data[5] are generally used to analyze the evolution of any social process; they are data resulting from the observations of entities, which composed the network that are measured repeatedly over time, for at least two or more distinct time.

For many years the longitudinal data collecting of social network has been very hard and for this reason the attention of SNA toward this kind of data has limited; in fact, at first the attention of researchers was mostly on single (i.e. cross-sectional) observations of networks.

[5] In appendix B there is a general description of relational data.

Longitudinal network data are typically collected as *panel data;* in this case, the network under study is composed by the same set of actors observed at least two consecutive time points, called the *panel waves*.

Starting in the 1980s, network panel data started to be collected more widely. An example could be the study of collaborations among the employees in a firm in which the collaborations may change over time but the group of actors under study remains the same.

In literature, a classical example of panel data is Freeman's (1980) EIES (Electronic Information Exchange System) data. The network is composed by 32 researchers, sociologists, anthropologists, mathematicians, psychologists, and statisticians, who participated in an early study on the effects of electronic information exchange (a precursor of email communication) over the course of an eighteen-month period.

In Figure 26 the networks observed at two time points are illustrated.

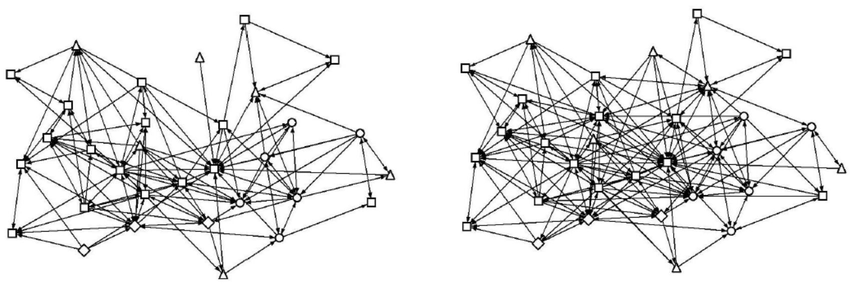


***Figure 25** – EIES friendship network in two moments of observation (Snijders, 1994).*

According to Moody (2005), a good way to visualize network changes "is to show how the network emerges over time by adding nodes and relations as they appear, but placing them in the display plane based on the final aggregate structure." (Moody et al., 2005).

To develop dynamic network images, it is necessary to understand how time is *encoded* in social networks; time can be interpreted as continuous or discrete parameter. When time is consider continuous, the visualization of the network changes consists in streaming relational events recorded with exact starting and ending times, whose visual representation should unfold as a continuous social process. This way of visualization needs the use of animation, which allows to mapping of empirical time. In this case it requires special media.

Moody (2005) utilized dynamic movies (SoNIA[6]) that allowed nodes to move as a function of relational change.

A discrete interpretation of time is cross-sectional *snapshots* of network. This represents a more common way of visualize the network changes, and consists into create static *snapshots* (that are configurations of a network at a particular moment of observation) at a fixed interval. In this case, the analysis focuses on the change from one configuration of network to another without consider the sequence of changes that generate change.

Moody (2005) referred to network flip that consisted in a kind of visualization in which the nodes remained in the same position while the arcs filled the holes among them.

In figure 26, another example of visualization is illustrated; the network is represented in two different moments of observation and the chosen approach to visualize them is an *aggregation* approach, in which all nodes remain in the same position determined from a network aggregated over all time points.

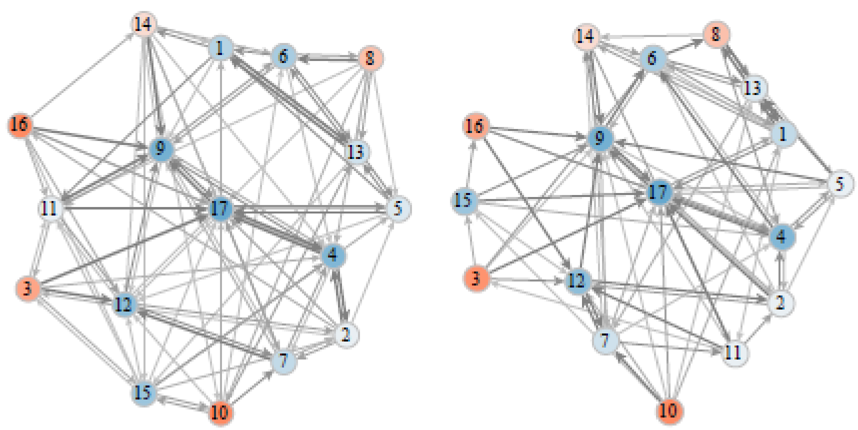


Figure 26 *– Example of visualization (Brandes and Nick, 2011).*

Besides visualization, the process composed by some snapshots generates insights into how network properties change over time, such as the average degree or clustering coefficient. A key benefit of this approach is that any measures of SNA can be applied to network.

[6] SoNIA (Social Network Image Animator) is a Java-based software package for visualizing of network evolution over time, development by Dan McFarland and Skye Bender-deMoll.

2.4 Network dependencies

The nature of networks leads to dependence between actors, and also to dependence between network ties.

In fact, social networks are characterized by several kinds of dependencies, which have been found empirically as well as theoretically (Snijders, 2011). Some kinds of these dependencies are *homophily*, *reciprocity* (Sahlins, 1972), that can be expressed by the saying “if you contact me, I contact you”, and *transitivity* (Davis, 1970), “friends of my friends are my friends”.

The *homophly principle* is defined as following: “Homophily is the principle that a contact between similar people occurs at a higher rate than among dissimilar people. The pervasive fact of homophily means that cultural, behavioral, genetic, or material information that flows through networks will tend to be localized” (McPherson et al., 2001). So, homophily is the tendency of similar actors to relate to each other. An example is a network of football fans; it is more likely that a fan of Milan team goes to watch a football match together another fan of his same team rather than a fan of enemy football team. This means that there is a tendency to form ties with who is similar according to a certain attribute, in this case favorite football team.

More complicated types of dependency involve more than two actors. The *reciprocation* is a basic feature of social networks, found previously by Moreno 1934. This dependency implies that an actor *i* that receives a tie by actor *j* then it is more likely thinking that *i* will reciprocity to *j*.

The well known of dependencies between two actors is *transitivity* of ties. Looking Figure 28, if an actor *i* and *j* know, and *j* knows also *h*, and there is a tendency toward transitivity, then it is more likely thinking that *i* will link to *h*.

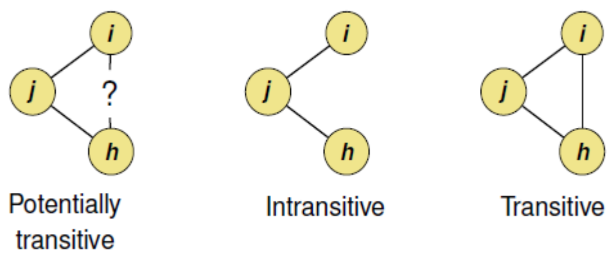


Figure 27 *– An example of transitivity effect (Snijders, 2012).*

In literature many ways to represent network dependencies in statistical models have been developed, accordingly various approaches have arisen. A first approach that is now considered as a relict is to incorporate the network structure through covariates[7].

Gulati and Gargiulo (1999) studied alliances between organizations and they tested the idea according to firms rely with whom to enter in alliances on the basis of information derived from the network. According to two authors, earlier observations of network can be used to produce covariates.

Another way is to represent network dependencies is to control it. The best-known example is permutational procedure, in which columns and rows of adjacency matrix are permuted simultaneously in such a way that network structure remains intact.

In the end, the third approach is to explicitly model the structural dependencies between tie variables. When the aim is that to study the network dynamics, the dependencies are spread out in time.

2.5 Stochastic models for network dynamics

When a network is observed over time, its relational structure may change: in the course of time, some ties could be created, others could be eliminated or could maintain constant.

Recent interest on longitudinal social networks revolves around understanding how networks change over time, so scientists seek to build models of social processes that result in observed structures.

Regarding statistical model, a first problem with which researchers have been confronted is whether network evolution can be see as one *jump*, or as the result of a *series* of small changes.

This has generated a first split: on one hand, several methods have been proposed for analyzing repeated observations on social networks using models in which changes are made in discrete steps from one observation moment to the next (Banks et al., 1997; Katz el al., 1959; Sanil et al., 1995),

[7] In statistics, a covariate is a variable that is possibly predictive of the outcome under study *(Wikipedia).*

on other hand, it is natural thinking that evolution process is not correlate with moments of observations, but as a result of continuous process.

The ones most directly amenable to statistical analysis are those in which network *X(t)* is a *continuous-time Markov chain*.

So, network is a changing simple directed graph, in which arcs represent social ties that can be regarded as *states*.

Friendship between people and pacts between companies are examples that allow understanding the concept of states.

The network is represented by node set $\{1, \ldots, n\}$ with tie variables x_{ij}, where x_{ij} is 1 if tie $i \rightarrow j$ is present or, x_{ij} is equal to 0 if tie $i \rightarrow j$ is absent. Tie variables are collected in *nxn* adjacency matrix. Self-ties are excluded, so that $x_{ii} = 0$ for all i.

For networks in which ties are states, the dependence among ties can be represented by assuming that changes are dependent upon the existing network structure (Snijders, 2009). This means to assume that the network is a Markov chain.

A Markov chain is a stochastic process[8], and within a social network context, saying that a Markov process means that the conditional distribution[9] of the changes at any moment depends only on the current network configuration, not on previous configurations. So dependencies between tie variables are represented by capital letters X_{ij} to expressing their stochastic nature.

In this class of models, the network is studied in each moments of observation and the basic idea is that in interval time between successive observations a continuous unobserved evolution takes place. In Figure 28, graphical representation of this evolutive scheme is depicted:

[8] A stochastic process is a family of random variables $\{X_t, t \in T\}$ defined on given probability space, indexed by the time variable t, where t varies over an index set T.

[9] The conditional probability is called transition probability at time t to state i to state j.

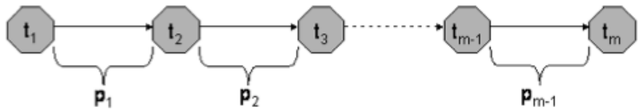


Figure 28 – *Series of moments of observation of network (Savoia, 2007).*

In Figure 28, there are *m* moments of observations (t_1, t_2, . . . , t_m) of network and among these moments there are *m-1* intervals (p_1, p_2, .., p_{m-1}) during which the network evolves in *continuous* and *unobserved* way. The idea of regarding the evolution of social phenomena as being the result of a continuous-time process, even through observations are made at discrete time points, was proposed already by Coleman (1964).

The simplest approach to construct dynamic network models is that called *independent arcs* because primary elements of the model are ties embedded and probabilities of tie changes. In this direction, Wasserman and Leenders works represent a step forward; they developed a model in which the probability of relational change depends by network structure; in particular, the *reciprocity model* (Wasserman, 1977, 1979, 1980) accounts for interdependencies between dyadic partners. This reflects reciprocity of relations, but not more complicated types of dependence.

Starting on these two works, Snjiders developed the actor-oriented model proposed for the first time in 1995 and subsequently modified and extended.

2.6 The actor-oriented model

Snijders (1995, 1996, 2001, 2005) and Snijders & Van Duijn (1997) introduced the actor-oriented model developed to describe and explain the network evolution over time.

Actor-oriented model interprets relational network evolution as the result of actors' choices to create, eliminate or maintain their ties within network, each of whom is individually optimizing his/her own utility.

For each change in relational network, the perspective is taken of the actor *whose tie* is changing. It is assumed that actor *i* controls the set of outgoing tie variables (collected in the *i'th* row of the adjacency matrix). Network changes occur step by step, that is only by one tie at a time. These mini-steps accumulate and can result in a big change. Given current network structure, actors act independently way, and apply a *myopic* strategy, so they consider only the situation obtained immediately after the mini-step.

The choice made by an actor to perform a certain change depends on actor's expectation of the utility of his/her state after the change. In fact, each knows relational structure of whole network and on the basis of it they evaluate their situation within network and operate changes in order to increase their *position*.

The moment in which actor *i* has the opportunity to change one of his/her ties, and the particular change that he/she makes, can depend on network structure and on attributes represented by observed covariates. The *moment* in which each actor can be change is stochastically determined by the *rate function*, while the particular *change to make* is modeled by *objective function*. This function is divided in three other functions: by *evaluation function* that models the satisfaction of actors for different possible configurations of network, by *endowment function* that is liked to gratification derived from different actions that have led to determinant configurations and, finally, by a *random component* that is a random variable indicating the part of actors' preference that is not represented by systematic components f_i and g_i. So the objective function is defined by its three components:

$$f_i\big(\beta, x(i \rightsquigarrow j)\big) + g_i(\gamma, x, j) + \varepsilon_i(t, x, j)$$

in which the term $i \rightsquigarrow j$ indicates change state of tie of actor *i* with actor *j*.

In the following paragraphs the rate function and each component of objective function have been explained.

2.6.1 Rate function

The rate function is the expected number of opportunities for change per unit of time, in other words it indicates how frequently actors make mini-steps.

This function is denoted by expression:

$$\lambda_i(\rho_m, \alpha, x)$$

and the rate function can be formally defined by:

$\lambda_i(x) = \lim_{dt \to 0} \frac{1}{dt} P\{X_{ij}(t+dt) \neq X_{ij}(t)\}$ for some $j \in \{1, \dots, g | X(t) = x\}$

This function depends by generic rate (ρ_m) of change and individual characteristics (α) of actors and network (x). The rate of change ρ_m is the average of tie that actors change between two subsequent observations. The parameter α represents the vector of actors' characteristics. Thus, when actors' characteristics are relevant, the rate of change varies among actors.

The simplest specification of the rate of change is that all actors have the same rate of change:

$$\lambda_i(\rho_m) = \rho_m$$

This means that for each actor, the probability that this actor makes a mini-step in the short time interval $(t, t+dt)$ is approximately $\rho\, dt$, and in a short time interval there is independence among actors in whether they take a mini-step. Then $\lambda_i(x) = \rho$ for all i. The waiting times D between successive mini-steps of each given actor then have the exponential distribution with probability density function $\rho e^{-\rho d}$ for $d > 0$, and the expected total number of mini-steps made by all actors between time points t_a and t_b is $g\rho(t_b - t_a)$: as is intuitively clear, this expected number is proportional to total number of actors g, proportional to the rate of change ρ, and proportional to the time length $t_b - t_a$.

2.6.2 Objective function: evaluation function

The basic idea of the actor-oriented model is that, when actor *i* has the opportunity to change in his outgoing, tie variables ($X_{i1},..,X_{ig}$), this actor selects change which gives the greatest increase in so-called objective function plus a random term.

Thus, the evaluation function represents the *preference distribution* of each actor over the set X of all possible networks.

The evaluation function can be formally defined by:

$$f_i = (\beta, x) = \sum_{k=1}^{L} \beta_k s_{ik}(x)$$

The functions $s_{ik}(x)$ represent meaningful aspects of the network, as seen from the viewpoint of actor *i*, and depend on the network but may also depend on actor attributes. The weights $\beta=(\beta_1,\ldots,\beta_L)$ are statistical parameters that represent the vector of structural/individual effects included in the model to determine the preference of actor.

Effects depending only on the network are called structural or endogenous effects, while effects depending only externally given attributes are called covariates or exogenous effects.

In the following the endogenous functions $s_{ik}(x)$ used in the application are shown (in the appendix C the remaining have been presented):

- *Density effect*: propensity of actor *i* to create arbitrary ties with any other members of the network; this effect is defined by outdegree:

  $s_{i1}(x) = \sum_j x_{ij}$

Figure 29 – *Representation of density effect.*

- *Reciprocity effect*: propensity of actor *i* to create a tie with an actor that is just linked to i; defined by the number of reciprocated ties:

  $s_{i2}(x) = \sum_j x_{ij}x_{ji}$

Figure 30 – *Representation of reciprocity effect.*

- *Transitivity effect*: tendency of actor *i* to form a tie with an actor that has ties with other actors with *i* is linked; defined by the number of transitive patterns in *i*’s ties:

  $s_{i3}(x) = \sum_{j,h} x_{ij}x_{ih}x_{jh}$

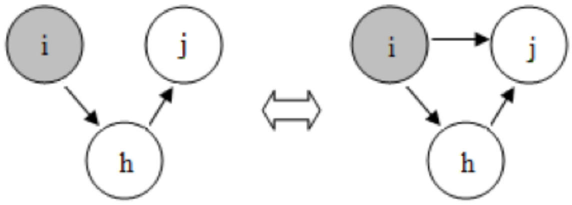


Figure 31 – *Representation of transitive effect.*

- *Balance effect*: tendency of actor *i* to create ties with structural similar actors; this effect is based on structural equivalence; it is defined by the similarity between outgoing ties of actor *i* and the outgoing ties of the other actors *j* to whom *i* is tied:

$$s_{i4}(x) = \sum_{j=1}^{g} x_{ij} \sum_{\substack{h=1 \\ h\neq i,j}}^{g} \left(b_0 - \left|x_{ih} - x_{jh}\right|\right)$$

where b_0 is a constant included to reduce the correlation between this effect and the density effect. Given that the density effect is included in the model, the value of b_0 only amounts to a re-parameterization of the model (viz., a different value for the parameter of the density effect). The proposed value is such that it yields a zero average for s_{i4} over the first $M-1$ observed networks $x(t_m)$ $(m = 1, \ldots, M-1)$ and over all actors, and is given by:

$$b_0 = \frac{1}{(M-1)g(g-1)(g-2)} \sum_{m=1}^{M-1} \sum_{i,j=1}^{g} \sum_{\substack{h=1 \\ h\neq i,j}}^{g} \left|x_{ih}(t_m) - x_{jh}(t_m)\right|$$

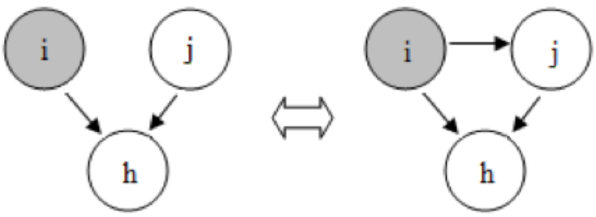


Figure 32 – *Representation of balance effect.*

For each actor, there are several exogenous effects related to characteristics of actors; in the following, those used in the application are shown:

- *The covariate-similarity effect*: its positive parameter implies that actors prefer to others with similar values on this variable (they have an individual attributes in common).
- *Covariate ego x covariate alter:* a positive effect means that actors with a higher value on the covariate will prefer ties to others who likewise have a relativity high value.
- *Characteristic of ego:* choice of actor ego to create or eliminate ties with other actors on the basis of his individual attribute.

2.6.3 Objective function: endowment function

Sometimes the order, in which changes could occur, makes a difference for the desirability of the states of the network (e.g. in the case of reciprocated ties). Then a specific effect may have a different intensity depending if the creation or elimination of a new tie is evaluated; in fact, often the removal of tie dues the loss of that of an actor has invested in terms of time and energy in a relation.

The endowment takes into account these differences and it can be defined conveniently as a weighted sum:

$$g_i(\gamma, x, j) = \sum_{h=1}^{H} \gamma_h r_{ijh}(x)$$

in which γ is the vector of parameters that determine the endowment function and represent the entity of difference between creation and elimination of a tie, while $r_{ijh}(x)$ includes a factor x_{ij} it refers to the gratification experienced for breaking a tie, whereas the inclusion of a factor $(1 - x_{ij})$ refers to gratification for creating a tie. The possible functions $r_{ijh}(x)$ are the following:

- *Breaking off a reciprocated tie*:
  $r_{ij1}(x) = x_{ij}x_{ji}$
- *Number of indirect links for creating a new tie*: representing the fact that indirect links (at geodesic distance 2) to another actor may facilitate the creation of a new tie:

$$r_{ij2}(x) = (1 - x_{ij}) \sum_h x_{ih}\, x_{hj}$$

- *Effect of dyadic covariate W on breaking off a tie:*

$$r_{ij3}(x) = x_{ij} w_{ij}$$

2.6.4 Complete model

Briefly, according to actor-oriented model the network evolution is interpreted, on the hand, by actors' evaluations that have the opportunities to change their outgoing ties within network, on the other hand, by frequency of these changes. The first action is modeled by objective function, composed by three components (evaluation function, endowment function, and random component), while the second one is modeled through rate function.

Two stochastic parameters associated to preferential structure of actors express the satisfaction for current state (β) and gratification for specific changes that have led this state (γ), while two parameters related to rate with which actors operate their changes represent generic rate of change (ρ) and actors' characteristics (α). In general, some simplifications are assumed: the random component is not modeled, there are not differences between created or eliminated tie (endowment function is not considered), and attributes of actors are not important for frequency of changes. Thus, the model considers that the objective function is expressed only by evaluation function and the rate function is the same for all actors (a model with a constant rate function is usually easier to explain and can be simulated in a simpler and quicker way).

Figure 34 shows complete (on the left hand) and reduced (on the right hand) formulas of objective function and rate function.

$f_i(\beta, x(i \rightsquigarrow j)) + g_i(\gamma, x, j) + \varepsilon_i(t, x, j)$	$f_i(\beta, x(i \rightsquigarrow j))$
$\lambda_i(\rho_m, \alpha, x)$	ρ_m

Figure 33 *– Complete (left side) and reduced (right side) formulas of objective and rate functions.*

Except in cases in which rate function depends on actors' characteristics, it is advisable to start modeling using a constant rate function and add the complexity of an non-constant rate function only after at a later stage. Analogous, except evident differences between creation and elimination of ties, it is advisable to start with endowment function nothing and add it only in a second moment.

2.6.5 Estimation: method of moments and Robbins-Monro algorithm

The actor-oriented model is too complicated for explicit calculation of possible evolutions (the space of possible evolutions is very large) and excepted values, but it can be simulated in a rather way. The method of moments is used for parameter estimation that identifies excepted values and the solution of equation of moments is approximated by iterative progressive algorithm. In the following, estimation method is explained for actors with a constant rate function ρ_m between two sub-sequent observations, and without an endowment function.

To explain the method of moments, it is made reference to Snijders (2006).

The objective function is given by expression:

$$f_i(\beta, x) = \sum_{k=1}^{L} \beta_k s_{ik}(x)$$

Greater values of β_k are expected to lead for all actors *i* to higher values of the statistics s_{ik} $(X(t_{m+1}))$, when starting from a given preceding network $x^{obs}(t_m)$. The observed networks are denoted by x^{obs}.

The principle of estimation now is to determine the parameters β_k in such a way that, summed over *i* and *m*, the expected values of these statistics are equal to the observed values. These observed target values are denoted:

$$s_k^{obs} = \sum_{m=1}^{M-1} \sum_{i=1}^{g} s_{ik}\left(x^{obs}(t_{m+1})\right) \quad (k = 1, \dots., L)$$

and collected in vectors s^{obs}. Since the expected values cannot be calculated explicitly, they are estimated from simulations.

These simulations run as follows:

1. For two digraphs x and y define their distance by:

$$\|x - y\| = \sum_{i,j} |x_{ij} - y_{ij}|$$

   and for *m=(1,…,M-1)* let cm be the observed distances:

$$c_m = \|x^{obs}(t_{m+1}) - x^{obs}(t_m)\|$$

2. Use given parameter vector $\beta=(\beta_1,\dots,\beta_L)$ and the fixed rate of change $\lambda_i(x)=1$.
3. Make the following steps independently for *m=1,…,M-1:*
   a) Define the time as 0 and start with the initial network:

$$X_m(0) = x^{obs}(t_m)$$

   b) Simulate the actor-oriented model $X_m(t)$ until the first time point, denoted R_m, where:

$$\|X_m(R_m) - x^{obs}(t_m)\| = c_m$$

4. Calculate for *k=1,...,L* the generated statistics:

$$S_k = \sum_{m=1}^{M-1} \sum_{i=1}^{g} s_{ik}\big(X_m(R_m)\big)$$

This simulation yields, for the input parameter vector β, as output the random variables $(S,R) = (S_1,\dots,S_L,R_1,\dots,R_{M-1})$. Note that the time parameter within the *m*'th simulation runs from 0 to R_m.

For the estimation procedure, it is desired to find the vector β for which the expected and observed vectors are the same:

$$\varepsilon_{\hat{\beta}} S = s^{obs}$$

that represents the *moment equation.*

The procedure of Snijders (2001) for approximating the solution to the moment equation is a stochastic iteration method consisted by a variation of the Robbins-Monro algorithm. This procedure is divided in three phases:

- The first phase is the purpose of roughly estimating the sensitivity of the expected value of S_k to variations in β_k;
- In the second phase the estimate is determined; and t

- The third phase is for checking the resulting estimate and calculating the standard errors.

2.7 The co-evolution of social networks and behavior dynamics

As mentioned before, if network is observed over time, relational network may change. This change may result from structural network mechanisms, like transitivity, popularity, and others, or from mechanisms that depend on *individual characteristic.*

In the last years, it has been understood that individual characteristics of actors play a fundamental role in network change.

The behavior of network structure is defined by influence and selection processes.

An extension of actor-oriented model can be used to analyze these two processes.

2.8 The influence and selection processes

A natural interdependence between network structure and individual characteristics of the network actors exists; the most well known pattern of this type is *network autocorrelation*.

To explain this phenomenon, it is necessary to take into account influence and selection processes.

In fact, when a network is observed over time, on one hand, characteristics of actors, pairs of actor, and structural positions of actors can affect the network evolution (Veenstra and Steglich, 2012); the selection process summarizes this type of dependencies. An example is homophily principle, previously introduced, according to which creation of a relationship is based on the similarity of two actors, also known as *preferential attraction*.

On other hand, networks can affect characteristics of actors and their behavioral development (Veenstra and Steglich, 2012). This kind of dependencies is summarized as influence process, defined as change in an actor's thoughts, feelings, attitudes, or behaviors because of interaction with another actor. An example is *assimilation principle*, according to two connected actors become similar over time, adapting their individual characteristics to match those of their social neighborhood (Steglich et al., 2006).

Veenstra and Steglich talk about these two processes and do a very important consideration.

They depicted transitions that can occur between two actors *i* and *j*; so, they supposed that, at first observation (configuration on the left of Figure 35), actor *i* considers *j* as a friend but *i* is not behaviorally similar to *j*. Whereas, at second observation (configuration on the bottom of the Figure 35), *i* again considers *j* as a friend but, this time, *i* is behaviorally similar to *j;* according to previous literature, a influence process exists; in fact, the actor *i* adapts its behavior to that of actor *j*, becoming so similar to actor *j* (transition (a) in Figure). It is possible to conclude that the friendship between *i* and *j* remains intact during the unobserved period.

In an alternative scenario (configuration on the top of Figure 34), the initial friendship between two actors could finish; after this, actor *i* may change his behavior (transition (c) on the right of Figure), and then *i* may have renewed the friendship with *j* (transition (d)). In this case, actor *i* changes his behavior when the relationship with *j* is absent.

So, through (a) influence process suggests that relationship is stable while behavior change; in contrary, through (b), (c), and (d), selection process suggests that behaviors remains similar but relations change (Veenstra and Dijkastra, 2011).

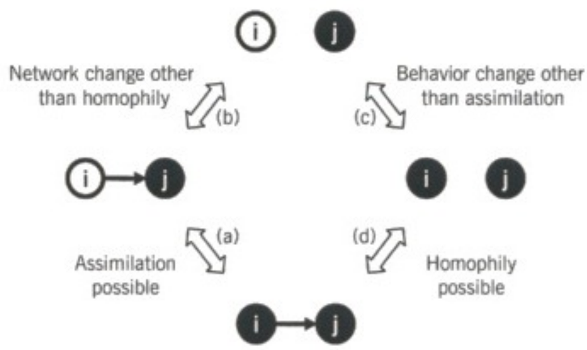


Figure 34 *– Elementary change process in a dyad (Veenstra and Steglich, 2012).*

So, it is very difficult to understand the kind of process that carries from first configuration to second configuration. In fact, the actor *i* may link to *j* through influence process (transition (a)), or, on the contrary, through selection process (transitions b and d).

In order to overcome this problem, it is necessary to take into account the possibility of unobserved change. This implies continuous time data collection or, if data are measured at discrete moments, at least continuous modeling (Veenstra R. and Steglich C., 2012).

The model expresses that, in response to the current network structure and the current behavior of the other individuals in the network, individuals can change either their peer network (make a new friend or break a relationship) or their behavior (increase or decrease in behavior) between two time points.

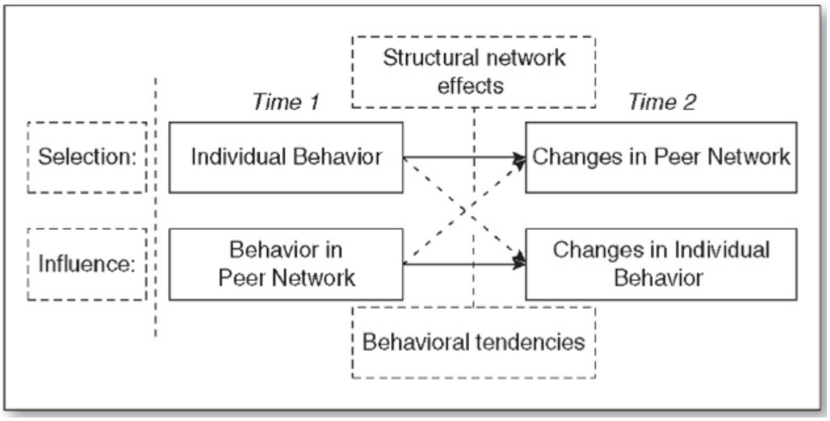


Figure 35 *– Representation of selection and influence processes (Veenstra and Dijkstra, 2011).*

The actor-oriented model considers existing of a continuous time change process, interpreting the discrete configuration that the network takes over time as the cumulative result of an unobserved sequence of *small changes*, resulting from choices made by actors between moments of observation.

2.9 Some cases of study

The actor-oriented model is a good tool to investigate tie's changes in a social network, so, over time, it has been applied to describe how real contexts work.

Adolescent friendships have received a great deal of attention in research as they are particularly adequate to conduct an application of actor-oriented model: making friends is an essential part of life for adolescents at school (it is easy to find data on them) and friendship networks change over time (the mechanisms that drive network evolution exist). Some studies from education science have shown the effects of friendship on learning, adaptation, and psychological health, others have revealed as some factors, such as alcohol, smoke or music, influence the choice of individuals to create ties. Thus, the main part of studies in which actor-oriented model is applied, has been focused on adolescents network.

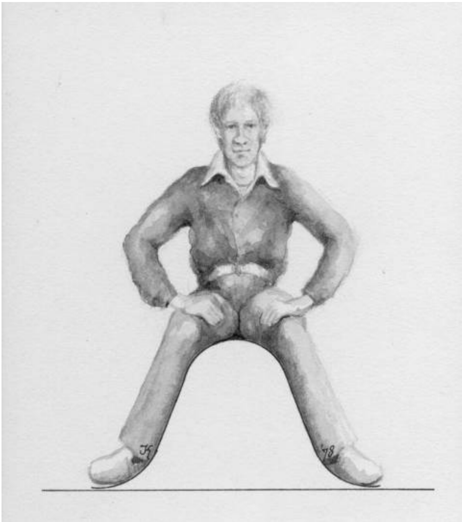
From 2001, research on networks and their dynamics is also flourishing in the strategy and organization literature; in particular, an important topic concerned the ongoing dynamics of networks that result from collaborative choice (among others, Ahuja, 2000; Gulati 1995, 1999; Gulati & Gargiulo, 1999; Hagedoorn, 2006; Powell, 1998). Over time, the idea has established that "today's choice of an alliance partner affects tomorrow's options as it changes the network structure and thereby the future alternatives and strategies of all fellow network members" (van de Bunt and Groenewegen, 2007). This concept has resulted the production of some studies in which the actor-oriented model was applied to organizational context. Van de Bunt and Groenewegen (2007) applied the model to a certain number of firms participating to the same project, in order to understand how these firms choose collaborative partners given their goals, their characteristics, and their network configuration.

Regarding the application of actor-oriented model to research networks, only recently a very few studies are presented in literature.

Kronegger et al. (2012) considered four scientific disciplines in the Slovene system of science to test small-world and preferential attachment processes. They utilized the actor-oriented model to identify the motivations that influence actors to form new co-authorship ties. The authors concluded that the formation of ties is consistent with small-world structure of networks, and, at the same time, the preferential attachment is far more complex than advocates of a global autonomous mechanism claim.

Katerndahl (2012) conducted a study over a 13-year period among faculty in the Department of Family and Community Medicine at the University of Texas Health Science Center at San Antonio to understand how the research collaboration network evolved within a department. In order to study changes in their patterns of connections and identification of network characteristics associated with the development of new connections, she entered networks obtained in SIENA software.

3. Software for stochastic model



A picture of proof Tom A.B. Snijders.

Jacqueline Kasemier.

3.1 Software for relational data

In order to facilitate the analysis of networks, their structure and their evolution, a variety of computer packages that can handle relational data have been developed.

The analyses routines on social networks are divided into three types of methods:

- Descriptive methods to calculate network statistics (e.g. centrality or transitivity);
- Procedure-based analysis based on more complex (iterative) algorithms (e.g. cluster analysis);
- Statistical modeling based on probability distributions (e.g. exponential random graph models or QAP correlation) and on network evolution (Snijders, 2001, 2005).

From generalist tools, such as UCINET (Borgatti et al., 1999), Pajek (Batagelj and Mrvar, 2007), NetMiner II (Cyram, 2003), StOCNET (Huisman and van Duijn, 2003), to more specialize applications, such as Netdraw (Borgatti, 2002), a variety of software solutions are available for network analysts.

In the following paragraphs, on the basis of previous division a cognitive framework of the currently available software has been introduced, giving a brief description of their goals, their main features, and especially their limits if the analyst would intend to carry out an analysis of dynamic type.

3.1.1 Cognitive framework of the current software.

The great interest about SNA has been consolidated through the increasing availability of a wide array of software packages for the automatic elaboration of relational data.

These kinds of software have been developed based on research in fields as diverse as computer science, bioinformatics and sociology.

Perhaps, the best known and most frequently used software package is UCINET (Borgatti et al., 1999) that is a comprehensive program for the analysis of social networks and other proximity data.

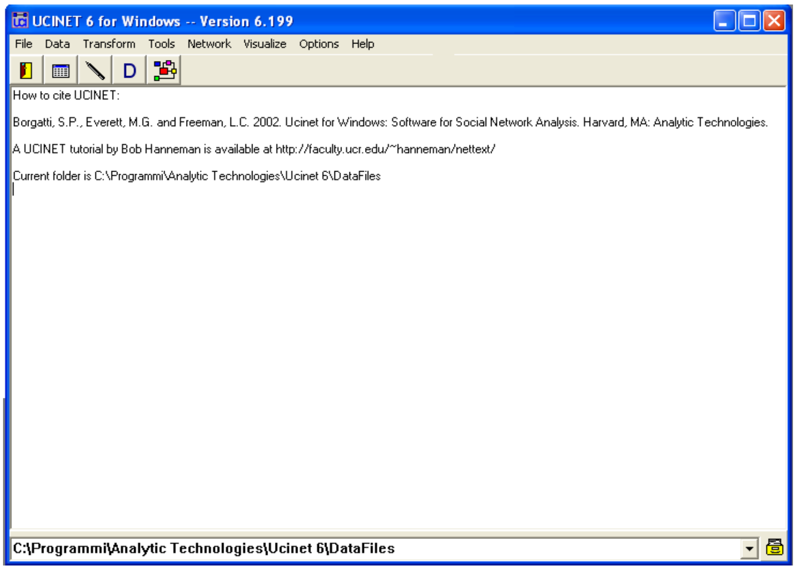


***Figure 36** – Home Screen of UCINET.*

The program covers a large number of network analytic routines for the detection of cohesive subgroups (cliques, clans) and regions (components, cores), for centrality analysis, for ego network analysis, and for structural holes analysis.

UCINET contains graphical tools to draw scatter plots, and tree diagrams but it does not contain graphical procedures to visualize networks; so, it has a speed button to execute the program NetDraw (Borgatti, 2002).

NetDraw includes some analysis procedures of networks, such as the implementation of centrality measures, the identification of the cliques and clusters, but mainly deals with the visualization of networks and various modifications can be applied to graphs, turning, by varying the size, changing the colour, the shape, the size of the nodes and arcs.

David Knoke (1982) studied the spread of administrative reform among city governments after the transformation of American municipal government. He considered ten organizations that were involved in the local political economy of social welfare services in a Midwestern city. Data derived from this study were loaded in Netdraw to build the network shown in Figure 37.

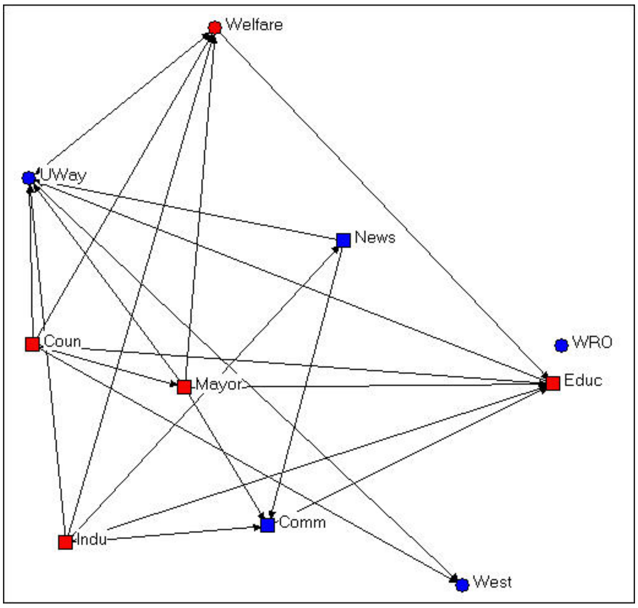


Figure 37 *– Example of network visualization with NetDraw (Hanneman and Riddle).*

So, each node represents the organizations, while the arcs are the information exchanged by organizations. The network was characterized adding the attributes related to each actor; so, to each node was associated a different color on the basis of type of organizations; so, those in red are governmental organizations (Welfare, Coun, Educ, Mayor, Indu), those in blue are non-governmental organizational (UWay, News, WRO, Comm, West). Besides, to each organization was assigned different shape on the basis that organization is generalist (i.e. perform a variety of functions and operate in several different fields) or it is specialist (e.g. work only in social welfare). So, the shape square was assigned to generalists and circle to specialists.

Pajek (the word in Slovenian for spider) (Batagelj and Mrvar, 2007) is another network analysis and visualization program, specially designed to handle large data sets.

In Pajek, a network is defined in accordance with graph theory, so nodes are actors and arcs represent ties between them. Each node can be characterized by its name, shape, and colour, whereas edges by thickness, colour, and label. In Figure 38, an example of input file in Pajek is shown.

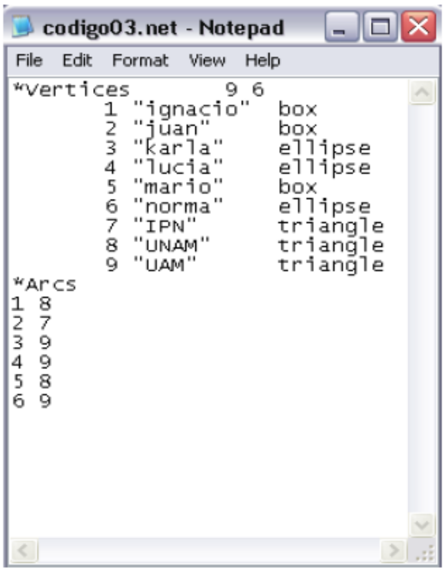


Figure 38 *– An example of input file for Pajek.*

The main goals in the design of Pajek are to facilitate the reduction of a large network into several smaller networks that can be treated further using more sophisticated methods, to provide user with powerful visualization tools, and to implement a selection of efficient network algorithms. Network data can be entered in several ways, for example it is possible to import data from software packages with other formats, such as UCINET file.

The very particularity of Pajek uses six different data structures: networks (nodes and arcs/edges), partitions (classifications of nodes, where each node is assigned exclusively to one class), permutations (reordering of nodes), clusters (subsets of nodes), hierarchies (hierarchically ordered clusters and nodes), and vectors (properties of nodes).

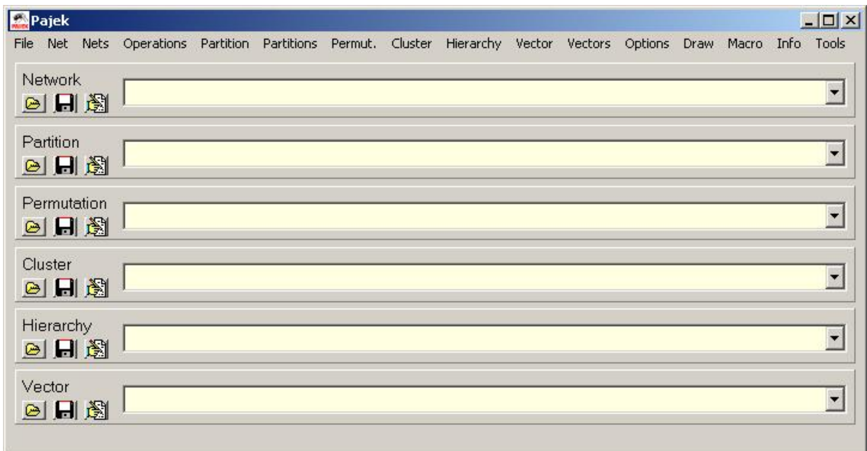


Figure 39 *– Home Screen of Pajek.*

Besides, network can be drawn in many different ways, so the analyst should rely on systematic rather than ad hoc principles for network drawing. For these reasons, Pajek contains automated procedures for finding an optimal layout that are a better way to obtain a basic layout than manual drawing, because the resulting picture depends less on the preconceptions and misconceptions of the investigator.

NetMiner II (Cyram, 2003) is another software that combines social network analysis and visual exploration techniques. It allows users to explore network data visually and interactively, and helps to detect underlying patterns and structures of the network. Like Pajek and NetDraw, NetMiner has advanced graphical properties. Moreover, almost all results are presented both textually and graphically, contrary to both other programs, where the user needs to request visualization of the results of a certain analysis.

3.2 A brief introduction to some tools for dynamic analyses

Existing computers packages for investigating relational are used to identify, represent, analyze, visualize, including mathematical models of social networks.

By using these software packages, researchers can investigate networks of different size, from small (e.g. the group of few persons) to very large (e.g. on line sites, trade between countries).

Another benefit related by utilization of these packages is the display of networks; the visualization of social networks helps the understanding of network data and calculation of some metrics. Besides, network analysis software are used when the network must be characterized with some attributes; so it is possible to change the shape, colors, size and other properties of the network representation.

These packages allow to make many activities but they have not always all tools necessary to carry out a dynamic analysis of networks because they analyze a single network, that is the network configuration in a certain moment of observation.

When scientists moved your attention on dynamics of social networks, need of tools that go beyond traditional SNA was born.

In response to these needs, a new variety of software has been presented for data collection, analysis, visualization and simulation.

Multi-agent dynamic-network simulation (MADN) systems are able to assess the dynamics of complex system observed in different moments in time. These computer simulations to predict vary scenarios that will happen, and to understand what is likely to happen (Carley et al., 2007).

"In MADN systems the actions performed by individual agents lead to changes in the underlying networks that then affect what actions agents take in the future" (Carley et al., 2007).

These findings have been used for DyNet[10], a simulation model used in integrated CASOS (Computational Analysis of Social and Organizational Systems) toolset. CASOS is software conceived on concept of *tool-chain* and it allows users to investigate change detection.

A software tool-chain consists of a number of small self-contained tools such as editors, compilers, debuggers and analysis software. Each of these tools might be developed as a separate product by different people and may vary in complexity, size and features.

In a similar manner, a tool-chain in support to dynamic analysis of social network needs to consist of a number of self-contained tools that support various steps of the process analysis (Carley et al., 2007).

CASOS group has developed a suite of tools that acts as a chain to extract networks from texts, analyze these networks, and then engage in what-if reasoning. This tool suite takes into account multi-mode, multi-link, and multi-time period data including attributes of nodes and edges.

Figure 40 shows all tools that are contained in CASOS: AutoMap for extracting networks from texts, ORA for longitudinal network analysis, and DyNet for what-if reasoning about the networks.

[10] DyNet is a complex system simulation model in which the social and knowledge networks co-evolve as agents interact, communicate, and engage in tasks (Carley et al., 2007).

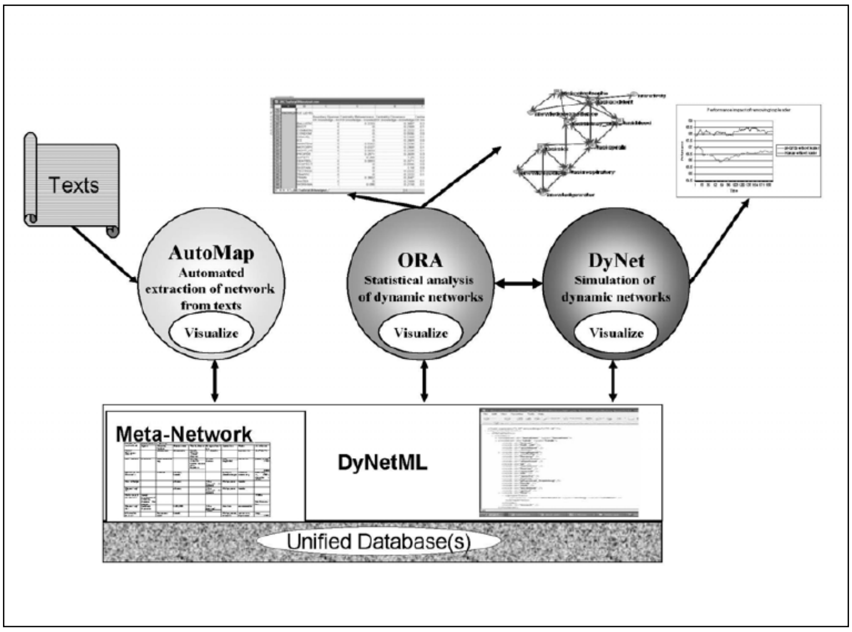


Figure 40 *– Toolchain for dynamic analysis of social networks (Carley, 2005).*

StOCNET, is an open software system to perform statistical analysis and estimation of models for the evolution of social networks according to the actor-oriented model of Snijders. This software allows to can estimate parameters for these structural forces by simulating *how* network evolved from one state into next state.

3.3 StOCNET: Software for statistical analysis of networks

At the Sunbelt XX conference, Zeggelink et al. (2000) announced the development of an open software system called StOCNET for the advanced statistical analysis of social networks.

A few years later, an update of the project was given by Huisman and Van Duijn (2002).

StOCNET is a computer program that carries out statistical estimation of models for repeated measures of social networks according to actor-oriented model of Snijders.

StOCNET does not contain procedures for the visualization of networks.

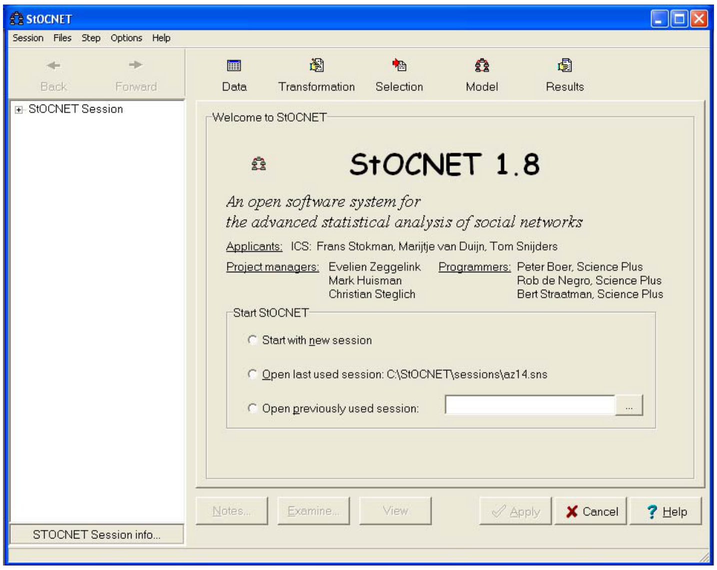


***Figure 41** – Home Screen of StOCNET.*

An analysis within StOCNET takes place within a so-called *session*, and consists of five sequential steps. In particular, five procedures allow to define data, transform data, select of subset to analyze, specify model and analyze it, and inspect of results. So, steps start with data definition and result in specified output, after which all or some steps can be repeated or skipped.

So, the first step is data definition, so specification and description of network(s) and actor attributes.

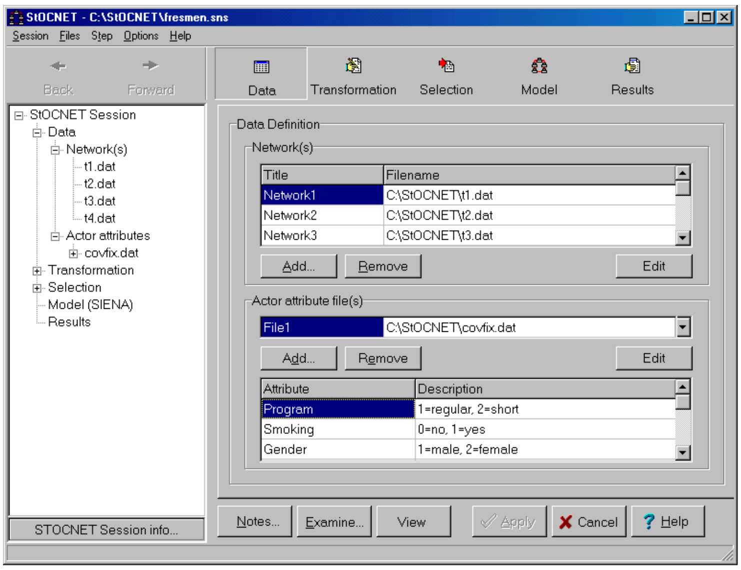


***Figure 42** – Screen of Data Session (Huisman and Van Duijn, 2003).*

Once defined data on which program will perform the calculations, it is possible to carry out some operations (transformation, selection) on data.

Then, it is necessary to choose program for data analysis; StOCNET contains five statistical modules:

- BLOCKS, for stochastical block modeling (Nowicki and Snijders, 2001);
- ULTRAS, for estimating latent transitive structures using ultrametrics (Schweinberger and Snijders, 2003);
- P2, for fitting the exponential random graph model p_2 (Van Duijn M.A.J et al., 2004);
- SIENA, for the analysis of longitudinal network data (Snijders, 2001, 2004);
- ZO for determining probability distributions of statistics of random graphs (Snijders, 1991; Molloy and Reed, 1995).

In the end, after specification parameters in the model specific user interface, and running the method, the last step is represented by the inspection of output and results from the analyses.

3.3.1 SIENA: Simulation Investigation for Empirical Network Analysis

For the aim of this doctoral thesis, *SIENA* (Simulation Investigation for Empirical Network Analysis) is chosen as analysis module.

The module SIENA allows to carry out the statistical estimation of models for the evolution of social networks according to actor-oriented model.

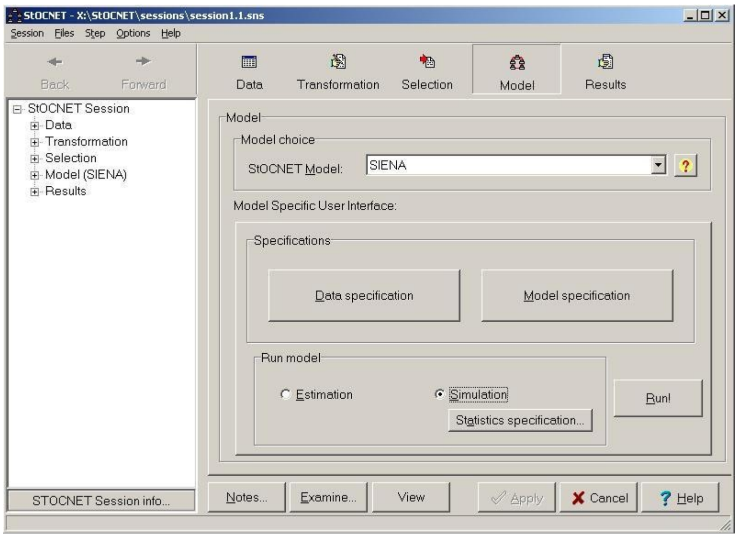


Figure 43 *– SIENA (Boer et al., 2006).*

If network evolution is studied and selection process is modeled, the dependent variable is the evolving relation network, represented by repeated measurements of network configurations (adjacency matrices actor for actor), while the characteristics of actors represent independent variables.

If influence process is considered and network evolution with the evolution of behavior of actors are analyzed, it is necessary to indicate at least one changing individual attribute as dependent variable.

Recovering what has been said about actor-oriented model, network evolution is modeled as the consequence of the choices of each actor to initiate or withdraw relations with other actors, choices driven by the aim to obtain a more rewarding configuration for the same actor (maximization of the objective function of each actor).

The data included are related to four aspects of the network and to characteristics of actors:

- The configurations of network in different moments of observation; the number of observations is at least two and for each moment of observation network is represented by an adjacency matrix *actor x actor* that gives information about actors and their relations. These matrices must be binaries, and generic element is equal to 1 if relation exists, 0 otherwise.
- Individual attributes of actors, called individual covariates; the possibility to upload constant or varying attributes (that vary over time) is provided. In the first case, there is a file with one row for each actor and one column valid for all moments of observation (in a friendship network, for example, a constant attribute is the gender of people that obviously remains the same for all observations of network). In the second case, there is one column for each interval between two successive observations and it has *column x column* the value related to moment of observation previous to interval (for instance, in a friendship network the age of actors represents an attribute that changes over time).
- Relational attributes of actors, called dyadic covariates; dyadic covariates are relations so they are represented by a matrix *actor x actor* (an example is the intensity of relations among actors).

- The changes of network composition that are times of composition change; often some actors belong to network only in certain moments of observation (for example, in a friendship network, it may happen that some actors become part of the network because they are presented by actors who were already part of the network). A way to represent this information could be that of considering the absent actors as rows and columns of zeros in adjacency matrix. A different way consists in a file of *n* lines with four numbers: the first two concern the joining: (1) the last observation moment at which actor is not yet observed and (2) the time of joining expressed as a fraction of the length of the period between two observations; the last two concern the leaving: (3) the last observation moment at which the actor is observed, and (4) the time of leaving also expressed as a fraction of the length of the period.

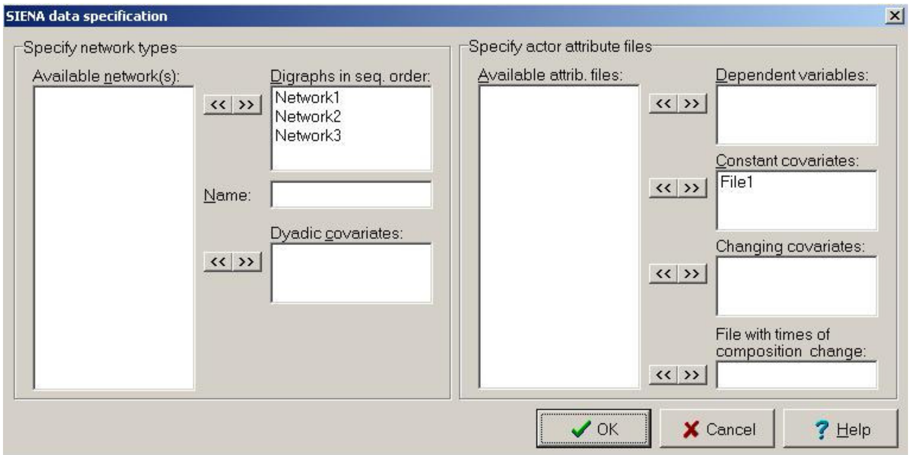


Figure 44 *– SIENA data specification.*

Once defined data, it is necessary to switch to specification of the model, (the effects that are taken into account must be indicated).

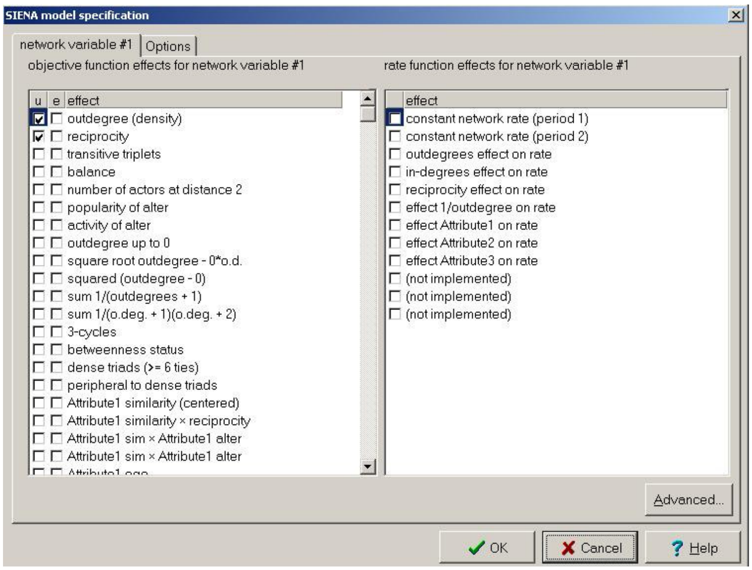


Figure 45 *– Objective and rate function effects.*

In particular, desired effects can be included in the objective function, (left side in Figure 45), and rate function (right side of Figure 45).

In left side, the effects can be specified as an evaluation effect (the first column indicated by u) or an endowment effect (the second column indicated by e). The endowment function represents part of the value of a tie that is lost when tie is broken, but that has not cost (or loss) when tie is created. The right side refers to frequency of such changes, that is the distribution function over time. Except in some cases in which is assumed a priori a difference in the frequency of changes between actors, only the rates of change of generic individual periods are inserted, so-called *rate parameters*.

It is advisable to start with a simple model that includes density and out degree effects (as default) and subsequently to complicate the model adding progressively others effects. The effects to include in the objective function may be network effects (e.g., reciprocity, transitivity), actor covariate effects (e.g., gender popularity, gender similarity), or dyadic covariate effects.

Together to selection of effects to include in the model, it is necessary to choose type of model; for non-directed networks (SIENA detects automatically when the networks all are non-directed), the model type has seven possible values:

- Forcing model: one actor takes initiative and unilaterally imposes that a tie is created or dissolved.

- Unilateral initiative and reciprocal confirmation: one actor takes initiative and proposes a new tie or dissolves an existing tie; if actor proposes a new tie, other has to confirm, otherwise tie is not created; for dissolution, confirmation is not required.
- Tie-based model: a random pair of actors is chosen, and the average change in objective function for toggling (i; j) and (j; i) is the log-odds of the probability of changing tie variable.
- Pair wise conjunctive model: a pair of actors is chosen and reconsider whether a tie will exist between them; tie will exist if both agree, it will not exist if at least one does not choose for it.
- Pair wise disjunctive (forcing) model: a pair of actors is chosen and reconsider whether a tie will exist between them; tie will exist if at least one of them chooses for the tie, it will not exist if both do not want it.
- Pair wise compensatory (additive) model: a pair of actors is chosen and reconsiders whether a tie will exist between them; this is based on the sum of their utilities for the existence of this tie.

The most fundamental option when using SIENA is estimation. The estimation is used to obtain estimates of selected effects.

As mentioned in previous paragraphs, the space of all possible evolutions of the network, so it is not realistic to carry out accurate calculations of excepted values; so, SIENA applies the methods of moments to identify the expected values that maximize similarity with observed data and approximates the solution of the equation moments through an iterative algorithm progressive (Robbins-Monro algorithm).

This simulation procedure follows three stages:

- Starting from observed values of the parameters of default effects, the parameters of other selected effects are supposed zeros.
- On this parametric basis, program simulates many casual evolutions and generates a large random sample of evolutions, by searching random parameters that come close to observed evolution.
- Finally, the distance of this sample from the actual is evaluated.

Based on the results obtained, it may decide to modify the model by adding and/or removing certain effects. The estimation procedure is the same and results obtained in the previous model become the starting parameters.

After the specification of estimation, program starts with the elaboration of the model and a window shows the progress:

- The column on left shows the parameter estimates in the different moments of elaboration;
- The middle column refers to autocorrelations of single parameters;
- In column on the right, values indicate deviations that should annul themselves.

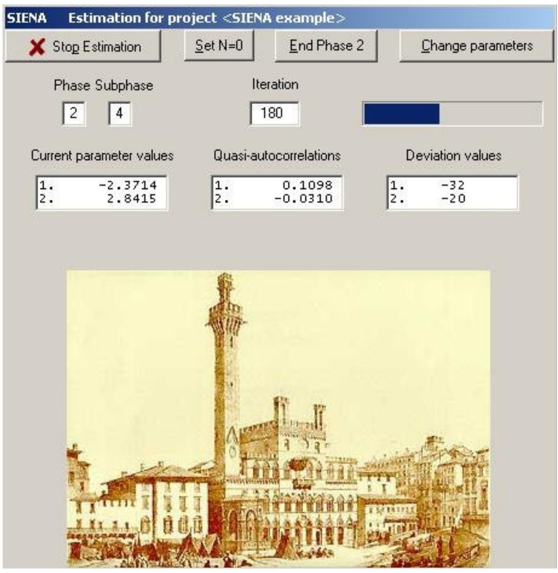


Figure 46 *– Executing.*

After executing, program produces output file that contains results. Proceeding by simulation, program takes random networks related to certain criteria. For each simulation slightly different results are obtained, as the case determines various configurations.

3.3.2 Results: descriptive statistics

After executing, program produces output file of results.

The first part of this output contains some descriptive statistics. In Figure 47, an example is shown.

```
@2
Change in networks:
-------------------

For the following statistics, missing values (if any) are not counted.

All 3 observed networks are symmetric.
Therefore, it is assumed that this is an analysis of an non-directed relation.

Network density indicators:
observation time             1       2       3
density                  0.035   0.054   0.109
average degree           1.702   2.596   5.234
number of ties              40      61     123
missing fraction          0.000   0.000   0.000

Edge changes between subsequent observations:
 periods        0 =>  0   0 =>  1   1 =>  0   1 =>  1  Distance     Missing
  1 ==>  2      1067          21         0        40        21         0 ( 0%)
  2 ==>  3      1005          62         0        61        62         0 ( 0%)
(The distances reported in the output file for conditional estimation
for the network variable refer to the total symmetric adjacency matrix,
and therefore are double the distance reported above.)

Dyad counts:
 observation    total    mutual    asymm.     null
     1.          1128       40         0      1088
     2.          1128       61         0      1067
     3.          1128      123         0      1005

Standard values for initial parameter values
-------------------------------------------------
constant network rate (period 1)                   1.7942
constant network rate (period 2)                   5.2804
degree (density)                        0.2666
(weighted average of estimates under the (trivial) independent edges model);
estimates and weights are as follows:
period              1        2
density par.est.   0.0000   0.5333
weight             0.5000   0.5000 )

Model descriptions are written to file rapporti_collaborazione_aziendale.log.

A list of objective function effects is given in file rapporti_collaborazione_aziendale.eff.
The numbers in this list can be used for specifying interaction effects (see the SIENA manual).
```

Figure 47 *– An example of results.*

The first statistics are the following:

- *Density*: the rapport between the number of present relations divided the number of potential relations;
- *Average degree*: the average number of relations per actor;
- *Number of ties*: the number of ties per each observation;
- *Missing* is the number of missing tie variables per each observation.

In the next, change statistics are calculated:

- *Changes in arcs* between subsequent observation: the number of ties that remain 0, that change from 0 to 1, from 1 to 0, and that remain 1.
- *The distance* is the total number of changes from 0 to 1 and from 1 to 0.

The second part refers to dyad and the descriptive statistics are the following:

- *Changes in dyads* between subsequent observations: indicate the number of dyads that change from one class to another; the classes are mutual (M), asymmetric (A), and null (N).

In the third part the changes in triplets are presented:

- *Changes in triplets* between subsequent observations: changes between triplets intransitive (I corresponds $x_{ij}=x_{jk}=1,\ x_{ik}=0$) or transitive (T corresponds $x_{ij}=x_{jk}=x_{ik}=1$), or other (O).

In the end, there are the values of rate parameters in each interval of time between subsequent observations.

In the third part of output file, the estimation results are shown. In Figure 48, an example is shown.

```
@2
Estimation results.
-------------------

Regular end of estimation algorithm.
Total of 1584 iteration steps.

@3
Estimates and standard errors

Rate parameters:
 0.1   Rate parameter period  1                        0.8452  (   0.1862)
 0.2   Rate parameter period  2                        2.3282  (   0.2907)

Other parameters:
 1. eval:  transitive triads                           0.4798  (   0.1402)


@3
Covariance matrices

Covariance matrix of estimates (correlations below diagonal):
     0.020

Derivative matrix of expected statistics X by parameters and
covariance/correlation matrix of X are written to file az 15.log.

Total computation time    14.62 seconds.
```

Figure 48 *– An example of result of estimation.*

From Figure 48, the rate parameters indicate the excepted number of changes of relations per time during two consecutive observations, and the estimate of selected effects. The positive value indicates that the effect plays a role in network evolution, whereas a negative value indicates that the effect does not. In the end, covariance matrix is illustrated; its values express the correlation between estimated values.

4. Collaborative networks in research

Over the course of scientific career there are opportunities for collaboration with other scientists, and wide variability in the extent to which individual scientists choose to collaborate. …There are others reasons to collaborate. First and foremost is the synergistic creativity that comes from working with others.

McCartey Christopher et al, 2012.

4.1 What is the research collaboration?

Scientific collaboration is a complex phenomenon that has been systematically studied in literature since 1960s. In this context, the first studies have been focuses primarily on search of *common* and *shared* way to define collaboration in research.

Regarding the word research, Must (2000) suggested important its features: “science is a collective, creative effort that cannot develop in *isolation*….. The fundamentals for an ample field of scientific research are *openness*, an opportunity to consult, *belief* on the research results of predecessors”. Thus, scientific activity implies the collaboration.

One speaks of research collaboration when at least two researchers decide to share their skills and knowledge to achieve a common scientific result; in practice, this scientific result translates in the production of a scientific paper.

Interactions among scientists with aim to produce a paper has for long been the essence of scientific practice (Melin and Personn, 1996), in every discipline as well as within and across geographic areas. Often, scientists talking to each other, and publishing an article, so over time the number of co-authored papers has recorded a continuous increase.

The first question linked to research collaboration concerns on *how closely* researchers have to work together for speaking of collaboration (Katz and Martin, 1997*)* and also, two or more researchers are *collaborators* or *co-authors*.

At the most *basic* level, research collaborators are scientists who work together in a project or paper *over time*; while scientists who have their names in a scientific article are defined as co-authors.

Research collaboration exists also at other levels: it can occurs between different research groups that belong of the same department, or between researchers that belong to different universities, or between sectors, or better across sectors (e.g. university and industry that is the collaboration between university scientists and scientists or professionals working in a company), and so on.

So research collaboration can occur either *between* (e.g. *inter*-national collaboration is a collaboration between scientists who work in different countries or *inter*-disciplinary collaboration involves the integration of knowledge from two or more disciplines), or *within* different levels (*intra*-department collaboration means collaboration between scientists that belong to a single department).

	Intra	Inter
Individual	–	Between individuals
Group	Between individuals in the same research group	Between groups (e.g., in the same department)
Department	Between individuals or groups in the same department	Between departments (in the same institution)
Institution	Between individuals or departments in the same institution	Between institutions
Sector	Between institutions in the same sector	Between institution in different sectors
Nation	Between institutions in the same country	Between institutions in different countries

Figure 49 *– Several levels of collaboration (Katz and Martin, 1997).*

Thus, collaborative research may be conceptualized as an *effort* done by scientists who may be from different disciplines, from different departments, either belonging to the same country or to more than one country, to same sector or to different sectors.

The light of the above considerations, research collaboration is very difficult to define, both because too many its characteristics must be considered and both because its boundaries are enough indefinites.

4.2 Research collaboration as research network

Social studies on research have a long interest in linking scientific collaboration to network structures of scientific community. In fact, especially in today's world, research activity is realized by collaboration among researchers, so it can be modeled as social network.

The idea of constructing a research collaborative is not *recent* (behind there is a long history).

Price, Garfield, Small, and Griffith represent real pioneers as they conceptualized scientific collaboration as network of scientists. These authors established important lines of activity, like the creation of bibliometrics[11] and scientometrics that are respectively two closely related approaches to measuring scientific publications and science in general.

The literature on research network is very wide and issues concerning its different aspects can be categorized into four sets (Katz and Martin, 1997). First concerns the question of how one can measure research collaboration, and, in particular, whether one can do so through the analysis of co-authored papers. A second set interests in factors that encourage the formation of research collaborations. Third set regards on the mechanisms in the formation of collaboration networks and processes leading to the observed structures. Least set investigates on the effects of collaboration on productivity of researchers.

4.2.1 How measuring scientific collaboration

The great diffusion of bibliographic databases that has made available records on scientific production, has favorite the idea to identify scientific collaboration as production of papers in common among researchers, designing research networks as *co-authorship networks* in which nodes represent authors and ties represent papers published in common by researchers.

During 1999s many studies focused on potential utility of co-authorship networks but starting from 2000s several scientists began the construction of large-scale networks.

Thus, the co-authorship network has been considered as the most *common* way to represent scientific collaboration, and *tangible* and *well-documented* form of social networks as relations among authors are documented by existent databases. For this reason, over time co-authorship networks have been studied from all points of view, and in all their aspects.

A number of studies have demonstrated that there has been large increase over time in the number of papers published by more authors in comparison to single authored papers. Newman (2001) tried that the average

[11] The bibliometrics is a discipline for quantitative evaluation of scientific literature with aim to analyze the dynamic of science.

number of co-authors per paper within computer science has been 2, and, during a period of five years, this average had increased to 4.

And then, scientometric works investigated co-authorship networks using quantitative methods, such as *co-authorship statistics*.

For instance, Newman (2004), utilizing data from three bibliographic databases, discussed the structures of three different co-authorship networks (in Figure 50); in this network, authors of several topics of research (physics, mathematics, and biology) are nodes, each topic is indicated by different shapes of nodes, and papers that they have published during 1997- 2002 represent ties.

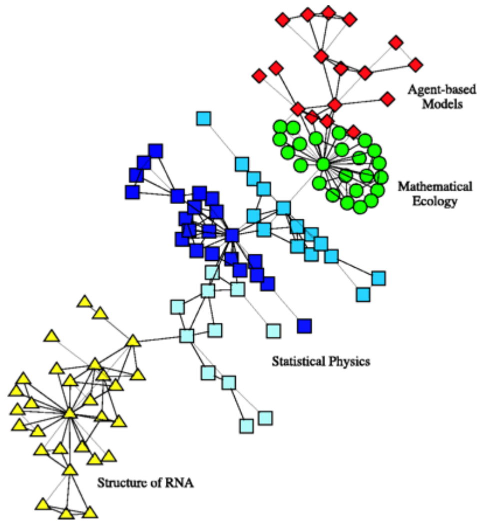


Figure 50 ***–*** *An example of co-authorship network (Newman, 2004).*

In order to highlight structural differences between networks of sub-communities corresponding to each topic research and characterize networks, Newman run an analysis on structure of obtained network, calculating statistical properties of networks, like distribution of numbers of co-authors in each of three fields studied, and many others statistics, such as the number of authors, the number of papers, the average collaboration, the average distance between authors, and other. Besides, he showed that co-authorship networks formed small worlds in which pairs of researchers were separated by a short path of intermediate acquaintances.

Another way to concept research network was through *citation networks* that are specific academic networks based on citation patterns among scientists in which nodes are papers, or journals, while links among them represent citations.

The literature on citation networks is very rich, and it can be split it according to three important ways:

- First, a previous works examined a single aspect of network, above all to rank journals in terms of their influential status; for example, Jobber and Simpson (1988) focused our work on influence of specific journals; in alternative, other works focused on more aspects treated in a independent way, and more few studies have analyzed the roles that journals play in their networks and their influence.
- Second, initially studies on citation research had focused on the study of network in one particular point in time. Recently researchers took the time dimension into account in order to investigate the dynamics aspects of citations (Hossian and Fazio, 2009).
- Third, in order to analyze the citation networks, first studies focused mainly on descriptive methodologies, which used some indicators of citation activity. Over time, the focus is moved on statistical methods because these are more adapts to investigate structure of network and its change over time.

An particular kind of citation network is co-citations, that is a network formed by links between authors established via the citation of their works in the same article; so this can mean that they are closely related to each other either because they belong to the same topic area or because their topic areas are closely connected.

In Figure 51, an example of co-citation network is showed, in which each node is identified with the name of authors who have written it and the date in which paper has been written, while the arcs represent the co-citations of paper.

To build the network, in total 5693 references were extracted and these references have been filtered of all articles that had less than six citations in the 133 papers sample.

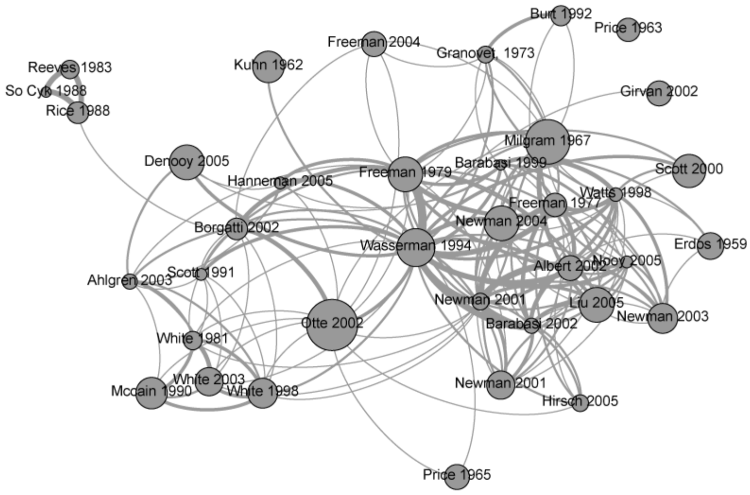


Figure 51 *– An example of co-citation network (Web).*

In definitive, co-authorship, and citation networks have been considered like a *reflection* of all academic links among researchers (Barabási et al, 2002).

In the most part of the studies scientific collaboration is simply given by co-authored in a paper and more for many years co-authorship networks have been used as *units* of collaboration.

In this way, all researchers who collaborate become co-authors. This consideration is not always true, as research collaboration does not always lead to *joint output*, such as the publication of a paper. Or more, publication of a paper is not always a result of research collaboration. Two researchers work very closely together but they decide to publish their findings in two separate papers because they operate into two different disciplinary sectors. Thus, two scientists have collaborated very intensively but, at the end, they have decided to publish two different works. On the contrary, two researchers that have not worked together, decide to link their findings to publish them in the same paper. In this case, scientists have not collaborated but they produce a joint paper. So, in the first case has sense to speaking of research collaboration, while in the second one speaking of co-authors, although often these two terms are considered as synonymous.

Melin and Persson (1996) have deepened this phenomenon and in their model reported the dependencies between collaboration and co-authorship (Figure 52), suggesting to utilize co-authorship in *conscious* way.

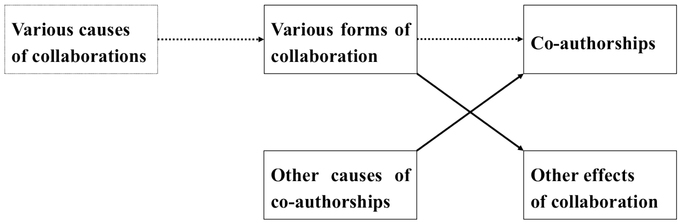


Figure 52 – *Co-authorships and its causes (Melin et al., 1996).*

Authors pointed out on many outputs, such as patents, that derived from research collaborations as well as there are many causes of co-authorships besides research collaboration, like "when research leaders demand to have their names on the articles without actually contributing to the specific work reported". Thus, two authors argued on how is risk to infer from co-authorship to collaboration and, more, on impossibility to identify the real reasons behind co-authored. In order to mitigate this risk, they suggested triangulating co-authorship with other indicators, to consider considerable periods of time, and to accept a certain level of uncertainty because in the most cases research collaborations lead to co-authorships.

Melin and Persson conducted an analysis limited to Umeå University, a small-scale survey, and concluded that only five percent of authors claimed to have collaborations that have not become co-authorship.

This result clearly contradicts those found by Laudel (2001b, 2002) in his work on interdisciplinary research collaboration. He investigated how a specific institution, so-called Collaborative Research Centre[12] (CRC), promotes interdisciplinary collaboration. He explored research collaboration undertaken between 57 German research groups in two CRC in an interdisciplinary field. Laudel combined *quantitative* and *qualitative* methods in order to identify the types of research collaborations between scientists of CRC. Qualitative method consisted of interviews with scientists (research groups and at least one group member, postdoctoral researcher or PhD student) about content and reward of their collaborations, while co-authorship and acknowledgement were used as an additional indicator. On the basis of these interviews with scientists and of kinds of their contributions that is depending on how a scientist is rewarded (as co-authorship or cited in

[12] Collaborative Research Centres are research networks that receive additional funding with aim to overcome the disciplinary and organizational barriers.

acknowledgements or with nothing at all), Laudel identified six kinds of research collaborations. He showed that about half of these collaborations are invisible in formal channels because they were not rewarded as co-authors or cited in acknowledgements, and that one third of collaborations were rewarded only by acknowledgements and not appeared as co-authors; in particular, all collaboration characterized by formal division of work had brought to co-authorship, while collaborations in which there were exchange of information, transfer of know-how and informal ideas were seldom recognized in a joint publication.

4.2.2 Factors encouraging the research collaboration

Numerous contributions focus on the study of elements that encourage research collaboration.

In his work, Beaver (2001) tried to investigate about the motivations according to which people collaborate in research, and, through a survey administered to his colleagues, he proposed 18 reasons (in Table 3):

1)	Access to expertise.
2)	Access to equipment, resources, or stuff one doesn't have.
3)	Improve access funds.
4)	To obtain prestige or visibility; for professional advancement.
5)	Efficiency: multiplies hands and minds; easier to learn the tacit knowledge that goes with a technique.
6)	To make progress more rapidly.
7)	To tackle "bigger" problems (more important, more comprehensive, more difficult, global).
8)	To enhance productivity.
9)	To get to know people, to create a network, like an "invisible college".
10)	To retool, learn new skills or techniques, usually to break into a new field, subfield, or problem.
11)	To satisfy curiosity, intellectual interest.
12)	To share the excitement of an area with other people.
13)	To find flaws more efficiently, reduce errors and mistakes.
14)	To keep one more focused on research, because others are counting on one to do so.
15)	To reduce isolation, and to recharge one's energy and excitement.
16)	To educate (a student, graduate student, or, oneself).
17)	To advance knowledge and learning.
18)	For fun, amusement, and pleasure.

Table 3 – *The purposes for which people collaborate in research (Beaver, 2001).*

The list of possible contributions factors is almost endless. Probably, factors identified by Beaver may occur frequently than others, but collaboration is an intrinsically social process as it is influenced by individual characteristics.

Some studies supported the idea that collaboration depends on the nature of the research. In fact, it is generally accepted that experimentalists tend to collaborate more than theoreticians, as, for the first, the use of large instrumentation is required.

Collaboration may also depend on how basic and applied is the research; applicative research tends to be more interdisciplinary because it requires a wide range of skills.

Then, the choice to collaborate also depends on characteristics of *discipline* that characterizes collaboration. For example, in discipline as sociology, sociologists are more likely to be collaborative than philosophers.

Another example is represented by some disciplines that require a team effort in order to conduct experiments. In this case the collaboration is driven by infrastructural needs.

In recent years, numerous political initiatives (through financing) have been launched to improve collaboration among research groups and international collaborations.

Besides, globalization has conceptually led the increase of the geographical diversity of collaborators, be they individuals, departments, or universities, supported by *web tools* (Internet, Email, Skype, and many others). This development of web tools and mental openness has facilitated communication, exchange of information, to inter and intra levels, and so, every kind of collaboration mentioned before, has been made possible. For instance, the advent of email had begun to increase diversity in geographical locations. Physical location is no longer a *barrier* to the free and easy exchange of information (Beaver, 2001), and it pushes scientists to collaborate at national and international levels.

4.2.3 Mechanisms driving co-authorship networks

On the study of representations of processes driving networks, several models have been proposed over years.

The *small-world* model inspired the work of de Sola and Kochen (1978), who partially formalized Milgram's work that represents one of the first and famous empirical studies on the structure of social networks. Milgram expressed the simple idea that any two individuals, selected randomly, are connected by a path of small number of intermediates. The experiment, conducted by him, showed that this number is about 6 and this notation became popular as *Six Degrees of Separation*.

The small world model, revisited by Wattz and Strogatz (1998), is a random graph generation model that produces graphs with small world properties. Intuitively, a small world network is any network where the level of *local clustering* (one's collaborators are also collaborators with each other) is high, but the average number of steps between actors is small.

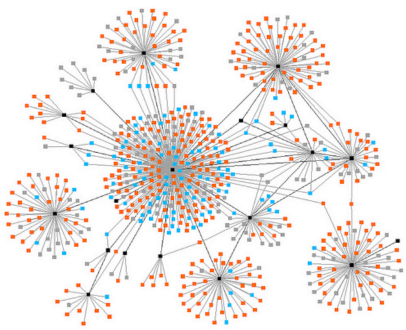


Figure 53 *– Example of small world structure.*

These properties were later used to identify small-world structure in measured networks defined on co-authorship of scientific publications (Newman 2001[13]; Moody 2004).

Formal modeling of cumulative advantage in terms of *preferential attachment* was brought by Barabási and Albert (1999) to study of social network; this process is based on the principle that *the rich get richer* and was originally proposed by Yule (1925). Barabási and Albert (1999)

[13] Newman (2001) shows that co-authorship networks form small worlds in which pairs of scientists are separated by only a short path of intermediate acquaintances.

investigated on a common property of many large networks in which degrees of nodes follow a *power-law distribution;* this feature was found to be the consequence of two mechanism: the growth of network is continuously and new nodes tend to connect to nodes characterized by a high number of links. The model was accepted and applied to structure of co-authorship networks (Barabási et al., 2002; Moody, 2004; Kronegger et al, 2011).

The application of small world and preferential attachment models reduces the generation of co-authorship network to a single mechanism ignoring the social context in which scientists work and their characteristics.

In other studies, authors found several features which lead to two or more researchers to collaborate, such as similar research topics (Kunh, 1996; Moody, 2004), while others authors tried that the collaboration is driven by departmental and institutional affiliation (Ziman, 1994).

Until recently, there were not methods for modeling the dynamics that drive the change of networks and actor attributes and organizational contexts were not seen as factors that influence change in networks.

This one has been changed by the development of the stochastic models (Snijders, 2001, 2005; Snijders et al. 2007; Steglich et al. 2010) that allows estimating complex models in which the change of network is driven by micro-mechanisms that depend on network and actor's characteristics.

4.2.4 Impact of co-authorship on researchers productivity

Another great part of literature concerns the effects of collaboration on productivity and on the impact of joint research.

Lotka represents the pioneer in the productivity of researchers and the findings of his work show that the number of authors producing *n* papers is proportional to $1/n^2$ (Lotka, 1926).

In the wake of Lotka's work (1926), many scientists investigated on tendency of authors to collaborate with prolific authors. Many studies confirmed that high productivity, in terms of published works, is correlated with high levels of collaboration. So, collaboration with high productivity scientists tends to increase individual productivity, and authors at all levels of productivity tend to collaborate more with highly productive authors than lower productivity authors.

Another finding besides enhancing personal productivity in research shows as the number of authors in an article is connected to acceptance of it for publication (Gordon, 1980). According to Gordon, a paper with multiple authored has the high degree of technical competence so it has more likely to be published.

Others studies have shown advantages of collaboration in terms of co-authored. For instance, Lawani (1988) argued that the number of co-authors is more correlated with impact of a paper; he demonstrated that the number of authors per paper increases, the impact, in terms of *earn* citations, also increases.

On the theme about the impact that the collaborations cause on performance of scientists, the current works have examined link between the extent of internalization of scientific paper and performance of authors of this paper.

An example is the paper of Abramo, D’angelo, and Solazzi (2011) who examined the international collaborations among Italian researchers about 26,000, of 82 different universities during years 2001-2005 and they confronted them with individual performance of each researcher. Authors showed that research productivity has positive effects on the degree of international collaboration of researchers because the increase of scientific outputs is correlated with increase of cross-national publications.

4.3 Evolution of research collaborative networks

All social and organizational networks, also research collaboration networks, evolve over time.

Xi and Tang (2004) showed a network organization via a case study; they investigated on network structure of an electric technology consulting company located in China, composed by multiple teams and ties among these teams; in particular, this network is regarded as a graph in which nodes represent those teams and members within teams, and arcs indicate links among them. According to Xi (2004), a network organization presents some characteristics (shown in Figure 54):

- Every node is one which is active and keeps moving;
- The network is formed by nodes and relations among them;
- In the structure of network, there exist some components which explicitly hierarchical while others are implicitly hierarchical;
- There are differences between the status and roles of nodes in network.

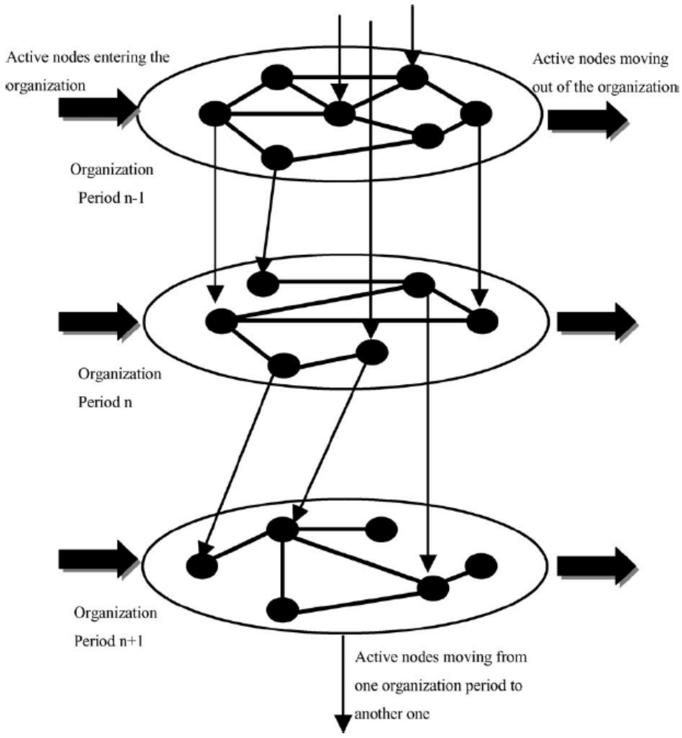


Figure 54 – *Scheme of dynamic connected network organization (Xi and Tang, 2004).*

These characteristics indicated that the structure of network is dynamic and organizational design is evolving over time.

This dynamic prospective is necessary when the aim is that of the understanding and catch network evolution caused by change of dyadic relationships and individual behaviours over time.

Each kind of relation is characterized by a different type and time of change; for instance, relationship among firms that take a part to a same project has contractual life, while friendship between two young people can be more instable, it can be gone on for years or finish immediately (Vignoli, 2008).

The changes in structure of relational network can be caused by dependencies that characterize network (introduced in previous chapter), such as reciprocity, transitivity, or that depend on individual characteristics of actors, such as similarity. In particular, the changes of relational structure are the consequence of selection process when the behavioural characteristics of the people determine a relational choice (e.g. Amelia smokes and in the next she becomes friend of Tom because also him smokes), or of the influence process when relation between two persons influences on their behavioural choices (e.g. Amelia smokes because her friends smoke).

Collaboration network represents a prototype of *evolving* networks because each researcher has the opportunity to collaborate with another researcher belonging to the same department or not, the same university or not, working in the same disciplinary or not.

The propensity of researcher to collaborate with another one is defined as the *willingness* of him to initiate collaboration with the given researcher. When choosing a collaborator, a scientist is influenced by several factors including economic dependence, mutual intellectual influence, social influence, mutual benefit, and trust.

Also, co-authorship and citation networks are characterized by continuous evolution because they are constantly expanded by addition of new papers and accordingly new links and authors (Barabási et al., 2002) or by new citations received, respectively.

In a research collaboration network could happen that a relation is created, eliminated or remained constant over time.

5. DIEG case study

Ties present in my department over time.

Raffaella Cicala.

5.1 Research questions

Social theory suggests that social networks generally develop not *randomly*. In some cases, people select and then decide to create links with key individuals having characteristics that make them attractive (i.e. to link with whom has many friends), or people form ties with individuals having similar characteristics to its (i.e. the friend of my friend is my friend). Over time a relational network is the dynamic result of what happens among people that can decide to create, eliminate or maintain ties with others.

Research networks represent an interesting example of dynamic social networks as they are characterized by spontaneous and not imposed relations among researchers.

In the past, the idea of analyzing research collaboration using bibliographic data was very diffuse since the availability of large bibliographic databases made information about the authors and their publications accessible to everyone, and so it was relativity easy to construct research networks with high reliability and large size. Thus, the writing of a paper was seen as research collaboration among authors.

Over time researchers have begun to investigate on other methods to measure scientific collaboration as not always the joint writing of a paper can be really research collaboration or, at contrary, it may happen that researchers who work closely have never published a paper together.

Another aspect on which the literature on research collaboration has focused on the *patterns* by which researchers choose to engage in a partnership or join a particular group.

Over the course of their scientific career researchers have opportunities to know other researchers and to decide if and with whom to collaborate. In this process several factors come into play and a wide variability of reasons that lead to a scientist to make this choice exists.

There are cultural differences between different disciplines regarding whether researchers collaborate with others. For instance, in the chemical field the creation of research team is strongly required to conduct complex laboratories experiments.

In addition to these reasons that regard infrastructural needs, it is well known that the synergistic creativity comes from working with others. In fact, if more researchers come from different sectors they can merge their different knowledge to produce a new knowledge. It is then well known that the splitting of work among many authors can generate increasing of scientific productivity.

On the other hand, reasons according to which researchers do not choose to publish co-authored papers occur. In fact, in some scientific fields co-authored is discouraged assigning more value to single author publications.

Once chosen whether to collaborate, researchers can choose with whom to collaborate. This choice depends on particular elements such as they work in the same disciplinary sector or if they are good level researchers. Particular attributes can play important roles in the choices of links. The similarity of attributes can be a decisive determinant in the mechanism of preference of research partners.

In Figure 61, an example of this similarity effect has been presented. Circles and square represent actors of the network with different attributes. If similarity effect plays a positive role in evolution of network, actor *i* will prefer actor *k* that is similar to it (both actors are represented by circles) to actor *j* that is not similar (*j* is a square).

Figure 55 – *Representation of similar effect (van de Bunt and Groenewegen, 2007).*

The status of a researcher represents another characteristic that can influence the choice to collaborate. So researchers could show their interest to create collaborations with high status researchers (i.e. a researcher with high status means that has many collaborators, it is popular) over those with a relatively low amount of status.

In Figure 62, the status effect has been shown: actor *i* can choose between actor *j* (with no collaborators) and actor *k* (with three collaborators). If the status effect is operating actor *i* will prefer actor *k* over actor *j.*

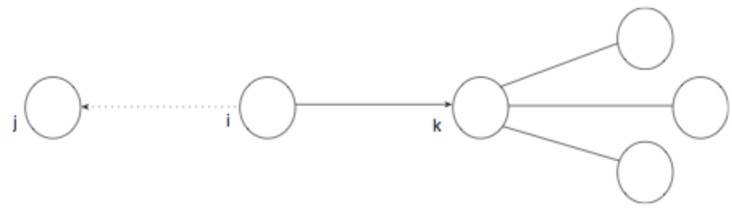


Figure 56 – *Representation of status effect (van de Bunt and Groenewegen, 2007).*

Taking into account the previous considerations, in the following, referring to a certain research unit, three question marks have to be considered:

- What are the patterns by which researchers choose to engage in a partnership and what are the elements by which a researcher chooses a research collaborator?
- How do these ties evolve over time and what influences them?
- How can be a research network represented and does it correct to represent it through co-authorship networks?

5.2 The unit of analysis

Researchers of DIEG[14] (Dipartimento di Ingegneria Economico-Gestionale - *Business and Management Engineering Department*) form the unit of analysis considered. The DIEG components (Professors, PhD students, Assistant professor) are, hence, actors of network (nodes of graph) and collaborations among them and with other external to DIEG researchers, are ties (edges of graph).

On the basis of uncertainties arising from literature on the question of how can measure research collaboration, and mostly whether it is correct to see research collaborative networks only as co-authorship, a double meaning has been assigned to research collaboration: on the one hand, scientific production is taken as an expression of the existence of a tie among authors

[14] In appendix D the description of DIEG is shown.

and, therefore, it is seen as research collaboration between them; on the other hand, a single paper does not attest research collaboration among their authors but collaboration exists if it lasts over time through the production of other papers.

So two kinds of network have been identified: co-authorship network that includes a set of authors and ties among them that represent the coauthored papers; collaborative network that includes a set of authors and ties represent the degree of collaboration among them.

Actors considered in case study are overall 76, including both members of the DIEG and who, belonging to other organizations, has collaborated with them.

The experiment has been conducted over 11-years period (from 2001 to 2011), a period characterized by the entry and exit from the department of some units.

The members of DIEG are characterized by different levels: professors (Full and Associate), Assistant professors, and PhD students. Among PhD students, only who has written one or more papers with professors or Assistant professors of DIEG has been considered.

Table 6 indicates the number of members (per career level and gender) and external researchers (a single category) in the time.

Measures	2001 *(n=17)*	2002 *(n=25)*	2003 *(n=28)*	2004 *(n=29)*	2005 *(n=41)*	2006 *(n=44)*	2007 *(n=48)*	2008 *(n=49)*	2009 *(n=60)*	2010 *(n=64)*	2011 *(n=76)*
Gender (female)	15 (2)	20 (5)	23 (5)	24 (5)	29 (12)	32 (12)	36 (12)	37 (12)	47 (13)	49 (15)	55 (21)
Rank											
Professor (Full and Associate)	6	6	8	8	8	10	10	10	10	10	10
Assistant professor	3 (1)	3 (1)	3 (1)	4 (1)	4 (4)	3 (5)	3 (5)	3 (5)	3 (5)	3 (5)	3 (5)
PhD student	0	(1)	(1)	(1)	0	1	1	1	(1)	(2)	(2)
Other roles*	1	1	1	1	1	1	1	1	1	1	1
External	5 (1)	10 (3)	11 (3)	11 (3)	19 (5)	21 (4)	24 (4)	25 (4)	36 (4)	38 (5)	44 (11)

*One member of DIEG is a researcher came from CNR (National Research Council) that is a public research organization.

Table 4 *– Description of DIEG department characteristics.*

To obtain the configuration of the department (people belonging to the DIEG and his/her career level) in each observation time, the official website[15] of the MIUR (Ministry of Education, University and Research) has been used.

DIEG's researchers have been described by different attributes, some constant and some varying over time: disciplinary sector, and institutional affiliation (internal or external to department) have been considered constant; professional rank, and scientific production have been obviously considered changing in the period of observation.

It is important to underline that the need to evaluate individual scientific research activity of people working in the universities through quantitative tools was born in the 50s. The Conference of Italian University Rectors (CRUI) adopted the *impact factor* as scientific evaluation tool. The impact factor is the best known and certainly the most discussed index in scientific evaluation. In reality, it was originally developed to measure the citation impact of journals, but after it is currently used for evaluating researchers despite the many problems that this use involves. It is identified by calculating the number of citations that the articles published in a specific journal received in the previous two years (or even just the previous year) and dividing the figure for the total number of articles published in the same journal in the two years under consideration. It is a purely quantitative tool and for this is characterized by numerous limits.

In order to overcome the limits of impact factor, in 2005 Jorge proposed a new index, called *H-index*, that attempts to measure both productivity of a researcher and the impact of his/her published works. It is defined as follows: a researcher has index h if h of his/her N_p papers have at least h citations each, and the others ($N_p - h$) papers have no more than h citations each.

[15] http://cercauniversita.cineca.it/php5/docenti/cerca.php?SESSION= is the link of MIUR in which you can enter the person's name and the year in which you are interested, to get the role and the department affiliation of the person.

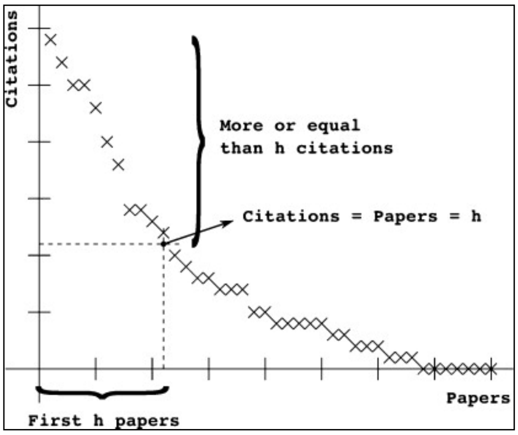


Figure 57 – *An example of H-index calculation.*

In the case study, it has been used the H-index for evaluating the scientific production of each researcher that has been considered as changing individual attribute along the time.

Information on actors and their papers (for each paper: year of publication, title, names of co-authors; for the researcher the H-index) have been obtained by well-known database *Scopus*, official source for Italian VTR (National Triennial Evaluation of Research). In dependence on Scopus characteristics, only publications on international journals have been taken into account.

5.3 Hypotheses assumed

On the unit of analysis presented, the following hypotheses have been assumed:

1. Co-authorship ties are permanent so they cannot be eliminated.
2. Collaborative tie exists when two researchers have published almost *two* papers and the time interval between the publications of these two papers must not exceed *5* years.
3. Papers published before the first observation period, have *not* been taken in consideration.
4. In measures of networks calculation, some *isolated* researchers have not been considered.

First hypothesis derives from assumption according to which once that a co-authorship tie (two authors write together a paper) has been formed, it exists for all successive observations and, thus, cannot be eliminated.

It is assumed that a collaborative tie between two researchers exists when (i) in observed period they are co-authors of least two papers; (ii) the time interval that passes between the publications of these two papers is less than 5 years. This second situation is always verified for the unit research along a period of observation.

From two first hypotheses, two types of ties *behave* according to a different way: in co-authorship network, ties once established can not be eliminated (i.e. when two or more authors write a paper together, the ties that are created between them can't be eliminated over time) and it remains in all successive observations; in collaborative network, on the contrary, ties can be eliminated (i.e. the ties after an interval of time in which authors do not publish together any papers, their collaboration is considered as exhausted). Thus, in the first case over time ties can be only increased in the number and in the strength, while in this second case ties can be created or eliminated or increased their strength over time.

Many researchers that compose the unit of analysis before the first observation just worked in the university field and, thus, some of them had already published articles. The second hypothesis is based on assumption to consider only papers that authors have published since 2001 (first year of observation). The H-index assigned to each researcher corresponds to real value that this researcher had obtained by publication of all his/her papers also from those published before 2001.

Finally, the third hypothesis regards the elimination of a number of researchers that are isolated during all study period in the sense that they have not published papers considered in Scopus. In fact, there is a part of researchers that since they come into play until the last observation are not involved in any tie with other, internal or external, components. This group is composed by researchers come from disciplinary sectors that are evaluated according to different criteria to those provided for scientific disciplines.

In many sectors of Engineering, the criterion of the evaluation of scientific productivity is called *bibliometric*. The bibliometric evaluation based on the publications of works on international journals is carried by different indicators that are recognized by database that receive a general consensus at the international level, such as Scopus, and that are validated by ANVUR[16]. Then, there are other disciplinary sectors of Engineering, called no bibliometric, for which the evaluation of productivity is performed in a different way.

In dependence of choice of using Scopus as a source of productivity information, some disciplinary sectors that are active in the department appear to be isolated and are not consider in the analysis.[17] Besides, the disciplinary sectors of this group are distant from other disciplinary sectors of components of unit analysis so it is justified the fact that they have not written any article with them.

[16] ANVUR is the Agency for the Evaluation of the University and Research (ANVUR) supervises the national public system of quality assessment of universities and research institutes

[17] Note that there are no papers written by researchers of isolated group with people of active group.

6. The Case Study: co-authorship networks

A co-authorship network is a social network consisting of a collection of researchers in which a link between two researchers is established by their co-authorship of one or more scientific papers.

Raffaella Cicala.

6.1 Methodology structure

Relational networks among DIEG's members and among them and external people from 2001 until 2011 have been constructed. As indicated, papers that researchers have published along time determinate edges of networks.

A sequence of snapshots, one for each observation period, by Pajek software has been generated; each snapshot allows highlighting new actors and new ties that characterize the network.

The first level of analysis consists on application of SNA techniques, so built networks have been analyzed in a *static* way; in fact, SNA offers some measures that yield aggregation information about whole network like density, average degree, indices of centralization, cliques, as well as information about position of a single researcher within network.

The trends of static measures over time have been drawn.

Then, using longitudinal data related to different observations, networks have been analyzed in a *dynamic* way: the actor-oriented model, proposed by Snijders (1996) has been adopted and SIENA software (Snijders, 2007) in which Snijders methodology has been used. Snijders's model allows representing changing networks over time as the result of actor's relational choices that decide to create or eliminated or no change their ties within network. Relational choices are defined determining the probability with which different choices can occur and identifying its preferential structure that is specifying *when* and *what* changes occur within of the network. As explained in chapter 2, actor's choices are modeled by an objective function, which can be interpreted as a measure of how the current network state is convenient for a single actor. Besides relational choices of actors, the model takes into account also the frequency of changes. This parameter is modeled by a rate function, which indicates the frequency with which actors get the opportunity to change their relational outgoing ties between two subsequent observations. The rate function of the network depends on rate generic of change and possibly by characteristics of actors and/or by their position within the network.

At the end, the findings have been interpreted for explaining the mechanisms that drive the network evolution.

6.2 The network construction

The networks have been drowning by software Pajek that allows to *freeze* the positions of nodes of the network in different observation times. In fact, by Pajek's Temporal Network function, a network is generated for each defined period, keeping fixed in the draft the position of nodes corresponding to different actors.

Pajek includes several types of network layout, and it is possible to choose that is the most adapt to personal needs: the algorithm of Kamada-Kawai[18], used especially for not very large network, has been chosen.

The features of Pajek allow obtaining more immediate visual interpretation of the network by a characterization of nodes assigning to them different shapes, colors, and sizes on basis of their attributes. In the case of analysis, nodes have been characterized in this way:

- Shape represents the professional rank;
- Color represents his/her disciplinary sector and institutional affiliation (white for externals to department);
- Size has been scaled by his/her h-index.

In the context of co-authorship network, tie strength is proportional to the number of common papers published by authors.

In Figure 64, shapes, color, and size are shown like example.

[18] The Kamada-Kawai layout algorithm is a force-directed layout algorithm that tries to place nodes with a distance corresponding to their graph theoretic distance between nodes (that is defined as length of shortest path between them) so it increases the readability, allowing the researcher to perceive the structures inherent in the network.

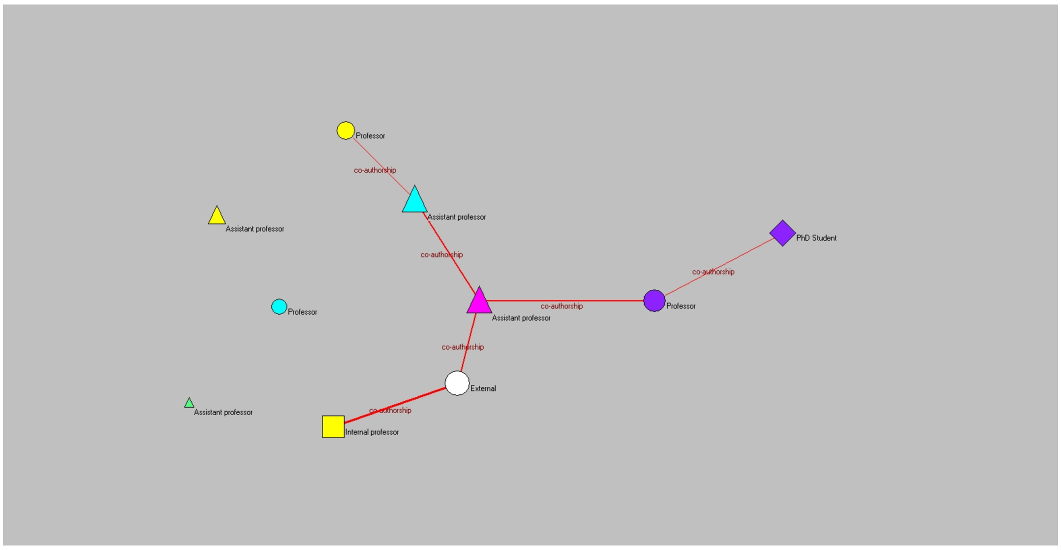


Figure 58 – *An example of co-authorship network.*

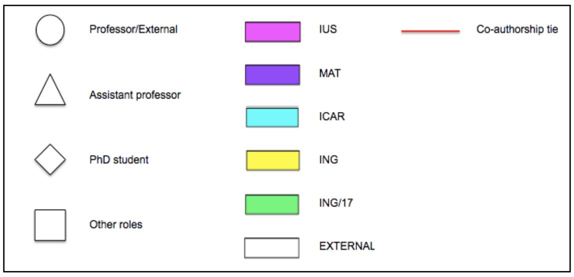


Figure 59 – *Legend of example in Figure 9.*

6.3 Co-authorship networks over study period

In Figure 66, research network of DIEG corresponding to 2011 are shown. The color of nodes indicates disciplinary afference, while red ties represent writing of a scientific paper. External authors are represented by white color because their disciplinary sectors are not considered.

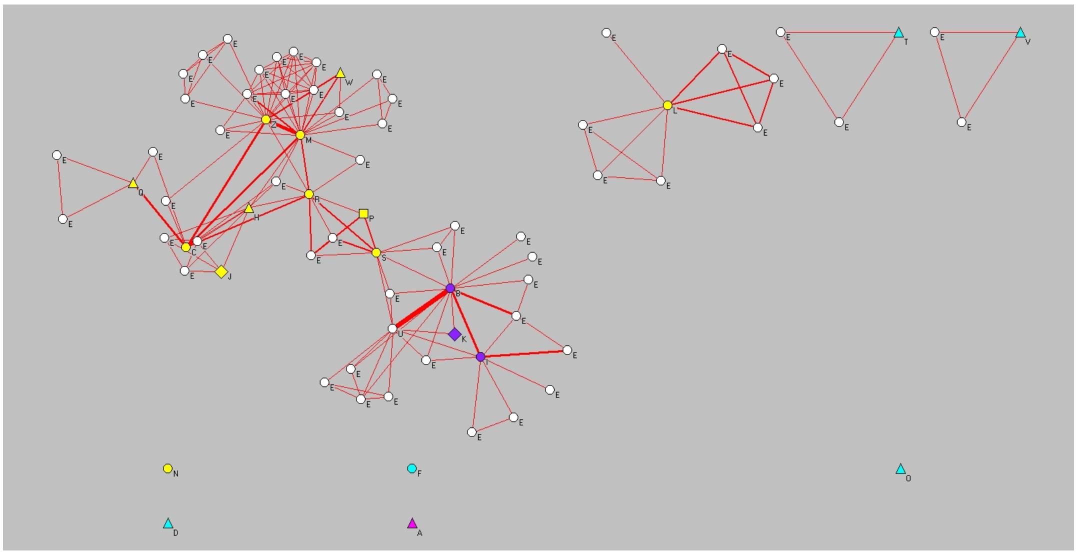


Figure 60 – *Co-authorship network in observation interval (2001-2011).*

In this last observation, 76 authors are considered, and co-authorship ties among them are 156 (network does not include external - external ties). The authors that belong to DIEG were divided into 5 disciplinary sectors, the rest are external to department. Besides, the network is composed by 12 Professors (15%), 8 Assistant Professors (11%), 2 PhD Students (3%), 1 researcher operating in other role (1%), and 53 external authors (70%).

Table 7 shows the cumulative pattern of papers (see hypothesis 1), authors, and papers with one author.

Year	Papers	Authors	Paper one author
2001	5	17	1
2002	13	25	4
2003	15	28	5
2004	16	29	6
2005	21	41	6
2006	24	44	6
2007	30	48	8
2008	36	49	10
2009	45	60	12
2010	54	64	16
2011	64	76	18

Table 5 – *Cumulative pattern of papers and authors over time.*

The number of papers, authors, and papers with one author increases gradually; in particular, the number of papers with one author grows less quickly than other two. The trends of the number of papers and authors are shown (Figure 67).

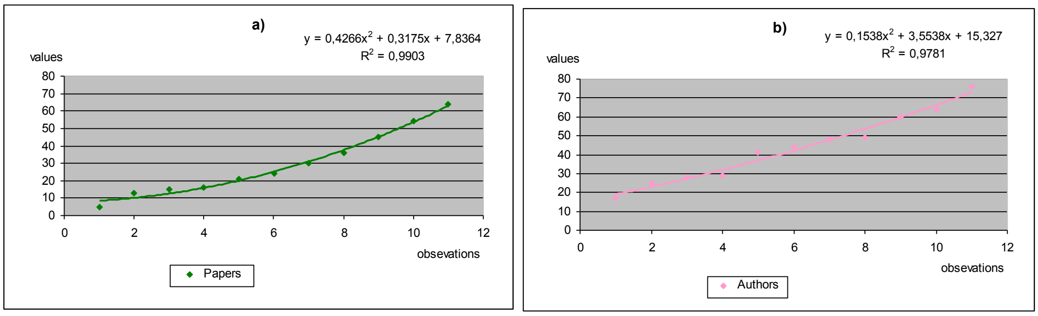


Figure 61 – *Cumulative distribution of papers (a)) and authors (b)) over time.*

In the following, the sequence of network configurations in the time is displayed[19].

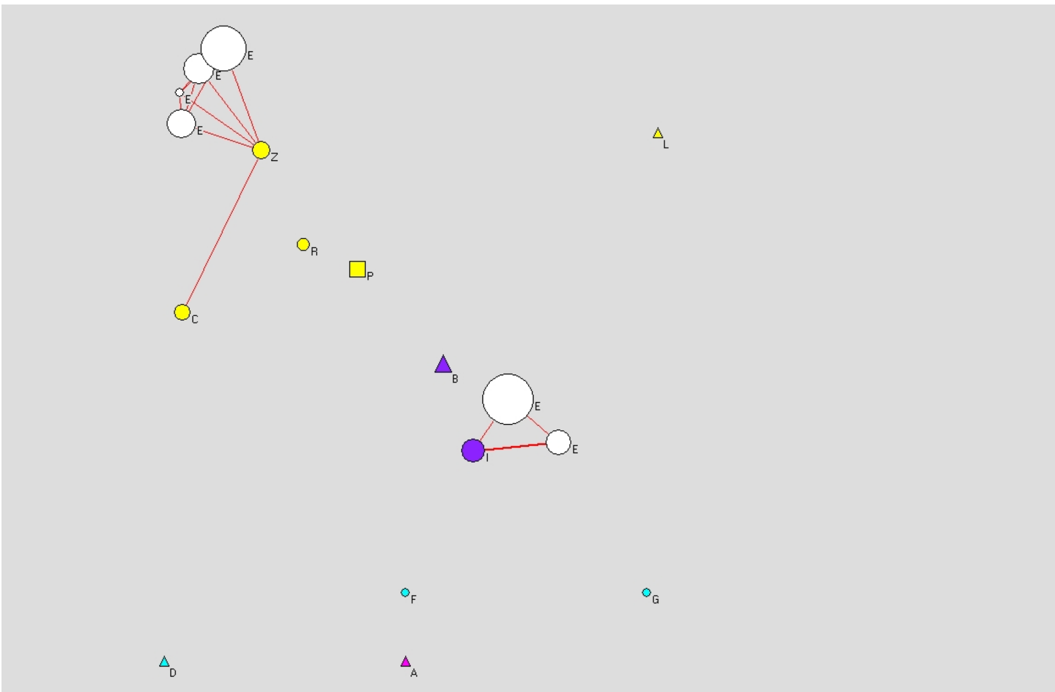


Figure 62 – *Co- authorship network in 2001 (first observation).*

[19] In each network, the size of node represents H-index of authors in considered period.

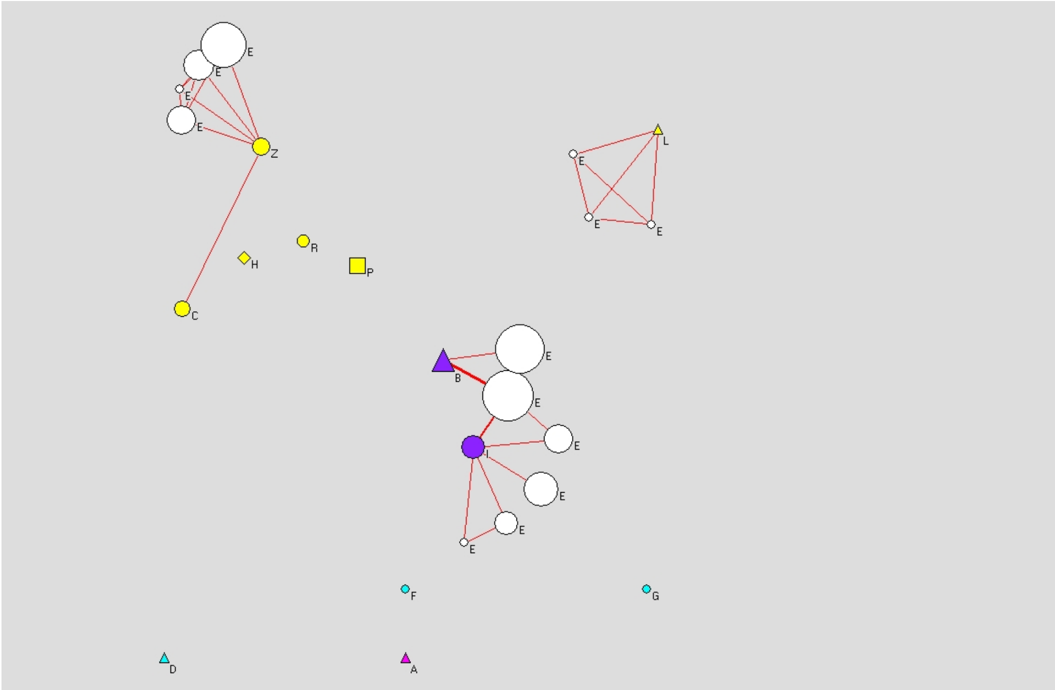


Figure 63 – *Co- authorship network in 2002 (second observation).*

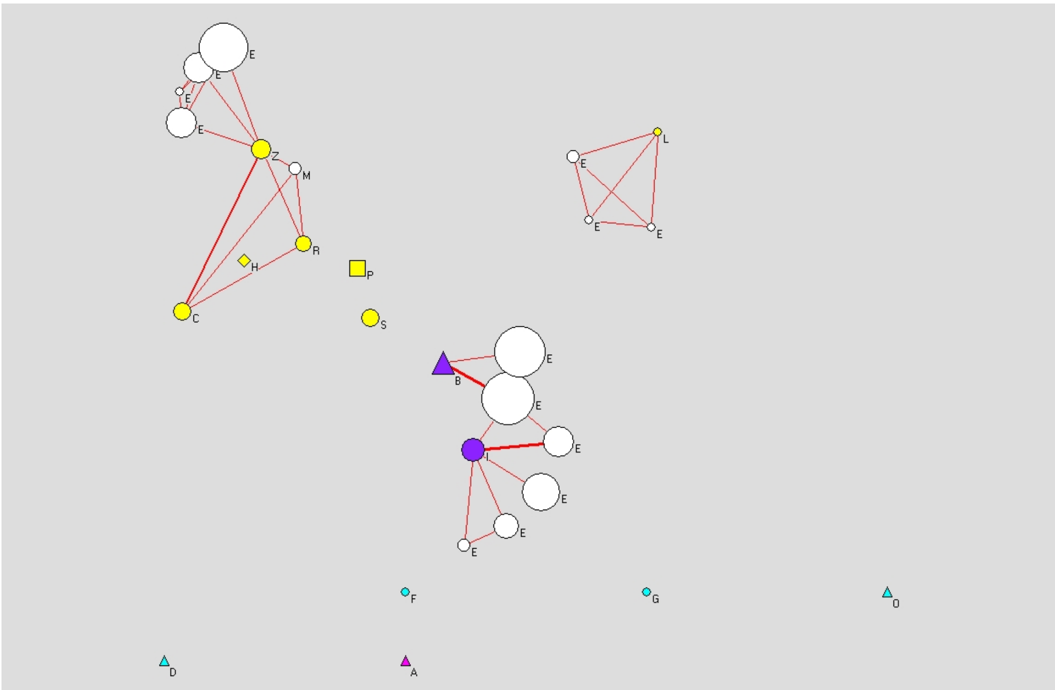


Figure 64 – *Co- authorship network in 2003 (third observation).*

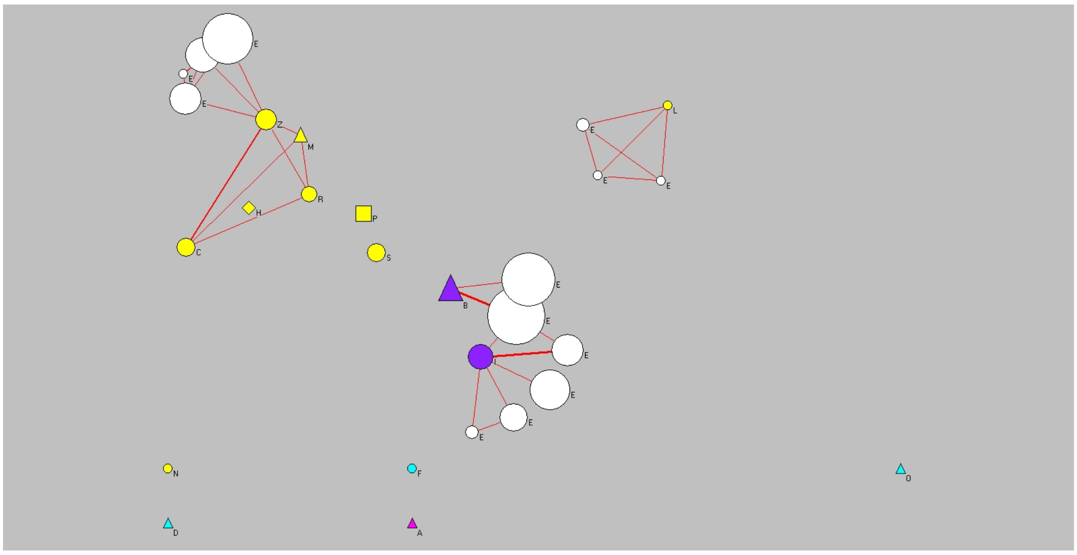


Figure 65 – *Co- authorship network in 2004 (fourth observation).*

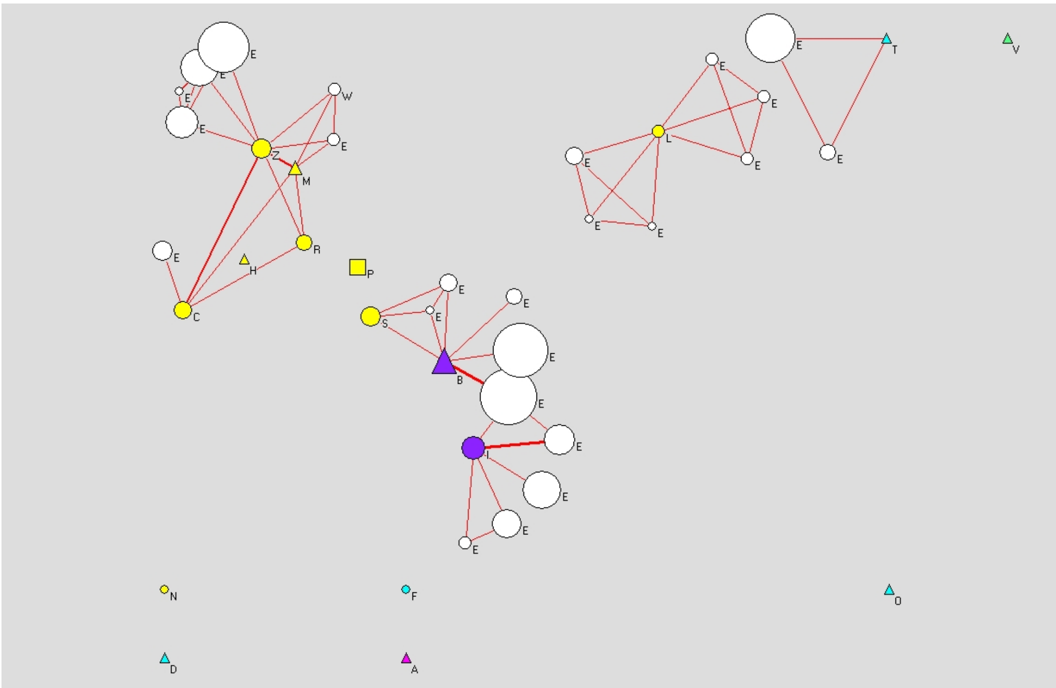


Figure 66 – *Co- authorship network in 2005 (fifth observation).*

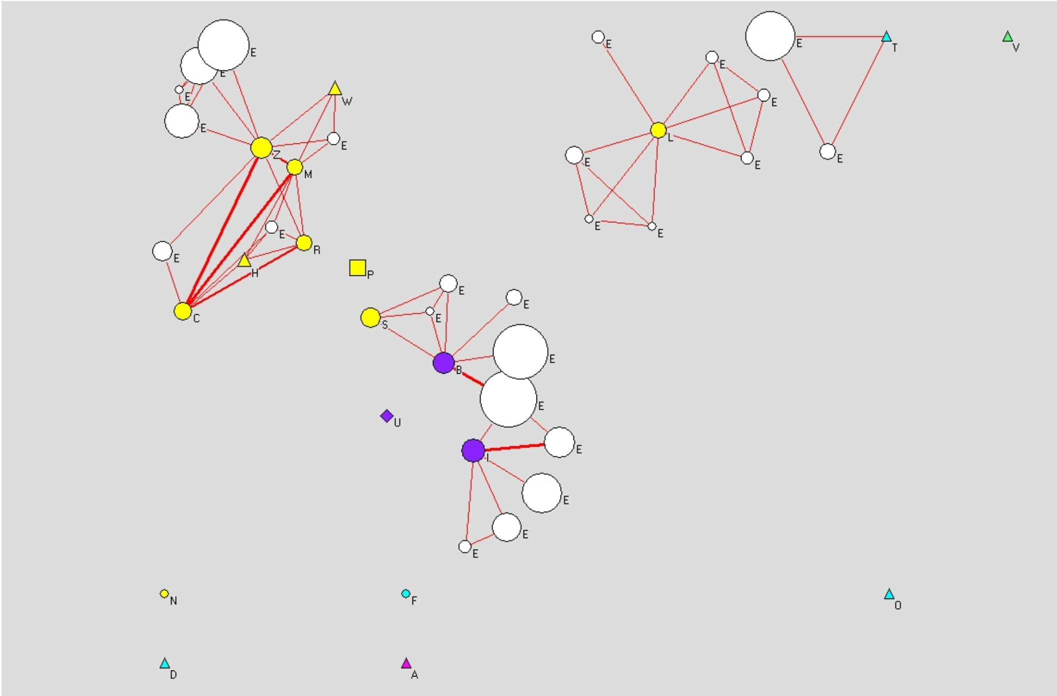


Figure 67 – *Co- authorship network in 2006 (sixth observation).*

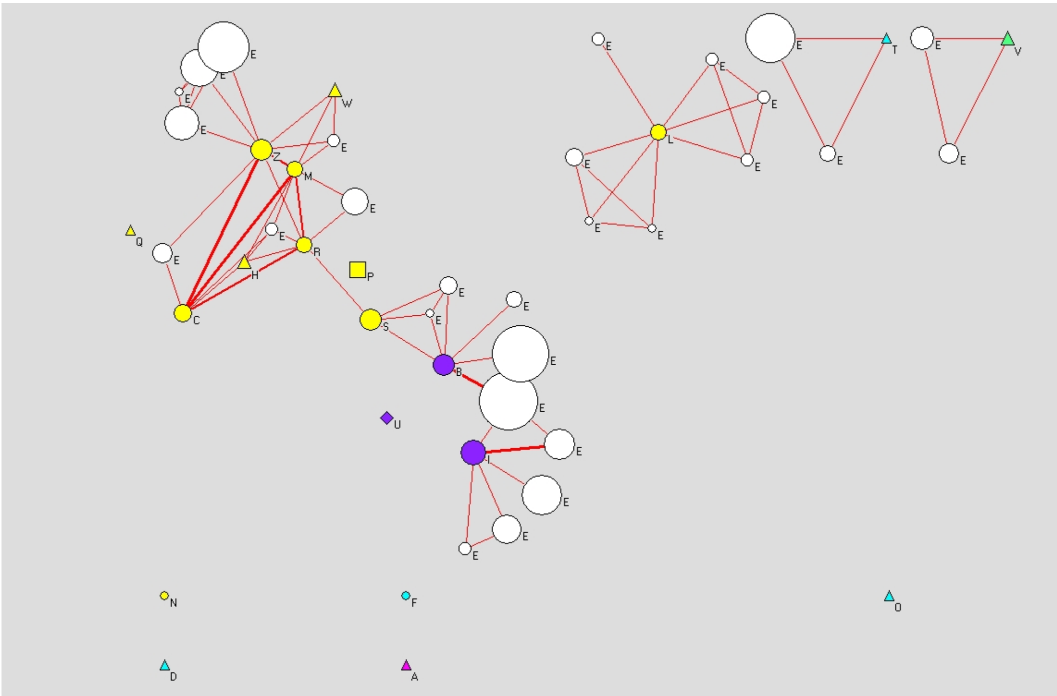


Figure 68 – *Co- authorship network in 2007 (seventh observation).*

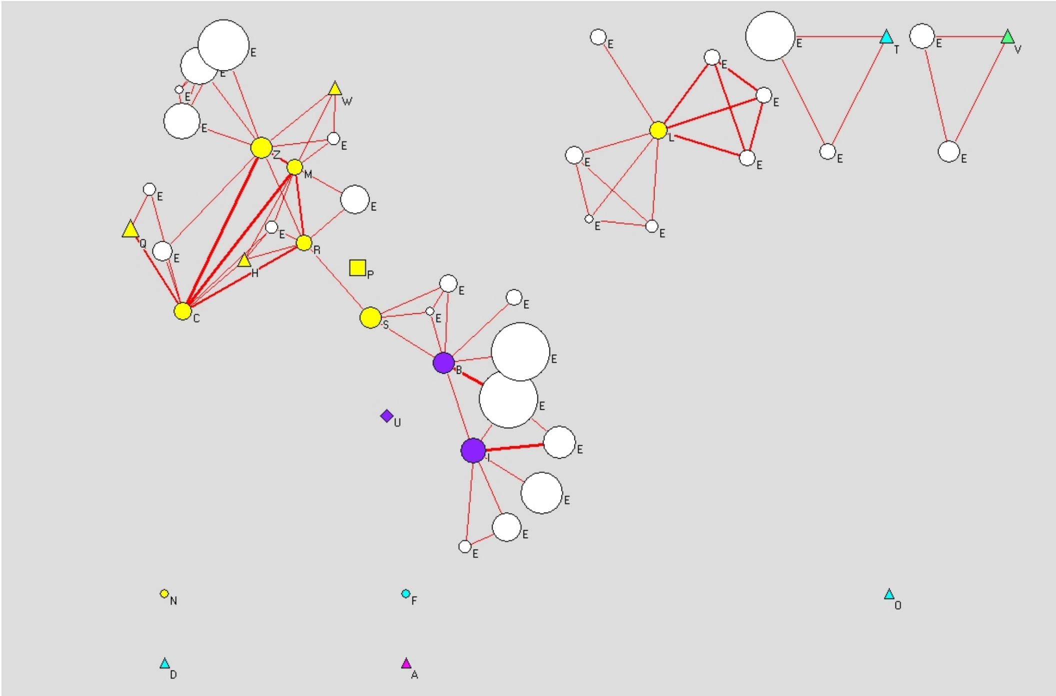


Figure 69 – *Co- authorship network in 2008 (eighth observation).*

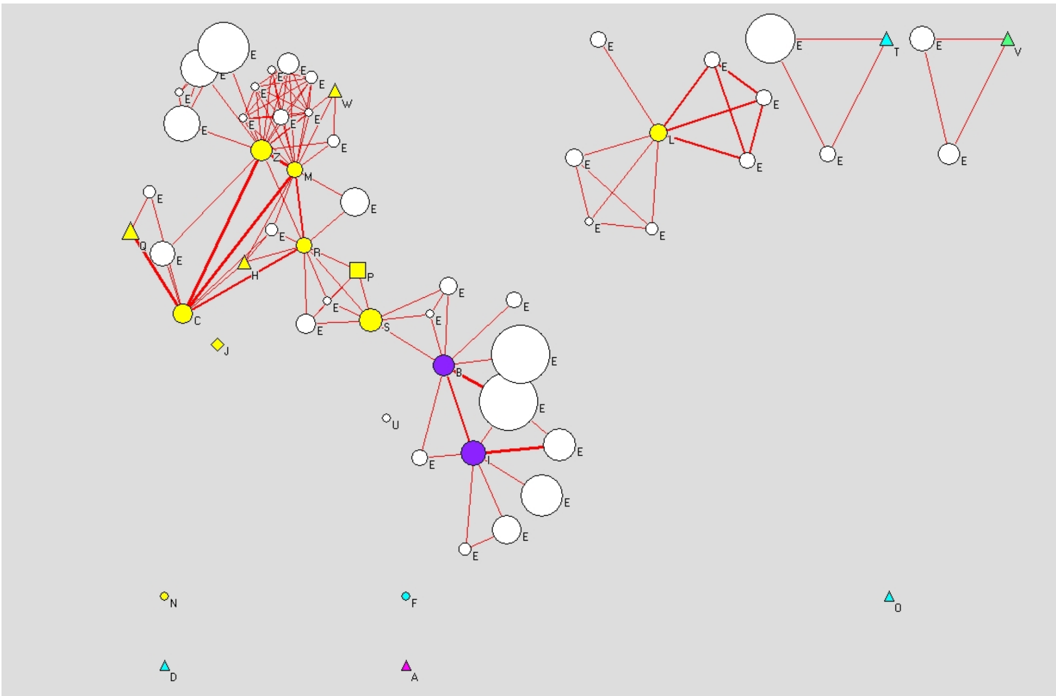


Figure 70 – *Co- authorship network in 2009 (ninth observation).*

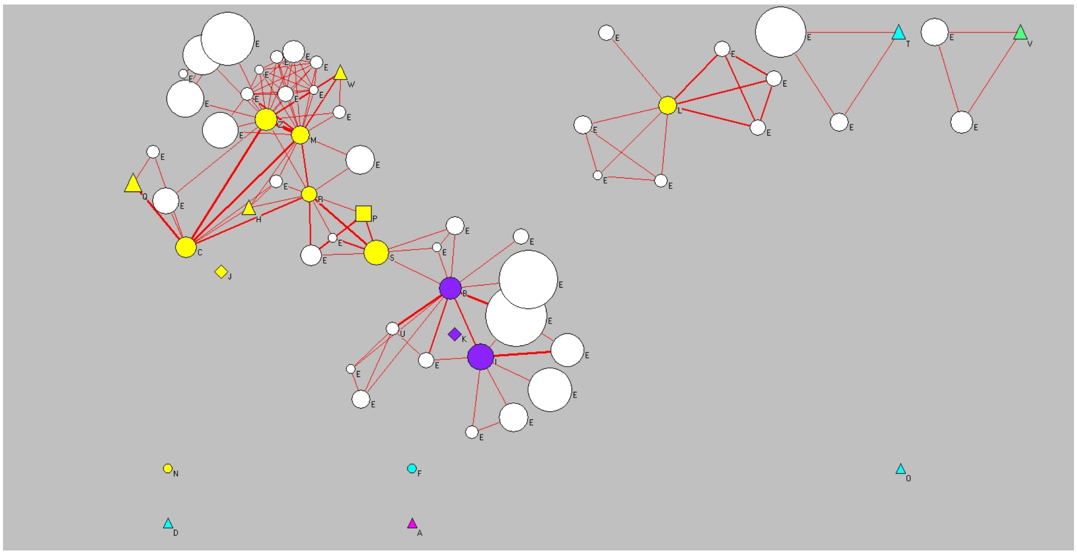


Figure 71 – *Co- authorship network in 2010 (tenth observation).*

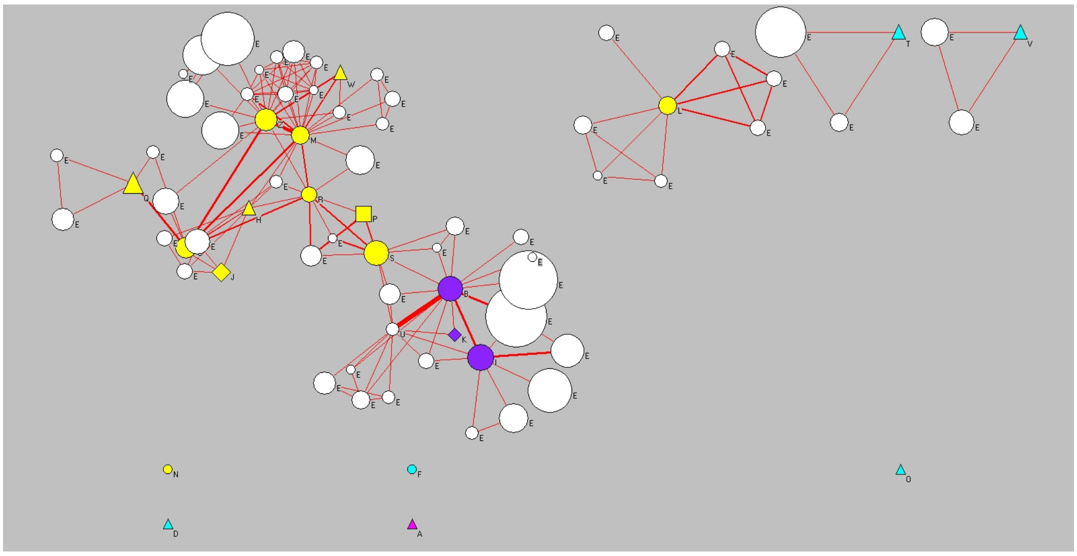


Figure 72 – *Co-authorship network in 2011 (eleventh observation).*

According to the hypothesis 3 (paragraph 6.3), static measures have been calculated without authors considered as isolated (A, D, F, G, N, O).

6.4 The static analysis

To perform the statistic analysis, network's properties are described on two levels:

- Global network properties, delineating the properties as a whole (number of authors, number of co-authored papers, density, average degree, clustering coefficient, inclusiveness index);

- Single actor properties, related to the analysis of properties of individual actors in a network (degree centrality, closeness centrality, betwenness centrality).

Besides structural properties of network are calculated such as identifying cohesive sub-groups (identification of clique).

6.4.1 Measures of cohesion

At macro level, measures found during the study period are summarized (Table 8).

Measures	2001 *(n=13)*	2002 *(n=21)*	2003 *(n=23)*	2004 *(n=23)*	2005 *(n=36)*	2006 *(n=39)*	2007 *(n=43)*	2008 *(n=44)*	2009 *(n=55)*	2010 *(n=59)*	2011 *(n=71)*
Total number of authors (female)	12 (1)	17 (4)	19 (4)	19 (4)	29 (7)	32 (7)	36 (7)	37 (7)	48 (7)	51 (8)	59 (12)
Total number of ties	15	28	31	31	53	62	68	72	115	125	157
Density*	0.19	0.13	0.12	0.12	0.08	0.08	0.07	0.07	0.07	0.07	0.06
Average degree*	2.28	2.60	2.64	2.64	2.80	3.15	3.16	3.27	4.18	4.23	4.42
Clustering coefficient of Wattz-Strogatz	0.95	0.88	0.89	0.89	0.89	0.84	0.83	0.84	0.87	0.87	0.86

*Calculated by formula for weighted undirected graphs (see paragraph 1.5.1).

Table 6 *– Measures of co-authorship networks without isolated authors over time.*

Over time the growth of the number of new links is higher than the growth of the number of new authors during study period. The number of authors starts to 13 and becomes 71, so it increases more than four times, while the number of ties increases more than ten times from 15 links in 2001 to 157 in 2011.

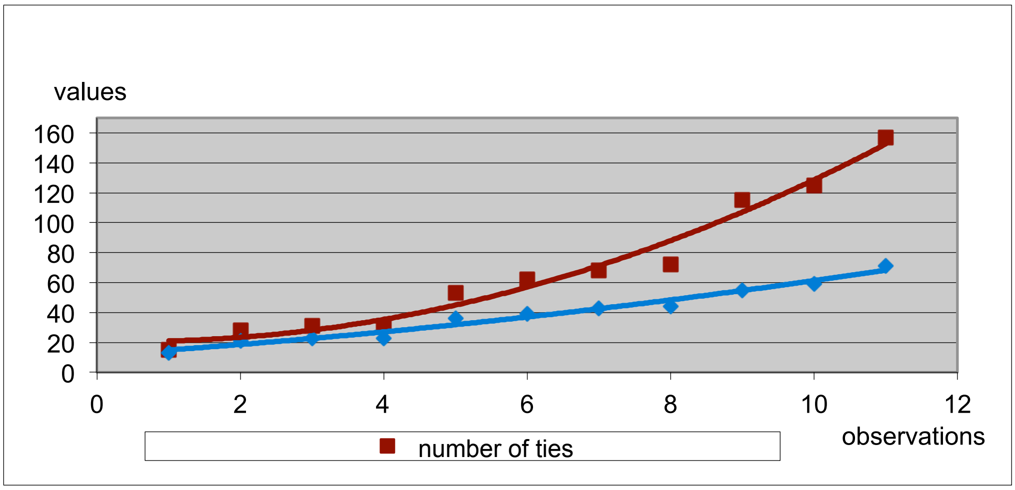


Figure 73 – *Trends of number of authors and their ties in study period.*

The first step of statistic analysis at macro level consists into calculate the degree of integration of network over time through the measures of density, average degree, clustering coefficient, transitivity, and inclusiveness.

The measure of its cohesion represents a very important aspect in the analysis of a network. Fixed the number of nodes, a larger value of the cohesion generally indicates that the network contains a larger number of ties. In the analysis of co-authorship networks, it indicates an increase of the level of scientific activity.

The density, one of the most well known measures for calculation of cohesion, is expressed by the percentage of all possible ties that are present in a network and it can vary between 0 and 1. It captures the idea that a network characterized by many ties has a close structure that is more cohesive (De Nooy et al., 2005).

A closely related measure of structural cohesion is the average degree of the network expressed by the average number of ties for single node. It is often a measure of the cohesion more intuitive than density.

Table 8 shows that over time the values of density decrease. In particular, it starts with value equal to 0.19 (i.e. means that are present the 19 percent of ties of all possible ties) that is a medium value of this index while in the last observation became 0.06 (i.e. the 6 percent of 4,970 possible ties).

It is very important to note that the density is influenced by *size* (number of actors) of network, in particular, it is inversely related to it: normally large networks are characterized by value of density lower than small networks because with increasing size of network, the number of possible ties increases rapidly with the number of actors.

Another factor that can influence the density is the *time*: more time provides to actors of the network more opportunities to build relationships.

As said before, in the first observation co-authorship network is composed by 13 authors while in the last one the number of authors becomes 71, so the reduction in the density must not be interpreted as a loss of cohesion of network because over time the size of network increases.

Besides, although the study period is quite extended, over time the density value does not increase because the network increases its size.

Figure 80 shows the trend of density.

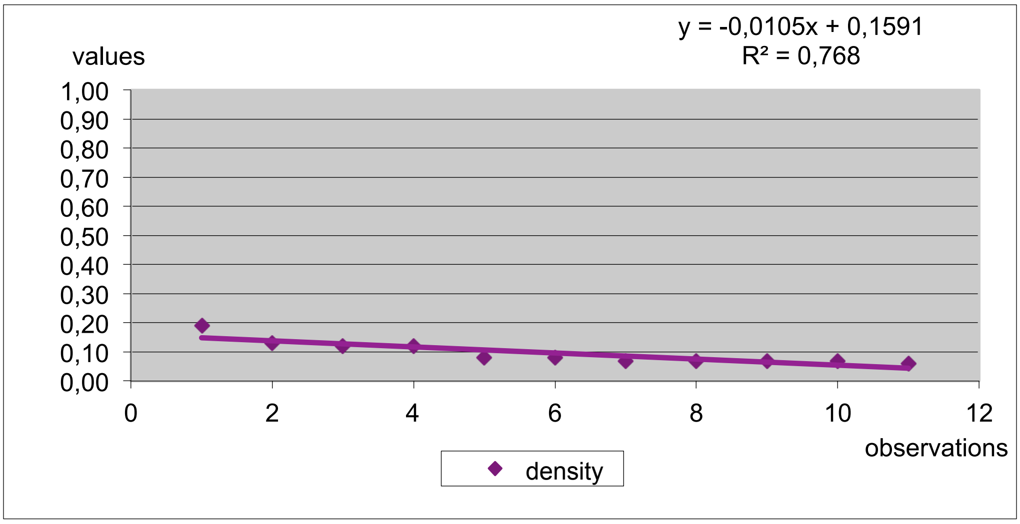


Figure 74 – *The trend of density in study period.*

The average degree is the average number of authors with whom one author has published papers during the study period. This measures gives more information than density because it does not depend on network size.

From Table 8, it starts with value equal 2.28 (i.e. in the first observation each author is linked in media with about 2 other authors) and becomes 4.42 in the last observation.

As shown in Figure 81, over the study period the average degree grows.

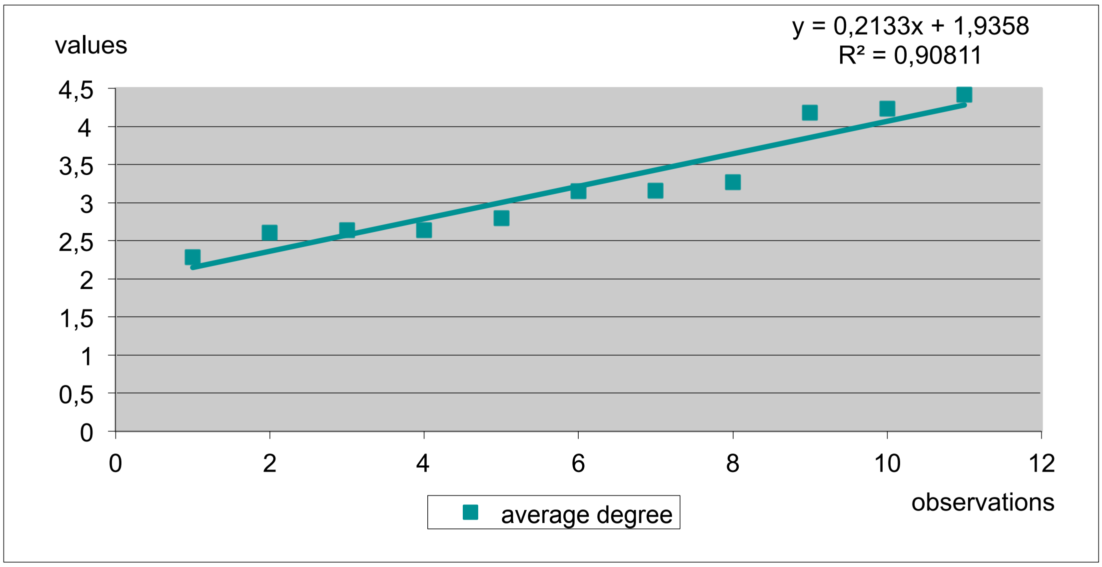


Figure 75 – *The trend of average degree in study period.*

This increase of average degree confirms that over study period the cohesion of network increases.

A measure linked to density is inclusiveness that gives information on the degree to which authors are involved in ties; in other words, it indicates how actors are not isolated, and so the proportion of authors that are actually connected.

In Table 9, the values of inclusiveness index are summarized.

Measures	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011
Total number of authors connected with others	13	18	20	20	33	36	40	42	53	58	71
Total number of authors not connected with others	4	3	3	3	3	3	3	2	2	1	0
Inclusiveness index	0.76	0.85	0.86	0.86	0.91	0.92	0.93	0.95	0.96	0.98	1

Table 7 – *The inclusiveness values over time.*

The values of inclusiveness are very high and they oscillate between 0.76 and 1. This means that the number of isolated authors is smaller than the number of connected authors, and in the last observation inclusiveness index reaches its maximum value (there is not isolated authors).

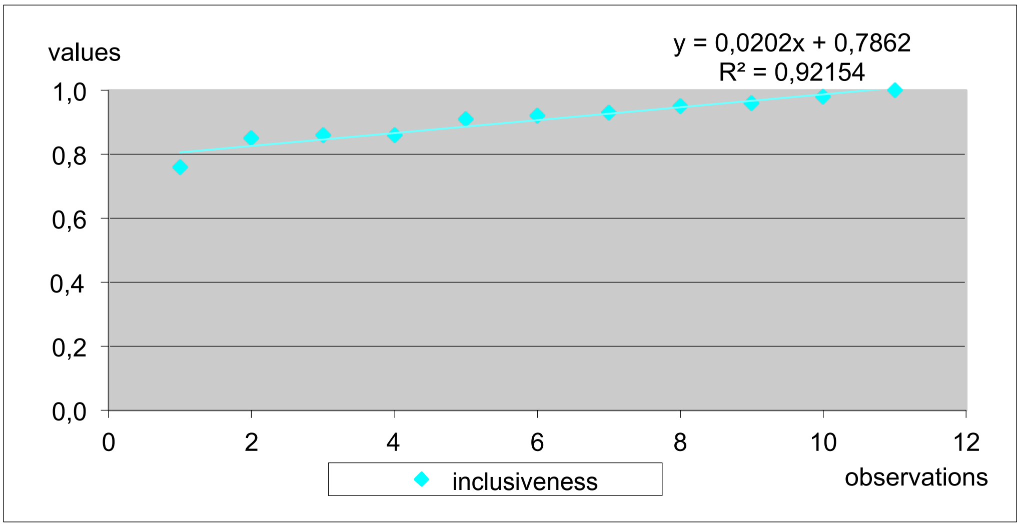


Figure 76 – *The trend of inclusiveness in study period.*

As mentioned in chapter 1, real social networks have an important property: they are clustered, it means that in the network there are local sub-networks (clusters) in which a degree, higher than average degree, occurs (Newman, 2001).

In research field, communities might form, as might form sets of researchers that work on particular arguments.

There are numerous criteria for identifing clustering in a network; one often used consists into examine the local neighborhood of an actor (that is all the actors who are directly connected to it), and to calculate the density in this neighborhood (but leaving out it). After doing this for all actors in the whole network, the degree of clustering is calculated as an average of all the neighborhoods.

Using Pajek, for each of the eleven networks studied, the values of *clustering coefficient of Wattz - Strogatz* are calculated. They denote the overall clustering coefficient that is simply the average of the densities of the neighborhoods of all authors calculated for each observation. As Table 8 shown, the values of Wattz-Strogatz coefficient are high; this indicates simply that, over time, are numerous papers having three or more authors.

In Figure 83, the trends of clustering coefficient are displayed.

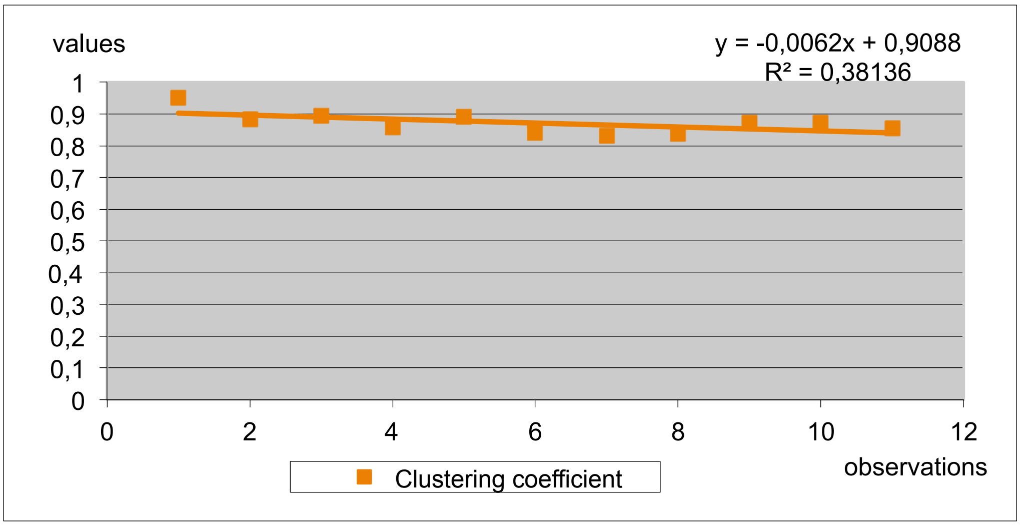


Figure 77 *– The trend of cluster coefficient in study period.*

Another index related to clustering coefficient is *transitivity* coefficient that gives information about the existence of ties among triplets of authors (ties between *i* and *j* and between *j* and *k* implies a tie between *i* and *k*). So, the transitivity indicates the fraction of connected triplets for each observation. A possible explanation can be given by an example. The author A writes a paper with authors B individually; B writes another paper with D, for transitivity, A collaborates with D. For each observation, the number of triplets with 3 legs and the percentage of these triples that are transitive have been calculated.

Measures	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011
Number of triplets with 3 legs	11	17	21	21	36	42	44	46	141	149	173
Transitivity (%)	73.3	51.5	46.6	44.6	36.7	32.8	29.9	26.9	35.4	31.5	26.2

Table 8 *– Transitivity of co-authorship network over time.*

The initial values are high. Starting from 2003, the networks are characterized by ties established among authors of DIEG working in the same disciplinary sector; thus, the high values of transitivity is explained as triplets of authors working at the same institution and in the same research field, and the result may be papers published by couple of three researchers of the triplet. Over time, these values decrease slightly in fact they remain high enough.

In conclusion the networks considered in study period are characterized by a *good* cohesion, and it indicates a *good* level of communications among actors and a *significant* scientific productivity.

6.4.2 Indices of centrality

At single actor level, the indices of centrality (degree, closeness, and betweenness) represent key factors (see paragraph 1.5.3).

The first index calculates the number of links that a node has with the others. In the context of co-authorship network, the *degree centrality* of an authors measures the number of authors with whom he/she has published his/her papers; so, being a central author means that the author has published with many other authors. Probably, this measure can be linked with rank of career.

The degree centrality has been determined with UCINET software and the results, thus, obtained are summarized in Table 11:

FREEMAN'S DEGREE CENTRALITY MEASURES:

	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011
B	0.000	4.000	4.000	5.000	8.000	8.000	8.000	9.000	11.000	17.000	26.000
c	1.000	1.000	4.000	4.000	6.000	12.000	12.000	15.000	16.000	16.000	16.000
H			0.000	0.000	0.000	3.000	3.000	3.000	3.000	3.000	7.000
I	3.000	7.000	7.000	7.000	7.000	7.000	7.000	8.000	10.000	10.000	12.000
J									0.000	0.000	4.000
K										0.000	2.000
L	0.000	3.000	3.000	3.000	6.000	7.000	7.000	10.000	10.000	10.000	10.000
M			3.000	3.000	6.000	11.000	12.000	12.000	20.000	25.000	29.000
P	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	4.000	7.000	7.000
Q							0.000	3.000	4.000	4.000	6.000
R	0.000	0.000	3.000	3.000	3.000	6.000	8.000	8.000	12.000	15.000	15.000
S			0.000	0.000	3.000	3.000	4.000	4.000	8.000	11.000	14.000
T					2.000	2.000	2.000	2.000	2.000	2.000	2.000
U						0.000	0.000	0.000	0.000	6.000	15.000
V					0.000	0.000	0.000	2.000	2.000	2.000	2.000
W					3.000	3.000	3.000	3.000	3.000	5.000	5.000
Z	5.000	5.000	8.000	8.000	11.000	12.000	12.000	12.000	20.000	24.000	24.000

Table 9 *– Degree centrality values of authors in each observation.*

It is possible to identify the most central authors C, M, Z, and B.

The values of degree centrality of author C are higher than those of B and M until 2007, while after they are lower than values of B and M. The degrees of centrality of authors B and M are characterized by very similar trends, and, in particular, over time their trends grow very quickly. During study period, the degree of centrality of author Z increases slowly.

The points corresponding to authors B and M fit the same regression equation: $y = 0.1981x^2 + 0.3106x – 0.1991$; while $y = 0.1072x^2 + 0.5133x – 4.4424$ is the curve regression for author Z, and $y = - 0,0688x^2 + 2,6343x – 3,2788$ is that for author C. In Figure 85, their trends are shown.

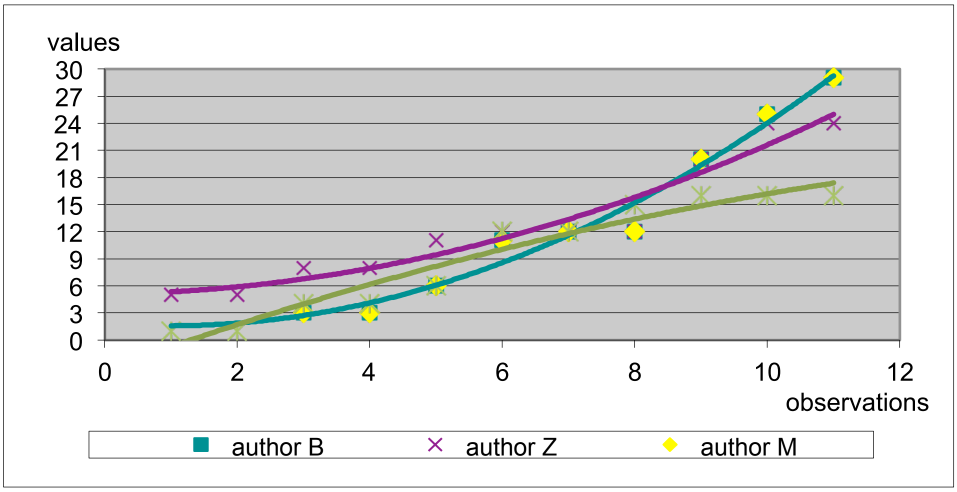


Figure 78 – *The trends of degree centrality of Z and B, C authors in study period.*

The authors B, C and Z belong to department starting from the first observation; the position of authors C and Z is Professor while author B initially Assistant professor becoming a Professor in sixth observation. In fourth observation, author M becomes a DIEG member as Assistant professor and in the sixth observation he/she becomes Professor. Z and C are characterized by higher values of degree than B and M in periods in which B and M are Assistant professors; starting from the sixths observation, degrees of M and B carries on to be the same while the Z and C values are close to them. This result is confirmed by sloops of curves. To have an average measures of slop from the first to fifth observation and from sixth to eleventh observation two linear regression equation have been calculated: $y_1 = 1.5 x – 0.4$ and $y_2 = 3,4 x – 12$. From these it is possible to try that in the second part of observation period (B is became a professor) the average slop is roughly double than that corresponding to the first (B is Assistant professor). The same situation occurs for M. Finally, for Z and C authors, who are always Professors in study period, the straight lines have positive slopes that increase in slow way.

H-index attribute does not seem to have a relevant influence on degree centrality of researchers. For instance, B and Z have both high H-index values, however, M, despite being one of the most degree central authors, is characterized by low H-index values. K (PhD student in all observations in which he/she is considered) is the less central author of the network, and his/her H-index value is very low during all study period. On the other hand, H-index for its definition strongly depends on length of research career.

In Figure 85, the trends of degree centrality of these authors, the three most centrals and the one less central, are shown.

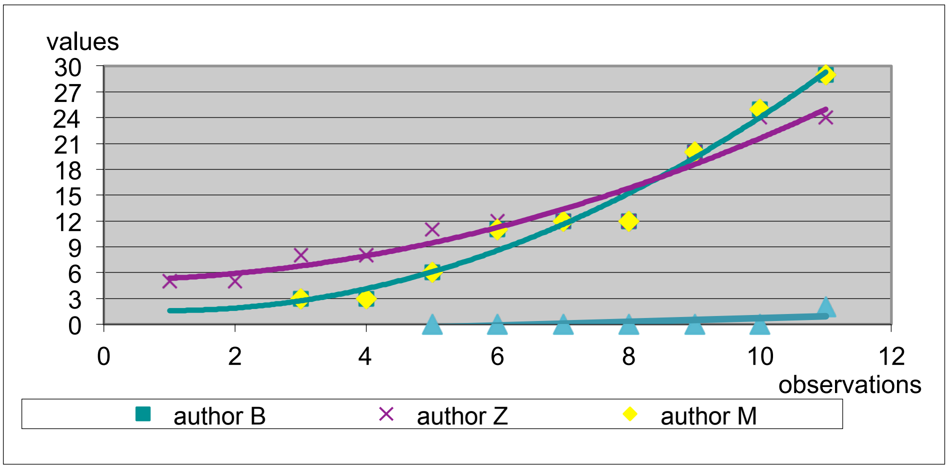


Figure 79 – *The trends of degree centrality of B, M, Z, and K authors over time.*

In order to give some global information about the networks, a histogram has been built. The axis *x* reports the 11 observations, while the axis *y* gives number of authors having a determined degree centrality.

Given the high variety of degree centrality values, these have been collected in 9 categories, and to each category a different color has been assigned (Figure 86).

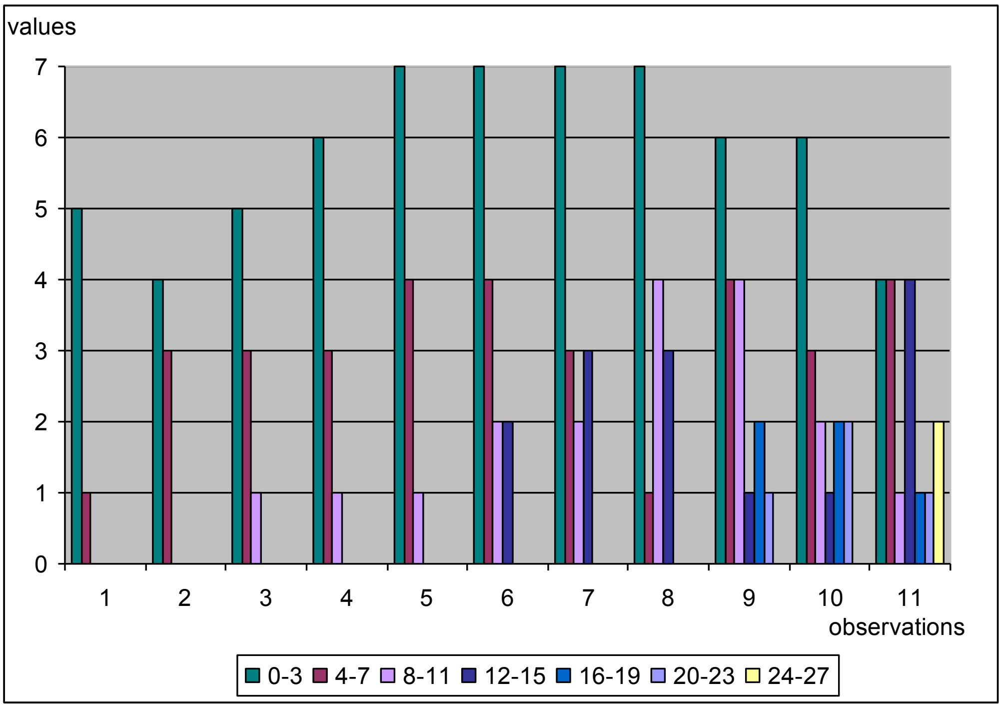


Figure 80 – *Percentages of categories in the last observations.*

In the first observation, five authors are characterized by a value of degree centrality that belongs to category 0-3 ties, and one authors is characterized by a value of degree centrality belonging to category 4-7.

Over time the number of categories present in each observation increases, until, in the last observation, 7 categories are present. This means that in the course of the observations the degree centrality increases (coherently with the fact during their carrier the authors have the opportunity to know new people), so the number of ties among authors grows, and this is coherent with the other results found until now: over time, authors create new ties and the cohesion of overall network increases.

The second index considered is *closeness centrality*. It expands the definition of degree centrality by focusing on how close a node is to all other nodes of the network. So, the intent behind this measure is to identify the nodes that could reach others quickly.

In the context of co-authorship network, higher is the closeness centrality of researcher simpler is for him to reach other researchers and then acquire scientific resources in a more efficient way. This could make to think that a linkage between closeness centrality and H-index of each author exists.

CLOSENESS CENTRALITY

	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011
B	0.000	6.944	6.111	6.117	3.964	3.424	5.257	5.541	5.202	5.482	6.151
C	12.000	6.173	6.180	5.823	3.825	3.555	5.243	5.570	5.263	5.524	6.228
H			0.000	0.000	0.000	3.529	5.116	5.416	5.138	5.375	6.130
I	9.091	7.092	6.215	6.183	3.955	3.418	5.066	5.422	5.075	5.326	6.261
J									0.000	0.000	5.863
K										0.000	5.882
L	0.000	5.556	5.000	4.762	3.333	3.030	2.778	2.703	2.083	1.923	1.563
M			6.180	5.823	3.825	3.555	5.250	5.563	5.299	5.561	6.352
P	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	5.222	5.482	6.066
Q							0.000	5.395	5.090	5.321	5.963
R	0.000	0.000	6.180	5.823	3.817	3.545	5.310	5.628	5.331	5.604	6.267
S			0.000	0.000	3.928	3.400	5.296	5.599	5.284	5.561	6.173
T					2.941	2.632	2.439	2.381	1.887	1.754	1.449
U						0.000	0.000	0.000	0.000	5.292	6.003
V					0.000	0.000	2.439	2.381	1.887	1.754	1.449
W					3.813	3.539	5.134	5.422	5.148	5.380	6.082
Z	12.500	6.250	6.250	5.882	3.842	3.562	5.257	5.570	5.305	5.566	6.239

***Table 10** – Closeness centrality values of authors in each observation.*

Three main groups that contain the most central authors can be determinate: one composed by C, I, and Z, the other formed by M, and R, and third composed by B and S. In the first observation authors C, Z, and I are characterized by the highest values of closeness centrality. From the third observation, their trends become very similar to those of the R and M. Finally, in the first observation in which they come into play, B and S start with a value of closeness equal to 0 that becomes very high in the remaining observations.

In Figures 87-89, trends of closeness centrality for B, C and R authors are shown, assumed to be representatives of the three different groups.

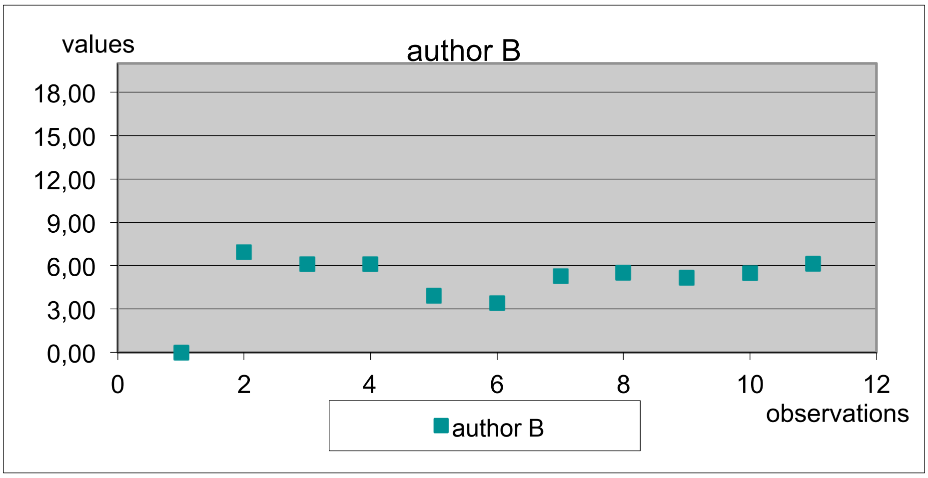


***Figure 81** – Trends of closeness centrality of B authors over time.*

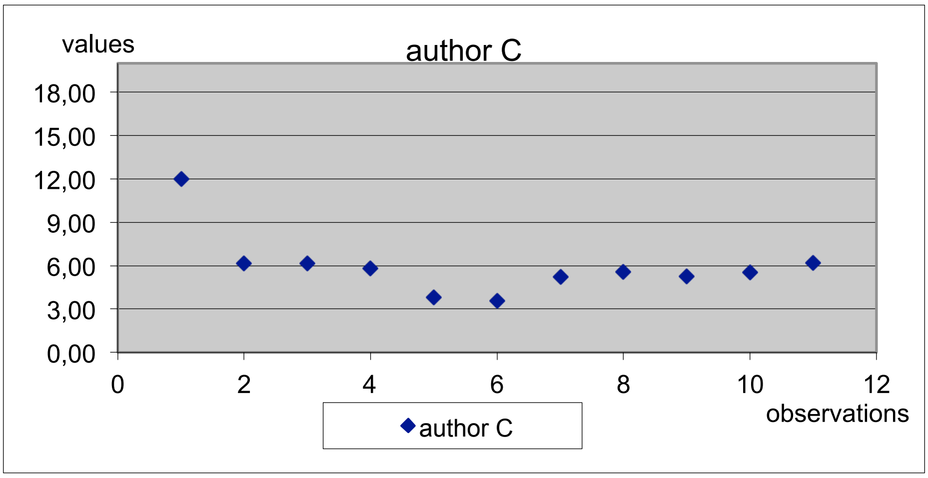


Figure 82 *– Trends of closeness centrality of C authors over time.*

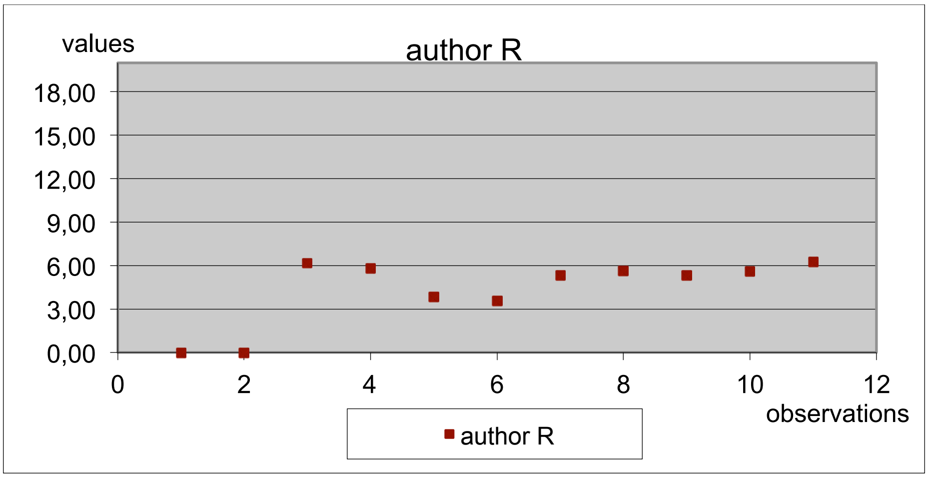


Figure 83 *– Trends of closeness centrality of R authors over time.*

Initially author C is characterized by a value of closeness centrality higher than B and R. From second observation the value of closeness centrality of author B becomes proximate to that of C author. From fourth observation the values of B, C, and R are very similar. Until fourth observation, C (and authors belong to his/her group) is the most central, after B, C, and R (and authors of the same his/her group) have similar values of closeness centrality so they represent the most central of the networks.

Tables 13-15 contain the values of h-index and closeness centrality for the most central authors, identified before, in each observation.

	2001		2002		2003		2004	
	H-index	Closeness	H-index	Closeness	H-index	Closeness	H-index	Closeness
B	3	0.000	5	6.944	5	6.111	6	6.117
C	3	12.000	3	6.173	4	6.100	4	5.823
I	7	9.091	7	7.092	7	6.215	7	6.183
M					2	6.100	2	5.823
R	2	0.000	2	0.000	3	6.180	3	5.823
S					3	0.000	4	0.000
Z	4	12.500	4	6.250	5	6.250	5	5.882

Table 11 *– H-index and closeness centrality of the most central authors from 2001 to 2005.*

	2005		2006		2007		2008	
	H-index	Closeness	H-index	Closeness	H-index	Closeness	H-index	Closeness
B	6	3.964	6	3.424	6	5.257	6	5.541
C	4	3.825	4	3.555	4	5.243	4	5.570
I	7	3.955	7	3.418	7	5.066	8	5.422
M	2	3.825	2	3.555	3	5.250	3	5.563
R	3	3.817	3	3.545	3	5.310	3	5.628
S	5	3.928	5	3.400	6	5.296	7	5.599
Z	5	3.842	5	3.562	6	5.257	6	5.570

Table 12 *– H-index and closeness centrality of the most central authors from 2005 to 2008.*

	2009		2010		2011	
	H-index	Closeness	H-index	Closeness	H-index	Closeness
B	6	5.202	6	5.482	7	6.151
C	5	5.263	5	5.524	5	6.228
I	8	5.075	8	5.326	8	6.261
M	3	5.299	4	5.561	4	6.352
R	3	5.331	3	5.604	3	6.267
S	7	5.284	7	5.561	7	6.173
Z	6	5.305	6	5.566	6	6.239

Table 13 *– H-index and closeness centrality of the most central authors from 2006 to 2011.*

In order to show the linkage between closeness centrality and H-index, the graphs containing over time the trends of H-index and closeness centrality of each most central author have been built.

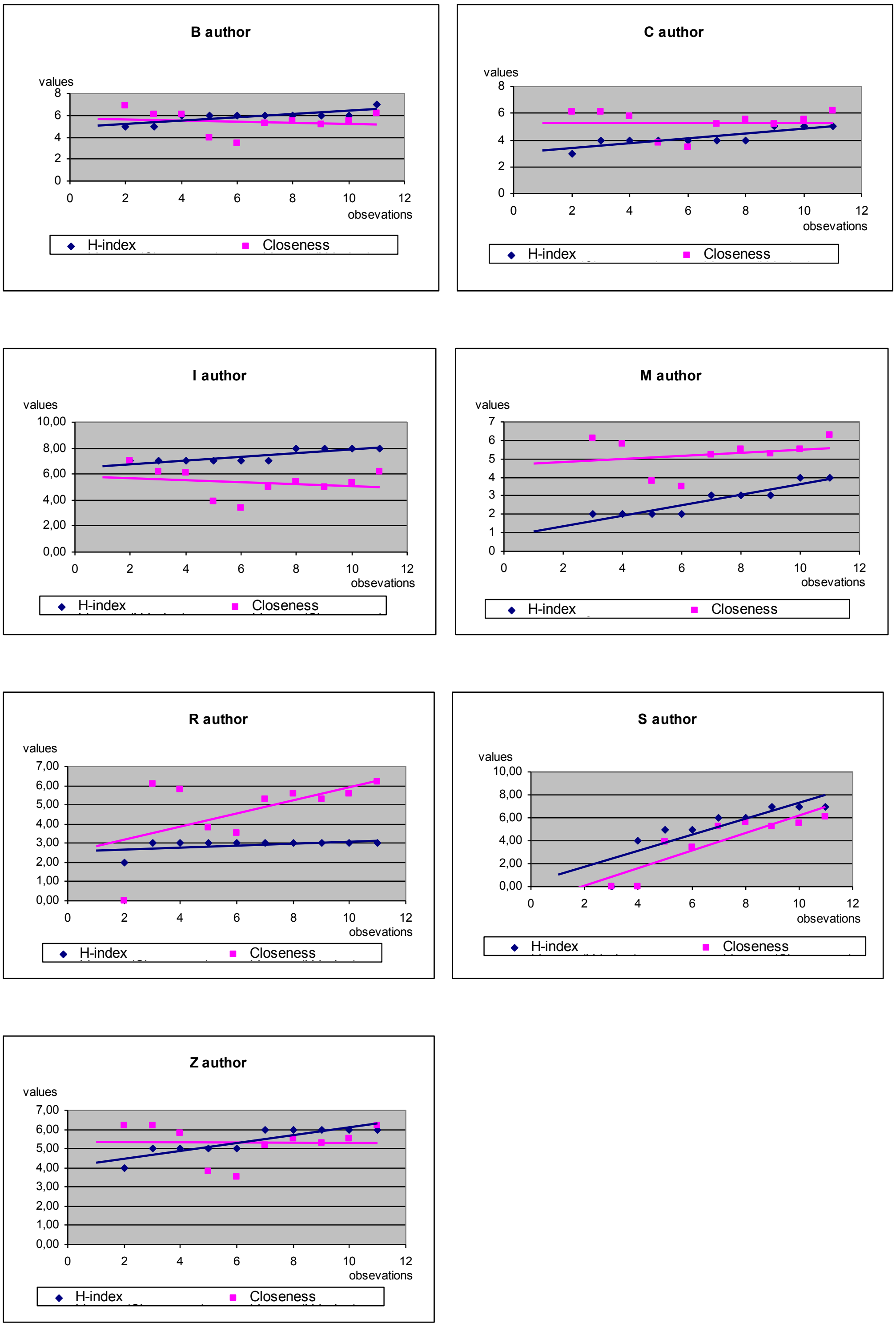


Figure 84 – *Trends of closeness centrality and H-index for each of the most central authors.*

For each author the values of closeness centrality and H-index corresponding to the first observation have not been considered for the hypothesis 2 (see paragraph 5.3) that concerns the restriction of not having regarded the information for the years prior to the first observation.

B, C, I, M, R, S, and Z have the values of closeness corresponding to fifth and sixth observations very similar and less than trend values.

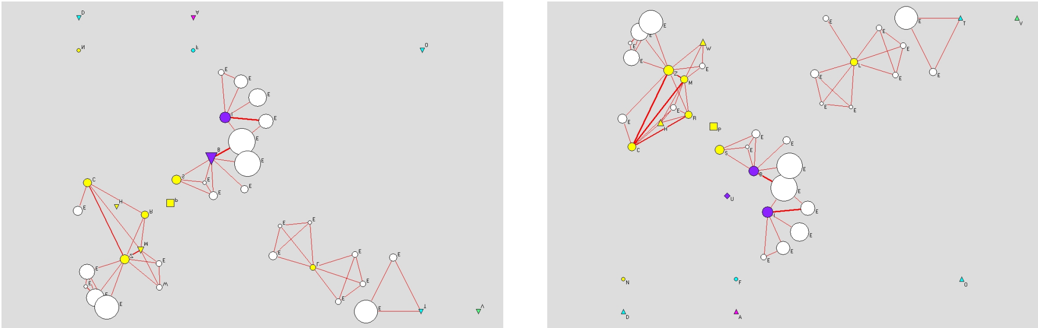


Figure 85 – *Co- authorship network in 2005 (on the left side) and in 2006 (on the right side).*

Looking to snapshots of network related to fifth and sixth observations, it is possible to note that in 2005 new authors become part of the network and each of them go to connect with just one old author belonging to network. So this entry causes a reduction of the closeness of all the components of the network making all them less reachable. In 2006 another author is added linking to just one isolated author of the network: this causes a further reduction of closeness centrality of all authors. In 2007 the authors R and S connect themselves by tie that acts as a bridge linking two large, before separated, parts of the network. This new tie makes all authors more reachable and then increasing their values of closeness centrality.

The trends of H-index and closeness centrality appear to be similar, indicating that the linkage between them exists.

The third centrality index calculated is *betweenness;* it is obtained by determining how often a particular node is found to appertain to the shortest path between pair of nodes in the network. That is, more times a node is present in shortest paths between pair of nodes higher is its betweenness centrality. A researcher characterized by high betweenness value will probably obtain easier knowledge and resource by the other researchers and then increase the quality of his/her papers.

FREEMAN BETWEENNESS CENTRALITY

	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011
B	0.000	0.000	0.000	0.000	62.000	62.000	286.000	161.000	263.500	394.167	360.250
C	0.000	0.000	0.000	0.000	9.667	33.667	75.667	87.833	86.500	137.500	293.867
H			0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	208.000
I	0.000	14.000	28.000	34.000	52.000	52.000	136.000	84.500	120.500	138.167	495.233
J									0.000	0.000	0.000
K										0.000	0.000
L	0.000	0.000	0.000	0.000	18.000	30.000	30.000	15.000	15.000	15.000	15.000
M			0.000	0.000	2.667	18.000	60.000	31.667	134.100	170.600	735.183
P	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Q							0.000	0.000	0.000	0.000	108.000
R	0.000	0.000	0.000	0.000	0.000	3.333	323.333	186.667	342.167	443.667	332.633
S			0.000	0.000	0.000	0.000	308.000	176.000	300.000	405.000	260.467
T					0.000	0.000	0.000	0.000	0.000	0.000	0.000
U						0.000	0.000	0.000	0.000	1.000	114.933
V					0.000	0.000	0.000	0.000	0.000	0.000	0.000
W					0.000	0.000	0.000	0.000	0.000	0.000	0.000
Z	4.000	4.000	24.000	24.000	46.000	67.000	191.000	104.167	224.767	276.767	285.067

***Table 14** – Betweenness centrality values of authors in each observation.*

The values do not allow easily identifying the most central authors. It is necessary to take a look to snapshots on networks over study period.

In 2007, a big increase of betweenness centrality of the authors B, R, and S occurs. B, R, and S represent key authors as they make a bridge between two large clusters of network (Figure 92).

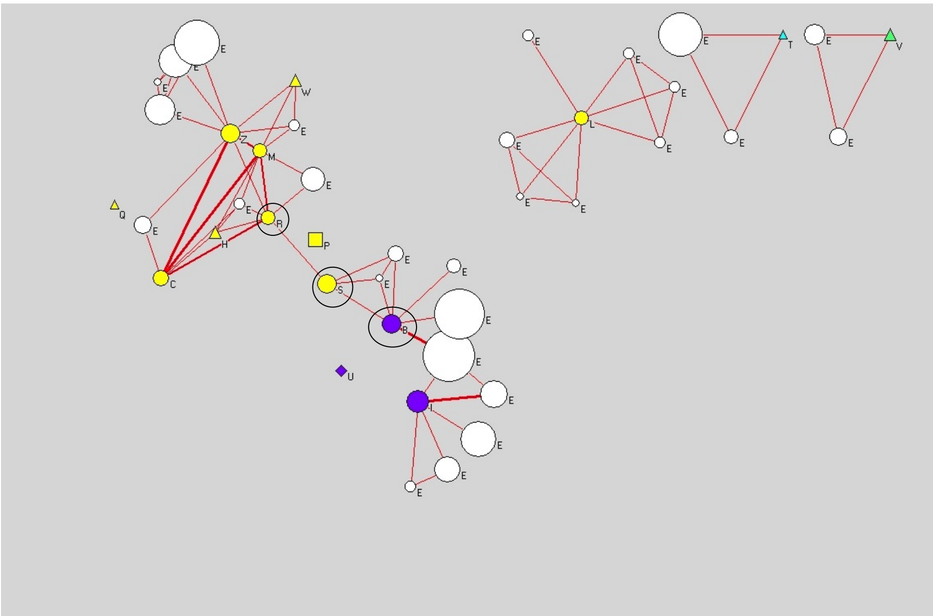


***Figure 86** – Highest values of betweenness centrality of B, R, and S.*

Over time network increases its size and cluster' densities increase. So the values of betweenness centrality of some authors belonging to these clusters grow. In particular, in the last observation B, C, M, S, R, Z, and I become the most centrals.

This result can be extended because the presence in a network of several central authors generally attests that is formed by several clusters joined by bridges.

Table 16 indicates the presence of some researchers J, K, T, U, and V that are the less central authors with reference to betweenness centrality.

For both the most central and the less central groups, trends of two authors, chosen as representative, are shown (Figure 93).

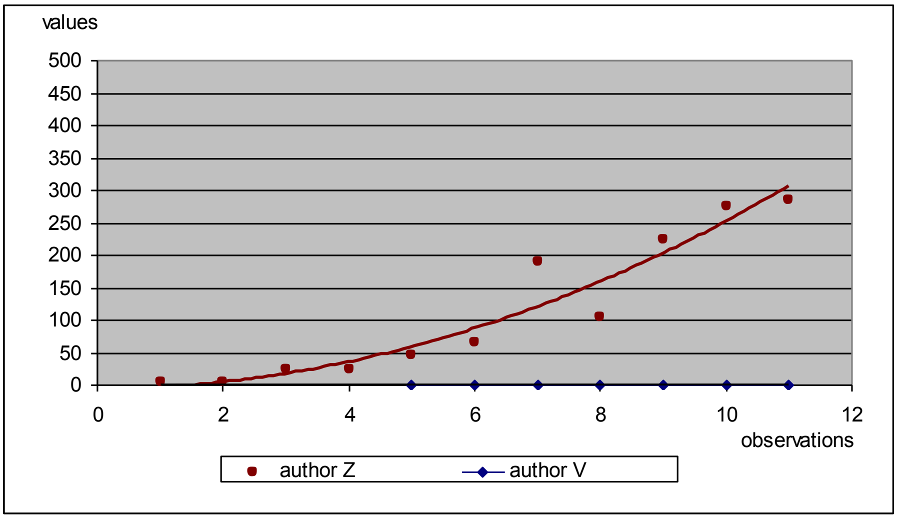


Figure 87 – *Trends of betweeness centrality of authors Z (representative of group of the most central authors) and V (representative of group of the less central authors).*

The author Z is characterized by low initial values of betweeness centrality but over time its value increases. His/her curve fits $y = 2{,}5639x^2 + 0{,}1788x - 5{,}3069$ with $R^2 = 0{,}915$. Instead, the V is characterized by values of betweeness that are zero in all study period.

6.4.3 Identification of cliques

The identification of network cliques (see paragraph 1.5.2.) can help to find cohesive groups of researchers. The characteristic of a clique is that each node appertaining to it is linked with to all other nodes. This makes clique identification a very important way to uncover meaningful groups in a network.

The tool used to identify cliques present in the networks is UCINET that provides an automatic identification of the cliques.

In Table 17, the number of cliques identified and the maximum size of them are indicated.

Measures	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011
Number of cliques	2	5	6	6	11	13	15	17	20	23	29
Number of cliques with maximum size	1	1	1	1	1	1	1	1	1	1	1
Maximum size of cliques	5	5	5	5	5	5	5	5	9	9	9

***Table 15** – Number of cliques identified over study period.*

In the following, for each network observation the cliques present have been highlighted by colored circles.

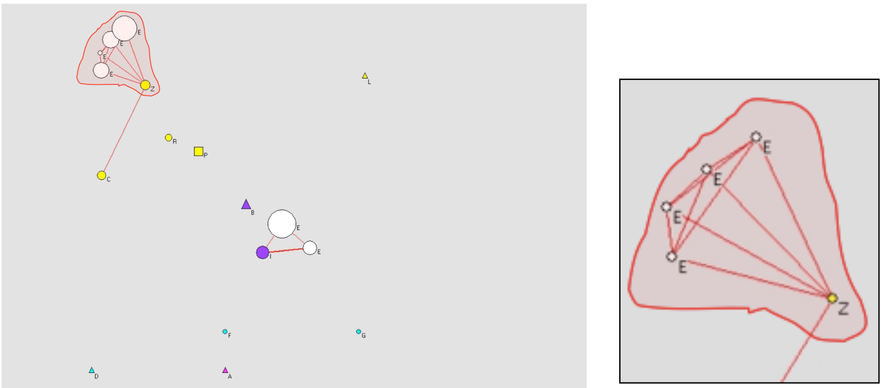


***Figure 88** – Clique with max size present in 2001 (on the left side) and its details.*

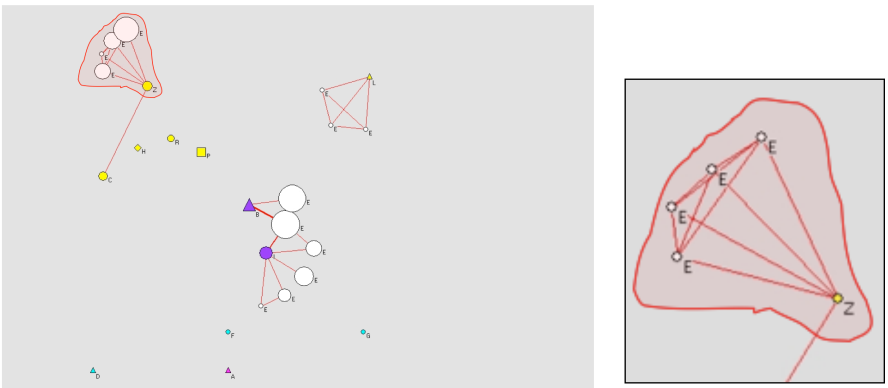


***Figure 89** – Clique with max size present in 2002 (on the left side) and its details.*

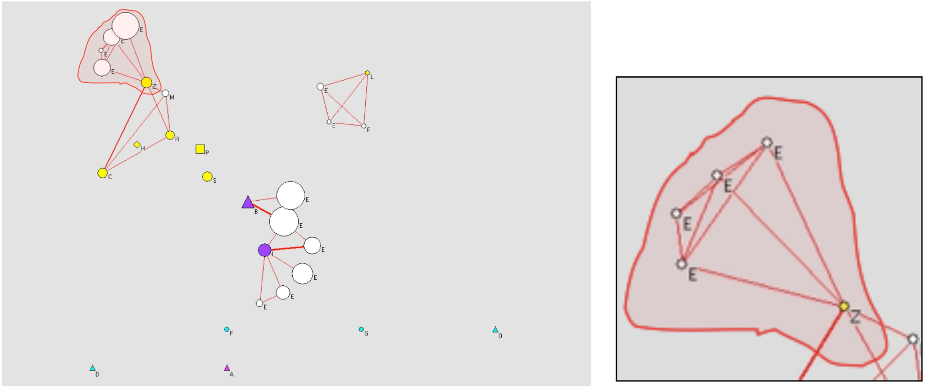


Figure 90 – *Clique with max size present in 2003 (on the left side) and its details.*

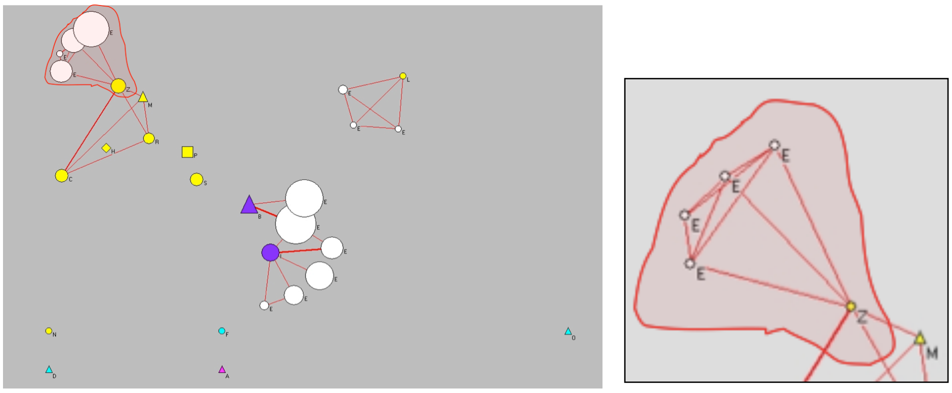


Figure 91 – *Clique with max size present in 2004 (on the left side) and its details.*

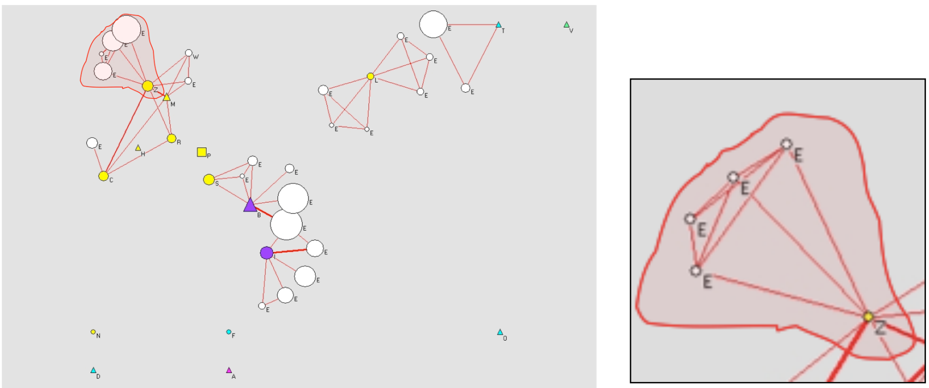


Figure 92 – *Clique with max size present in 2005 (on the left side) and its details.*

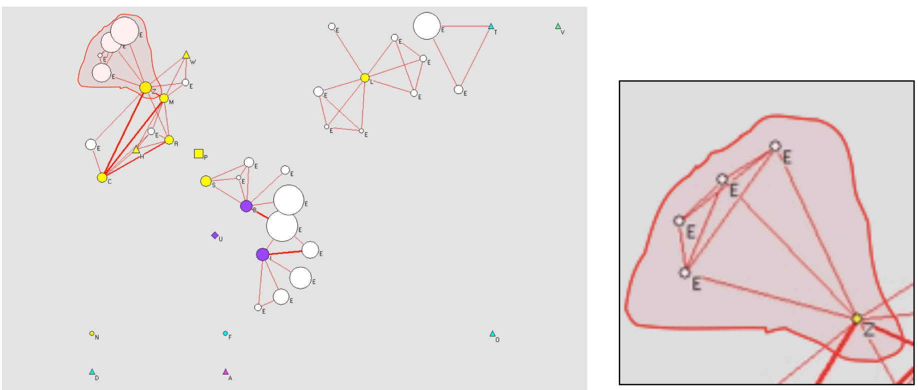


Figure 93 – *Clique with max size present in 2006 (on the left side) and its details.*

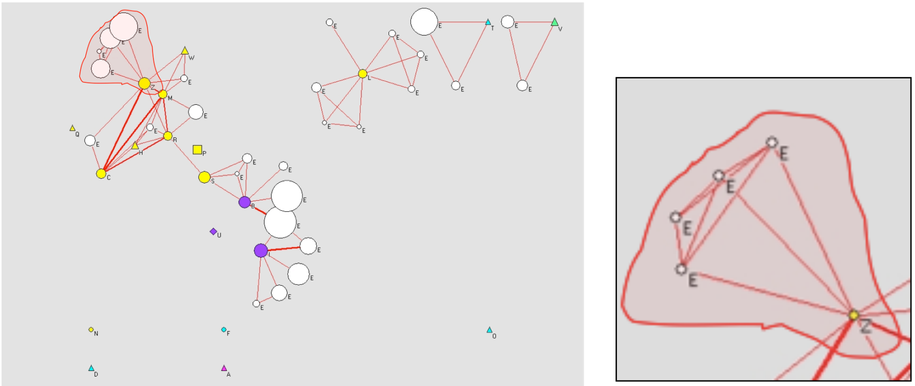


Figure 94 – *Clique with max size present in 2007 (on the left side) and its details.*

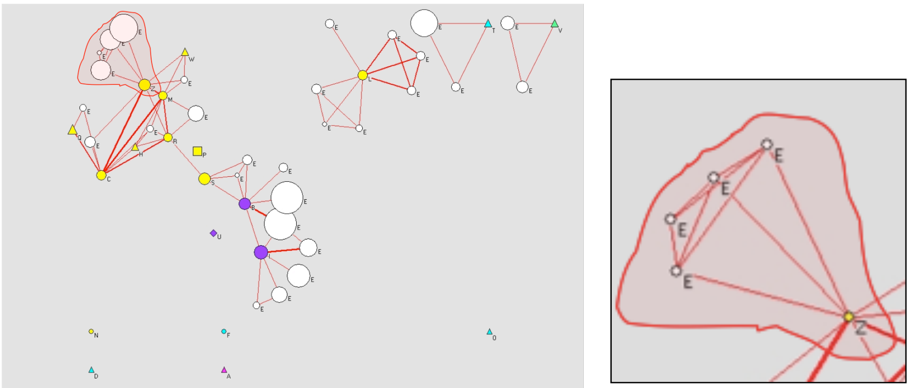


Figure 95 – *Clique with max size present in 2008 (on the left side) and its details.*

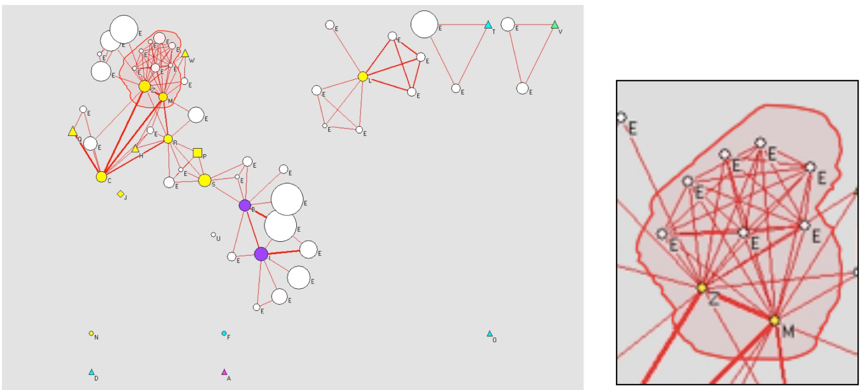


Figure 96 – *Clique with max size present in 2009 (on the left side) and its details.*

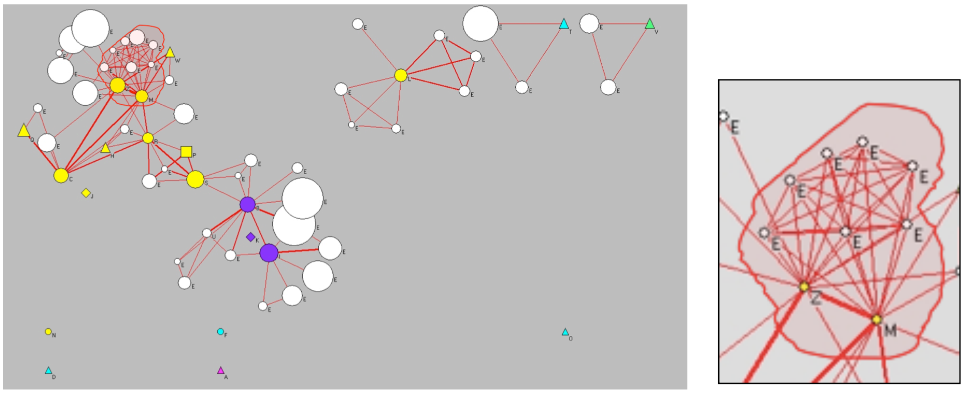


Figure 97 – *Clique with max size present in 2010 (on the left side) and its details.*

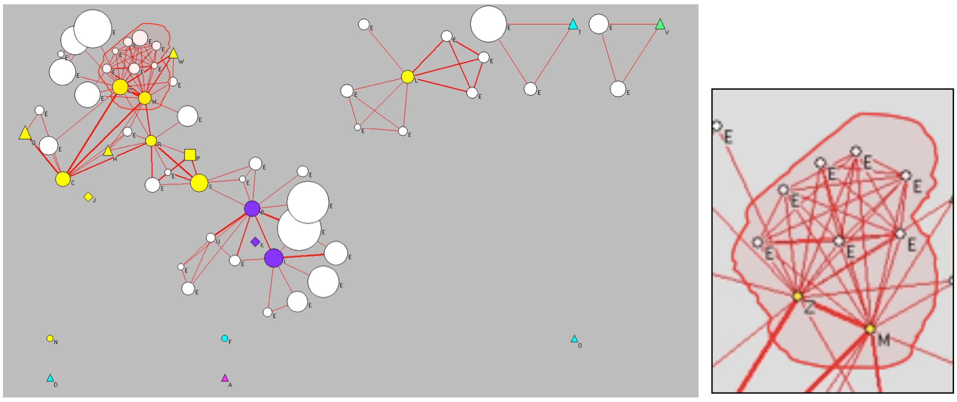


Figure 98 – *Clique with max size present in 2011 (on the left side) and its details.*

The cliques identified have an average size of 4 authors and the largest one consists of 9 authors. The authors of the cliques are primarily Professors and external authors.

The cliques identified have a common characteristic: each of them is composed by members of department and two or more external authors. When there is more than one author of department they belong to the same disciplinary sector.

In 2005 only exception occurs: there is a clique that includes two authors, S and B, of department that belong two different disciplinary sectors and an external author. Analyzing this clique, B is a professor in MAT discipline while the author S is a professor in ING discipline. This means that two authors have joined their different skills; the first one is an expert in modeling while the second in economic theory, in order to produce a unique work.

6.5 The dynamic analysis

Briefly (for details see the paragraph 2.6), given a set of subsequent observations of a social network, the corrisponding panel data represent snapshots of a dynamic process driven by actors trying to optimize some their objective function, both with regard to their own network position and their own behavior.

The aim of the method consists into estimate the variables and the coefficients of this objective function starting from data obtained from observed situations of the networks.

This result can be achieved by simulating the underlying process, and optimizing the fit between simulated process and real process through maximum likelihood criteria.

Data shown in the previous paragraph have been, then, entered in SIENA software to identify the effects that drive the network evolution.

6.5.1 Input data

Several data formats in SIENA are allowed one of these is the binary adjacency matrix (SIENA can not work with weight ties).

So, networks have been represented by binary adjacency matrices, one for each observation, in which each element represents scientific paper: if the actors *i* and *j* have been co-authored in one or more papers, the element x_{ij} is equal to 1, conversely, if they have not published any papers together, the element x_{ij} is zero.

As mentioned in previous chapters, social theory suggests that network development is determined for particular reasons (Katerndahl, 2011). In research networks, researchers create ties selecting their *key* collaborators upon some characteristic such as their disciplinary sector, or their carrier levels. Besides, ties can be established through a conscious or unconscious desire to link to researchers with similar attitudes.

In order to capture the influence that some attributes can have on selection of co-authors, the characteristics of authors, described through files (one for each attribute), have been entered into SIENA. Note that, for each attribute, a distinction between constant (one value per author valid for all observations) and changing (*M-1* vectors that correspond to the *M-1* transitions between subsequent observations) has been introduced.

Disciplinary sector and membership to the department are assumed to be constant attributes, while level of career and H-index are changing attributes.

Finally, the file describing composition change, that is when the authors joined and/or left the department, has been created. This file contains *n* lines (one line for each author) with on each line four numbers. The first two concern joiners, the last two concern leavers: the first number indicates the last observation moment at which the actor is not yet observed, the second one specifies the time of joining (expressed as a fraction of the length of the period), the third indicates the last observation moment at which the actor is observed, and the fourth indicates the time of leaving (also expressed as a fraction).

In Table 18, an example of composition change file for a general actor is presented, in which the number of observations is considered to be 5.

Example of file with times of composition change				
Present at all five observation times	0	1.0	5	0.0
Joining in period 2 at fraction 0.6 of length of period	2	0.6	5	0.0
Leaving in period 3 at fraction 0.4 of length of period	0	1.0	3	0.4
Joining in per. 1 (0.7) and leaving in per. 4 (0.2)	1	0.7	4	0.2
Leaving in per. 2 (0.6) and joining in per. 3 (0.8)	3	0.8	2	0.6

Table 16 ***–*** *Example of composition change file.*

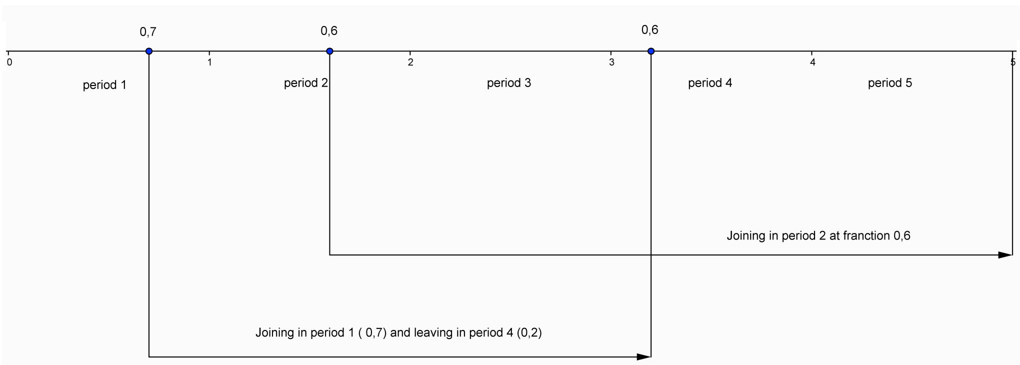


Figure 99 – *Representation of 2^{nd} and 4^{th} rows of Table 18.*

SIENA needs to know how to combine and use the imported and selected data: observations of the network must be entered in the correct chronological sequence and attributes must be entered in right way including whether constant or variable and whether independent or dependent.

In social *selection process*, the individual attributes are independent variables ties and hence the evolution of relation network constitutes the dependent variables. While in a social *influence process*, the attributes of the actors are dependent variables and they change over time.

In the following, the attributes, both constant and changing, have been considered as independent variables because the aim is that to understand how authors, on the basis of their attributes, *selected* their co-authors.

6.5.2 Changes in the network

For hypothesis 3 (see paragraph 6.3), in the following isolated authors A, D, F, G, N, O have not been considered.

After defining the input data, the next step is to specify the effects (see appendix C) that are considered influencing the network evolution. Starting from a first model in which some (also only one) of chosen effects are included, in successive steps, on the basis of results obtained, a second model in which some effects can be excluded and/or other effects can be included is considered; the process continues subsequently complicating the model through the progressive addition of other effects until to all chosen effects are included; also considering different combinations of effects.

The choice of the effects to include in each model has been performed on general evaluations that depend on:

- Network typology (actors and ties);
- Network characteristics (attributes).

Following this logic, for co-authorship research network *outdegree* (this effect is as a default in the model), *balance, Same home institution, H-index similarity* have been considered and four models were sequentially run by SIENA.

Model 1: only constant control rate parameters related to each transition (that indicates the expected, thought estimated, average frequency of unobserved changes per author within the networks) have been selected.

Model 2: two parameters, *outdegree* (a measure of density) and *balance*, have been added. The first one indicates the tendency to create arbitrary ties, while the *balance* parameter captures the tendency of authors to prefer others who makes same choice as them (create new co-authorship ties with authors that have ties with common co-authors).

The next two models consider the possible influence that author's characteristics could have on their choices to operate.

Model 3: *Same home institution* parameter (non-changing characteristic) has been inserted. When this parameter is positive it expresses the tendency of the authors to be tied to others with exactly the same value on the attribute. So, on the basis of the attribute chosen, the parameter describes the impact of belonging to the same institution as predictor of the tendency to create new ties.

Model 4: *H-index similarity* effect has been added. A positive similarity implies that the authors prefer ties to others with similar values on this attribute. So, the *H-index similarity* parameter has been tested in order to verify if it can be seen as an element of attractiveness.

Taking in account the typology of network, the *Pairwise conjunctive model* (see paragraph 3.3 for list of all possible kinds of models and their meanings) has obviously been chosen to explain the dynamic of changes as two or more authors choose to write one or more papers together in common agree; so, ties among them exist only if they are both in agreement.

In Table 19, during study period the dynamics of networks are summarized.

Changes		2001-02	2002-03	2003-04	2004-05	2005-06	2006-07	2007-08	2008-09	2009-10	2010-11
Network											
	Joined	8	2	0	13	3	4	1	11	4	12
	Left	0	0	0	0	0	0	0	0	0	0
Ties											
	0 -> 1	13	3	0	22	9	6	6	43	10	32
	1 -> 1	15	28	31	31	53	62	68	72b	115	125
	1 -> 0	0	0	0	0	0	0	0	0	0	0

Table 17 **–** *Annual changes in ties over time.*

This Table contains the annual changes between subsequent observations (joined, left, created, maintained, and eliminated ties).

It results that all periods are characterized by the entry of several new resources; the label 0 -> 1 specifies the number of created ties between two subsequent observations; the label 1 -> 1 indicates the number of ties maintained between two observations; 1 -> 0 specifies the number of eliminated tie between two subsequent observations. For eliminated ties related row contains only zero values because in co-authorship networks (see hypothesis 1), ties cannot be eliminated. The value of total number of ties is not decreasing over time.

6.5.3 The mechanisms driving co-authorship networks

The development of network has been considered into three significant instants: the born of the department (2001), the end of period 2002 – 2006, and the end of period 2006 – 2011. The period from 2002 to 2006 has been characterized by entry of all members having high level of carrier (Professors, Assistant professors), while the period 2007 – 2011 has been characterized by a large growth of general research activity. So, in SIENA three networks, respectively corresponding to the 2001, to the 2006, and to 2011, have been entered.

Effects	Model 1	Model 2	Model 3	Model 4
Rate parameters				
$r_{2001\text{-}2006}$	0.12	0.28	0.36	0.35
$r_{2006\text{-}2011}$	0.20	0.26	2.31	2.31
Networks measures				
Outdegree (density)		0.67	1.15	- 0.39
Balance		14.99	22.19	20.08
Authors characteristics				
Same home institution			- 0.99	- 1.17
H-index similarity				2.20

Table 18 *– Models predicting co-authorship network evolution from 2001 to 2011.*

In the transition from 2001 to 2006 the general rate parameter is smaller than for 2006 to 2011. In particular, in the model 4 $r_{2001\text{-}2006}$ is equal to 0.35, so in the first transition there is an average of 0.35 changes per author, while $r_{2006\text{-}2011}$ is 2.31, so in the second transition the average changes became 2.31. These low values denote some level of stability in the networks, that is a little tendency of the authors to create ties over time.

In all models, the values of *outdegree* parameter are very low, and in model 4 the corresponding value is negative; the formation of a tie implies some *costs* (in terms of time, effort, and resources) per each author; for this reason, researchers have limited number of different co-authors. This implies a negative value of degree.

The values of *balance* are, instead, positive and high, so forming ties is more likely when the opportunity for closure exists. In particular, in all models these values are the highest among effects.

With reference to *same home institution* parameter, negative values show that authors tend mainly to create new ties with authors that do not belong to same institution.

Finally, *H-index similarity* parameter presents a positive value, this means that the authors form ties with others that are characterized by similar H-index values.

To identify tendencies of actors' action, it is sufficient to interpret only parameters of *significant* effects.

According to the results, researchers prefer to publish papers with others having co-authors in common.

The triplet formed by B, S, and U authors represents an example. S becomes a component of network in 2003 in which S and B are not linked (Figure 106a). In fifth observation, S and B write a paper together so they link (Figure 106b); in 2006, also U becomes a component of DIEG but he/she is not linked with anyone until 2009 (Figure 106c). In 2010, B and U join by various papers (Figure 106d). At this point, S and U have B as common co-authors and, in 2011, S and U link too (Figure 106e).

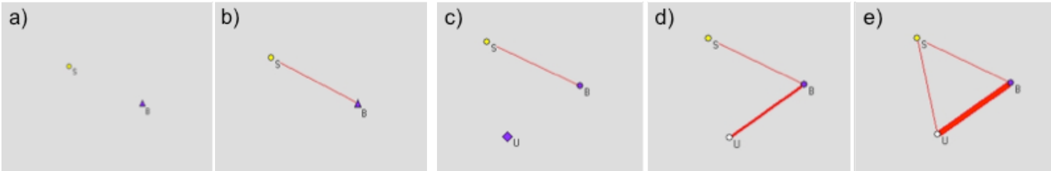


Figure 100 – *Balance effect for B, S, and U authors.*

The same process operates on triplet composed by B, U authors, and I. In 2001, B and I belong to network but there is no tie between them (Figure 107a). In 2006 U becomes part of network and there are no ties among B, I, and U (Figure 107b). In 2009, B and I write some papers together. In 2010, also B and U join by various papers (Figure 107c), and, thus, U and I have B as common co-authors; in 2011, U and I link too (Figure 107d).

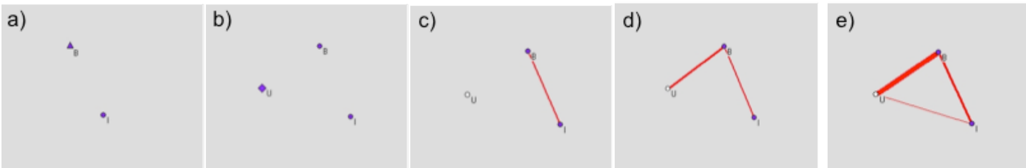


Figure 101 – *Balance effect for B, I, and U authors.*

In conclusion, U operates in both cases following the balance effect, as he/she choses his/her co-authors forming triplets.

6.6 Co-authorship networks of DIEG

Co-authorship networks of DIEG have been realized not considering the external authors. They are, therefore, sub networks of the co-authorship networks before considered. The choice to consider the co-authorship networks of DIEG allows a better detecting of dynamics within department.

In the following, the DIEG networks for each observation are shown (Figures 108-119) and measures calculated for extended networks are recalculated and briefly described.

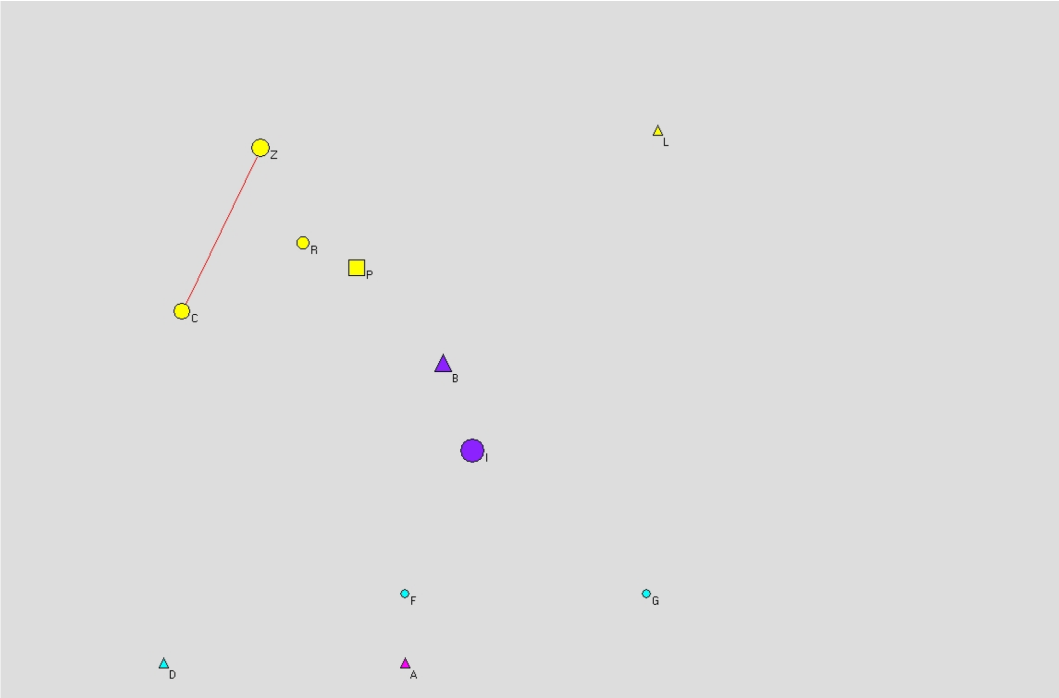


Figure 102 – *Co-authorship network of DIEG in 2001.*

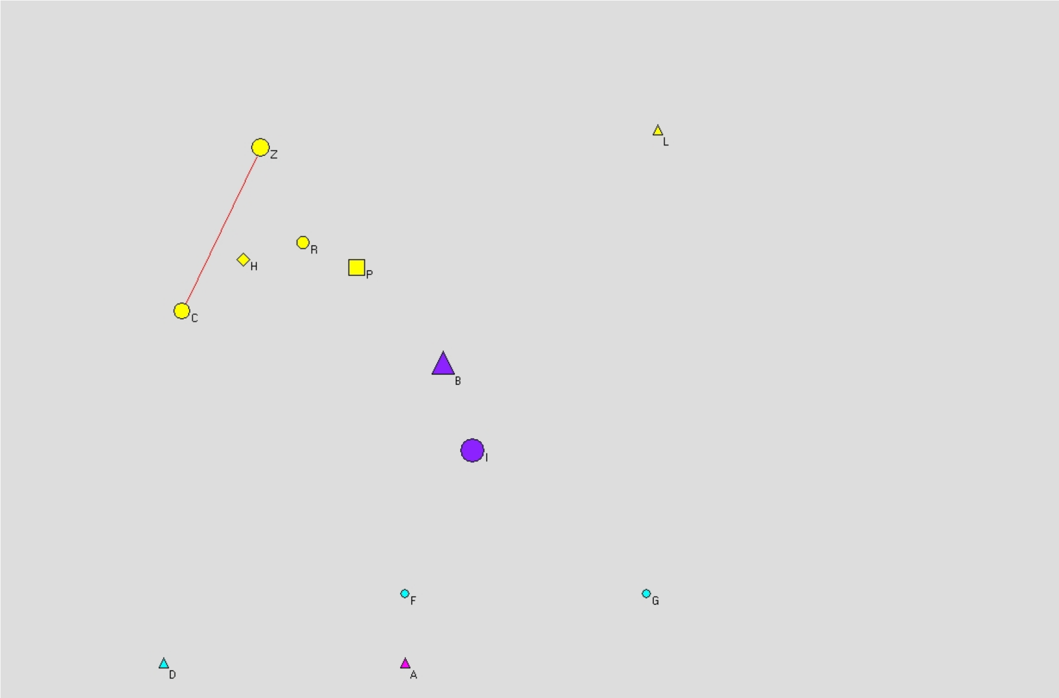


Figure 103 – *Co-authorship network of DIEG in 2002.*

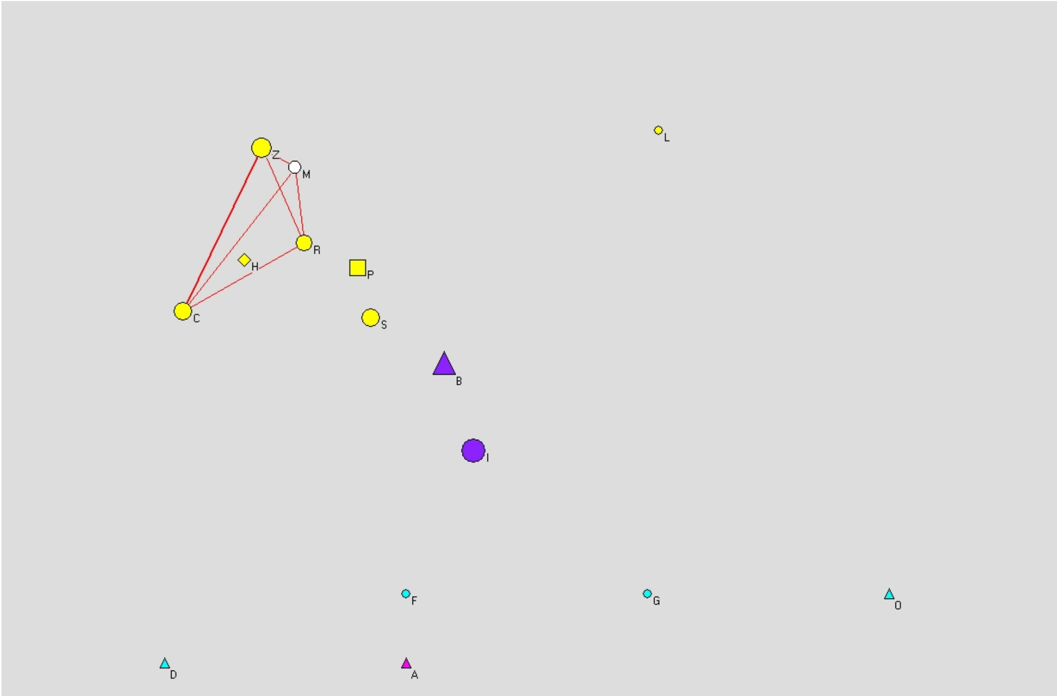


Figure 104 – *Co-authorship network of DIEG in 2003.*

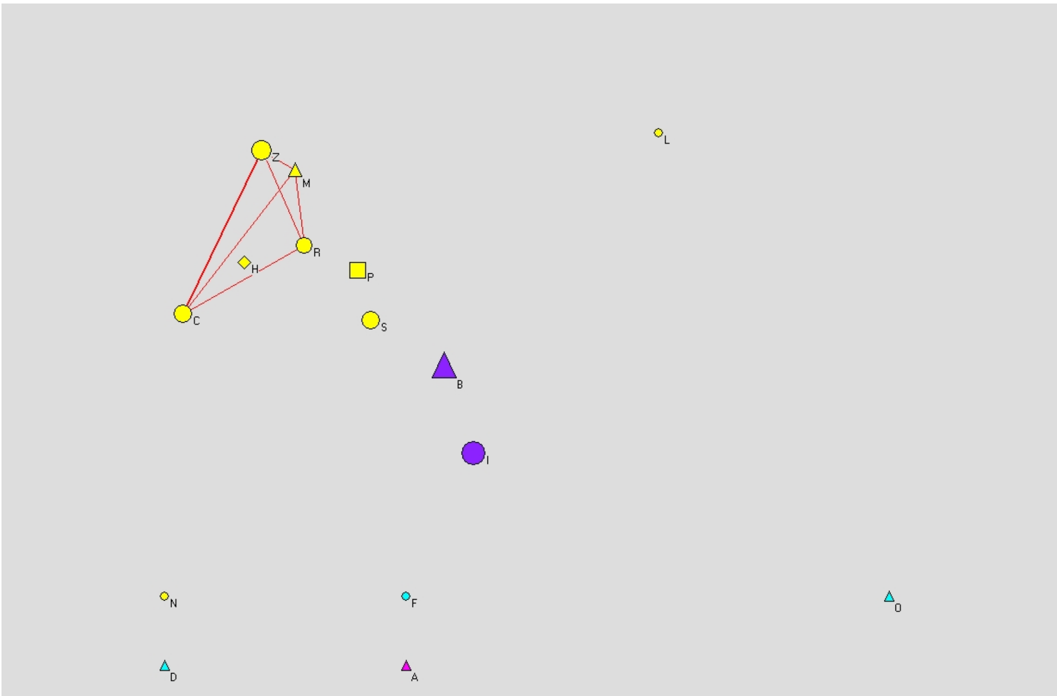


Figure 105 – *Co-authorship network of DIEG in 2004.*

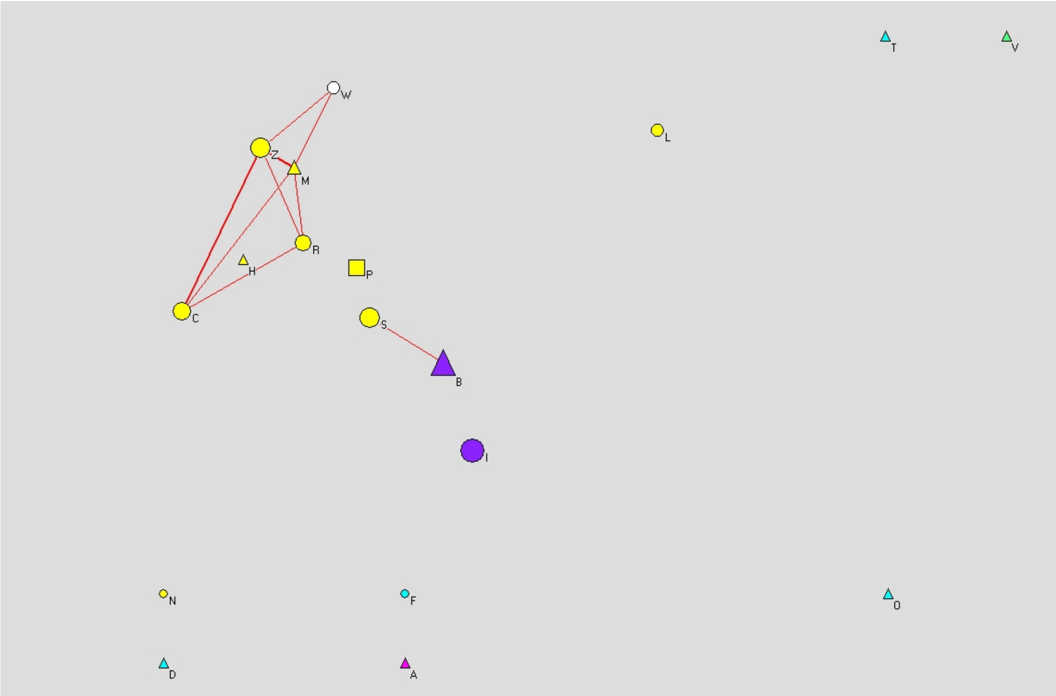


Figure 106 – *Co-authorship network of DIEG in 2005.*

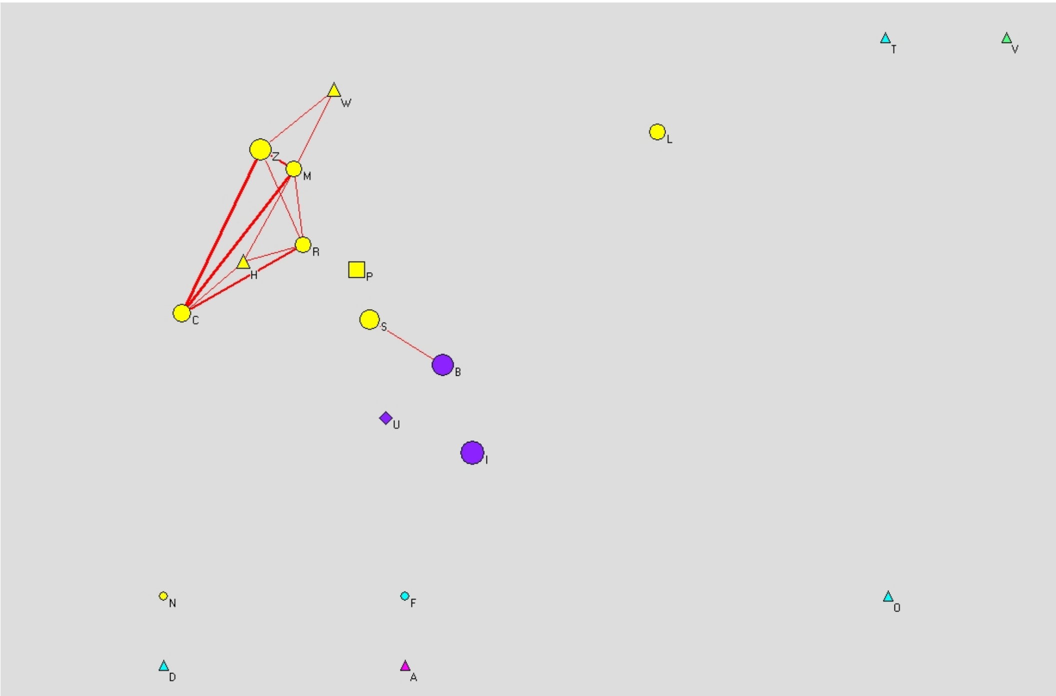


Figure 107 – *Co-authorship network of DIEG in 2006.*

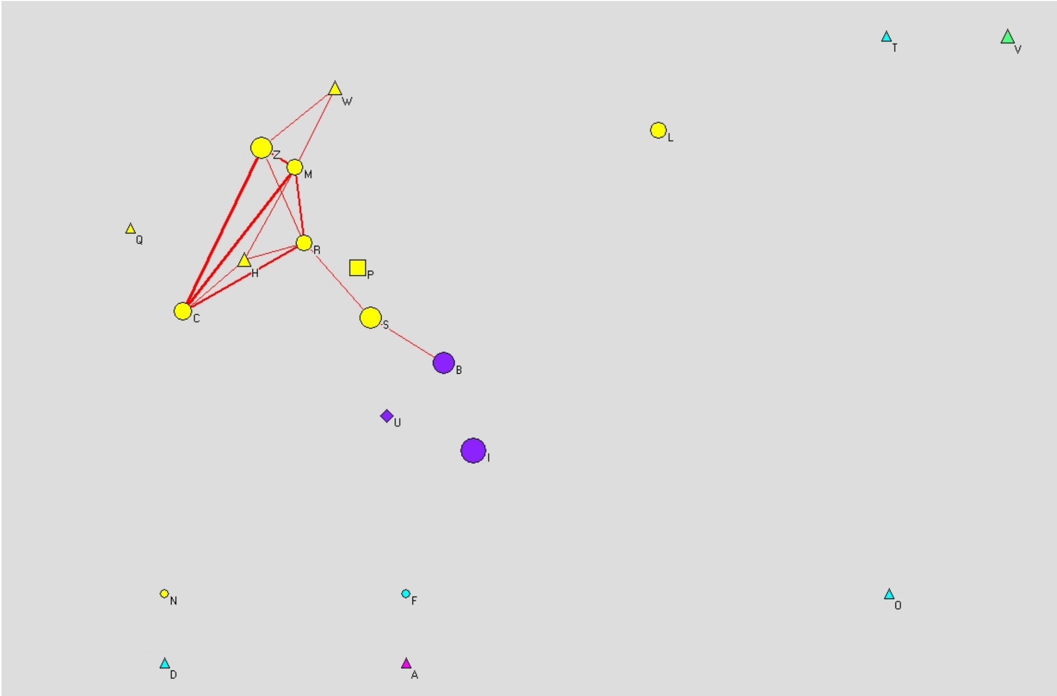


Figure 108 – *Co-authorship network of DIEG in 2007.*

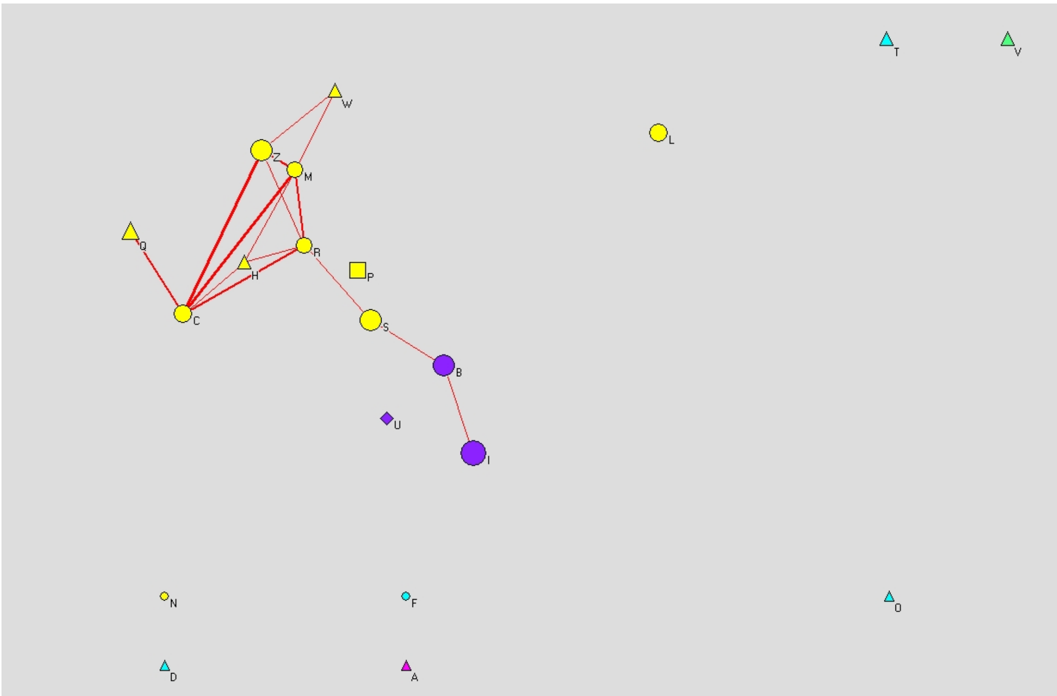


Figure 109 – *Co-authorship network of DIEG in 2008.*

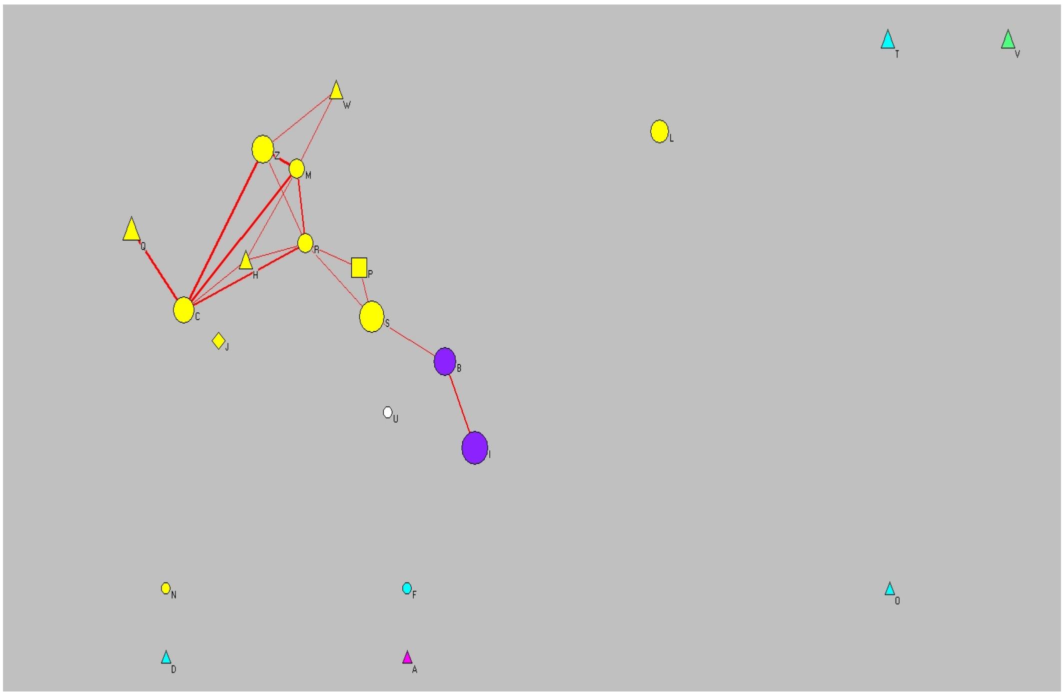


Figure 110 – *Co-authorship network of DIEG in 2009.*

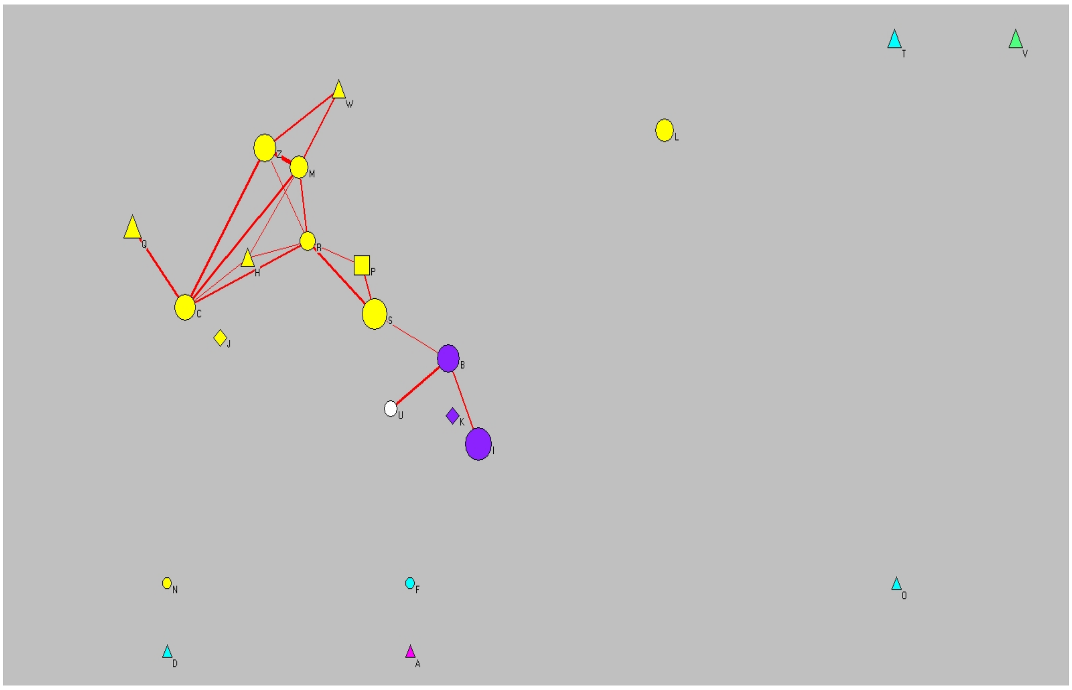


Figure 111 – *Co-authorship network of DIEG in 2010.*

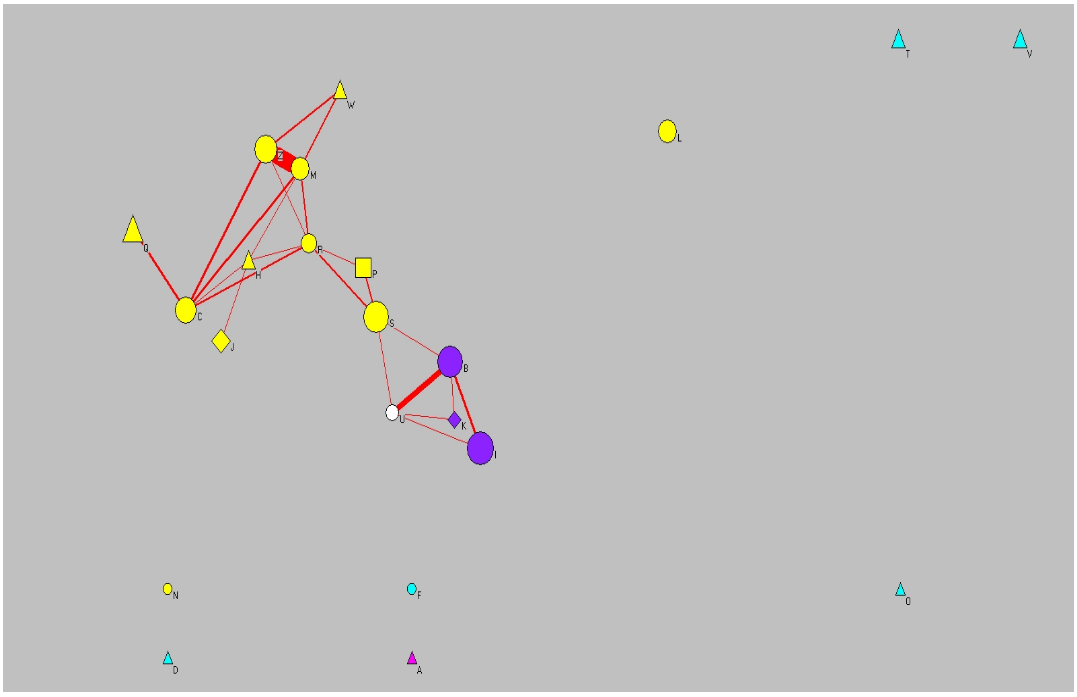


Figure 112 – *Co-authorship network of DIEG in 2011.*

The elimination of external components determinates a reduction in the number of ties (network cohesion accordingly decreases) and an increase in number of isolated authors, distinguishing the different characteristic respect to those authors considered before: in the first case being isolated means to be not co-authored only respect to authors that belong to DIEG in the second case being isolated means to be not co-authored with authors within and outside of DIEG.

6.6.1 Some measures of cohesion

Also for calculation of the following measures it was thought not to consider the authors A, D, F, G, N, O (see hypothesis 4).

Besides, the measures linked to number of authors and to number of ties show a reduction in dependence of their reduction.

Measures	2001 complete co-authorship	2001 DIEG co-authorship	2002 complete co-authorship	2002 DIEG co-authorship	2003 complete co-authorship	2003 DIEG co-authorship	2004 complete co-authorship	2004 DEIG co-authorship
Total number of authors	12 (1)	7	17 (4)	7 (1)	19 (4)	9 (1)	19 (4)	9 (1)
Total number of ties	15	1	28	1	31	6	31	6
Density	0.19	0.05	0.13	0.04	0.12	0.13	0.12	0.13
Average degree	2.28	0.30	2.60	0.28	2.64	1.17	2.64	1.17

Table 19 ***– Measures of complete and DIEG co-authorship networks from 2001 to 2004.***

Measures	2005 complete co-authorship	2005 DIEG co-authorship	2006 complete co-authorship	2006 DIEG co-authorship	2007 complete co-authorship	2007 DIEG co-authorship	2008 complete co-authorship	2008 DEIG co-authorship
Total number of authors	29 (7)	9 (4)	32 (7)	10 (4)	36 (7)	11 (4)	37 (7)	11 (4)
Total number of ties	53	9	62	12	68	13	72	15
Density	0.08	0.16	0.08	0.21	0.07	0.12	0.07	0.14
Average degree	2.80	1.92	3.15	2.73	3.16	1.68	3.27	1.96

Table 20 ***– Measures of complete and DIEG co-authorship networks from 2005 to 2008.***

Measures	2009 complete co-authorship	2009 DIEG co-authorship	2010 complete co-authorship	2010 DIEG co-authorship	2011 complete co-authorship	2011 DIEG co-authorship
Total number of authors	48 (7)	11 (5)	51 (8)	11 (6)	59 (12)	11 (6)
Total number of ties	115	17	125	18	157	23
Density	0.07	0.14	0.07	0.13	0.06	0.17
Average degree	4.18	2.10	4.23	2.08	4.42	2.72

Table 21 ***– Measures of complete and DIEG co-authorship networks from 2009 to 2011.***

Over time the number of authors and ties continue to be characterized by positive trends, although they vary between lower values than those calcualted for the network including external authors. In Figure 119 their trends are shown.

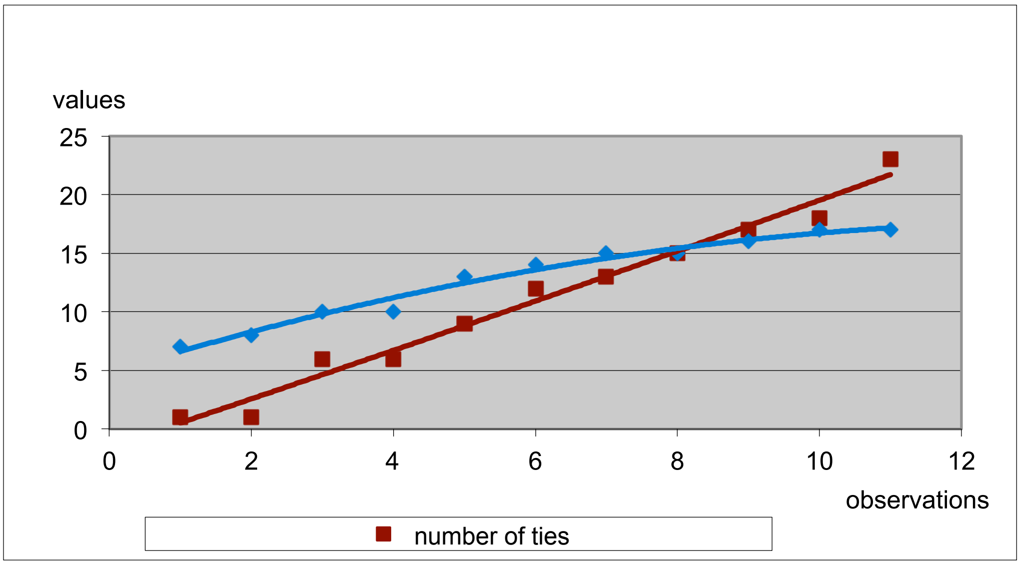


Figure 113 – *Trends of number of authors and their ties over time for DIEG network.*

The density presents a positive trend (from 0.05 to 0.17), while for complete co-authorship network was negative. This is probably due to the fact that the increase of number of authors and ties are less large compared to the case of networks including external authors. In fact, for DIEG networks the number of authors starts from value equal to 7 until to 17 while in network with external authors it varies between 13 and 71. The number of ties begins with 1 and becomes 23, while for networks with external authors it started from 14 until to 157. In Figure 120 the density trends for two cases are shown.

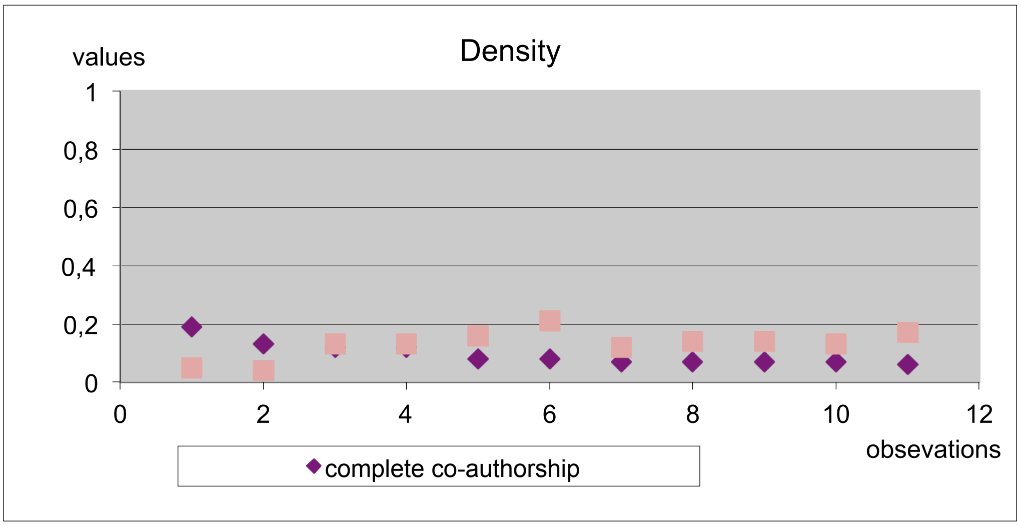


Figure 114 – *Trends of density over time for complete and DIEG co-authorship networks.*

The average degree has a positive trend: it starts from value 0.30, it means that the authors have an average of written papers much less than 1, reaches in the last observation the value of 2.72.

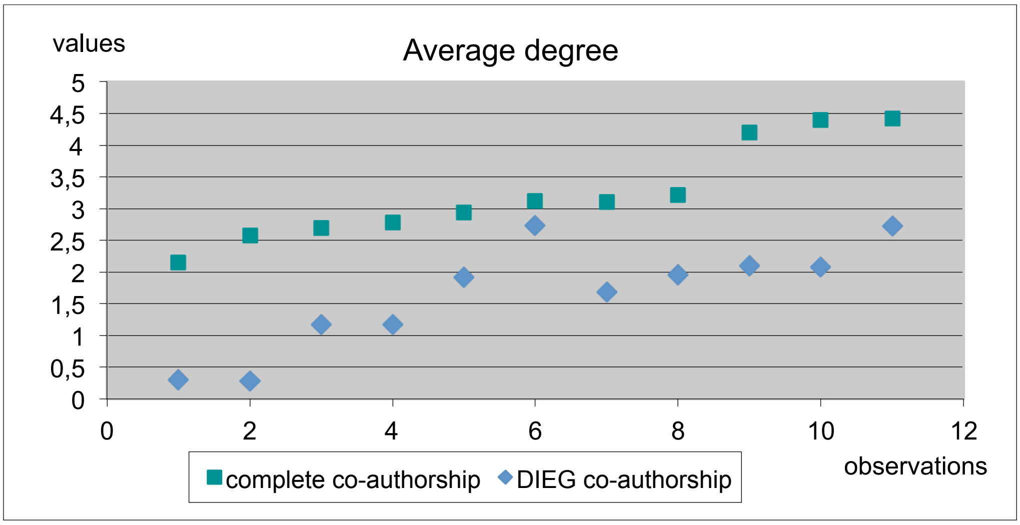


Figure 115 ***–*** *Trends of average degree over time for complete and DIEG co-authorship networks.*

The values of inclusiveness, presented in Tables 24-26, confirm an increase of cohesion for DIEG network.

Measures	2001 complete co-authorship	2001 DIEG co-authorship	2002 complete co-authorship	2002 DIEG co-authorship	2003 complete co-authorship	2003 DIEG co-authorship	2004 complete co-authorship	2004 DEIG co-authorship
Total number of connected authors	13	2	18	2	20	4	20	4
Total number of not connected authors	4	5	3	6	3	6	3	6
Inclusiveness index	0.76	0.29	0.85	0.25	0.86	0.40	0.86	0.40

Table 22 ***–*** *The inclusiveness values for complete and DIEG co-authorship networks from 2001 to 2004.*

Measures	2005 complete co-authorship	2005 DIEG co-authorship	2006 complete co-authorship	2006 DIEG co-authorship	2007 complete co-authorship	2007 DIEG co-authorship	2008 complete co-authorship	2008 DEIG co-authorship
Total number of connected authors	33	7	36	8	40	8	42	11
Total number of not connected authors	3	6	3	6	3	7	2	4
Inclusiveness index	0.91	0.54	0.92	0.57	0.93	0.53	0.95	0.73

Table 23 *– The inclusiveness values for complete and DIEG co-authorship networks from 2005 to 2008.*

Measures	2009 complete co-authorship	2009 DIEG co-authorship	2010 complete co-authorship	2010 DIEG co-authorship	2011 complete co-authorship	2011 DIEG co-authorship
Total number of connected authors	53	11	58	12	71	14
Total number of not connected authors	2	5	1	5	0	3
Inclusiveness index	0.96	0.69	0.98	0.71	1	0.82

Table 24 *– The inclusiveness values for complete and DIEG co-authorship networks from 2009 to 2011.*

In Figure 122, trends of inclusiveness for co-authorship networks with external authors and for DIEG co-authorship networks are displayed.

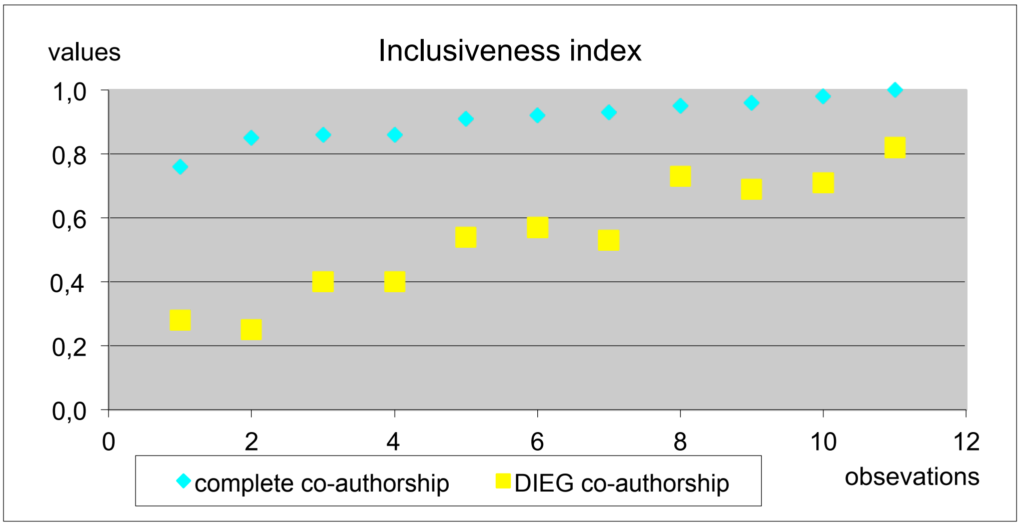


Figure 116 *– Trends of inclusiveness over time for complete and DIEG co-authorship networks.*

The values of inclusiveness are different as the number of authors that do not linked with others (authors linked with external authors) increases.

The results indicate that the networks composed by only DIEG authors have a number of ties (at the last observation a difference is equal to 135 ties) less than those of the networks of DIEG and external authors. This is coherent with result previous indicated that the co-writing of papers is few diffuse among the members of DIEG. However, over time the aggregation of network increases but its value remains very low.

6.6.2 Centrality indices

With reference to centrality indices, the results obtained are very different from those for network with external authors. In Table 27 the values of degree centrality are shown.

FREEMAN'S DEGREE CENTRALITY MEASURES:

	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011
B	0.000	0.000	0.000	0.000	1.000	1.000	1.000	2.000	3.000	6.000	13.000
C	1.000	1.000	4.000	4.000	4.000	9.000	9.000	11.000	12.000	12.000	12.000
H		0.000	0.000	0.000	0.000	2.000	2.000	2.000	2.000	2.000	3.000
I	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.000	2.000	2.000	4.000
J									0.000	0.000	0.000
K										0.000	2.000
L	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
M			3.000	3.000	5.000	9.000	9.000	9.000	10.000	13.000	15.000
P	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	2.000	4.000	4.000
Q							0.000	2.000	3.000	3.000	3.000
R	0.000	0.000	3.000	3.000	3.000	5.000	6.000	6.000	8.000	10.000	10.000
S			0.000	0.000	1.000	1.000	2.000	2.000	4.000	6.000	9.000
U							0.000	0.000	0.000	3.000	8.000
V					0.000	0.000	0.000	0.000	0.000	0.000	0.000
T					0.000	0.000	0.000	0.000	0.000	0.000	0.000
W						2.000	2.000	2.000	2.000	4.000	4.000
Z	1.000	1.000	4.000	4.000	6.000	7.000	7.000	7.000	8.000	11.000	11.000

***Table 25** – Degree centrality values of DIEG network in each observation.*

The most central authors are C, M, R, and Z. This result does not agree with the previous ones. In fact in that the most centrals authors were B, M, and Z. This means that C, and R are more integrated in DIEG network than in extented network, while B who, on the contrary, is more integrated in complete co-authorship network than in DIEG network (he/she has written many papers in collaboration with external authors). Besides, in the case of network that included the external authors, the less central author was K who now presents a value of degree centrality different to 0.

These results indicate that some researchers of DIEG have a good propension to work with external researchers, the others have an interaction that is much close within the department.

The situation is very close to that shown for network with external authors; it differs only for author I. This means that in the DIEG network the authors are reachable in the same way. It depends on the fact that external authors are not integrated in the DIEG researchers' network; in other word, their ties are not included in *knit*. Table 28 shows the values obtained regarding to closeness centrality.

CLOSENESS CENTRALITY

	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011
B	0.000	0.000	0.000	0.000	8.333	7.143	11.200	14.000	14.019	14.159	19.753
C	16.667	14.286	14.286	14.286	11.009	9.929	12.069	15.385	15.306	14.953	21.053
H		0.000	0.000	0.000	0.000	9.790	11.667	14.433	14.286	13.913	19.753
I	0.000	0.000	0.000	0.000	0.000	0.000	0.000	12.963	12.931	13.008	20.513
J									0.000	0.000	17.204
K										0.000	17.391
L	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
M			14.286	14.286	11.111	10.000	12.174	15.385	15.306	14.953	21.918
P	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	14.706	14.545	19.512
Q							0.000	14.141	14.019	13.675	18.182
R	0.000	0.000	14.286	14.286	11.009	9.859	12.174	15.556	15.625	15.385	21.622
S			0.000	0.000	8.333	7.143	11.765	14.894	15.000	14.953	20.513
U							0.000	0.000	0.000	13.008	18.605
V					0.000	0.000	0.000	0.000	0.000	0.000	0.000
T					0.000	0.000	0.000	14.286	0.000	0.000	0.000
W					10.909	9.790	11.667	14.286	14.151	13.793	19.048
Z	16.667	14.286	14.286	14.286	11.111	9.929	12.069	15.217	15.152	14.815	20.513

Table 26 *– Closeness centrality values of DIEG network in each observation.*

Table 29 summarizes the values for betweenness centrality.

FREEMAN BETWEENNESS CENTRALITY

	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011
B	0.000	0.000	0.000	0.000	0.000	0.000	0.000	8.000	9.000	19.000	12.500
C	0.000	0.000	0.000	0.000	0.000	0.000	2.000	10.500	12.000	13.500	16.667
H			0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	12.000
I	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	12.000
J									0.000	0.000	0.000
K										0.000	0.000
L	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	15.000
M			0.000	0.000	1.000	1.000	5.000	6.500	7.500	8.500	26.500
P	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Q							0.000	0.000	0.000	0.000	0.000
R	0.000	0.000	0.000	0.000	0.000	0.000	10.000	18.000	24.000	30.000	23.500
S			0.000	0.000	0.000	0.000	6.000	14.000	16.000	24.000	15.500
T					0.000	0.000	0.000	0.000	0.000	0.000	0.000
U						0.000	0.000	0.000	0.000	0.000	2.500
V					0.000	0.000	0.000	0.000	0.000	0.000	0.000
W					0.000	0.000	0.000	0.000	0.000	0.000	0.000
Z	0.000	0.000	0.000	0.000	1.000	1.000	2.000	3.000	3.500	4.000	2.833

Table 27 *– Betweenness centrality values of DIEG network in each observation.*

The most central authors are B, C, R, and S. This result agrees with the previous ones as in extend co-authorship networks the most centrals authors were B, R, and S. For both kinds of network, in the last observation the number of the most central authors extends. In fact, in collaborative networks they correpond to B, C, H, I, L, M, R, and S. The authors H belongs for the first time to groups of the most centrals authors and this means that he/she has ties that survive over time.

6.6.3 Dynamic analysis

The co-authorship networks of DIEG do not included external authors so each author is characterized by disciplinary sector in which he/she works and level of his/her career held in each observation. In this sense, the choice of effects that must be included in vary models has been driven by these attributes.

In Table 30, the annual changes are reported.

Changes	2001-02	2002-03	2003-04	2004-05	2005-06	2006-07	2007-08	2008-09	2009-10	2010-11
Network										
Joined	1	2	0	3	1	1	0	1	1	0
Left	0	0	0	0	0	0	0	0	0	0
Ties										
0 -> 1	0	5	0	3	3	1	2	2	1	5
1 -> 1	1	1	6	6	9	12	13	15	17	18
1 -> 0	0	0	0	0	0	0	0	0	0	0

Table 28 ***–*** *Changes in DIEG co-authorship networks from 2001 to 2011.*

As for the case of complete co-authorship networks, the networks are characterized by the entry of several authors, no authors leave them, and eliminated ties are zeros. The number of created ties increases over time but it is lower than that corresponding to complete co-authorship networks. The number of maintained ties is much lower than that of complete case because of possibility to eliminate ties.

Four models described before have been entered in SIENA and, in Table 31 the results obtained are shown.

Effects	Model 1	Model 2	Model 3	Model 4
Rate parameters				
$r_{2001\text{-}2006}$	0.80	0.77	0.73	0.70
$r_{2006\text{-}2011}$	0.90	0.82	0.74	0.73
Networks measures				
Balance		2.17	3.60	3.56
Authors characteristics				
Disciplinary sector similarity			13.31	13.77
Career ego x career alter				0.04

Table 29 ***–*** *Models predicting DIEG network from 2001 to 2011.*

Model 1: simply the *rate parameters* have been included. The results of these parameters indicate that the change rate decreases from first interval of time to the next.

Model 2: the effect of *balance* has been added; the *balance* is a positive effect but it is low value, so in a little way the authors tend to form ties with whom they share other ties.

Model 3: the effect of *disciplinary sector similarity* has been added. The *disciplinary sector similarity* is used to test the tendency of authors who have similar professional experience to create a new co-authorship tie or who have different professional experience. The value obtained is positive and very high; in fact, it represents the effect that has the most weight respect to the other effects considered, and this means that the authors tend to create ties with other authors that work in the same disciplinary sector. This result has been underlined from observations.

Model 4: finally, the *career ego x career alter* has been entered. The *level of carrier ego x alter* is used to test if the tendency of authors with a higher level of carrier to prefer or not the ties to others who likewise have a relatively high professional level. The corresponding value is positive though it is very low; this means that a little number of ties is created among authors with the similar carrier level. Observing the networks over time, the structures existing are mainly characterized by authors that are only Professors; besides, there are few other structure composed by authors with different levels of carrier (Professors plus Assistant Professors or Professors plus PhD students). This last parameter tests if the authors that have a high level of career are more attractive that those with less level of career.

7. The Case Study: collaborative networks

A collaborative network, like co-authorship network, is a social network consisting made up of a set of researchers and a set of ties among them. The difference of collaborative ties consists in the fact that they take into account the continuity of scientific production.

Raffaella Cicala.

7.1 Collaborative networks over study period

The collaborative networks over study period have been built in which components of DIEG (Full and Associate professors, PhD students, Assistant professor) and people with they collaborate represent actors of network, while the collaborations between them and with other people, are ties. In collaborative networks, two or more authors are linked if they have published two or more papers and the interval in time between the publication of these two or more articles must not exceed 5 years (see hypothesis 1 in chapter 6.3). Accordingly, in this case elimination of ties is allowed, and each author can decide to change or not change its ties, that is he/she can create, eliminate, or maintain his/her ties.

As co-authorship network, each node has been characterized by shape (it represents the professional rank), by color (it represents disciplinary sector and institutional affiliation), and by size (scaled by h-index).

The tie color is blue (it is different from that chosen for co-authorship), and its strength represents the frequency with which authors collaborate and it defines the quality of scientific relationship between them.

In Figure 123, the collaborative network is shown, in which the color of nodes is the same of that assigned for co-authored network as it indicates the disciplinary afference, while the color and meaning of links is different as its meaning is different. So it blue has been chosen to represent collaboration between two or more researchers. External authors are represented by white color because their disciplinary sector is not considered.

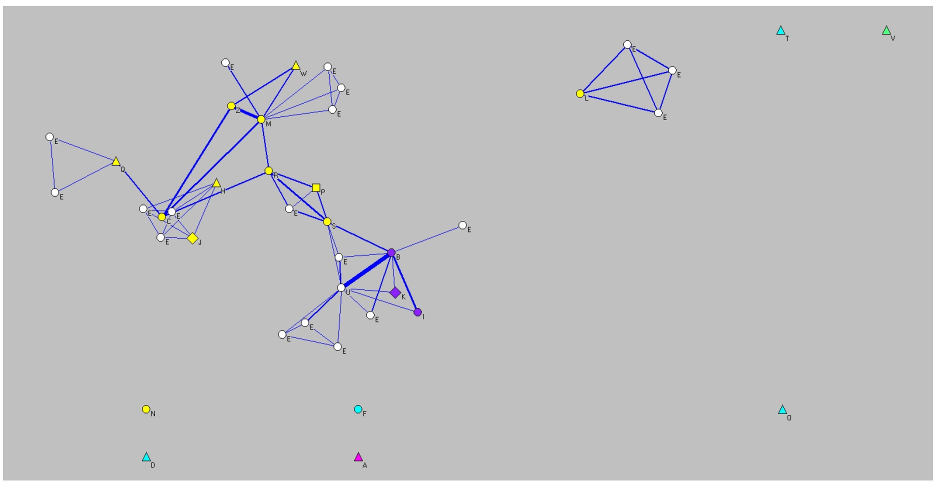


Figure 117 – *Collaborative network in observation interval (2001-2011).*

In the last observation, 41 authors are considered, and co-author ties among them are 57 (these ties do not include the ties that external authors have with other externals authors).

Table 32 shows the number of papers, authors, and papers with one author for each observation.

Year	Papers	Authors	Paper one author
2001	5	17	1
2002	8	20	3
2003	2	15	1
2004	1	16	1
2005	5	27	0
2006	3	25	0
2007	6	26	2
2008	6	24	2
2009	9	34	2
2010	9	31	4
2011	10	41	2

Table 30 *– Number of papers and authors over study period.*

To make a visual comparison, the sequence of collaborative networks (right side), from 2001 to 2011, are displayed with co-authorships networks (left side) obtained before.

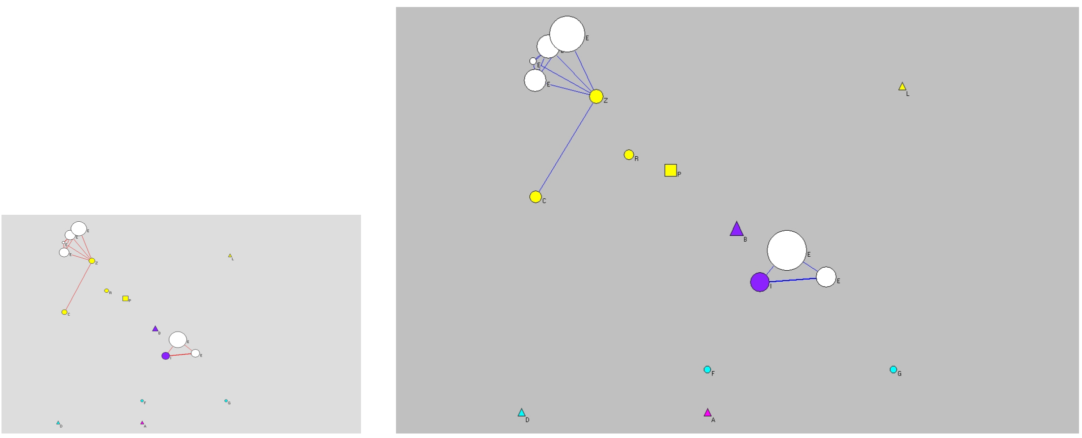


Figure 118 – *Co-authorship network (smaller figure) and collaborative network in 2001.*

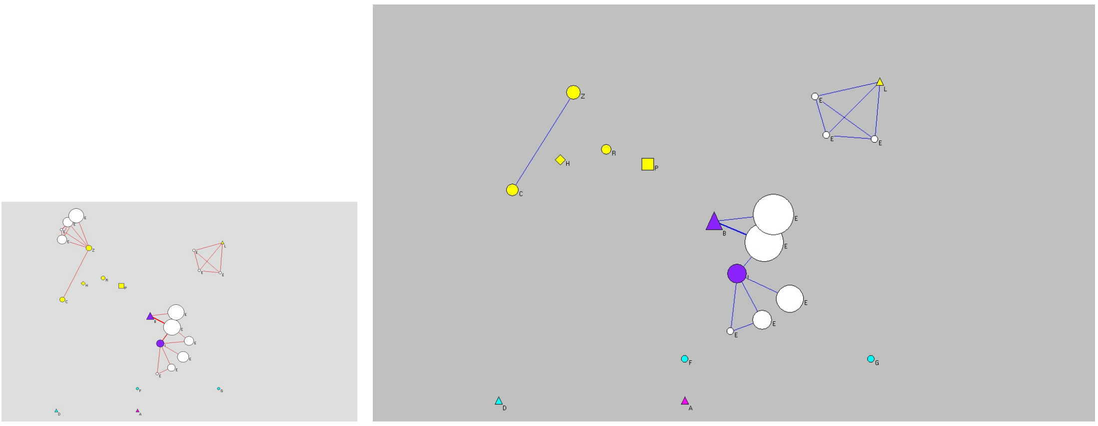


Figure 119 – *Co-authorship network (smaller figure) and collaborative network in 2002.*

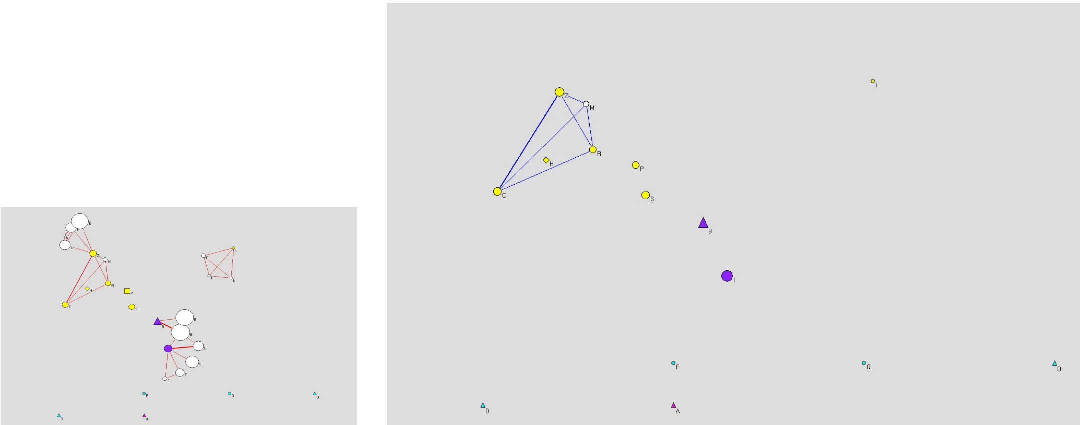


Figure 120 – *Co-authorship network (left side) and collaborative network (right side) in 2003.*

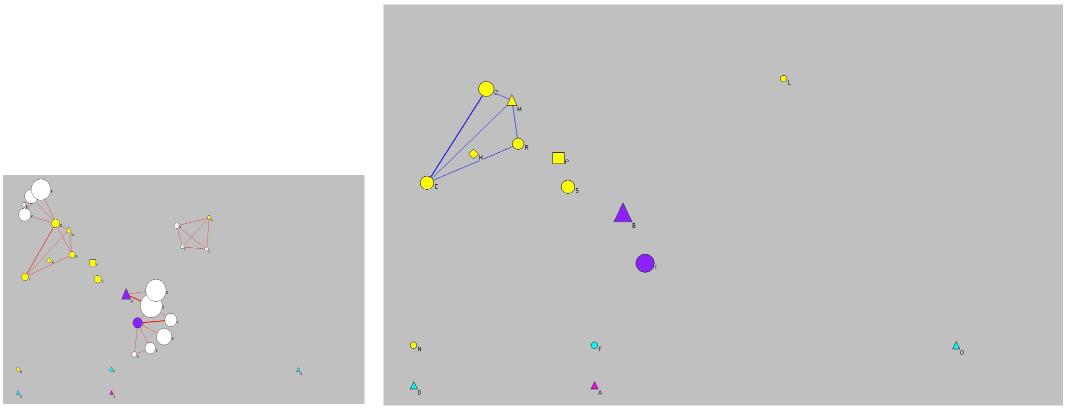


Figure 121 – *Co-authorship network (smaller figure) and collaborative network in 2004.*

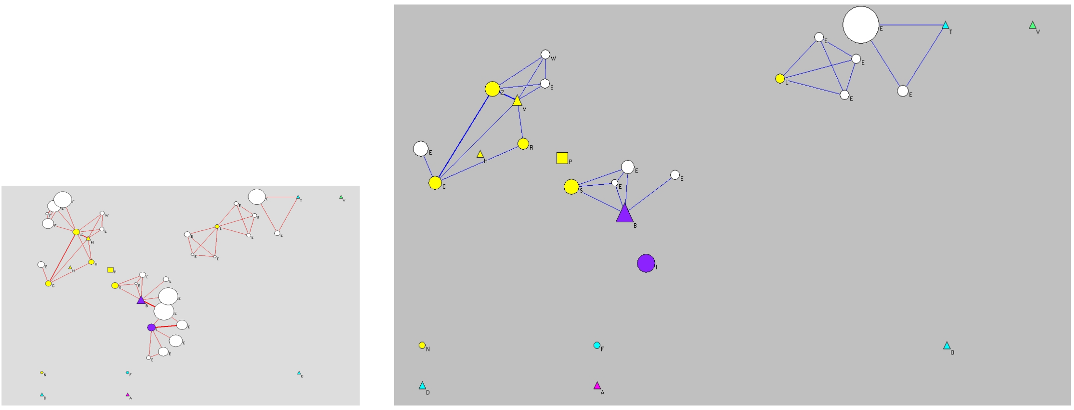


Figure 122 – *Co-authorship network (smaller figure) and collaborative network in 2005.*

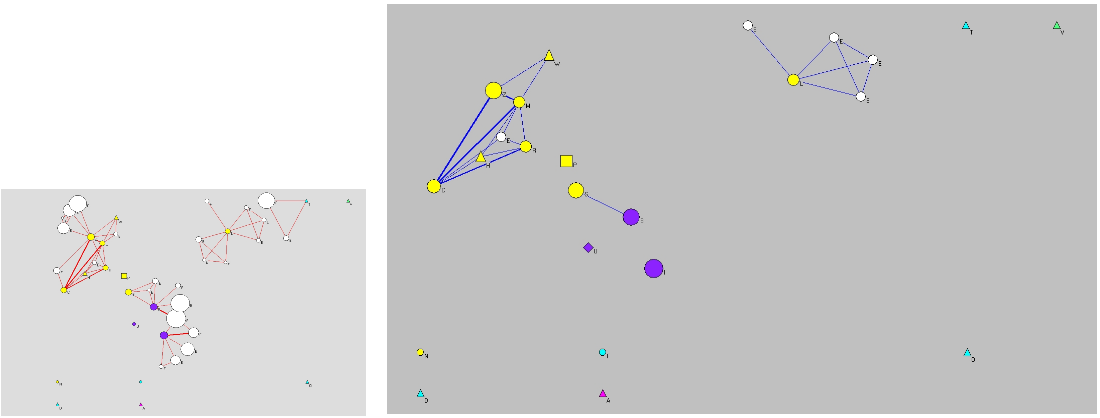


Figure 123 – *Co-authorship network (smaller figure) and collaborative network in 2006.*

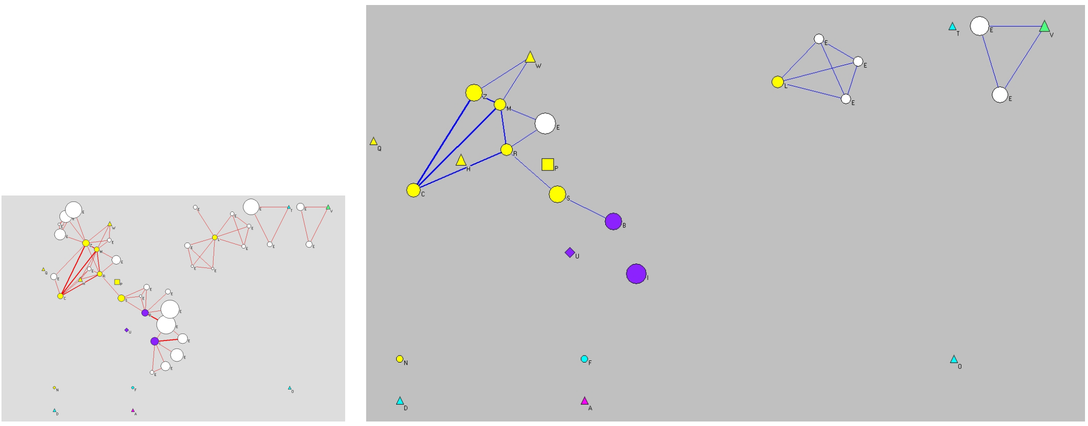


Figure 124 – *Co-authorship network (smaller figure) and collaborative network in 2007.*

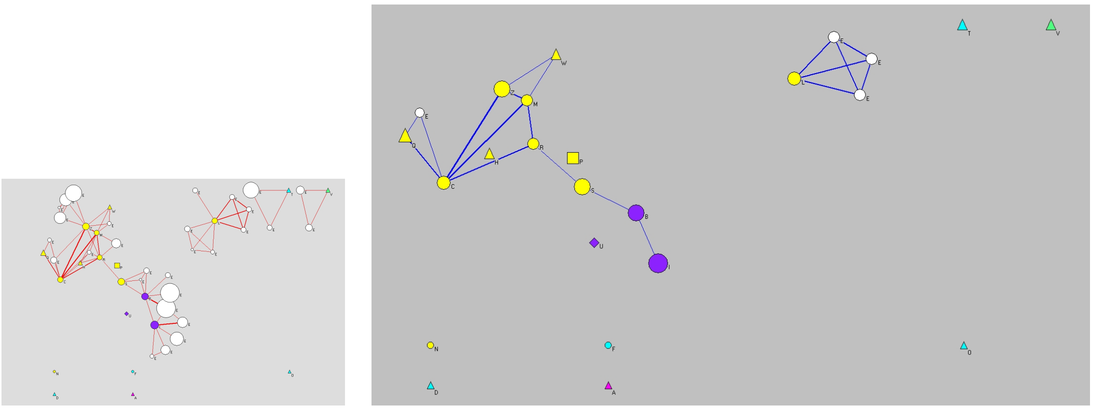


Figure 125 – *Co-authorship network (smaller figure) and collaborative network in 2008.*

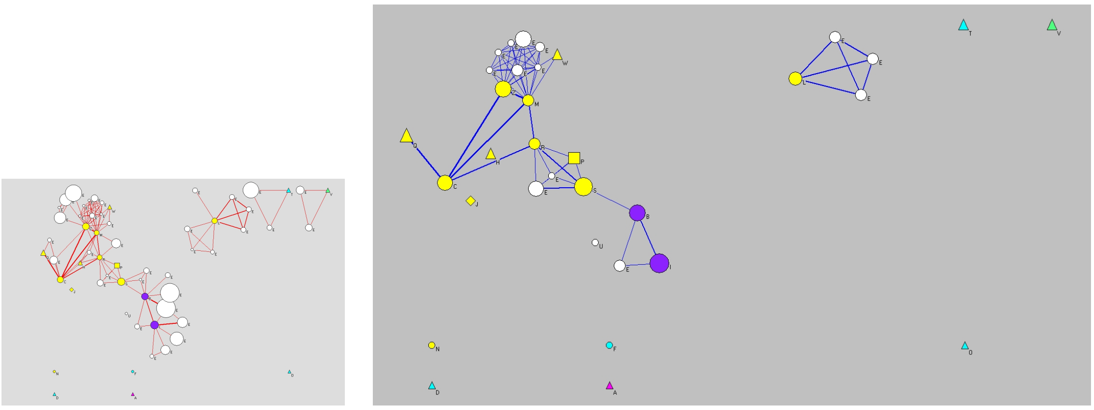


Figure 126 – *Co-authorship network (smaller figure) and collaborative network in 2009.*

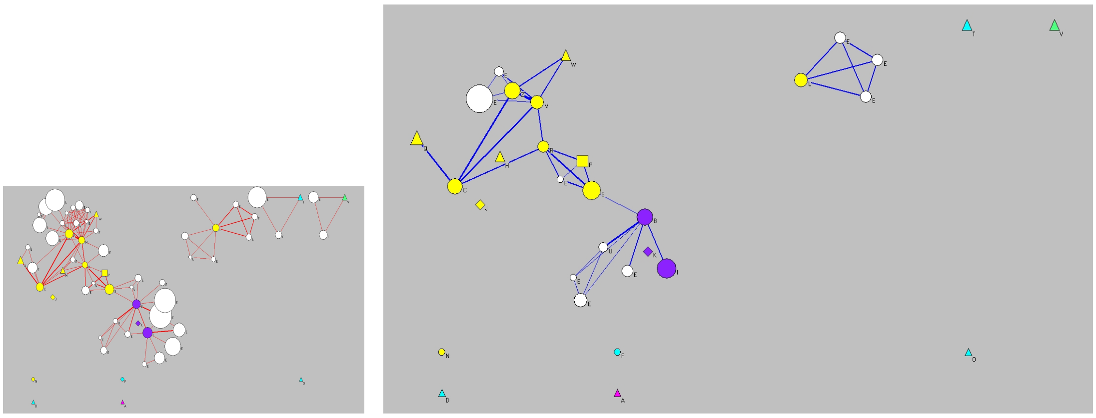


Figure 127 – *Co-authorship network (smaller figure) and collaborative network in 2010.*

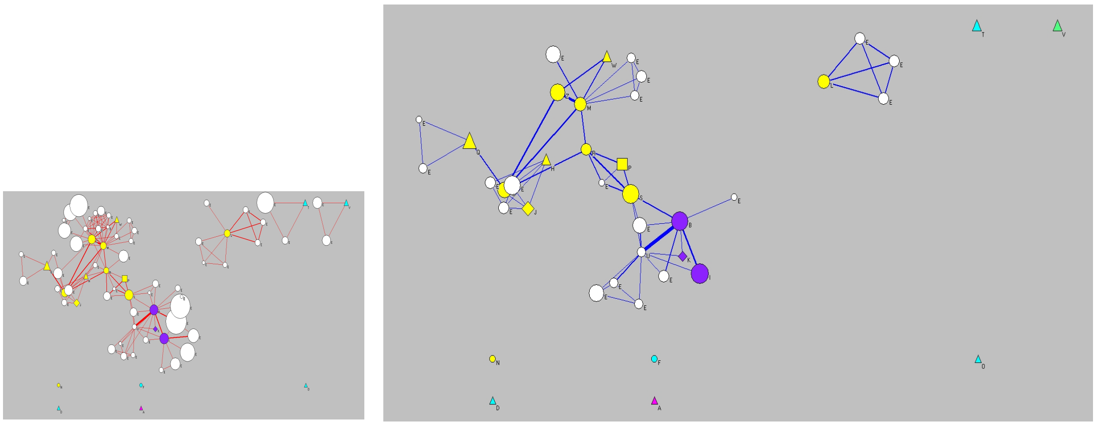


***Figure 128 –** Co-authorship network (smaller figure) and collaborative network in 2011.*

Respect to the case of co-authorship networks, there is not a uniform growth: in some observations ties are formed, in others ties are eliminated and, then, in others they remain the same. In the complex, in collaborative networks a large number of ties does not appear.

The snapshots of collaborative networks show that exists a group (A, D, F, G, N, O), the same group identified in the case of co-authorship networks, composed by isolated researchers. Thus, in the next the measures have been calculated without to consider the researchers of this group.

7.1.1 Measures of cohesion

In Table 33, the values obtained for collaborative networks are displayed.

Measures	2001 *(n=13)*	2002 *(n=16)*	2003 *(n=10)*	2004 *(n=10)*	2005 *(n=23)*	2006 *(n=19)*	2007 *(n=21)*	2008 *(n=19)*	2009 *(n=29)*	2010 *(n=26)*	2011 *(n=36)*
Total number of authors	12 (1)	14 (2)	9 (1)	9 (1)	17 (6)	15 (4)	17 (4)	15 (4)	24 (5)	19 (7)	24 (12)
Total number of ties	15	15	6	5	27	21	20	19	62	34	59
Density*	0.19	0.12	0.13	0.11	0.10	0.12	0.09	0.11	0.15	0.10	0.09
Average degree*	2.28	1.80	1.17	0.99	2.20	2.16	1.80	1.98	4.20	2.50	3.15
Clustering coefficient of Wattz-Strogatz	0.95	0.93	1.00	1.00	0.91	0.82	0.79	0.62	0.86	0.73	0.82
Transitivity	0.73	0.75	1.00	0.50	0.73	0.43	0.40	0.26	0.26	0.27	0.30

*Calculated by formula for weighted undirected graphs.

***Table 31 –** Measures of collaborative networks without isolated authors over time.*

All values increase and decrease in no-uniform way as external researchers, and papers are not *cumulates*, they can operate in one observation and disappear in the successive observation. Accordingly, also measures follow trends that increase and decrease during study period.

Also, the density is not characterized by uniform trend: it starts from 0.18 but it becomes 0.09 in the last observation, varying between higher and lower values. This depends on the fact that the collaboration is closely linked with external people.

Regarding average degree, values found trace a no uniform trend too. The first value is 2.16, after it decreases and increases and, then, decreases. In 2009 average degree reaches its maximum value due to the entry of a groups of 7 external people.

The coefficient clustering starts with high value, and, in third and fourth observation (in which unique researchers that are linked belong to the same single cluster), it reaches its maximum values. After, coefficient clustering decreases but, however, it maintains high value.

Until 2005 the main part of present triplets is transitivity, while, to start from 2008, triplets become no transitive.

Finally, for each observation the inclusiveness index has been calculated, and found values, presented in Table 34, are high. Also in collaborative networks the researchers result to have a high degree of connection.

Measures	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011
Total number of authors connected with others	9	13	4	4	19	14	15	14	24	21	34
Total number of authors not connected with others	4	3	6	6	4	5	6	5	5	5	2
Inclusiveness index	0.60	0.81	0.40	0.40	0.82	0.73	0.71	0.73	0.82	0.80	0.94

***Table 32** – The inclusiveness values over time.*

7.1.2 Centrality indices

In Table 35, the values of degree centrality for collaborative networks are shown.

FREEMAN DEGREE CENTRALITY

	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011
B	0.000	2.000	0.000	1.000	3.000	1.000	1.000	2.000	4.000	8.000	17.000
C	1.000	1.000	3.000	4.000	5.000	10.000	8.000	11.000	11.000	11.000	10.000
H			0.000	0.000	0.000	3.000	0.000	0.000	0.000	0.000	4.000
I	3.000	3.000	0.000	0.000	0.000	0.000	0.000	1.000	1.000	2.000	4.000
J									0.000	0.000	4.000
K										0.000	2.000
L	0.000	3.000	0.000	0.000	3.000	4.000	3.000	6.000	6.000	6.000	6.000
M			3.000	3.000	6.000	9.000	9.000	8.000	16.000	15.000	17.000
P	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	4.000	4.000	4.000
Q							0.000	3.000	3.000	3.000	4.000
R	0.000	0.000	3.000	3.000	2.000	5.000	6.000	5.000	9.000	11.000	9.000
S			0.000	0.000	3.000	1.000	2.000	2.000	7.000	8.000	11.000
T					2.000	0.000	0.000	0.000	0.000	0.000	0.000
U						0.000	0.000	0.000	0.000	5.000	16.000
V					0.000	0.000	2.000	0.000	0.000	0.000	0.000
W					3.000	2.000	2.000	2.000	2.000	4.000	4.000
Z	5.000	1.000	4.000	4.000	6.000	6.000	6.000	6.000	14.000	12.000	10.000

Table 33 *– Degree centrality values for collaborative networks over study period.*

The most centrals authors are B, C, M, and Z. so, for these researchers, ties established are durative.

Passing to closeness centrality, in Table 36 the found values are indicated.

CLOSENESS CENTRALITY

	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011
B	0.000	7.143	0.000	10.000	8.333	5.556	6.402	9.137	8.805	9.124	7.625
C	12.000	6.667	14.286	12.500	6.213	7.660	6.176	9.626	9.333	9.091	7.642
H		0.000	0.000	0.000	0.000	7.595	0.000	0.000	0.000	0.000	3.125
I	9.091	7.692	0.000	0.000	0.000	0.000	0.000	8.780	8.358	8.651	7.322
J									0.000	0.000	3.125
K										0.000	7.322
L	0.000	7.692	0.000	0.000	5.263	6.667	5.263	6.250	3.846	4.348	3.030
M			14.286	12.500	6.231	7.692	6.604	9.574	9.556	9.158	7.675
P	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	9.121	9.058	7.609
Q							0.000	9.278	8.805	8.621	7.338
R	0.000	0.000	14.286	12.500	6.176	7.627	6.604	9.626	9.492	9.328	7.795
S			0.000	0.000	5.263	5.556	6.522	9.424	9.211	9.294	7.795
U							0.000	0.000	0.000	8.772	7.642
V					0.000	0.000	5.000	0.000	0.000	0.000	0.000
T					5.000	0.000	0.000	0.000	0.000	0.000	0.000
W					6.176	7.563	6.502	9.231	9.032	8.711	7.322
Z	12.500	6.667	14.286	12.500	6.213	7.595	6.522	9.375	9.302	8.834	7.384

Table 34 *– Closeness centrality values for collaborative networks over study period.*

Three groups of the most central researchers have been identified: one composed by C, and Z, another formed by M, and R, and, finally, that constituted by B, and S.

For betweenness centrality, in Table 37 values found are summarized.

FREEMAN BETWEENNESS CENTRALITY

	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011
B	0.000	0.000	0.000	0.000	0.000	0.000	0.000	8.000	34.000	62.000	61.500
C	0.000	0.000	0.000	0.500	5.500	1.833	1.500	16.000	22.000	20.000	70.000
H			0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
I	0.000	2.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
J									0.000	0.000	0.000
K										0.000	0.000
L	0.000	0.000	0.000	0.000	0.000	3.000	0.000	0.500	15.000	15.000	0.000
M			0.000	0.500	4.500	5.333	7.500	7.500	79.643	39.000	106.000
P	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Q							0.000	0.000	0.000	0.000	44.000
R	0.000	0.000	0.000	0.000	0.000	0.333	10.500	18.000	84.000	63.500	143.000
S			0.000	0.000	0.000	0.000	6.000	14.000	48.000	60.500	153.000
T					0.000	0.000	0.000	0.000	0.000	0.000	0.000
U						0.000	0.000	0.000	0.000	28.000	79.500
V					0.000	0.000	0.000	0.000	0.000	0.000	0.000
W					0.000	0.000	0.000	0.000	0.000	0.000	0.000
Z	4.000	0.000	0.000	0.000	2.000	0.500	0.500	1.500	11.643	4.000	2.000

Table 35 *– Betweenness centrality values for collaborative networks over study period.*

The most central researchers are M, R, S, and U as they are characterized by the highest values of betweenness centrality; in the last observation, they reach their highest values.

7.1.3 Identification of cliques

As the case of co-authorship network, the last part of static analysis consists into identify cliques present.

In Table 38, the number of cliques identified and the maximum size of them are indicated.

Measures	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011
Number of cliques	2	3	1	2	11	4	5	4	6	7	11
Number of cliques with maximum size	1	1	1	2	1	3	1	1	1	3	1
Maximum size of cliques	5	4	4	3	5	4	4	4	9	4	5

Table 36 *– Number of cliques identified over study period.*

For each observation, the cliques present in collaborative networks (blue colored circles) are shown.

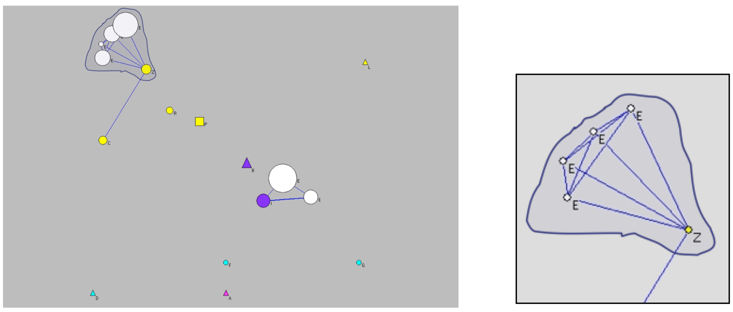


Figure 129 *– Cliques of collaborative networks and their details in 2001.*

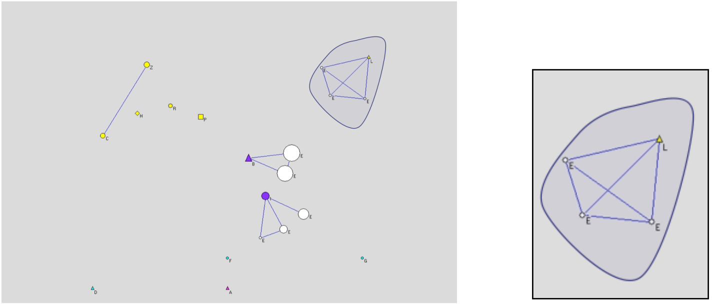


***Figure 130** – Cliques of collaborative networks and their details in 2002.*

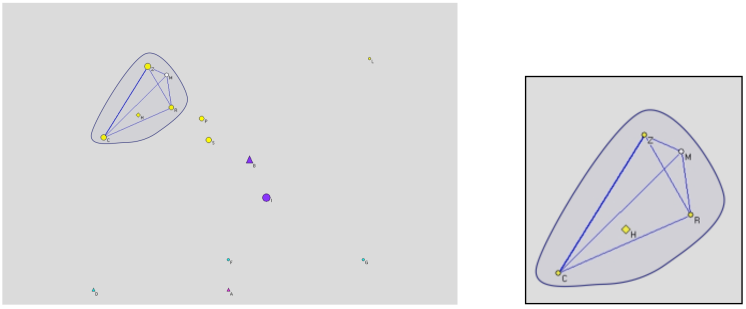


***Figure 131** – Cliques of collaborative networks and their details in 2003.*

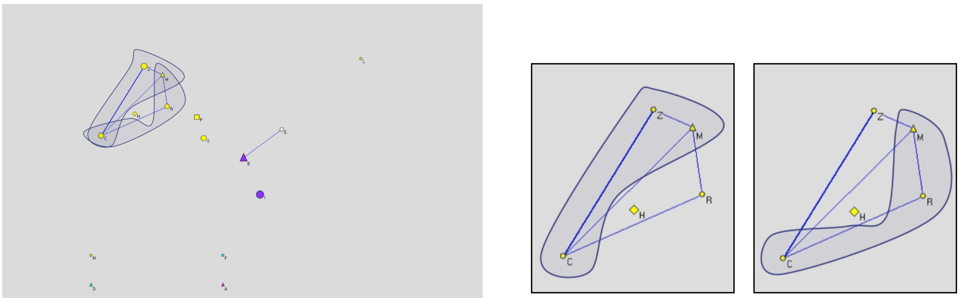


***Figure 132** – Cliques of collaborative networks and their details in 2004.*

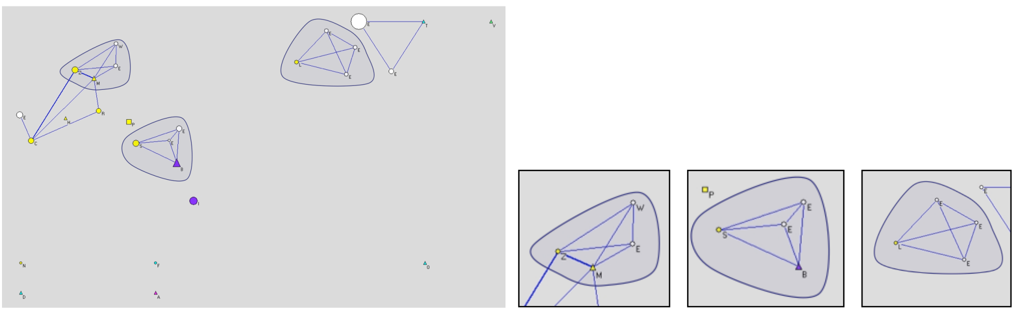


***Figure 133** – Cliques of collaborative networks and their details in 2005.*

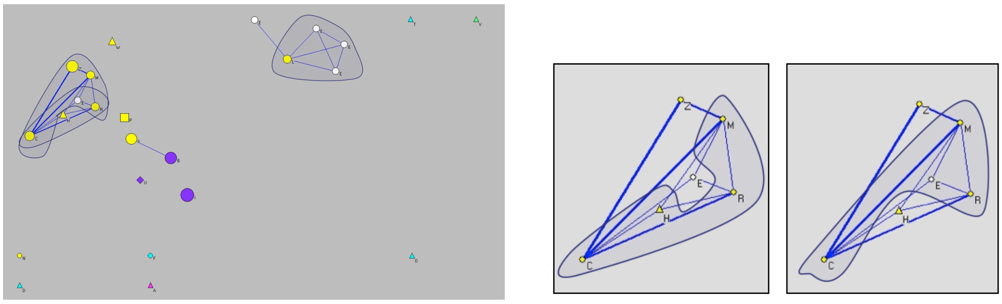


Figure 134 *– Cliques of collaborative networks and their details in 2006.*

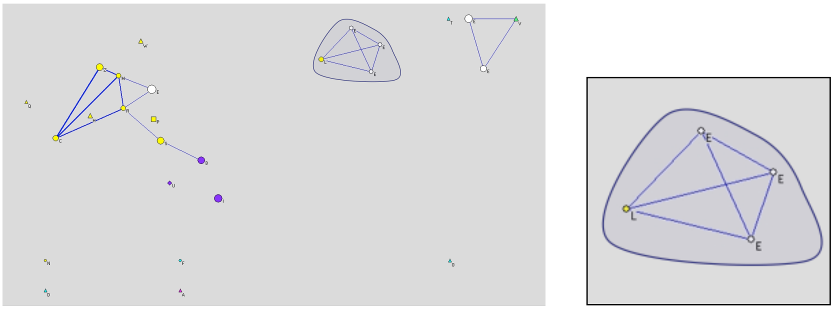


Figure 135 *– Cliques of collaborative networks and their details in 2007.*

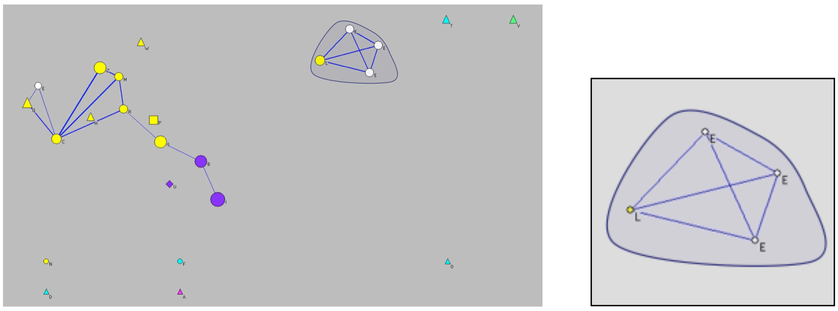


Figure 136 *– Cliques of collaborative networks and their details in 2008.*

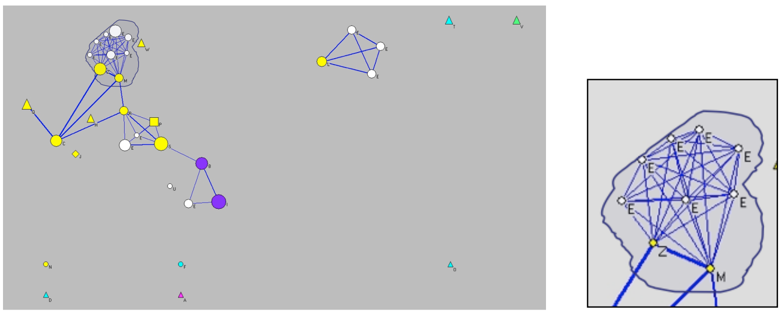


Figure 137 *– Cliques of collaborative networks and their details in 2009.*

Figure 138 *– Cliques of collaborative networks and their details in 2010.*

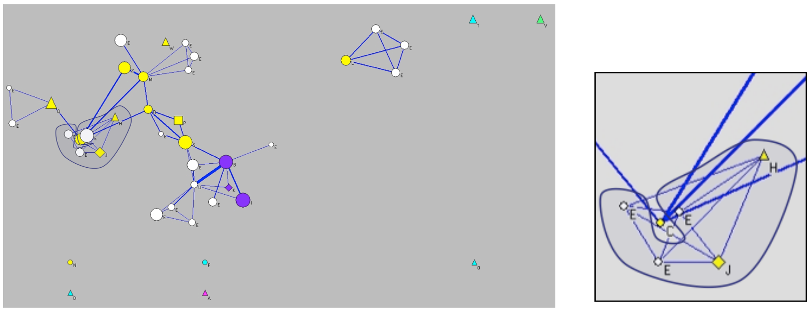


Figure 139 *– Cliques of collaborative networks and their details in 2011.*

7.2 Dynamic analysis: results found

The next step consists into identify the effects that drive collaborative network evolution. In the following, longitudinal data have been used to study network evolution and to catch the mechanism that drive this evolution.

For hypothesis 3 (see paragraph 6.3), the isolated authors A, D, F, G, N, O have not been considered.

As for the co-authorship network, it starts with a first model (simple model) that contains fewer (also only one) effects; successive, on the basis of results obtained, this first model has been complicated through the progressive addition of others effects until to all chosen effects are included. The effects to include in the models are the same of those chosen for the co-authorship networks as the network typology and characteristics are the same.

Thus, *degree* (this effect is as a default in the model), *balance, Same home institution, H-index similarity* effects have been considered and four models have been sequentially run by SIENA.

Model 1: only constant rate parameters corresponding to each transition have been included.

Model 2: two parameters, *outdegree* and *balance*, have been added.

Model 3: *Same home institution* parameter has been inserted.

Model 4: *H-index similarity* effect has been added.

For explain the dynamic of changes the Pairwise conjunctive model has been chosen.

In Table 39, the annual changes between subsequent observations are shown.

Changes		2001-02	2002-03	2003-04	2004-05	2005-06	2006-07	2007-08	2008-09	2009-10	2010-11
Network											
	Joined	9	2	0	13	3	4	1	11	5	12
	Left	6	8	0	0	6	2	3	1	8	2
Ties											
	0 -> 1	22	11	19	36	22	26	24	86	12	96
	1 -> 1	8	1	1	12	16	10	10	30	44	44
	1 -> 0	20	29	11	11	32	28	26	4	72	72

Table 37 ***–*** *Annual changes in collaborative ties in study period.*

In almost all periods of observation the number of researchers that left the networks is different to zero as well as the number of eliminated ties (indicated by 1 -> 0). Besides, the number of maintained ties (indicated by 1 -> 1) is high enough, and this indicates that there are many durative ties. Besides, remembering that collaborative ties are not cumulative, the number of created collaborative ties increases over time.

In definitive, the collaborative networks are characterized by researchers that join or left them, and that create or eliminate or maintain their ties.

In Table 40, weights of selected effects are summarized.

The rate parameter $r_{2001\text{-}2006}$ corresponding to transition 2001 - 2006 is smaller than that of transition 2006 - 2011. In particular, in all models it remains about 2.5, so in the first transition there is an average of 2.5 changes per researcher, while $r_{2006\text{-}2011}$ is about 6.5, so in the second transition the average changes became 6.5. So, in 2006-2001 researchers operate many more changes respect to 2001-2006.

Effects	Model 1	Model 2	Model 3	Model 4
Rate parameters				
$r_{2001\text{-}2006}$	2.54	2.56	2.57	2.21
$r_{2006\text{-}2011}$	6.54	6.84	6.56	5.56
Networks measures				
Outdegree (density)		- 3.22	- 3.12	- 3.41
Balance		- 18.87	- 19.68	- 20.99
Authors characteristics				
Same home institution			- 0.17	- 0.24
H-index similarity				5.56

Table 38 *– Models predicting collaborative network from 2001 to 2011.*

In all models, the values of *outdegree* parameter are negative, and this depends on costs of the formation of a tie in terms of time, effort, and resources.

Also the values of *balance* result negative. Probably, the elimination of some ties not durative causes a reduction of closure. Thus, the *balance* represents the effect that weighs in a negative way respect to all others.

With reference to *same home institution* parameter, negative values can be thinking that the authors tend mainly to create new ties with authors that do not belong to same institution.

Finally, only *H-index similarity* parameter presents a positive value, this means that this represents the only effect that plays a positive role on network evolution. So, the researchers form ties with others that are characterized by similar H-index values.

7.3 Collaborative networks of DIEG

As in the case of co-authorship networks, in order to investigate on dynamics of DIEG, the sub-networks of the collaborative networks have been realized without to consider external components.

In the following, the DIEG collaborative networks for each observation are shown (Figures 146-156) and measures calculated for extended networks have been recalculated and briefly described.

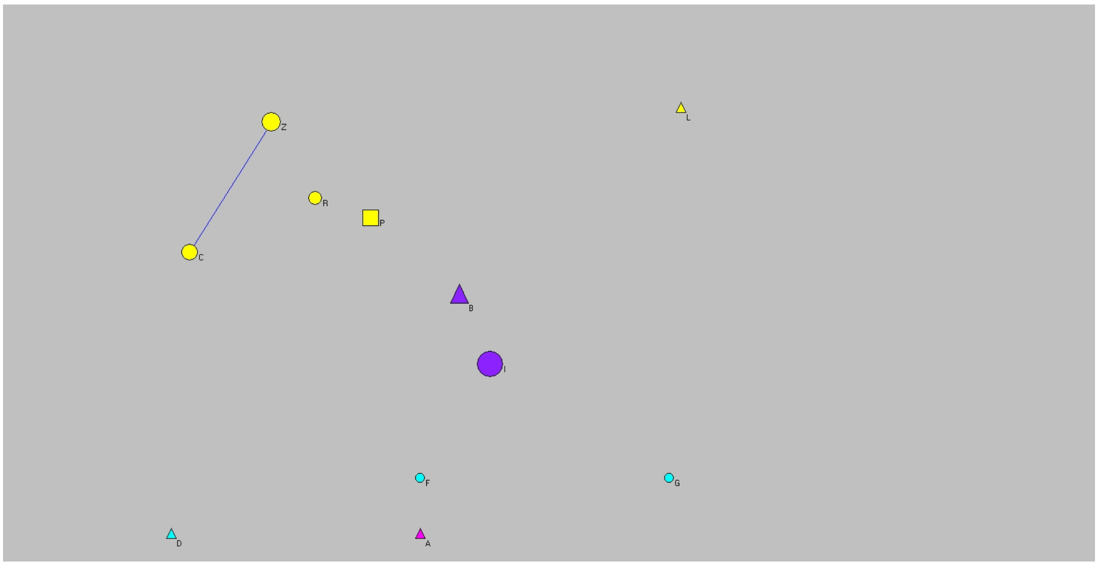


Figure 140 *– Collaborative DIEG network in 2001.*

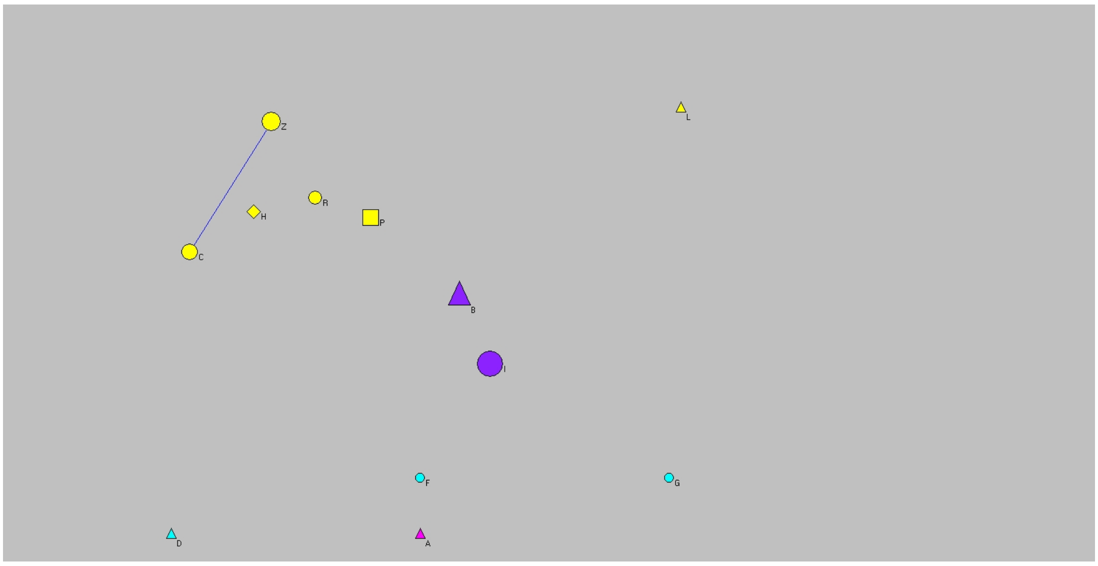


Figure 141 *– Collaborative DIEG network in 2002.*

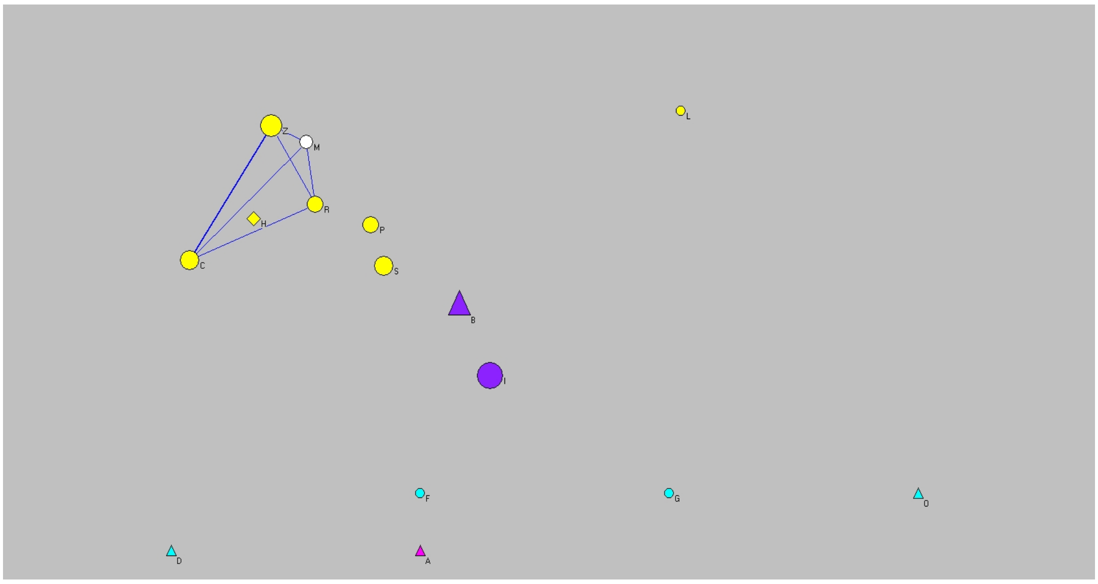


Figure 142 ***–*** *Collaborative DIEG network in 2003.*

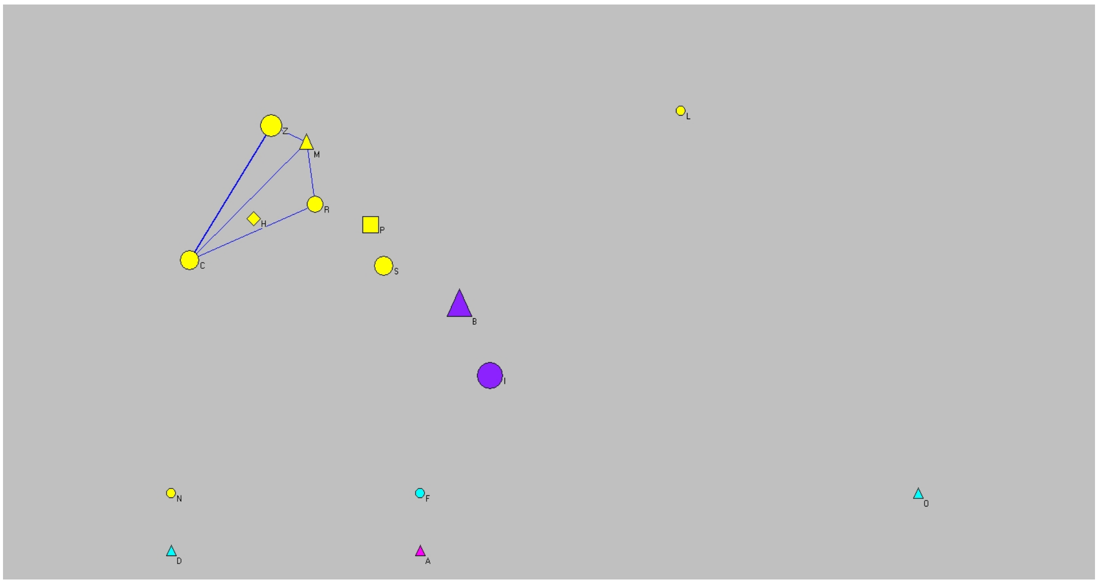


Figure 143 ***–*** *Collaborative DIEG network in 2004.*

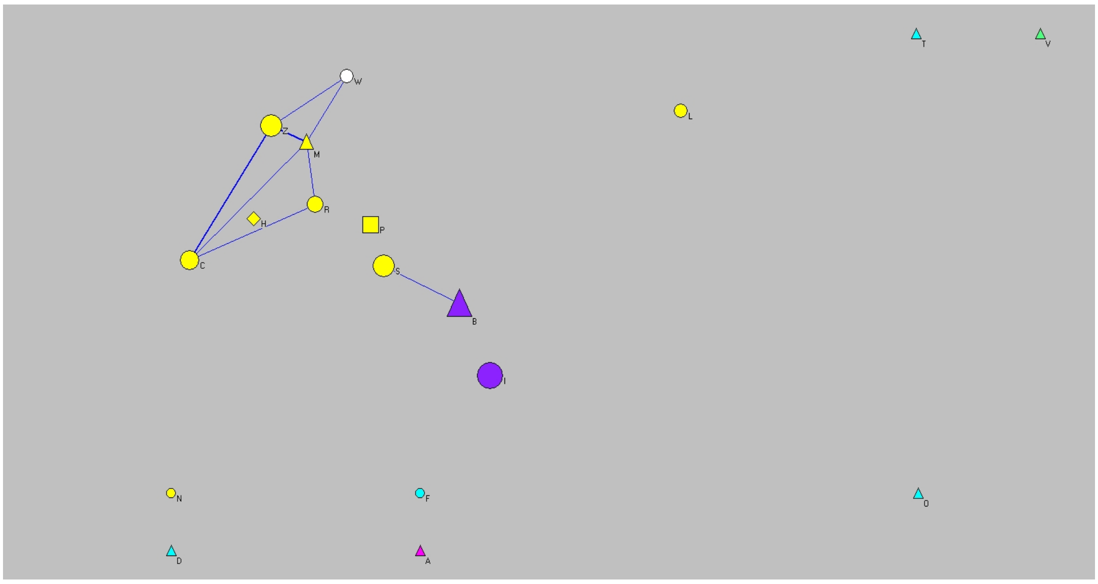


Figure 144 **–** *Collaborative DIEG network in 2005.*

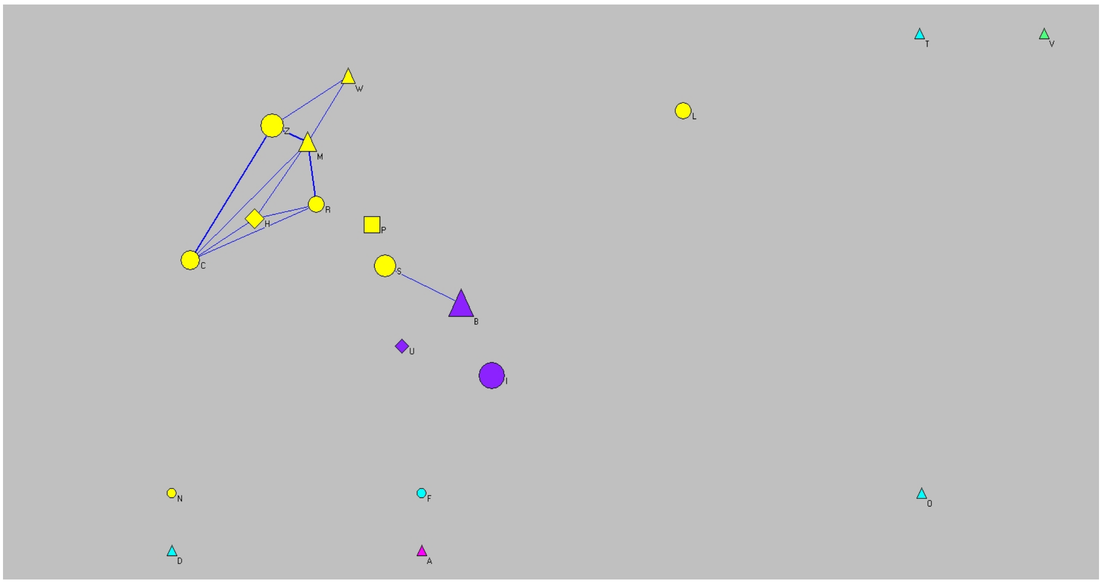


Figure 145 **–** *Collaborative DIEG network in 2006.*

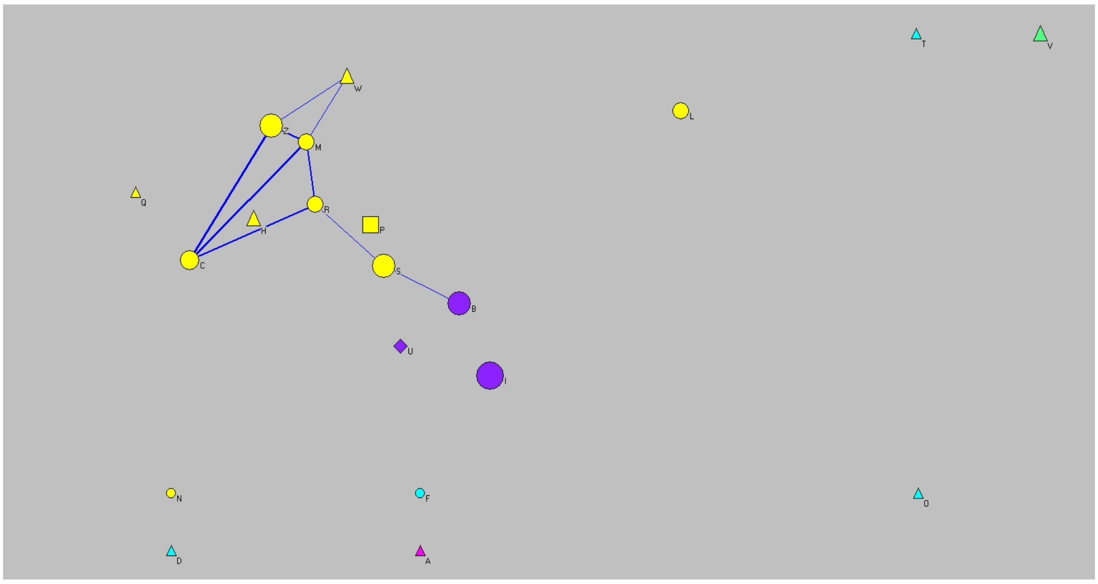


Figure 146 *– Collaborative DIEG network in 2007.*

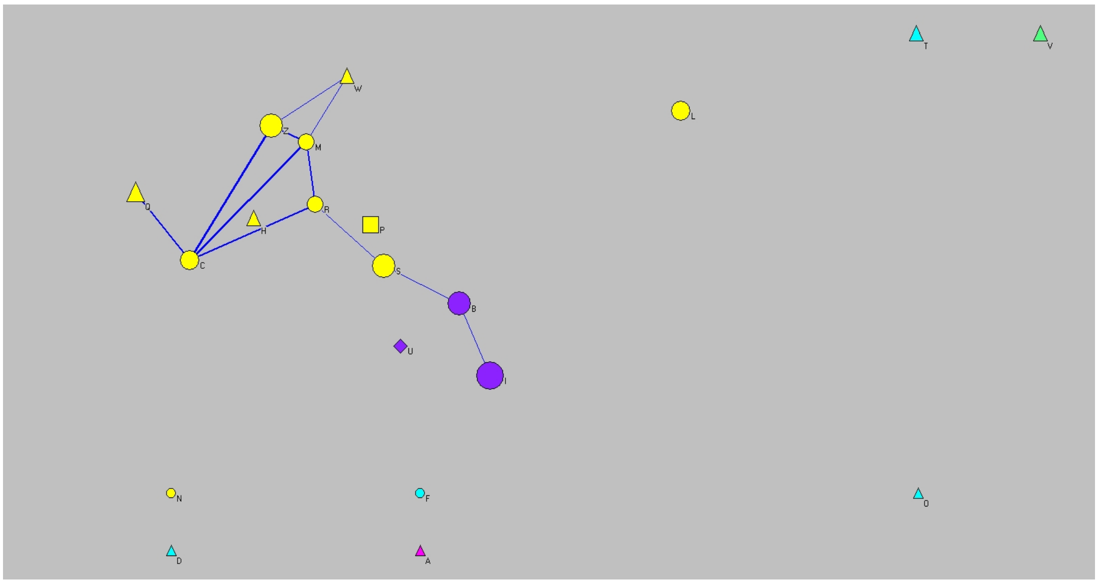


Figure 147 *– Collaborative DIEG network in 2008.*

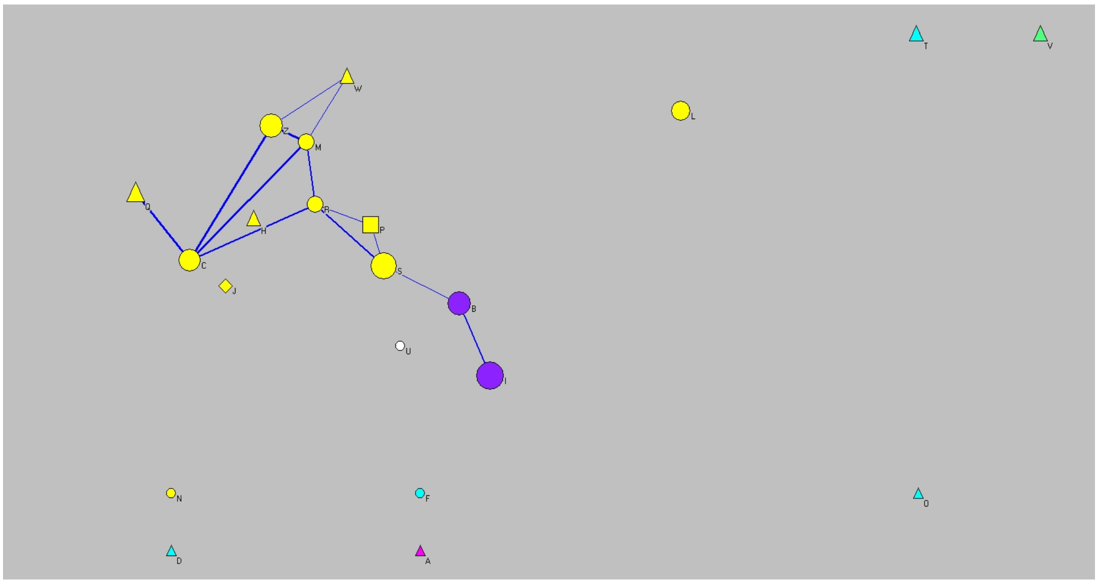


Figure 148 *– Collaborative DIEG network in 2009.*

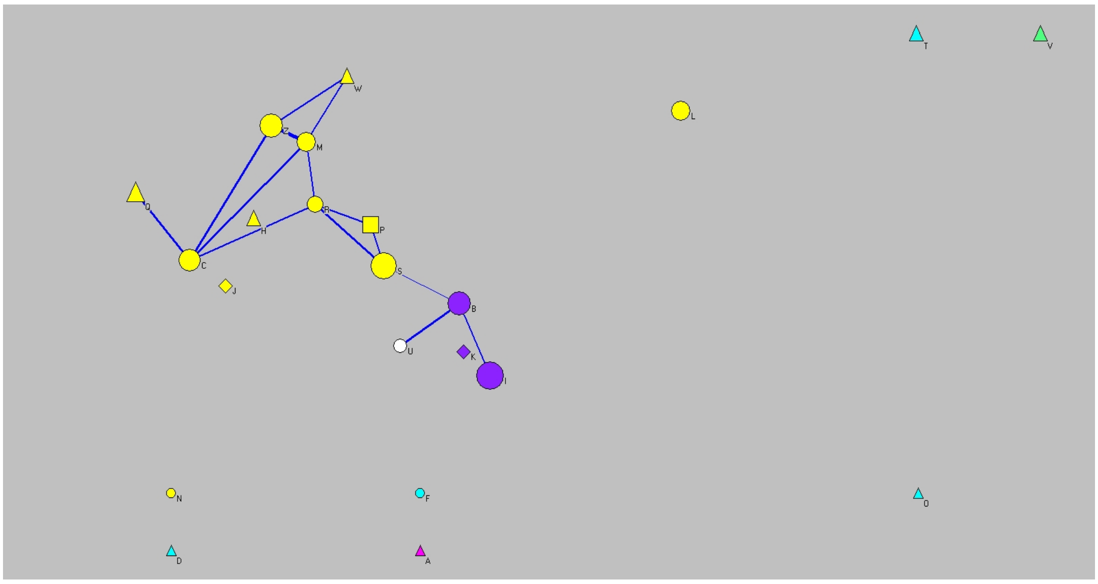


Figure 149 *– Collaborative DIEG network in 2010.*

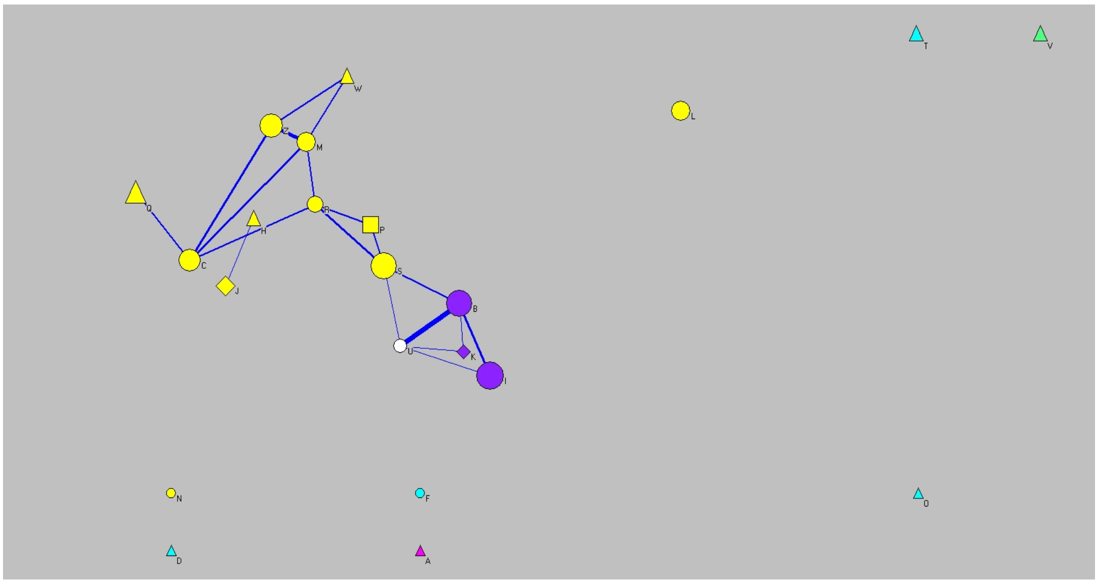


Figure 150 *– Collaborative DIEG network in 2011.*

The elimination of the external components determinates a reduction in the number of ties.

7.3.1 Measures of cohesion

In the calculation of the following measures the authors A, D, F, G, N, O have not been considered (see hypothesis 4 paragraph 6.3).

Tables 41-43 show measures calculated and these are displayed together the results for complete collaborative networks.

Measures	2001 complete collaborative	2001 DIEG collaborative	2002 complete collaborative	2002 DIEG collaborative	2003 complete collaborative	2003 DIEG collaborative	2004 complete collaborative	2004 DEIG collaborative
Total number of authors	12 (1)	7	14 (2)	7 (1)	9 (1)	9 (1)	9 (1)	9 (1)
Total number of ties	14	1	14	1	6	6	6	5
Density	0.18	0.05	0.12	0.05	0.13	0.13	0.13	0.11
Average degree	2.16	0.30	1.80	0.35	1.17	1.17	0.99	0.99

Table 39 *– Measures of complete and DIEG collaborative networks from 2001 to 2004.*

Measures	2005 complete collaborative	2005 DIEG collaborative	2006 complete collaborative	2006 DIEG collaborative	2007 complete collaborative	2007 DIEG collaborative	2008 complete collaborative	2008 DEIG collaborative
Total number of authors	17 (6)	9 (4)	15 (4)	10 (4)	17 (4)	11 (4)	15 (4)	11 (4)
Total number of ties	5	8	27	11	21	9	20	11
Density	0.11	0.10	0.10	0.12	0.12	0.08	0.09	0.10
Average degree	2.20	1.20	2.16	1.30	1.80	1.12	1.98	1.40

***Table 40** – Measures of complete and DIEG collaborative networks from 2005 to 2008.*

Measures	2009 complete collaborative	2009 DIEG collaborative	2010 complete collaborative	2010 DIEG collaborative	2011 complete collaborative	2011 DIEG collaborative
Total number of authors	24 (5)	11 (5)	19 (7)	11 (6)	24 (12)	11 (6)
Total number of ties	62	13	34	14	59	19
Density	0.15	0.10	0.10	0.10	0.09	0.14
Average degree	4.20	1.50	2.50	1.60	3.15	2.24

***Table 41** – Measures of complete and collaborative DIEG networks from 2009 to 2011.*

Comparing to complete collaborative networks, the values obtained are lower as some ties are not present. For this, the values of density and average degree for DIEG collaborative networks result lower than those of complete collaborative networks.

The values of inclusiveness, presented in Tables 44-46, show that over study period there is an increase of cohesion for DIEG collaborative network.

Measures	2001 complete collaborative	2001 DIEG collaborative	2002 complete collaborative	2002 DIEG collaborative	2003 complete collaborative	2003 DIEG collaborative	2004 complete collaborative	2004 DEIG collaborative
Total number of connected authors	9	2	13	2	4	4	19	4
Total number of not connected authors	4	5	3	6	6	6	4	6
Inclusiveness index	0.60	0.29	0.81	0.25	0.40	0.40	0.82	0.40

***Table 42** – The inclusiveness values for complete and DIEG collaborative networks from 2001 to 2004.*

Measures	2005 complete collaborative	2005 DIEG collaborative	2006 complete collaborative	2006 DIEG collaborative	2007 complete collaborative	2007 DIEG collaborative	2008 complete collaborative	2008 DEIG collaborative
Total number of connected authors	19	7	14	8	15	7	14	9
Total number of not connected authors	4	6	5	6	6	7	5	6
Inclusiveness index	0.82	0.54	0.73	0.57	0.71	0.47	0.73	0.60

***Table 43** – The inclusiveness values for complete and DIEG collaborative networks from 2005 to 2008.*

Measures	2009 complete collaborative	2009 DIEG collaborative	2010 complete collaborative	2010 DIEG collaborative	2011 complete collaborative	2011 DIEG collaborative
Total number of connected authors	24	10	21	11	34	14
Total number of not connected authors	5	6	5	5	2	3
Inclusiveness index	0.82	0.62	0.80	0.64	0.94	0.82

***Table 44** – The inclusiveness values for complete and DIEG collaborative networks from 2009 to 2011.*

The results indicate that the DIEG collaborative networks have a number of ties less than those of the complete networks. This means that the co-writing of papers is few diffuse among members of DIEG. However, over time the aggregation of network increases but its value remains very low.

7.3.2 Dynamic analysis

The last step of analysis consists into identify the mechanisms that drive network evolution. Also in this case, the isolated authors A, D, F, G, N, O have not been considered.

The effects to include in the models are the same of those chosen for the DIEG co-authorship networks because the network's typology and characteristics are the same. Starting by a first simple model (that contains fewer, also only one, effects), successive, others effects have progressively been added until to all selected effects are included. Thus, *balance,* *disciplinary sector similarity,* and *career ego x career alter* effects have been considered.

Model 1: only constant rate parameters corresponding to each transition have been included.

Model 2: *balance* has been added.

Model 3: *disciplinary sector similarity* parameter has been inserted.

Model 4: *career ego x career alter* has been added.

Besides, the Pairwise conjunctive model has been chosen.

In Table 47, the annual changes between subsequent observations are indicated.

Changes		2001-02	2002-03	2003-04	2004-05	2005-06	2006-07	2007-08	2008-09	2009-10	2010-11
Network											
	Joined	1	2	0	3	1	1	0	1	1	0
	Left	0	0	0	0	0	0	0	0	0	0
Ties											
	0 -> 1	0	5	0	3	3	2	3	2	1	5
	1 -> 1	1	1	5	5	5	5	6	9	11	12
	1 -> 0	0	0	1	0	3	3	1	0	0	0

Table 45 **–** *Changes in DIEG collaborative networks from 2001 to 2011.*

In all periods of observation the number of researchers that left the network is zero. The DIEG collaborative networks are characterized by researchers that join, and that create or eliminate or maintain their ties.

The four models, indicated before, have been sequentially run by SIENA, and the results obtained are summarized in Table 48.

Effects	Model 1	Model 2	Model 3	Model 4
Rate parameters				
$r_{2001\text{-}2006}$	0.46	0.49	0.48	0.44
$r_{2006\text{-}2011}$	1.57	1.52	1.44	1.45
Networks measures				
Balance		- 4.06	5.64	5.65
Authors characteristics				
Disciplinary sector similarity			5.91	5.91
Career ego x career alter				0.01

***Table 46** – Models predicting DIEG collaborative networks.*

In the first model, the *rate parameters* indicate that the change rate increases from first interval of time to the next. In fact, over observations the number of ties increases as researchers have more opportunities to change their ties.

The effect of *balance* is a positive effect and its values in all models are high. This indicates that researchers tend to form ties with whom they share other ties.

Disciplinary sector similarity, added to start from model 3, is the effect that weighs more than other effects considered. This means that the authors tend to create ties with other authors that work in the same disciplinary sector. This result has been underlined from observations.

Finally, the *level of carrier ego x alter* is used to test if the tendency of authors with a higher level of carrier to prefer or not the ties to others who likewise have a relatively high professional level. The corresponding value is positive though it is very low; this means that a little number of ties is created among authors with the similar carrier level.

8 Conclusions

The dynamic social networks are networks composed by social relations among actors that take in account changes of these relations over time: in a certain interval of time actors can decide to create, not create or eliminate their ties with others. Among large variety of possible changing networks, research network emerging from collaboration among researchers represents an interesting example to study.

Raffaella Cicala.

8.1 Developed activities

The dynamic social networks are networks composed by social relations among actors that take into account changes of these relations over time: over a certain interval of time the actors can decide to change their system of ties.

Among large variety of possible dynamic networks, research networks emerging from scientific collaborations have been chosen as study objective. They represent a typical affiliation network, in which a link between two researchers corresponds to a scientific collaboration between them. The great diffusion of digital libraries that allowed a wide availability of information on scientific production has encouraged the construction of research networks as *co-authorship* networks that are considered, also today, the most common and well-documented form of scientific collaboration. In reality, the production of a paper it is not always an index of collaboration.

A review of the literature in the field of collaboration in research, its interpretation in terms of social network, and linkage between co-authorship and collaboration has been analyzed.

Given dynamic nature of research networks, the literature concerning their changing has been revisited.

The identification of collaborative network existing within DIEG, a department of University Federico II of Naples, and its evolution along a period of time represent the case study developed. Two different definitions of scientific collaboration have been compared: the first linked to the production of papers that actors have in common (co-authorship networks), the second taking into account the continuity of scientific production (collaborative networks). In order to verify the main differences that these two interpretations determine on characteristics of co-authorship and collaborative networks, static and dynamic analyses have been applied.

8.2 Summarizing results

With reference to co-authorship networks, information that are used are: who publishes a paper and with who, and when they have published their papers. For collaborative networks, another information is necessary: the time-distances between successive publications.

Researchers have been characterized by several attributes, such as home institution, disciplinary sector, H-index, and the level of career.

The analysis on networks over study period has been conducted in order to understand: the reasons for which authors are mutually related; the reasons for which authors produce papers with other colleagues of the department; the reasons for which they interact with researchers external to department; the degree of integration among researchers within department; finally, are there attributes of researchers that influence decisions?

	2001		2002		2003		2004	
Measures	co-authorship	collaborative	co-authorship	collaborative	co-authorship	collaborative	co-authorship	collaborative
Total number of authors	13	13	21	16	23	10	23	10
Total number of ties	15	15	28	15	31	6	31	5
Density	0.19	0.19	0.13	0.12	0.12	0.13	0.12	0.11
Average degree	2.28	2.28	2.60	1.80	2.64	1.17	2.64	0.99
Clustering coefficient	0.95	0.95	0.88	0.93	0.89	1.00	0.89	1.00
Transitivity	0.73	0.73	0.51	0.75	0.46	1.00	0.44	0.50

***Table 47** – Measures of co-authorship and collaborative networks without isolated authors from 2001 to 2004.*

	2005		2006		2007		2008	
Measures	co-authorship	collaborative	co-authorship	collaborative	co-authorship	collaborative	co-authorship	collaborative
Total number of authors	36	23	39	19	43	21	44	19
Total number of ties	53	27	62	21	68	20	72	19
Density	0.08	0.10	0.08	0.12	0.07	0.09	0.07	0.11
Average degree	2.80	2.20	3.15	2.16	3.16	1.80	3.27	1.98
Clustering coefficient	0.89	0.91	0.84	0.82	0.83	0.79	0.84	0.62
Transitivity	0.37	0.73	0.33	0.43	0.30	0.40	0.27	0.26

***Table 48** – Measures of co-authorship and collaborative networks without isolated authors from 2005 to 2008.*

	2009		2010		2011	
Measures	co-authorship	collaborative	co-authorship	collaborative	co-authorship	collaborative
Total number of authors	55	29	59	26	71	36
Total number of ties	115	62	125	34	157	59
Density	0.07	0.15	0.07	0.10	0.06	0.09
Average degree	4.18	4.20	4.23	2.50	4.42	3.15
Clustering coefficient	0.87	0.86	0.87	0.73	0.86	0.82
Transitivity	0.35	0.26	0.31	0.27	0.26	0.30

Table 49 *– Measures of co-authorship and collaborative networks without isolated authors from 2009 to 2011.*

The static analysis has shown that co-authorship and collaborative networks are characterized by dissimilar cohesion (Tables 49-51).

In co-authorship, values have indicated a good cohesion over time. In other words, co-authorship networks are characterized by a good communication among researchers.

All values corresponding to collaborative networks have not monotone trends as external researchers, and consequently papers are not *cumulates*; they can operate in one observation and disappear in the successive observation. This result has shown that a large part of co-authorship ties is not durative representing occasional collaborations.

In order to understand mechanisms that drive the researchers' choices, networks have been entered in SIENA.

In particular, on the basis of characteristics of networks and on attributes of researchers, the choice of effects to include in models has been conducted. Selected effects have been added in progressive way.

Longitudinal analysis has highlighted differences for two networks (Table 52): for co-authorship, the tendency to create new ties with whom shares co-authors; for collaborative, the decisions to create ties are influenced by H-index so some authors tend to link with others that have similar H-index.

	Model 1		Model 2		Model 3		Model 4	
Measures	co-authorship	collaborative	co-authorship	collaborative	co-authorship	collaborative	co-authorship	collaborative
Rate parameter								
$r_{2001\text{-}2006}$	0.12	2.54	0.28	2.56	0.36	2.57	0.35	2.21
$r_{2006\text{-}2011}$	0.20	6.54	0.26	6.84	2.31	6.56	2.31	5.56
Network measures								
Outdegree			0.67	- 3.22	1.15	- 3.12	- 0.39	- 3.41
Balance			14.99	- 18.87	22.19	- 19.68	20.08	- 20.99
Author characteristics								
Same home institution					- 0.99	- 0.17	- 1.17	- 0.24
H-index similarity							2.20	5.56

Table 50 ***– Dynamic results for co-authorship and collaborative networks.***

The last part of analysis has been conducted to investigate on the degree of aggregation of DIEG's people, so for each observation new networks have been built in which external authors have been eliminated. The same statistics for complete networks have been then calculated (Tables 53-55).

	2001		2002		2003		2004	
Measures	DIEG co-authorship	DIEG collaborative	DIEG co-authorship	DIEG collaborative	DIEG co-authorship	DIEG collaborative	DIEG co-authorship	DEIG collaborative
Total number of authors	7	7	7 (1)	7 (1)	9 (1)	9 (1)	9 (1)	9 (1)
Total number of ties	1	1	1	1	6	6	6	5
Density	0.05	0.05	0.04	0.04	0.13	0.13	0.13	0.11
Average degree	0.30	0.30	0.28	0.28	1.17	1.17	1.17	0.99

Table 51 ***– Measures of DIEG co-authorship and collaborative networks without isolated authors from 2001 to 2004.***

	2005		2006		2007		2008	
Measures	DIEG co-authorship	DIEG collaborative	DIEG co-authorship	DIEG collaborative	DIEG co-authorship	DIEG collaborative	DIEG co-authorship	DEIG collaborative
Total number of authors	9 (4)	9 (4)	10 (4)	10 (4)	11 (4)	11 (4)	11 (4)	11 (4)
Total number of ties	9	8	12	11	13	9	15	11
Density	0.16	0.10	0.21	0.12	0.12	0.08	0.14	0.10
Average degree	1.92	1.20	2.73	1.30	1.68	1.12	1.96	1.40

Table 52 ***– Measures of DIEG co-authorship and collaborative networks without isolated authors from 2005 to 2008.***

Measures	2009 DIEG co-authorship	2009 DIEG collaborative	2010 DIEG co-authorship	2010 DIEG collaborative	2011 DIEG co-authorship	2011 DIEG collaborative
Total number of authors	11 (5)	11 (5)	11 (6)	11 (6)	11 (6)	11 (6)
Total number of ties	17	13	18	14	23	19
Density	0.14	0.10	0.13	0.10	0.17	0.14
Average degree	2.10	1.50	2.08	1.60	2.72	2.24

***Table 53** – Measures of DIEG co-authorship and collaborative networks without isolated authors from 2009 to 2011.*

For dynamic analysis, given additional information for each author, different effects have been taken into account. In fact, using as effects disciplinary sectors and academic roles of authors, the investigation has been concentrated on:

- To link with whom has higher level of career;
- To link with who works in the same disciplinary sector.

So, the models that have been entered in SIENA are different respect to the previous case.

In particular, the results obtained have shown that over study period DIEG co-authorship and collaborative networks are driven by same mechanisms; in particular, researchers tend to publish with others that operate to the same disciplinary sector. There are no particular differences in behavior with respect to the role of scientific collaboration and level of carrier. At last, there are subjects that performing the role of intermediaries writing one and more papers with both members belonging to DIEG both with external members.

Measures	Model 1 DIEG co-authorship	Model 1 DIEG Collaborative	Model 2 DIEG co-authorship	Model 2 DIEG collaborative	Model 3 DIEG co-authorship	Model 3 DIEG collaborative	Model 4 DIEG co-authorship	Model 4 DIEG collaborative
Rate parameter								
$r_{2001-2006}$	0.80	0.46	0.77	0.49	0.73	0.48	0.70	0.44
$r_{2006-2011}$	0.90	1.57	0.82	1.52	0.74	1.44	0.73	1.45
Network measures								
Balance			2.17	- 4.06	3.60	5.64	3.56	5.65
Author characteristics								
Disciplinary sector similarity					13.31	5.91	13.77	5.91
Career ego x career alter							0.04	0.01

Table 54 ***–*** *Dynamic results for DIEG co-authorship and collaborative networks.*

Thus, the results corresponding to model 4 underline that collaborations are lasting within DIEG.

8.3 Final remarks

The results found for two kinds of network are representative of two different phenomena.

In particular, the static analysis showing differences on structures demonstrates that speaking of co-authorship it is not the same of speaking of scientific collaboration. Also, dynamic analysis demonstrates that co-authorship and collaborative networks are driving by two different effects.

In conclusion, even co-authorship networks were indeed no more than partial indicator of scientific collaboration, studying of this phenomenon allows a deep insight into measurable interaction between communications.

8.4 Future developments

Possible directions of future research have been identified.

Once to represent the mechanisms that drive the networks, a model has been built able to make, by chance, predictions about the configuration that the network could take in future time.

For instance, in the network shown in Figure 157, the creation of new collaboration is influenced in a positive way by membership to the same research groups. The researchers *k* and *i* belong to the same research group and *i* has the opportunity to decide a collaboration choosing between *j* and *k.*

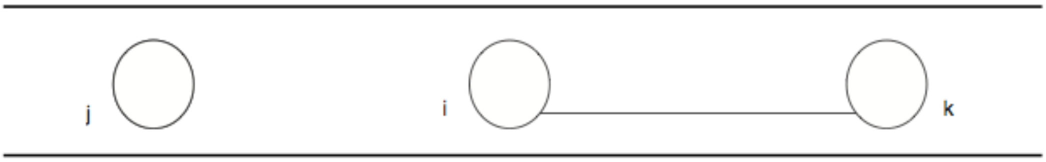


Figure 151 – *The initial network.*

The researcher *i* will decide to create collaboration with *k* that belongs to the same research group rather than with *j* that belongs to a different research group. So it might be that as long as it is still possible to collaborate with potential partners that are member of the own group, *i* chooses to collaborate with them rather with non members.

After *i*'s choice, the network will be the following:

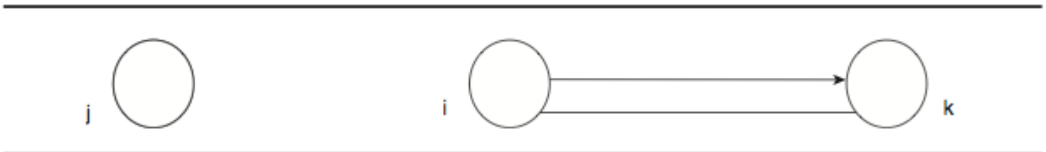


Figure 152 – *Configuration of the network in future observation.*

Besides, in line as a part of literature on research network, an interesting development could be to investigate on effects linked to scientific collaboration on productivity of researchers.

Finally, the unit of analysis focuses on scientific production within a single department and this involves that for external components, ties that they have with other external people have not been considered; so, the analysis could be to extend including ties that external ties that external components have with other external people. This can be increased the size of unit of study. Besides, external people have not been characterized by their sector disciplinary.

Appendix A

The origins of Graph Theory

One of the first and most well known applications of graph theory was built in 1736 by L. Euler, to solve so-called the problem of the bridges of Konigsberg.

The river Pregel (in Russian Pregolja), coming from the east across Lithuania and enters the Russian enclave, sandwiched between Lithuania and Poland, formerly known as East Prussia, which in Kaliningrad (former Königsberg) and its center in the suburb Pilau its main port. The two branches of the river (Staraya and Novaya Pregolja Pregolja) crossed (and cross) the city of Königsberg (forming an island in the middle of it) and then become a single course, result in the Vistula Lagoon to the west of the city and bring their waters, through it, to the Baltic Sea.

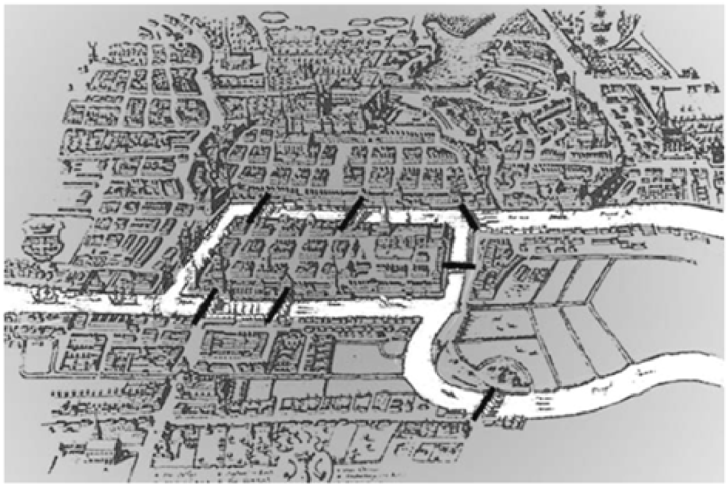


Figure 153 *– The river Pregel (Newman et al, 2006).*

The territory of Königsberg covered four main macro-zones, the left bank, the central island and the tongue of land between the two branches of the Pregel confluent, connected to each other at the time of Euler, by seven bridges.

Figure 154 *– The representation of river Pregel as a graph (Newman et al, 2006).*

Euler described the problem using a graph (Figure 160) in which nodes indicated the two shores and islands, and arcs represented the bridges; he showed that this problem had no solutions. Also Euler showed that, in order to admit a solution, each node would have to have a greater number of arcs connected to it.

The problem of search on a graph of a path that uses all arcs one and once it is known in literature as the search for a *Eulerian path*.

Over time, graph theory has found numerous applications in various fields of science and technology, ranging from chemistry to physics, engineering, electrical and the electronics engineering, the company organization, but also anthropology, social psychology, communication, on-line business, and many others.

Appendix B

Network data

Many kinds of data sets exist but they can be classified according to two their components: *cross-sectional*, which refers to the number of units of analysis included in the data set, and *temporal*, which refers to the number of time periods included in the data set.

In order to propose a class of actor-oriented statistical model, Van De Bunt and Snijders studied a friendship networks; they considered a group of university freshmen that did not know each other at the first measurement, in seven points in time during 1994 and 1995.

So, the unit of analysis, that is the number of students under observations, is the cross-sectional component, while the points of observation period represent temporal component.

According this distinction, it is possible to identify two important types of data set:

- *Cross-sectional data* consists in a collection of units of analysis for a single moment of time, a time interval, or a period. In Figure 161, an example of cross-sectional data is showed (Castilla E. J., 2007). The variable *y* is measured in time T_4 for five different cases.

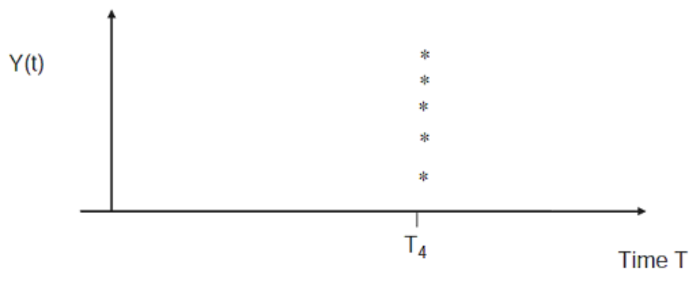


Figure 155 – *An example of cross sectional data for five cases (Castilla E. J., 2007).*

- *Longitudinal data* are data resulting from the observations of subjects that are measured repeatedly over time, for at least two or more distinct time. In Figure 162, an example of longitudinal data is depicted (Castilla E. J., 2007), in which two cases are observed in the five discrete points of time.

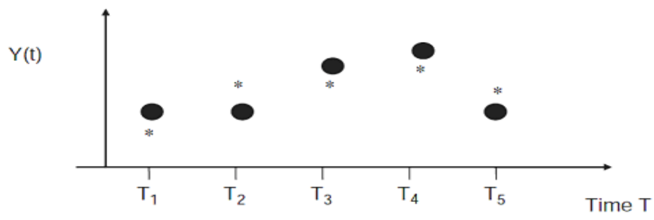


Figure 156 – *An example of longitudinal data with five time points of times (Castilla E. J., 2007).*

Longitudinal network data can be collected using different methods, like questionnaire, interview, and so on.

Appendix C

Effects of actor-oriented model

The potential effects s_{ik} for generic actor *i* in the objective function are the following:

1. *density effect*, defined by the out-degree
$s_{i1}(x) = x_{i+} = \sum_j x_{ij}$;

2. *reciprocity effect*, defined by the number of reciprocated ties
$s_{i2}(x) = \sum_j x_{ij}\, x_{ji}$;

3. *transitivity effect*, defined by the number of transitive patterns in i's relations (ordered pairs of actors (j, h) to both of whom i is tied, while also j is tied to h),
$s_{i3}(x) = \sum_{j,h} x_{ij}\, x_{ih}\, x_{jh}$;

4. *balance*, defined by the similarity between the outgoing ties of actor i and the outgoing ties of the other actors j to whom i is tied,

$$s_{i4}(x) = \sum_{j=1}^{n} x_{ij} \sum_{\substack{h=1 \\ h \neq i,j}}^{n} (b_0 - \mid x_{ih} - x_{jh} \mid) \ ,$$

where b_0 is a constant included to reduce the correlation between this effect and the density effect, defined by

$$b_0 = \frac{1}{(M-1)n(n-1)(n-2)} \sum_{m=1}^{M-1} \sum_{i,j=1}^{n} \sum_{\substack{h=1 \\ h \neq i,j}}^{n} \mid x_{ih}(t_m) - x_{jh}(t_m) \mid \ .$$

5. *number of distances two effect*, or indirect relations effect, defined by the number of actors to whom i is indirectly tied (through one intermediary, i.e., at sociometric distance 2),
$s_{i5}(x) = \#\{j \mid x_{ij} = 0,\, \max_h(x_{ih}\, x_{hj}) > 0\}$;

6. *popularity effect*, defined by $1/n$ times the sum of the in-degrees of the others to whom i is tied,
$s_{i6}(x) = \frac{1}{n} \sum_j x_{ij}\, x_{+j} = \frac{1}{n} \sum_j x_{ij} \sum_h x_{hj}$

7. *activity effect*, defined by $1/n$ times the sum of the out-degrees of the others to whom i is tied,
$s_{i7}(x) = \frac{1}{n} \sum_j x_{ij}\, x_{j+} = \frac{1}{n} \sum_j x_{ij} \sum_h x_{jh}$

8. *out-degree up to c*, where c is some constant, defined by
$s_{i8}(x) = \max(x_{i+}\, , c)$;

9. *square root out-degree - c× o.d.*, where c is some constant, defined by
$s_{i9}(x) = \sqrt{x_{i+}} - c x_{i+}$,
where c is chosen to diminish the collinearity between this and the density effect;

10. *squared (out-degree - c)*, where c is some constant, defined by
$s_{i10}(x) = (x_{i+} - c)^2$,
where c is chosen to diminish the collinearity between this and the density effect.

11. *number of 3-cycles*,
$s_{i11}(x) = \sum_{j,h} x_{ij}\, x_{jh}\, x_{hi}$;

Appendix D

DIEG (Business and Management Engineering Department) profile

The DIEG is formed in March 2000 by two groups of Full and Associate professors and Assistant professors with old experience in research and teaching; the first group came from the Department of Computer Science and Systems (Economic-Management and Operational Research areas) and the second one (legal and valuation areas) derived from the Institute of Law topics.

In order to focus attention on economic and financial topics of engineering, on the use of quantitative methods for decision support, on organizational, management (both at the operational and strategic) and the legal aspects DIEG was born.

Now, the scientific-disciplinary sectors in the DIEG are as follows:

- ICAR/22 (H15X) – Valuation: the scientific-disciplinary contents concern the theoretical assumptions and methodologies on estimates of costs, prices, rates of return of property, investments, equipment, businesses as well as for determination of compensation, duties, fees, with the purpose of formulation of value judgments and of cost-effectiveness in the civil, territorial, industrial. The disciplinary interests extend to issues of environmental economics and, at specifically methodological, to analysis of the feasibility of projects and plans and the evaluation of their economic and extra-economic approaches.
- ING-IND/17 – Industrial Engineering: the sector concerns the methods and general principles that govern the planning, design, construction and management of industrial plants (or production systems). This sector includes the following main areas: analysis and design of industrial plants, including the feasibility study, the choice of the location and the economic evaluation of the initiative, analysis and design of general services facility, including methods of technical-economic optimization, analysis and design of processes and production technologies; ergonomic and safety

analysis and design in production systems, management of production systems, including the management of the quality and maintenance, logistics of industrial plants, including management and material handling, automation of production systems, including analysis of cost-effectiveness of integrated and flexible systems and industrial instrumentation for automatic control of the process.

- ING-IND/35 (I27X) – Economic-Management Engineer: the sector includes the skills for the integration of design, economic, organizational and management aspects in the engineering field. In it one can identify two major thematic strands. The first strand is addressed to the integration of economic and management knowledge oriented to design, highlighting the economic implications of the projects, the relationships between design choices and business performance, the relationship between the design and implementation of innovations, the financing arrangements of the projects, the connection with the context in which the firm operates. The second strand explores the different skills that characterize the engineering management, integrating, for each of them, the economic, organizational and technological skills through an approach in which the following components of the engineering culture co-exist: the finalization of design, optics based on the theory of systems and control, the emphasis on modeling and quantitative methods, the integration between theoretical models and empirical verification.
- IUS/01 (N01X) – Private Law: the sector includes studies related to the system of private law which emerges from the rules of the Civil Code and the laws complementary to it. The studies relate also to the civil law, the rights of individuals, the family, the right to information technology and bio-law.
- MAT/09 (A04B) – Operational Research: the sector concerns the decision-making processes in organizations, models and methods to predict the behavior of them, in particular those related to the growth of their complexity, to evaluate the consequences of certain decisions and to identify the decisions that optimize their

performance. The basic methods include theory and optimization algorithms, graph theory and network flows, game theory and decision-making. The issues under study include systems of production, transport, distribution and logistics support for goods and services, planning, organization and management of activities, projects and systems, in all the different phases that characterize the decision-making process: problem definition, its mathematical formulation, formulation of constraints, objectives and action alternatives, development of solution algorithms, evaluation, implementation and certification procedures and found solutions. Teaching skills also cover all the institutional aspects of basic mathematics.

References

Abramo, G., D'Angela, C.A., & Solazzi, M., (2011). The relationship between scientist's research performance and the degree of internationalization of their research. Scientometrics, 86: 629–643.

Backstrom, L., Huttenlocher, D., Kleinberg J., & Lan, X., (2006). *Group Formation in Large Social Networks: Membership, Growth, and Evolution.* Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining ACM Press, New York, NY, USA: 44–54.

Barabasi, A.L., & Albert, R., (1999) *Emergence of scaling in random networks*. Science, 286: 509–512.

Barabasi, A.L., Jeong, H., Neda, Z., Ravasz, E., Schubert, A., & Vicsek, T., (2002). *Evolution of the social network of scientific collaborations.* Physica A, 311: 590–614.

Barnes, J. A., (1954). *Class and committee in a Norwegian Island parish.* Human Relations, 7: 39–58.

Beaver D. D., (2001). *Reflections on Scientific Collaboration (and its Study): Past, Present, and Future.* Scientometrics, 52: 365–377.

Benson, J., (1987). *Working More Creatively with Groups*. Tavistock.

Boer, P., Huisman, M., Snijders, T. A.B., Steglich, C. E.G., Wichers, L. H.y., & Zeggelink, E. P.H., (2006). *An open software system for the advanced statistical analysis of social networks.* User Manual version StOCNET 1.7.

Borgatti, S.P., Everett, M.G., & Freeman, L.C., (2007). *UCINET 6.0 for Windows: Software for social network analysis.* Harvard: Analytic Technologies.

Borgatti, S. P., Mehra, A., Brass, D., & Labianca, G ., (2009). *Network Analysis in the Social Sciences.* Science, 323: 892–895.

Borgatti, S. P., & Foster, P. C., (2003). *The Network Paradigm in Organizational Research: A Review and Typology.* Journal of Management, 29: 991–1013.

Brandes, U., Indlekofer, N., & Mader, M., (2011). *Visualization Methods for Longitudinal Social Networks and Stochastic Actor-Oriented Modeling.* Submitted to Social Networks.

Brandes, U., & Nick, B., (2011). *Asymmetric Relations in Longitudinal Social Networks*. Iee transactions on visualization and computer graphics, 17: 2283–2290.

Calero, C., Buter, R., Cabello Valdés, C., & Noyons, E., (2006). *How to identify research groups using publication analysis: an example in the field of nanotechnology.* Scientometrics, 66: 365–376.

Carley, K. M., (2005). *Dynamic Network Analysis for Counter-Terrorism.* Carnegie Mellon University, Pittsburgh, Pennsylvania.

Carley, K. M., (2006). *A dynamic network approach to the assessment of terrorist groups and the impact of alternative courses of action.* Visualizing Network Information Meeting Proc., RTO-MP-IST-063.

Carley, K. M., Diesner, J., Reminga, J., & Tsvetovat, M., (2007). *Toward an interoperable dynamic network analysis toolkit*. Decision Support Systems, 43: 1324–1347.

Cartwright, D., & Zander, A., (1968). *Group dynamics: Research and TheoryHarper and Row.* New York, NY.

Castilla, E. J., (2007). *Dynamic Analysis in the Social Sciences*. Oxford, U. K.: Elsevier Academic Press.

Chiesi, A., (2007). *L'analisi dei reticoli*. Metodologia delle scienze umane. Milano: Franco Angeli.

Coulon, F., (2005). *The use of Social Network Analysis in Innovation Research: A literature review.* Lund University, Sweden.

Crossley, N., Prell, C. & Scott, J., (2009). *Social Network Analysis: Introduction to Special Edition*. Methodology Innovation, 4: 1–7.

De Nooy, W., Mrvar, A. & Batagelj,. V., (2005). *Exploratory Social Network Analysis with Pajek*. Cambridge University Press.

Diesner, J., Frantz, T.L. & Carley, K.M., (2005) *Communication networks from the Enron email corpus*. Journal of Computational & Mathematical Organization Theory, 11: 201–228.

Doreian, P., & Stokman, F.N. ,(1997). *Evolution of Social Networks.* Amsterdam etc.: Gordon and Breach.

Ehrlich, K., & Carboni, I., (2005). *Inside Social Network Analysis*. Copyright 2005, IBM. All rights reserved. 05–10, IBM, Amherst, MA.

Federico, P., Aigner, W., Miksch, S., Windhager, F., & Zenk, L., (2011). *A visual analytics approach to dynamic social networks.* In Proceedings of the 11th International Conference on Knowledge Management and Knowledge Technologies (i-KNOW), Special Track on Theory and Applications of Visual Analytics (TAVA). i-KNOW ’11, 47: 1–8, New York, NY, USA, 2011. ACM.

Forsyth, D. R., (2009). *Group dynamics* (4th ed. Belmont). CA: Wadsworth.

Fredericks, K. A. & Durland, M. M., (2005). *The Historical Evolution and Basic Concepts of Social Network Analysis.* New Direction for Evaluation, 107: 15–23.

Freeman, L.C. & Freeman, S.C., (1980). *A semi-visible college: Structural effects of seven months of EIES participation by a social networks community*. Electronic Communication: Technology and Impacts (M. M. Henderson and M. J. McNaughton, eds.). AAAS Symposium, 52: 77–85. American Association for Advancement of Science, Washington, D.C.

Garcia, S., (2013). *Identifying chance influencers through social network analysis*. Internet http://www.philosophyib.com/3/wholebrain/identifying-change-influencers#!prettyPhoto.

Gibson, D., (2005). *Taking turns and talking ties*. American Journal of Sociology, 110: 1561–97.

Glänzel, W., & Schubert, A., (2005). *Analyzing Scientific Networks through Co-authorship.* Handbook of Quantitative Science and Technology Research, pp.257–276.

Gordon, M.D., (1980). *A critical Reassessment of Inferred Relations Between Multiple Authorship, Scientific Collaboration, the Production of Papers and their Acceptance for Publication*. Scientometrics 2, pp. 193–210.

Gould, R., (1995). *Insurgent Identities*. Chicago, IL: University of Chicago, Press.

Hanneman, R., & Riddle, M., (2005). *Introduction to social network methods.* Chapter 10: Centrality and power. http://faculty.ucr.edu/~hanneman/nettext/C4_netdraw.html

Homans, G., (1950). *The Human Group.* New York, NY: Harcourt, Brace.

Hossian, L., & Fazio, D., (2009). *The social networks of collaborative process.* Journal of High Technology Management Research, 20: 119–130.

Huisman, M., & van Duijn, M. A.J., (2005). *Software for Social Network Analysis*. In P.J. Carrington, J. Scott, and S. Wasserman (Eds.) Models and methods in social network analysis (pp. 270-316). New York: Cambridge University Press.

Huisman, M., & van Duijn, M. A.J., (2003). *StOCNET: Software for the statistical analysis for social networks.* Connections, 25: 7–26.

Izquierdo, Luis R. & Hanneman, Robert A., (2006). *Introduction to the formal analysis of social networks using mathematica.* University of California, Riverside.

Kamada, T., & Kawai, S., (1989). An algorithm for drawing general undirected graphs. Information Processing Letters, 31: 1–7.

Katerndahl, D., (2012). *Evolution of the research collaboration network in a productive department.* Journal of Evaluation in Clinical Practice, 18, 195–201.

Katz, J.S., & Martin, B.R., (1997). *What is research collaboration?*. Research Policy, 26: 1–18.

Katz, L., & Powell, J.H., (1955). *Measurement of the tendency toward reciprocation of choice.* Sociometry, 18: 659–665.

Katz, L., & Proctor, C.H, (1959). *The concept of configuration of interpersonal relations in a group as a time-dependent stochastic process*. Psychometrika, 24: 317–327.

Knoke, D., (1982). *The Spread of Municipal Reform: Temporal, Spatial, and Social Dynamics.* American Journal of Sociology, 87: 1314–1339.

Kronegger, L., Mali, F., Ferligoj, A. & Doreian,. P., (2012). *Collaboration structures in Slovenian scientific communities.* Scientometrics, 90: 631–647.

Kuhn, T. S., (1996). *The structure of scientific revolutions.* Chicago: University of Chicago Press.

Laudel, G., (2001b). *Collaboration, creativity and rewards: why and how scientists collaborate.* International Journal of Technology Management, 22.

Laudel, G., (2002). *What do we measure by co-authorships?.* Research Evaluation, 11: 3–15.

Lawani, S.M., (1986). *Some Bibliometric Correlates of Quality in Scientific Research.* Scientometrics, 9: 13–25.

Leinhardt, S., (1977). *Social Networks: A Developing Paradigm*. New York, NY: Academic Press.

Liao, C. H., & Yen, H. R., (2012). *Quantifying the degree of research collaboration: A comparative study of collaborative measures.* Journal of Infometrics, 6: 27–33.

Lotka, A.J., (1926). *The frequency Distribution of Scientific Productivity.* Journal of Washington Academy of Science, 16: 317–323.

Martino, F., & Spoto, A., (2006). *Social Network Analysis: A brief review and further perspectives in the study of Information Technology.* PsychNology Juornal, 4: 53–86.

McCartey, C., Jawitz, J. W., Hopkins, A., & Goldman, A., (2013). *Predicting author h-index using characteristics of the co-author network.* Scientometrics, 96: 467–483.

McCulloh, I., (2009). *Detecting Changes in a Dynamic Social Network.* Unpublished PhD thesis, Carnegie Mellon University.

McCulloh, I., & Carley, K., (2009). *Longitudinal Dynamic Network Analysis.* Tech. rep., Institute for Software Research, School of Computer Science, Carnegie Mellon University.

McPherson, M., Smith-Lovin, L., & Cook J. M., (2001). *Birds of a Feather: Homophily in Social Networks.* Annual review of Sociology, 27: 415–44.

Melin, G., & Persson, O., (1996). *Studying Research Collaboration Using Co-Authorship*. Scientrometrics, 36: 363–377.

Milgram, S., (1967). *The small world problem*. Pysichal Today, 2: 60–67.

Mills, T. M., (1967). *The Sociology of Small Groups.* Englewood Cliffs, N.J.: Prentice-Hall.

Moody, J., (2004). *The structure of a social science collaboration network: disciplinary cohesion from 1963 to 1999*. American Sociological Review, 69: 213–238.

Moody, J., McFarland, D., & Bender-deMoll, S., (2005). *Dynamic network visualization*. American Journal of Sociology 110, pp. 1206–1241.

Must, Ü., (2000). *Cooperation of Estonian research institutions: comparative study of co-authorship and co-partnership.* Proceedings of Second Berlin workshop on Scientometrics & Informetrics: Collaboration in Science and in Technology: 163–171, September 13, Berlin (Germany).

Nakai, M., & Ke, W., (2009). *Statistical Models for Longitudinal Data Analysis.* Applied Mathematical Sciences, 40: 1979–1989.

Newcomb, T.M., (1961). *The Acquaintance Process.* New York: Holt, Rinehart and Winston.

Newman, M.E., (2001). The structure of *Scientific collaboration networks*. Proceeding of the National Academy of Science, 98: 404–409.

Newman, M., Barabasi, A., & Watts, D. J., (2006). *The Structure and Dynamics of Networks.* Princeton University Press, 575.

Nordlie, P.G. , (1958). *A longitudinal study of interpersonal attraction in a natural group setting*. PhD Thesis, University of Michigan.

Pattinson, P., & Wasserman, S., (1995). *Constructing Algebraic Models for Local Social network Using Statistical Method*. Journal of Mathematical Psychology, 39: 57–72.

Perianes-Rodríguez, A., Olmeda-Gómez, C., & Moya-Anegón, F., (2010). Scientometrics, 82: 307–319.

Pravdić, N., & Oluić-Vuković, V. , (1986). *Dual approach to multiple authorship in the study of collaboration/scientific output relationship*. Scientometrics, 10: 259–280.

Prell, C., (2012). *Social Network Analysis.* Thousand Oaks, CA: Sage.

Robins, G., & Martina, M., (2007). *Advances in exponential random graph (p*) models.* Social Networks 29, pp. 169–72.

Rousseau, R., (2011). *Comments on the modified collaborative coefficient.* Scientometrics, 87: 171–174.

Sanil, A., Banks, D., & Carley, K.M. (1995). *Models for evolving fixed node networks: Model fitting and model testing.* Social Networks, 17: 65–81.

Savoia, L., (2007). *L'analisi della dinamica del network con SIENA*. In: Salvini, A., Analisi delle reti sociali. Teorie, metodi, applicazioni, Milano: Franco Angeli.

Scott, J., (1991). *Social network analysis*. London: Sage.

Scott, J., & Carrington, P.C., (2011). *The SAGE Handbook of Social Network Analysis.* Thousand Oaks, CA: Sage.

Snijders, T.A.B., (1996). *Stochastic actor-oriented dynamic network analysis.* Journal of Mathematical Sociology, 21: 149–172.

Snijders, T.A.B., (2001). *The statistical evaluation of social network dynamics*. Sociological Methodology, 31: 361–395.

Snijders, T. A.B., (2005). *Models for longitudinal network data.* Chapter 11 in Carrington, P.J., Scott, J., & Wasserman, S.. Models and Methods in Social Network Analysis. New York, NY: Cambridge University Press.

Snijders, T. A.B., Van de Bunt, G.G., & Steglich, C. E.G., (2010). *Introduction to stochastic actor-based models for network dynamics.* Social Networks, 32: 44–60.

Snijders, T. A.B. & Van de Bunt, G. G. (2011). *Statistical Methods for Network Dynamics.* Annual Review Sociological, 37: 131–153.

Snijders, T. A. B., & Van Duijn, M.A.J., (1997). *Simulation for Statistical Inference in Dynamic Network Models*. In Simulating Social Phenomena, (Ed. R. Conte, R. Hegselmann, and P. Tera) Berlin: Springer.

Snijders, T. A.B., Steglich, C. E.G., Schweinberger, M., & Huisman, K., (2008). *Manual for SIENA.* Groninger, Oxford: University of Groninger.

Sonnenwald, D. H., (2007). *Scientific Collaboration: A Synthesis of Challenges and Strategies.* Annual Review of Informational Science and Technology, 4: 643–681.

Stadtfeld, C., (2012*). Events in Social Networks A Stochastic Actor-oriented Framework for Dynamic Event Processes in Social Networks*. Karlsruhe: KIT Scientific Publishing.

Steglich, C., Snijders, T. A.B., & West, P., (2006). *Applying SIENA. An Illustrative Analysis of the Coevolution of Adolescents' Friendship Network, Taste in Music, and Alcohol Consumption*. Methodology, 2: 48–56.

Van de Bunt, G.G., & Groenewegen, P., (2007). *An Actor-Oriented Dynamic Network Approach: the Case of Interorganizational Network Evolution*. Organizational Research Methods, 10: 463–482.

Van de Bunt, G.G., Van Duijn, M. A.J., & Snijders, T. A.B., (1999). *Friendship networks through time: An actor-oriented statistical network model*. Computational and Mathematical Organization Theory, 5: 167–192.

Veenstra, R., and Dijkstra, J.K. (2011). *Transformations in Peer Networks.* Chapter 7 (pp. 135-154) in B. Laursen and W.A. Collins (eds.), Relationship Pathways: From Adolescence to Young Adulthood. New York: Sage.

Vignoli, M. (2008). *La struttura e l'evoluzione del network di ricercatori e l'impatto sulle performance.* Unpublished PhD thesis.

Wagner, C., (2005). *Six case studies of international collaboration in science.* Scientometrics, 62: 3–26.

Wasserman, S., & Faust, K., (1994). *Social Network Analysis: Methods and Applications.* Cambridge, England: Cambridge University Press.

Wattz, D.J., & Strogatz, S., (1998). *Collective dynamics of smallworld networks*. Nature, 393: 440–442.

Ziman, J., (1994). *Prometheus bound. Science in dynamic steady state.* Cambridge: Cambridge University Press.

Xi, Y., & Fangcheng, T., (2004). *Multiplex Multi-Core Pattern of Network Organizations: An Exploratory Study.* Computational & Mathematical Organization Theory, 10: 179–195.